Preface
The 2nd International Workshop on Knowledge Discovery in Healthcare Data (KDH)

Introduction

The notion of a learning healthcare system has been put forward to denote the translation of routinely collected data into knowledge that drives the continual improvement of medical care by seamlessly embedding learned best practices in the healthcare delivery process. This notion has been described in many forms, but each follows a similar cycle of assembling, analyzing and interpreting data from multiple sources (clinical records, guidelines, patient-provided data including wearables, omic data, etc..), followed by feeding the acquired knowledge back into clinical practice. This framework aims to provide personalised recommendations and decision support tools to aid both patients and care providers, to improve outcomes and personalise care.

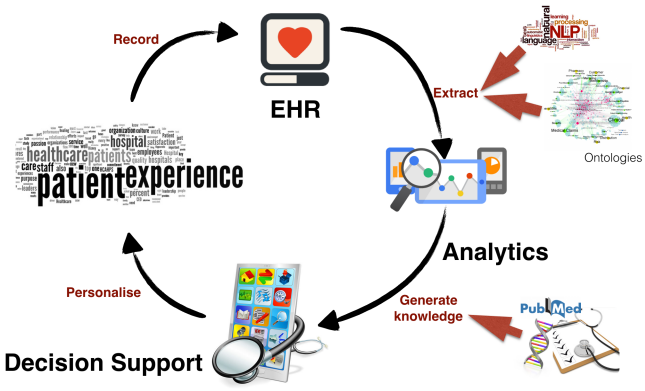

Figure 1: Aspects of knowledge discovery in healthcare

The idea of a learning healthcare system poses grand technical challenges in terms of: 1) data extraction, organisation and assembly of the large amounts structured and free-text data contained within the data sources, 2) near real time analytics and knowledge discovery from the large, temporal and uncertainty-ridden healthcare data and 3) the design of tools supporting clinical decision making as well as self management and care by patients in an autonomous and efficient manner, without jeopardising existing clinical workflows or the privacy of patient data. Therefore, the notion of the learning healthcare system encompasses research in prominent areas of Artificial Intelligence including language engineering, data mining, knowledge representation and reasoning, learning and autonomous systems.

The workshop received 13 submissions that were peer-reviewed by at least three reviewers each. After the review phase, 2 long papers and 5 short papers were accepted for presentation at the workshop. Among the accepted papers, the current trend of applying deep learning can be seen here as well, three papers use deep learning methods on health care data, while other methods used are: case-based reasoning, natural language processing or time series analysis.

Invited Speakers

1. **Daniel Sonntag, German Research Centre of Artificial Intelligence, Saarbruecken, Germany** s

   Bio: Daniel Sonntag (German Research Center for Artificial Intelligence, DFKI) is a Principal Researcher and Research Fellow. He has been teaching since 2010 at Saarland University and the Technical University of Kaiserslautern. His research interests include multimodal and mobile AI-based interfaces, natural language processing, dialogue systems, commonsense modelling, and semantic machine learning methods for cognitive computing and improved usability. This includes intelligent user interfaces (IUIs), multimodal multisensor interfaces for medical and health systems in particular, common-sense and (interactive) machine learning methods for human computer interfaces, knowledge discovery, information extraction, and cognitive modelling with ontologies.

   He has published over 120 scientific articles, and has been recipient of the German High Tech Champion Award in 2011 and the AAAI Recognition and IAAI Deployed Application Award in 2013. He is an editorial board member of the German Journal on Artificial Intelligence (KI). Currently, he leads both national and European projects from the Federal Ministry of Education and Research, the Federal Ministry for Economic Affairs and Energy, and Horizon 2020.

   Title: Multimodal Multisensor Interfaces for Medical and Health Systems

   Abstract: We discuss the trends of multimodal-multisensor interfaces of medical and health systems and emphasizes the theoretical foundations of multimodal

interfaces and systems in the healthcare domain, namely multimodal interaction, distributing multimodal processing into applications, and multisensory-multimodal facilitation of cognition in medical and health systems. We aim to provide a better basis for motivating and accelerating future interfaces for medical and health systems. Therefore, we will provide many examples of existing and futuristic systems. The goal is to create a path for understanding how to design more effective medical systems in the future. The main applications are medical knowledge acquisition by intelligent user interfaces; networked embedded systems development and sensors development in activity monitoring of humans by non-intrusive sensors; and knowledge integration towards clinical decision support.

2. **Yuan-Fang Li, Monash University, Victoria, Australia**

   Bio: Yuan-Fang Li is a senior lecturer at Faculty of Information Technology, Monash University, Australia. He received his PhD in computer science from National University of Singapore in 2006. His research interests include knowledge graphs, knowledge representation and reasoning, ontology languages, and software engineering.

   Title: Who does what: expertise discovery in biomedical research

   Abstract: For a large research institution and a broad research discipline such as the healthcare and life sciences, it is a highly important and very challenging task to capture researchers' expertise, and to match researchers by expertise to assist in identifying inter-disciplinary collaboration opportunities and in making informed policy decisions. The challenges are multi-dimensional, stemming from the needs to (a) provide thorough coverage of the breadth and depth of the disciplinary areas, (b) develop accurate representation of researcher's expertise, and (c) process large volumes of data efficiently. Medical Subject Headings (MeSH), a comprehensive taxonomy for the life sciences, has been widely used for indexing MEDLINE publications. In this talk, I will present a novel framework for capturing and matching research expertise based on knowledge discovered from publications and encoded in MeSH.

Accepted Papers

The following full papers presenting original research works were accepted. Chen et al. describe a prototype mobile application for use by patients with type 1 diabetes to inform exercise decisions based on blood glucose levels. The paper describes a case-based reasoning application for the personalised recommendation for diabetics during exercise.

Rubin et al. present a deep-learning approach for heart sound classifications. Specifically, the sound data is first converted into a heat-map using popular MFCC approach, then a convolutional neural network is trained to do the classification. The approach has been applied on 2016 PhysioNet Computing in Cardiology challenge and achieved a reasonably good overall score - 8th out of 48 teams.

Choo et al. explore using Virtual Reality as part of the treatment for Agoraphobia and Social phobia. In the approach the patients are exposed to their phobia in a virtual environment. The intuition is that exposure to their phobia can help be overcome the irrational fear. The approach is part of a wider Cognitive behavioural therapy approach and has the benefit that the exposure in virtual space can be better controlled and supervised.

Nguyen et al. The paper presents a deep learning approach to handle irregularities and missing data in clinical time series, for mortality prediction.

Ormandy et al. present a technique to learn a representation of treatment and diagnoses using the well-known skip-gram model applied to ICD9 codes. The authors focus on patient similarity based on their medication and diagnosis.

Wang et al. describe a study on psyco-env corpus, which aims at annotating published studies for facilitating knowledge discovery on pathologies of mental diseases. The corpus of this paper focuses on the correlations between mental diseases and environmental factors. In addition to the corpus, the open source annotation tool should have broad value beyond this use case.

Zhu et al. introduce a graph based approach to mining adverse drug events from MEDLINE papers. The authors use a publicly available database to mine for adverse drug events, compare their graph based clustering approach to other methods. Their results show a slight improvement in accuracy over baseline methods. Short papers report on work in progress, descriptions of available datasets, as well as data collection efforts. Short papers can also be position papers regarding potential research challenges.

We very much appreciate the support of the workshop coordinator, Tianqing Zhu as well as this year's conference chair Fahiem Bacchus and program chair Carles Sierras.

We sincerely hope that the participants enjoy this year's workshop program and that this collection of papers will inspire and encourage more AI-related research for and within healthcare in the future.

Zina Ibrahim, Honghan Wu,
Kerstin Bach, Richard Dobson,
Spiros Denaxas, Nirmalie Wiratunga,
Stewart Massie, and Sadiq Sani

London/Trondheim/Aberdeen, July 2017