

Proceedings of the 2nd International Workshop on Explainable and Interpretable Machine Learning (XI-ML) – PREFACE

Martin Atzmueller^{1,2}, Tomáš Kliegr³ and Ute Schmid⁴

¹*Institute of Computer Science, Osnabrück University, Osnabrück, Germany*

²*German Research Center for Artificial Intelligence (DFKI), Osnabrück, Germany*

³*Department of Information and Knowledge Engineering, Faculty of Informatics and Statistics, Prague University of Economics and Business, Czech Republic*

⁴*Cognitive Systems, University of Bamberg, Bamberg, Germany*

1. Introduction

With the current scientific discourse on explainable AI (XAI), algorithmic transparency, interpretability, accountability and finally, explainability of algorithmic models and decisions, the XI-ML workshop on explainable and interpretable machine learning tackles these themes from the modeling and learning perspective. In particular, it targets interpretable methods and models being able to explain themselves and their output, respectively. The workshop aims to provide an interdisciplinary forum to investigate fundamental issues in explainable and interpretable machine learning as well as to discuss recent advances, trends, and challenges.

With this edition of the workshop, we aimed to provide a discussion platform for topics related to XAI. The main emphasis of the call was on approaches such as rule learning and pattern mining for explainable and interpretable machine learning (cf., e.g., [1, 2, 3]). An equally important emphasis was put on cognitive approaches [4, 5] and human concept learning and contrastive explanation [6]. The workshop did not intend to cover only supervised (classification approaches) but also research on XAI in the unsupervised domain such as clustering [7], or subgroup discovery [8]. The third focus area was explanations of black box models as well as applications of all of the above (cf., [9, 10]).

The first edition of the XI-ML (Explainable and Interpretable Machine Learning) workshop was held on September 21, 2020, at the 43rd German Conference on Artificial Intelligence, Bamberg, Germany. This second edition of the workshop was held on September 20, 2022, in co-location with the 45th German Conference on Artificial Intelligence, Trier, Germany (held virtually).

Proceedings of the 2nd International Workshop on Explainable and Interpretable Machine Learning (XI-ML)


✉ martin.atzmueller@uni-osnabrueck.de (M. Atzmueller); tomas.kliegr@vse.cz (T. Kliegr);

ute.schmid@uni-bamberg.de (U. Schmid)

🌐 <https://martin.atzmueller.net/> (M. Atzmueller); <http://kliegr.eu/> (T. Kliegr)

🆔 0000-0002-2480-6901 (M. Atzmueller); 0000-0002-7261-0380 (T. Kliegr); 0000-0002-1301-0326 (U. Schmid)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

For the workshop, there were six accepted papers out of eight submissions in total. Below, we structure these according to the general topics of *explanation & methods* as well as *explanation & applications* with their specific methodological foci in the context of explainable and interpretable machine learning, as well as individual applications, respectively.

- Explanation & Methods:
 - *Evolutionary Counterfactual Visual Explanation* by Jacqueline Höllig, Steffen Thoma and Cedric Kulbach. The paper proposes a genetic algorithm for the important problem of explaining image classification models using counterfactuals.
 - *Self-explaining variational Gaussian Processes for transparency and modelling of prior knowledge* by Sarem Seitz. This paper presents an approach for combining Gaussian Processes with recent algorithms proposed in XAI to study feature importance.
 - *Study on Criteria for Explainable AI for Laypeople* by Thorsten Zylowski. This paper presented a survey aimed at identifying criteria important for the development of XAI solutions for the public. The results of this survey show that explanations for AI decisions are widely sought for. The authors suggest that global explanations are instrumental in building initial user trust, while local explanations probably contribute to maintaining trust.
- Explanation & Application:
 - *Imitation Learning of Logical Program Policies for Multi-Agent Reinforcement Learning* by Manuel Eberhardinger, Johannes Maucher and Setareh Maghsudi. This paper deals with making decisions in Multi-agent systems interpretable by using using the Logical Program Policies, a recently proposed approach originally developed for few-shot Bayesian imitation learning [11].
 - *Mining Interesting Outlier Subgraphs In Attributed Graphs* by Ahmad Mel and Tijn De Bie. The paper presents an approach for identifying interesting outlier subgraphs in attributed graphs. The algorithm searches for k interesting anomalous subgraphs by scanning the space of candidates using a branch-and-bound beam search procedure. The topic is well-motivated, the approach described algorithmically and also exemplified.
 - *Multi-Perspective Anomaly Detection on Bipartite Multi-Layer Social Interaction Networks* by Asep Maulana and Martin Atzmueller. This paper focuses on anomaly detection in complex networks, specifically bipartite multi-layer networks for modeling multiple relations. It presents an interactive approach for characterizing anomalies from different explanatory perspectives.

Acknowledgments

We would like to thank the KI'22 organization team for their support throughout the event.

References

- [1] M. Atzmueller, F. Puppe, H.-P. Buscher, Exploiting Background Knowledge for Knowledge-Intensive Subgroup Discovery, in: Proc. IJCAI, Edinburgh, Scotland, 2005, pp. 647–652.
- [2] M. Atzmueller, J. Baumeister, F. Puppe, Quality measures and semi-automatic mining of diagnostic rule bases, in: Proc. INAP, number 3392 in LNAI, Springer, 2005, pp. 65–78.
- [3] J. Fürnkranz, T. Kliegr, A brief overview of rule learning, in: Proc. RuleML: Rule Technologies: Foundations, Tools, and Applications., Springer, 2015, pp. 54–69.
- [4] L. Ai, S. H. Muggleton, C. Hocquette, M. Gromowski, U. Schmid, Beneficial and harmful explanatory machine learning, *Machine Learning* 110 (2021) 695–721.
- [5] T. Kliegr, Š. Bahník, J. Fürnkranz, A review of possible effects of cognitive biases on interpretation of rule-based machine learning models, *Artificial Intelligence* 295 (2021).
- [6] J. Rabold, M. Siebers, U. Schmid, Generating contrastive explanations for inductive logic programming based on a near miss approach, *Machine Learning* 111 (2022) 1799–1820.
- [7] J. Žárský, G. Lopez, T. Kliegr, Explainability of text clustering visualizations—twitter disinformation case study, *IEEE Computer Graphics and Applications* 42 (2022) 8–19.
- [8] M. Iferroudjene, C. Lonjarret, C. Robardet, M. Plantevit, M. Atzmueller, Methods for explaining top-n recommendations through subgroup discovery, *Data Mining and Knowledge Discovery* (2022) 1–40.
- [9] D. Müller, M. März, S. Scheele, U. Schmid, An interactive explanatory ai system for industrial quality control, in: Proc. AAAI, volume 36, 2022, pp. 12580–12586.
- [10] L. Beranová, M. P. Joachimiak, T. Kliegr, G. Rabby, V. Sklenák, Why was this cited? explainable machine learning applied to covid-19 research literature, *Scientometrics* 127 (2022) 2313–2349.
- [11] T. Silver, K. R. Allen, A. K. Lew, L. P. Kaelbling, J. Tenenbaum, Few-shot bayesian imitation learning with logical program policies, in: Proc. AAAI, volume 34, 2020, pp. 10251–10258.

Editors

- Martin Atzmueller, Osnabrück University & DFKI, Germany
- Tomáš Kliegr, Prague University of Economics and Business, Czech Republic
- Ute Schmid, University of Bamberg, Germany

Program Committee of XI-ML 2022

- Klaus-Dieter Althoff, University of Hildesheim & DFKI, Germany
- Henrik Boström, KTH Royal Institute of Technology, Sweden
- Amit Dhurandhar, IBM TJ Watson Research Center, USA
- Johannes Fürnkranz, Johannes Kepler University, Linz, Austria
- Martin Holena, Czech Academy of Sciences, Czech Republic
- Grzegorz Nalepa, Jagiellonian University, Poland
- Stefano Teso, KU Leuven, Belgium
- Filip Železný, Czech Technical University, Czech Republic