Onto Explicative Data Mining: Exploratory, Interpretable and Explainable Analysis

Martin Atzmueller
Tilburg University, Cognitive Science and Artificial Intelligence, Tilburg, Netherlands
m.atzmuller@uvt.nl

Data mining systems are commonly applied to obtain a set of *novel*, *potentially useful*, and ultimately *interesting* patterns from a given (large) data set [1]. This can be achieved using *exploratory* methods, e. g., utilizing descriptive data mining techniques like methods for association rule mining [2] or subgroup discovery, e. g., [3], [4]. While the resulting patterns are typically interpretable, the large results sets in pattern mining, i. e., a large sets of potentially interesting patterns that the user needs to assess, require further *exploration* and *interpretation* techniques. Such problems also occur for other complex models in data mining which often require *explanations* of their results and/or structure. In this context, *explanation-aware* approaches have been a prominent research direction in artificial intelligence and data science, e. g., [5]–[8]. Recently, the concept of transparent and explainable models has also gained a strong focus and momentum in the data mining and machine learning community, e. g., [9]–[11].

We introduce *explicative data mining* as a comprehensive paradigm, tackling all these different aspects. Similar to the philosophical process of *explication* cf. [12], [13] which aims to make the implicit explicit, explicative data mining aims to model, describe and explain the underlying structure in the data. By targeting interpretable (and transparent) models utilizing exploratory and explanation-aware methods, these can be constructed and inspected on different layers and levels. This ranges from pure data summarization to pattern-based exploratory data mining. Furthermore, these features also provide for different options for including the human in the loop, e. g., using visualization methods.

We outline and discuss the explicative data mining paradigm in detail: We introduce foundational aspects, present the respective approaches summarized above, and discuss current perspectives and challenges. In particular, we focus on explicative data mining methods, e. g., in the areas of pattern mining [14]–[16] and feature engineering for machine learning approaches [17], [18], with exemplary applications in network analysis and anomaly detection. The pattern mining methods are discussed in the context of the VIKAMINE[1] system [19]. Furthermore, we also discuss the relation to incorporating prior knowledge, e. g., in the form of knowledge graphs [20]–[22] into the data mining process [18]. This enables hybrid approaches that incorporate semantic knowledge [23] into the process, e. g., supporting modeling and explanation methods.

[1] www.vikamine.org

References

[1] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery: An Overview," in *Advances in Knowledge Discovery and Data Mining*. AAAI Press, 1996, pp. 1–34.
[2] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in *Proc. VLDB*. Morgan Kaufmann, 1994, pp. 487–499.
[3] S. Wrobel, "An Algorithm for Multi-Relational Discovery of Subgroups," in *Proc. PKDD*. Springer, 1997, pp. 78–87.
[4] M. Atzmueller, "Subgroup Discovery – Advanced Review," *WIREs Data Mining and Knowledge Discovery*, vol. 5, no. 1, pp. 35–49, 2015.
[5] J. L. Kolodner, "Making the Implicit Explicit: Clarifying the Principles of Case-based Reasoning," *Case-based Reasoning: Experiences, Lessons & Future Directions*, pp. 349–370, 1996.
[6] M. R. Wick and W. B. Thompson, "Reconstructive Expert System Explanation," *Artif. Intell.*, vol. 54, no. 1-2, pp. 33–70, 1992.
[7] T. Roth-Berghofer, S. Schulz, D. Leake, and D. Bahls, "Explanation-Aware Computing," *AI Magazine*, vol. 28, no. 4, 2007.
[8] M. Atzmueller and T. Roth-Berghofer, "The Mining and Analysis Continuum of Explaining Uncovered," in *Proc. AI*, 2010.
[9] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," in *Proc. ACM SIGKDD*. ACM, 2016, pp. 1135–1144.
[10] X. Li and J. Huan, "Constructivism Learning: A Learning Paradigm for Transparent Predictive Analytics," in *Proc. SIGKDD*. ACM, 2017, pp. 285–294.
[11] O. Biran and C. Cotton, "Explanation and Justification in Machine Learning: A Survey," in *IJCAI-17 Workshop on Explainable AI*, 2017.
[12] R. Carnap, "Logical foundations of probability," 1962.
[13] P. Maher, "Explication Defended," *Studia Logica*, vol. 86, no. 2, pp. 331–341, 2007.
[14] M. Atzmueller, "Detecting Community Patterns Capturing Exceptional Link Trails," in *Proc. IEEE/ACM ASONAM*. IEEE Press, 2016.
[15] M. Atzmueller, A. Schmidt, B. Kloepper, and D. Arnu, "HypGraphs: An Approach for Analysis and Assessment of Graph-Based and Sequential Hypotheses," in *Postproceedings NFMCP 2016*. Springer, 2017.
[16] M. Atzmueller, D. Arnu, and A. Schmidt, "Anomaly Detection and Structural Analysis in Industrial Production Environments," in *Proc. IDSC*, Salzburg, Austria, 2017.
[17] M. Atzmueller, N. Hayat, A. Schmidt, and B. Klöpper, "Explanation-Aware Feature Selection using Symbolic Time Series Abstraction: Approaches and Experiences in a Petro-Chemical Production Context," in *Proc. IEEE INDIN*. Boston, MA, USA: IEEE Press, 2017.
[18] M. Atzmueller and E. Sternberg, "Mixed-Initiative Feature Engineering Using Knowledge Graphs," in *Proc. K-CAP*. ACM, 2017.
[19] M. Atzmueller and F. Lemmerich, "VIKAMINE - Open-Source Subgroup Discovery, Pattern Mining, and Analytics," in *Proc. ECML/PKDD*. Springer, 2012.
[20] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia - A Crystallization Point for the Web of Data," *Web Semantics*, vol. 7, no. 3, pp. 154–165, 2009.
[21] J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum, "YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia," *Artificial Intelligence*, vol. 194, pp. 28–61, 2013.
[22] X. Wilcke, P. Bloem, and V. de Boer, "The Knowledge Graph as the Default Data Model for Learning on Heterogeneous Knowledge," *Data Science*, no. Preprint, pp. 1–19, 2017.
[23] P. Ristoski and H. Paulheim, "Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey," *Web Semantics*, vol. 36, pp. 1–22, 2016.

Preprint of: Martin Atzmueller (2017) Onto Explicative Data Mining: Exploratory, Interpretable and Explainable Analysis.
Proc. Dutch-Belgian Database Day (DBDBD), Eindhoven University of Technology, Eindhoven, 2017.