

Semantic Data Mining: A Brief Outline

Martin Atzmueller¹[0000-0002-2480-6901],
Grzegorz J. Nalepa²[0000-0002-8182-4225],
Szymon Bobek²[0000-0002-6350-8405], and
Nada Lavrac³[0000-0002-9995-7093]

¹ Semantic Information Systems Group, Osnabrück University, Germany

`martin.atzmueller@uni-osnabrueck.de`

² GEIST Research Group, Jagiellonian University, Poland

`gjn@gjn.re`, `szymon.bobek@uj.edu.pl`

³ Jožef Stefan Institute, Ljubljana, Slovenia

`nada.lavrac@ijs.si`

When considering knowledge discovery in databases, data mining, and associated machine learning and data analytic methods, the general goal of data mining is to uncover novel, interesting, and ultimately understandable patterns, relating to valuable, useful and implicit knowledge [10]. Considering the development of data mining in the last decades, it can be observed that not only the addressed data mining tasks were more restricted, but also the applied data mining workflows were simpler than today. Thus, recent advances of data mining and machine learning address new challenges in its practical use for data analysis. This relates to, for example, novel processing, mining and learning methods and approaches, as well as large-scale and complex data representations [1, 8, 11, 19], which also includes important aspects of interpretability [21, 34] and explainability [1, 26, 28].

Using semantic information such as domain/background knowledge in data mining is a promising emerging direction for addressing these problems [3, 22, 33], where the domain knowledge is typically represented in a knowledge repository, such as an ontology, or a knowledge base [9, 25, 27, 30]. The main aspect of semantic data mining [2, 9, 15–17, 20, 23, 24], is the explicit integration of this knowledge into the data mining and knowledge discovery modeling step, where the algorithms for data mining/modeling or post-processing make use of the formalized knowledge to improve the overall results. There has been growing interest in this issue, e.g., [3–5, 18, 22, 31], in various domains, e. g., in the medical domain [4, 7, 12, 13, 17, 29] but also in human behavior analysis and industrial applications [5, 6, 14, 32, 35].

In summary, the term *semantic data mining* can be interpreted rather broadly as being concerned with the integration of semantic/domain knowledge into the data mining/knowledge discovery process, where in the respective methods and approaches, “semantic information” or “declarative knowledge” is meaningfully integrated into the data mining process. For example, this can relate to ontologies or to other declarative and/or rule-based mechanisms and formalizations w.r.t. feature construction and engineering, the semantics of attributes, and different post-processing approaches etc.

SEDAMI Workshop 2021 – Contextualization and Preface

The goal of the SEDAMI 2021 workshop is to offer an interdisciplinary forum for researchers working in the fields of semantic data mining. With this workshop we thus aim to get an insight into the current status of research in this area. We focus mainly on methods that allow include/utilize/exploit semantic information and domain knowledge in the context of machine learning and data mining. The workshop seeks for contributions on methods, techniques and applications that are both domain-specific but also transversal to different application domains. In particular, this includes contributions that aim to focus on semantic data mining for providing and/or enhancing interpretability, the introduction and preservation of knowledge, as well as the provisioning of explanations.

Submissions and Sessions

This proceedings volume comprises the papers of the SEDAMI 2021 workshop. In total, we received 7 submissions, from which we were able to accept five submissions based on a rigorous reviewing process.

Based on the set of accepted papers, we set up two sessions. The first session discusses the foundations of semantic data mining. The work *Meta-Interpretive Learning meets Neural Networks* by Victor Guimarães and Vítor Costa discusses a structure learning system based on meta-interpretive learning. The paper *Towards Explainable Relational Boosting via Propositionalization* by Blaž Škrlj and Nada Lavrač describes an approach improving black-box classifiers' interpretability in a relational setting using propositionalization, also combining XGBoost with SHAP. In *Declarative Knowledge Discovery in Databases via Meta-Learning - Towards Advanced Analytics*, Dietmar Seipel and Martin Atzmueller propose a novel approach for declarative knowledge discovery in databases enabling advanced analytics via the concept of meta-learning.

The second session is concerned with modeling and application of semantic data mining. In *Interpretable Knowledge Mining for Heart Failure Prognosis Risk Evaluation* by Shaobo Wang, Guangliang Liu, Wenyan Zhu, Zengtao Jiao, Haichen Lv, Jun Yan and Yunlong Xia, a pipeline to mine interpretable knowledge from electronic health records in the context of Heart Failure (HF) prognosis risk evaluation is proposed. Finally, the paper *Knowledge-Augmented Induction of Complex Networks on Supply-Demand-Material Data* by Dan Hudson, Leonid Schwenke, Stefan Bloemhevel, Arnab Ghosh Chowdhury, Nils Schut and Martin Atzmueller presents a method for matching items in a database according to their attributes, using knowledge of sub-contexts within the problem domain. The goal is to improve the specificity and relevance of matches, specifically within a challenging domain, i. e., supply chain modeling.

We thank all the participants of the workshop for their contributions and the organizers of the IJCAI 2021 conference for their support. Additionally, we want to thank the reviewers for their careful help in selecting and improving the accepted workshop papers.

We are looking forward to a very exciting and interesting workshop.

Osnabrück, Ljubljana, Krakow – August 2021

Martin Atzmueller, Grzegorz J. Nalepa, Szymon Bobek, Nada Lavrac

References

1. Atzmueller, M.: Declarative Aspects in Explicative Data Mining for Computational Sensemaking. In: Seipel, D., Hanus, M., Abreu, S. (eds.) Proc. International Conference on Declarative Programming. pp. 97–114. Springer, Heidelberg, Germany (2018)
2. Atzmueller, M., Lemmerich, F., Reutelshoefer, J., Puppe, F.: Wiki-enabled semantic data mining—task design, evaluation and refinement. In: Proc. International Workshop on Design, Evaluation and Refinement of Intelligent Systems (DERIS), vol. CEUR-WS. vol. 545 (2009)
3. Atzmueller, M., Puppe, F., Buscher, H.P.: Exploiting Background Knowledge for Knowledge-Intensive Subgroup Discovery. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05). pp. 647–652. Edinburgh, Scotland (2005)
4. Atzmueller, M., Seipel, D.: Declarative specification of ontological domain knowledge for descriptive data mining (extended version). In: Proceedings of the 18th International Conference on Applications of Declarative Programming and Knowledge Management. Springer (2008)
5. Atzmueller, M., Sternberg, E.: Mixed-initiative feature engineering using knowledge graphs. In: Proceedings of the 9th International Conference on Knowledge Capture (K-Cap). ACM Press, New York, NY, USA (2017)
6. Bobek, S., Nalepa, G.J., Ślażyński, M.: Heartdroid—rule engine for mobile and context-aware expert systems. *Expert Systems* **36**(1), e12328 (2019)
7. Cespivova, H., Rauch, J., Svatek, V., Kejkula, M.: Roles of medical ontology in association mining crisp-dm cycle. In: Proceedings of the ECML/PKDD 2004 Workshop on Knowledge Discovery and Ontologies. Pisa, Italy (2004)
8. Che, D., Safran, M., Peng, Z.: From big data to big data mining: challenges, issues, and opportunities. In: International conference on database systems for advanced applications. pp. 1–15. Springer (2013)
9. Dou, D., Wang, H., Liu, H.: Semantic data mining: A survey of ontology-based approaches. In: Proceedings of the 2015 IEEE 9th international conference on semantic computing (IEEE ICSC 2015). pp. 244–251. IEEE (2015)
10. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery: An Overview. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) *Advances in Knowledge Discovery and Data Mining*, pp. 1–34. AAAI Press (1996)
11. Guven, C., Seipel, D., Atzmueller, M.: Applying ASP for Knowledge-Based Link Prediction with Explanation Generation in Feature Rich Networks. *IEEE Transactions on Network Science and Engineering* **8**(2) (April–June 2021)
12. Kralj, J., Robnik-Sikonja, M., Lavrac, N.: Netsdm: Semantic data mining with network analysis. *Journal of Machine Learning Research* **20**(32), 1–50 (2019)
13. Kuo, Y.T., Lonie, A., Sonenberg, L., Paizis, K.: Domain ontology driven data mining: A medical case study. In: DDDM '07: Proceedings of the 2007 International Workshop on Domain Driven Data Mining. pp. 11–17. ACM, New York, NY, USA (2007)
14. Kutt, K., Drazyk, D., Bobek, S., Nalepa, G.J.: Personality-based affective adaptation methods for intelligent systems. *Sensors* **21**(1), 163 (2021)
15. Lavrač, N., Škrlić, B., Robnik-Šikonja, M.: Propositionalization and embeddings: two sides of the same coin. *Machine Learning* **109**(7), 1465–1507 (2020)
16. Lavrač, N., Vavpetič, A.: Relational and semantic data mining. In: International Conference on Logic Programming and Nonmonotonic Reasoning. pp. 20–31. Springer (2015)
17. Lavrač, N., Vavpetič, A., Soldatova, L., Trajkovski, I., Novak, P.K.: Using ontologies in semantic data mining with segs and g-segs. In: International Conference on Discovery Science. pp. 165–178. Springer (2011)
18. Ławrynowicz, A.: *Semantic Data Mining - An Ontology-Based Approach*, Studies on the Semantic Web, vol. 29. IOS Press (2017)

19. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *nature* **521**(7553), 436–444 (2015)
20. Liu, H.: Towards semantic data mining. In: 9th International Semantic Web Conference (ISWC2010). pp. 7–11 (2010)
21. Molnar, C., Casalicchio, G., Bischl, B.: Interpretable machine learning—a brief history, state-of-the-art and challenges. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 417–431. Springer (2020)
22. Nalepa, G.J.: Modeling with Rules Using Semantic Knowledge Engineering, Intelligent Systems Reference Library, vol. 130. Springer (2018)
23. Nalepa, G.J., Bobek, S., Kutt, K., Atzmueller, M.: Semantic data mining in ubiquitous sensing: A survey. *Sensors* **21**(13), 4322 (2021)
24. Nalepa, G.J., Kutt, K., Bobek, S.: Mobile platform for affective context-aware systems. *Future Generation Computer Systems* **92**, 490–503 (2019)
25. Ristoski, P., Paulheim, H.: Semantic web in data mining and knowledge discovery: A comprehensive survey. *Web Semantics* **36**, 1–22 (2016)
26. Roscher, R., Bohn, B., Duarte, M.F., Garcke, J.: Explainable machine learning for scientific insights and discoveries. *Ieee Access* **8**, 42200–42216 (2020)
27. von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Garcke, J., Bauckhage, C., Schuecker, J.: Informed machine learning – a taxonomy and survey of integrating knowledge into learning systems (2020)
28. Schwenke, L., Atzmueller, M.: Show Me What You’re Looking For: Visualizing Abstracted Transformer Attention for Enhancing Their Local Interpretability on Time Series Data. In: Proc. 34th International Florida Artificial Intelligence Research Society Conference (FLAIRS-2021). FLAIRS, North Miami Beach, FL, USA (2021)
29. Sikora, M., Wróbel, Ł., Gudyś, A.: Guider: a guided separate-and-conquer rule learning in classification, regression, and survival settings. *Knowledge-Based Systems* **173**, 1–14 (2019)
30. Sirichanya, C., Kraisak, K.: Semantic data mining in the information age: A systematic review. *International Journal of Intelligent Systems* (2021)
31. Svátek, V., Rauch, J., Ralbovský, M.: Ontology-enhanced association mining. In: *Semantics, Web and Mining*. LNCS, vol. 4289, pp. 163–179 (2005)
32. Szpyrka, M., Brzywczy, E., Napieraj, A., Korsiński, J., Nalepa, G.J.: Conformance checking of a longwall shearer operation based on low-level events. *Energies* **13**(24), 6630 (2020)
33. Vavpetič, A., Lavrač, N.: Semantic data mining system g-SEGS. In: In proceedings of the Workshop on Planning to Learn and Service-Oriented Knowledge Discovery (PlanSoKD-11), ECML PKDD conference, Athens, Greece, September 5-9. pp. 17–29 (2011)
34. Vollert, S., Atzmueller, M., Theissler, A.: Interpretable Machine Learning: A Brief Survey From the Predictive Maintenance Perspective. In: Proc. IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2021). IEEE (2021)
35. Weidner, D., Atzmueller, M., Seipel, D.: Finding maximal non-redundant association rules in tennis data. In: Hofstedt, P., Abreu, S., John, U., Kuchen, H., Seipel, D. (eds.) *Declarative Programming and Knowledge Management - Conference on Declarative Programming, DECLARE 2019, Unifying INAP, WLP, and WFLP, Revised Selected Papers*. Lecture Notes in Computer Science, vol. 12057, pp. 59–78. Springer (2019)

**SEDAMI 2021: International workshop on Semantic Data Mining,
held (online) at IJCAI 2021 on 20th of August 2021**

Editors

- Martin Atzmueller, Osnabrück University, Germany
- Grzegorz J. Nalepa, Jagiellonian University, Poland
- Szymon Bobek, Jagiellonian University, Poland
- Nada Lavrač, Jožef Stefan Institute, Slovenia

Program Committee

- Klaus-Dieter Althoff, University of Hildesheim & DFKI, Germany
- Martin Atzmueller, Osnabrück University, Germany
- Przemysław Biecek, Warsaw University of Technology, Poland
- Szymon Bobek, Jagiellonian University, Poland
- João Gama, University of Porto, Portugal
- Nada Lavrač, Jožef Stefan Institute, Slovenia
- Stan Matwin, Dalhousie University, Canada
- Grzegorz J. Nalepa, Jagiellonian University, Poland
- Sławomir Nowaczyk, Halmstad University, Sweden
- Jose Palma, Universidad de Murcia, Spain
- Juan Pavon, Universidad Complutense de Madrid, Spain
- Marc Plantevit, Université Lyon, France
- Eric Postma, Tilburg University, The Netherlands
- Céline Rouveirol, Université Sorbonne Paris Nord, France
- Marek Sikora, Silesian University of Technology, Poland
- Blaž Škrlj, Jožef Stefan Institute, Slovenia