

# Applied Descriptive Pattern Mining

Martin Atzmueller<sup>1</sup> and Florian Lemmerich<sup>2</sup>

<sup>1</sup> University of Kassel, Research Center for Information System Design,  
Ubiquitous Data Mining Team, Chair for Knowledge and Data Engineering,  
Wilhelmshöher Allee 73, 34121 Kassel, Germany

<sup>2</sup> University of Wuerzburg, Chair for Artificial Intelligence and Applied  
Computer Science, Am Hubland, 97074

**Abstract.** Pattern mining, and more specifically descriptive pattern mining aims at extracting relevant "nuggets" of knowledge from a set of data. In this context, the focus is on descriptive applications, i.e., for explorative data mining, e.g., extracting knowledge for humans, instead of considering mainly automatic approaches, e.g., for classification. Prominent approaches for descriptive pattern mining include frequent pattern mining, supervised descriptive rule induction, and subgroup discovery but also descriptive community mining approaches.

The aim of this tutorial is to provide a general, comprehensive overview of the state-of-the-art of descriptive pattern mining. We consider different types of data such as structured (tabular) data, texts, and networks.

## References

- [Agrawal & Srikant] R. Agrawal, R. Srikant (1994) Fast Algorithms for Mining Association Rules. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) Proc. 20th Int. Conf. Very Large Data Bases, (VLDB). pp. 487–499. Morgan Kaufmann
- [Atzmueller 2007] M. Atzmueller (2007) Knowledge-Intensive Subgroup Mining – Techniques for Automatic and Interactive Discovery, Vol. 307 of Dissertations in Artificial Intelligence-Infix (Diski), IOS Press
- [Atzmueller et al. 2004] M. Atzmueller, F. Puppe, H.-P. Buscher (2004) Towards Knowledge-Intensive Subgroup Discovery, Proc. LWA 2004, pp. 117–123.
- [Atzmueller & Mitzlaff 2011] M. Atzmueller and F. Mitzlaff (2011) Efficient Descriptive Community Mining. Proc. 24th International FLAIRS Conference, pages 459-464, Palo Alto, CA, USA. AAAI Press.
- [Atzmueller et al. 2011] M. Atzmueller, D. Benz, S. Doerfel, A. Hotho, R. Jäschke, B.-E. Macek, F. Mitzlaff, C. Scholz, and G. Stumme (2011) Enhancing Social Interactions at Conferences. *it - Information Technology*, (53)3:101–107, Oldenbourg Wissenschaftsverlag GmbH
- [Atzmueller & Roth-Berghofer 2010] M. Atzmueller, T. Roth-Berghofer (2010) The Mining and Analysis Continuum of Explaining Uncovered. Proc. 30th SGAI International Conference on Artificial Intelligence (AI-2010)
- [Atzmueller & Lemmerich 2009] M. Atzmueller, F. Lemmerich (2009) Fast Subgroup Discovery for Continuous Target Concepts. Proc. International Symposium on Methodologies for Intelligent Systems, Vol. 5722 of LNCS, Springer, Berlin, pp. 1–15.

- [Atzmueller et al. 2009] M. Atzmueller, F. Lemmerich, B. Krause, and A. Hotho (2009) Who are the Spammers? Understandable Local Patterns for Concept Description. In Proc. 7th Conference on Computer Methods and Systems, Krakow, Poland. Oprogramowanie Nauko-Techniczne.
- [Atzmueller & Puppe 2008] M. Atzmueller and F. Puppe (2008) A Case-Based Approach for Characterization and Analysis of Subgroup Patterns. *Journal of Applied Intelligence*, 28(3):210-221
- [Atzmueller & Puppe 2006] M. Atzmueller and F. Puppe (2006) SD-Map - A Fast Algorithm for Exhaustive Subgroup Discovery. Proc. 10th European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD 2006), pp. 6-17, Heidelberg, Germany. Springer Verlag
- [Atzmueller et al. 2008] M. Atzmueller, P. Kluegl, and F. Puppe (2008) Rule-Based Information Extraction for Structured Data Acquisition using TextMarker.
- [Atzmueller et al. 2006] M. Atzmueller, J. Baumeister, and F. Puppe (2006) Inrospective Subgroup Analysis for Interactive Knowledge Refinement. Proc. 19th International Florida Artificial Intelligence Research Society Conference 2006 (FLAIRS-2006), 402-407, AAAI Press
- [Atzmueller & Puppe 2005] M. Atzmueller and F. Puppe (2005) Semi-Automatic Visual Subgroup Mining using VIKAMINE. *Journal of Universal Computer Science*, 11(11):1752-1765, 2005.
- [Atzmueller et al. 2005a] M. Atzmueller, F. Puppe, and H.-P. Buscher (2005) Exploiting Background Knowledge for Knowledge-Intensive Subgroup Discovery. Proc. 19th International Joint Conference on Artificial Intelligence (IJCAI-05), 647-652, Edinburgh, Scotland
- [Atzmueller et al. 2005b] M. Atzmueller, J. Baumeister, A. Hemsing, E.-J. Richter, and F. Puppe (2005) Subgroup Mining for Interactive Knowledge Refinement. In Proc. 10th Conference on Artificial Intelligence in Medicine AIME 05), LNAI 3581, pp. 453-462, Heidelberg, Germany, Springer Verlag.
- [Atzmueller et al. 2005c] M. Atzmueller, J. Baumeister, and F. Puppe (2005) Quality Measures and Semi-Automatic Mining of Diagnostic Rule Bases (extended version). Proc. 15th International Conference on Applications of Declarative Programming and Knowledge Management (INAP 2004), 3392:65-78
- [Atzmueller et al. 2005d] M. Atzmueller, F. Puppe, and H.-P. Buscher (2005) Profiling Examiners using Intelligent Subgroup Mining. In Proc. 10th International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2005), pp. 46-51, Aberdeen, Scotland
- [Atzmueller et al. 2003] M. Atzmueller, J. Baumeister, and F. Puppe (2003) Evaluation of two Strategies for Case-Based Diagnosis handling Multiple Faults. Proc. 2nd Conf. Professional Knowledge Management (WM2003), Luzern, Switzerland
- [Batal & Hauskrecht 2010a] I. Batal, M. Hauskrecht (2010) A concise representation of association rules using minimal predictive rules. *Machine Learning and Knowledge Disc.*, 87-102
- [Batal & Hauskrecht 2010b] I. Batal, M. Hauskrecht (2010) Constructing classification features using minimal predictive patterns. Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, 869-877
- [Bayardo et al. 1999] R. Bayardo, R. Agrawal, D. Gunopulos (1999) Constraint-based rule mining in large, dense databases. Proc. 15th IEEE International Conference on Data Engineering

- [Bringmann et al. 2011] B. Bringmann, S. Nijssen, and A. Zimmermann (2011) Pattern-based Classification: A Unifying Perspective. arXiv:1111.6191
- [Clauset et al. 2004] A. Clauset, M. E. J. Newman, C. Moore (2004) Finding Community Structure in Very Large Networks. arXiv:cond-mat/0408187
- [Dakin 1965] R. J. Dakin (1965) A tree-search algorithm for mixed integer programming problems. *The computer journal* 8.3, 250-255.
- [Fayyad et al. 1996a] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth (1996) From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17, 37–54
- [Fayyad et al. 1996b] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth (1996) 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
- [Garriga et al. 2008] G. C. Garriga, P. Kralj, and N. Lavrac (2008) Closed Sets for Labeled Data. *Journal of Machine Learning Research*, 9, 559–580
- [Geng & Hamilton 2006] L. Geng and H. J. Hamilton (2006) Interestingness Measures for Data Mining: A Survey. *ACM Computing Surveys*, 38(3)
- [Girvan & Newman 2002] M. Girvan, M. E. J. Newman (2002) Community Structure in Social and Biological Networks, *PNAS* 99 (12)
- [Grosskreutz & Paurat 2011] H. Grosskreutz, D. Paurat (2011) Fast and memory-efficient discovery of the top-k relevant subgroups in a reduced candidate space. *Machine Learning and Knowledge Discovery in Databases*. Springer Berlin Heidelberg, 533-548.
- [Grosskreutz & Rueping 2009] H. Grosskreutz and S. Rueping. On Subgroup Discovery in Numerical Domains. *Data Mining and Knowledge Discovery*, 19(2), 210–226
- [Grosskreutz et al. 2008] H. Grosskreutz, S. Rueping, and S. Wrobel (2008) Tight Optimistic Estimates for Fast Subgroup Discovery. In *Proc. ECML/PKDD*, volume 5211 of *LNAI*, pages 440–456, Heidelberg, Germany, Springer Verlag.
- [Guns et al. 2011] T. Guns, S. Nijssen, and L. De Raedt (2011) Itemset mining: A constraint programming perspective.” *Artificial Intelligence* 175.12, 1951-1983.
- [Kavsek & Lavrac 2004] B. Kavsek and N. Lavrac (2004) Analysis of Example Weighting in Subgroup Discovery by Comparison of Three Algorithms on a Real-Life Data Set. In *Proc. Workshop on Advances in Inductive Rule Learning*, at *ECML/PKDD*
- [Kavsek et al. 2003] B. Kavsek, N. Lavrac, and V. Jovanoski (2003) APRIORI-SD: Adapting Association Rule Learning to Subgroup Discovery. In *Proc. 5th Intl. Symposium on Intelligent Data Analysis*, pages 230–241, Heidelberg, Germany, Springer Verlag.
- [Kloesgen 1996] Kloesgen, W. (1996) Explora: A Multipattern and Multistrategy Discovery Assistant. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., editors, *Advances in Knowledge Discovery and Data Mining*, pp. 249–271. AAAI Press.
- [Kluegl et al. 2009] P. Kluegl, M. Atzmueller, and F. Puppe (2009) Meta-Level Information Extraction. *The 32nd Annual Conference on Artificial Intelligence*, Springer, Berlin
- [Knobbe & Ho 2006] A. Knobbe and E. Ho (2006) Pattern teams. In *Knowledge Discovery in Databases: PKDD 2006*, pages 577–584, Heidelberg, Germany, Springer Verlag.
- [Lancichinetti 2009] A. Lancichinetti, S. Fortunato (2009) Community Detection Algorithms: A Comparative Analysis. arXiv:0908.1062

- [Leman et al. 2008] D. Leman, A. Feelders, and A. Knobbe (2008). Exceptional Model Mining. In Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, volume 5212 of Lecture Notes in Computer Science, pages 1–16. Springer.
- [Mitzlaff et al. 2011] F. Mitzlaff, M. Atzmueller, D. Benz, A. Hotho, and G. Stumme (2011) Community Assessment using Evidence Networks. In Analysis of Social Media and Ubiquitous Data, volume 6904 of LNAI
- [Morishita & Sese] S. Morishita, J. Sese (2000) Traversing Itemset Lattices with Statistical Metric Pruning. In Proc. of ACM SIGMOD, 226-236
- [Newman 2006] M. E. Newman (2006) Modularity and Community Structure in Networks. PNAS, 103(23), 8577-8582.
- [Novak et al. 2009] P. Kralj Novak, N Lavrac, and G. I. Webb (2009) Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. The Journal of Machine Learning Research 10, 377-403
- [Puppe et al. 2008] F. Puppe, M. Atzmueller, G. Buscher, M. Huettig, H. Lührs, and H.-P. Buscher (2008) Application and Evaluation of a Medical Knowledge-System in Sonography (SonoConsult). In Proc. 18th European Conference on Artificial Intelligence (ECAI 2008), pp. 683-687
- [Scholz et al. 2011] C. Scholz, S. Doerfel, M. Atzmueller, A. Hotho, and G. Stumme (2011) Resource-Aware On-Line RFID Localization Using Proximity Data. Proceedings of the ECML/PKDD 2011, 129–144, Springer Verlag, Heidelberg, Germany
- [Tuzhilin 2002] A. Tuzhilin (2002) Handbook of Data Mining and Knowledge Discovery, chapter 19.2.2: Usefulness, Novelty, and Integration of Interestingness Measures. Oxford University Press, New York
- [Wang et al. 2006] K. Wang, Y. Jiang, A. Tuzhilin (2006) Mining Actionable Patterns by Role Models. In: Proc. 22nd International Conference on Data Engineering (ICDE). pp. 16–. ICDE '06, IEEE Computer Society, Washington, DC, USA
- [Wassermann & Faust 1994] S. Wasserman, and K. Faust (1994) Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences. Cambridge University Press, 1 edition.
- [Wrobel 1997] S. Wrobel (1997) An Algorithm for Multi-Relational Discovery of Subgroups. In Proc. 1st Europ. Symp. Principles of Data Mining and Knowledge Discovery, pages 78–87, Heidelberg, Germany. Springer Verlag.