Exceptional Model Mining in Ubiquitous and Social Environments

Martin Atzmueller M.ATZMULLER@UVT.NL
Tilburg University (TiCC), Warandelaan 2, 5037 AB Tilburg, The Netherlands

Keywords: exceptional model mining, subgroup discovery, community detection, social interaction networks

Abstract

Exceptional model mining in ubiquitous and social environments includes the analysis of resources created by humans (e. g., social media) as well as those generated by sensor devices in the context of (complex) interactions. This paper provides a structured overview on a line of work comprising a set of papers that focus on local exceptionality detection in ubiquitous and social environments and according complex social interaction networks.

1. Introduction

In ubiquitous and social environments, a variety of heterogenous multi-relational data is generated by sensors and social media (Atzmueller, 2012a). Then, a set of complex social interaction networks (Atzmueller, 2014), capturing distinct facets of the interaction space (Mitzlaff et al., 2014). Here, *local exceptionality detection* – based on *subgroup discovery* (Klösgen, 1996; Wrobel, 1997; Atzmueller, 2015) and *exceptional model mining* – provides flexible approaches for data exploration, assessment, and the detection of unexpected and interesting phenomena.

Subgroup discovery is an approach for discovering interesting subgroups – as an instance of *local pattern detection* (Morik, 2002). The interestingness is usually defined by a certain property of interest formalized by a quality function. In the simplest case, a binary target variable is considered, where the share in a subgroup can be compared to the share in the dataset in order to detect (exceptional) deviations. More complex target concepts consider sets of target variables. In particular, *exceptional model mining* (Leman et al., 2008; Duivesteijn et al., 2012; Duivesteijn et al., 2016) focuses on more complex quality functions.

Appearing in *Proceedings of Benelearn 2017*.

As a revision of (Atzmueller, 2016b), this paper summarizes formalizations and applications of subgroup discovery and exceptional model mining in the context of social interaction networks.

2. Methods

Social interaction networks (Atzmueller, 2014; Mitzlaff et al., 2011; Mitzlaff et al., 2013) focus on user-related social networks in social media capturing social relations inherent in social interactions, social activities and other social phenomena which act as proxies for social user-relatedness.

Exploratory data analysis is an important approach, e. g., for getting first insights into the data. In particular, descriptive data mining aims to uncover certain patterns for characterization and description of the data and the captured relations. Typically, the goal of description-oriented methods is not only to find an actionable model, but also a human interpretable set of patterns (Mannila, 2000).

Subgroup discovery and exceptional model mining are prominent methods for local exceptionality detection that can be configured and adapted to various analytical tasks. Local exceptionality detection especially supports the goal of explanation-aware data mining (Atzmueller & Roth-Berghofer, 2010), due to its more interpretable results, e. g., for characterizing a set of data, for concept description, for providing regularities and associations between elements in general, and for detecting and characterizing unexpected situations, e. g., events or episodes. In the following, we summarize approaches and methods for local exceptionality detection on attributed graphs, for behavioral characterization, and spatio-temporal analysis. Furthermore, we address issues of scalability and large-scale data processing.

Preprint of: Atzmueller M (2017) Exceptional Model Mining in Ubiquitous and Social Environments. Appearing in: Proceedings Benelearn 2017. Eindhoven, The Netherlands, 2017

2.1. Descriptive Community Detection

Communities can intuitively be defined as subsets of nodes of a graph with a dense structure in the corresponding subgraph. However, for mining such communities usually only structural aspects are taken into account. Typically, no concise nor easily interpretable community description is provided.

In (Atzmueller et al., 2016a), we focus on description-oriented community detection using subgroup discovery. For providing both structurally valid and interpretable communities we utilize the graph structure as well as additional descriptive features of the graph's nodes. We aim at identifying communities according to standard community quality measures, while providing characteristic descriptions at the same time. We propose several optimistic estimates of standard community quality functions to be used for efficient pruning of the search space in an exhaustive branch-and-bound algorithm. We present examples of an evaluation using five real-world data sets, obtained from three different social media applications, showing runtime improvements of several orders of magnitude. The results also indicate significant semantic structures compared to the baselines. A further application of this method to the exploratory analysis of social media using geo-references is demonstrated in (Atzmueller, 2014; Atzmueller & Lemmerich, 2013). Furthermore, a scalable implementation of the described description-oriented community detection approach is given in (Atzmueller et al., 2016b), which is also suited for large-scale data processing utilizing the Map/Reduce framework (Dean & Ghemawat, 2008).

2.2. Characterization of Social Behavior

Important structures that emerge in social interaction networks are given by subgroups. As outlined above, we can apply community detection in order to mine both the graph structure and descriptive features in order to obtain description-oriented communities. However, we can also analyze subgroups in a social interaction network from a compositional perspective, i. e., neglecting the graph structure. Then, we focus on the attributes of subsets of nodes or on derived parameters of these, e. g., corresponding to roles, centrality scores, etc. In addition, we can also consider sequential data, e. g., for characterization of exceptional link trails, i. e., sequential transitions, as presented in (Atzmueller, 2016a).

In (Atzmueller, 2012b), we discuss a number of exemplary analysis results of social behavior in mobile social networks, focusing on the characterization of links and roles. For that, we describe the configuration, adaptation and extension of the subgroup discovery methodology in that context. In addition, we can analyze multiplex networks by considering the match between different networks, and deviations between the networks, respectively. Outlining these examples, we demonstrate that local exceptionality detection is a flexible approach for compositional analysis in social interaction networks.

2.3. Exceptional Model Mining for Spatio-Temporal Analysis

Exploratory analysis on ubiquitous data needs to handle different heterogenous and complex data types. In (Atzmueller, 2014; Atzmueller et al., 2015), we present an adaptation of subgroup discovery using exceptional model mining formalizations on ubiquitous social interaction networks. Then, we can detect locally exceptional patterns, e. g., corresponding to bursts or special events in a dynamic network. Furthermore, we propose subgroup discovery and assessment approaches for obtaining interesting descriptive patterns and provide a novel graph-based analysis approach for assessing the relations between the obtained subgroup set. This exploratory visualization approaches allows for the comparison of subgroups according to their relations to other subgroups and to include further parameters, e. g., geo-spatial distribution indicators. We present and discuss analysis results utilizing a real-world ubiquitous social media dataset.

3. Conclusions and Outlook

Subgroup discovery and exceptional model mining provide powerful and comprehensive methods for knowledge discovery and exploratory analyis in the context of local exceptionality detection. In this paper, we presented according approaches and methods, specifically targeting social interaction networks, and showed how to implement local exceptionality detection on both a methodological and practical level.

Interesting future directions for local exceptionality detection in social contexts include extended postprocessing, presentation and assessment options, e. g., (Atzmueller et al., 2006; Atzmueller & Puppe, 2008; Atzmueller, 2015). In addition, extensions to predictive modeling, e. g., link prediction (Scholz et al., 2013; Atzmueller, 2014) are interesting options to explore. Furthermore, extending the analysis of sequential data, e. g., based on Markov chains as exceptional models (Atzmueller et al., 2016c; Atzmueller, 2016a; Atzmueller et al., 2017), as well as group and network dynamics (Atzmueller et al., 2014; Kibanov et al., 2014) are further interesting options for future work.

References

Atzmueller, M. (2012a). Mining Social Media. *Informatik Spektrum*, *35*, 132 – 135.

Atzmueller, M. (2012b). Mining Social Media: Key Players, Sentiments, and Communities. *WIREs Data Mining and Knowledge Discovery*, *2*, 411–419.

Atzmueller, M. (2014). Data Mining on Social Interaction Networks. *JDMDH*, *1*.

Atzmueller, M. (2015). Subgroup Discovery. *WIREs Data Mining and Knowledge Discovery*, *5*, 35–49.

Atzmueller, M. (2016a). Detecting Community Patterns Capturing Exceptional Link Trails. *IEEE/ACM ASONAM*. Boston, MA, USA: IEEE.

Atzmueller, M. (2016b). Local Exceptionality Detection on Social Interaction Networks. *ECML-PKDD 2016* (pp. 485–488). Springer.

Atzmueller, M., Baumeister, J., & Puppe, F. (2006). Introspective Subgroup Analysis for Interactive Knowledge Refinement. *AAAI FLAIRS* (pp. 402–407). AAAI Press.

Atzmueller, M., Doerfel, S., & Mitzlaff, F. (2016a). Description-Oriented Community Detection using Exhaustive Subgroup Discovery. *Information Sciences*, *329*, 965–984.

Atzmueller, M., Ernst, A., Krebs, F., Scholz, C., & Stumme, G. (2014). On the Evolution of Social Groups During Coffee Breaks. *WWW 2014 (Companion)*. New York, NY, USA: ACM Press.

Atzmueller, M., & Lemmerich, F. (2013). Exploratory Pattern Mining on Social Media using Geo-References and Social Tagging Information. *IJWS*, *2*.

Atzmueller, M., Mollenhauer, D., & Schmidt, A. (2016b). Big Data Analytics Using Local Exceptionality Detection. In *Enterprise Big Data Engineering, Analytics, and Management*. IGI Global.

Atzmueller, M., Mueller, J., & Becker, M. (2015). *Exploratory Subgroup Analytics on Ubiquitous Data*, vol. 8940 of *LNAI*. Heidelberg, Germany: Springer.

Atzmueller, M., & Puppe, F. (2008). A Case-Based Approach for Characterization and Analysis of Subgroup Patterns. *Applied Intelligence*, *28*, 210–221.

Atzmueller, M., & Roth-Berghofer, T. (2010). The Mining and Analysis Continuum of Explaining Uncovered. *AI-2010*. London, UK: SGAI.

Atzmueller, M., Schmidt, A., & Kibanov, M. (2016c). DASHTrails: An Approach for Modeling and Analysis of Distribution-Adapted Sequential Hypotheses and Trails. *WWW 2016 (Companion)*. ACM Press.

Atzmueller, M., Schmidt, A., Kloepper, B., & Arnu, D. (2017). HypGraphs: An Approach for Analysis and Assessment of Graph-Based and Sequential Hypotheses. *New Frontiers in Mining Complex Patterns*. Heidelberg, Germany: Springer.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*, *51*, 107–113.

Duivesteijn, W., Feelders, A., & Knobbe, A. J. (2012). Different Slopes for Different Folks: Mining for Exceptional Regression Models with Cook's Distance. *ICDM* (pp. 868–876). ACM Press, New York.

Duivesteijn, W., Feelders, A. J., & Knobbe, A. (2016). Exceptional Model Mining. *DMKD*, *30*, 47–98.

Kibanov, M., Atzmueller, M., Scholz, C., & Stumme, G. (2014). Temporal Evolution of Contacts and Communities in Networks of Face-to-Face Human Interactions. *Sci China Information Sciences*, *57*.

Klösgen, W. (1996). Explora: A Multipattern and Multistrategy Discovery Assistant. In *Advances in Knowledge Discovery and Data Mining*. AAAI.

Leman, D., Feelders, A., & Knobbe, A. (2008). Exceptional Model Mining. *PKDD* (pp. 1–16). Springer.

Mannila, H. (2000). Theoretical Frameworks for Data Mining. *SIGKDD Explor.*, *1*, 30–32.

Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., & Stumme, G. (2011). *Community Assessment using Evidence Networks*, vol. 6904 of *LNAI*. Springer.

Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., & Stumme, G. (2013). User-Relatedness and Community Structure in Social Interaction Networks. *CoRR/abs*, *1309.3888*.

Mitzlaff, F., Atzmueller, M., Hotho, A., & Stumme, G. (2014). The Social Distributional Hypothesis. *Journal of Social Network Analysis and Mining*, *4*.

Morik, K. (2002). *Detecting Interesting Instances*, vol. 2447 of *LNCS*, 13–23. Springer Berlin Heidelberg.

Scholz, C., Atzmueller, M., Barrat, A., Cattuto, C., & Stumme, G. (2013). New Insights and Methods For Predicting Face-To-Face Contacts. *ICWSM*. AAAI.

Wrobel, S. (1997). An Algorithm for Multi-Relational Discovery of Subgroups. *Proc. PKDD-97* (pp. 78–87). Heidelberg, Germany: Springer.