Anomaly Detection and Structural Analysis in Industrial Production Environments

Martin Atzmueller
Tilburg University (TiCC),
Jheronimus Academy of Data
Science, University of Kassel (ITeG)
m.atzmueller@uvt.nl

David Arnu
RapidMiner GmbH
Dortmund, Germany
darnu@rapidminer.com

Andreas Schmidt
University of Kassel (ITeG)
Kassel, Germany
schmidt@cs.uni-kassel.de

***Abstract*—Detecting anomalous behavior can be of critical importance in an industrial application context. While modern production sites feature sophisticated alarm management systems, they mostly react to single events. Due to the large number and various types of data sources a unified approach for anomaly detection is not always feasible. One prominent type of data are log entries of alarm messages. They allow a higher level of abstraction compared to raw sensor readings. In an industrial production scenario, we utilize sequential alarm data for anomaly detection and analysis, based on first-order Markov chain models. We outline hypothesis-driven and description-oriented modeling options. Furthermore, we provide an interactive dashboard for exploring and visualization of the results.**

Keywords—anomaly detection; exceptional model mining; sequence mining; sequential patterns; industry 4.0

I. INTRODUCTION

In many industrial areas, production facilities have reached a high level of automation: sensor readings are constantly analyzed and may trigger various forms of alarms. Hence, knowledge about the respective processes is crucial, e.g., targeting the topological structure of a plant, sequences of operator notifications (alarms), and unexpected (critical) situations. Then, the analysis of (exceptional) sequential patterns is an important task for obtaining insights into the process and for modelling predictive applications. The research project *Early detection and decision support for critical situations in production environments* (short FEE) aims at detecting critical situations in production environments as early as possible and to support the facility operator in handling these situations, e.g., [14]. In abnormal situations, typically such a large number of notifications is generated, that it often cannot be physically assessed by the operator [2]. Therefore, appropriate abstractions and analytics methods are necessary to adapt from a reactive to a proactive behavior. The consortium of the FEE project consists of several partners also including application partners from the chemical industry. These partners provide the use cases for the project and background knowledge about the production process which is important for designing suitable analytical methods.

This paper presents the implementation of a comprehensive modelling approach for anomaly detection and analysis of observed "reference" sequential patterns, based on methods for modelling and comparing hypotheses on sequence data [16, 17]. Implemented as a new RapidMiner operator and embedded in an analytical process, we demonstrate its application.

II. RELATED WORK

The investigation of sequential patterns and sequential trails are interesting and challenging tasks in data mining and network science, in particular in graph mining and social network analysis, e.g., [5, 9]. A general view on modeling and mining of ubiquitous and social multi-relational data is given in [5] focusing on social interaction networks. Here, dynamics and evolution of contact patterns [9, 23, 29], for example, and their underlying mechanisms, e.g., [33] are analyzed. However, the analysis in these contexts focuses on aggregated sequential data. Navigational patterns, as sequential (link) patterns in online systems, have been analyzed and modeled, e.g., in [35, 40]. In contrast to that, our approach focuses on the modeling and comparing sequential patterns (hypothesis) in a graph-based network representation.

In a previous work [16] we have presented the DASHTrails approach that incorporates probability distributions for deriving transitions utilizing HypTrails [39]. Based on that, the HypGraphs framework [17] provides a more general modeling approach. Using general weight-attributed network representations, we can infer transition matrices as graph interpretations; HypGraphs consequently also relies on Markov chain modeling [29, 35] and Bayesian inference [35, 41].

Sequential pattern analysis has also been performed in the context of alarm management systems, where sequences are represented by the order of alarm notifications. Folmer et al. [21] proposed an algorithm for discovering temporal alarm dependencies based on conditional probabilities in an adjustable time window. To reduce the number of alarms in alarm floods Abele et al. [2] performed root cause analysis with a Bayesian network approach and compared different methods for learning the network probabilities. Vogel-Heuser et al. [41] proposed a pattern-based algorithm for identifying causal dependencies in the alarm logs, which can be used to aggregate alarm information and therefore reduce the load of information for the operator. In contrast to those approaches, we provide a systematic approach for the analysis of sequential transition matrices and its comparison relative to a set of hypotheses. Thus, similar to evidence networks in the context of social networks, e.g., [32], we model transitions assuming a certain interpretation of the data towards a sequential representation. Then, we can identify important influence factors.

This work was funded by the BMBF project FEE under grant number 01IS14006

Preprint of: Atzmueller M, Arnu D, Schmidt A (2017)
Anomaly Detection and Structural Analysis in Industrial Production Environments.
In: Proc. International Data Science Conference (IDSC 2017), Salzburg, Austria

Process Mining [1] aims at the discovery of business process related events in a sequence log. The assumption is that event logs contain fingerprints of a business process, which can be identified by sequence analysis. One task of process mining is conformance checking [34, 37] which has been introduced to check the matching of an existing business process model with the segmentation of the log entries. Compared to these approaches, we do not use any apriori knowledge about business processes to create our hypothesis. Also, our hypothesis does not necessarily need to conform to an existing business process.

One definition of an anomaly or outlier is, that "an outlier is an observation that differs so much from other observations as to arouse suspicion that it was generated by a different mechanism". Thus, we understand an anomaly as a real-world situation, that could be represented as one or more outliers. In the literature those two terms are often used interchangeably. Then, interesting, important or exceptional groups [36] can be identified. Classic approaches for anomaly detection provide a classification of anomalous and normal events. In the industrial context finding either significant changes in the multivariate sensor readings or managing hundreds of univariate scores for single sensors is also a challenge, e.g., described in [30]. In contrast to approaches for anomaly detection that only provide a classification of anomalous and normal events, we can assess different anomaly hypotheses: Applying the proposed approach, we can then generate an anomaly indicator – as a potential kind of second opinion method for assessing the state of a production plant that can help for indicating explanations [15] and traces of unusual alarm sequences in the plant. Also, using the network representation, we can analyze anomalous episodes relative to structural (plant topology) as well as dynamic (alarm sequence) episodes.

III. METHOD

The detection and analysis of irregular or exceptional patterns, i.e., anomalies, in complex-structured heterogeneous data is a novel research area, e.g., for identifying new and/or emerging behavior, or for identifying detrimental or malicious activities. The former can be used for deriving new information and knowledge from the data, for identifying events in time or space, or for identifying interesting, important or exceptional groups. In this paper, we focus on a combined detection and analysis approach utilizing heterogeneous data. That is, we include semi-structured, as well as structured data for enhancing the analysis. Furthermore, we also outline a description-oriented technique that does not only allow the detection of the anomalous patterns, but also its description using a given set of features. The latter relates to the context of descriptive pattern mining. In particular, the concept of exceptional model mining, e.g., [8, 25, 27] suitably enables such description-oriented approaches, adapting methods for the detection of interesting subgroups (that is, subgroup discovery) with more advanced target concepts for identifying exceptional (anomalous) groups.

In our application context of an industrial production plants in an Industry 4.0 context, cf., [20, 42], we based our anomaly detection system on the analysis of the plant topology and alarm logs as well as on the similarity based analysis of metric sensor readings. The combined approach will compare the insights of the two methods.

1) *Anomaly Analytics on Sequential Data*

For sequential data, we formulate the "reference behavior" by collecting episodes of normal situations, which is typically observed for long running processes. Episodes of alarm sequences (formulated as hypotheses) can be compared to the normal situations in order to detect deviations, i.e., abnormal episodes. We map these sequences to transitions between functional units of an industrial plant, applying the modeling approach described below. The results can also be used for diagnostics, by inspecting the transitions in detail.

Following HypGraphs [18] and DASHTrails [17], we model transition matrices given a probability distribution of certain states. The steps we need to perform, as shown in Fig. 1, are:

1) Modeling: Determine a transition model given the respective weighted network using a transition modeling function $\tau : \Omega \times \Omega \rightarrow R$. Transitions between sequential states $i,j \in \Omega$ are captured by the elements $m_{i,j}$ of the transition matrix M, i.e., $m_{i,j} = \tau(i,j)$. Then, we collect sequential transition matrices for the given network (data) and hypotheses.
2) Estimation: Apply HypTrails [39] on the given data transition matrix and the respective hypotheses, and return the resulting evidence.
3) Analysis: Present the results for semi-automatic introspection and analysis, e.g., by visualizing the network as a heatmap or characteristic sequence of nodes.

We can model (derived) transition matrices corresponding to the observed data, e.g., given frequencies of alarms on measurement points, as well as hypotheses on sequences of alarms. For data transition matrices, we need to map the transitions into derived counts in relation to the data; for hypotheses, we provide (normalized) transition probabilities.
In summary, we utilize Bayesian inference on a first-order Markov chain model. As an input, we provide a (data) matrix, containing the transitional information (frequencies) of transition between the respective states, according to the (observed) data. In addition, we utilize a set of hypotheses given by (row-normalized) stochastic matrices, modelling the given hypotheses. The estimation method outputs an evidence value, for each hypothesis, that can be used for ranking. Also, using

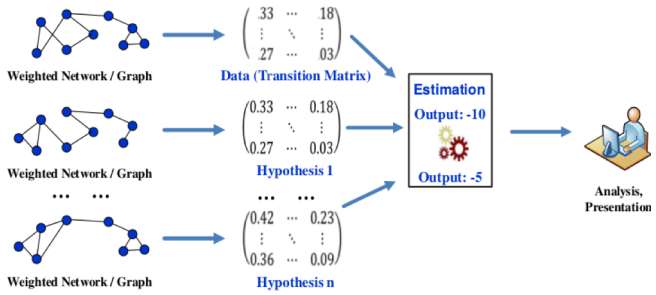

Figure 1: Overview on the HypGraphs modeling and analysis process [17].

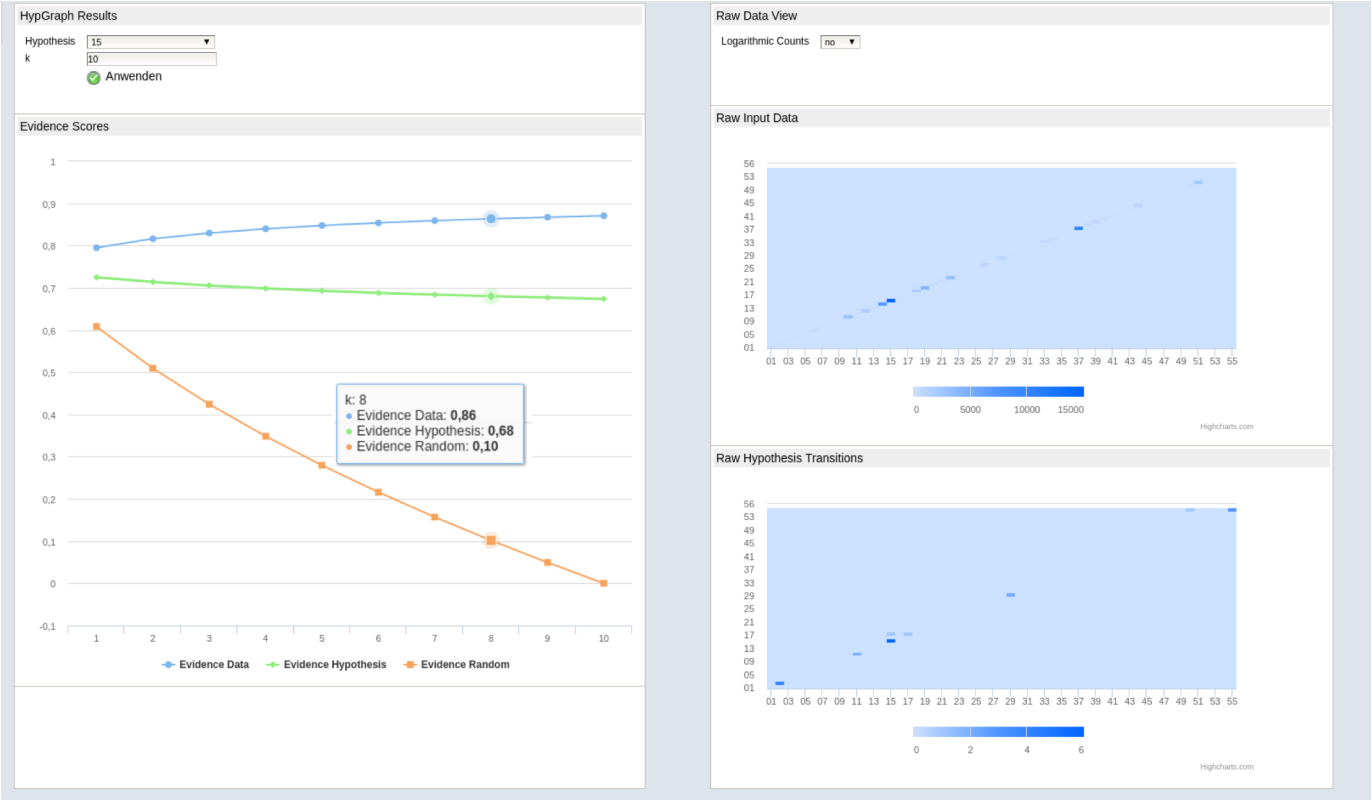

Figure 2: RapidMiner Dashboard showing the HypGraphs transition scores (left) and the raw transition matrices (right), as transitions between different components of a plant visualized as a heatmap. It is possible to select a specific hypothesis for which the evidence scores are calculated and displayed.

the evidence values, we can compare the hypotheses in terms of their significance.

For modeling, we use the freely available RapidMiner [31] extension of HypGraphs[1], that calculates the evidence values for different believe weights *k* and compares them directly with the given hypothesis and a random transition as a lower bound. The evidence scores and transition matrices are displayed in an interactive dashboard, as shown in Figure 2.

As an extension to the hypothesis-based approach, we can furthermore include descriptive information, that is, features of the dataset for identifying patterns capturing anomalous behavior [11]. For that, we can consider the transition matrix as a graph, for which we can then we include node and/or edge labels. That is, the edges of the graph (modeling specific transitions between nodes) are labeled according to descriptive properties, e.g., capturing properties of the specific sequences the transitions were derived from. Then, using a specific set of labels we can select a set of edges, that is, all edges having the respective label set, inducing a subgraph which corresponds to a set of transitions having the respective labeling. Then, we can define an anomaly pattern as the respective label set and its corresponding (induced) subgraph, covering a subset of nodes and set transitions, respectively. In that way, we can not only include anomalous episodes in the sequential anomaly detection step, but we can include descriptive information for enabling further inspection, explanation and/or exemplification [10, 15] by the operator or the process engineer. Thus, the descriptive

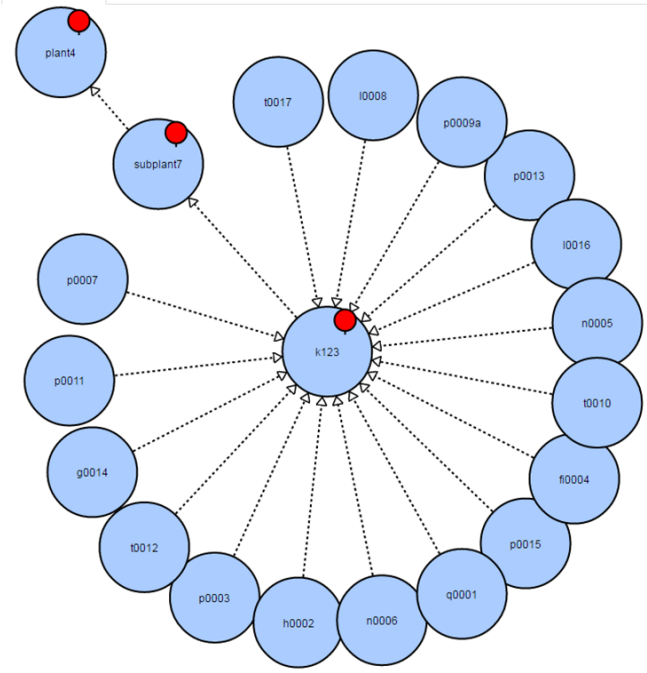

Figure 3: Example of a (conceptual) knowledge graph [14].

[1] https://github.com/rapidminer/rapidminer-extension-hypgraphs/releases

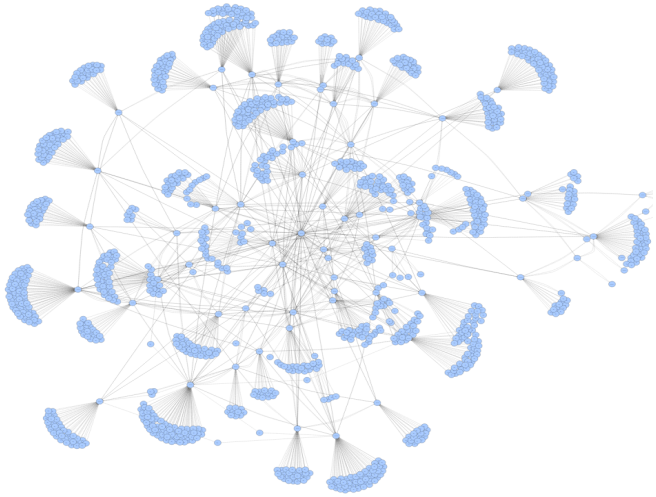

Figure 4: Larger (anonymized) example of a knowledge graph.

features can, e.g., indicate important indicators for anomalies like sensors or alarm related labels as proxies for specific faults. The descriptive information can, for example, be derived by the integration of unstructured information such as plant topology information derived from a Knowledge Graph [14]. Such a graph can be constructed from heterogeneous information, such as sensor measurements, alarm logs, process and instrumentation diagrams (P&IDs), shift books, operation manuals etc. Figure 3 shows an abstracted example of a knowledge graph showing the conceptual plant-subplant relations and measurements [14], while Figure 4 shows an anonymized example of a larger plant context.

2) *Anomaly Detection on Metric Data*

For detecting outliers on the numeric sensor data we apply the Local Outlier Factor (LOF) algorithm by Breunig et al. [18], as implemented in the Anomaly Detection extension [31] for RapidMiner. The algorithm estimates local deviations of the data points using a defined distance function. It compares the local density of a point to the density of its *k* nearest neighbors. Due to the nature of the provided sensor data, the concept of a locally sensitive algorithm is useful, because with different set points (for plant operation) range and characteristics of the sensor readings can vary greatly. The outlier scores can be calculated for either all available sensors, for certain subgroups, or single sensors, depending on the desired granularity.

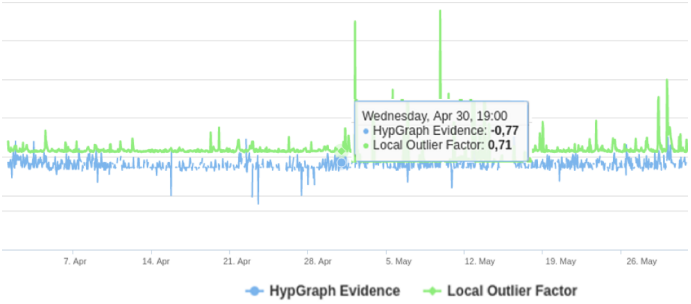

Figure 5: Overview of outlier scores and HypGraphs evidence values. A high outlier score indicates anomalous sensor readings, while a low the evidence score indicates deviating alarm sequences

IV. Process Model & Implementation

Distributed storage and computation systems, can handle the evaluation of several years of production data. In the context of the FEE project we want to build upon a two-layered computation architecture to enable the plant operators and system engineers to design and test their processes on a local machine and execute memory and computational intensive processes in the Hadoop infrastructure [19, 24].

The first part of the analytical workflow is to build the transition network for training and testing the hypotheses. We build these hypotheses on real plant data and calculate the transition matrices for hourly time slots over a period of two months. In the same way, after further preprocessing (smoothing and down-sampling) we aggregate the raw sensor data. The calculated outlier score is shown in Figure 5, together with the evidence scores. A high outlier score indicates possible anomalous sensor readings and a low evidence score indicates deviating transition patterns in the alarm sequences. As one can see, the two values are apparently not strongly correlated. But this shows, that the two algorithms monitor different aspects of the plant behavior: the lowest evidence scores occur in the early morning hours with nearly no alarm transitions (when normally several dozen alarms are recorded per hour).

For further inspecting the outlier scores, we have built another dashboard. This shows the *k* highest outlier score for single sensor readings for a selected time segment, for example by clicking on a specific time point in the dashboard from Figure 5. In addition, this board shows the associated sensor readings. With this drill-down from a high level of abstraction for a whole processing unit down to single sensor readings, a process engineer is able to identify and inspect possible critical situations in a convenient way.

V. Conclusion and Future Work

This paper presented a sequential modelling and anomaly analytics approach in an industrial application context. Based on first order Markov chain models and methods for modelling and comparing networks and transition matrices, we sketched an approach for comparing hypotheses with observed “reference” sequential patterns. Furthermore, we described the extension to integrating descriptive information in the sequential modeling approach. In addition, we also showed a comparison between our approach and an established outlier detection algorithm. It became evident that both methods target different aspects of detecting anomalous behavior.

For future work, we aim at extending the proposed approach by integrating the knowledge gained from the conceptual knowledge graph, e.g., by grouping and analyzing the outlier scores for the sensors associated with specific functional units. We also plan to integrate the system into the Big data architecture proposed in [24]. As outlined above, we want to extend that two-layered computation architecture for enabling flexible and powerful Big data processing approaches, also including large-scale descriptive subgroup [26], sequence [38, 43], and graph mining methods [13] for efficient exceptional model mining and anomaly analytics.

REFERENCES

[1] Aalst, W.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Berlin (2011)

[2] Abele, L., Anic, M., Gutmann, T., Folmer, J., Kleinsteuber, M., Vogel-Heuser, B.: Combining Knowledge Modeling and Machine Learning for Alarm Root Cause Analysis. In: Proc. IFAC Volumes, 46(9):1843–1848. International Federation of Automatic Control (2013)

[3] Akoglu, L., Tong, H., Koutra, D.: Graph Based Anomaly Detection and Description. Data Min Knowl Disc 29(3), 626–688 (May 2015)

[4] Amer, M., Goldstein, M.: Nearest-Neighbor and Clustering-based Anomaly Detection Algorithms for Rapidminer. In: Proc. of the 3rd RapidMiner Community Meeting and Conference (RCOMM 2012). pp. 1–12 (2012)

[5] Atzmueller, M.: Analyzing and Grounding Social Interaction in Online and Offline Networks. In: Proc. ECML PKDD. LNCS, vol. 8726, pp. 485–488. Springer, Heidelberg, Germany (2014)

[6] Atzmueller, M.: Data Mining on Social Interaction Networks. Journal of Data Mining and Digital Humanities 1 (June 2014)

[7] Atzmueller, M.: Subgroup Discovery - Advanced Review. WIREs: Data Mining and Knowledge Discovery, (5)1:35—49 (2015)

[8] Atzmueller, M.: Detecting Community Patterns Capturing Exceptional Link Trails. Proc. IEEE/ACM ASONAM, IEEE Press, Boston, MA, USA (2016)

[9] Atzmueller, M.: Local Exceptionality Detection on Social Interaction Networks. In: Proc. ECML PKDD 2016: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Springer, Heidelberg, Germany (2016)

[10] Atzmueller, M., Baumeister, J., Puppe, F.: Introspective Subgroup Analysis for Interactive Knowledge Refinement. In: Proc. FLAIRS Conference, pp. 402-407, AAAI Press, Palo Alto, CA, USA (2006)

[11] Atzmueller, M., Doerfel, S., Mitzlaff, F.: Description-Oriented Community Detection using Exhaustive Subgroup Discovery. Information Sciences, 329, 965-984. (2016)

[12] Atzmueller, M., Doerfel, S., Hotho, A., Mitzlaff, F., Stumme, G.: Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles. In: Modeling and Mining Ubiquitous Social Media, LNAI, vol. 7472. Springer, Heidelberg, Germany (2012)

[13] Atzmueller, M., Mollenhauer, D., Schmidt, A.: Big Data Analytics Using Local Exceptionality Detection. In: Enterprise Big Data Engineering, Analytics, and Management, IGI Global, Hershey, PA, USA, 2016.

[14] Atzmueller, M., Kloepper, B., Mawla, H.A., Jäschke, B., Hollender, M., Graube, M., Arnu, D., Schmidt, A., Heinze, S., Schorer, L., Kroll, A., Stumme, G., Urbas, L.: Big Data Analytics for Proactive Industrial Decision Support: Approaches First Experiences in the Context of the FEE Project. atp edition 58(9):62-74 (2016)

[15] Atzmueller, M., Roth-Berghofer, T.: The Mining and Analysis Continuum of Explaining Uncovered. Proc. 30th SGAI International Conference on Artificial Intelligence (2010)

[16] Atzmueller M, Schmidt A, Kibanov M. DASHTrails: An Approach for Modeling and Analysis of Distribution-Adapted Sequential Hypotheses and Trails. In: Proc. WWW 2016 (Companion). IW3C2 / ACM, New York, NY, USA (2016)

[17] Atzmueller M., Schmidt A., Kloepper B., Arnu D.: HypGraphs: An Approach for Modeling and Comparing Graph-Based and Sequential Hypotheses. In: Proc. ECML PKDD Workshop on New Frontiers in Mining Complex Patterns (NFMCP). Riva del Garda, Italy (2016).

[18] Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: OPTICS-OF: Identifying Local Outliers, pp. 262–270. Springer, Berlin/Heidelberg (1999)

[19] Dean, J., Ghemawat, S.: Mapreduce: Simplified Data Processing on Large Clusters. Communications of the ACM 51(1), 107–113 (2008).

[20] Folmer, J., Kirchen, I., Trunzer, E., Vogel-Heuser, B., Pötter, T., Graube, M., Heinze, S., Urbas, L., Atzmueller, M., Arnu, D: Challenges for Big and Smart Data in Process Industries. *atp edition*, 01/02 (2017)

[21] Folmer, J., Schuricht, F., Vogel-Heuser, B.: Detection of Remporal Dependencies in Alarm Time Series of Industrial Plants. Proc. IFAC, pp. 24–29, International Federation of Automatic Control (2014)

[22] Hawkins,D.: Identification of Outliers. Chapman and Hall, London, UK (1980)

[23] Kibanov, M., Atzmueller, M., Scholz, C., Stumme, G.: Temporal Evolution of Contacts and Communities in Networks of Face-to-Face Human Interactions. Science China Information Sciences 57 (2014)

[24] Klöpper, B., Dix, M., Schorer, L, Ampofo, A., Atzmueller, M., Arnu, D., Klinkenberg, R.: Defining Software Architectures for Big Data Enabled Operator Support Systems. In: Proc. INDIN. IEEE Press, Boston, MA, USA (2016)

[25] Leman, D., Feelders, A., Knobbe, A.: Exceptional Model Mining. In: Proc. ECML PKDD, pp. 1-16, Springer, Heidelberg, Germany (2008)

[26] Lemmerich, M., Atzmueller, M., Puppe, F.: Fast Exhaustive Subgroup Discovery with Numerical Target Concepts. Data Mining and Knowledge Discovery, (30):711-762 (2016)

[27] Lemmerich, M., Becker, M., Atzmueller, M.: Generic Pattern Trees for Exhaustive Exceptional Model Mining. Proc. ECML PKDD 2012, pp. 277-292, Springer, Heidelberg, Germany (2012)

[28] Lempel, R., Moran, S.: The Stochastic Approach for Link-Structure Analysis (SALSA) and the TKC Effect. Computer Networks 33(1), 387–401 (2000)

[29] Macek, B.E., Scholz, C., Atzmueller, M., Stumme, G.: Anatomy of a Conference. In: Proc. ACM Hypertext. pp. 245–254. ACM Press, New York, NY, USA (2012)

[30] Martí, L., Sanchez-Pi, N., Molina, J.M., Garcia, A.C.B.: Anomaly Detection based on Sensor Data in Petroleum Industry Applications. Sensors 15(2), 2774–2797 (2015)

[31] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T.: Yale: Rapid Prototyping for Complex Data Mining Tasks. In: Proc. KDD. pp. 935–940. ACM, New York, NY, USA (2006)

[32] Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., Stumme, G.: Community Assessment using Evidence Networks. In: Analysis of Social Media and Ubiquitous Data. LNAI, vol. 6904 (2011)

[33] Mitzlaff, F., Atzmueller, M., Hotho, A., Stumme, G.: The Social Distributional Hypothesis. SNAM 4(216) (2014)

[34] Munoz-Gama, J., Carmona, J., van der Aalst, W.M.P.: Single-Entry Single-Exit Decomposed Conformance Checking. Inf. Syst. 46, 102–122 (2014)

[35] Pirolli, P.L., Pitkow, J.E.: Distributions of Surfers' Paths Through the World Wide Web: Empirical Characterizations. WWW 2(1-2) (1999)

[36] Ranshous, S., Shen, S., Koutra, D., Harenberg, S., Faloutsos, C., Samatova, N.F.: Anomaly Detection in Dynamic Networks: A Survey. WIREs: Comput. Statistics 7(3), 223–247 (2015)

[37] Rozinat, A., Aalst, W.: Conformance Checking of Processes Based on Monitoring Real Behavior. Information Systems 33(1), 64–95 (2008)

[38] Seipel, D., Köhler, S., Neubeck, P., Atzmueller, M.: Mining Complex Event Patterns in Computer Networks. In: New Frontiers in Mining Complex Patterns (NFMCP), Springer, Heidelberg, Germany (2013)

[39] Singer, P., Helic, D., Hotho, A., Strohmaier, M.: Hyptrails: A Bayesian Approach for Comparing Hypotheses about Human Trails. In: Proc. WWW. ACM, New York, NY, USA (2015)

[40] Singer, P., Helic, D., Taraghi, B., Strohmaier, M.:Memory and Structure in Human Navigation Patterns. PLoS ONE 9(7) (2014)

[41] Strelioff, C.C., Crutchfield, J.P., Hübler, A.W.: Inferring Markov Chains: Bayesian Estimation, Model Comparison, Entropy Rate, and Out-of-Class Modeling. Physical Review E 76(1), 011106 (2007)

[42] Vogel-Heuser, B., Schütz, D., Folmer, J.: Criteria-based alarm flood pattern recognition using historical data from automated production systems (aps). Mechatronics 31, 89–100

[43] Weiss, C. H., Atzmueller, M.: EWMA Control Charts for Monitoring Binary Processes with Applications to Medical Diagnosis Data. Qual. Reliab. Engng. Int., 26: 795–805 (2010)