Mining Behavioural Patterns from Event Data to Enable Context-Aware Root Cause Analysis (Extended Abstract)

Greg Van Houdt
Faculty of Business Economics
UHasselt - Hasselt University
Hasselt, Belgium
greg.vanhoudt@uhasselt.be

I. Introduction

Process mining [1] concentrates on the extraction of insights and knowledge on business processes from event logs. With businesses evolving into a more digital environment due to, for example, the Internet of Things (IoT), it has become easier than ever to capture event data [2]–[4]. This poses a great advantage as event data is actionable and has strong explanatory power, as seen in process discovery [5].

Unfortunately, there are drawbacks to this digital environment as well. First of all, IoT may enhance data collection in terms of volume, but also granularity. Nevertheless, there is no guarantee that business users view a process on the same level as data collection occurs on. Second, the volume of data available can quickly overwhelm analysts in understanding the order of and reasons behind the events. This is due to a critical characteristic of IoT sensor data: the lack of an explicit case notion. It adds additional complexity to the analyses.
Ultimately, the large amount of data intrinsically has the potential to understand processes better, especially when combined with root cause analytics. Against this background, the goal of this PhD research project is threefold:

1) Investigating the possibilities of event log abstraction to elevate logs to a higher granularity level;
2) Studying context changes in the data taking into account the relations between interconnected business objects;
3) Exploring feature engineering and selection methods for a diagnostic analysis. This is to aid business users in understanding why particular events take place in their processes and to guide them on how to either stimulate or prevent a new occurrence of such events.

II. Research Gaps

In each of the three aforementioned research goals, several challenges can be identified. The first goal covers event log abstraction or augmenting the event log to a higher granularity level. To start, an evaluation of abstraction quality of existing pattern detection techniques is required. Abstraction utilises input patterns of events to be elevated into a single high-level activity or system state on the one hand, but also an algorithm to replace these patterns in the event log on the other hand. In this stage, it is also vital to keep track of the quality of this high-level event log in terms of fitness, precision and complexity.

The second goal tackles an investigation of context changes in the masses of data. Event data typically originates from process-aware information systems [6] storing a great deal of information. However, capturing *all* data does not mean it is all useable. In most cases, data integration is key to perform the appropriate analysis for a particular problem due to a large possible number of data sources. Only with a correct integration of data can we link all business objects and their corresponding triggered high-level activities. Important elements to consider are anomaly detection [7] and concept drift [8].

Finally, goal three shifts focus to root cause analysis. Despite the importance of feature selection in root cause analysis, there is a lack of attention being attributed to the specification of features [9]–[11]. For example, trends in certain variables or events should also be considered to be a potential feature. An important question to be asked is to which extent this feature selection could be automated and, of course, to which degree the domain expert must be kept in the loop. In that regard, there is a difference between feature selection [12] and feature engineering [13] which remains to be explored in this context.

III. Methodology

This PhD research project shall follow the principles of design science research (DSR). DSR is centred around the development and study of artefacts aiming to solve a problem taking the problem context into account [14]. Additionally, we attempt to draw methodological plans of action from fields in which event data is also known. Examples are visual analytics [15] and complex event processing [16].

For goal 1, we start with an analysis of already developed abstraction techniques on their ability to elevate event logs (in the process mining context) to a higher level of granularity. The next stage would imply moving beyond process mining and generalise to recurrent sequences. Goal 2 is about detecting shifts in the context or behaviour of the system. As a starting point, we can, e.g., evaluate existing concept drift techniques. Alternatively, we can train a generative model on the low-level data on different time windows and check if these models differ in structure significantly. Detected changes can be represented as new events in the data. Finally, in goal 3, we start again with an evaluation of existing techniques to discover areas of improvement regarding their performance in finding root causes. This should be done for both low-level data as well as the abstracted data to see if there are differences in performance. But not only the impact of abstraction should be tested. Attention must be given to the impact of the context change events as well.

This research aims to be valuable not only in theory, but also in practice. To that regard, we strive for a high degree of applicability in industry. This implies that validation in practice is key. To that end, we have identified a number of partners to aid us with validation processes.

IV. Related Work

At the time of writing, there is already a strong basis of event abstraction techniques present. The work of van Zelst et al. [17] provides a taxonomy on these techniques based on several properties, for example, the supervision strategy. We have a profound research interest in the unsupervised techniques. Most interesting to us are the local process models [18], global trace segmentation [19], the RefMod-miner [20], and combination based behavioral pattern mining [21]. One can then utilise the pattern-based abstraction approach designed by Mannhardt et al. [22] to obtain the high-level event log.

Having achieved a more simplified event log by removing part of the clutter, it should be easier to link the higher-level activities with each other in terms of cause-effect relationships. Two domains are interesting to investigate further in this regard, namely visual analytics and root cause analysis. Visual Analytics is a vast domain in the sense that it is a combination of several research areas, namely visualisation, data mining and statistics. At the same time, one cannot forget the importance of the human factor [23]. It can be defined as the science of analytical reasoning facilitated by interactive visual interfaces [15], which makes the way of processing data and information transparent [24]. A root cause is the most fundamental reason for an undesirable condition or problem which, if eliminated or corrected, would have prevented it from existing or occurring. Therefore, the root cause is always negative and usually defined in terms of specific or systematic factors. Root cause analysis can be used to identify the most apparent improvement opportunities by tagging current obstacles to efficient operations or activities [25].

References

[1] W. M. P. van der Aalst, *Process Mining: Data Science in Action*. Berlin, Heidelberg: Springer, 2016.

[2] I. Lee and K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Business Horizons*, 2015.

[3] I. Lee, "The Internet of Things for enterprises: An ecosystem, architecture, and IoT service business model," *Internet of Things*, vol. 7, p. 100078, 9 2019.

[4] D. J. Teece and G. Linden, "Business models, value capture, and the digital enterprise," *Journal of Organization Design*, vol. 6, no. 1, p. 8, 2017.

[5] W. M. P. van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 9, pp. 1128–1142, 2004.

[6] M. Dumas, W. M. van der Aalst, and A. H. M. ter Hofstede, *Process-Aware Information Systems: Bridging People and Software through Process Technology*. Wiley Online Library, 2005.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[8] G. Widmer and M. Kubat, "Learning in the presence of concept drift and hidden contexts," *Machine learning*, vol. 23, no. 1, pp. 69–101, 1996.

[9] S. Suriadi, C. Ouyang, W. M. Van Der Aalst, and A. H. Ter Hofstede, "Root cause analysis with enriched process logs," in *Lecture Notes in Business Information Processing*, 2013.

[10] B. F. Hompes, A. Maaradji, M. La Rosa, M. Dumas, J. C. Buijs, and W. M. Van Der Aalst, "Discovering causal factors explaining business process performance variation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017.

[11] N. Gupta, K. Anand, and A. Sureka, "Pariket: Mining business process logs for root cause analysis of anomalous incidents," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.

[12] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys*, vol. 50, no. 6, p. 94, 2018.

[13] F. Nargesian, H. Samulowitz, U. Khurana, E. B. Khalil, and D. Turaga, "Learning feature engineering for classification," in *IJCAI International Joint Conference on Artificial Intelligence*, 2017.

[14] P. Johannesson and E. Perjons, *An introduction to design science*. Springer, 2014.

[15] J. J. Thomas and K. A. Cook, "Illuminating the path: The research and development agenda for visual analytics," *IEEE Computer Society*, 2005.

[16] D. Luckham, "The power of events: An introduction to complex event processing in distributed enterprise systems," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008.

[17] S. J. van Zelst, F. Mannhardt, M. de Leoni, and A. Koschmider, "Event abstraction in process mining: literature review and taxonomy," *Granular Computing*, May 2020.

[18] N. Tax, N. Sidorova, R. Haakma, and W. M. van der Aalst, "Mining local process models," *Journal of Innovation in Digital Ecosystems*, vol. 3, no. 2, pp. 183–196, Dec. 2016.

[19] C. W. Günther, A. Rozinat, and W. M. Van Der Aalst, "Activity mining by global trace segmentation," in *International Conference on Business Process Management*. Springer, 2009, pp. 128–139.

[20] J.-R. Rehse and P. Fettke, "Clustering Business Process Activities for Identifying Reference Model Components," in *Business Process Management Workshops*, F. Daniel, Q. Z. Sheng, and H. Motahari, Eds. Cham: Springer International Publishing, 2019, vol. 342, pp. 5–17, series Title: Lecture Notes in Business Information Processing.

[21] M. Acheli, D. Grigori, and M. Weidlich, "Efficient Discovery of Compact Maximal Behavioral Patterns from Event Logs," in *Advanced Information Systems Engineering*, P. Giorgini and B. Weber, Eds. Cham: Springer International Publishing, 2019, vol. 11483, pp. 579–594, series Title: Lecture Notes in Computer Science.

[22] F. Mannhardt, M. de Leoni, H. A. Reijers, W. M. P. van der Aalst, and P. J. Toussaint, "From Low-Level Events to Activities - A Pattern-Based Approach," in *Business Process Management*, M. La Rosa, P. Loos, and O. Pastor, Eds. Cham: Springer International Publishing, 2016, vol. 9850, pp. 125–141, series Title: Lecture Notes in Computer Science.

[23] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler, "Challenges in visual data analysis," in *Proceedings of the International Conference on Information Visualisation*, 2006.

[24] D. Keim, G. Andrienko, J. D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual analytics: Definition, process, and challenges," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008.

[25] P. F. Wilson, L. D. Dell, and G. F. Anderson, "Root Cause Analysis: A Tool for Total Quality Management," *Journal For Healthcare Quality*, 1996.