Value Behaviour Analysis of Event Attributes in Process Mining (Extended Abstract)

Jonas Cremerius[1]

[1]Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

1. Introduction

Discovering and analysing business processes are important tasks for organizations. Process mining bridges the gap between process management and data science by discovering process models using event logs derived from real-world data [1]. Besides mandatory event attributes like case identifier, activity, and timestamp, additional event attributes can be present, such as human resources, costs, and laboratory values.

Event attributes represent a valuable data-source to understand the process under investigation. For example, read/write operations of activities can be identified and visualized in a discovered process model [2]. Additionally, event attributes can be used in the field of process outcome prediction as additional features [3].

However, we see a lack in the analysis of the actual value behaviour of event attributes. We argue, that event attributes can be associated to multiple events representing different activities. How event attribute values differ among activities and how they change within the process is what we define as value behaviour analysis.

Especially in healthcare processes, a huge amount of data is generated, such as vital signs or laboratory measurements representing a patient's wellbeing. When analysing treatment processes, one might be interested in how these measurements develop depending on which treatment activities were conducted. This can help to evaluate, for example, different treatment paths in the sense, that one path might result in better patient wellbeing than others.

This problem leads us to the following research question: **How can we analyse the value behaviour of event attributes?**

In this PhD project, I focus on healthcare processes extracted from the Medical Information Mart for Intensive Care (MIMIC-IV) database, which provides a holistic representation of the hospital treatment process of patients undergoing different treatments [4]. Nevertheless, the developed methods can be applied to different domains, such as manufacturing, where event attributes are available.

ICPM 2022 Doctoral Consortium and Tool Demonstration Track
jonas.cremerius@hpi.de (J. Cremerius)

2. Related Work

The analysis of event attributes has been approached from different perspectives in the literature.

A prominent application is the identification of decision rules, such as in data-aware heuristic mining [2]. Regarding the exploration of event attributes, the multi-perspective process explorer allows investigating the distribution of each event attribute at each activity [5]. In [6], the access to event attributes is described and annotated to the process model, describing the data object lifecycle of each event attribute. The field of predictive process analytics makes use of event attributes as well, by incorporating them in different machine learning models, such as process outcome prediction [3].

Event attributes have also been classified in several ways for different analysis purposes. For instance, [7] classifies event attributes into controllable and non-controllable, where the first one can be altered during a process execution and the other cannot. Another contribution classifies event attributes according to a medical standard (openEHR) to find dependencies within and between these categories [8].

Even though it has been identified, that event attributes can be altered during the process, its analysis is limited to read/write access or distribution visualization at single activities. Therefore, this PhD project aims to provide different methods to understand the value behaviour of event attributes throughout the process.

3. Research Roadmap

The doctoral thesis will provide a framework for the value behaviour analysis of event attributes following a design science approach. In this section, we present the current state of research and plans for enhancing the methodology.

3.1. Data-Enhanced Process Models

The beginning of the value behaviour analysis of event attributes started with the question, if it is possible to increase the visibility of event attributes in process models. To achieve that, we designed Data-Enhanced Process Models, which allow calculating and displaying event attribute aggregations for each activity in the process model. For example, if a patient is registered at the beginning of treatment, an activity could include the patient's age as an event attribute. One can now calculate the mean age as an event attribute aggregation and represent it directly in the process model.

As there might be lots of event attributes available in an event log, we further developed three mechanisms to support event attribute selection in Data-Enhanced Process Models. These are based on the data type, process characteristic, and degree of variability of event attributes [9].

3.2. Change Detection in Dynamic Event Attributes

With Data-Enhanced Process Models, we allow exploring the value behaviour of event attributes over the process. However, it is not possible to make concrete statements regarding their value behaviour. In the previous work, we defined the process characteristic *dynamic* for event attributes [9]. *Dynamic* event attributes change throughout the process and are associated to multiple activities in an event log. These are especially interesting, as their changing behaviour can be analysed. For example, when patients undergo a certain treatment, such as dialysis, certain laboratory values are expected to change from beginning until end of treatment, such as the creatinine value [10].

Therefore, we developed a method based on statistical significance tests, allowing to identify changes of *dynamic* event attributes throughout the process [11]. It allows identifying changes in three dimensions: the event attribute, directly and eventually follows relations, and trace variants. Thus, we are able to provide detailed information about the location of changes in a process-oriented way, so we can say, that an event attribute is changing between two activities in certain trace variants.

3.3. Patient Discharge Classification based on the Hospital Treatment Process

The analysis methods provided so far allow retrieving novel insights about the value behaviour of event attributes. In addition to that, we wanted to demonstrate how the process behaviour of the values can be utilized in the field of predictive process analytics in healthcare.

Predicting the discharge location of patients after hospital treatment is a popular research topic. So far, the status of the patient at the end of treatment is used as input for most models. However, the treatment process has not been considered yet. Our approach proposes to incorporate the conducted treatment activities and development of the patient's state in form of event attributes to make the decision of the most suitable discharge location, such as home or skilled nursing facility [12]. We could demonstrate, that the development of the patient's state during the process and the respective activities in the hospital have a considerable impact on the discharge location prediction.

3.4. Validation Method

The approaches presented are validated on MIMIC-IV, providing real-world healthcare process data. Thus, we generated reproducible event logs out of MIMIC-IV and applied the proposed methods on them. In addition to that, we consulted medical experts to validate the correctness of the findings. It is planned to validate the results on a second dataset with event attributes available, such as the Sepsis event log. Furthermore, we want to demonstrate, that the approach is able to solve concrete problems by conducting case studies.

3.5. Next Steps

Our plan is to extend the methods to analyse the value behaviour of event attributes. One plan is to identify correlating event attribute changes, so we can say, that two or more event attributes are changing together at the same location. It would also be interesting to cluster cases by their changing behaviour, such that it is possible to characterize changing and non-changing cases. Furthermore, we think that there is potential in the field of process variant analysis, where the value behaviour can be compared between multiple process variants. As this thesis focusses on healthcare processes, we also would like to cope with the challenges presented in [13], where we, for example, focus on dealing with reality by looking at real-world data and aiming to provide white-box approaches which are understandable for medical experts.

4. Conclusion

In this doctoral thesis, we want to enable process analysts and domain experts to understand the value behaviour of event attributes within a process. To achieve that, we propose a framework with different methods to look at the value behaviour from different perspectives. This can help to confirm expected behaviour, but also allows deriving novel insights about the data associated to process activities. The results are validated based on real-world data and domain expertise.

References

[1] W. van der Aalst, Process Mining, Springer Berlin Heidelberg, 2016.

[2] M. de Leoni, W. van der Aalst, Data-aware process mining: Discovering decisions in processes using alignments, in: Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 1454–1461.

[3] A. Leontjeva, I. Kuzovkin, Combining static and dynamic features for multivariate sequence classification, 2016, pp. 21–30.

[4] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L. A. Celi, R. Mark, Mimic-iv, 2020.

[5] F. Mannhardt, M. de Leoni, H. Reijers, The multi-perspective process explorer, in: CEUR Workshop Proceedings, volume 1418, 2015.

[6] D. Bano, F. Zerbato, B. Weber, M. Weske, Enhancing discovered process models with data object lifecycles, in: 2021 IEEE 25th International Enterprise Distributed Object Computing Conference (EDOC), 2021, pp. 124–133.

[7] Z. Bozorgi, I. Teinemaa, M. Dumas, M. La Rosa, A. Polyvyanyy, Process mining meets causal machine learning: Discovering causal rules from event logs, 2020, pp. 129–136.

[8] H. Xu, J. Pang, X. Yang, J. Yu, X. Li, D. Zhao, Modeling clinical activities based on multi-perspective declarative process mining with openEHR's characteristic, BMC Medical Informatics and Decision Making 20 (2020).

[9] J. Cremerius, M. Weske, Supporting domain data selection in data-enhanced process models, in: Wirtschaftsinformatik 2022 Proceedings 3, 2022.

[10] N. Amin, R. T. Mahmood, M. Asad, M. Noorulamin, Evaluating urea and creatinine levels in chronic renal failure pre and post dialysis: A prospective study 2 (2014) 2330–4596.

[11] J. Cremerius, M. Weske, Change detection in dynamic event attributes, in: C. Di Ciccio, R. Dijkman, A. del Río Ortega, S. Rinderle-Ma (Eds.), Business Process Management Forum, Springer International Publishing, Cham, 2022, pp. 157–172.

[12] J. Cremerius, M. König, C. Warmuth, M. Weske, Patient discharge classification based on the hospital treatment process, in: J. Munoz-Gama, X. Lu (Eds.), Process Mining Workshops, Springer International Publishing, Cham, 2022, pp. 314–326.

[13] J. Munoz-Gama, et al., Process mining for healthcare: Characteristics and challenges, Journal of Biomedical Informatics 127 (2022) 103994.