

Towards Semantic-driven, Declarative and Interactive Process Mining

Christian Dormagen¹

¹Otto-Friedrich Universität Bamberg, Kapuzinerstraße 16 96047 Bamberg, Germany

Abstract

Semantic knowledge about organizational processes, available either explicitly in information systems, extracted as event log traces, or implicitly as expert knowledge, is underutilized in process mining. This PhD proposal aims to advance process mining by integrating implicit and explicit business process knowledge into a semantically enriched event log that links data from different sources. The challenges of noisy event logs and data integration require an optimal level of abstraction. We aim to extend the scope of the semantic event log by incorporating different levels of abstraction, hierarchical structures and process perspectives such as macroscopic system-level behavior. Central to our strategy is the development of interactive declarative process discovery methods using the semantic event log, combining inductive learning, logical reasoning as well as explainable and interactive machine learning. The goal is to enable incremental correction and refinement of process models by domain experts, with feedback formalized in the semantic event log as background knowledge. This research aims to bridge theoretical advances with practical applications, evaluated in collaboration with industry partners, to promote more effective process discovery, exploration and understanding.

Keywords

Semantic Process Mining, Explanatory Interactive Machine Learning, Event Log Abstraction

1. Introduction

Classical process mining methods typically treat information within event logs as abstract tokens, lacking a deeper exploitation of the underlying semantics governing process behavior. This leads to discovered process models potentially being semantically unsound, for example allowing a *Send Order* before a *Create Order* event in a *purchase-to-pay (P2P)* process due to artifacts in the data. Explicit incorporation of semantic knowledge about cause-effect relations between events can identify and mitigate such errors. Another challenge arises when unifying possibly heterogeneous data from various sources. Different sites of one company may use varied namespaces or record data at different granularity levels. Integrating this information requires a semantic understanding of the underlying entities and their hierarchical relationships.

These problems exist despite the fact that semantic information is usually available in some form, either encoded in the ERP system supporting process execution, made explicit in technical documents [1], or implicit as expert knowledge of the human actors executing the process. This highlights the need to move beyond classical process mining and towards semantic process

ICPM Doctoral Consortium and Demo Track 2023

✉ christian.dormagen@uni-bamberg.de (C. Dormagen)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

mining, which has recently been recognized by the process mining community [2, 1, 3, 4]. These considerations motivate our research questions:

- RQ1: Can we extract and encode existing semantic knowledge in a semantic event log?
- RQ2: Can we further enrich the semantic event log with additional information to provide different perspectives on the process behavior beyond typical event sequences?
- RQ3: Can we use the information encoded in the semantic event log to enable new types of process analysis and exploration beyond the traditional workflow perspective?
- RQ4: Can we develop declarative process discovery methods on the foundation of the semantic event log to discover more accurate process models?

We outline our approach to extend current event log abstraction methods to explicitly capture system behavior and enrich the event log with contextual background knowledge, transforming it into a semantic event log. We then discuss the semantic foundations for formalizing and encoding this knowledge. Finally, we discuss process discovery methods based on the semantic event log and propose a practical evaluation concept for our approach.

2. Methodology

2.1. Event log abstraction

Traditional event log abstraction methods transform event logs from a lower to a higher level of granularity [5]. This is necessary because the granularity of raw data from ERP systems often exceeds the level of detail required for analysis and contains artifacts from human error or interactions between different information systems. These methods typically aim to cluster multiple events representing an activity or sub-process within the target process. Note that this inherently assigns semantics to the data through a hierarchy of events. If we cannot infer this information from available context or expert knowledge, existing discovery methods can still uncover it.

We are particularly interested in event log abstraction methods that go beyond adjusting log granularity and instead transform the information to capture a novel view of process behavior. An example of this is the system-level event discovery method proposed by Bakullari et al. [6]. This method abstracts a classical event log into a log of system-level events that highlight patterns of interest across multiple cases, such as high resource workload or atypical delays in specific steps. We want to explicitly incorporate this system-level behavior into our semantic event log. As a result, we plan to extend the method by introducing new types of system-level patterns for discovery, such as batching, the simultaneous execution of an event type across multiple cases. Since batching behavior can lead to delays or spikes in workload, it is of interest to explain delays in a process [7, 8]. We expect that explicitly discovering such patterns and enriching the semantic log with them will enable us to reason more easily about the dependencies between individual cases and the overall systemic behavior within a process.

2.2. Semantic grounding

To capture the semantics of an event log and of contextual background knowledge, we aim to formally define meaningful representations inspired by [3, 9, 1, 10, 2, 11] that encode causal,

hierarchical and temporal relationships between events and attribute types typical of event logs. We are interested in both structural knowledge about the domain of the process and knowledge of control-flow dependencies. Structural knowledge refers to the types of entities and relations which the process manipulates. It includes for example hierarchical sub/superclass relationships among events or resources. This knowledge facilitates seamless integration of diverse data sources and grants flexibility in adjusting the level of event or resource granularity during analysis. Control-flow dependencies are for example self-imposed via business rules or exist as *laws of nature* [12]. These relations enforce a strict order of event execution and any trace violating the imposed event order is inherently erroneous. Our earlier example of *Send Order* event not being possible before a corresponding *Create Order* event is such a *law of nature* relation, while a business rule might dictate that 'Any bank transfer above a certain value must first be checked by a resource with a sufficiently high level of authorization.' We can derive such knowledge from specific details about the organization in which the process is executed. In addition, we can refer to generic knowledge available about the type of process. For example, process types such as *P2P* and *Order-to-Cash (O2C)* are ubiquitous and have a wealth of typical or expected process behavior to reference.

Various logical formalisms are commonly used to specify knowledge in this context. For expressing time-variant properties, LTLf is a frequently used formalism and is widely used in declarative process mining as the underlying logic of DECLARE [13]. Relational knowledge, on the other hand, is often specified using Description Logics (DL) [14]. DLs are a family of knowledge representation languages widely used in the Semantic Web to formalize and specify ontologies, which we plan to use in order to encode our extracted knowledge. Ontologies are closely related to knowledge graphs through their role in providing a logical framework for representing structured knowledge in such graphs. We plan to realize our semantic event log in the form of a knowledge graph encoding, integrating contextual background knowledge from linked data sources such as formal ontologies from the semantic web. Furthermore, a knowledge graph representation may offer solutions to recognized issues with classical event logs, including con/divergence and the limitation in capturing multiple case concepts or process views within a single data structure [15].

Of particular interest to us is the causal and temporal specification of system-level event patterns. Our focus will be on investigating extended semantics of window operators. Bakullari et al.[6] utilize a simple window function for pattern discovery that divides a log into fixed-size windows. Our plan is to develop generalized window semantics to enable the representation and discovery of more complex behavior patterns. Consequently, we are particularly interested in languages that support expressive temporal window operators, such as LARS[16], which supports time-based, tuple-based, partition-based, and filter-based window operators.

2.3. Declarative interactive process mining

Based on the enriched semantic log, we plan to develop declarative process discovery and analysis methods, combining inductive learning (e.g. Inductive Logic Programming (ILP) [17]), logical inference methods and explanatory interactive machine learning [18].

We plan to use ILP for discriminative mining in the analysis of system-level behavior, made explicit as system-level events by splitting the log into traces affected and unaffected by a system-

level event of interest. This can allow us to learn a discriminative process model and gain further insight into the factors that lead to a particular system-level event and its consequences. For example, we can split the log into traces that are affected by a high workload of a particular resource and use the learned discriminative model to guide the experts using the system in restructuring the process to prevent such events from occurring.

Our declarative process discovery algorithm will leverage the enhanced semantic representation and contextual background knowledge. This knowledge constrains the set of discoverable process models, aiming to improve process discovery. Furthermore, it facilitates the identification of new and implicitly hidden relationships shared between event patterns or process instances. Attempts at incorporating richer constraints over data, classes or relations of events into declarative process mining have already been made, most notably in the context of *object-centric behavioral constraints* [19, 20] for the DECLARE [13] language. The ideas presented there can serve as a starting point and to inform our own declarative miner. Furthermore, the semantic information is crucial as a foundation for the realization of Explainable and Interactive Machine Learning (XIML) within process mining. This enables a human-in-the-loop approach that incorporates explanations and considers the collaborative engagement of domain experts. While initial approaches to implementing XIML methods in process mining exist, they either lack the ability to make additional knowledge explicit for later process discovery [21] or are post-hoc methods that modify an existing process model without integration into the discovery process itself [22].

Our plan is to implement an XIML method inspired by the CAIPI algorithm [18, 23], enabling expert users to provide feedback that is then fed back into the semantic log. This approach facilitates incremental correction and refinement of process models during process discovery. Consequently, the contextual background is progressively extended by domain experts, enhancing the discovery, analysis, and exploration of process models.

2.4. Evaluation

We will evaluate our method on synthetic logs for correctness of implementation. For comparison with established methods we will also apply it to existing event logs, including the BPIC19 log [24], as it is a *P2P* log with substantial contextual knowledge. Working with an industry partner, we have access to real-world *P2P* and *O2C* event logs from production ERPs. In addition, a case study involving industry experts will provide feedback on our XIML based declarative process discovery method, validating the correctness and practicality of our semantically enriched event log and discovery approach.

Acknowledgments

This project is funded by the Bavarian research program (BayVFP) under KIGA (DIK0313).

References

- [1] O. Nykänen, A. Rivero-Rodriguez, P. Pileggi, P. A. Ranta, M. Kailanto, J. Koro, Associating event logs with ontologies for semantic process mining and analysis, in: Proceedings of the 19th International Academic

- Mindtrek Conference, ACM, 2015. doi:10.1145/2818187.2818273.
- [2] A. K. A. de Medeiros, W. van der Aalst, C. Pedrinaci, Semantic process mining tools: core building blocks, in: 16th European Conference on Information Systems, 2008. URL: <https://oro.open.ac.uk/23397/>.
 - [3] S. Esser, D. Fahland, Multi-dimensional event data in graph databases, *Journal on Data Semantics* 10 (2021).
 - [4] A. K. A. de Medeiros, C. Pedrinaci, W. M. P. van der Aalst, J. Domingue, M. Song, A. Rozinat, B. Norton, L. Cabral, An outlook on semantic business process mining and monitoring, in: *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*, Springer Berlin Heidelberg, 2007, pp. 1244–1255.
 - [5] S. J. van Zelst, F. Mannhardt, M. de Leoni, A. Koschmider, Event abstraction in process mining: literature review and taxonomy, *Granular Computing* 6 (2020) 719–736. doi:10.1007/s41066-020-00226-2.
 - [6] B. Bakullari, W. M. P. van der Aalst, High-level event mining: A framework, 2022.
 - [7] P. Waibel, C. Novak, S. Bala, K. Revoredo, J. Mendling, Analysis of business process batching using causal event models, in: *Lecture Notes in Business Information Processing*, Springer International Publishing, 2021.
 - [8] E. L. Klijn, D. Fahland, Performance mining for batch processing using the performance spectrum, in: *Business Process Management Workshops*, Springer International Publishing, 2019.
 - [9] K. Okoye, A.-R. Tawil, U. Naeem, E. Lamine, Semantic Process Mining Towards the Discovery And Enhancement Of Learning Models Analysis, 2016, pp. 121–164. doi:10.13140/RG.2.1.4793.4325.
 - [10] A. Rebmann, H. van der Aa, Extracting semantic process information from the natural language in event logs, in: *Advanced Information Systems Engineering*, Springer International Publishing, 2021.
 - [11] O. Kingsley, A.-R. H. Tawil, U. Naeem, S. Islam, E. Lamine, Using semantic-based approach to manage perspectives of process mining: Application on improving learning process domain data, in: *2016 IEEE International Conference on Big Data (Big Data)*, 2016, pp. 3529–3538. doi:10.1109/BigData.2016.7841016.
 - [12] G. Adamo, C. D. Francescomarino, C. Ghidini, F. M. Maggi, Beyond arrows in process models: A user study on activity dependences and their rationales, *Information Systems* 100 (2021).
 - [13] M. Pesic, H. Schonenberg, W. M. van der Aalst, DECLARE: Full support for loosely-structured processes, in: *11th IEEE International Enterprise Distributed Object Computing Conference (EDOC 2007)*, IEEE, 2007.
 - [14] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, P. F. Patel-Schneider (Eds.), *The Description Logic Handbook*, 2 ed., Cambridge University Press, Cambridge, UK, 2007.
 - [15] W. M. P. van der Aalst, Object-centric process mining: Dealing with divergence and convergence in event data, in: *Software Engineering and Formal Methods*, Springer International Publishing, 2019.
 - [16] H. Beck, M. Dao-Tran, T. Eiter, LARS: A logic-based framework for analytic reasoning over streams, *Artificial Intelligence* 261 (2018) 16–70. doi:10.1016/j.artint.2018.04.003.
 - [17] S. Muggleton, Inductive logic programming, *New Generation Computing* 8 (1991).
 - [18] S. Teso, K. Kersting, Explanatory Interactive Machine Learning, in: V. Conitzer, G. K. Hadfield, S. Vallor (Eds.), *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019, Honolulu, HI, USA, January 27-28, 2019*, ACM, 2019, pp. 239–245. doi:10.1145/3306618.3314293.
 - [19] G. Li, R. M. de Carvalho, W. M. P. van der Aalst, Automatic discovery of object-centric behavioral constraint models, in: *Business Information Systems*, Springer International Publishing, 2017.
 - [20] A. Artale, A. Kovtunova, M. Montali, W. M. P. van der Aalst, Modeling and reasoning over declarative data-aware processes with object-centric behavioral constraints, in: *Lecture Notes in Computer Science*, Springer International Publishing, 2019, pp. 139–156. doi:10.1007/978-3-030-26619-6_11.
 - [21] P. M. Dixit, H. M. W. Verbeek, J. C. A. M. Buijs, W. M. P. van der Aalst, Interactive data-driven process model construction, in: *Conceptual Modeling*, Springer International Publishing, 2018.
 - [22] P. M. Dixit, J. C. A. M. Buijs, W. M. P. van der Aalst, B. F. A. Hompes, J. Burman, Using domain knowledge to enhance process mining results, in: *Lecture Notes in Business Information Processing*, Springer International Publishing, 2017, pp. 76–104. doi:10.1007/978-3-319-53435-0_4.
 - [23] E. Slany, Y. Ott, S. Scheele, J. Paulus, U. Schmid, Caipi in practice: Towards explainable interactive medical image classification, 2022. doi:10.48550/ARXIV.2204.02661.
 - [24] B. van Dongen, Bpi challenge 2019, 2019. doi:10.4121/UUID:D06AFF4B-79F0-45E6-8EC8-E19730C248F1.