

Leveraging Desirable and Undesirable Event Logs in Process Mining Tasks

Ali Norouzfifar
RWTH Aachen University, Germany

ALI.NOROUZIFAR@PADS.RWTH-AACHEN.DE

Wil van der Aalst
RWTH Aachen University, Germany

WVDAALST@PADS.RWTH-AACHEN.DE

Abstract

Traditional process mining techniques utilize one event log as input to offer organizational insights. In many applications, information regarding undesirable process aspects may exist. However, the literature lacks a comprehensive overview of their integration into process mining tasks. In our paper, we explore leveraging data from both desirable and undesirable event logs to augment existing process mining tasks and develop innovative applications. Our aim is to systematically outline the potential for enhancements in this realm.

Keywords: Process Mining, Desirable and Undesirable Behavior, Process Discovery

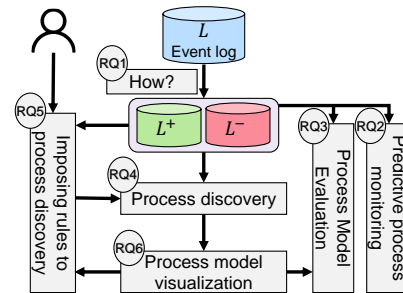


Figure 1: Overview of the research questions leveraging the desirable and undesirable event logs.

1. Introduction

Data extracted from information systems constitutes a rich and valuable resource, providing possibilities for diverse analyses. Process mining, as a broad discipline, encompasses a variety of applications applied to event data to extract meaningful insights [1]. In our research, we delved into leveraging information related to desirable and undesirable event logs to extract valuable insights and assist organizations in improving their processes by addressing performance and compliance problems. In Fig. 1, an overview of the research questions targeting desirable and undesirable cases is illustrated.

2. Potential Research Questions

Identification of Desirable and Undesirable Event Logs (RQ1) Some potential approaches to derive the desirability of the cases include domain-specific labeling [2], assessing the adherence to the normative behavior [3], rule-checking techniques [3], identification of outliers or strange cases [4], and automated detection of control flow variability [5]. Given that automatic labeling of cases as *desirable*

or *undesirable* relies on interpretations and specific scenarios, user input may be necessary to choose an appropriate method. In [5], a framework for the identification of control flow variations across continuous dimensions like duration of cases is proposed. This framework takes an event log and a continuous dimension. Considering the cases are sorted based on their assigned value, a sliding-window-based algorithm utilizing the earth mover’s distance is employed to find the change points in the control flow. Further analysis of the identified segments helps to categorize the cases into desirable or undesirable.

Predictive Process Monitoring (RQ2) Assuming that the desirability of the cases is known, the event logs can be encoded as suitable features for the machine learning techniques and the state-of-the-art architectures can be used to obtain predictive models or recommendations [6]. The process mining field can contribute to the improvements by providing meaningful features [7]. Extracting explainability from such predictive models helps to establish trustworthy predictions [8]. In addition to predictive models, many process variant analysis techniques from

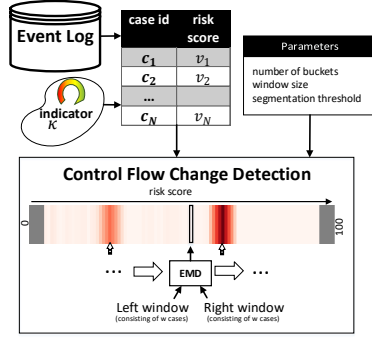


Figure 2: Process variant identification framework across continuous dimensions [5]

the literature could be used to effectively compare the event logs [9].

Process Model Evaluation (RQ3) Assessing a process model’s capability to differentiate between desirable and undesirable behavior is challenging. Fitness is an evaluation metric used to assess the re-playability of the observed behavior on the discovered model [3]. In [10], evaluation metrics are suggested, utilizing fitness criteria to ascertain model fit with the desirable event log while not aligning with the undesirable event log. Another approach proposed in [11] checks if the generalization of behavior allowed by the model conflicts with the undesirable behavior generated artificially from the desirable event log. Adaptations are required to make them applicable in scenarios with desirable and undesirable event logs.

Process Discovery (RQ4) The goal is to discover process models that support a desirable event log while avoiding an undesirable event log. Limited research has been conducted to involve desirable and undesirable event logs in process discovery. Declarative process discovery approaches like [2] and [12] discover constraint-based models from desirable and undesirable event logs. In [13] and [14], the discovery of procedural process models using these event logs is investigated. The IMbi algorithm is another relevant discovery technique introduced in [10]. In each recursion, the algorithm finds a process structure that has a low cost based on the desirable event log and a high cost based on the undesirable event log. In Fig. 3, one recursion of the proposed algorithm is illustrated [10]. The *ratio* parameter controls the involvement of the undesirable event log in the process discovery.

Imposing Rules to Process Discovery (RQ5) User-defined rules or discovered rules from event logs can help to enhance the quality of the discovered pro-

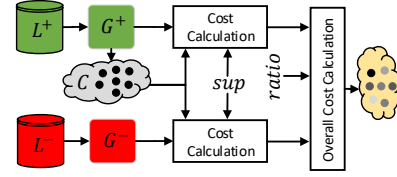


Figure 3: One recursion of IMbi framework [10], discovering a process model to support the desirable event log while avoiding the undesirable event log.

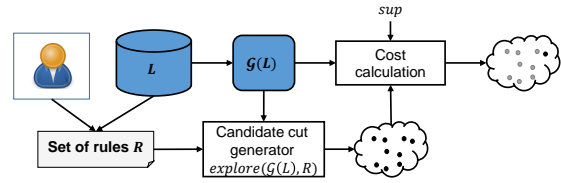


Figure 4: One recursion of IMr framework [15], allowing for a set of rules R as input.

cess models. Existing process discovery frameworks often disregard other valuable sources such as domain knowledge and documentation within business processes that provide insights into how the process functions. In [15], the IMr algorithm is proposed as a generalization of the IMbi framework that is capable of considering some declarative rules as input. One recursion of the IMr framework is illustrated in Fig. 4. This approach can be extended to scenarios with two event logs as input. In addition, a set of rules that can discriminate between the desirable and undesirable event logs can help to discover better process models.

Process Model Visualization (RQ6) The obtained insights from the previous research questions could be accompanied by visualization techniques. Conformance checking techniques from the literature or user feedback can be used to further improve the process models and align them better with reality.

3. Conclusion

Six potential research questions regarding the utilization of desirable and undesirable event logs in the process mining field are introduced in this extended abstract. Extending existing process mining frameworks with the integration of this information could yield enhanced insights.

Acknowledgments

This research was supported by the research training group “Dataninja” (Trustworthy AI for Seamless Problem Solving: Next Generation Intelligence Joins Robust Data Analysis) funded by the German federal state of North Rhine-Westphalia.

References

- [1] Wil M. P. van der Aalst. *Process Mining - Data Science in Action, Second Edition*. Springer, 2016.
- [2] Tijds Slaats, Søren Debois, and Christoffer Olling Back. Weighing the pros and cons: Process discovery with negative examples. In *International Conference on Business Process Management*, pages 47–64. Springer, 2021.
- [3] Josep Carmona, Boudewijn F. van Dongen, Andreas Solti, and Matthias Weidlich. *Conformance Checking: Relating Processes and Models*. Springer, Cham, 1st edition, 2018. ISBN 3319994131.
- [4] Jochen De Weerd, Seppe K. L. M. vanden Broucke, Jan Vanthienen, and Bart Baesens. Active trace clustering for improved process discovery. *IEEE Trans. Knowl. Data Eng.*, 25(12): 2708–2720, 2013.
- [5] Ali Norouzifar, Majid Rafiei, Marcus Dees, and Wil M. P. van der Aalst. Process variant analysis across continuous features: A novel framework. In *Enterprise, Business-Process and Information Systems Modeling - 25th International Conference, BPMDS 2024, and 29th International Conference, EMMSAD 2024, Proceedings*, volume 511 of *Lecture Notes in Business Information Processing*, pages 129–142. Springer, 2024.
- [6] Irene Teinemaa, Marlon Dumas, Marcello La Rosa, and Fabrizio Maria Maggi. Outcome-oriented predictive process monitoring: review and benchmark. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 13(2):1–57, 2019.
- [7] Mozghan Vazifehdoostirani, Laura Genga, and Remco M. Dijkman. Encoding high-level control-flow construct information for process outcome prediction. In *4th International Conference on Process Mining, ICPM 2022*, pages 48–55. IEEE, 2022.
- [8] Riza Velioglu, Jan Philip Göpfert, André Artelt, and Barbara Hammer. Explainable artificial intelligence for improved modeling of processes. In *Intelligent Data Engineering and Automated Learning - IDEAL 2022, Manchester, UK, Proceedings*, volume 13756 of *Lecture Notes in Computer Science*, pages 313–325. Springer, 2022.
- [9] Farbod Taymouri, Marcello La Rosa, Marlon Dumas, and Fabrizio Maria Maggi. Business process variant analysis: Survey and classification. *Knowledge-Based Systems*, 211:106557, 2021.
- [10] Ali Norouzifar and Wil M. P. van der Aalst. Discovering process models that support desired behavior and avoid undesired behavior. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, SAC 2023*, pages 365–368. ACM, 2023.
- [11] Seppe KLM vanden Broucke, Jochen De Weerd, Jan Vanthienen, and Bart Baesens. Determining process model precision and generalization with weighted artificial negative events. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1877–1889, 2013.
- [12] Federico Chesani, Chiara Di Francescomarino, Chiara Ghidini, Giulia Grundler, Daniela Loreti, Fabrizio Maria Maggi, Paola Mello, Marco Montali, and Sergio Tessaris. Shape your process: Discovering declarative business processes from positive and negative traces taking into account user preferences. In *International Conference on Enterprise Design, Operations, and Computing*, pages 217–234. Springer, 2022.
- [13] Hernán Ponce De León, Lucio Nardelli, Josep Carmona, and Seppe KLM vanden Broucke. Incorporating negative information to process discovery of complex systems. *Information Sciences*, 422:480–496, 2018.
- [14] Evelina Lamma, Paola Mello, Marco Montali, Fabrizio Riguzzi, and Sergio Storari. Inducing declarative logic-based models from labeled traces. In *Business Process Management: 5th International Conference, BPM 2007. Proceedings*, pages 344–359. Springer, 2007.

- [15] Ali Norouzifar, Marcus Dees, and Wil M. P. van der Aalst. Imposing rules in process discovery: An inductive mining approach. In *Research Challenges in Information Science - 18th International Conference, RCIS 2024, Proceedings, Part I*, volume 513 of *Lecture Notes in Business Information Processing*, pages 220–236. Springer, 2024.