Enhancing Support for Individual Analysts in Process Mining

Lisa Zimmermann[1]

[1]University of St. Gallen, Rosenbergstr. 30, 9000 St. Gallen, CH

Abstract

The adoption of process mining technology is steadily increasing in practice. However, individual analysts confronted with questions about a process execution lack adequate support for various tasks. Therefore, this doctoral project aims to explore the specific support needs of individuals and offer methods and operational assistance for the analysis phase in process mining projects. Employing an empirical approach and analyzing behavior observations, interviews, and questionnaire results, we seek to identify the challenges of analysts and develop practice-oriented guidance. The envisioned methods, guidelines, and operational support tools will aid analysts throughout the entire analysis process, from formulating questions to evaluating their answers.

Keywords

Process Mining, Individual Perspective, End-User Support, Challenges

1. Introduction

Research in the field of process mining (PM) has so far mainly focused on technical aspects [1], while less attention is given to the perspective of analysts who apply PM algorithms, asses discovered process models, formulate and test hypotheses, and define and evaluate statistics to answer a PM question. Indeed, prior work pointed out that the analysis process oftentimes remains emergent, ad-hoc, and manual in practice and highlighted the need for more advanced support for analysts [2, 3, 4]. Initial efforts exist that started addressing this gap by, for example, developing (semi-)automated analysis mechanisms [5, 6], by suggesting artifacts (*dashboards*) to communicate improvement opportunities [7], or by capturing the analysis process allowing analysts to follow and replay executed operations [4]. Besides these first efforts, there is still relevant research potential for better understanding the work practices of PM analysts and for suggesting support systems. In particular, when contemplating a PM analysis in which analysts employ different PM techniques, assess results, apply filters, re-apply PM techniques, etc., to answer a question, we currently lack a comprehensive understanding of how this analysis process unfolds in practice, where the challenges lie, and how it can be effectively supported.

Within this doctoral project, we aim to close this gap by (1) contributing to a better understanding of the analysis process and (2) developing methodological guidance and software-based support for novice analysts.

ICPM Doctoral Consortium and Demo Track 2023
lisa.zimmermann@unisg.ch (L. Zimmermann)
0000-0002-6149-7060 (L. Zimmermann)

2. Motivation and Research Questions

The motivation behind the aforementioned goals stems from the recognition that performing a PM analysis poses challenges for analysts, especially for those who are less experienced (novices). Existing research (e.g., [2, 3, 8, 9]) has already highlighted the necessity for enhanced guidance across various PM tasks. However, to confirm the relevancy, we conducted a preliminary validation by answering the following initial research question (RQ).

RQ0: **What are the challenges perceived by individual process analysts during a PM project?**

To answer this question, we followed grounded theory, a qualitative research approach that suggests how to retrieve so far unknown concepts and theories from empirical data [10]. We conducted semi-structured interviews and identified 23 challenges that are faced during PM projects [11]. The existence of challenges, such as formulating PM questions, navigating an analysis with limited domain knowledge and prior experience, or concluding an analysis and finding an answer [11], confirms the need for a better understanding of the individual level of PM. This understanding is required to identify the causes of challenges and develop tailored support for analysts.

In this project, we focus on the analysis phase, including its influencing factors and its output, but excluding PM project tasks detached from the analysis phase, such as data preparation. Guided by the outcomes of the interview study, combined with findings from existing literature (esp. [2, 12, 13]), we structure the outline of the further research plan based on the following considerations.

First, we could infer that the available tools and PM techniques, the properties of the event log to be analyzed, and the PM question influence the analysis phase and are often given from the start and adopted over several iterations. While the quality of event logs already received recognizable attention in the research community [14] and tools are constantly enhanced [1], the PM question and support for analysts in deriving such a question, have received less attention.

Second, we learned that the *analysis behavior* is highly analyst-specific, emergent, and manual, as analysts apply sequences of operations, i.e., data manipulations, creation and interpretation of representations, formulation and testing of hypotheses, etc. [15, 16], to answer (business) questions. Additionally, many analysis steps depend on prior experience and domain knowledge [12]. While ideas have been proposed to automate selected analysis tasks (e.g., [5, 6]) general support in guiding analysts on how to answer PM questions is missing.

Third, analysts need to translate their observations into insights and ultimately answer the questions of the stakeholders. In practice, these insights often trigger process transformations that increase the efficiency or quality of the process. Based on our knowledge, attempts to streamline the qualitative evaluation of PM results exist [17], but support in assessing the quality of PM insights objectively beyond the model quality is currently lacking.

In order to suggest support for analysts comprehensively, this project will consider all three areas with the following RQs.

RQ1: **How can analysts be supported in deriving PM questions for their analysis?**

RQ2: **How can analysts be supported in adopting effective analysis behavior during the mining and analysis phase to answer PM questions?**

RQ3: **How can analysts be supported in evaluating the answers derived during the analysis?**

3. Research Outline and Proposed Solution

By answering the aforementioned RQs, our goal is to develop a comprehensive framework to support PM analysts. Based on our current research status, we envision the framework to consist of, at least, the following methods and tools.

3.1. RQ1 - PM Research Question

To address RQ1, we will develop a PM question bank containing questions retrieved from literature and practice, enriched by a taxonomy providing a question classification system. With the combination of these two components, we aim to support analysts in the identification of questions based on their analysis goals. To allow access to the taxonomy and the question bank, we plan to implement an interface enabling analysts to retrieve questions by selecting categories they are interested in. They can adopt questions from the question bank either directly, or use them as inspiration to formulate their own questions. Additionally, analysts will be enabled to categorize questions and ensure a common understanding of the question and the underlying PM concepts (e.g., bottlenecks, rework) based on the definitions of the categories of the taxonomy.

Method To collect a representative set of questions, we already performed a systematic literature review of case studies and *Business Process Intelligence Challenge*[1] reports. The results of an online data collection complement the set. To develop the taxonomy, we will thoroughly follow the guidelines proposed by [18]. We will evaluate our taxonomy including its interface with experts and in a case study together with an industry partner to demonstrate its usefulness for the support of analysts.

3.2. RQ2 - Effective Analysis Behavior

To support analysts during the analysis (cf. RQ2), we first need to enhance the existing understanding of what analysts do, how they do it, and also why they do it. Based on such an enhanced understanding of the analysis process, we plan to propose a collection of *effective analysis patterns*. In a first version, we will integrate these patterns into the taxonomy interface (cf. RQ1), linked to questions they can be applied for. For a second version, proposing these patterns in the form of traceable and re-playable analysis building blocks (similar to the idea introduced in [4] for capturing analysis operations retrospectively – but for our case applied proactively) seems a promising direction.

[1] https://www.tf-pm.org/competitions-awards/bpi-challenge

Method To first enhance the existing understanding of the analysis phase and identify low-level analysis tasks that could compose the analysis patterns, we will rely on behavioral data gathered from analysts while performing an analysis. This data contains screen recordings, enriched by concurrent think-aloud data. Following the design science research principles, we will then propose, and subsequently evaluate, a mapping of analysis tasks to questions of the question bank (cf. RQ1). Whether the mapping is useful and supports analysts during the question answering will be evaluated with the help of controlled experiments.

3.3. RQ3 - Evaluation of PM Answers

Our plan to address RQ3 is preliminary. We envision developing evaluation guidelines that help analysts objectively assess their PM answers following the design science research methodology. We will start by defining the problem space. To this end, a case study can help to identify the context and concrete gaps in the evaluation of different PM results. Afterward, we can propose adequate support measures, such as a method or rules that should be followed to ensure high-quality PM results that help process stakeholders achieve their goals. Compared to the authors in [17] who suggest guidelines for the qualitative evaluation of PM results together with a stakeholder group, we will focus on a quantitative assessment applicable to an individual analyst. Based on our knowledge, this will be one of the first works explicitly linking PM questions and answers and therefore providing analysts with guidance regarding objective assessment measures. Similar to the results of RQ1, we plan to evaluate the proposed guidelines both with experts as well as in a real-world setting (e.g., a case study with an industry partner).

Acknowledgments

I thank my PhD supervisor Barbara Weber, for her valuable guidance, and Francesca Zerbato, for her consistent support throughout this research.

This work is supported by the ProMiSE project funded by the SNSF under Grant No.: 200021_197032.

References

[1] W. M. van der Aalst, J. Carmona, Process mining handbook, Springer Nature, 2022.
[2] M. L. v. Eck, X. Lu, S. J. Leemans, W. M. Van Der Aalst, PM^2: a process mining project methodology, in: CAiSE, Springer, 2015, pp. 297–313.
[3] T. Grisold, J. Mendling, M. Otto, J. vom Brocke, Adoption, use and management of process mining in practice, Bus. Process Manag. J. 27 (2021) 369–387.
[4] F. Zerbato, A. Burattin, H. Völzer, P. N. Becker, E. Boscaini, B. Weber, Supporting provenance and data awareness in exploratory process mining, in: CAiSE, Springer, 2023, pp. 454–470.
[5] L. Barbieri, E. Madeira, K. Stroeh, W. van der Aalst, A natural language querying interface for process mining, J Intell Inf Syst (2022) 1–30.

[6] A. Seeliger, A. Sánchez Guinea, T. Nolle, M. Mühlhäuser, Processexplorer: intelligent process mining guidance, in: BPM, Springer, 2019, pp. 216–231.
[7] K. Kubrak, F. Milani, A. Nolte, A visual approach to support process analysts in working with process improvement opportunities, Bus. Process Manag. J. 29 (2023) 101–132.
[8] N. Martin, D. A. Fischer, G. D. Kerpedzhiev, K. Goel, S. J. Leemans, M. Röglinger, W. M. van der Aalst, M. Dumas, M. La Rosa, M. T. Wynn, Opportunities and challenges for process mining in organizations: results of a delphi study, Bus. Inf. Syst. Eng. 63 (2021) 511–527.
[9] F. Zerbato, P. Soffer, B. Weber, Initial insights into exploratory process mining practices, in: BPM, Springer, 2021, pp. 145–161.
[10] A. Strauss, J. M. Corbin, Grounded theory in practice, Sage, 1997.
[11] L. Zimmermann, F. Zerbato, B. Weber, Process mining challenges perceived by analysts: An interview study, in: BPMDS, Springer, 2022, pp. 3–17.
[12] F. Zerbato, P. Soffer, B. Weber, Process mining practices: evidence from interviews, in: BPM, Springer, 2022, pp. 268–285.
[13] F. Zerbato, J. J. Koorn, I. Beerepoot, B. Weber, H. A. Reijers, On the origin of questions in process mining projects, in: EDOC, Springer, 2022, pp. 165–181.
[14] H. M. Marin-Castro, E. Tello-Leal, Event log preprocessing for process mining: a review, Applied Sciences 11 (2021).
[15] C. Capitán-Agudo, M. Salas-Urbano, C. Cabanillas, M. Resinas, Analyzing how process mining reports answer time performance questions, in: BPM, Springer, 2022, pp. 234–250.
[16] E. Sorokina, P. Soffer, I. Hadar, U. Leron, F. Zerbato, B. Weber, Pem4ppm: A cognitive perspective on the process of process mining, in: BPM, Springer, 2023, pp. 465–481.
[17] J. J. Koorn, I. Beerepoot, V. S. Dani, X. Lu, I. Van de Weerd, H. Leopold, H. A. Reijers, Bringing rigor to the qualitative evaluation of process mining findings: an analysis and a proposal, in: ICPM, IEEE, 2021, pp. 120–127.
[18] R. C. Nickerson, U. Varshney, J. Muntermann, A method for taxonomy development and its application in information systems, Eur J Inf Syst 22 (2013) 336–359.