Understanding and Improving Process Analysts' Comprehension of Process Mining Visualizations

Femke Pieters[1,*]

[1]*UHasselt, Digital Future Lab, Agoralaan Building D, 3590 Diepenbeek, Belgium*

Abstract

Process mining (PM) has become fundamental for data-driven process analysis. It enables organizations to identify inefficiencies and areas for improvement. Central to the effectiveness of PM is the use of visualizations and the ability of process analysts to comprehend the visualizations generated by PM algorithms. This doctoral research aims to address the challenge of enhancing the comprehensibility of PM visualizations for process analysts. The study has two key objectives: (1) to establish an improved understanding of how process analysts interpret PM visualizations and (2) to design and evaluate novel visualizations that improve comprehensibility. Through a combination of theoretical frameworks and empirical studies, this research develops an assessment framework to evaluate the comprehensibility of PM visualizations. Additionally, the research proposes novel visualization designs, which are assessed through a series of experiments to determine their impact on analysts' performance. The findings contribute to the field of PM by providing insights that can guide the design of more user-friendly visualizations, which facilitates better decision-making and process improvement within organizations.

Keywords

Process Mining, Comprehensibility, Visualization, Visual Analytics, Process Analyst

1. Introduction

Organizations commonly conduct process analysis to pinpoint operational problems and areas for improvement [1]. These analyses are increasingly being performed in a data-driven way given the widespread presence of process execution data captured by information systems, i.e. fine-grained data reflecting how a process is being performed in reality. Process mining (PM) enables such data-driven process analyses that use process execution data to, e.g., discover the activity order in a process or expose bottlenecks [2].

The great majority of PM algorithms provide a visual representation as an output such as a process model showing the activity order [3]. To actually contribute to organisational value creation by, e.g., identifying root causes of operational problems or process improvement opportunities, process analysts who use PM in their profession need to make sense of the provided PM output [4, 5]. Hence, it is crucial that PM visualizations are comprehensible for process analysts such that the visualization can effectively and efficiently satisfy their information needs [6]. The incomprehensibility of current PM visualizations has emerged as an important challenge for the use of PM in organizations [7]. Strikingly, there is hardly any understanding on how process analysts comprehend PM visualizations [8], even though this is critical to guide the (re)design of PM visualizations.

In this doctoral research, we aim to gather an in-depth understanding of process analysts' comprehension of PM visualizations to create novel PM visualizations with improved comprehensibility. Besides providing a highly relevant knowledge contribution, the novel visualizations enable process analysts to better comprehend the provided output, turning PM into a more powerful instrument to support data-driven process improvement. The two *research objectives* are (RO1) establish an in-depth understanding of process analysts' comprehension of PM visualizations and, (RO2) create and evaluate novel PM visualizations, designed to improve comprehensibility for process analysts.

The remainder of this extended abstract is structured as follows. In Section II, we describe our research methodology. An overview of related work is given in Section III. Finally, we conclude our research intents in Section IV.

2. Methodology

This research will take both a theoretical and empirical perspective. The proposed research objectives will be achieved following five iterative steps.

(1) Identify common PM visualizations and related process analysis tasks (preparatory to RO1 and RO2). To inform the selection of PM visualizations in subsequent research steps, the most common process analysis tasks and associated PM visualizations are identified. Published PM case studies will be collected from various sources. For each collected case study, the process analysis tasks that are conducted are recorded, together with the PM visualizations that have been used. This step results in a structured overview of common process analysis tasks and the associated PM visualizations used for this task. We will validate the overview by (a) consulting the main commercial tools and (b) conducting semi-structured interviews with academics and process analysts from industry.

(2) Create an assessment framework to assess the comprehensibility of PM visualizations, based on theory (RO1). A rapid review will be conducted to identify theories, guidelines and principles on (process) visualizations and comprehensibility. As relevant literature is scattered over various fields such as process modeling, information visualizations, visual analytics and human-computer interaction, a systematic literature review is infeasible here. Before conducting the rapid review, we will develop an appropriate protocol building upon Klerings et al. [9]. The data will be coded using the Gioia method [10], using Atlas.ti as a tool [11]. Drawing upon the results of the rapid review, an assessment framework for the assessment of the comprehensibility of PM visualizations (from a theoretical perspective) will be composed. An example theory that could be part of the assessment framework is the cognitive load theory. This states that a visualizations should help analysts process information without overloading their working memory. The cognitive load can be reduced by, for example, using color codes or showing the right amount of details [8]. The goal of our assessment framework is to be able to assess from a theoretical perspective (the components of) a visualization based on its comprehensibility.

(3) Exploratory study on process analysts' comprehension of PM visualizations (RO1). To understand how process analysts comprehend PM visualizations, an exploratory study, in which process analysts will perform certain tasks to locate particular information in the visualization, will be designed and conducted. While performing these tasks, data is collected using three complementary data collection methods: (a) clickstream traces, (b) eye-tracking and (c) concurrent think-aloud. Using these rich datasets, we will identify action patterns and intents of process analysts and relate this to their task performance. The assessment framework marking the output of the previous step will be refined with the outcomes of step 3.

(4) Assess the comprehensibility of current PM visualizations (RO2). The comprehensibility of the most common PM visualizations identified in step 1 will be assessed using the assessment framework created in step 2 and 3. To limit the impact of personal bias, dual data extraction will be used. Based on the assessment, various areas to improve PM visualizations to enhance their comprehensibility will emerge. This constitutes the basis for a research agenda containing a structured overview of areas for future research. This research agenda will not only inform step 5, but will also be highly relevant for the research community at large. While the importance of devoting more attention to the visualization of PM output is widely recognised at a conceptual level, this research agenda will provide specific directions to move forward. In this way, it has the potential to generate significant impact on the future development of the research field.

(5) Create and evaluate novel PM visualizations (RO2). Open challenges regarding the comprehensibility of current PM visualizations emerging from step 4 will be selected and tackled. In particular, the two visualizations that are commonly used and show the largest potential to enhance their comprehensibility will be selected. To create the novel PM visualizations, the procedure to design cognitively efficient

visualizations by Ware [12] will be used. An intermediate evaluation with a convenience sample of master students will be performed. Concurrent think-aloud will be used, i.e. studentds and academics are requested to perform the PM task using the created prototype while talking aloud. This intermediate evaluation can provide areas to further refine the visualizations to obtain a mature prototype. Afterwards, a between-subjects experimental design is set up, in which we evaluate the impact of the novel PM visualizations on their comprehensibility for process analysts. Participating analysts will be randomly assigned to one of two groups: the experimental or the control group. The visualization is the treatment variable, but both groups will first undergo a baseline assessment to measure their initial level of PM visualization comprehension. Following this baseline assessment, the control group will perform the experimental tasks using the original PM visualizations and the experimental group will perform the same tasks using the novel visualizations. Similar to step 3, the tasks will take the form of search and recognition tasks. Task performance metrics such as accuracy, speed, and subjective ratings of comprehension will be recorded or calculated. Clickstream traces and eye-tracking data will be collected to contextualise the analysis. A statistical analysis (ANOVA) will be performed to compare the performance of the experimental and control group on the collected task performance metrics, thereby evaluating the impact of the novel PM visualizations on their comprehensibility for process analysts.

3. Related work

Until now, only a limited number of works in PM literature have considered the perspective of the user of PM algorithms. For instance, Zerbato et al. [13] identify behavioral patterns of PM users (amongst which one process analyst from industry) that emerge during the early stages of a PM analysis. While this paper takes the perspective of PM users, it does not provide specific insights into how process analysts comprehend PM visualizations.

Prior research from the process modeling domain has shown that a variety of factors have an impact on process model comprehension, such as the modeling language used, the visual layout and the size of the model [14] However, it is not clear to which extent these findings can be transferred to PM as PM visualizations differ from traditional process models. For example, PM visualizations are embedded in interfaces that offer interaction functionalities. Traditional process models are less dynamic [2, 15].

In more recent years, synergies between PM and the research fields of visual analytics and information visualization have been explored. For example: Rehse et al. [16] use a visual analytics framework to assess how current PM tools visualize the output of conformance checking, which is a very common class of PM. This assessment highlights the need to develop and evaluate novel PM visualization idioms (i.e. forms or types of visualizations) for conformance checking output [16]. Another recent paper in this stream of literature is Yeshchenko and Mendling [17], which reviews literature to map contributions on the visualization of event sequences from information visualization and PM [17]. Future research can build upon their initial observations to create novel PM visualizations. Additionally, in one systematic study of work practices in PM, the focus lies on both visualizations and tasks. The researchers identified that problems beyond the control-flow perspective, particularly from the case perspective, are often analyzed using general-purpose visualization techniques. These techniques may lack the sophistication needed to support deeper analytical insights [18].

As mentioned in previous studies, there is a cognitive challenge faced by analysts in interpreting PM outputs. In this context, the work from Sorokina et al. [19] contributes by proposing the PEM4PPM model, a theoretical foundation that explains the cognitive processes of individual analysts during the process of process mining (PPM). Cognitive steps such as task understanding, goal refinement, data exploration and hypothesis testing play important roles in shaping how analysts approach PM tasks. As PM visualizations are central to this process, there is a growing recognition that improving their comprehensibility can significantly enhance analysts' performance and effectiveness PM tools in organizational settings [19].

4. Conclusion

In this doctoral research, we aim to understand and improve the comprehensibility of PM visualizations. First, we provide a highly relevant knowledge contribution by proposing an assessment framework to assess the comprehensibility of PM visualizations. Having a better understanding of PM visualization comprehension allows for furthering the impact of PM in organizations as process analysts can make more sense of the output of PM algorithms. Second, we design novel visualizations and we empirically research if the comprehensibility of these visualizations is improved.

References

[1] M. Dumas, M. La Rosa, J. Mendling, H. Reijers, Fundamentals of Business Process Management, Springer, 2018. doi:10.1007/978-3-662-56509-4.

[2] W. Aalst, Process Mining: Data Science in Action, Springer, 2016. doi:10.1007/978-3-662-49851-4.

[3] T. Gschwandtner, Visual analytics meets process mining: Challenges and opportunities, Lecture Notes in Business Information Processing 244 (2017) 142–154. doi:10.1007/978-3-319-53435-0_7.

[4] P. Badakhshan, B. Wurm, T. Grisold, J. Geyer-Klingeberg, J. Mendling, J. v. Brocke, Creating business value with process mining, The Journal of Strategic Information Systems 31 (2022) 2023.

[5] M. Eck, X. Lu, S. Leemans, W. Aalst, Pm^2: A process mining project methodology, Lecture Notes in Computer Science (2015) 297–313. doi:10.1007/978-3-319-19069-3_19.

[6] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas, H. Ziegler, Visual analytics: Scope and challenges, First publ. in: Lecture notes in computer science, No. 4404 (2008), pp. 76-90 1 (2008). doi:10.1007/978-3-540-71080-6_6.

[7] N. Martin, D. Fischer, G. Kerpedzhiev, K. Goel, S. Leemans, M. Roeglinger, W. Aalst, M. Dumas, M. La Rosa, M. Wynn, Opportunities and challenges for process mining in organizations: Results of a delphi study, Business Information Systems Engineering 63 (2021) 1–17. doi:10.1007/s12599-021-00720-0.

[8] J. Mendling, D. Djurica, M. Malinova Mandelburger, Cognitive effectiveness of representations for process mining, Lecture Notes in Computer Science (2021) 17–22. doi:10.1007/978-3-030-85469-0_2.

[9] I. Klerings, S. Robalino, A. Booth, C. Escobar-Liquitay, I. Sommer, G. Gartlehner, D. Devane, S. Waffenschmidt, Rapid reviews methods series: Guidance on literature search, BMJ evidence-based medicine 28 (2023). doi:10.1136/bmjebm-2022-112079.

[10] D. Gioia, K. Corley, A. Hamilton, Seeking qualitative rigor in inductive research, Organizational Research Methods 16 (2013) 15–31. doi:10.1177/1094428112452151.

[11] K. Charmaz, Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis, volume 1, SAGE Publications Ltd, 2006.

[12] C. Ware, Information Visualization: Perception for Design: Second Edition, Morgan Kaufmann, 2004.

[13] F. Zerbato, P. Soffer, B. Weber, Initial insights into exploratory process mining practices, Lecture Notes in Business Information Processing (2021) 145–161. doi:10.1007/978-3-030-85440-9_9.

[14] K. Figl, Comprehension of procedural visual business process models: a literature review, Business & Information Systems Engineering 59 (2017) 41–67.

[15] J. S. Yi, Y. a. Kang, J. Stasko, J. Jacko, Toward a deeper understanding of the role of interaction in information visualization, IEEE Transactions on Visualization and Computer Graphics 13 (2007) 1224–1231. doi:10.1109/TVCG.2007.70515.

[16] J.-R. Rehse, L. Pufahl, M. Grohs, L.-M. Klein, Process mining meets visual analytics: The case

of conformance checking, Proceedings of the 56th Hawaii International Conference on System Sciences (2022). doi:`10.48550/arXiv.2209.09712`.

[17] A. Yeshchenko, J. Mendling, A survey of approaches for event sequence analysis and visualization, Information Systems 120 (2023) 102283.

[18] C. Klinkmüller, R. Müller, I. Weber, Mining process mining practices: An exploratory characterization of information needs in process analytics, Lecture Notes in Computer Science (2019) 322–337.

[19] E. Sorokina, P. Soffer, I. Hadar, U. Leron, F. Zerbato, B. Weber, Pem4ppm: A cognitive perspective on the process of process mining, Lecture Notes in Computer Science (2023) 465–481. doi:`10.1007/978-3-031-41620-0_27`.