Interest-Driven Recommendations To Support Time Performance Analyses

Carlos Capitán-Agudo[1,*]

[1]*SCORE Lab, University of Seville, Seville, Spain*

Abstract

This paper outlines a doctoral research work that investigates how to support analyses that address time performance questions. To do that, it is investigated how process analysts behave to answer them. Aligned with insights of their behaviors, a system will be developed to recommend the most interesting insights for a certain analyst, with the objective of automating the analysis process.

Keywords

Time performance analysis, BPI challenge, Business questions, Interestingness

1. Introduction

In recent years, there has been an increasing development of process mining techniques with the objective of discovering, monitoring, and improving business processes [1]. These techniques usually focus on specific aspects of the processes, such as the time performance of process executions or the distribution of activities. However, these techniques have the problem that they require extensive previous knowledge to be applied. This problem has inspired the creation of various types of support to facilitate the use of these techniques.

On the one hand, methodologies and case studies have been developed to understand how these techniques are applied. For example, in [2] the authors evaluated the maturity of the field from a practical perspective, taking into account the adoption of tools and the use of process mining methodologies throughout the years. In [3] the authors evaluated qualitatively reports submitted to an annual competition called the Business Process Intelligence Chanllenge (BPIC). In that competition, participants receive an event log provided by a real organization along with specific business questions, and they submit a report with the analyses performed to answer them. Specifically, the objective in [3] was to understand the use of visual techniques in process mining. Furthermore, in [4] the authors performed an empirical study to observe how analysts tackle the task of initial exploration. In their study, they conducted interviews with 12 participants where they asked a general business question. To answer it, the participants examined and analyzed an event log.

On the other hand, other authors have focused on developing tool support to assist exploratory analyses. For instance, in [5] the authors developed a system that receives an event log and

✉ ccagudo@us.es (C. Capitán-Agudo)

🆔 0000-0003-2772-1740 (C. Capitán-Agudo)

recommends subsets that could contain interesting insights. To achieve this, they applied trace clustering techniques to split the log into trace subsets. Subsequently, they rank these subsets based on their level of deviation in Key Performance Indicators (KPI) or Process Performance Indicators (PPI), as well as trace diversity, from the most to the least diverse and deviating. Additionally, in [6] the authors proposed a solution to support the awareness and the track of steps in exploratory analyses. To accomplish that, they integrate the system into a process mining tool, where it is built a replayable history of the user interactions with the tool. They also provide views where the goals of the user and the applied data operations of the tool are represented as graphs.

The previous research approaches have facilitated the application of process mining techniques from a global or exploratory perspective. However, it is unknown how analysts use these techniques to address specific business questions, which usually play an important role during analysis [2]. Furthermore, no tool has been developed to provide recommendations to address specific business questions. The research goal of this doctoral research work is to extend existing tools by providing a recommendation tool to address time performance questions, with the objective of automating the analysis to answer time performance questions. The recommendations will be aligned with the behavior of the analysts. In particular, the tool support is focused on time performance questions referred to aspects such as calculating cycle times of the process, finding bottlenecks, calculating waiting times of resources, or determining time performance differences. Time performance is one of the most frequent problems in process mining [7]. From this main goal, the following research questions should be answered to achieve it:

RQ1: How do process analysts behave when answering to time performance questions?
RQ2: How can we support time performance analyses through recommendations aligned with their behaviors to automate the analysis process?

By addressing the first question, we can derive comprehensive knowledge that represents the proper approaches for conducting these analyses. To answer the second question, a tool support will be developed to make recommendations about the analysis, with the objective of automating part of the analyst process. These recommendations will be aligned with the knowledge discovered in the first question. The following objectives have been defined to answer them:

OBJ1: Understand how process analysts behave when dealing with time performance questions
OBJ2: Provide support through recommendations aligned with the analyst behaviors to automate the process

2. Methodology

This doctoral research work will follow the principles of Design Science Research (DSR) to develop the needed support [8]. This research methodology focuses on the creation of artifacts that improve the quality of life of our society through technological progress. As artifact is understood any element containing knowledge such as models, methods, or theories. The first

step of DSR is to identify a specific problem and collect data related to it. Secondly, the objectives and requirements to solve that problem must be defined. Thirdly, the data collected in the first step is analyzed, and the artifact is designed. Fourthly, the usefulness of the artifact to address the problem is evaluated. Finally, the process is documented and the results are communicated to other researchers.

Taking into account the previous methodology, the following steps will be performed to answer the questions. By thoroughly reviewing related literature, we will collect analyses that address time performance questions. Specifically, we will collect these analyses from different sources such as reports or screen and audio recordings. The objectives have already been defined and highlighted in the Introduction. To achieve these objectives, we will firstly conduct a mixed quantitative-qualitative study of these analyses to understand how they are performed. Secondly, we will qualitetively develop a set of codes related to the actions that were performed in the analyses inspired by Grounded Theory [9]. Afterwards , we will apply both quantitative and qualitative methods to obtain conclusions of the codes (e.g., clustering techniques, similarity measures). These conclusions will be a set of insights describing the behavior of analysts to address time performance questions. Once, these insights are obtained, a recommender system will be developed to help analysts answer them. This system will support users throughout the process, offering automatic recommendations for doing time performance analyses to answer business questions. These recommendations will be aligned with the insights of the analyst behaviors. A study involving experts in process mining will be conducted to evaluate the tool

3. Research state

The first step of this research has been completed, and the specific problem has been highlighted in the Introduction.

The data collection step of the investigation is close to completion, and a significant portion of the required data has been gathered. To gather the reports, we have analyzed the questions related to all the BPICs identifying those that were related to time performance. Subsequently, we collected reports submitted to the BPIC that had a clear section to answer time performance questions. In total, four BPI challenges had time performance questions, and 110 answers were selected. With respect to the recordings, a dataset composed of 12 videos of analysts answering time performance questions has been gathered, where the conduct of the analysts was captured in real time during their analysis process.

The study of the previous data is in advanced state. With respect to the answers of the BPICs, they were independently coded by two researchers who shared their codings, and they have already been analyzed. These results results were published in the BPM 2022 [10], and are described next. A set of codes of 55 different operations was obtained (e.g., Filter traces) and a set of 137 different ways to implement these operations (e.g. Filter traces by activities), which were called variants. These operations had different purposes such as analyze time (e.g., Calculate cycle time) or manipulate the data (e.g., Filter traces). Moreover, we applied the Kmeans algorithm to classify the answers in different groups. The best results were obtained for four clusters using the Silhouette index as evaluation measure. Each group represented an answer type that were characterized using depending on the operations used along with their

frequencies. We also measured the similarities between the answers depending on different contexts(e.g., the author profile, the objective of the question) to see their influence. These insights have recently been expanded with the dataset of recordings. This dataset could enrich and generalize them by adding the conduct of analysts. This extension of the results is under revision in a scientific journal.

Finally, the development of the recommender system is in an early stage [11]. We are currently designing the recommendation process and establishing its alignment with the previous insights. One way to perform the recommendations could be to assess the operations related to process metrics and manipulation methods of the discovered catalog that are interesting for a time performance analysis of a specific user. This evaluation could be performed using measures that assess the interestingness of association rules [12], which have already been used to recommend filtering methods that provide the most interesting data subsets or summaries in data mining [13]. These measures assess the interestingness of the data based on various aspects, including conciseness (few attribute-value pairs in the result) and generality (a result that encompasses a significant subset of the dataset). Thus, the challenge is to make them suitable for process mining and the mentioned operations.

Acknowledgments

References

[1] W. M. P. van der Aalst, Process Mining – Data Science in Action, Second Edition, Springer, 2016. doi:10.1007/978-3-662-49851-4.

[2] F. Emamjome, R. Andrews, A. H. ter Hofstede, A Case Study Lens on Process Mining in Practice, in: OTM Conferences, 2019, pp. 127–145.

[3] C. Klinkmüller, R. Müller, I. Weber, Mining Process Mining Practices: An Exploratory Characterization of Information Needs in Process Analytics, in: Int. Conf. on Business Process Management (BPM), 2019, pp. 322–337.

[4] F. Zerbato, P. Soffer, B. Weber, Initial insights into exploratory process mining practices, in: BPM Forum, 2021, pp. 145–161.

[5] A. Seeliger, A. Sánchez Guinea, T. Nolle, M. Mühlhäuser, ProcessExplorer: Intelligent Process Mining Guidance, in: Business Process Management (BPM), 2019, pp. 216–231.

[6] F. Zerbato, A. Burattin, H. Völzer, P. N. Becker, E. Boscaini, B. Weber, Supporting provenance and data awareness in exploratory process mining, in: Int. Conf. on Advanced Information Systems Engineering (CAiSE), 2023, pp. 454–470.

[7] M. L. van Eck, X. Lu, S. J. J. Leemans, W. M. van der Aalst, PM^2: A Process Mining Project Methodology, in: Int. Conf. on Advanced Information Systems Engineering (CAiSE), 2015, pp. 297–313.

[8] K. Peffers, T. Tuunanen, M. A. Rothenberger, S. Chatterjee, A Design Science Research Methodology for Information Systems Research, J. Manag. Inf. Sys. 24 (2007) 45–77.

[9] K. Stol, P. Ralph, B. Fitzgerald, Grounded theory in software engineering research: a critical review and guidelines, in: ICSE, 2016, pp. 120–131.

[10] C. Capitán-Agudo, M. Salas-Urbano, C. Cabanillas, M. Resinas, Analyzing how process mining reports answer time performance questions, in: Business Process Management (BPM), 2022, pp. 234–250.

[11] C. Capitán-Agudo, M. Salas-Urbano, C. Cabanillas, M. Resinas, Recommending interesting results in process mining analysis, in: JCIS 2022, 2022.

[12] R. J. Hilderman, H. J. Hamilton, Knowledge Discovery and Measures of Interest, Kluwer Academic Publishers, 2001.

[13] T. Milo, C. Ozeri, A. Somech, Predicting "What is interesting" by mining interactive-data-analysis session logs, in: Int. Conf. on Extending Database Technology (EDBT), 2019, pp. 456–467.