Exploring Expert Behavior of Process Mining Analysts

Jessica Van Suetendael[1]

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

Abstract

Within the field of Process Mining, the research topic known as Process of Process Mining emerged. Process of Process Mining seeks to comprehend process mining analysts' thought processes and behavioral patterns. A better understanding of the analysts' behavior allows for the design of more effective process mining tools and the construction of more efficient process mining training. The behavioral differences between expert and non-expert process mining analysts have not yet been studied within this domain. Therefore, my PhD research aims to conceptualize behavior in Process of Process Mining research and identify expertise-related behavioral differences among process mining analysts.

Keywords

Process Mining Behavior, Expert Behavior, Exploratory Process Mining

1. Introduction

While process mining research has focused extensively on the technology side, studies related to cognitive labor by Process Mining (PM) analysts are rather limited [10]. Therefore, this project studies behavioral differences among PM analysts of varying levels of expertise. We argue that different levels of expertise are related to different behaviors of PM analysts, and understanding these differences is crucial for designing more effective process mining tools and constructing more efficient process mining training.

The hypothesis for this relation between expertise performance and behavior is rooted in the research on expert performance and the concept of situation awareness. Expert performance has been defined in the literature as consistently exhibiting superior performance for domain-representative tasks [5]. A critical framework to explain the source of expert performance is the theory of Situation Awareness (SA) [4]. SA refers to an up-to-date understanding of the state of the world and forms a critical concept in the field of human decision-making. SA consists of three levels: perception of the elements in the environment, comprehension of the current situation, and projection of the future state, given specific actions. The relation to expertise lies in the fact that experts have a more effective situation assessment process, resulting in higher SA, which leads to better actions to tackle the challenge ahead [3]. Therefore, differences between experts and non-experts are reflected in the actions performed to assess the current situation and to solve the problem at hand.

Studying PM analysts' actions aligns with Process of Process Mining, which examines individual behaviors during process mining. Insights gained from better understanding PM analysts' behavior facilitate the development of supportive tools [11]. Previous research on expertise in Process of Process Mining was inconclusive and rather limited [10]. Furthermore, expertise was defined as years of experience, which is known to be a questionable proxy for expertise [5]. Therefore, this project will focus on the relationship between behavior and expertise to advance Process of Process Mining's state of the art by uncovering the PM analyst's behavior using event data [11].

ICPM 2024 Doctoral Consortium, October 14–18, 2024, Kongens Lyngby, Denmark
jessica.vansuetendael@uhasselt.be (J. V. Suetendael)
0009-0002-6017-323X (J. V. Suetendael)

2. Research Objectives

This research aims to conceptualize behavior in Process of Process Mining research and identify expertise-related behavioral differences among PM analysts. This research is motivated by the hypothesis that understanding expertise-related differences among PM analysts improves training and tool support, enhancing data-driven process analysis. Methodologically, the study uses ideas from behavioral informatics and process research employing a multi-modal data collection approach that combines event data and verbal reports. In terms of scope, this project restricts its focus to the context of exploratory process mining, which is defined as tasks aimed at gaining insights and identifying process patterns from data with open-ended questions; it excludes predictive process mining and non-process related analyses.

There are four research objectives defined for this research:

- RO1: Develop a vocabulary of behavior exhibited by PM analysts.
- RO2: Develop an approach to measure behavior and performance of PM analysts.
- RO3: Collect behavioral data of PM analysts with varying degrees of expertise.
- RO4: Identify expertise-related differences in PM analysts' behavior.

3. Methodology

The work plan for this project is organized into work packages one to four, corresponding to research objectives one to four, respectively.

3.1. Workpackage 1: Develop a Vocabulary of Behavior Exhibited by PM Analysts

The first work package aims to develop a PM analyst ethogram cataloging possible behaviors during exploratory process mining tasks, regardless of expertise. A literature review of exploratory process mining case studies will be conducted to create the initial ethogram, as case studies provide more detailed behavior descriptions than regular research papers. The ethogram will be refined through semi-structured interviews with experienced PM analysts, who will recount past exploratory process mining analyses. When the behavior or intent is unclear, follow-up questions will be asked. After each interview, the ethogram will be updated by adding intent to existing behavior descriptions as well as adding new behaviors. Additional interviews will be conducted iteratively until the ethogram no longer needs to be updated and data saturation is reached.

The updated ethogram's usability for encoding behavior in exploratory process mining will be validated. PM analysts will screen-record a real-life exploratory process analysis and then review it with a researcher to discuss their actions and intent. The researcher will code the behavior using the ethogram. The ethogram is ready if it successfully encodes the behavior; otherwise, it needs further updates.

3.2. Workpackage 2: Develop an Approach to Measure Behavior of PM Analysts

The goal of this work package is to create a precise conceptual definition of behavior in the context of exploratory process mining and an accompanying approach to record such behavior.

Firstly, a conceptual definition of behavior in the context of exploratory process mining will be devised. The definition of behavior in Behavioral Informatics [2] will be used as a starting point and adapted to the context of Process of Process Mining. Additionally, the idea of internalizing situational and sequential context from Pentland et al. [8] will also be integrated. The output of this task shifts Process of Process Mining research from an action to a behavior perspective. Based on the preliminary research, an initial structure has been constructed which needs to be further refined during the project execution. Behavior is currently operationalized as a behavioral vector represented by three components: Subject, Intent, and Context. Intent represents the goal of the behavior. Context is conceptualized as a sequence of actions described by a separate action vector. Each element of this vector consists of an Object, an Action, the Impact of the action, and a Timestamp, mixing both sequential and situational context.

To record behavior during exploratory process mining in a controlled manner, specific process mining tasks need to be constructed. Each task consists of a specific exploratory process mining question, a process log, and a set of (hidden) challenges/peculiarities in the process logs. The type of process mining questions will be inspired by the insights from work package 1. To increase sensitivity to detect different expertise levels, the factors "task type", "task complexity" and "task environment" will be varied. These three factors are known to influence the task's cognitive load [6]. For "task type", we distinguish two types: schema-based and non-schema-based. Schema-based tasks require the application of a fairly fixed procedure to solve the problem. Non-schema-based tasks require the analyst to go beyond the directly provided information and to use some creativity to solve the task [9]. For the factor "task complexity", three levels of complexity will be introduced, determined by the number of hidden challenges incorporated into the task. In total six different tasks will be constructed, applying a full-factorial design over the factors task type and task complexity. Finally, the task environment, e.g. time constraints, will be varied across the six tasks.

After developing the method for recording process mining behavior, the method will be tested through a pilot study. The pilot study will entail a limited number of participants executing the six designed process mining tasks. A mix of participants with different professional backgrounds, ages, and years of experience will be selected. The participants will execute the designed tasks on separate days to minimize the impact of fatigue. For each task, the participant will receive the process log which needs to be analyzed using bupaR, which allows the direct export of an action log. A screen recording of the exploratory process analysis will be made. Final answers to the task's questions will be recorded in an online form. After task completion, a retrospective think-aloud will be conducted with the participant while reviewing the screen recording. Together with the ethogram, the think-aloud transcript will be encoded and the behavioral vectors will be constructed. The action log exported with bupaR will be used to construct the action vectors for the behavior context.

For every process mining task, an online form will collect the PM analyst's answer to the related question. These answers will be graded by multiple evaluators to eliminate evaluator bias. Since experts are defined as PM analysts who consistently show superior performance, two measures per participant will be computed. For each participant, the grades for all their completed tasks will be aggregated into the mean or median grade to express the level of performance. Additionally, the standard deviation or interquartile range will be used to express the level of performance consistency. Do note that our goal is to identify experts rather than predict expertise, which allows us to focus directly on performance rather than some precursor of expertise such as the commonly used, albeit questionable [5], concept of experience. The goal of the pilot study is to identify unexpected issues prior to the full-scale experimental study and evaluate the sensitivity to detect differences in expertise level. If the pilot study is insufficiently capable of distinguishing high from low expertise levels, alterations will be made. This could imply changes to the number or complexity of the tasks or the operationalization of expert performance measurements. The results of the pilot study will inform the correct course of action.

3.3. Workpackage 3: Collect Behavioral Data of PM Analysts with Varying Degrees of Expertise

This work package aims to set up a large-scale observational study to collect behavioral data on PM analysts with varying expertise during multiple exploratory process mining tasks. A similar setup to the pilot study will be followed for the experimental study. Participants will be invited to solve the six validated exploratory process mining tasks from the previous work package. Each participant will tackle each task on a separate day to minimize fatigue. The screen recordings and think-aloud retrospective transcripts will be used to identify the different behaviors, while the R history log will be used to reconstruct the sequential and situational context. The online form with the answers will be used and corrected by multiple evaluators with extensive process mining experience, and a performance and performance consistency score will be computed for each participant. Background and experience-related information about the participants will be collected through a brief survey at the start of the experiment.

Additionally, initial insights into the analyst's behavior during exploratory process mining tasks will be generated using the four different techniques suited for the analysis of ethogram-encoded data. First, kinematic diagrams will be constructed. These diagrams, known for their role in illustrating motion and dynamics in ethological studies [7], will be adapted to depict behavior types and their transitions during process analysis. This provides a visual depiction of the sequence of behaviors by the PM analyst and makes the PM analyst's behavioral dynamics insightful.

Secondly, time budgets will be formulated to gain a better understanding of the time distribution spent on activities undertaken by PM analysts [7]. Time budgets allow for systematic exploration of how PM analysts allocate their time, prioritize tasks, and navigate the dynamic landscape of exploratory process analysis. As time budgets only show the duration of activities, timeband plots will also be generated to gain a deeper understanding of the sequencing of activities. These plots offer a detailed and intuitive way to showcase the temporal aspects of the analyst's work, enabling a comprehensive analysis of how time is utilized throughout the course of their exploratory analysis.

Finally, behavioral kinetics diagrams will be constructed to show each behavior type individually over time, considering all PM analysts. If, at a specific point in time, none of the PM analysts are engaged in a particular behavior, the graph for that behavior will have a value of zero. Conversely, if three PM analysts exhibit the behavior at a certain point in time, the graph for that behavior will have a value of three. This allows us to visualize at which times during the analysis certain behaviors are more prominent than others. By doing so, we aim to gain insights into the specific phases of the exploratory process analysis when particular behaviors are more prominent.

3.4. Workpackage 4: Identify Expertise-Related Differences in PM Analysts' Behavior

This work package analyzes expertise-related behavioral differences among PM analysts using data from the previous work package. Expertise will be measured through average performance and performance consistency, and participants will be clustered accordingly to represent a specific type of expertise: e.g. consistent high-performers, inconsistent high-performers, consistent low-performers, and inconsistent low-performers.

First, narrative networks will be used to uncover behavioral differences between the different expertise levels with respect to the manifestation of a specific type of behavior. Driving questions include: do PM analysts perform different actions for the same behavior, in a different order, on different parts of the data, …? Narrative networks are weighted, directed graphs where the node represents event categories and the edges represent sequential relationships between events. Recent research [8] illustrated that increasing the granularity of these events - i.e. incorporating situational context - holds great potential to disentangle and visualize organizational processes. In our setting, a separate narrative network per expertise level will be constructed for a specific behavior. Nodes will represent the actions taken by the analyst, defined at a fine-grained level, e.g. "compute the mean of variable X in dataset Y" instead of "compute descriptive statistics", while edges represent the order between these actions. Because Pentland et al. [8] showed that adding contextual elements can increase the structure and interpretability of narrative networks, we will use this to identify the key contextual elements to analyze the behavior. Contextual elements will be incorporated into the action description and if the structure of the narrative network increases, then the contextual element is considered important to the analysis of behavior and behavioral differences. Furthermore, the final narrative networks will be analyzed and interpreted, as well as scanned for deviant behavior (represented as isolated parts of a graph). Finally, the narrative networks of different expertise level will be compared to identify structural differences.

Second, a mutually exciting multivariate spatial point process model [1] will be used to analyze differences in how behavior evolves throughout the exploratory process mining task. More specifically, we estimate the probability that a specific behavior occurs at a given time during the analysis. Additionally, we are also interested in the factors that influence this probability and how these probabilities differ among different levels of expertise. The basic idea behind the proposed method is to model the probability that a specific behavior occurs at time t given a history H of past behaviors. The proposed model is multivariate in the sense that we are modeling the probabilities for multiple behavior types simultaneously. Because the probability of a particular behavior at time t is influenced by the occurrence of behaviors prior to time t, the model is referred to as "an exciting point process model". The spatial dimension of the model will be exploited in a similar innovative way as in [1] to model the effect of contextual similarity between past and current behavior. From the estimated parameters, it is possible to determine the baseline probabilities of specific behaviors, as well as how past behavior influences these probabilities, how fast the influence of past behavior decays, and to what extent changes in situational context (e.g. characteristics of the working data) trigger a different kind of behavior.

Acknowledgments

This study was supported by the Special Research Fund (BOF) of Hasselt University under Grant No. BOF23OWB03.

References

[1] Aggarwal, V., Hwang, E. H., & Tan, Y. (2021). Learning to be creative: a mutually exciting spatiotemporal point process model for idea generation in open innovation. Information Systems Research, 32(4), 1214-1235.

[2] Cao, L. (2010). In-depth behavior understanding and use: the behavior informatics approach. Information Sciences, 180(17), 3067-3085.

[3] Endsley, M.R., 1995. Toward a Theory of Situation Awareness in Dynamic Systems. Hum Factors 37, 32–64. https://doi.org/10.1518/001872095779049543

[4] Endsley, M.R. (2018). Expertise and Situation Awareness. In Ericsson, K. A., Hoffman, R. R., & Kozbelt, A. (Eds.), The Cambridge handbook of expertise and expert performance. Cambridge University Press.

[5] Ericsson, K.A. (2018). An Introduction to the Second Edition of The Cambrigde Handbook of Expertise and Expert Performance: Its Development , Organization and Content. In Ericsson, K. A., Hoffman, R. R., & Kozbelt, A. (Eds.), The Cambridge handbook of expertise and expert performance. Cambridge University Press.

[6] Kelton, A. S., Pennington, R. R., & Tuttle, B. M. (2010). The effects of information presentation format on judgment and decision making: A review of the information systems research. Journal of Information Systems, 24(2), 79-105.

[7] Peltier, C., Lejeune, F. X., Jorgensen, L. G., Rametti-Lacroux, A., Tanguy, D., Godefroy, V., ... & ECOCAPTURE study group. (2022). A temporal classification method based on behavior time series data in patients with behavioral variant of frontotemporal dementia and apathy. Journal of Neuroscience Methods, 376, 109625.

[8] Pentland, B. T., Recker, J., Wolf, J. R., & Wyner, G. (2020). Bringing context inside process research with event data. Journal of the Association for Information Systems, 21(5), 5.

[9] Ritchi, H., Jans, M., Mendling, J., & Reijers, H. A. (2020). The influence of business process representation on performance of different task types. Journal of Information Systems, 34(1), 167-194.

[10] Sorokina, Elizaveta, et al. "PEM4PPM: A Cognitive Perspective on the Process of Process Mining." International Conference on Business Process Management. Cham: Springer Nature Switzerland, 2023.

[11] Zerbato, F., Zimmermann, L., Völzer, H., & Weber, B. (2023). ProMiSE: Process Mining Support for End-Users.