# Noise Detection in Task Mining for RPA Implementation

Tom Hohenadl[1,*]

[1]*Business Administration and Business Informatics, Catholic University of Eichstätt-Ingolstadt, Germany*

**Abstract**

Robotic Process Automation (RPA) has gained significant attention in recent years as a technology enabling organizations to automate repetitive tasks and improve operational efficiency. Task mining, a technique used to capture users interactions with software systems, plays a crucial role in understanding user behavior and identifying automation opportunities in RPA implementations. However, task mining data often contains noise, which is erroneous or irrelevant action data, that can affect the accuracy and reliability of analysis results and synthesized bots. This extended abstract presents a research proposal focused on noise filtering techniques specifically tailored for user interaction data recorded for RPA implementation.

**Keywords**

Robotic Process Automation (RPA), Task mining, Noise Filtering

## 1. Introduction

Task mining is a growing research area and technology domain. It is a part of process mining and Business Process Management (BPM). Both academia and industry have shown growing interest in understanding user actions in business processes [1]. Currently, process mining and BPM techniques are mainly focused on information systems [2]. Most techniques in the process mining domain focus on static information provided from system logs [3]. Yet, capturing ad-hoc user actions and user processes is a BPM problem to be solved in future research [1]. Task mining delivers process insights on a user interaction level through recording of clicks or keystrokes and adds significant benefits for process enhancement [4]. In addition, the analysis of task mining logs facilitates the implementation of Robotic Process Automation (RPA) scripts, so called bots. Where currently a lot of manual initiative is necessary for identifying RPA bot opportunities and their implementation [5], better bots can be created utilizing task mining methods [6]. Leno et al. [7] combined the interconnection of task mining and RPA by creating the Robotic Process Mining (RPM) framework.

While task mining focuses on capturing user actions in user interaction logs, discovering user processes, checking behavior conformance, and enhancing manual processes [8], RPM uses task mining techniques to generate bots without the need for manual analysis or implementation

*ICPM Doctoral Consortium and Demo Track 2023*

*Corresponding author.

tom.hohenadl@stud.ku.de (T. Hohenadl)

0000-0002-7501-97803 (T. Hohenadl)

effort, thereby automating user behavior across different information systems [9]. Although bots created using the available methods can work, they are still susceptible to erroneous or irrelevant behavior recorded in the logs [10]. This behavior is defined as noise, e.g. attribute noise [11], and is not relevant for analysis of manual processes or implementation [10]. When implementing this behavior into bots automatically, the benefits of RPA, such as efficiency and quality [12], are marginal due to erroneous RPA bot executions, which lead to failures and increased maintenance effort [13]. These factors diminish the advantages of RPA compared to manual work.

The consequences of faulty bot implementation, driven by noisy user interaction logs, highlight two core areas of focus for this research proposal: the identification of noise-indicating properties within user actions recorded in interaction logs, and the development of methods for cleaning such noise from these logs. The primary objective of the research is to create an artifact capable of identifying such noisy actions so that subsequently automated RPA bots are not affected by erroneous or irrelevant user behavior. In addition, the purpose of this research project is to evaluate the transformation of actual user actions into automated RPA routines during bot implementation and to determine the actions that are important for automation.

## 2. Research Proposal and Research Questions

As outlined by Mayr et al. [8] there are three particular challenges to overcome in task mining research at the moment: segmentation of logs, privacy preservation of user related information, and noise filtering. This proposal focuses on the noise filtering aspects of the described challenges.

So far, the preprocessing of user interaction logs is performed within the developed RPM recorders, e.g. actionLogger [14] or smartRPA [9]. These developed Design Science Research (DSR) artifacts use hard-coded static rules to remove unwanted or incorrect user behavior. On the one hand, Leno et al. mention exemplary rules, such as repeated navigation or copy actions without pasting [14]. On the other hand, Agostinelli et al. identify the quality of the recorded fine-grained data as a weakness of the RPM artifact [9], so that not all contingencies may be mapped to static rules. Thus, this proposal entails research targeting the detection and removal of noise in user interaction logs and is grounded on the following research questions:

- What user actions are relevant for RPA?
- How can irrelevant actions be identified and removed after recording users and before RPA bot creation?
- Do existing algorithms for noise detection from the process mining domain exceed the capabilities of hard-coded static rule based filters from existing RPM DSR artifacts?
- Is it possible to create a better filtering method based on the results of the previous research questions?

## 3. Research Methodology

Initially, a structured literature review based on guidelines by Kitchenham et al. [15] was conducted. This literature review was used to identify user action properties described in the

existing literature. Based on the acquired data a taxonomy of user actions in manual processes was created following the directive of Nickerson et al. [16].

Following the literature review and the identification of action categories, a semi-structured interview study is used to identify translation patterns. A translation is the mapping of user actions to bot functionality. The interview study is based on the methods described in Saldana [17]. The guiding questions for the interview is "Which user actions, intrinsic to users daily routines, are being incorporated into the operational framework of an RPA bot?". By answering this question and encoding the interviewees' answers, an overview of actions that are valuable in RPA bots, as well as actions that are nonfunctional, can be created.

Based on the theoretical findings of the first two research questions, a comparative analysis is conducted to examine existing process mining noise detection methods and their applicability for task mining. Initially, students at the researchers institute are instructed to perform a range of manual processes on their digital devices and record the corresponding actions to create a reference data set. This data set serves as a baseline for the subsequent comparison of algorithms. By applying Leno's et al. [18] and Agostinelli's et al. [19] hard-coded static noise filtering rules on the data set, an initial base line to remove noise or irrelevant events from the log can be established. The disparity between the original logs and the filtered logs will subsequently be employed to identify appropriate noise detection algorithms. A set of outlier and noise detection algorithms is reviewed by Koschmieder et al. [11] . The set of publicly available algorithms, i.e. source code or DSR artifacts, will be used as an initial set of comparable algorithms to the baseline data. Furthermore, a literature review and forward backward search will be used to identify further relevant noise filtering algorithms.

The final objective of this study is to develop a DSR artifact aimed at eliminating the reliance on hard-coded filtering. The findings from the comparative study of available algorithms as well as the interview study will be used to elaborate noise filtering techniques for user interaction logs. Based on these results, a DSR noise filtering artifact will be developed. The reference data set will be labeled in collaboration with the students and will be used as a basis for evaluation using scores such as precision, recall or the F-score. The artifact aims to establish a robust data preprocessing foundation for the automated creation of RPA bots by leveraging filtered user interaction logs.

## 4. Current Research Status

The initial literature review to identify user actions currently processed in task mining and RPM has been conducted. The result is a taxonomy containing six categories of value adding user actions and a category for not relevant actions. These value adding actions categories are opening, navigating, transforming, transferring, concluding and closing actions. Furthermore, the category of empty actions, i.e. doing nothing at all, was identified. However, a validation of these categories through practical use cases or empirical evaluation is still necessary.

Following the literature review, the initial two stages of the qualitative interview study have been accomplished. The first stage involved the formulation of interview questions and the identification of the target group of participants, specifically focusing on RPA developers and researchers in the fields of RPM or task mining. In the second stage, interviews were conducted

with a total of five RPA developers, four task mining researchers and four RPA business experts. Following the interviews, an initial round of process coding is conducted. This will be followed by the categorization of the codes and a second cycle coding approach.

## 5. Conclusion

In the evolving research and industry landscape, the automation of user processes is achieved using RPA, which has opened up new avenues for improving process efficiency and effectiveness. RPA implementation, in turn, is further enhanced by leveraging task mining and the RPM framework. While research on task mining is increasing and industry interest shifts towards automated automation, the task mining, RPA and RPM domain still has challenges. The described research proposal aims at improving the automatic creation of RPA bots by removing noise from user interaction logs. Consequently, this work addresses relevant aspects in the convergence of task mining, RPA and RPM, bridging gaps in the pursuit of efficient and effective automation solutions.

## References


[1] I. Beerepoot et al, The biggest business process management problems to solve before we die, Computers in Industry 146 (2023) 103837. doi:10.1016/j.compind.2022.103837.

[2] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers, Fundamentals of business process management, softcover re-print of the hardcover 2nd edition 2018 ed., Springer, Berlin, 2018. URL: https://link.springer.com/book/10.1007/978-3-662-56509-4.

[3] W. van der Aalst et al, Process mining manifesto, in: F. Daniel, K. Barkaoui, S. Dustdar (Eds.), Business Process Management Workshops, Lecture Notes in Business Information Processing, Springer, Berlin, 2012, pp. 169–194.

[4] W. van der Aalst, Process mining - Data science in action, second edition ed., Springer, Berlin and Heidelberg, 2016. doi:10.1007/978-3-662-49851-4.

[5] W. van der Aalst, On the pareto principle in process mining, task mining, and robotic process automation, DATA 2020 - Proceedings of the 9th International Conference on Data Science, Technology and Applications (2020). doi:10.5220/0009979200050012.

[6] S. Agostinelli, A. Marrella, M. Mecella, Research challenges for intelligent robotic process automation, Business Process Management Workshops, BPM 2019 (2019) 12–18. URL: doi.org/10.1007/978-3-030-37453-2_2.

[7] M. Dumas, M. La Rosa, V. Leno, A. Polyvyanyy, F. M. Maggi, Robotic process mining, in: W. M. P. van der Aalst, J. Carmona (Eds.), Process Mining Handbook, Springer International Publishing, Cham, 2022, pp. 468–491. URL: doi.org/10.1007/978-3-031-08848-3_16.

[8] A. Mayr, L.-V. Herm, J. Wanner, Christian Janiesch, Applications and challenges of task mining: A literature review, ECIS 2022 Research-in-Progress Papers (2022). URL: https://aisel.aisnet.org/ecis2022_rip/55.

[9] S. Agostinelli, M. Lupia, A. Marrella, M. Mecella, Automated generation of executable rpa scripts from user interface logs, in: A. Asatiani (Ed.), BPM 2020 Blockchain and RPA Forum

Proceedings, volume 393 of *Lecture Notes in Business Information Processing*, Springer International Publishing, Cham, 2020, pp. 116–131. URL: doi.org/10.1007/978-3-030-58779-6_8.

[10] S. Agostinelli, M. Lupia, A. Marrella, M. Mecella, Smartrpa: A tool to reactively synthesize software robots from user interface logs, in: S. Nurcan, A. Korthaus (Eds.), Intelligent Information Systems, volume 424 of *Lecture Notes in Business Information Processing*, Springer International Publishing and Imprint Springer, Cham, 2021, pp. 137–145. URL: doi.org/10.1007/978-3-030-79108-7_16.

[11] A. Koschmider, K. Kaczmarek, M. Krause, S. J. van Zelst, Demystifying noise and outliers in event logs: Review and future directions, in: A. Marrella, B. Weber (Eds.), Business Process Management Workshops, volume 436 of *Springer eBook Collection*, Springer International Publishing and Imprint Springer, Cham, 2022, pp. 123–135. URL: doi.org/10.1007/978-3-030-94343-1_10.

[12] A. Meironke, S. Kühnel, How to measure rpa's benefits? a review on metrics, indicators, and evaluation methods of rpa benefit assessment, in: Wirtschaftsinformatik 2022 Proceedings, 2022, pp. 1–19. URL: https://aisel.aisnet.org/wi2022/bpm/bpm/5.

[13] P. Noppen, I. Beerepoot, I. van de Weerd, M. Jonker, H. A. Reijers, How to keep rpa maintainable?, in: D. Fahland, C. Ghidini, J. Becker, M. Dumas (Eds.), Business Process Management, volume 12168 of *Springer eBook Collection*, Springer International Publishing and Imprint Springer, Cham, 2020, pp. 453–470. URL: doi.org/10.1007/978-3-030-58666-9_26.

[14] V. Leno, A. Polyvyanny, M. Dumas, F. Maggi, Action logger: Enabling process mining for robotic process automation, in: Proceedings of the Dissertation Award, Doctoral Consortium, and Demonstration Track at BPM 2019, 2019, pp. 124–128. URL: https://bia.unibz.it/esploro/outputs/conferenceProceeding/Action-logger-Enabling-process-mining-for/991006186495001241.

[15] B. Kitchenham, Procedures for performing systematic reviews, Keele, UK, Keele Univ 33 (2004). URL: http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf.

[16] R. C. Nickerson, U. Varshney, J. Muntermann, A method for taxonomy development and its application in information systems, European Journal of Information Systems 22 (2013) 336–359. doi:10.1057/ejis.2012.26.

[17] J. Saldaña, The coding manual for qualitative researchers, 4e ed., SAGE, Los Angeles and London and New Delhi and Singapore and Washington DC and Melbourne, 2021.

[18] V. Leno, A. Polyvyanyy, M. La Rosa, M. Dumas, F. M. Maggi, Action logger: Enabling process mining for robotic process automation, in: Proceedings of the Dissertation Award, Doctoral Consortium, and Demonstration Track at BPM 2019, volume 2420 of *CEUR Workshop Proceedings*, CEUR-WS, 2019, p. 5. URL: https://hdl.handle.net/10863/19700.

[19] S. Agostinelli, A. Marrella, M. Mecella, Exploring the challenge of automated segmentation in robotic process automation, in: Research Challenges in Information Science 2021, 2021, pp. 38–54. URL: doi.org/10.1007/978-3-030-75018-3_3.