6-18-2022

Applications and Challenges of Task Mining: A Literature Review

Alexander Mayr
TU Dortmund University, alexander.mayr@uni-wuerzburg.de

Lukas-Valentin Herm
University of Würzburg, lukas-valentin.herm@uni-wuerzburg.de

Jonas Wanner
University of Würzburg, jonas.wanner@uni-wuerzburg.de

Christian Janiesch
TU Dortmund University, christian.janiesch@tu-dortmund.de

APPLICATIONS AND CHALLENGES OF TASK MINING: A LITERATURE REVIEW

Research in Progress

Alexander Mayr, TU Dortmund University, Dortmund, Germany, alexander.mayr@tu-dortmund.de

Lukas-Valentin Herm, University of Würzburg, Würzburg, Germany, lukas-valentin.herm@uni-wuerzburg.de

Jonas Wanner, University of Würzburg, Würzburg, Germany, jonas.wanner@uni-wuerzburg.de

Christian Janiesch, TU Dortmund University, Dortmund, Germany, christian.janiesch@tu-dortmund.de

Abstract

Task mining is a technological innovation that combines current developments in process mining and data mining. Using task mining, the interactions of workers with their workstations can be recorded, processed, and linked with the business data of the organization. The approach can provide a holistic picture of the business processes and related tasks. Currently, there is no overview of application scenarios and the challenges of task mining. In our work, we reflect application scenarios as well as technological, legal, and organizational challenges of task mining using a structured literature review. The application areas include discovery of automation potentials, monitoring, as well as optimization of business processes. The challenges include the cleansing, collection, data protection, explainability, merging, organization, processing, and segmentation of task mining data.

Keywords: task mining, literature review, applications, challenges, research agenda.

1 Introduction

Task mining (TM) is an innovation that combines current developments from the fields of process mining (PM) and machine learning. In this approach, analog to the event logs of PM (van der Aalst et al., 2012) user interface data of users at workstations during the processing of defined tasks is collected by different tools. With this data, methods of data mining can be used to link the respective manual actions of workers – that may not have been defined or foreseen in the process – to the corresponding business processes and thus to trace the effects on the respective company goals (Celonis, 2020).

TM can be employed in for a variety of applications and it is unclear if the common scenarios of process improvement still hold. Further, those scenarios are confronted with several technological, legal, and business challenges. Due to the currency of this innovation, there is no survey of the above-mentioned aspects of TM available yet. Consequently, we aimed to close the gap with a structured literature analysis. We research and derive potential application scenarios as well as the challenges of the technological innovation of TM employed in these scenarios. Thus, this paper's research can be summarized with the following research questions:

RQ 1 *Which application scenarios exist in the context of task mining?*

RQ 2 *Which challenges are associated with the technological innovation of task mining?*

We first present the theoretical foundations of the approaches of PM and TM (Section 2). Then, we define our methodology (Section 3), before we present the synthesis and evaluation of the extracted literature (Sections 4) to answer our research questions. Concluding, we present a summary, limitations, and an outlook to highlight open research items (Section 5).

2 Process Mining & Task Mining

Process mining is an approach to extract information from process data and analyze it systematically to monitor and optimize real business processes or to derive new value-adding processes (van der Aalst, 2016). For all techniques of PM, it is assumed that it is possible to record events sequentially through executed activities. In addition, information such as timestamps, resources involved, and data elements captured with the event are recordable (van der Aalst et al., 2012). Nevertheless, not all information is stored within the respective process-based systems. In particular, this applies to system interfaces. So, there might be discrepancies between the operational reality and the recorded process data that distorts the information obtained from the process logs (van der Aalst, 2012).

TM provides a complementary approach to extend PM to the collection and processing of low-level user interaction data. This user interaction data covers the individual steps a user takes when using their workstation (Celonis, 2020). The aim is to obtain a holistic picture of the business processes to optimize them at task level in terms of value creation. For this purpose, the recorded data is extended by common context data. Analog to PM, the recorded interaction data is processed using various data mining tools. In this way, the individual successive activities can be classified and linked to the respective resources such as users or systems. By linking users and business processes, user activities can be compared with the business data, so that the impact of every reaction associated to every task to the flow of business processes can be traced. Further, the collected user interaction data can be used to gather information about the variety of possible approaches of users when using software to perform defined tasks (Leno et al., 2020). Thus, TM allows for several process optimizations that extend the possibilities of traditional PM. The systems at hand can be comprehensively analyzed and then optimized by uncovering the core steps of each process and associated workarounds. Furthermore, repetitive tasks can be identified which can be substituted by automation approaches. By linking interaction data and business data, one can optimize (sequences of) activities to execute tasks.

3 Research Methodology

Literature Search. We used the literature search framework according to vom Brocke et al. (2009) for our structured literature review. We aimed to identify research contributions dealing with TM and focused on peer-reviewed contributions from the following databases: ACM Digital Library and IEEEXplore for computer science-related contributions, AIS eLibrary and ScienceDirect for information system-related contributions as well as ESCBOhost Business Source Premier and Web of Science for economics-related contributions. We used the following pseudocode for our search term: *((’process mining’ OR ’task mining’ OR ‘desktop activity mining’) AND (’log’ OR ’task’ OR ’challenge’ OR ‘application’)).* While including ‘process mining’ may sound counterintuitive, it enabled us to collect more relevant TM research as many authors link their research to the general term of PM. Appendix A comprises the employed search queries in more detail (Mayr et al., 2022).

First, we queried all databases with the presented search term, which resulted in 720 potentially relevant publications. Based on multiple selection criteria (only contributions in English or German; removal of duplicates) and an abstract and keyword analysis, 132 contributions were then considered for further literature search. We then performed a full-text analysis, which yielded 51 contributions that we considered relevant as they addressed the application and challenges of TM. Based on these contributions, we performed a forward search using Google Scholar to find currently published research that cite the discovered contributions. We further examined this combined body of literature with a backward search. By performing a full-text analysis and re-using the same selection criteria as before, our forward and backward searches yielded an additional 72 publications. As a result, we consider 123 publications as relevant for the literature synthesis. The distribution of contributions according to the different steps is shown in Figure 1.

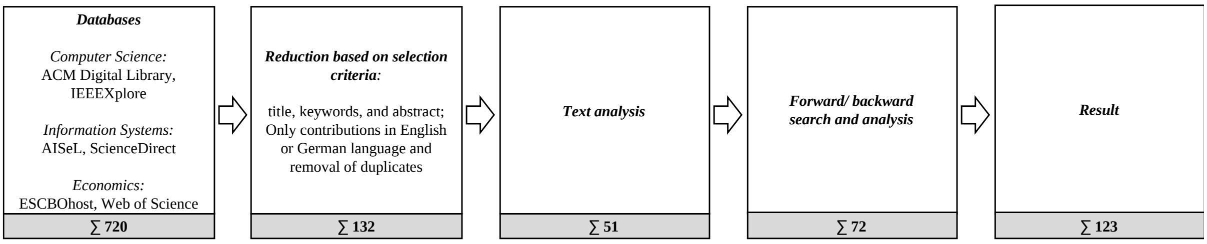

Figure 1. Overview of literature review.

Synthesis of Literature. We used a concept matrix-based approach according to Webster and Watson (Webster and Watson, 2002). Firstly, we analyzed all contributions to find classification dimensions. Similar dimensions were grouped together (e.g., “change management” and “human resources” are grouped into the subdimension *Organization*). Subsequently, we analyzed all 123 contributions again according to the identified dimensions. As introduced earlier, we distinguish *Applications* and *Challenges* as dimensions, which we further segregate into sub-dimensions to capture the specific concepts that are relevant in a particular context. We divided applications into *Discovery of Automation Potentials*, *Monitoring*, and *Optimization*. Accordingly, we divided challenges into *Cleansing*, *Collection*, *Data Protection*, *Explainability*, *Merging*, *Organization*, *Processing*, and *Segmentation*. We further detail and describe the sub-dimensions in the following two sections. For the concept matrix that led to this distinction, see Appendix B (Mayr et al., 2022).

Descriptive Statistics. The examined literature addresses applications (n=34) and/or challenges (n=110). We also found that it primarily deals with the technological challenges (n=156). Other challenges have been discussed at a lesser rate (n=67). Figure 2 provides an overview of the distribution per subdimensions. Some contributions were classified into several (sub-)dimensions.

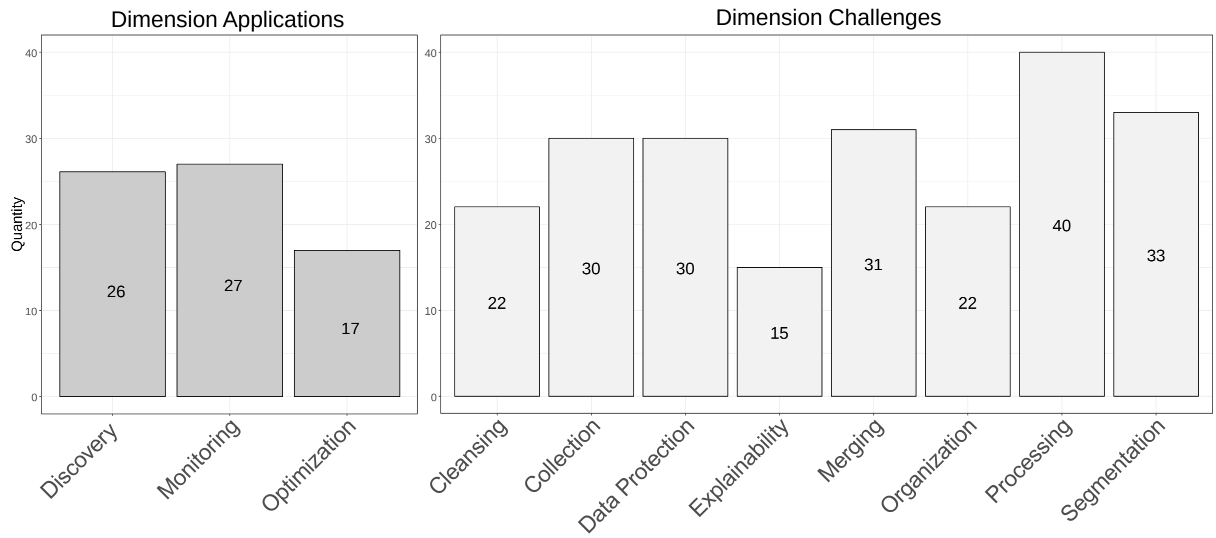

Figure 2. Descriptive statistics on the sub-dimensions.

4 Application Scenarios and Challenges

4.1 Application Scenarios

We identified three general application scenarios for the collection and analysis of user interaction data. These are the applications of monitoring, optimizing, and discovering of processes.

Monitoring of Processes. Most organizations define standard process flows, work instructions, or best practices to be embedded in a system. However, in practice many defined organizational processes are supplemented by manual processes that are allowed to bypass defined organizational regulations (so-called shadow processes), resulting in unanticipated changes. These, can endanger the security of the system because users could perform fraudulent tasks (Kerremans et al., 2020). TM can check compliance with pre-defined corporate guidelines (van der Aalst, 2012) to check for compliance with guidelines (Bezerra and Wainer, 2011), to detect of potentially fraudulent input (Kerremans et al., 2020), or to compare and analyze the same processes across different employees, business units, subsidiaries, or countries to identify dependencies (Fischer et al., 2019).

Optimization of Processes. The efficiency of business processes is a critical success factor for organizations. In addition to increasing process effectiveness through qualitative means such as project groups and surveys, this can be achieved through analytical instruments. Formal theories and techniques are used to model and derive the process design. The focus is on the identification of process parameters extracted by TM. In addition, the collected data in combination with event logs allow for business process simulations. This enables to uncover process optimization potentials and to pre-check planned process changes (Zhu et al., 2014).

Furthermore, bottlenecks such as shortage or interface errors can be reliably identified (van der Aalst, 2012). In practice, organizations use a variety of information sources to gain a comprehensive insight into a business process to create a simulation model. This includes process documentation, expert interviews, and the observation of employees during the execution of process steps (Zhu et al., 2014). However, this is subject to biases. Interviews with experts can lead to contradictory information. The perception of experts, that is human perception in general, tends to be subjectively distorted. Likewise, there are differences between a simulation and reality, as employees often work faster and more efficiently when monitored (Hawthorne effect). Therefore, additional event logs created by PM must be used to create more accurate simulations (Rozinat et al., 2009). The procedures associated with the process can be recorded at any time without changing the work environment of the people involved (Martin et al., 2014). This results in increased data quality. Instead, the collected user interaction data can be used to gather information about the variety of possible approaches of users when using software to perform defined tasks (Leno et al., 2020). By linking the manual actions and their effects on business objectives, process optimizations can be made and stalled in the organization as best-practice solutions (Celonis, 2020).

Discovery of Automatable Process Tasks. TM is suitable for the identification of automatable processes, like in the context of robotic process automation (RPA). Current research in process automation aims at supporting more complex and less tightly defined tasks using artificial intelligence technology, such as text analysis of documentation (Leopold et al., 2018). The goal is that process automation tools learn to adapt and handle non-standard cases by observing (and mimicking) human behavior. Additionally, the interaction between process automation and humans opens up further potential (Astromskis et al., 2015). The dilemma of process automation arises from the fact that the corresponding interaction data is only available after the automation has been completed. This can lead to limited business value of the automation for new cases. To this issue, TM can offer a solution (Kerremans et al., 2020). The collected process and task data can be interpreted in the process automation context as operational data. It is also possible to check, feed, and optimize process automation already in operation with the aggregated TM information (Aguirre and Rodriguez, 2017). In addition, TM allows to improve the success rate of automation through visualization and a connected, better understanding of process interrelationships and workflows. This allows process automation to adapt to necessary process changes without losing relevance. A further advantage of the connection of process and TM in the process automation context is the support in the definition and prioritization of automation options at the process step level (van der Aalst et al., 2018).

4.2 Challenges

We could identify eight generic challenges of TM. These are data collection, cleansing, merging, segmentation, processing, data protection, model explainability, and practical implications for organizational use.

Collecting the Data. The data collected during TM can originate from a variety of potential sources. In addition to event logs, which contain detailed information about activities performed in systems (van der Aalst, 2016), TM also processes user interaction data. Firstly, this data must be collected from a variety of different tools. The input data of the workstations are collected by mouse trackers and key loggers. Further interactions with the respective systems can also be recorded. These include the recording and processing of all entered texts, the screen surface in the form of screenshots, as well as information from e-mail traffic and print logs (Epure et al., 2015, van der Aalst et al., 2018).

New technologies including the processing of RFID information (Linn et al., 2018), all ubiquitous mobile systems, as well as alternative input methods such as touch screens should be considered (Gerke et al., 2009). Furthermore, the derivation of process or context information via the sensor technology of smart products such as wearables as well as the facial expressions of employees at the workplace is conceivable (de Medeiros et al., 2005, Shinyama et al., 2018). These tools must reduce the recording to relevant, that is meaningful and value-adding actions. This leads to a challenge since the evaluation of the meaningfulness of actions is only possible if the respective context is known. In addition, repetitive and non-value-adding activities such as copying data twice without pasting or navigating between worksheet cells without editing the data must be detected and filtered out (Leno et al., 2020). Also, the granularity of the collected data must be considered. It should be fine enough to fully reconstruct the task at hand, which is crucial for the discovery of rule-based decision logic or performance analysis. Likewise, TM tools must be able to record actions in a way that allows the impact of actions on different platforms, systems, and in different contexts to be captured even if process times of several employees are mapped (van Eck et al., 2016). Furthermore, tools should record in a format that is supported by other tools to enable interoperability.

Cleansing of Data. Next, the data must be filtered and cleansed to enable processing. This cleansing processes leads to different challenges such as noise within the captured data. This noise hast to be defined, recognized, and adjusted (Leno et al., 2020). A possible approach to solving this task is to treat noise as chaotic events that can occur anytime (Leno et al., 2019). However, events can be misinterpreted as chaotic events if the captured interaction data has a high variance. Accordingly, the shape of data must also be considered when cleansing. This includes changing values and sources of data (Leno et al., 2020). Additionally, some attributes and sources occur less frequently (data sparsity). These can be mistakenly interpreted as noise when they occur (van der Aalst et al., 2012). Furthermore, data that is not noise may be redundant and should be removed from the dataset as well. This occurs, for example, if a user performs a task step twice due to a self-inflicted error. Accordingly, frequency-based approaches for filtering the corresponding data are only suitable to a limited extent. Accordingly, it may be necessary to manually check low-frequency behavior for conformity (van der Aalst et al., 2012). Sequential pattern mining techniques, which distinguish between intentional events and outlier events, are also only of limited suitability, since, analogous to rare attributes and sources, corresponding data can be interpreted as outliers and therefore be filtered in error (Leno et al., 2020). As a result, additional process knowledge of analysts is necessary to identify and filter noise, outliers, and redundant data in TM (Tax et al., 2019).

Merging of Data. The collected and purified data from the different sources must be merged afterward (Wen et al., 2007). This is a complex task in practice. The collected data often show a heterogeneous structure. This is, for example, because the different data sources sometimes produce very different data sets and different identifiers can be used in comparable data sets. Another problem is the different degrees of granularity, quality, and completeness of the data (van der Aalst, 2012). The subsequent derivation of the corresponding information may require considerable effort (Pospiech et al., 2014) or even be impossible (van der Aalst et al., 2012). Information systems often store information in unstructured form (Xu and Liu, 2019), which can make the intended meaningful merging of log files significantly more difficult (Banziger et al., 2018). Data can be collected across many tables or must be extracted from different messages by subsystems. The increasing prevalence of service technology, cloud computing, and supply chain integration also means that the corresponding event logs may be available from different organizations. This leads to the fact that the problem of the different data formats and structures continuously increase, and the respective sub-steps of the processes are partly recorded by different systems. Furthermore, the linking of event data with context data implied by TM can lead to a significant increase in variables, which significantly increases the complexity of the analysis or even makes it impossible. The merging of data should be done in a way that the final data sets are suitable for further processing in corresponding frameworks (Claes and Poels, 2012). The data structure should be adapted so that the combination with other analysis methods is possible (van der Aalst et al., 2012).

Segmentation of Data. In practical applications, the processing of collected data is complicated by the variety of the observed behavior, which is due to the large number of consecutive sequences in which activities are carried out. This can lead to the results of the evaluation being imprecise or difficult to understand. The segmentation of the data into homogeneous groups can help to overcome this problem (Măruşter and Beest, 2009). The merged and not preprocessed data consists of a sequence of events recorded during one session (Leno et al., 2020). In this session, a user may have performed different executions of one or more tasks. Accordingly, information about several tasks is available in the respective logs. The interaction data contained in these logs reflects the specific order in which the user performed the tasks. Accordingly, the logs must be segmented into individual tracks so that each track corresponds to the execution of a task, across different systems and applications. Heuristic approaches such as timestamp-based limitations are a solution but may be unreliable due to users often needing different times for the same processes (Leno et al., 2020).

The problem of segmentation can be solved in principle with different approaches of correlating uncorrelated or unlabeled events in logs used for PM (De Koninck and De Weerdt, 2019). However, the available approaches require specific conditions such as a defined process model or the absence of cycles or repetitions (Bayomie et al., 2019). Moreover, current approaches provide relatively unreliable solutions (Leno et al., 2020). Therefore, the approaches seem to be only partly suitable for the segmentation of TM datasets, since these properties contradict some of the necessary goals, such as the optimization of process models at the task level and the identification of tasks that can be automated. In the context of RPA, a routine defined by a bad segmentation can lead to high costs, especially if unsupervised bots are used (Leno et al., 2020). A possible solution is the approach of explicit inclusion of expert knowledge (Ferreira and Gillblad, 2009).

Processing of Data. To process these extensive and complex event logs, a suitable data mining tool is required- Generally, most existing algorithms scale very poorly with the number of activities (Prodel et al., 2018). These algorithms are therefore of limited use in the context of TM. Accordingly, there are several approaches to process complex and large TM datasets to overcome the complexity of the datasets. Other approaches consider formalizations of data extraction and task identification as a problem of extracting attributes as process components and relationships between process components using sequencing techniques to reduce the effort and increase the accuracy of the formalized constructs (Banerjee and Gupta, 2015). Another approach is periodic TM developed for process optimization in of software development. Here, periodic tasks together with information about their periods and response profiles are obtained of real-time systems (Li et al., 2015).

Nevertheless, we could not identify a generally valid approach for processing the interaction data. However, there are several specific solutions for PM for certain application areas or industries. For a detailed overview of these areas and tools and frameworks related to them, refer to Iegorov et al. (2017). The choice of the tool to be used for processing the data is complicated by the fact that there is no comprehensive framework for benchmarking existing algorithms (van der Aalst et al., 2012).

Data Protection. Data protection includes the protection of collected data from third parties and informal self-determination. TM datasets enable making conclusions about the work of individual units through the analysis of the metadata of the individual process steps (Dakic et al., 2018b). The use of this data is often unavoidable for achieving consistent results (Mannhardt et al., 2019). The comparatively high degree of transparency increases the inherent risk of violating privacy (Dakic et al., 2018b). This risk also increases if the collection or analysis of the respective data is carried out by third parties. This must be prevented in terms of data protection. According to Mendes and Vilela (2017), there are eight strategies for the protection of privacy in conformity with the General Data Protection Regulation (GDPR) and can be regarded as requirements for privacy-observing TM systems. The amount of personal data should be minimized and abstracted, and not be freely accessible. Likewise, a legally compliant data protection policy should be established, which allows the user to have traceable control and knowledge about the handling of their personal data. These strategies can be transferred to event logs and thus represent the prerequisite for the design of privacy-preserving PM (Dakic et al., 2018b) and thus also for privacy-preserving TM.

These requirements can limit the quantity of data and thus the quality of the results that can be generated by data mining models. This can lead to trade-offs between the accuracy of the results and data protection. To resolve these, different approaches exist to address the challenge of data mining in compliance with data protection. These are summarized in the literature under the term privacy-preserving data mining (Mannhardt et al., 2019). On the one hand, there are some techniques during the actual data collection process while at the same time preserving the value of the collected data as best as possible (Hoepman, 2014). These include processes that distort the original data in such a way that it can no longer be used by third parties (Mannhardt et al., 2019), for example by adding (multiplicative) noise (Zaman et al., 2019). On the other hand, according to Mannhardt et al. (2019), there are special techniques for protecting privacy when publishing or forwarding and processing corresponding data records (Kim and Winkler, 2003, Li et al., 2007, Machanavajjhala et al., 2007, Samarati and Sweeney, 1998, Xiao and Tao, 2006) as well as to ensure privacy in data mining outputs (Aggarwal and Philip, 2008, Dwork, 2006, Verykios, 2013). There are also several frameworks for privacy-preserving data mining. These aim to find the most suitable of the methods mentioned depending on the method used and the area of application. At present, there is no specific framework for TM that respects data protection.

Model Explainability. The use of machine learning is often indispensable for processing the extensive datasets of TM or at least represents a valuable tool for detecting hidden relationships or anomalies in the dataset. Due to their high complexity, many machine learning models represent so-called black boxes (Shoshani, 1982). Understanding how the outputs of the respective models are generated can have a similar relevance in a variety of applications as the accuracy of the prediction itself. Explainability of machine learning applications is also laid down by law (Ribeiro et al., 2016). This makes the economic use of black-box approaches more difficult, resulting in a practical challenge. GDPR obliges organizations to make the results of corresponding models traceable. The research area of explainable AI (XAI) has been established (Holzinger et al., 2017) to extract concepts or explanatory models from the input data, which can be interpreted by humans. Currently, there are no XAI guidelines that ensure legal and socio-technical explainability in TM or PM.

Practical Implications for Organizational Use. In addition to the technological, legal, and human aspects, TM also poses several other practical challenges for use in organizations. With the processing of all interaction data, TM represents a significant change from traditional data collection and processing methods. Accordingly, the use of TM in an organizational context places high demands on the change management of the respective organization. This is primarily due to the employees' fear of being overtaxed by new requirements resulting from the change, fear of criticism of the previous working process due to the detailed analysis through TM, and fear of possible job loss due to the automation with RPA (Eaton, 2010). Accordingly, the implementation of TM is a complex task for the management, especially concerning employee motivation (Helmke et al., 2013). When TM is implemented in an organization, further negative implications can be associated from the use of the system. As mentioned earlier, one of the objectives of TM is to create new standardized processes. However, the formalization of processes also implies some negative effects. These include the inhibitory effect of too narrowly setting standards on creativity, innovative strength, and flexibility in solving complex tasks (Lauer, 2019). Additionally, it has been demonstrated that formalizing processes carries the risk that employees will reduce performance to the minimum and limit initiatives above the set level (Mintzberg, 2000).

5 Limitations and Research Agenda

The structured literature review enabled us to derive potential applications of TM for monitoring, optimization as well as the discovery of potential for automation. Furthermore, we were able to identify technological, legal, and organizational challenges. Our research is not without limitations as we focused on the derivation of the state-of-the-art within TM. That is, we conducted a literature review with a representative scope of literature, meaning there is a possibility that we did not consider all relevant research contributions.

Based on the discussion above, we identify the following open research items (RI) in the sense of a research agenda for TM. They stem from those sub-dimensions that we found to be the most prevalent for information systems research in the ongoing discussion. We acknowledge that for other fields, particularly data science, other priorities may be reasonable.

RI 1: Research in the Context of Practical Applications. We identified several Applications of TM. However, the practical application and its impact on organizations have hardly been researched compared to other aspects (Dakic et al., 2018a). While there is apparently a lot of movement in the vendor market, actionable examples and guidelines are scarce. This deficit should be compensated by appropriate more research on the applications of TM.

RI 2: Frameworks for the Comparison of Existing Methods and Software. As explained in the Processing Challenge, there are several different methods for handling the collected interaction data. However, the ensuing results are highly dependent on the scope, quality, and structure of the corresponding dataset. Furthermore, the selection and application of the respective tool often requires extensive expertise in the respective business processes as well as the existing methods for processing and analyzing data. The selection is further complicated by the fact that there is no comprehensive framework for benchmarking existing algorithms (van der Aalst et al., 2012). This research gap must be closed to validate future and existing approaches and make them better comparable.

RI 3: Framework in the Context of Explainability Methods. As detailed in the Model Explainability Challenge, the explainability of the outputs generated by the respective data mining methods must conform to certain guarantees due to legal regulations. As many current state-of-the-art methods rely on machine and in particular deep learning peeking into the black box becomes a necessity. There is a multitude of corresponding methods to ensure the transparency of the corresponding models. During our literature analysis, however, we could not identify a specific framework in the context of PM or TM. Future research should check for the suitability of common XAI frameworks for PM as well as TM based not only on data interpretability but also on initial findings of social evaluations of model explainability (Wanner et al., 2022).

RI 4: Frameworks for the Comparison of Data Protections. As explained in the Data Protection Challenge, privacy of data is a major challenge in the context of TM. In response, there are a variety of potential methods to ensure it. These should be combined in a specific framework. The framework for privacy-preserving PM of (Mannhardt et al., 2019) could be transferable to a large extent due to the similarity of PM and TM. However, the unique aspects of TM are not yet considered explicitly and would need to be addressed.

Further, it should be noted that there is a general need for research in the context of TM in addition to the research agenda outlined above. Due to the high degree of innovation in TM, there is hardly any specific literature. Accordingly, there is a need for research for all aspects explained in the paper.

References

Aggarwal, C. C. and Philip, S. Y. (2008). *Privacy-preserving data mining: models and algorithms*. Boston, MA: Springer.

Aguirre, S. and Rodriguez, A. (2017). "Automation of a business process using robotic process automation (RPA): A case study," In: *Proceedings of the Workshop on engineering applications*, pp. 65-71.

Astromskis, S., Janes, A. and Mairegger, M. (2015). "A process mining approach to measure how users interact with software: an industrial case study," In: *Proceedings of the 2015 International Conference on Software and System Process*, pp. 137-141.

Banerjee, A. and Gupta, P. (2015). Extension to alpha algorithm for process mining. *International Journal of Engineering and Computer Science*, 4 (09), 14383-14386.

Banziger, R., Basukoski, A. and Chaussalet, T. (2018). "Discovering Business Processes in CRM Systems by leveraging unstructured text data," In: *Proceedings of the IEEE 4th International Conference on Data Science and Systems*, pp. 1571-1577.

Bayomie, D., Di Ciccio, C., La Rosa, M. and Mendling, J. (2019). "A Probabilistic Approach to Event-Case Correlation for Process Mining," In: *Proceedings of the International Conference on Conceptual Modeling*, pp. 136-152.

Bezerra, F. a., bio and Wainer, J. (2011). Fraud detection in process aware systems. *International Journal of Business Process Integration and Management*, 5 (2), 121-129.

Celonis (2020). *What is Task Mining?*: https://www.celonis.com/process-mining/what-is-task-mining/.

Claes, J. and Poels, G. (2012). "Merging Computer Log Files for Process Mining: An Artificial Immune System Technique," In: *Proceedings of the Business Process Management Workshops*. Ed. by F. Daniel, K. Barkaoui and S. Dustdar, pp. 99-110.

Dakic, D., Stefanovic, D., Cosic, I., Lolic, T. and Medojevic, M. (2018a). "Business Process Mining Application: A Literature Review," In: *Proceedings of the Proceedings of the 29th DAAAM International Symposium*, pp. 866-875.

Dakic, D., Stefanovic, D., Cosic, I., Lolic, T. and Medojevic, M. (2018b). "Business Process Mining Application: A Literature Review," In: *Proceedings of the 29th DAAAM International Symposium*, pp. 866-875.

de Koninck, P. and de Weerdt, J. (2019). "Scalable Mixed-Paradigm Trace Clustering using Super-Instances," In: *Proceedings of the 2019 International Conference on Process Mining (ICPM)*, pp. 17-24.

de Medeiros, A. K. A., van Dongen, B. F., van der Aalst, W. M. P. and Weijters, A. J. M. M. (2005). "Process Mining for Ubiquitous Mobile Systems: An Overview and a Concrete Algorithm," In: *Proceedings of the Ubiquitous Mobile Information and Collaboration Systems*. Ed. by L. Baresi, S. Dustdar, H. C. Gall and M. Matera, pp. 151-165.

Dwork, C. (2006). Differential privacy, in automata, languages and programming. *International Colloquium on Automata, Languages, and Programming (ICALP). Lecture Notes in Computer Scienc*, 4052, 112.

Eaton, M. (2010). Why change programmes fail - Mark Eaton examines why many improvement initiatives stall. *Training Journal*, 53.

Epure, E. V., Martin-Rodilla, P., Hug, C., Deneckere, R. and Salinesi, C. (2015). "Automatic process model discovery from textual methodologies," In: *Proceedings of the 9th International Conference on Research Challenges in Information Science*, pp. 19-30.

Ferreira, D. R. and Gillblad, D. (2009). "Discovering process models from unlabelled event logs," In: *Proceedings of the International Conference on Business Process Management*, pp. 143-158.

Fischer, M., Imgrund, F., Janiesch, C. and Winkelmann, A. (2019). Directions for future research on the integration of SOA, BPM, and BRM. *Business Process Management Journal*, 25 (7), 1491-1519.

Gerke, K., Claus, A. and Mendling, J. (2009). "Process mining of RFID-based supply chains," In: *Proceedings of the 2009 IEEE Conference on Commerce and Enterprise Computing*, pp. 285-292.

Helmke, S., Brinker, D. and Uebel, M. (2013). Aufgaben im Change Management zur erfolgreichen Einfuhrung von CRM. In *Effektives Customer Relationship Management*. Wiesbaden: Springer, pp. 277-287.

Hoepman, J.-H. (2014). "Privacy Design Strategies," In: *Proceedings of the ICT Systems Security and Privacy Protection*. Ed. by N. Cuppens-Boulahia, F. Cuppens, S. Jajodia, A. Abou El Kalam and T. Sans, pp. 446-459.

Holzinger, A., Biemann, C., Pattichis, C. S. and Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? *arXiv preprint arXiv:1712.09923*.

Iegorov, O., Torres, R. and Fischmeister, S. (2017). "Periodic task mining in embedded system traces," In: *Proceedings of the 2017 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, pp. 331-340.

Kerremans, M., Searle, S., Srivastava, T. and Iijima, K. (2020). *2021 Market guide for process mining*. https://www.gartner.com/en/documents/3991229/market-guide-for-process-mining.

Kim, J. and Winkler, W. (2003). *Multiplicative noise for masking continuous data*. Research Report Series (Statistics #2003-01).

Lauer, T. (2019). *Change Management: Grundlagen und Erfolgsfaktoren*. Berlin, Heidelberg: Springer.

Leno, V., Polyvyanyy, A., Dumas, M., La Rosa, M. and Maggi, F. M. (2020). Robotic Process Mining: Vision and Challenges. *Business & Information Systems Engineering*, 1-14.

Leno, V., Polyvyanyy, A., La Rosa, M., Dumas, M. and Maggi, F. M. (2019). "Action logger: Enabling process mining for robotic process automation," In: *Proceedings of the 17th International Conference on Business Process Management, (BPM'19), Vienna, Austria*, pp. 124-128.

Leopold, H., van der Aa, H. and Reijers, H. A. (2018). "Identifying Candidate Tasks for Robotic Process Automation in Textual Process Descriptions." Cham, pp. 67-81.

Li, J., Wang, H. J. and Bai, X. (2015). An intelligent approach to data extraction and task identification for process mining. *Information Systems Frontiers*, 17 (6), 1195-1208.

Li, N., Li, T. and Venkatasubramanian, S. (2007). "t-closeness: Privacy beyond k-anonymity and l-diversity," In: *Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering*, pp. 106-115.

Linn, C., Zimmermann, P. and Werth, D. (2018). "Desktop activity mining-a new level of detail in mining business processes," In: *Proceedings of the Workshops der Informatik 2018-Architekturen, Prozesse, Sicherheit und Nachhaltigkeit*, pp. 245-258.

Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1 (1), 3-es.

Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M. and Michael, J. (2019). Privacy-preserving Process Mining: Differential. *Informatik Spektrum*, 42 (5), 349-351.

Martin, N., Benoit, D. and Caris, A. (2014). "Event log knowledge as a complementary simulation model construction input," In: *Proceedings of the 2014 4th International Conference On Simulation And Modeling Methodologies, Technologies And Applications (SIMULTECH)*, pp. 456-462.

Măruşter, L. and Beest, N. (2009). Redesigning business processes: A methodology based on simulation and process mining techniques. *Knowledge and Information Systems*, 21 (3), 267-297.

Mayr, A., Herm, L.-V., Wanner, J., and Janiesch, C. (2022). Appendix for Mayr et al. "Applications and Challenges of Task Mining: A Literature Review". EUDAT B2SHARE. https://doi.org/10.23728/b2share.4d33fde314e84f2cb63f73707518b032.

Mendes, R. and Vilela, J. P. (2017). Privacy-preserving data mining: methods, metrics, and applications. *IEEE Access*, 5, 10562-10582.

Mintzberg, H. (2000). *The rise and fall of strategic planning*: Pearson Education.

Pospiech, S., Mielke, S., Mertens, R., Pelke, M., Jagannath, K. and Stadler, M. (2014). "Exploration and analysis of undocumented processes using heterogeneous and unstructured business data," In: *Proceedings of the 2014 IEEE International Conference on Semantic Computing*, pp. 191-198.

Prodel, M., Augusto, V., Jouaneton, B., Lamarsalle, L. and Xie, X. (2018). Optimal process mining for large and complex event logs. *IEEE Transactions on Automation Science and Engineering*, 15 (3), 1309-1325.

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016). ""Why should I trust you?" Explaining the predictions of any classifier," In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135-1144.

Rozinat, A., Wynn, M. T., van der Aalst, W. M. P., ter Hofstede, A. H. M. and Fidge, C. J. (2009). Workflow simulation for operational decision support. *Data & Knowledge Engineering,* 68 (9), 834-850.

Samarati, P. and Sweeney, L. (1998). "Generalizing data to provide anonymity when disclosing information," In: *Proceedings of the PODS, Vol. 98*, pp. 275487-275508.

Shinyama, Y., Arahori, Y. and Gondow, K. (2018). "CodeCAM: capturing programmer's reaction during coding session," In: *Proceedings of the 2018 IEEE Workshop on Mining and Analyzing Interaction Histories (MAINT)*, pp. 1-5.

Shoshani, A. (1982). "Statistical Databases: Characteristics, Problems, and some Solutions," In: *Proceedings of the VLDB, Vol. 82*, pp. 208-222.

Tax, N., Sidorova, N. and van der Aalst, W. M. P. (2019). Discovering more precise process models from event logs by filtering out chaotic activities. *Journal of Intelligent Information Systems,* 52 (1), 107-139.

van der Aalst, W. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems*, 3 (2), 1-17.

van der Aalst, W., Adriansyah, A., de Medeiros, A. K. A. and al., e. (2012). "Process Mining Manifesto," In: *Proceedings of the Business Process Management Workshops*. Ed. by F. Daniel, K. Barkaoui and S. Dustdar. Berlin, Heidelberg, pp. 169-194.

van der Aalst, W. M. P. (2016). *Process Mining: Data Science in Action*. 2nd Edition. Berlin: Springer.

van der Aalst, W. M. P., Bichler, M. and Heinzl, A. (2018). Robotic process automation. *Business & Information Systems Engineering: The International Journal of Wirtschaftsinformatik,* 60 (4), 269-272.

van Eck, M. L., Sidorova, N. and van der Aalst, W. M. P. (2016). "Enabling process mining on sensor data from smart products," In: *Proceedings of the 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*, pp. 1-12.

Verykios, V. S. (2013). Association rule hiding methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 3 (1), 28-36.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A. and others (2009). "Reconstructing the giant: on the importance of rigour in documenting the literature search process," In: *Proceedings of the European Conference on Information Systems*, pp. 2206-2217.

Wanner, J., Herm, L.-V., Heinrich, K. and Janiesch, C. (2022). A social evaluation of the perceived goodness of explainability in machine learning. *Journal of Business Analytics,* 5 (1).

Webster, J. and Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.

Wen, L., van Der Aalst, W. M. P., Wang, J. and Sun, J. (2007). Mining process models with non-free-choice constructs. *Data Mining and Knowledge Discovery*, 15 (2), 145-180.

Xiao, X. and Tao, Y. (2006). "Anatomy: Simple and effective privacy preservation," In: *Proceedings of the 32nd international conference on Very large data bases*, pp. 139-150.

Xu, J. and Liu, J. (2019). A Profile Clustering Based Event Logs Repairing Approach for Process Mining. *IEEE Access,* 7, 17872-17881.

Zaman, R., Cuzzocrea, A. and Hassani, M. (2019). "An Innovative Online Process Mining Framework for Supporting Incremental GDPR Compliance of Business Processes," In: *Proceedings of the 2019 IEEE International Conference on Big Data (Big Data)*, pp. 2982-2991.

Zhu, X., Recker, J. C., Zhu, G. and Maria, F. a., via (2014). Exploring location-dependency in process modeling. *Business Process Management Journal*, 20 (6), 794-815.