Mining for User-Defined Categorizations as an Approach for Process Simplification in Business Process Discovery

Full research paper

Leonard Nake
Chair for Information Management
Martin Luther University Halle-Wittenberg
Halle (Saale), Germany
Email: leonard.nake@wiwi.uni-halle.de

Stephan Kuehnel
Chair for Information Management
Martin Luther University Halle-Wittenberg
Halle (Saale), Germany
Email: stephan.kuehnel@wiwi.uni-halle.de

Abstract

Business process discovery approaches analyze event logs to create process models describing the current state of the underlying processes. Because this type of process mining is especially relevant for low-structured processes, there are approaches designed to deal with such processes by simplifying the resulting model. Such simplifications are primarily applied with metrics based on the frequency of observed behaviours. However, a high frequency of a certain behaviour is not synonymous with high relevance to the user. Consequently, this paper applies a design science research approach to develop and implement a business process discovery approach mining for user-defined categories to guarantee relevance to the respective user. This categorization approach is demonstrated in a case study that serves as the basis for an evaluation regarding its perceived usefulness in an expert survey. Although the method currently requires the user to categorize activities before simplifying, it is perceived to be useful by experts.

Keywords business process discovery, design science research, process simplification, relevance, activity aggregation

1 Introduction

Since business processes are the key to organizing, improving and understanding procedures in organizations (Weske 2012), their quality is highly important for decision-makers. Process mining techniques improve the quality of business processes by revealing actual behaviours in organizations instead of depending on modelled behaviours (Mans et al. 2015). Specifically, business process discovery, one of the three main types of process mining, deals with the generation of business process models in the control-flow perspective through the analysis of event logs (Van der Aalst 2016). The resulting models describe the status quo of observed behaviour in the underlying organization, which improves the basis for decision-making. However, in order to enable optimal decisions based on complex business process models, providing the correct behaviour is insufficient as providing behaviour relevant for the respective decision is essential.

Many business process discovery approaches create inadequate results when dealing with spaghetti processes, a type of process characterized by a low structure, which leads to highly complex business process models (Günther 2009; Van der Aalst 2016). Because of this, methods have been introduced that are specifically designed to address such processes by undertaking simplifications of behaviour in the model. Such simplifications are undertaken either by filtering the event log before the discovery (Van der Aalst 2016) or by abstracting or aggregating less significant elements during the discovery itself. However, the significance used by these methods is primarily based on the frequency of these elements (Günther and Van der Aalst 2007; Leemans et al. 2014; Weijters and Van der Aalst 2006). While this seems practical in order to guarantee a generic business process discovery with no prior information required, it can lead to challenges during the simplification. Challenges arise since a high frequency of an element is not synonymous with a high relevance of this element. For instance, the low-frequency behaviour of the process might be the most interesting for an auditor (Van der Aalst 2016). If the user is interested in this low-frequency behaviour, few if any simplifications are possible because a frequency-based simplification abstracts such behaviour. However, due to the complexity of spaghetti process models, simplifications are necessary to understand the discovered process model. Hence, for visualizing low-frequency behaviour or behaviour with differing levels of frequency existing approaches provide insufficient results. To the best of our knowledge, this problem can be solved only by filtering irrelevant behaviour based on information such as activity names. This can lead to a loss of potentially relevant information such as relations between behaviour in the process model and abstracted behaviour or to transformed frequency values due to the loss of the broader process context through filtering. Additionally, such filtering is not saved in the event log, meaning that simplified process models created with filtering may be difficult to reproduce. To address this research gap, we raise the following research question (RQ):

RQ: How to conceptualize and design a business process discovery approach focused on spaghetti processes that emphasizes relevant behaviour for the respective task and user by simplification to improve the quality of decisions based on mined process models?

Therefore, we present the following contributions in this paper:

- A design theory to improve the quality of decisions made by a person who must make a decision based on a spaghetti process model. This is expected to be achieved by improving the visualization of the process regarding the perceivability of task-relevant patterns and by minimizing cognitive effort.
- A method called 'categorization approach' that realizes the design theory and the design requirements by instantiating their respective design principles.
- A positive evaluation of the perceived usefulness of our method implemented in a software artefact and demonstrated in a case study.

The paper is structured as follows: In section 2, we discuss the relevant theoretical background of process mining and business process discovery. In section 3, we present the applied design science research method in detail. Section 4 introduces a tentative design theory with design requirements and design principles that answer our RQ. Section 5 describes the created categorization approach and presents a case study to explain the method implemented in a software prototype. Section 6 deals with an evaluation of the categorization approach using an expert survey. In section 7, related work is discussed. Finally, section 8 summarizes our findings and discusses limitations.

2 Theoretical Background

The goal of process mining is to improve the business processes of organizations by expanding traditional approaches of process science by providing data-centric techniques. There are three main types of process mining: discovery, conformance and enhancement of business processes. All three use event logs to analyze the current state of business processes for several analysis techniques (Van der Aalst 2016).

Event logs are automatically generated by many organizations' IT systems and represent the procedures and transactions observed. They contain information about a single process divided into process instances, also called cases, in which events that represent activities are saved. Typically, each event consists of at least an event ID, a timestamp and an activity name and can be extended by further attributes such as resources and costs. The XES standard, which describes the structure of event logs, allows for further extending event data with additional attributes (Verbeek et al. 2010).

In the context of process mining, processes are often classified as either spaghetti processes or lasagna processes. While lasagna processes are well-structured and can, therefore, be used in various ways, mining spaghetti processes poses several challenges. Because of their low structure, which leads to a high level of complexity in mined process models, the business process discovery is primarily applicable to these spaghetti processes (Van der Aalst 2016).

As mentioned, business process discovery is one of the three main types of process mining. Its approaches analyze the data of an event log to mine a process model that captures the described behaviour (Van der Aalst 2016). The kind of representation of business processes depends on the approach used. The graph-based languages most commonly used to represent business processes are BPMN and Petri-Nets (Burattin 2015). In such graph-based representations, a node typically represents an activity, while an arc represents the relation between two activities. Approaches designed to mine spaghetti processes often utilize process graphs that discover only causal dependencies between activities to reduce complexity (Agrawal et al. 1998; Günther and Van der Aalst 2007).

Typically, business process discovery approaches consider behaviour with very low frequency to be noise, which has to be removed from the resulting model because of its variability (Van der Aalst 2016). Nevertheless, low-frequency behaviour can be relevant for the respective user who has to solve a specific task using the model. For tasks related to auditing, for instance, low-frequency behaviour is especially relevant (Van der Aalst 2016). Generally, the relevance of behaviour in the process is not easily predictable because it depends on the respective task to be solved using the process model (Petrusel 2013).

3 Research Method

To provide a structure as well as to ensure scientific rigour in designing the categorization approach, the design science research method, following the work of Vaishnavi and Kuechler (2015), was applied as shown in figure 1. A complete cycle of this method comprises five steps:

- Awareness of problem
- Suggestion
- Development
- Evaluation
- Conclusion

During the cycle, a design theory, consisting of design requirements and design principles, was conceptualized since complete design theories need to have requirements and components that together embody a general design solution for a class of problems (Baskerville and Pries-Heje 2010). The design requirements define the objectives of a design theory and serve as meta-requirements for the software artefact (Baskerville and Pries-Heje 2010; Walls et al. 1992). The defined design principles are prescriptive and state how an artefact should be instantiated to fulfil the design requirements (Fu et al. 2016). The following sections describe our results for the finalized design science research cycle.

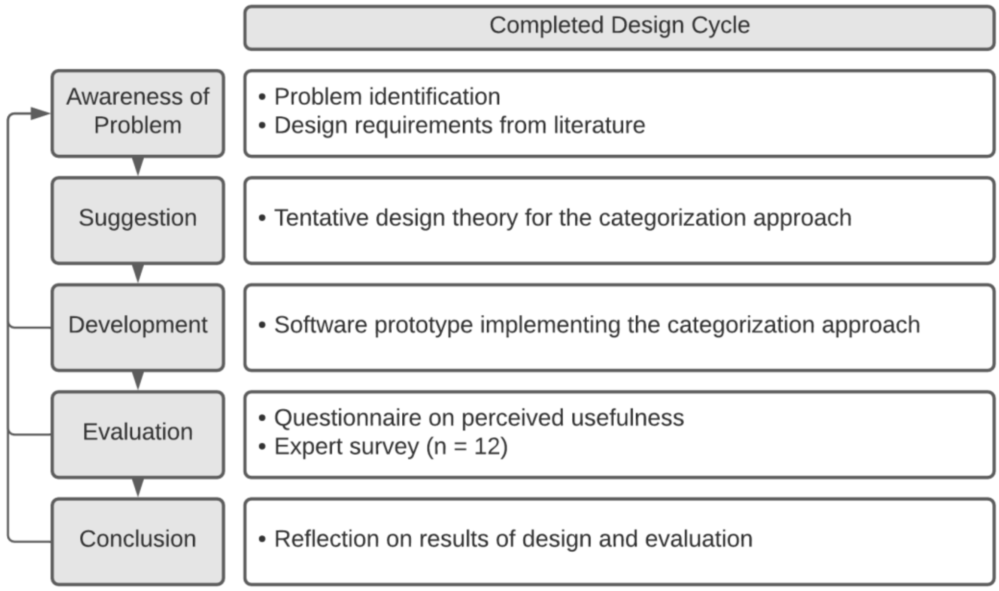

Figure 1: Design Science Research Cycle for the Categorization Approach

4 Suggestion for a Tentative Design Theory for the Categorization Approach

To define the design principles and the overall design theory, the approach developed by Möller et al. (2020) was used as guidance. In this specific context, design requirements and design principles are derived from literature or other appropriate sources of design knowledge. Hence, the defined design principles are used to provide design knowledge before the actual conceptualization and design of the software artefact (Möller et al. 2020).

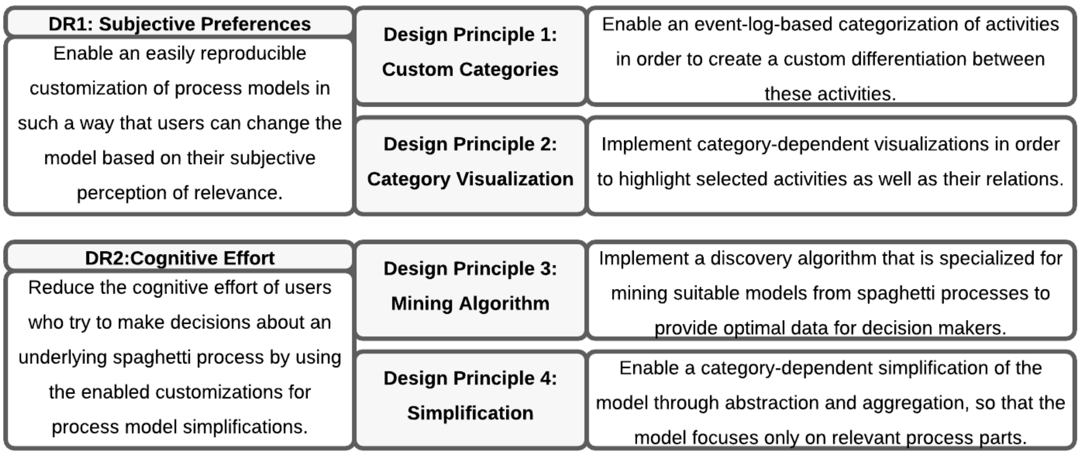

Figure 2: Design Theory Comprising Design Requirements and Design Principles

The design theory, illustrated in figure 2, is defined to achieve the overall goal of the approach, which is to improve the quality of decisions made by a user who wants to make decisions based on a mined spaghetti process model. In a suitable visualization of a business process, the task-relevant patterns have to be perceivable with a minimum of cognitive effort (Ware 2020). Hence, the improvement of decision quality is expected to be attained by including subjective preferences in the discovery to customize process models for the respective task as well as to reduce the cognitive effort required to make decisions by simplifications. To conceptualize these design requirements, we propose four prescriptive design principles following the definition by Fu et al. (2016).

Design requirement 1 including its design principles 1 and 2 describes the fundamental logic of the design theory. The identified problem is to be solved by a categorization of activities that takes place in the event log. While filtering techniques often allow the inclusion or exclusion of specific activities in a simplification, the definition of an XES extension was developed so that the categories are permanently saved in the event log to guarantee the reproducibility of process models (design principle 1). Once activities are categorized or a categorization is obtained from the event log, these categories are subsequently used to visually highlight selected process behaviour as a first step (design principle 2).

The categorization is also the basis and prerequisite for the actual simplification proposed to solve the identified problems. As described in the previous section, the approach is aimed at spaghetti processes since their low structure makes simplifications necessary. Hence, the implemented discovery algorithm must be specialized to deal with such processes (design principle 3). As mining suitable models from spaghetti processes usually requires a simplification, the categories from DP1 are also used for this simplification. They are defined directly by the user, meaning that the conducted simplification focuses on process behaviour that the user perceives to be relevant (design principle 4).

5 Development of the Categorization Approach Implemented in the Software Artefact 'CatMiner'

5.1 The Categorization Approach

The categorization approach instantiates the categorization through an XES extension that allows saving multiple categories in a list element. These categories are saved in string elements with the key attribute 'cat:Category' and a value attribute representing the name of the category. In the implemented software artefact, categories can be assigned to each activity using checkboxes in a table (design principle 1). After assignment or extraction from the event log, the categories are highlighted with different colours in the model since colours are highly effective for nominal information coding (design principle 2) (Ware 2020). It allows for a business process discovery by using an algorithm based on the fuzzy miner from ProM with the difference that the clustering does not depend on frequency-based significance and correlation metrics (design principle 3). Instead, activities belonging to selected categories are shown in the process model, while all other activities are clustered to maintain the relations between process model behaviour and abstracted behaviour (design principle 4).

Following the definition of a process model abstraction from Polyvyanyy et al. (2008), the parameters of the abstraction function used in the categorization approach are a process model and an abstraction setting. The abstraction setting defines a subspace of the following abstraction criteria values:

- K {asc} is an abstraction criterion named category significance. It is a Boolean variable, meaning that $K \in \{0, 1\}$. Only nodes $N \in N_A$ (activities) have a category significance. If the respective node is categorized as a selected category, then $K := 1$, otherwise $K := 0$.
- U {asc} is an abstraction criterion named edge utility. It is a real number between 0 and 1 for E, a set of directed edges between nodes, such that $E \subseteq N \times N$. It is influenced by frequency-based significance and correlation.

5.2 A Case Study Using the Categorization Approach Implemented in the Software Prototype 'CatMiner'

The tentative design theory was used to develop a Java application called 'CatMiner', which implements the method. To demonstrate the categorization approach implemented in the software prototype, an example process and a scenario are used. The example process being used is the process for reimbursement of travel costs for Eindhoven University of Technology (IPCM Conference 2020). The process starts with a travel permit that must either be rejected or approved by university authorities. If it is approved, the next step is to undertake the trip. Following the trip, a claim has to be filed to receive reimbursement. The process comprises 51 activities and a high number of arcs among those activities. Figure 3 shows the process graph mined by the fuzzy miner with no simplifications regarding the activities. The edge utility threshold of the model is 0.2, meaning that all edges with a lower utility are abstracted. While the reimbursement process is not the most extreme example of a spaghetti process, the resulting process model is certainly complex enough to impede an understanding of the underlying process by examining only the model. Hence, to fully understand the process and, therefore, to improve the quality of decisions based on the process model, a simplification seems to be useful.

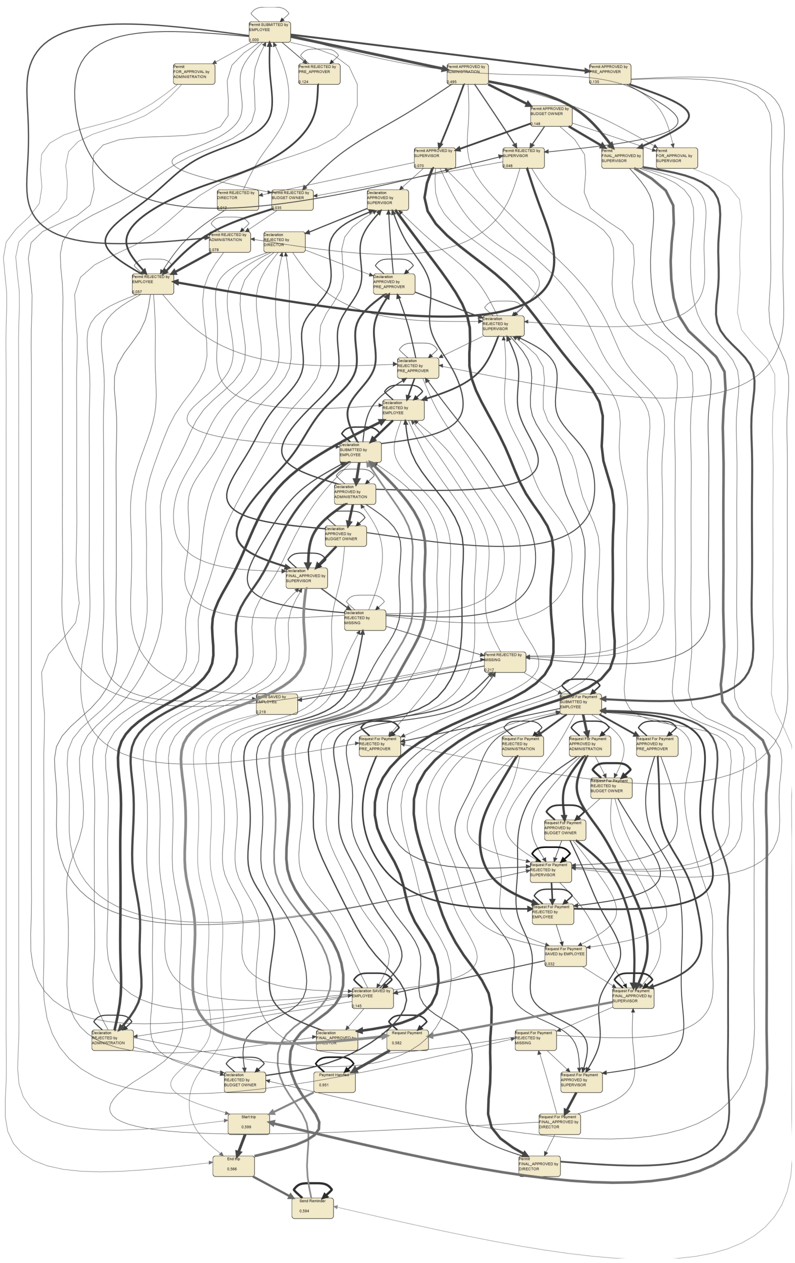

Figure 3: Complete Model Mined by the Fuzzy Miner

Behaviour related to the travel permit takes place at the beginning of the reimbursement process. An employee who plans a trip needs to fully understand the procedure because a rejection of the travel permit can terminate the process. Therefore, adequate visualization of this behaviour is important. Approval for a travel permit is required only for international trips, and rejection is relatively rare. This means that all activities related to the rejection of a travel permit have a low frequency. Hence, most simplification approaches would abstract these activities. However, the low frequency does not necessarily mean that these activities are irrelevant to the user. The relevance depends on the goal the respective user aims to achieve when examining the process model. Relevant questions related to the rejection of travel permits for an employee who wants to understand the process and make decisions based on the process model could be:

- Which authority is most likely to reject the travel permit?
- Who decides to either approve or reject the travel permit?
- Is it likely that the travel permit will be rejected?

For an employee who seeks to make informed decisions about the trip, a model must provide suitable answers. It is difficult to answer the questions above based on figure 3 because of the complexity of the full process model. However, it is impossible to answer the questions if the used simplification approach abstracts activities related to the rejection of the travel permit due to their low frequency. A process analyst attempting to create a suitable process model for employees can use the categorization approach. How to categorize the activities in the event log is a subjective decision, but a categorization of activities related to the rejection of the travel permit to answer the questions seems useful. The categorization was conducted by extracting the activity names from the event log and then categorizing the activities based on their names, as described in the pseudocode in figure 4. The category 'permit rejected' includes all activities that are directly related to the rejection of the travel permit. The category 'permit general' includes all activities related to the travel permit but not to the approval or the rejection of the permit.

```
Begin
        Input ActivityName
        If ActivityName contains "Permit" and !(contains "Approved" or "Approval") Then
                If ActivityName contains "Rejected" Then
                        Category := "Permit Rejected"
                Else
                        Category := "Permit General"
End
```

Figure 4: Conducted Categorization in Pseudocode

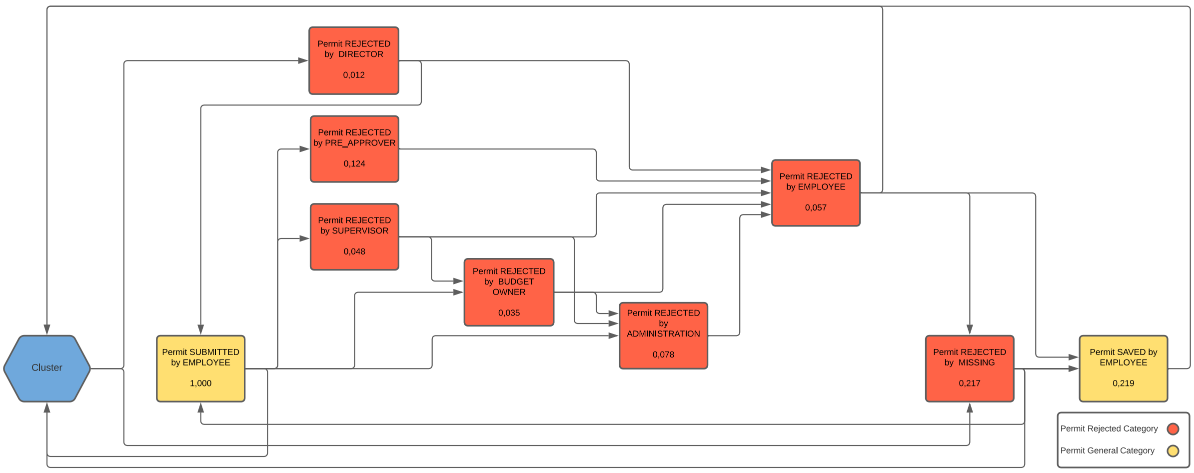

Figure 5: Process Graph Mined by the Categorization Approach Implemented in the Software Artefact 'CatMiner'

Figure 5 shows the process graph mined using the categorization approach implemented by the developed software artefact. It uses the same edge utility threshold as the model in figure 3, meaning that the relations between the activities are the same. The category 'permit rejected' is highlighted in red, and the category 'permit general' is highlighted in yellow to improve distinguishability. The activities that are not classified as one of the two selected categories are clustered to maintain the relations between the process model and aggregated behaviour. While the primarily frequency-dependent significance of the activities calculated by the fuzzy miner is not used for the abstraction and aggregation, it is included on every activity as additional information regarding its frequency.

The process graph in figure 5 focuses on all activities related to the rejection of the travel permit and the travel permit in general. Unrelated activities are aggregated. The relevance of the activities in the model is guaranteed because the simplification depends on the user's input.

6 Evaluation of the Categorization Approach

The categorization approach was evaluated by an expert survey. The experts in process science were shown a process graph mined by the categorization approach implemented in the software artefact ‘CatMiner’ as well as a corresponding full model of the underlying business process. The scenario and the models used for this evaluation are presented in section 5.2. Then, they were asked to rate six statements designed to evaluate the perceived usefulness (Davis 1989) of the categorization approach on verbal-numeric seven-point rating scales. The statements were: 1) The implemented categorization approach enables me to accomplish tasks more quickly; 2) Using the implemented categorization approach improves my job performance; 3) Using the implemented categorization approach increases my productivity; 4) Using the implemented categorization approach enhances my effectiveness on the job; 5) Using the implemented categorization approach makes it easier to do my job; and 6) Overall, I find the implemented categorization approach useful in my job. The sample size for this expert survey was 12, following the ‘10 ± 2 rule’ (Hwang and Salvendy 2010), which states that the rate of problem identification among experts increases only marginally once the sample size exceeds 12.

The responses to the items in the evaluation are shown in the appendix. They range from 4 to 7, with an overall average of 5.9, indicating that the experts perceived the categorization approach to be useful. Statement 6, in particular, received high agreement, with an average score of 6.33. Statement 2 received the lowest agreement, with an average score of 5.58. This seems plausible as the tasks that the abstracted model specializes in answering could, in theory, also be answered with the full model since it includes all the relevant information needed to do so. Overall, no survey results disagreed with the perceived usefulness of the approach. Figure 6 shows the responses to the survey analyzed using box plots.

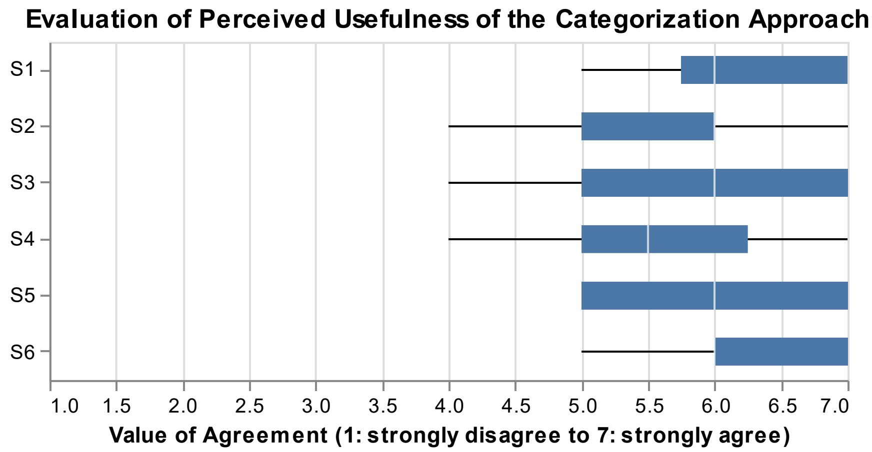

Figure 6: Results of the Evaluation

Another possibility for evaluating the categorization approach could be to measure the degree of simplification of complexity conducted in the case study, for example, by examining the NOA (number of activities) metric (Cardoso et al. 2006). The full model in figure 3 consists of a high number of activities; more precisely, its NOA is 51. The exemplary application of the categorization approach reduced the NOA to 10, counting the cluster as an activity, or by about 80.39 %, while retaining the relevant activities. The conducted aggregation of activities additionally reduces the number of arcs that add complexity to the full process model.

7 Related Work

In this section, we give an overview of related approaches in the relevant research areas. Firstly, we discuss related work in the business process discovery field. In the second part of this section, we discuss related approaches that adapt the business process discovery by incorporating the concept of relevance to simplify discovered process models.

Traditional business process discovery approaches have attempted to implement the behaviour observed in event logs as a single and static process model. This can be sufficient for well-structured processes but leads to highly complex models when dealing with low structure in business processes (Günther 2009; Van der Aalst 2016). The fuzzy miner from ProM, which is an adaptive approach for process simplification, was one of the first business process discovery approaches to use simplifications to reduce the complexity of the resulting process model (Günther and Van der Aalst 2007). It is closely related to our work since it also aggregates activities. The limitation of this approach is that the conducted simplifications are based primarily on the frequency of model elements. The approach will, therefore, always target low-frequency behaviour. However, such low-frequency behaviour can be relevant to the specific user and task as it is often relevant in auditing, for instance (Van der Aalst 2016). Nevertheless, the algorithm used for the categorization approach is similar to the algorithm of the fuzzy miner, as well as the abstraction of edges based on the edge utility criterion. However, our approach uses the category significance criterion based on user-defined categories to aggregate activities instead of depending on frequency metrics. Another related approach in the process mining context is filtering, especially filtering based on activity names. Plug-ins that implement such filtering allow the user to select which activities to include in the business process discovery (Van der Aalst 2016). In fact, using filtering in combination with business process discovery algorithms enables the user to build models similar to those of the categorization approach. One difference in our approach is that the categories used are permanently saved in the event log, making the approach directly readable from the event log and mined models easily reproducible through the saved categorizations. Another difference is that filtering leads to a loss of information such as relations between behaviour in the process model and abstracted behaviour and the loss of the broader process context, which can transform frequency values. Additionally, the categorization approach allows for nominal information coding using colour, which can be especially helpful in large process models. Finally, all business process discovery approaches able to mine simplified models from the event log of an underlying process can be considered related work (Leemans et al. 2014; Weijters and Van der Aalst 2006). The differences lie in the method of the respective approach. While the other approaches use metrics that are primarily based on frequency, the categorization approach uses user-defined categories to guarantee relevancy to the user. Additionally, these other approaches abstract activities in their simplification by elimination, while the categorization approach aggregates activities into clusters to maintain the relations of behaviour in the model to abstracted behaviour.

We also consider approaches that argue for incorporating relevance in the simplification of process models to be relevant work. Bobrik et al. (2007) argue for personalized views of processes since the preferred level of detail in process models depends on the user. They propose an approach that uses a process model as input and simplifies it based on parameters such as the involvement of workers in the underlying activities. However, the approach is not a process mining approach since it does not use real-world data from event logs, but conducts simplifications of existing process models. Kuehnel (2019, 2020) proposed an approach that simplifies processes based on an eXtensible Event Stream extension, which is similar to the more general ‘cat:Category’ extension of the categorization approach. To conduct an economic assessment of business process compliance, this method eliminates behaviour that is not labelled as relevant from a compliance perspective. Therefore, the resulting process model comprises only process elements relevant to compliance. However, this method can be applied in the compliance context only. Additionally, the elimination of behaviour leads to a higher loss of information compared to an approach that uses aggregation. Similar to our reasoning, Stierle et al. (2021) argue that to improve business processes as a whole, the discovered process models should not be dependent on frequency-based metrics only. Hence, they propose an approach that determines the relevance of individual activities by analysing process performance indicators. Using these indicators, the process model can be simplified regarding relevance instead of frequency. However, while the inclusion of process performance indicators seems to be a promising concept, we consider the relevance of an activity to be task-dependent. Therefore, there will be cases where a user deems behaviour with high process performance indicators to be irrelevant for a specific task and vice versa.

8 Conclusion

Approaches designed to deal with low-structured processes simplify the resulting process models. These simplifications are applied primarily with metrics based on the frequency of observed behaviour. Since a high frequency of a certain behaviour is not synonymous with a high relevance to the user, the simplification can abstract relevant behaviour for the specific user and task. Our goal in this paper is to design and implement a process simplification approach in the business process discovery field that emphasizes relevant behaviour for the respective task and user in models of spaghetti processes to improve the quality of decisions based on the respective process model.

To answer the research question, we followed the design science research approach proposed by Vaishnavi and Kuechler (2015). This includes the definition of a tentative design theory consisting of two design requirements and four design principles. The design theory was then used to create the categorization approach. Finally, the perceived usefulness of a model resulting from the software artefact 'CatMiner' implementing the categorization approach was evaluated with a predominantly positive outcome. Our design theory can be used and adapted by practitioners and scientists to develop new software implementations and similar approaches. In addition, the design requirements, design principles, and the overall design theory contribute to the prescriptive knowledge base of the IS community (Gregor and Hevner 2013).

There are limitations regarding our results that must be considered for an adequate interpretation. Firstly, the conceptualization of design theories has the inherent weakness of subjectivity regarding the design decisions. The resulting design requirements and design principles, as well as implementation details, could be different if other designers were to solve the same problem with the same means. However, we underpinned our design theory methodologically by considering the methods of Möller et al. (2020) for supportive design approaches as well as Fu et al. (2016) for the articulation of prescriptive design principles. Secondly, although the evaluation provided insight regarding the perceived usefulness of the categorization approach, a quantitative evaluation of cognitive effort and decision quality is subject to future research. Additionally, our evaluation results depend on the sample, meaning that other participants or a different sample size could lead to a different outcome. Nevertheless, by applying a common evaluation rule in addition to selecting experts, we believe our insights to be sound. Finally, a weakness of the proposed categorization approach is that it requires a categorization of activities in the event log before the business process discovery. At the moment, this can be achieved only by manually labelling activities in the developed software prototype. However, it is possible to automate such categorizations through algorithms for common problem cases. The design and implementation of such algorithms is subject to future research. Additionally, IT systems producing event logs could be adapted to categorize activities at the time of creation as an addition to information such as timestamp, activity name, or resource. Another weakness of the current categorization approach is that it requires activity names in the event log to be understandable to the user for the categorization. This might not be the case in some IT systems and could be remedied only by an adaptation of the respective system.

9 References

Agrawal, R., Gunopulos, D., and Leymann, F. 1998. "Mining Process Models from Workflow Logs," Lecture Notes in Computer Science (1377), pp. 469-483.

Baskerville, R., and Pries-Heje, J. 2010. "Explanatory Design Theory," BISE (2:5), pp. 271-282.

Bobrik, R., Reichert, M., and Bauer, T. 2007. "View-Based Process Visualization," Lecture Notes in Computer Science (4714), pp. 88-95.

Burattin, A. 2015. *Process Mining Techniques in Business Environments – Theoretical Aspects, Algorithms, Techniques and Open Challenges in Process Mining*. Basel: Springer.

Cardoso, J., Mendling, J., Neumann, G., and Reijers, H.A. 2006. "A Discourse on Complexity of Process Models," Lecture Notes in Computer Science (4103), pp. 117-128.

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," MIS Quarterly (13:3), pp. 319-340.

Fu, K. K., Yang, M. C., and Wood, K. L. 2016. "Design Principles: Literature Review, Analysis, and Future Directions," Journal of Mechanical Design (138:10), e1.

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," MIS Quarterly (37:2), pp. 337-355.

Günther, C. 2009. "Process mining in flexible environments," dissertation, Eindhoven University of Technology.

Günther, C., and Van der Aalst, W. 2007. "Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics," Lecture Notes in Computer Science (4714), pp. 328-343.

Hwang, W., and Salvendy, G. 2010. "Number of people required for usability evaluation: the 10±2 rule," Communications of the Association for Computing Machinery (53:5), pp. 130-133.

IPCM Conference. 2020. "BPI Challenge." (https://icpmconference.org/2020/bpi-challenge, accessed August 11, 2021)

Kuehnel, S. 2019. "Wirtschaftliche Bewertung und Analyse von Business Process Compliance: Ein Ansatz basierend auf Basic Control Flow Patterns und Extensible Event Streams," dissertation, Martin Luther University Halle-Wittenberg.

Kuehnel, S. 2020. “Economic Assessment and Analysis of Business Process Compliance: An Approach based on Basic Control Flow Patterns and Extensible Event Streams,” Proceedings of the Best Dissertation Award at BPM (18), pp. 41-45.

Leemans, S., Fahland, D., and Van der Aalst, W.M.P. 2014. „Exploring Processes and Deviations,” Lecture Notes in Business Information Processing (202), pp. 46-50.

Mans, R., Van der Aalst, W.M.P., and Vanwersch, R. 2015. *Process Mining in Healthcare – Evaluating and Exploiting Operational Healthcare Processes*. Berlin, Heidelberg: Springer.

Möller, F., Guggenberger, T., and Otto, B. 2020. “Towards a Method for Design Principle Development in Information Systems,” Design Science Research in Information Systems and Technology, pp. 208-220.

Petrusel, R. and Mendling, J. 2013. “Eye-Tracking the Factors of Process Model Comprehension Tasks,” Advanced Information Systems Engineering (7908), pp. 224-239.

Polyvyanyy, A., Smirnov, S., and Weske, M. 2008. “Process Model Abstraction: A Slider Approach,” International IEEE Enterprise Distributed Object Computing Conference.

Stierle, M., Weinzierl, S., Harl, M., and Matzner, M. 2021. “A technique for determining relevance scores of process activities using graph-based neural networks,” Decision Support Systems (144).

Vaishnavi, V., and Kuechler, W. 2015. *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, Boca Raton: CRC Press LLC.

Van der Aalst, W.M.P. 2016. *Process Mining – Data Science in Action*. Berlin, Heidelberg: Springer.

Verbeek, H., Buijs, J., Van Dongen, B., and Van der Aalst, W. 2010. “XES tools,“ Proceedings of the CAiSE Forum.

Walls, J. G., Widmeyer, G. R., and El Sawy, O. A. 1992. “Building an Information System Design Theory for Vigilant EIS,” Information Systems Research (3:1), pp. 36-59.

Ware, C. 2020. *Information Visualization – Perception for Design*. Burlington, MA: Morgan Kaufmann.

Weijters, A.J.M.M., van der Aalst, W.M.P., and De Medeiros, A.K. 2006. “Process Mining with the Heuristics Miner Algorithm,” BETA Publicatie: Working Papers (166) ; Technische Universiteit Eindhoven.

Weske, M. 2012. *Business Process Management – Concepts, Languages, Architectures*. Berlin, Heidelberg: Springer.

Appendix

Questions	A	B	C	D	E	F	G	H	I	J	K	L
1) The implemented categorization approach enables me to accomplish tasks more quickly.	6	5	5	6	5	6	7	7	7	6	7	6
2) Using the implemented categorization approach improves my job performance.	5	5	5	4	6	6	7	6	5	6	6	6
3) Using the implemented categorization approach increases my productivity.	6	5	5	5	5	7	7	4	6	7	7	6
4) Using the implemented categorization approach enhances my effectiveness on the job.	5	5	5	4	6	6	7	6	7	5	7	5
5) Using the implemented categorization approach makes it easier to do my job.	6	5	5	6	6	5	7	7	7	5	7	5
6) Overall, I find the implemented categorization approach useful in my job.	6	7	5	6	6	6	7	7	6	6	7	7

Table 1. Results of the Expert Survey