Towards an Evaluation Framework for Process Mining Systems

Irina Ailenei[1], Anne Rozinat[2], Albert Eckert[3], and Wil van der Aalst[1]

[1] Eindhoven University of Technology, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands. i.m.ailenei@student.tue.nl, w.m.p.v.d.aalst@tue.nl
[2] Fluxicon Process Laboratories. anne@fluxicon.com
[3] Siemens AG, Corporate Technology, Munich, Germany. albert.eckert@siemens.com

Abstract. Process mining is an emerging topic in the BPM marketplace. Recently, several (commercial) software solutions have become available. Due to the lack of an evaluation framework, it is very difficult for potential users to assess the strengths and weaknesses of these process mining tools. As the first step towards such an evaluation framework, we developed a set of process mining use cases and validated these use cases by means of expert interviews and a survey. We present the list of use cases and discuss the insights from our empirical validation.

Keywords: Business Process Intelligence, Process mining, Use cases, Evaluation framework.

1 Introduction

The area of Process Mining has attracted the attention of both researchers and practitioners. As a consequence, a significant number of algorithms and tools were developed. For instance, the academic process mining tool ProM Version 5.2 contains more than 280 pluggable algorithms, developed to provide a wide range of functionalities and techniques. Additionally, commercial process mining tools have emerged on the market and often use their own standards and naming. For a potential user, this situation is quite confusing and it is difficult to choose the most suitable process mining tool or algorithm for the task at hand.

Our goal is to develop an evaluation framework that can be used to assess the strengths and weaknesses of different process mining tools. We will then apply this evaluation framework to compare commercial process mining tools that are currently available on the market. Therefore, the main questions of this project are:

1. What are typical process mining *use cases*?
2. Which *process mining tools* are suitable for which use case?

As *process mining tool* we consider any software that is able to extract process models from raw event logs (without having to manually create a model beforehand). As process mining *use cases* we consider typical applications of process mining functionality in a practical situation.

Consider Figure 1, which illustrates that the use of any process mining tool will be carried out in a certain context. We can assume that the context of the person using

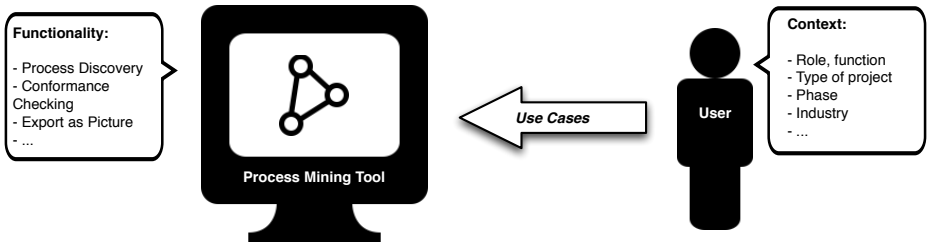

Fig. 1. Use cases for process mining may vary depending on the context

process mining has an influence on which type of functionality is considered most important. For example, the role or function a person fulfills in their organization might impact the type of analysis that the user is interested in (e.g., an auditor would be more interested in checking the compliance of processes whereas a process analyst will be mostly focused on process improvement). Another example is the type of project: In a process improvement project a user is likely to be more focused on diagnosing process bottlenecks and inefficiencies whereas in an IT re-implementation project the main goal might be to extract the current processes in an executable process modeling language such as BPMN. Even within one project, process mining could be used in different phases (e.g., as a quick-scan in the beginning of an improvement project or as a means to validate the actual improvements at the end of the project).

In this paper, we report on the development of an evaluation framework by defining and categorizing use cases for process mining. To ensure that the list of use cases is as complete and as relevant as possible, we validate these use cases by expert interviews with practitioners and a survey. During the validation, we also capture information about the context of the user to find out how their role affects the importance they give to the different use cases. These use cases will then form the basis for a detailed evaluation of current process mining tools in the market.

The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 describes the approach that we followed to define and validate the process mining use cases. Section 4 introduces our list of process mining use cases in detail. In Section 5, we then describe how we validated these use cases through expert interviews and a survey. Finally, in Section 6 we give an outlook on how we are currently detailing and applying our evaluation framework for the assessment of different commercial process mining tools.

2 Related Work

As process mining is an emerging topic, little work has been done on the systematic identification of use cases. Lion's share of process mining literature focuses on process discovery. Several authors describe how to evaluate discovered process models [9, 3–5, 7, 6]. For example, in [7] an evaluation framework is defined. The framework provides an extended set of tests to judge the quality of process mining results. One of the problems is a lack of commonly agreed upon benchmark logs. This year's Business

Processing Intelligence Challenge (BPIC) aims to address this problem by providing a reference log.

Unlike the approaches aiming to judge the quality of the discovered process model [3–5, 7, 6], we focus on the different functionalities related to process mining. Clearly, this extends beyond pure control-flow discovery.

Our approach to define and validate use cases is related to [8] (e.g., conducting interviews with BPM experts). However, in [8] the focus is on business process model abstraction rather than process mining. Also related are the evaluations done in the context of the workflow patterns [2].

3 Approach

One of the challenges of our study was to decide which approach we are going to follow in defining and validating the list of use cases to be used for the tools evaluation. Since there was no standard reference for process mining use cases, we followed an inductive approach, similar to the one described in [8], which aimed at defining a list of process mining functionalities needed in practice that is as complete and relevant as possible. Figure 2 illustrates the sequence of steps that constitute the approach we followed.

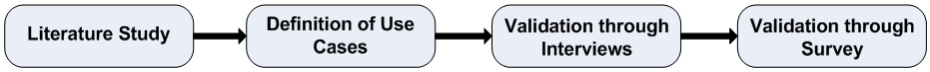

Fig. 2. The four phases of the approach

Literature study The purpose of the literature study was to get an overview about the existing functionality available in the context of process mining. In order to do this, we looked at the functionality provided by the process mining tool ProM [1] and focused our attention on academic articles about process mining techniques as well as on marketing brochures and descriptions of a couple of commercial process mining tools present on the web.

Definition of Use Cases The next step was the definition of an initial list of process mining use cases. We consider a use case to represent the use of a concrete process mining functionality with the goal to obtain an independent and final result. Therefore, actions performed before the actual analysis, like the import of the event log or filtering, are not included in our list. When defining the list of use cases, we used the classification of process mining techniques described in [9]. Figure 3 is a simpler representation of this classification and also shows our scope in relation with the entire classification. The definition of use cases is thus restricted to the offline analysis and does not include any techniques that deal with prediction, detection or recommendation. This limitation was introduced due to the inability of evaluating

the systems participating in the study in an online analysis environment. The description and examples of each use case are introduced in Section 4.

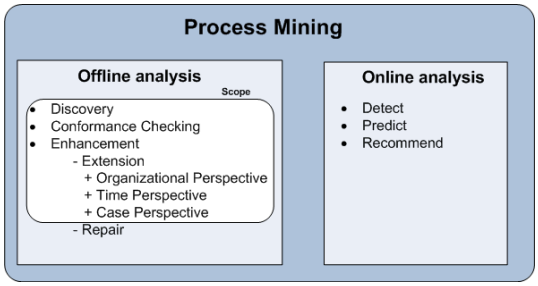

Fig. 3. The project's scope in the context of process mining

Validation through Interviews The great number of existing process mining techniques and the lack of a standard list of use cases led to the need of validating the defined list. We started our validation phase by conducting a series of ten semi-structured interviews with practitioners having process mining expertise. First, we wanted to verify the understandability of the descriptions of the use cases by asking them to provide examples with situations in which each use case would be useful. Second, the goal of the interviews was to validate the list of use cases by removing the use cases that the participants considered irrelevant, and by determining whether there are use cases missing from the initial set. Furthermore, we wanted to find out whether there are differences between the importance of each use case for different categories of end users. One lesson learnt from the interviews was that participants have the tendency of saying that all use cases are equally important. As a result of this observation, we deviated from the approach described in [8], where use cases were just classified as important or not important, and instead used the sorting method for *ranking* the use cases based on their importance. The findings of the interviews are presented in detail in Section 5.1.

Validation through Survey Distributing a survey among people familiar with the field of process mining was the most suitable method to collect a larger number of responses for the validation phase. In total, we obtained 47 responses. The main goals of the survey were to capture the context of the respondents by asking for their role and domain, get the use cases rankings, and find out what additional functionality not covered by the list of use cases is considered important and should be included in our tool evaluation. The results of the survey are discussed in Section 5.2.

The list of validated use cases will serve as a basis for a detailed evaluation of a couple of commercial process mining systems. For this purpose, an evaluation framework has been developed. This framework incorporates next to the description and the example for each use case, also related assumptions and a set of acceptance criteria used to decide whether the use case is supported or not by a tool.

4 Use Cases

This section introduces the list of process mining use cases by providing a short description of each use case. A more complete presentation, containing in addition a practical example for every use case, is given in Appendix 1. The use cases are grouped into the categories described in [9]. Section 4.1 contains use cases belonging to the process discovery part, subsection 4.2 focuses on the conformance checking use cases, while sections 4.3, 4.4, 4.5 present the use cases related to the organizational, the time, and the case perspective.

4.1 Discovery

The use cases belonging to this category are focused on the control flow perspective of the process. The user gets a clear understanding of the analyzed process by looking at its structure, frequent behavior and at the percentages of cases following every discovered path.

Use case 1: Structure of the process. Determine the structure of an unknown process or discover how a process looks like in practice.

Use case 2: Routing probabilities. Get a deeper understanding of the process by looking at the probabilities of following one path or another after a choice point.

Use case 3: Most frequent path in the process. Discover what is the path in the process that is followed by the highest percentage of cases.

Use case 4: Distribution of cases over paths. Discover common and uncommon behavior in the process by looking at the distribution of cases over the possible paths in the process.

4.2 Conformance Checking

This category consists of use cases which have the purpose of checking whether the process has the intended behavior in practice. The use cases pertaining to this category have in common that in order to execute them one needs an additional input besides the event log of the process to be analyzed. This input may be a reference model of the process or a rule which the discovered process has to be checked against.

Use case 5: Exceptions from the normal path. Discover the outliers of the process by looking at the exceptional behavior observed in practice.

Use case 6: The degree in which the rules are obeyed. Check whether the rules and regulations related to the process are obeyed.

Use case 7: Compliance to the explicit model. Compare the documented process model with the real process as observed in the event log.

4.3 Enhancement - Extension - Organizational Perspective

The focus of the use cases included in this category is on the organizational analysis. The outcome of executing these use cases provides the user with an insight in the issues related to the resource perspective of the process.

Use case 8: Resources per task. Discover the relation between resources and tasks.

Use case 9: Resources involved in a case. Discover the group of resources involved in

solving a particular case.
Use case 10: Work handovers. Manage resource location or determine possible causes for quality and time issues by looking at how work is transferred between resources.
Use case 11: Central employees. Determine who the central resources for a process are by analyzing the social network based on handovers of work.

4.4 Enhancement - Extension - Time Perspective

As performance-related insights are most valuable, most of the use cases related to enhancement correspond to the time perspective.
Use case 12: Throughput time of cases. Determine the time that passed since the start of a case in process until its completion.
Use case 13: Slowest activities. Discover potential time problems by looking at the slowest activities in the process.
Use case 14: Longest waiting times. Determine delays between activities by analyzing the waiting times before each activity.
Use case 15: Cycles. Learn whether additional delays occur in the process due to cycles.

Use case 16: Arrival rate of cases. Determine the frequency with which new cases arrive in the process.
Use case 17: Resource utilization rate. Determine what are the utilization rates of the resource i.e, measure the fraction of time that a resource is busy.
Use case 18: Time sequence of events. Get a deeper understanding on the organization of a process by looking at the time sequence of activities for a specific case. (e.g. Gant-graph for activities).

4.5 Enhancement - Extension - Case Perspective

The case perspective of the process is represented by a single use case.
Use case 19: Business rules. Discover what are the process attributes that influence the choice points and what are the conditions for following one branch or another.

5 Validation of the use cases

The use cases were validated by conducting ten interviews (Section 5.1) and by distributing a survey (Section 5.2) among process mining users and experts.

5.1 Interviews

We conducted in total ten interviews with process mining users and domain experts. The interviews can be divided into two categories:(1) interviews aimed at gaining some qualitative feedback on the understandability of the use cases and (2) interviews which were focused on obtaining a ranking of the use cases based on their importance for the interviewees and on identifying missing use cases.

(1) Based on the feedback received from the first type of interviews (in total: four) two non-relevant use cases were removed from the list, the descriptions of a couple of

use case were refined and a short motivation was added for each remaining use case. The two irrelevant use cases referred to the possibility of identifying the paths in the process taking most time and to the possibility of visualizing the list of process attributes stored in the event log. The aim of refining the use case descriptions and of adding the motivation dimension was to increase the understandability and clarity of what each use case is about and what its practical purpose is.

(2) In the second type of interviews (in total: six) we asked the interviewees to sort the list of cases in the order of their importance in practice and on discovering any missing use cases. Moreover, we were interested in gaining additional insights on what are the functionalities that a process mining tool should provide to its users. These interviews were structured in three parts. The first part aimed at getting information about the experience of the interviewee in the context of process mining and about the added value that process mining brings to their work. Secondly, the interviewees were shown the list of use cases and were asked to assign to each use case a score from 1 to 19 based on its importance (1 being the most important). The last part of the interview was meant to summarize the discussion, to learn about possible use cases missing from the initial list and about additional functionality that interviewees consider useful in a process mining tool. The complete summary of the outcomes of these six interviews can be found in Appendix 2.

The six interviews we conducted were balanced from the point of view of the interviewee's role in the context of using process mining techniques. Three of the persons interviewed were process analysts and the other three were auditors. The second dimension we took into account when selecting the interviewees was the domain they belong to. In this context we aimed at having a broader range of domains and therefore we talked with people working in the banking industry, healthcare, public sector, and business process consulting.

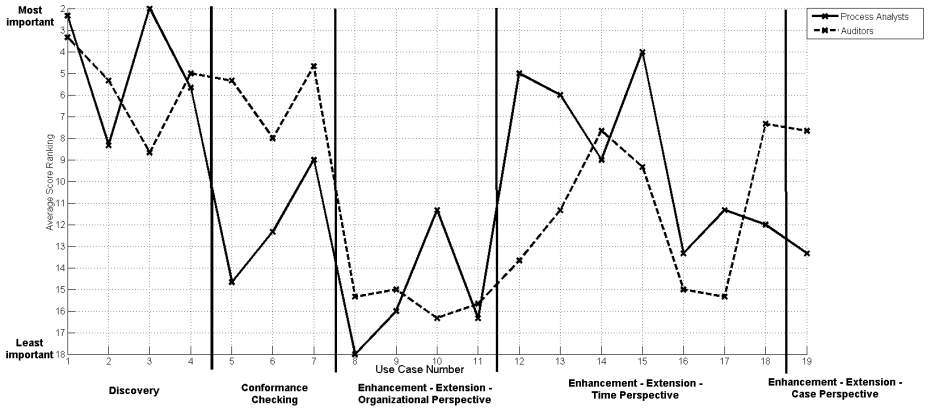

Fig. 4. Use cases ranking results from the interviews with process analysts and auditors

Figure 4 depicts the profiles of process analysts and auditors based on the use case rankings collected from our interviews. On the x-axis we refer to use case numbers, while the y-axis represents the averages of the scores the use cases were assigned during the interviews. The graphic shows there are some differences in ranking the use cases based on the profile of the respondents. For instance, use case 12 (Throughput time of cases) is one of the most important use cases according to the process analysts group, while the auditors consider this quite irrelevant in practice. The opposite holds for use case 5 (Exceptions from the normal path), which is ranked as highly important by the auditors and less important by the process analysts.

Furthermore, the top five and bottom five use cases were extracted for each category of respondents (cf. Table 1 and Table 2). Our expectations regarding the difference in needs of people having different roles are confirmed by comparing the top five use cases for each category. The contents of the top rankings are quite different, except for two use cases that are considered important by all: discovering the structure of a process and looking at the distribution of cases over the paths in the process.

When comparing the rankings of the least interesting use cases, one can also identify some similarities. Four use cases are common for both rankings. Respondents, independent of their role, consider that determining the group of resources performing a task and the group of resources involved in a case, as well as looking at the central employees of a process and at the arrival rate of cases in the process are less relevant use cases.

Table 1. Top 5 and Bottom 5 Use Cases for Process Analysts

Top 5 Use cases	**Bottom 5 Use Cases**
U3. Most frequent path in the process	U8. Resources per task
U1. Structure of the process	U11. Central employees
U15. Cycles	U9. Resources involved in a case
U12. Throughput time of cases	U5. Exceptions from the normal path
U4. Distribution of cases over paths	U16. Arrival rate of cases U19. Business rules

Table 2. Top 5 and Bottom 5 Use Cases for Auditors

Top 5 Use cases	**Bottom 5 Use Cases**
U1. Structure of the process	U10. Work handovers
U7. Compliance to the explicit model	U11. Central employees
U4. Distribution of cases over paths	U8. Resources per task
U2. Routing probabilities	U17. Resource utilization rate
U5. Exceptions from the normal path	U9. Resources involved in a case U16. Arrival rate of cases

5.2 Survey

As a next step, we designed and distributed a survey to collect a larger number of responses. The survey contained all the questions addressed during the interviews, but

also additional ones, which serve the purpose of capturing more detailed information about the end user's need in terms of process mining functionality. The complete contents of the survey is given in Appendix 3, while Appendix 4 presents the results for the entire survey.

This section presents the results obtained for a selection of the questions asked. We focus on the role and activity domain of the respondents, the ranking of the use cases, the identification of missing use cases and the possible functionality important for a process mining tool but not covered in the list of use cases.

From this survey, we received 47 responses. Although this number of responses is not enough to obtain statistically significant results, the survey results can provide useful qualitative feedback to validate our use cases. The highest percentages of responses we received are from people working in domains like academia (43%, 20 responses), information technology(21%, 10 responses), business process management consulting (19%, 9 responses), and banking (6%, 3 responses). The distribution over the roles shows a high percentage of researchers (51%, 24 responses), followed by process analysts (28%, 13 responses), process managers (9%, 4 responses), and consultants (6%, 3 responses).

The scores obtained by each use case based on the rankings were computed both over all responses and based on the role of the respondent. The score of a use case is the average of all scores registered from all rankings of the respondents belonging to the same role (the lower the score the more important is the use case). Based on these scores, we generated the graph depicted in Figure 5, which presents the profiles of the four most representative roles among the respondents.

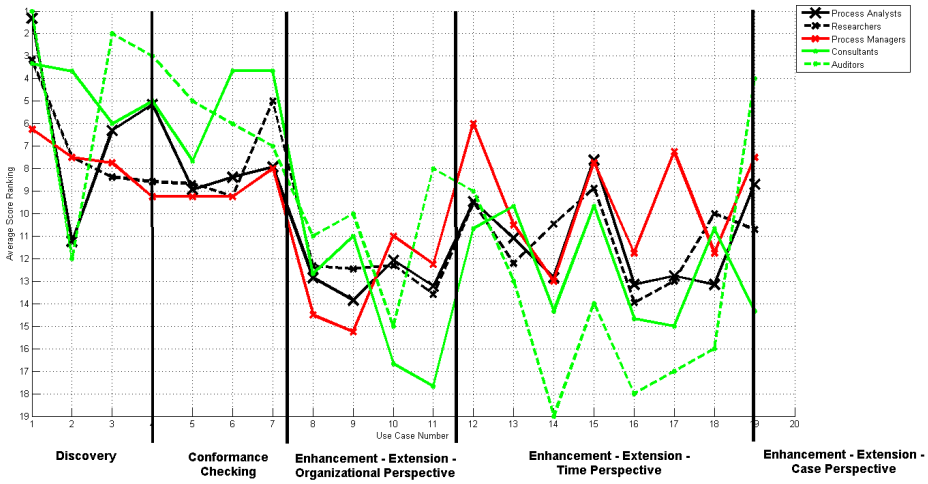

Fig. 5. Use cases ranking results based on respondents roles

Again, the results confirmed our expectation that the way users rank the use cases differs based on the role they have. It is interesting to see that use case 6 (The degree in

which rules are obeyed) is considered medium important by researchers, process analysts and process managers while consultants view it as an essential use case. The same observation holds for use case 17 (Resource utilization rates); process managers view it as a highly relevant use case, while the respondents belonging to the other categories have a different opinion.

However, similarities in the ranking are also quite frequent. For instance, use case 1 (Structure of the process) is graded as one of the most important use cases by all the roles. Similarly, use cases 3 (Most frequent path in the process) and 7 (Compliance to the explicit model) are present in the tops of all rankings. The lower parts of the four rankings also share common use cases. Examples are use case 11 (Central employees) and use case 16 (Arrival rate of cases).

The rankings obtained for the use cases were also grouped based on the domains of activity of the respondents. The results show few differences between the three domains considered (academia, information technology and business process management consulting). The profiles of the domains are depicted in Figure 9 in Appendix 4.

Table 3 presents the results of rankings of the use cases based on the survey responses. We make the distinction between use cases relevant for all the roles, use cases less relevant for all the roles and use cases relevant only for some specific roles. This distinction was made by considering relevant the top nine use cases from the aggregated rankings of each role and less relevant the remaining ten use cases.

Four use cases (U1, U3, U4, and U7) are considered important by all the groups of respondents, while six use cases (U8, U9, U10, U14, U16, and U18) are rated as less important by all the groups. It is interesting to note that there are two use cases (U13 and U17) that are relevant for only one of the categories of respondents. The opposite holds for use cases U5, U6, U12, and U15, which resulted to be important for three out of the four categories of respondents.

Table 3. Aggregated results survey

Uses case relevant for all roles
U1. Structure of the Process
U3. Most frequent path in the process
U4. Distribution of cases over paths
U7. Compliance to the explicit model

Use cases less relevant for all roles
U8. Resources per task
U9. Resources involved in a case
U10. Work handovers
U14. Longest waiting times
U16. Arrival rate of cases
U18. Time sequence of events

Use case	**Relevant for**
U2. Routing probabilities	researchers, pr managers, consultants
U5. Exceptions from the normal path	researchers, pr analysts, consultants
U6. The degree in which the rules are obeyed	researchers, pr analysts, consultants
U12. Throughput time of cases	researchers, pr analysts, pr managers
U13. Slowest activities	consultants
U15. Cycles	researchers, pr analysts, pr managers, consultants
U17. Resource utilization rate	pr managers
U19. Business rules	pr analysts, pr managers

For the question asking whether there are any missing use cases, 58% of the respondents answered no, while 42% suggest new use cases. Among these suggestions, the measurement of different KPIs (cost, quality, flexibility, etc), the creation of a simulation model, and the online analysis of an event log with the purpose of making predictions were mentioned. Since our scope is limited to the process mining techniques that perform an offline analysis of processes and the last two indications we received are related to the online type of analysis, they are not considered for new use cases. The suggestion related to the KPIs measurement does however fit in our scope, but at the moment is too vague and general to be transformed in a testable use case.

The answers regarding the additional functionalities that a process mining system should offer to its users can be grouped into the following categories: input and output capabilities, the ability to filter and cluster data, the integration with external systems like databases, BPM tools, ERP, CRM, etc, animation capabilities, and the support for large input event logs. This information will be used as basis for the extended evaluation of the process mining tools in the following phases of the project.

5.3 Conclusions of the validation phase

The use cases ranking results derived from the survey are in line with the ones resulted from the interviews, in the sense that respondents having different roles have different needs in terms of process mining functionality. This is reflected in the scores assigned to the use cases. Another similarity between the results of the two validation steps is the fact that use case 1 (Structure of the process) was considered overall the most important one, while use cases 11 (Central employees) and 16 (Arrival rate of cases) are the least significant ones.

Based on the feedback received during the validation phase of our approach, we removed two irrelevant use cases, we rephrased all the use cases descriptions that were unclear, and we obtained a classification of use cases based on their importance for different roles.

The outcome of the interviews and survey was the validated list of process mining use cases. By validated, we mean use cases properly formulated, understandable, and corresponding to the needs of process mining users. Additional developments of the use cases needed for the practical tool evaluation are described in section 6.

6 Future Work

In this paper we presented the method we used to define and validate a list of process mining use cases. We employed an exploratory approach to collect a comprehensive set of process mining functionalities needed in practice. We started by looking at the literature in the domain of process mining and the functionality available in ProM. The next step was the definition of a set of use cases grouped according to the classification of process mining techniques given in [9]. We then validated the use cases by means of ten semi-structured interviews with domain experts and process mining users and by a survey.

The outcome of this study, namely the validated list of process mining use cases, is a part of a broader project that aims at evaluating a set of commercial process mining systems. The evaluation is done by judging whether a system provides support for each of the use cases in the list.

To do this, the use cases are currently further refined by assumptions and detailed acceptance criteria to allow for an unambiguous and repeatable evaluation. For example, use case 1 (Structure of the process) will be tested based on detailed acceptance criteria that determine which kinds of behavioral patterns [2] can be discovered by the tool. Additional to the complete use cases framework, we developed a set of benchmark event logs as part of our experimental setup for the evaluation.

So far we used our framework to evaluate two process mining tools: Futura Reflect by Futura Process Inteligence and ProcessAnalyzer by QPR. Based on the use cases we created a comprehensive set of event logs to test the functionality. Our initial findings show that the approach indeed reveals relevant strengths and weaknesses of the different tools. Currently, we are working on the evaluation of two other systems: ARIS Performance Process Manager (PPM) by Software AG and Flow by Fourspark.

Acknowledgements. We want to thank the practitioners who were willing to support us in an expert interview and everyone who took the time to fill out our survey.

References

1. Process Mining. http://www.processmining.org/.
2. W.M.P. van der Aalst, A.H.M. ter Hofstede, B. Kiepuszewski, and A.P. Barros. Workflow Patterns. *Distributed and Parallel Databases*, 14(1):5–51, 2003.
3. A. K. Alves de Medeiros and C. W. Günther. Process mining: Using cpn tools to create test logs for mining algorithms. *Proceedings of the Sixth Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools*, 2005.
4. S. Goedertier, D. Martens, J. Vanthienen, and B. Baesens. Robust Process Discovery with Artificial Negative Events. *Journal of Machine Learning Research*, 10:1305–1340, 2009.
5. J. Munoz-Gama and J. Carmona. A Fresh Look at Precision in Process Conformance. In R. Hull, J. Mendling, and S. Tai, editors, *Business Process Management (BPM 2010)*, volume 6336, pages 211–226, 2010.
6. A. Rozinat and W.M.P. van der Aalst. Conformance Checking of Processes Based on Monitoring Real Behavior. *Information Systems*, 33(1):64–95, 2008.
7. A. Rozinat, A.K. Alves de Medeiros, C.W. Günther, A.J.M.M. Weijters, and W.M.P. van der Aalst. The Need for a Process Mining Evaluation Framework in Research and Practice. In M. Castellanos, J. Mendling, and B. Weber, editors, *Informal Proceedings of the International Workshop on Business Process Intelligence (BPI 2007)*, pages 73–78. QUT, Brisbane, Australia, 2007.
8. S. Smirnov, H.A. Reijers, T. Nugteren, and M. Weske. Business process model abstraction: Theory and practice. Technical report, Hasso Plattner Institute Postdam, 2010.
9. W. M. P. van der Aalst. *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

Appendix 1 - Process Mining Use Cases Framework

Use case 1: Structure of the process

Description: Determine the structure of an unknown process or discover how a process looks like in practice.
Example: This use case could be used in order to mine the structure of an unknown process. By looking at a discovered process model it is possible to get an overview on how things are really happening in practice. Moreover, the discovered model can be used for communication and redesign purposes.

Use case 2: Routing probabilities

Description: Get a deeper understanding of the process by looking at the probabilities of following one path or another after a choice point.
Example: From the user's point of view it might be useful to get a better grasp on the structure and behavior of the process by looking at the routing probabilities in choice points. These probabilities might be used as KPIs when monitoring the process.

Use case 3: Most frequent path in the process

Description: Discover what is the path in the process that is followed by the highest percentage of cases.
Example: The user is able to get a clearer understanding of their process by knowing the path of events followed by the majority of cases. This information can be used for redesign purposes, since optimizing the most frequent path in the process might bring benefits to the overall performance.

Use case 4: Distribution of cases over paths

Description: Discover common and uncommon behavior in the process by looking at the distribution of cases over the possible paths in the process.
Example: By visualizing the percentage of process instances following each possible path the user can decide to focus on the most frequent behavior. This also allows determining whether there are a lot of variations in the process and possibly representing a starting point for a redesign meant to reduce these variations.

Use case 5: Exceptions from the normal path

Description: Discover the outliers of the process by looking at the exceptional behavior observed in practice.
Example: Deviations from the normal behavior can be an indication that the guidelines are not followed in practice or that things are not implemented in a correct way. Visualizing the outliers of a process may trigger a process redesign.

Use case 6: The degree in which the rules are obeyed

Description: Check whether the rules and regulations related to the process are obeyed. **Example:** Checks whether the business rules and principles are respected in practice and assess whether they are necessary to ensure the integrity and the quality of the process.

Use case 7: Compliance to the explicit model

Description: Compare the documented process model with the real process as observed in the event log.
Example: By executing this use case scenario, the user can detect possible inconsistencies between the real process and the theoretical one, which may then either lead to the enforcement of the documented process or to an update of the documentation.

Use case 8: Resources per task

Description: Discover the relation between resources and tasks.
Example: This use case gives an overview on both the group of resources working on the process and the set of tasks specific to each of these resources. This is a starting point for understanding the organizational dimension of a process. Additionally, this use case tells who is responsible for each task in isolation and who is sharing the responsibility for a certain activity.

Use case 9: Resources involved in a case

Description: Discover the group of resources involved in solving a particular case. **Example:** This information could be used to track down problems related to a particular process instance and to find the source of error or the cause of the delay among the resources involved in the execution of the case.

Use case 10: Work handovers

Description: Manage resource location or determine possible causes for quality and time issues by looking at how work is transferred between resources.
Example: This information can be used to improve the distribution of work or for analyzing quality, time and cost problems.

Use case 11: Central employees

Description: Determine who the central resources for a process are by analyzing the social network based on handovers of work.

Example: Knowing who the central employees of a process provides insight into the resource perspective and the way the process is organized. The central employee can be seen as the middle-man in the process, as a hub for work handover. Being aware of this information, the users can further determine for instance how sensitive the process is in case the central employee leaves the company.

Use case 12: Throughput time of cases

Description: Determine the time that passed since the start of a case in process until its completion.
Example: Statistics related to throughput time serves as a starting point for more in depth performance analysis.

Use case 13: Slowest activities

Description: Discover potential time problems by looking at the slowest activities in the process.
Example: The user can identify where problems might exist by considering the activities that take most time.

Use case 14: Longest waiting times

Description: Determine delays between activities by analyzing the waiting times before each activity.
Example: Performance problems caused by a bad synchronization of tasks or by an unsuitable resource allocation are reflected by the long waiting times specific to some activities.

Use case 15: Cycles

Description: Learn whether additional delays occur in the process due to cycles.
Example: This kind of behavior observed within in a process might be an indicator that in practice things are not happening as they should happen. Loops introduce additional time delays, decrease the efficiency of the overall process, might be a sign that an error was made, and can cause the occurrence of bottlenecks.

Use case 16: Arrival rate of cases

Description: Determine the frequency with which new cases arrive in the process.
Example: The pattern followed by cases arriving in the process is important information that is taken into consideration when a new process redesign is applied.

Use case 17: Resource utilization rate

Description: Determine what are the utilization rates of the resource i.e, measure the fraction of time that a resource is busy.
Example: This information provides insight from two perspectives. First, the existence of overutilized resources might also indicate possible bottlenecks in the process. Opposite, resources with a very low utilization are a sign that the organization of the process is inefficient and that a redesign should be considered.

Use case 18: Time sequence of events

Description: Get a deeper understanding on the organization of a process by looking at the time sequence of activities for a specific case. (e.g. Gant-graph for activities).
Example: By having the possibility of visualizing the time sequence of events, the user can get a better overview on the overlapping activities in the process. This representation captures extra-information about the process that is lost in the control flow model representation.

Use case 19: Business rules

Description: Discover what are the process attributes that influence the choice points and what are the conditions for following one branch or another.
Example: The user can discover which data fields stored in the log determine the choice between different branches in the process.

Appendix 2 - Interviews Summary

Interview 1

Role: Process Analyst
Industry: Banking

1. **Introduction**
   - **Experience with Process Mining**
     The interviewee was involved in a pilot process mining project within a Dutch bank. The purpose of the project was to analyze an internal HR process using two process mining tools: ProM and Futura Reflect. The process was simple and well-defined. At the end, the results obtained from the two systems were compared.
   - **Benefits brought by Process Mining**
     According to the interviewee the main benefits that process mining brings are the possibility of visualizing the actual process and the detection of bottlenecks. He also mentioned the role of process mining in the objective analysis of the AS-IS situation and as a basis for redesign and the possibility of analyzing different parts of a process in different levels of depth.
2. **Results ranking list of use cases**

Table 4. Results ranking - Interview 1

Order number	1	2	3	4	5	6	7	8	9	10
Use case number	1	3	15	13	4	12	2	7	14	16

Order number	11	12	13	14	15	16	17	18	19
Use case number	17	18	19	5	6	9	10	11	8

3. **Additional insights**
   In the interviewee's opinion process mining should provide a mechanism of performing root-cause analysis when problems are discovered. At the moment process mining stays in an experimental phase because a link between objective insights and further follow-up is required. It is often the case that managers or process analysts get a large amount of information about the analyzed process, but this information is not translated into solutions. The interviewee also mentioned the possibility of combing and comparing the objective data obtained by applying process mining techniques with subjective data with the purpose of gaining more accuracy when analyzing a process. On a different perspective, people seem more enthusiastic for use on the long term (for projects) than on a daily basis. No missing use cases were identified.
   As a conclusion of this interview, process mining can be used in three different directions: start point for a structural change, auditing, and as a stand-alone functionality (similar to the concept of quick-scan).

Interview 2

Role: Process Analyst
Industry: Business Process Management Consultancy

1. **Introduction**
   - **Experience with Process Mining**
     The experience of the interviewee in the process mining context consists of attending a presentation on Futura Reflect. He has however no practical experience with using it.
   - **Benefits brought by Process Mining**
     The most significant benefit of process mining is that its results are based on evidence and facts, making it an accurate technique. Additional advantages are related to the time and cost perspectives: process mining is faster and cheaper than the other techniques employed in the process analysis phase.
2. **Results ranking list of use cases**

Table 5. Results ranking - Interview 2

Order number	1	2	3	4	5	6	7	8	9	10
Use case number	3	12	6	15	1	13	2	4	14	18

Order number	11	12	13	14	15	16	17	18	19
Use case number	10	5	17	19	16	9	7	8	11

3. **Additional insights**
   No missing use cases were identified. According to the interviewee process mining has a lot of potential, but many people do not see it now.

Interview 3

Role: Process Analyst
Industry: Healthcare

1. **Introduction**
   - **Experience with Process Mining**
     The interviewee was involved in the project aimed at analyzing a gynecology-oncology process using techniques implemented in ProM. The process spans over several departments in the hospital (multi-disciplinary), contains a lot of diagnosis steps and decision making and covers a large number of patients. The analysis showed differences between the protocol and what happens in practice. The first step of the analysis was to look at the general structure of the process and then in the second phase problems related to throughput time, repetitions and resource allocation were deeper inspected. The focus of the analysis was on the most frequent behavior.

- **Benefits brought by Process Mining**
  Process mining takes less time and gives insight into the process. These are according to the interviewee the main advantages. He also mentioned the ease of detecting patterns compared to the manual operation and the benefits provided by the simulation aspect of process mining.

2. **Results ranking list of use cases**

Table 6. Results ranking - Interview 3

Order number	1	2	3	4	5	6	7	8	9	10
Use case number	1	7	3	4	15	10	12	13	14	17

Order number	11	12	13	14	15	16	17	18	19
Use case number	2	11	19	18	16	9	8	5	6

3. **Additional insights**
   No missing use cases were identified. According to the interviewee process mining can also be viewed as a communication tool. Once again, he mentioned the usefulness of coupling simulation with process mining.

Interview 4

Role: Auditor
Industry: Public Sector

1. **Introduction**
   - **Experience with Process Mining**
     The experience of the interviewee with process mining techniques consists of the analysis of the TBU (Tegemoetkoming Buitengewone Uitgaven) process with the purpose of getting better insight and checking whether what happens in practice is 100% conform to the theoretical process. The project had three goals: to identify that cases that do not complete, but get stuck in the process,to check its correctness, and to look into the time aspect. The process was spread over multiple IT systems, so gathering the data and linking all systems was a bit of a problem before starting the actual analysis.
   - **Benefits brought by Process Mining**
     The most important advantages specific to process mining are the complete transparency it offers and the full insight into processes. In the future process mining might be also used as a tool for continuous monitoring.
2. **Results ranking list of use cases**
3. **Additional insights**
   The ease of using a process mining tool represents an important aspect for the interviewee. In his opinions, systems should provide ways of representing the results in a fashion that is more presentable to managers. No missing use cases were found.

Table 7. Results ranking - Interview 4

Order number	1	2	3	4	5	6	7	8	9	10
Use case number	5	14	19	18	6	7	2	1	4	9

Order number	11	12	13	14	15	16	17	18	19
Use case number	11	15	13	10	16	12	17	8	3

Interview 5

Role: Auditor
Industry: Public Sector

1. **Introduction**
   - **Experience with Process Mining**
     The interviewee has no practical experience with using process mining. However, he supervised interviewee 6 in the project made for analyzing a purchase process from a SAP system from an auditing perspective. The control flow of the process and conformance checking aspects were investigated. The interviewee also collaborated with interviewee 4 in the project that analyzed the TBU process.
   - **Benefits brought by Process Mining**
     According to the interviewee, process mining enables its users to see what is exactly happening in a process. This way one can discover strange things taking place in practice, which can constitute exceptions from the normal path of events. Moreover, using the filtering functionalities, one could choose to look into cases which are more important of significant than others.
2. **Results ranking list of use cases**

Table 8. Results ranking - Interview 5

Order number	1	2	3	4	5	6	7	8	9	10
Use case number	1	2	3	4	7	18	19	15	5	14

Order number	11	12	13	14	15	16	17	18	19
Use case number	13	16	8	6	17	12	11	10	9

3. **Additional insights**
   Additional features that a process mining tool should provide, in the interviewee's opinion, are the continuous monitoring of a process and the possibility of receiving alerts every time something is not working as it should. He also mentioned the importance of the user interface and the large amount of time it takes to prepare the log for the analysis phase. No additional use case was identified.

Interview 6

Role: Auditor
Industry: Public Sector

1. **Introduction**
   - **Experience with Process Mining**
     The interviewee analyzed from an auditing perspective a purchase process from a SAP system using process mining techniques. This was done in the context of the interviewee's master thesis "Automated Analysis of Business Processes for IT auditing". As an IT-auditor working in the public sector, the interviewee is still using process mining to investigate processes.
   - **Benefits brought by Process Mining**
     As benefits of process mining the interviewee mentioned the possibility of seeing what is actually happening in practice, and from an auditing perspective the ability of comparing the explicit process with the one discovered from the event log.
2. **Results ranking list of use cases**

Table 9. Results ranking - Interview 6

Order number	1	2	3	4	5	6	7	8	9	10
Use case number	1	4	7	3	6	5	2	15	12	13

Order number	11	12	13	14	15	16	17	18	19
Use case number	14	18	19	17	8	9	10	16	11

3. **Additional insights**
   According to the interviewee process mining could be used in the context of continuous monitoring, but setting up the environment for this still requires some manual steps. Important features that process mining software should incorporate are preprocessing of data and documentation and help functions. The interviewee also mentioned the necessity of repeating identical steps when performing the same type of analysis on different events log and suggested the possibility of recording the actions performed when setting up an analysis, saving them, and then using the resulting sequence of saved actions for further setting analysis, without the need of manually performing them again for every new analysis.

Appendix 3 - Survey Contents

1. In which industry do you work?
2. Which role fits you best?
   (a) Process Analyst
   (b) Auditor
   (c) Process Manager
   (d) Researcher
   (e) Other (Please specify which)
3. Do you have any practical experience with process mining?
   (a) No, not yet.
   (b) Yes, I have used it. (Please specify also the number of projects)
4. If you answered Yes to Question No. 3, for which purpose did you use process mining?
5. What do you think is the biggest benefit of process mining?
6. Please rank all the following process mining use cases according to their importance/relevance to you. To do this: first read the complete list of use cases and identify the most important use case. Then, drag it to the right side and repeat the process for the remaining use cases.
   1 Determine the structure of an unknown process or discover how a process looks like in practice.
   2 Get a deeper understanding of the process by looking at the probabilities of following one path or another after a choice point.
   3 Discover what is the path in the process that is followed by the highest percentage of cases.
   4 Discover common and uncommon behavior in the process by looking at the distribution of cases over the possible paths in the process.
   5 Discover the outliers of the process by looking at the exceptional behavior observed in practice.
   6 Check whether the rules and regulations related to the process are obeyed.
   7 Compare the documented process model with the real process as observed in the event log.
   8 Discover the relation between resources and tasks.
   9 Discover the group of resources involved in solving a particular case.
   10 Manage resource location or determine possible causes for quality and time issues by looking at how work is transferred between resources.
   11 Determine who the central resources for a process are by analyzing the social network based on handovers of work.
   12 Determine the time that passed since the start of a case in process until its completion.
   13 Discover potential time problems by looking at the slowest activities in the process.
   14 Determine delays between activities by analyzing the waiting times before each activity.

15 Learn whether additional delays occur in the process due to cycles.
16 Determine the frequency with which new cases arrive in the process.
17 Determine what are the utilization rates of the resource i.e, measure the fraction of time that a resource is busy.
18 Get a deeper understanding on the organization of a process by looking at the timely sequence of activities for a specific case. (e.g. Gant-graph for activities).
19 Discover what are the process attributes that influence the choice points and what are the conditions for following one branch or another.

7. Do you think there is any use case missing?
   (a) No
   (b) Yes (Please specify which)
8. How many of the processes in your area contain parallelism (activities that can be performed in parallel)?
   (a) None
   (b) Up to 20%
   (c) Between 20 and 50%
   (d) Between 50 and 80%
   (e) More than 80%
9. A typical process in your area has the following characteristics:
   (a) Number of cases (process instances) per year:
   (b) Number of activities per case:
10. Which additional functionalities, not covered in the use cases presented in question 6, do you consider important for a process mining tool? (for example specific import or export capabilities)
11. Is there anything else you would like to add?

Appendix 4 - Survey Results

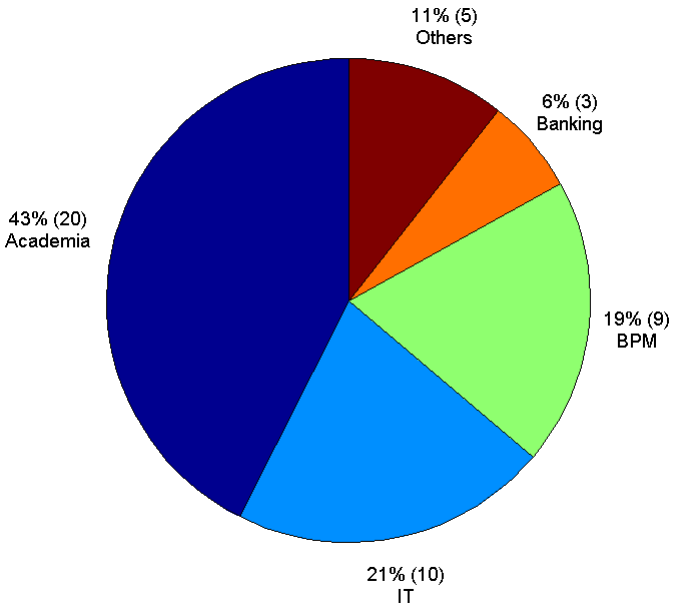

Fig. 6. In which industry do you work?

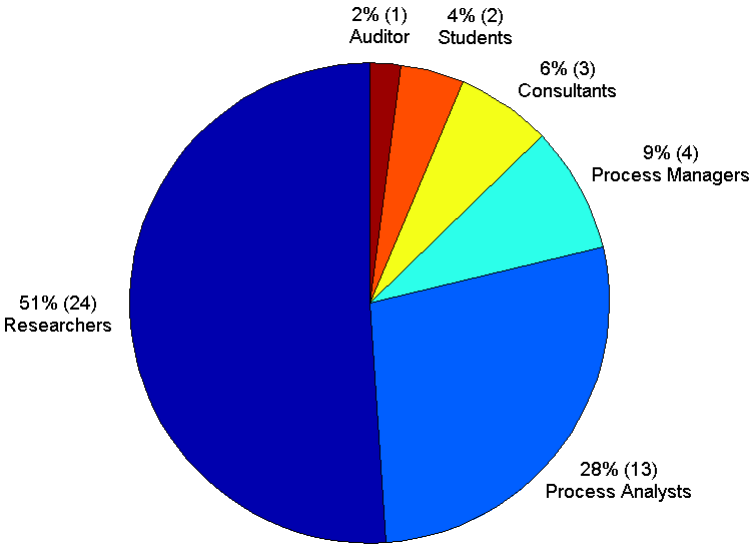

Fig. 7. Which role fits you best?

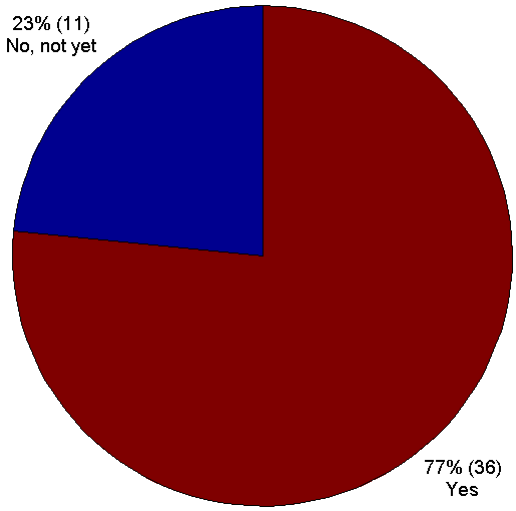

Fig. 8. Do you have any practical experience with process mining?

Table 10. For which purpose did you use process mining?

Discovery	Discover and show the true business flow within companies
	Find the frequent patterns in the event log
	Discover processes from log traces
	Mine the actual used process model out the history data (Costs)
	Automated business process discovery
Conformance checking	Conformance between event logs and Petri net specifications
	Compliance analysis
	Evaluate the compliance of processes with best practices
Performance analysis	Identify inefficiency
	Resource usage analysis
	Duration analysis
Organizational analysis	Show the maximum number of involvement of resource
Monitoring	Develop a real-time process monitoring solution

Table 11. What do you think is the biggest benefit of process mining?

Functional qualities	Visualization of processes, analysis	
	Extracting process models	
	Exploring and discovering processes	
	Process performance	
	Get real key figures about your process	
Non-functional qualities	**Objectivity**	Objective unbiased information
		Objectivity on how the process is executed in real-life
	Accuracy	Accurate information based on actual facts
	Speed	Speed of getting results
		Saving time
	Transparency	Making real processes transparent
		Transparent view and analysis of business processes

Table 12. Overall results use cases rankings

Rank	Use Case	Score
1	U1. Structure of the process	2.3
2	U7. Compliance to the explicit model	5.9
3	U3. Most frequent path in the process	7.1
4	U4. Distribution of cases over paths	7.2
5	U2. Routing probabilities	8.2
6	U5. Exceptions from the normal path	8.3
7	U6. The degree in which the rules are obeyed	8.7
8	U15. Cycles	8.8
9	U12. Throughput time of cases	9.4
10	U19. Business rules	10
11	U18. Time sequence of events	11.5
12	U13. Slowest activities	11.6
13	U14. Longest waiting times	11.9
14	U17. Resource utilization rate	12.4
15	U10. Work handovers	12.5
16	U8. Resources per task	12.6
17	U9. Resources involved in a case	12.7
18	U11. Central employees	12.8
19	U16. Arrival rate of cases	13.8

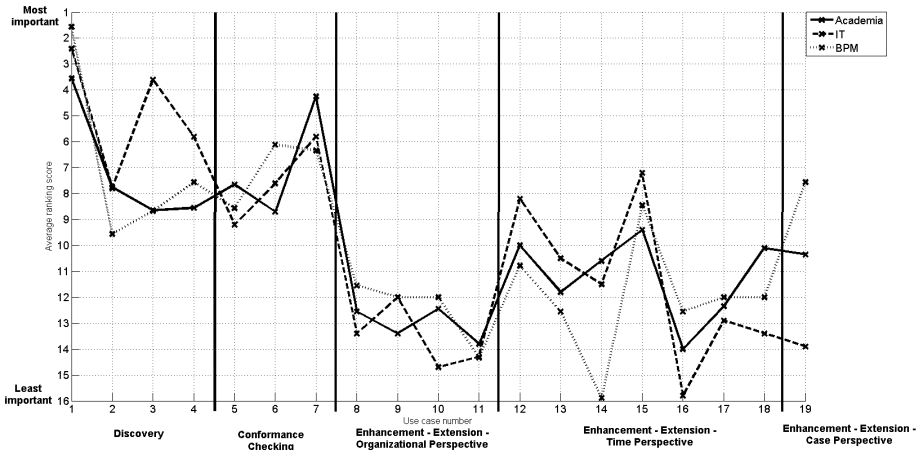

Fig. 9. Use cases ranking results based on respondents activity domains

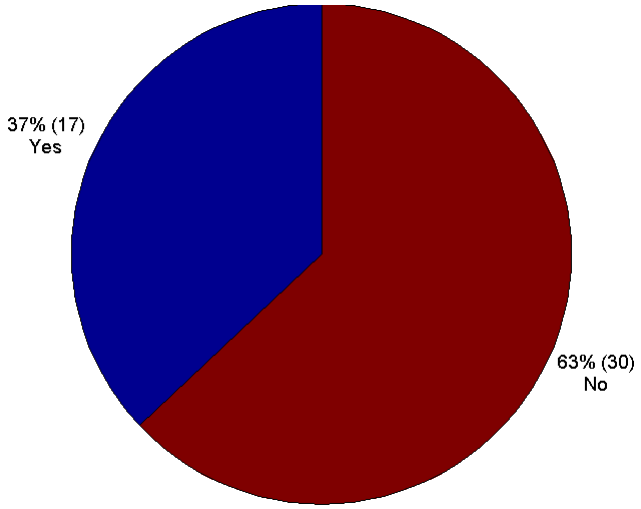

Fig. 10. Do you think there is any use case missing?

Table 13. Missing use cases

Measure different KPIs	Measure key performance indicators (costs, quality, flexibility)
	Calculating costs based on events (Activity Based Costing etc.)
	Analysis of attributes (Costs)
Build a simulation model	Build simulation models
	Use process mining in combination with simulation for forward decisional support analysis
	Create a simulation model
Use model to make predictions	Predicting the result of a process
	Analyze an event log online to make predictions for the future
Preprocessing	Trace clustering
	Filtering irrelevant cases out of the event log

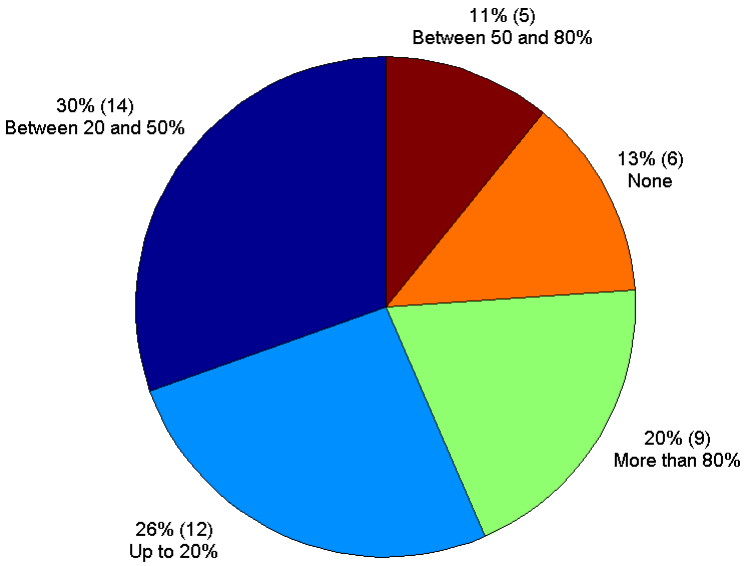

Fig. 11. How many of the processes in your area contain parallelism?

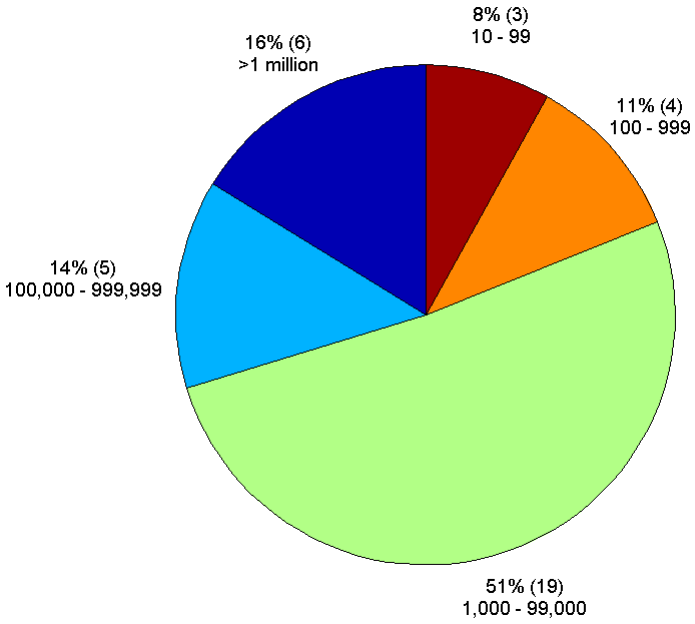

Fig. 12. Typical process - Number of cases (process instances) per year

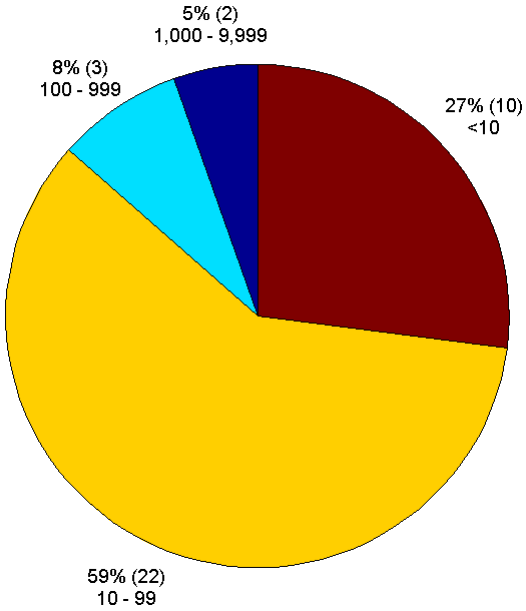

Fig. 13. Typical process - Number of activities per case

Table 14. Which additional functionalities do you consider important for a process mining tool?

Input and output capabilities	Export to excel
	XPDL export
	Import/export enhancements are always welcome
	Export capabilities for figures and results
	Easy export of output, both graphical as results
	Event log to EPC or BPMN export results as pdf or html
	Export to BPM tools
Filter and cluster data	Clustering to identify similar patterns in the process
	Intelligent filters
Integration with external systems	Plugins for existing BPMS, such as Oracle
	Database connection
	Integration with statistical tools (for example: R)
	Integrated data mining tools to analyze the patterns in a process
	Integration with ERP, CRM, etc systems
Animation capabilities	Replay
	See the animation of a process
Support for large input event logs	Visualization for the large-volume of process instance data
	Support for high amount of cases