A Survey of Process Mining Competitions: the BPI Challenges 2011–2018

Iezalde F. Lopes, Diogo R. Ferreira

Instituto Superior Técnico (IST), University of Lisbon, Portugal
iezalde.lopes@tecnico.ulisboa.pt

Abstract. In recent years, several advances in the field of process mining, and even data science in general, have come from competitions where participants are asked to analyze a given dataset or event log. Besides providing significant insights about a specific business process, these competitions have also served as a valuable opportunity to test a wide range of process mining techniques in a setting that is open to all participants, from academia to industry. In this work, we conduct a survey of process mining competitions, namely the Business Process Intelligence Challenge, from 2011 to 2018. We focus on the methods, tools and techniques that were used by all participants in order to analyze the published event logs. From this survey, we develop a comparative analysis that allows us to identify the most popular tools and techniques, and to realize that data mining and machine learning are playing an increasingly important role in process mining competitions.

Keywords: Process Mining, Data Mining, Machine Learning.

1 Introduction

The field of data science is thriving with competitions where participants are asked to perform challenging tasks involving the analysis of real-world data. In the field of process mining, the competition that has brought the community together around real-world event logs is the Business Process Intelligence Challenge (BPI challenge). Since 2011, the BPI challenge has been providing event logs that have served as a testbed for innumerous process mining techniques.

An interesting aspect of the BPI challenge is that it has drawn the attention not only of academic researchers, but also of practitioners working at the intersection between business process management and data analysis. Furthermore, the BPI challenge has served as an introduction to many students entering the field of process mining, and some of its event logs, namely the one from the BPI Challenge 2012, have become a standard of reference for many authors.

In this work, we provide a summary of each BPI challenge, followed by an overview of the tools and techniques that have been most used across all BPI challenges. This paper is intended to share our findings and provide a sense of what sort of techniques and approaches the community is making use of when dealing with real-world event logs.

2 BPI Challenges

In this section, we provide a summary of each BPI challenge, including a brief description of the business process domain, the business questions (if applicable), the winning submissions, and an overall impression about the approaches that were used to address the challenge.

2.1 BPI Challenge 2011

In the first edition, the BPI challenge involved an event log from a Dutch Academic Hospital. The event log concerned the diagnosis and treatments performed on patients in a Gynecology department. The names that were given to those treatments do not seem to follow a strict format. This led to a relatively large number of different task names, which means that any direct control-flow analysis of the event log is likely to yield a spaghetti model.

In this first edition, participants could focus on a specific aspect to analyze it in depth, or they could focus on a broader range of aspects without going into much detail. The winning authors were J.C. Bose and W. van der Aalst [1], who used the enhanced fuzzy miner [2] and trace alignment techniques [3] in ProM to group homogenous cases. In the end, they were able to present a compact process model. The trace alignment analysis carried out by the team also yielded common patterns of execution and exceptional/rare behavior.

In general, all authors used ProM combined with different preprocessing techniques to create subsets of cases which could be mined to obtain an understandable process model.

2.2 BPI Challenge 2012

In the second edition, the BPI challenge involved an event log from a Dutch Financial Institute. The event log concerned an application process for personal loans. Here, three distinct subprocesses could be identified by the event prefixes *A_* (for application states), *O_* (offer states), and *W_* (work items).

In this challenge, there were four business questions of interest to the process owner: (1) estimating the total cycle time; (2) determining which resources generate the highest activation rate of applications; (3) discovering the process model; and (4) identifying which decisions have greater influence on the process flow.

The winning authors were A. D. Bautista et al. [4] (from a New York-based consulting firm), who used Disco to understand the process, and decision trees to segment loan applications according to their approval result.

In general, authors used both ProM and Disco, and they also focused on several process mining perspectives (control-flow, organizational, performance) using a variety of analysis plug-ins. Preprocessing did not play such a large role as in the first edition, since three different subprocesses were already identified in the event log.

2.3 BPI Challenge 2013

In the third edition, the BPI challenge involved an event log from Volvo IT Belgium. The log contained events from an incident management system called VINST. Each event refers to a change in the status/sub-status of an incident.

In this challenge, there were four business questions of interest to the process owner: (1) whether incidents are pushed too often to second- and third-line support; (2) whether there is ping-pong behavior between teams; (3) whether the wait-user status reveals performance problems; and (4) whether process instances conform across departments.

The winning authors were C. J. Kang et al. [5] (a team from a South Korean university), who used a footprint matrix to capture the activity precedence. They found that departments were not conforming to each other, and they analyzed the process at each department with Disco.

In this challenge, there was a noticeable trend towards the use of statistics. In a sense, this was to be expected since the business questions required ranking and comparison among process instances and case attributes. Disco and ProM were the most popular tools. Both the control-flow perspective and the organizational perspective played a key role.

2.4 BPI Challenge 2014

In the fourth edition, the BPI challenge involved an event log from Rabobank Group ICT. The log concerned ITIL processes such as interaction management, incident management and change management.

The main goal was to predict the workload of the Service Desk (SD) and IT Operations (ITO) when a new change is introduced. There were four business questions involving: (1) identification of impact patterns; (2) impact of such patterns on workload; (3) improvement of service level after each change; and (4) creative analysis, where participants could pursue other insights.

The winning authors were P. Buhler et al. [6] (from the same New York-based firm as in 2012), who used custom metrics to measure performance and improvements after a change. Using decision trees, they classified the impact of those changes on workload. Also, using a multinomial logistic regression model, they were able to determine the probability of a specific change resulting in a given impact.

In this challenge, a new student category was introduced, which targeted BSc, MSc, PhD students. In this category the winning authors were G. Cacciola et al. [7], who used several custom plots to address the business questions. They also used Disco for performance analysis.

In general, most authors turned to data mining tools and techniques to answer the business questions. It is the first time that we see data mining tools being preferred over process mining tools, although Disco and ProM still played an important role in the analysis.

2.5 BPI Challenge 2015

In the fifth edition, the BPI challenge concerned the application process for construction permits in five Dutch municipalities. There were six business questions involving: (1) roles of people involved in the process; (2) possible improvements to the organizational structures; (3) changes in the process due to relocation of employees; (4) effect of outsourcing in organizational structures; (5) throughput times; and (6) control-flow for each municipality.

The winning author was U. van der Ham [8] (an independent consultant), who used Disco to analyze the average throughput time per municipality, and decision trees in WEKA to predict resource assignment. Using concept drift analysis [9] in ProM, the author identified the major changes in the process.

In the student category, the winning authors were I. Teinemaa et al. [10], who used the Kleinberg algorithm [11] to analyze the organization structure and identify key resources. They also analyzed handover of work with Disco and performed concept drift analysis [9] in ProM. Using the heuristics miner [12] and log replay in ProM did not reveal significant differences between the municipalities.

In this challenge, the business questions touched the control-flow, organizational and performance perspectives of process mining. Hence, most authors made use of process mining tools such as Disco, ProM and Minit. The use of data mining techniques (e.g. decision trees) was present in some submissions but had only a secondary role.

2.6 BPI Challenge 2016

In the sixth edition, the BPI challenge involved some very large event logs (~1 GB) from the Dutch Employee Insurance Agency (UWV). The logs concerned the customer interaction through different channels (website, messages, and call center) when applying for unemployment benefits. The main goal was to provide insights about the way the website was used.

In this challenge, there were six business questions: (1) identification of usage patterns on the website; (2) change of usage patterns over time; (3) transitions from the website to other channels; (4) change in customer behavior after using other channels; (5) customer behavior leading to complaints; (6) any new insights that could be obtained from the event log.

In this edition, a sponsor provided participants with the opportunity to apply for a free license of a software tool (Minit) for use in the BPI challenge.

The winning author was again U. van der Ham [13], who used mostly a spreadsheet-based analysis in Excel to collect statistics related to the business questions. The author also used IBM Watson [14] to try to find correlations between complaints and case attributes such as the number of visits to specific webpages.

In the student category, the winning authors were S. Dadashnia et al. [15], who clustered the traces and then used Disco to analyze the control-flow. In addition, they created a predictive model using a recurrent neural network (RNN) [16] to predict the next user action. They also used sequence clustering in ProM [17] to derive usage patterns.

The event logs for this challenge were much larger than in previous challenges (one of the logs had over 7 million events). Due to the log size and characteristics, it was difficult to apply standard process mining techniques, so most authors focused on simpler techniques based on filtering and clustering. There were some attempts at using machine learning (namely neural networks) but this was not very effective on this dataset.

2.7 BPI Challenge 2017

In the seventh edition, the BPI challenge involved the same Dutch Financial Institute as in the BPI Challenge 2012, and the event log concerned an upgraded version of the loan application process.

There were four business questions: (1) throughput times, in particular the time a customer is waiting for the bank and vice-versa; (2) influence of multiple information requests on offer acceptance; (3) comparison between single-offer and multiple-offer customers; (4) any other interesting trends.

In this edition, there were three categories – student, academic and professional – and there was a record number of submissions (23 in total). The challenge also gave participants the opportunity to use tools provided by the sponsors (Minit and Celonis).

In the student category, the winning authors were E. Povalyaeva et al. [18], who created a BPMN diagram in Celonis based on their log analysis, and then assessed conformance using the same tool. They used ProM for concept drift analysis, and Disco for performance analysis. The authors also used random forests to analyze the process outcome based on several features.

In the academic category, the winning authors were A. Rodrigues et al. [19], who used Disco, ProM and Yasper. Disco was used to obtain a process model. ProM was used for conformance checking by calculating the fitness and precision of a Petri net model, and also replaying the log on that model. The authors also modeled parts of the process in BPMN. Yasper was used to assess the performance perspective together with ProM and Disco.

In the professional category, the winning authors were L. Blevi et al. [20] who combined process mining with KPMG's customer experience methodology. Using Power BI, the authors provided several statistics. They also obtained a process model using fuzzy miner in ProM. Using R and Microsoft Azure Machine Learning Studio, the team created predictive models using logistic regression, random forests, and neural networks.

In general, there was a lot of process mining analysis, combined with data mining and machine learning techniques such as decision trees, random forests, etc. A wide variety of process mining tools were used (Disco, ProM, Celonis, Minit, etc.). Despite the large number of submissions, the analysis was mostly focused on the business questions, with similar results being reported by all authors.

2.8 BPI Challenge 2018

In the eighth edition, the BPI challenge involved an event log from the European Agricultural Guarantee Fund, in Germany. The event log concerns annual payments to farmers. The workflow is based on document types; each document has a state that allows some actions to be performed.

In this challenge, there were four business questions: (1) detection of undesired cases (e.g. late payment); (2) improving the sampling of applications selected for inspection; (3) differences between departments and relation to undesired outcomes; (4) differences across time.

Again, participants were given the opportunity to use tools provided by the sponsors (Minit and Celonis).

In the student category, the winning authors were J. Brils et al. [21], who obtained a process model with Disco and collected several statistics after filtering the event log with Python. For outcome prediction, they tried several machine learning techniques available in RapidMiner (naive Bayes, logistic regression, neural networks, decision trees, etc.).

In the academic category, the winning authors were S. Pauwels and T. Calders [22], who used the competition to test their own method to detect concept drift. Their model is based on Bayesian networks and, after training it, they were able to detect two drift points. They also used attribute-density plots to find differences between departments.

In the professional category, the winning authors were L. Wangikar et al. [23], who used Celonis to obtain a process model. They used predictive models based on binomial logistic regression to detect undesired outcomes. They also analyzed concept drift in ProM and conformance checking in myInvenio to detect changes in the control-flow.

In this edition, data mining techniques were predominant over process mining. This was due to the nature of the business questions, which involved prediction. These questions received a lot of attention from participants. Process mining was used to get an idea of the control-flow, and to analyze concept drift. Subsequent analysis was performed using machine learning.

3 Comparative Analysis

In this section, we analyze the tools and techniques that were used in the BPI challenges, not only by the winners, but across all submissions. We also analyze the use of specific ProM plug-ins, and of data mining/machine learning techniques.

3.1 Techniques

Typically, process mining focuses on the control-flow, organizational, and performance perspectives. Some BPI challenges involve all these perspectives, while others involve only some of them. In general, however, all BPI challenges go beyond those perspectives to include additional types of analysis, as shown in Table 1.

Table 1. Techniques used in the BPI challenges 2011–2018.

	2011	2012	2013	2014	2015	2016	2017	2018	Total
Control-flow discovery	X	X	X	X	X	X	X	X	**8**
Trace clustering	X	X	X	X		X	X	X	**7**
Social network analysis	X	X	X	X	X	X	X		**7**
Performance perspective	X	X	X	X	X		X	X	**7**
Log statistics		X	X	X	X	X	X	X	**7**
Conformance checking		X	X	X	X		X	X	**6**
Predictive modeling				X	X	X	X	X	**5**
Dotted chart analysis		X	X		X	X	X		**5**
Plotting/visualization			X	X		X	X	X	**5**
Trace alignment	X	X	X				X		**4**
Concept drift analysis					X		X	X	**3**
Spreadsheet-based analysis		X	X			X			**3**

Control-flow discovery techniques were the most used. This is to be expected since no process flow was ever provided in a BPI challenge, and one needs to understand the business process to be analyzed. In general, all authors used at least one control-flow mining technique, either to understand the business process or because a business question required it. Heuristics-based techniques [12] and fuzzy miners [24] were the most used for process discovery.

Trace clustering [25] is also among the most used techniques. Here the main reason is the fact that the derived process models were often too complex to be understood and there was a need to divide the event log into smaller and more homogenous groups of cases in order to mine an understandable process model.

Social network analysis gained an increased popularity, mostly because of the implementation of such techniques in the ProM framework [26], which made them easy to use. The analysis on a resource perspective was also made possible by the data available in the competitions.

Although many authors worked on the performance perspective, most participants only analyzed it thoroughly if a business question required it. Otherwise, basic performance statistics were provided, which were obtained using some tool (more on this below).

Statistics about the event log have been widely collected, and exploratory analysis is the main reason why participants use those statistics. Then techniques such as predictive modeling, data plotting/visualization have also gained special attention. Predictive modeling techniques have been used mainly when a business question required it, but they have also been helpful to study the control-flow perspective.

3.2 Tools

In Table 2, a list of the most popular tools in the BPI challenges is presented. Disco (commercial) and ProM (academic) were definitely the most used. Despite the sponsoring of other tool providers in some editions of the BPI challenge (e.g. Celonis, Minit), Disco was still the most used. However, the use of competing tools seems to be growing, especially in recent years, possibly due the influence of academic programs that provide students and researchers with the opportunity to use such tools.

Table 2. Popular tools used in the BPI challenges 2011–2018.

	2011	2012	2013	2014	2015	2016	2017	2018	Total
Disco		3	10	9	6	4	18	1	**51**
ProM	3	5	6	3	7	5	17	3	**49**
Excel		1	7	6	4	2	10		**30**
R			2	3	1	2	8		**16**
Celonis						1	13	1	**15**
Python			1				8	1	**10**
WEKA				3	2	1	2		**8**
Oracle		1	1	2	1		2		**7**
RapidMiner			1	1	1		2	1	**6**
Java			1	2	1	1			**5**
SQL Server				2		1	1		**4**
Minit					1	1	1		**3**
C#				1	1		1		**3**

In general, it was observed that the use of these tools follows a cascading pattern: (1) Disco and/or ProM are used to get an overview of the process; (2) additional statistics are collected from the event log using tools such as Excel and R; and (3) other tools are selected depending on the challenge and on the nature of the business questions. In some cases, authors have used database engines (e.g. Oracle, SQL Server) to compute statistics about the event log.

3.3 ProM plug-ins

Since ProM is a framework that includes a wide variety of plug-ins, it is interesting to check which of those plug-ins have been most used. In Table 3, the focus is on the use of ProM plug-ins only. Here, the heuristics miner [12], the dotted chart analysis [27], and the social network miner [26] take the podium, with a significant lead over the remaining ones.

Table 3. ProM plug-ins used in the BPI challenges 2011–2018.

	2011	2012	2013	2014	2015	2016	2017	2018	Total
Heuristics miner	2	3	3	1	2		2		**13**
Dotted chart analysis		2	1		4	1	5		**13**
Social network miner	1	1	3		2	1	3		**11**
Inductive miner					1		4		**5**
Fuzzy miner			1	2			2		**5**
Alpha miner		1	1				2		**4**
Sequence clustering			1			2	1		**4**
Trace alignment	1	1	1				1		**4**
Concept drift					1		1	1	**3**
Organizational miner					1		2		**3**
Filter log simple heurist.		1	1		1				**3**
Guide Tree Miner	1	1							**2**
LTL-checker	1	1							**2**
Originator-by-task	1						1		**2**
Pattern abstractions	1	1							**2**
Trace align. w/ guide tree	1	1							**2**

It is interesting to note that the social network miner is one of the top plug-ins and has been used even in BPI challenges where there were no business questions involving the organizational perspective. It seems that the social network miner is very useful to complement and/or corroborate results obtained in other analysis perspectives.

Following the top three, we find the inductive miner [28] and the fuzzy miner [24], which are two popular control-flow discovery techniques. Both allow some form of abstraction over the control-flow behavior (process trees in the inductive miner, and activity clusters in the fuzzy miner). Finally, it is worth noting that trace alignment [3] and concept drift [9] have been playing an increasingly important role.

3.4 Data Mining/Machine Learning

Since we came across several data mining/machine learning techniques during our survey, it is interesting to analyze their use in the BPI challenges. Table 4 presents the data mining/machine learning techniques used across all BPI challenge submissions.

Although only two BPI challenges (2014 and 2018) included a prediction goal, we can see in Table 4 that this type of analysis was performed in editions where it was not apparently required. Decision trees are by far the most used technique.

Participants used tools (for example, RapidMiner) which provide the implementation of the technique, which had only to be parametrized and/or customized to the data. Decision trees were mostly used for classification purposes across the BPI challenges. For example, in the BPI Challenge 2017, decision trees were used to classify the incompleteness of a loan application.

Table 4. Data mining/machine learning techniques used in the BPI challenges 2011–2018.

	2011	2012	2013	2014	2015	2016	2017	2018	Total
Decision Trees		1		4	3		3	1	12
Logistic Regression				2			2	2	6
Random Forest				1			3	1	5
Neural Network						1	2		3
Linear Regression				2					2
Support Vector Machine				1			1		2
Sequential Pattern Mining				1	1				2
Sequence Classification				1	1				2
Naïve Bayes				1				1	2
Ada Boost				1					1
Apriori Algorithm				1					1
Association Rules				1					1
Multilayer Perceptron				1					1
Binary Segmentation					1				1
K-means Clustering						1			1
Bayesian Networks								1	1
Generalized Linear Model								1	1
Deep Learning								1	1
Gradient Boosted Trees								1	1

Sometimes, decision trees are used to address specific business questions; other times, they are used to complement the analysis. About half of the times decision trees were used, it was due to the fact that there was a business question requiring to predict a behavior; in the other half, the aim was to provide additional insight to the analysis previously carried out during the process mining phase.

Logistic Regression and Random Forests are also worth mentioning. We have noted that they are usually picked as the second choice after decision trees. They were used for classification tasks as well as for predicting behavior.

In general, data mining and/or machine learning techniques tend to be used to enhance the business process analysis. When there is some aspect that cannot be explained by process mining techniques alone, data mining/machine learning techniques come as an aid to understand those issues.

4 Conclusion

The BPI challenges have been not only a testbed for process mining techniques, but have also brought many other approaches into the realm of process mining. Besides the analysis of the control-flow, organizational and performance perspectives, the business questions associated with the BPI challenges often require the use of data mining and machine learning techniques.

Examples are the use of decision trees to find the most important factors that influence the process outcome, and the use of neural networks for next-step prediction. Besides supervised machine learning, unsupervised techniques also play key role, especially with the use of clustering as a means of preprocessing to better understand the process and facilitate the analysis of the event log.

Having observed a growing use of data mining in process mining competitions, we expect that, in the future, the use of data mining techniques will become as important as the use of process mining techniques in the analysis of the event logs.

References

1. Bose, R.P.J.C., van der Aalst, W.M.P.: Analysis of patient treatment procedures: The BPI Challenge case study. In: LNBIP 99. Springer (2012).
2. Li, J., Bose, R.P.J.C., van der Aalst, W.M.P.: Mining context-dependent and interactive business process maps using execution patterns. In: LNBIP 66. Springer (2011).
3. Bose, R.P.J.C., van der Aalst, W.M.P.: Trace Alignment in Process Mining: Opportunities for Process Diagnostics. In: LNCS 6336. Springer (2010).
4. Bautista, A.D., Wangikar, L., Akbar, S.M.K.: Process Mining-Driven Optimization of a Consumer Loan Approvals Process: The BPIC 2012 Challenge Case Study. In: LNBIP 132. Springer (2013).
5. Kang, C.J., Kang, Y.S., Lee, Y.S., Noh, S., Kim, H.C., Lim, W.C., Kim, J., Hong, R.: Process Mining-based Understanding and Analysis of Volvo IT's Incident and Problem Management Processes. In: CEUR Workshop Proceedings Vol-1052 (2013).
6. Buhler, P., O' Callaghan, R., Aubry, S., Dejoy, D., Kuo, E., Shoup, N., Khosla, I., Ginsburg, M., Hartman, N., Mcbride, N.S.: Service Desk and Incident Impact Patterns Following ITIL Change Implementation. In: BPI Challenge 2014 (2014).
7. Cacciola, G., Conforti, R., Nguyen, H.: Rabobank: A Process Mining Case Study BPI Challenge 2014 Report. In: BPI Challenge 2014 (2014).
8. van der Ham, U.: Benchmarking of Five Dutch Municipalities with Process Mining Techniques Reveals Opportunities for Improvement. In: BPI Challenge 2015 (2015).
9. Bose, R.P.J.C., van der Aalst, W.M.P., Žliobaitė, I., Pechenizkiy, M.: Handling Concept Drift in Process Mining. In: LNCS 6741. Springer (2011).

10. Irene Teinemaa, Leontjeva, A., Masing, K.-O.: BPIC 2015: Diagnostics of Building Permit Application Process in Dutch Municipalities. In: BPI Challenge 2015 (2015).
11. Kleinberg, J.M.: Authoritative Sources in a Hyperlinked Environment. In: Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms (1998).
12. Weijters, A.J.M.M., van der Aalst, W.M.P., Alves de Medeiros, A.K.: Process Mining with the HeuristicsMiner Algorithm. (2006).
13. van der Ham, U.: Marking up the right tree: understanding the customer process at UWV. In: BPI Challenge 2016 (2016).
14. Chen, Y., Argentinis, E., Weber, G.: IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research. Clin. Ther. 38, (2016). https://doi.org/10.1016/j.clinthera.2015.12.001.
15. Dadashnia, S., Niesen, T., Hake, P., Fettke, P., Mehdiyev, N., Evermann, J.: Identification of Distinct Usage Patterns and Prediction of Customer Behavior. In: BPI Challenge 2016 (2016).
16. Evermann, J., Rehse, J.-R., Fettke, P.: A Deep Learning Approach for Predicting Process Behaviour at Runtime. In: LNBIP 281 (2017).
17. Veiga, G.M., Ferreira, D.R.: Understanding Spaghetti Models with Sequence Clustering for ProM. In: LNBIP 43. Springer (2010).
18. Povalyaeva, E., Khamitov, I., Fomenko, A.: BPIC 2017: Density Analysis of the Interaction With Clients. In: BPI Challenge 2017 (2017).
19. Rodrigues, A.M.B., Almeida, C.F.P., Saraiva, D.D.G., Moreira, F.B., Spyrides, G.M., Varela, G., Krieger, G.M., Peres, I.T., Dantas, L.F., Lana, M., Alves, O.E., França, R., Neira, R.A.Q., Gonzalez, S.F., Fernandes, W.P.D., Barbosa, S.D.J., Poggi, M., Lopes, H.: Stairway to value: mining a loan application process. In: BPI Challenge 2017 (2017).
20. Blevi, L., Robbrecht, J., Delporte, L.: Process mining on the loan application process of a Dutch Financial Institute. In: BPI Challenge 2017 (2017).
21. Brils, J.H.H., van den Elsen, N.A.F., de Priester, J., Slooff, T.A.: Business Process Intelligence Challenge 2018: Analysis and Prediction of Undesired Outcomes. In: BPI Challenge 2018 (2018).
22. Pauwels, S., Calders, T.: Detecting and Explaining Drifts in Yearly Grant Applications. In: BPI Challenge 2018 (2018).
23. Wangikar, L., Dhuwalia, S., Yadav, A., Dikshit, B., Yadav, D.: Faster Payments to Farmers: Analysis of the Direct Payments Process of EU's Agricultural Guarantee Fund. In: BPI Challenge 2018 (2018).
24. Günther, C.W.;, van der Aalst, W.M.P.: Fuzzy mining-adaptive process simplification based on multi-perspective metrics. In: LNCS 4714. Springer (2007).
25. Song, M., Günther, C.W., van der Aalst, W.M.P.: Trace Clustering in Process Mining. In: LNBIP 17. Springer (2009).
26. Song, M., van der Aalst, W.M.P.: Towards comprehensive support for organizational mining. Decis. Support Syst. 46, (2008).
27. Mans, R.S., Schonenberg, M.H., Song, M., van der Aalst, W.M.P., Bakker, P.J.M.: Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital. In: CCIS 25. Springer (2008).
28. Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P.: Discovering Block-Structured Process Models from Event Logs - A Constructive Approach. In: LNCS 7927 (2013).