EuroVis Workshop on Visual Analytics (2016)
N. Andrienko and M. Sedlmair (Editors)

Visual Analytics in Process Mining: Classification of Process Mining Techniques

Simone Kriglstein[1], Margit Pohl[1], Stefanie Rinderle-Ma[2], and Magdalena Stallinger[2]

[1]Faculty of Informatics, Vienna University of Technology, Austria
[2]Faculty of Computer Science, University of Vienna, Austria

Abstract

The increasing interest from industry and academia has driven the development of process mining techniques over the last years. Since the process mining entails a strong explorative perspective, the combination of process mining and visual analytics methods is a fruitful multidisciplinary solution to enable the exploration and the understanding of large amounts of event log data. In this paper, we propose a first approach how process mining techniques can be categorized with respect to visual analytics aspects. Since ProM is a widely used open-source framework which includes most of the existing process mining techniques as plug-ins, we concentrate on the plug-ins of ProM as use case to show the applicability of our approach.

Categories and Subject Descriptors (according to ACM CCS): H.4.m [Information Systems Applications]: Miscellaneous—

1. Introduction

Business processes describe activities relevant to a company's business, their execution order, the data flow between the activities, as well as the invocation of services and human resources. If business processes are implemented through Process-Aware Information Systems (PAIS) their execution can be logged in so called event logs by means of observing and storing events thrown by the activities at runtime. Process mining has arisen as a bundle of techniques that enables the discovery and analysis of process-relevant information from event logs [vdA11]. Event logs contain information on one or several processes. A process consists of event data of process instances or cases, i.e., the different executions of this process. A case consists of events that reflect the activity executions. Each event comprises several attributes. At minimum, the application of process mining techniques requires a case ID as well as the timestamps and labels of the executed activities. It is to note that process mining techniques typically count occurrences of events and their order relation in order to derive the structure of the process model. In addition, the actors that have executed a certain activity might be logged in the associated events as well.

Aside the algorithmic aspect, process mining entails a strong explorative perspective, i.e., process models are discovered but might be subject for further analysis. Hence, visual analytics has been named as one of the most crucial challenges in the context of process mining [vdAea11]. Visual analytics refers to the visual exploration of data (i.e., event logs) in order to provide users insights into the structure of the data, to enable them to draw conclusions, and to interact with the data [Kei02, KKEM10]. However, a systematic analysis on how visual analytics is realized in existing process mining software such as Celonis [Cel16], Fluxicon Disco [Flu12], Perceptive Process Mining [Lex16], QPR Process Analyzer [QPR16], SNP Business Process Analysis [SNP16], and ProM [Pro10a] is missing.

In this paper, we conducted an analysis of the process mining techniques in order to identify how visual analytic aspects are supported. The contribution of this paper is twofold. First, we propose a first approach how process mining techniques can be categorized with respect to visual analytics aspects. This categorization schema can be used as foundation for further research and can help to discuss and evaluate process mining techniques as well as can give a deeper understanding of the interplay of process mining and visual analytics. For this purpose, our schema groups the process mining techniques according to different promising categories to cover the representation component, interaction component, and the process mining perspective. Second, to show the applicability of our approach we decided to

© The Eurographics Association 2016.

DOI: 10.2312/eurova.20161123

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

concentrate on the plug-ins of ProM as use case. ProM is a widely used open-source framework (cf. [CP13]) which supports most of the existing process mining techniques in the form of plug-ins and is used by practitioners and academics. We investigate the most used ProM plug-ins (as identified in a study by Claes and Poels [CP13]) to give an overview of the supported visualization types and interaction techniques, what kind of input they require, what kind of output they generate, and for which process mining perspectives they were developed. Such an overview can be helpful as orientation for the future development of such plug-ins and how they can be improved.

2. Related Work

In recent years, researchers have tried to generalize the results of investigations in information visualization and visual analytics to develop systematic frameworks, especially for task descriptions (e.g., [AA05, BM13, SHB*14]). The goal of these generalization processes is to develop more systematic frameworks as foundations for future research and guidance for designers. Andrienko and Andrienko [AA05] developed a very detailed overview of tasks which can be supported by visualizations. Brehmer and Munzner [BM13] describe a multi-tier model of visualization tasks. Based on Brehmer and Munzner [BM13], Sedlmair et al. [SHB*14] developed a framework for visual parameter space analysis. All these approaches are supposed to form a basis for scientific research, so that investigations can relate to a consistent set of conceptual ideas. In addition, design decisions can be supported by such approaches because existing research is consolidated systematically in such frameworks. The research presented in this paper follows these lines and tries to come up with a similar framework in a more restricted area – plug-ins for process mining in business process analysis. This research can help to assess existing plug-ins to find out which interactions they support, what kind of input they need and what kind of output they generate. In addition, it can help to identify gaps in existing plug-ins and to clarify which output categories are combined with which interaction possibilities.

For the classifications of representation and interaction components, different categories were proposed in the last years (see, e.g., [AES05, GZ09, Kei02, Nor05, Maz09, Shn96, WY04, YKSJ07]). For our classification, we applied the data type taxonomy presented by Shneiderman [Shn96] in combination with the categories for interaction techniques presented by Yi et al. [YKSJ07]. Based on our previous work, we find the latter very valuable because it based on users' intents to consider higher-level user tasks than on low-level techniques provided by the system.

3. Methodology

The plug-ins were analyzed by two researchers in an iterative process. The analysis itself focused on the most used plug-ins in ProM 6.1 [Pro10b] as revealed by a study of Claes and Poels [CP13]. Claes and Poels asked 90 researchers and practitioners from the process mining community about their opinions regarding process mining and the ProM framework. Results revealed that 16 process mining technique plug-ins were rated as most frequently used. Each of this plug-ins requires different types of input data (e.g., event logs) to create an output of a specific type (e.g., Petri Net). For analyzing the output different visualization and interaction strategies are provided by the plug-ins. For the inspection of the plug-ins, one small test data set (including 2 cases and 10 events) and two larger test data sets (including 1000 cases and 10845 events / 1104 cases and 11855 events) were used. The resulting output of each plug-in was then investigated with respect to the following categories:

Data type: Shneiderman [Shn96] presents seven data types for visualizations: *One-dimensional* (e.g., lists of data items), *Two-dimensional* (e.g., geographic maps), *Three-dimensional* (e.g., 3D computer models), *Temporal* (e.g., timeline), *Multi-dimensional* (e.g., bar charts), *Tree* (e.g., dendrogram), and *Network* (e.g., node-link diagram). These categories were used to classify the visualizations of the outputs.

Interaction techniques: Yi et al. [YKSJ07] propose seven different interaction techniques categories: *Select* (e.g., mark data items), *Explore* (e.g., panning), *Reconfigure* (e.g., spatial arrangements), *Encode* (e.g., change representation type), *Abstract/Elaborate* (e.g., details-on-demand), *Filter* (e.g., change range or condition), and *Connect* (e.g., linking and brushing). These categories were used for analyzing the different interaction techniques which are provided by the plug-ins for interacting with the visualized output.

Process mining perspectives: van der Aalst [vdA11] presents four process mining perspectives: *Control-flow* (focus on the control-flow, e.g., ordering of activities), *Organizational* (focus on information about resources, e.g., involved actors), *Case* (focus on the properties of cases), and *Time* (focus on timing and frequency of events). The output types were classified based on these perspectives to identify for which purpose the plug-ins were developed.

4. Results and Discussion

Table 1 and Table 2 give an overview of the required input data, the corresponding output types, and the categorization of the output type with respect to the data type taxonomy, interaction strategies, and process mining perspectives for each of the investigated plug-ins. In the following, findings regarding the output type, input data, data type, and their interaction strategies are discussed in more detail:

Output type: From the 16 process mining technique plug-ins, two plug-ins were excluded since the output of these two plug-ins were only provided in textual form. For the

© The Eurographics Association 2016.

	Control-flow Perspective				**Org. Perspective**
Output	HeuristicsNet	Petri Net	Metrics-Repository	TSMiner-TransitionSystem	Social Network (SN)
Data Type for Visualization					
Multi-dimensional			✓		
Network	✓	✓	✓	✓	✓
Interaction Techniques for Visualization					
Select	✓	✓	✓	✓	✓
Explore	✓	✓	✓	✓	✓
Reconfigure	✓	✓	✓	✓	✓
Encode	✓		✓		✓
Abstract/Elaborate	✓	✓	✓	✓	✓
Filter			✓	✓	✓
Connect	✓	✓	✓	✓	
Required Input	Event Log	Event Log	Event Log	Event Log	Event Log
Plug-ins	Heuristics Miner; Flex.Heuristics Miner	Mine for a Petri Net using α-Algorithm	Mine for a Fuzzy Model	Mine Transition System	Mine for a Handover-of-Work SN; Mine for a Working-Together SN

Table 1: *Categorization of the output types with respect to the data type taxonomy, interaction strategies as well as process mining perspectives (control-flow and organizational perspective). For each output type the corresponding plug-ins and the required input data are listed.*

	Case Perspective			**Time Perspective**		
Output	PNRepResult	AlignmentTree	Event Log	DottedChart-Model	Fuzzy-Animation	TimeTransition-SystemAnnotation
Data Type for Visualization						
One-dimensional			✓		✓	
Temporal				✓	✓	
Multi-dimensional	✓		✓			
Network		✓			✓	✓
Interaction Techniques for Visualization						
Select	✓	✓	✓	✓		✓
Explore	✓	✓	✓	✓	✓	✓
Reconfigure		✓		✓		✓
Encode		✓	✓	✓	✓	✓
Abstract/Elaborate	✓	✓	✓	✓	✓	✓
Filter		✓		✓		✓
Connect			✓	✓		✓
Required Input	Event Log ∧ Petri Net	Guide Tree	Event Log	Event Log	Event Log ∧ MutableFuzzyGraph	Event Log ∧ Transition System with Event Playload
Plug-ins	Replay a Log on Petri Net for Conf. Analysis	Trace Alignment (with Guide Tree)	Add Artificial Events; Filter Log Using Simple Heuristics	Analyze using Dotted Chart	Animate Event Log in Fuzzy Instance	Analyze Transition System

Table 2: *Categorization of the output types with respect to the data type taxonomy, interaction strategies as well as process mining perspectives (case and time perspective). For each output type the corresponding plug-ins and the required input data are listed.*

remaining 14 process mining technique plug-ins, 11 different output types were identified.

For three output types (*MetricsRepository*, *Event Log*, and *FuzzyAnimation*), the corresponding plug-ins offer different visualization types in order to uncover new aspects of relationships. For example, the *Mine for a Fuzzy Model* plug-in includes a matrix representation, a node-link diagram, and different chart diagrams to support analysts to explore processes derived from event logs (see, e.g., Figure 1).

© The Eurographics Association 2016.

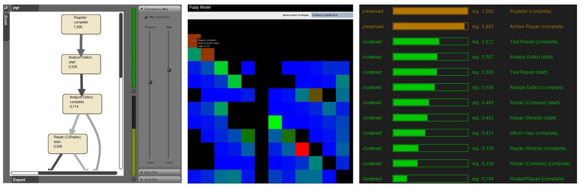

Figure 1: *Examples of visualization types provided by the* Mine for a Fuzzy Model *plug-in [Pro09b]: (left) node-link diagram to illustrate the process model derived from the event logs, (center) matrix representation to present measurement values for each event, and (right) bar chart to compare the significance of the event classes.*

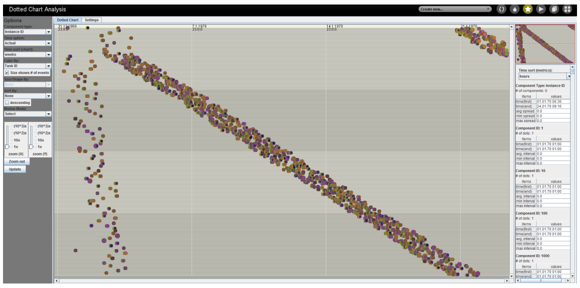

Figure 2: *Example of the visualization provided by the* Analyze using Dotted Chart *plug-in [Pro09a].*

Input data: An *Event Log* is the required input for most of the analyzed output types (7 from 11 output types). Further three output types require an *Event Log* in combination with another type as input. For example, the plug-in *Animate Event Log in Fuzzy Instance* additionally requires a so-called *MutableFuzzyGraph* as input.

Data type for visualizations: Most output types are of the data type *Network* (8 for 11) for which mainly node-link diagrams are used for visualization (e.g., for representations the control-flow of processes or of relationships between the actors).
Output types from the data type *Multi-dimensional* are often visualized by means of different chart diagrams (e.g., line chart or bar chart), for example, to analyze additional measurements of the event logs. Lists of data items or a tachometer to present the progression of animated events are examples for visualizations of outputs from data type *One-dimensional*.
For the data type *Temporal*, single events with respect to their timestamps are animated or are visualized along a time axis (see, e.g., Figure 2). An interesting observation was that not all output types which are concerned with the *Time* perspective, provide the data type *Temporal*. For example, the output type *TimeTransitionSystemAnnotation* only uses the data type *Network* and time information is only presented as annotation to a node.

Interaction techniques for visualization: Interaction techniques of the type *Abstract/Elaborate* as well as *Explore* are supported by all plug-ins and from the type *Select* by almost all (13 from 14). Furthermore, almost all output types, which are visualized through node-link diagrams (7 from 8), provide interaction techniques of type *Reconfigure*. All plug-ins which support the *Time* and *Organizational* perspective, facilitate *Encode* interactions and all plug-ins which support the *Control-flow* perspective offer the type *Connect*.
We observed that *Connect* is always provided in connection with *Abstract/Elaborate* (e.g., overview and detail view in combination with linking and brushing). A further interesting observation was that the plug-in *Animate Event Log in Fuzzy Instance*, which combines animation with a node-link diagram, provides interaction techniques mainly for controlling the animation but less for the exploration of the node-link diagram (e.g., to select a node or to change the arrangement of the nodes or links).

It should be kept in mind that for the analysis of the plug-ins it was sometimes challenging to identify the exact meaning of the measurements and variables due to the lack of a detailed documentation. A description of the goal of the plug-in, the provided visualizations, and about the presented data would be helpful, especially for users who are not so familiar with the different process mining techniques. Furthermore, it should be noted that the categories are partially overlapping and not exhaustive. Nevertheless, they provide a good characterization of process mining techniques to identify which visual analytics aspects they support.

5. Conclusions

Since to our best knowledge there exists no categorization schema for process mining techniques with focus on visual analytic aspects, we proposed a first approach how such techniques can be categorized. To show the applicability of our approach, we analyzed the most used ProM plug-ins. Our work aims at providing an orientation for the future development of such plug-ins. As future work we plan to investigate and compare also further categories. Moreover, such an overview can be used as a foundation for identifying visual analytics requirements for process mining techniques. Therefore, user studies will be necessary in order to verify how such categorizations can help for improving process mining technique plug-ins on the one hand and to identify requirements for the future development on the other hand.

6. Acknowledgments

Simone Kriglstein and Margit Pohl were supported by CVAST (funded by the Austrian Federal Ministry of Science, Research, and Economy in the exceptional Laura Bassi Centres of Excellence initiative, project nr: 822746).

© The Eurographics Association 2016.

References

[AA05] ANDRIENKO N., ANDRIENKO G.: *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, 2005. 2

[AES05] AMAR R., EAGAN J., STASKO J.: Low-level components of analytic activity in information visualization. In *Proc. of the IEEE Symposium on Information Visualization* (2005), INFOVIS '05, IEEE Computer Society. 2

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2376–2385. 2

[Cel16] CELONIS: Celonis process mining. http://www.celonis.de/en/solutions/processes Accessed: February, 2016, 2016. 1

[CP13] CLAES J., POELS G.: Process mining and the ProM framework: An exploratory survey. In *Business Process Management Workshops*, La Rosa M., Soffer P., (Eds.), vol. 132 of *Lecture Notes in Business Information Processing*. Springer, 2013, pp. 187–198. 2

[Flu12] FLUXICON PROCESS LABORATORIES: Fluxicon disco. https://fluxicon.com/disco/ Accessed: February, 2016, 2012. 1

[GZ09] GOTZ D., ZHOU M. X.: Characterizing users' visual analytic activity for insight provenance. *Information Visualization 8*, 1 (2009), 42–55. 2

[Kei02] KEIM D.: Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics 8*, 1 (Jan. 2002). 1, 2

[KKEM10] KEIM D., KOHLHAMMER J., ELLIS G., MANSMANN F. (Eds.): *Mastering The Information Age - Solving Problems with Visual Analytics*. Eurographics, November 2010. 1

[Lex16] LEXMARK INTERNATIONAL, INC.: Perceptive process mining. http://www.lexmark.com/en_us/products/software/workflow-and-case-management/process-mining.html Accessed: February, 2016, 2016. 1

[Maz09] MAZZA R.: *Introduction to Information Visualization*, 1st ed. Springer, 2009. 2

[Nor05] NORTH C.: *Handbook of Human Factors and Ergonomics*, 3rd ed. John Wiley & Sons, 2005. 2

[Pro09a] PROCESS MINING GROUP, MATH&CS DEPARTMENT, EINDHOVEN UNIVERSITY OF TECHNOLOGY: Dotted chart analysis. http://www.processmining.org/online/dottedchartanalysis Accessed: February, 2016, 2009. 4

[Pro09b] PROCESS MINING GROUP, MATH&CS DEPARTMENT, EINDHOVEN UNIVERSITY OF TECHNOLOGY: Fuzzy miner. http://www.processmining.org/online/fuzzyminer Accessed: February, 2016, 2009. 4

[Pro10a] PROCESS MINING GROUP, EINDHOVEN TECHNICAL UNIVERSITY: ProM. http://www.promtools.org/doku.php?id=start Accessed: February, 2016, 2010. 1

[Pro10b] PROCESS MINING GROUP, EINDHOVEN TECHNICAL UNIVERSITY: Prom 6.1. http://www.promtools.org/doku.php?id=prom61 Accessed: February, 2016, 2010. 2

[QPR16] QPR SOFTWARE PLC: Qpr processanalyzer. http://www.qpr.com/products/qpr-processanalyzer Accessed: February, 2016, 2016. 1

[SHB*14] SEDLMAIR M., HEINZL C., BRUCKNER S., PIRINGER H., MÖLLER T.: Visual parameter space analysis: A conceptual framework. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2161–2170. 2

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. of the IEEE Symposium on Visual Languages* (1996), VL'96, IEEE Computer Society. 2

[SNP16] SNP | THE TRANSFORMATION COMPANY: Snp business process analysis. http://products.snp-ag.com/business-process-analysis Accessed: February, 2016, 2016. 1

[vdA11] VAN DER AALST W. M. P.: *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1st ed. Springer, 2011. 1, 2

[vdAea11] VAN DER AALST W. M. P., ET AL.: Process mining manifesto. In *Business Process Management Workshops - BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I* (2011), pp. 169–194. 1

[WY04] WARD M., YANG J.: Interaction spaces in data and information visualization. In *Proc. of the Sixth Joint Eurographics - IEEE TCVG Conference on Visualization* (2004), VISSYM'04, Eurographics Association, pp. 137–146. 2

[YKSJ07] YI J. S., KANG Y. A., STASKO J., JACKO J.: Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (Nov. 2007), 1224–1231. 2

© The Eurographics Association 2016.