

Multi-perspective Process Mining

Felix Mannhardt

Department of Mathematics and Computer Science, Eindhoven University of
Technology

Abstract. Process mining methods analyze an organization's processes by using process execution data. During the handling of a process instance data about the execution of activities is recorded. Process mining uses such data to gain insights about the real execution of processes. In this thesis, we address research challenges in which a multi-perspective view on processes is needed and that look beyond the control-flow perspective, which defines the sequence of activities of a process. We consider problems in which multiple interacting process perspectives — in particular control-flow, data, resources, time, and functions — are considered together. The contributed methods span several types of process mining: two are concerned with conformance checking, two are process discovery techniques, and one is a decision mining method. All methods have been implemented, evaluated, and applied in the context of four case studies.

1 Introduction

The efficient and effective handling of its processes is essential for the success of an organization. This thesis [4] is about *process mining*: Analyzing the processes of an organization by using data recorded about their execution. Due to the growing computing power and storage capacity of today's IT systems, organizations have the opportunity to *store information about all their activities*. The amount of data being stored about process executions is rapidly growing. Process execution data can be seen as collection of log traces that contain at least: the timestamps of activity executions and the names or identifiers of the occurred activities. *Process mining* leverages such unbiased execution data to analyze the actual execution of processes [1]. Take for example the simplified process of a patient's trajectory through a hospital that is depicted in Figure 1. Execution data of such a process can be used to *discover* a process model suitable for analysis or to check *conformance* between prescribed behavior that has been modeled and the actual execution.

Often, process mining methods solely use the activity names and the timestamps of events recorded in execution traces. Other aspects of the process execution are then overlooked. This thesis contributes process mining techniques that make use of *additional data* to analyze a process from *multiple perspectives*. Typical examples for additional data that is considered in this thesis are identifiers of resources that execute an activity (e.g., humans, machines), input data used to execute an activity (e.g., patient age, loan amount), output data generated

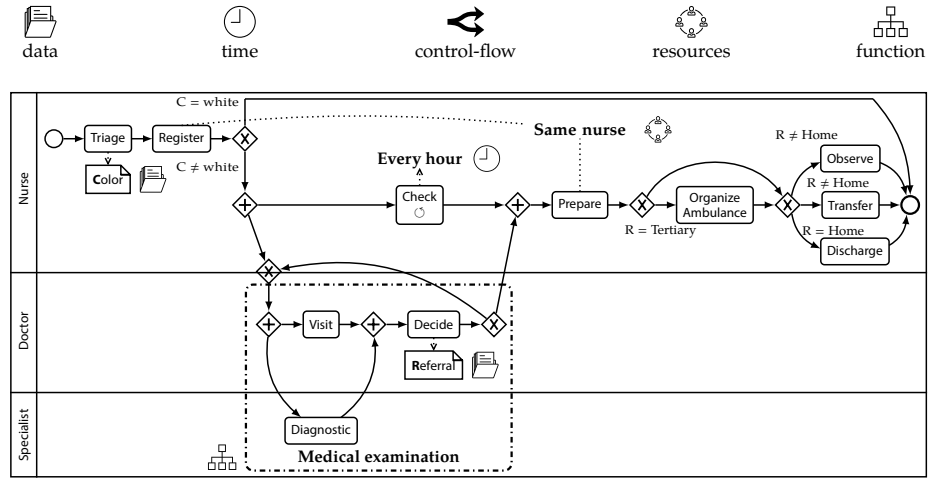


Fig. 1. Multiple perspectives on a process can be used for process mining. The control-flow perspective can be augmented with other perspectives such as resources, data, time and functional hierarchy.

by activity executions (e.g., decisions, outcomes), and information on the relation between multiple events (e.g., activity lifecycles). Process models are also not restricted to express only the control-flow of a process. Real-life activities rarely are atomic constructs. Often, there is a hierarchy of activities: multiple activities executed together form an activity on a higher level of abstraction. Decision rules based on data associated to the process instance and contextual information can be included (e.g., only certain patients require an ambulance). The five basic perspectives depicted in Figure 1 — the control-flow perspective, the resource perspective, the data perspective, the time perspective, and the function perspective — are often considered in the literature on BPM, process modeling, and process mining [1, 12] and are the basis for our contributions.

Our main research goal was to develop discovery, enhancement, and conformance checking methods that consider the *interaction of multiple perspectives on the process*. We aimed to advance the use of multi-perspective information for all three types of process mining instead of focusing on one specific type. Moreover, we targeted problems in which *multiple perspectives on a process are viewed together*, e.g., data objects that influence the routing of activities, routing that influences the possible resources, routing that depends on time constraints (e.g., fast vs. normal procedure). Starting from the premise that efficient, effective and usable tools are essential to facilitate the adoption of research results, we aimed for the development of tools that can deal with realistic event logs in an efficient and effective manner. Finally, we aimed to show the practical applicability of methods in real-world scenarios.

2 Contributions

We categorize our five main contributions along the three main types of process mining: conformance, enhancement, and discovery.

Conformance. We contribute two methods for multi-perspective conformance checking, i. e., the diagnosis and quantification of discrepancies between the real execution as recorded by information systems and the desired execution as specified by process models.

- A method that computes an optimal, multi-perspective, balanced alignment. The alignment relates the behavior modeled in a multi-perspective process model with the behavior observed in an event log and enables to determine a fitness score between model and log. We denoted the method as *balanced*, since it balances deviations on the different process perspectives and provides an optimal explanation for the observed behavior in terms of an execution trace of the multi-perspective process model. Deviations that occur on the control-flow perspective may be explained by wrongly recorded data values and vice versa. The technique enables to specify statements such as "Skipping activity *Check* is more severe than executing activity *Check* too late" and "Executing activity *Decide* by a different doctor than activity *Visit* is less severe than sending patients with the triage color *Red* to their home". The method has been published in [8] and is implemented in the *DataReplayer* package of ProM 6.7.
- A method to measure the precision of multi-perspective process models with regard to an event log. The precision of a process model can be seen as the fraction of the possible behavior allowed by the model in relation to what has actually been observed, as recorded in the event log. Our method is the *first proposal* to measure precision for multi-perspective process models and generalizes existing precision measures by taking the rules and data values of the multi-perspective process into account. Compared to the state-of-the-art our method is able to answer questions such as "What is the difference in precision between process model A with data rules and process model B without data rules?". The method has been published in [10] and is implemented in the *DataReplayer* package of ProM 6.7.

Discovery. We contribute two multi-perspective process discovery and one enhancement method. The proposed methods leverage the additional information recorded in data attributes (also denoted as event payload) of the event log or use domain knowledge on all process perspectives to discover better process models and enhance existing models. Our methods discover *integrated models* in which multiple perspectives on the process are *intertwined with the control-flow*.

- A method for data-aware heuristic process discovery that aims to reveal infrequent conditional behavior by using recorded data attributes. Data- and control-flow are learned together. The proposed method employs classification techniques to discover conditional dependencies based on the attribute

values recorded in the event log. It adds infrequent behavior to the process model such as, e.g., characterized by the following statements "In a few cases patients are assigned a white triage color and leave the hospital" and "Sometimes as a specific nurse reverses the order of the Diagnostics and Visit activity". The method has been published in [2] and is implemented in the *DataAwareCNetMiner* package of ProM 6.7.

- The Guided Process Discovery (GPD) method discovers a mapping between occurrences of low-level events and high-level activities instances of the process (i.e., functional perspective) in order to improve the quality of existing process discovery methods. The method uses *multi-perspective activity patterns* to specify domain knowledge on the function perspective of the process. Activity patterns encode the assumptions on how high-level activities of the process manifest themselves in terms of recorded low-level events. An optimal mapping between all activity patterns and the low-level event log is established through an alignment. Here, we compute the alignment not for diagnostic purposes but to create an abstracted event log. Based on this abstracted event log, we discover a high-level process model that can be validated on the low-level log using an model expansion step. Using GDP can lead to a considerable improvements in the model quality as perceived by stakeholders. The method has been published in [3, 11] and is implemented in the *DataAwareCNetMiner* package of ProM 6.7.

Enhancement. Regarding the enhancement category of process mining, we contribute a method to discover potentially overlapping decision rules in process models based on an event log. Existing techniques only return rules that assume completely deterministic decisions. We observed that this assumption often does not hold due to missing data relevant for the actual decision making is unavailable or non-deterministic business rules. The method builds upon standard classification techniques and makes an effort to introduce overlap by reclassifying instances that were previously misclassified. The method balances precision and fitness of a process model with regard to an event log. When rules are overlapping two or more possible routing options can be chosen non-deterministically. As result, the process model is less precise but fits the observations better.

Implementation and Applications. Next to their implementation in the open source framework ProM, we integrated the functionality of the proposed methods in two interactive tools: the *Multi-perspective Process Explorer* and the *Interactive Data-aware Heuristic Miner*. Both tools reached a high level of maturity and were published in the BPM demo track [6, 7]. We applied all proposed methods in four case studies conducted in several organizations (e.g., [5]). For each case study, we obtained real-life event data, identified relevant process questions, and showed that the application of our methods is *feasible* and provides *valuable insights*.

3 Conclusion

Leveraging knowledge from such recorded data is widely acknowledged to be an important challenge. Process mining is part of this trend towards organizations that are driven by data. Process mining methods, such as our contributions, operate on event logs that contain traces recorded from the execution of a process. There are many potential benefits by making decisions about the design and optimization of organizational processes more evidence-based, i.e., based on the actual execution of processes as recorded in event logs rather than based on assumptions and feelings of stakeholders. In the light of this, our contributions can be used to get more reliable diagnostics about the process from data ([8, 10]) and to discover more understandable (i.e., structured according to domain knowledge [3, 11]), complete (i.e., including potentially interesting infrequent process behavior [2]) and balanced (i.e., between a fitting and precise model [9]) process models from data.

References

1. van der Aalst, W.M.P.: *Process Mining - Data Science in Action*, Second Edition. Springer (2016)
2. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Data-driven process discovery - revealing conditional infrequent behavior from event logs. In: CAiSE 2017. LNCS, vol. 10253, pp. 545–560 (2017)
3. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P., Toussaint, P.J.: Guided process discovery - A pattern-based approach. *Inf Syst* **76**, 1–18 (2018)
4. Mannhardt, F.: *Multi-perspective Process Mining*. Ph.D. thesis, Eindhoven University of Technology (2018)
5. Mannhardt, F., Blinde, D.: Analyzing the trajectories of patients with sepsis using process mining. In: RADAR+EMISA 2017. CEUR Workshop Proceedings, vol. 1859, pp. 72–80. CEUR-WS.org (2017)
6. Mannhardt, F., de Leoni, M., Reijers, H.A.: The multi-perspective process explorer. In: Daniel, F., Zugal, S. (eds.) BPM 2015 Demos. CEUR Workshop Proceedings, vol. 1418, pp. 130–134. CEUR-WS.org (2015)
7. Mannhardt, F., de Leoni, M., Reijers, H.A.: Heuristic mining revamped: An interactive data-aware and conformance-aware miner. In: BPM 2017 Demos. CEUR Workshop Proceedings, CEUR-WS.org (2017)
8. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Balanced multi-perspective checking of process conformance. *Computing* **98**(4), 407–437 (2016)
9. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Decision mining revisited - discovering overlapping rules. In: CAiSE 2016. LNCS, vol. 9694, pp. 377–392. Springer (2016)
10. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Measuring the precision of multi-perspective process models. In: BPM 2015 Workshops. LNBIP, vol. 256, pp. 113–125. Springer (2016)
11. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P., Toussaint, P.J.: From low-level events to activities - A pattern-based approach. In: BPM 2016. LNCS, vol. 9850, pp. 125–141. Springer (2016)
12. Scheer, A.W.: *ARIS — Business Process Modeling*. Springer (2000)