

From Zero to Hero: A Process Mining Tutorial

Andrea Janes¹, Fabrizio Maria Maggi², Andrea Marrella³, and Marco Montali⁴

¹ Free University of Bozen-Bolzano, Bolzano, Italy, andrea.janes@unibz.it

² University of Tartu, Estonia f.m.maggi@ut.ee

³ Sapienza University, Rome, Italy, marrella@diag.uniroma1.it

⁴ Free University of Bozen-Bolzano, Bolzano, Italy, marco.montali@unibz.it

Abstract. Process mining is an emerging area that synergically combines model-based and data-oriented analysis techniques to obtain useful insights on how business processes are executed within an organization. This tutorial aims at providing an introduction to the key analysis techniques in process mining that allow decision makers to discover process models from data, compare expected and actual behaviors, and enrich models with key information about the actual process executions. In addition, the tutorial will present concrete tools and will provide practical skills for applying process mining in a variety of application domains, including the one of software development.

1 Introduction

Process mining [2] is a recent research discipline that sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. Through process mining, decision makers can discover process models from data, compare expected and actual behaviors, and enrich models with information retrieved from data. This, in turn, provides the basis for understanding, maintaining, and enhancing processes based on reality.

Since process mining has many applications, this tutorial aims at encouraging participants to apply it in new fields, in which it has not been applied so far. In particular, first, we will introduce the process mining framework, the main process mining techniques and tools, and the different phases of event data analysis through process mining. Second, we will carry out an hands-on session using concrete process mining tools, considering business use cases, as well as the particular scenario of software processes. Finally, we will discuss common pitfalls and critical issues and will give suggestions on how to mitigate them.

2 Process Mining Framework

The reference framework for process mining is depicted in Figure 1. On the one hand, process mining considers conceptual models describing processes, organizational structures, and the corresponding relevant data. On the other hand, it focuses on the real execution of processes, as reflected by the footprint of reality logged and stored by the software systems in use within an enterprise. For

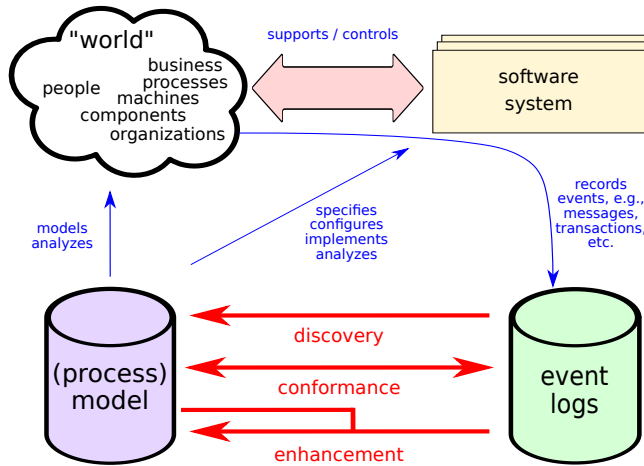


Fig. 1: The reference framework for process mining [1]

process mining to be applicable, such information has to be structured in the form of explicit *event logs*. In fact, all process mining techniques assume that it is possible to record the sequencing of relevant events occurred within an enterprise, such that each event refers to an activity (i.e., a well-defined step in some process) and is related to a particular case [1]. Events may have additional information stored in event logs. In fact, whenever possible, process mining techniques use extra information such as the exact timestamp at which the event has been recorded, the resource (i.e., person or device) that generated the event, the event type in the context of the activity transactional lifecycle (e.g., whether the activity has been started, canceled, or completed), or data elements recorded with the event (e.g., the size of an order).

The three main types of process mining techniques are marked by the three, thick red arrows in the bottom part of Figure 1. *Discovery* starts from an event log and automatically produces a process model that explains the different behaviors observed in the log, without assuming any prior knowledge on the process. The vast majority of process discovery algorithms focus on the discovery of the process control-flow, towards generating a model that indicates the allowed sequences of activities according to the log. *Conformance checking* compares an existing process model and an event log for the same process, with the aim of understanding the presence and nature of *deviations*. Conformance checking techniques take as inputs an event log and a process model, and return indications related to the adherence of the behaviors contained in the log to the prescriptions contained in the model. *Enhancement* improves an existing process model using information recorded in an event log for that process. The inputs of enhancement techniques are a process model and an event log, and the output is a new process model that incorporates and reflects new information extracted from the data. The first important class of enhancement techniques is

that of *extension*, where the input process model is not altered in its structure, but is extended with additional perspectives, using information present in the log. A second important class of enhancement techniques is that of *repair*, where deviations detected by checking the conformance of the input event log to the input process model are resolved by suitably modifying the process model.

The presented tutorial will demonstrate that process mining *can* also be used in the context of software development [9], trying to focus on aspects that occur with a certain regularity and with a finite number of possible activities so that repeating cases within an actual “process” can be discovered. One example will be the *user behavior analysis* [9], i.e., the study on how the users interact with software applications. In particular, we will study an example that can be found in [3], in which the authors applied process mining to study and to improve the user interface of an ERP software.

2.1 Process Mining Tools

A plethora of process mining techniques and technologies have been developed and successfully employed in several application domains⁵. The process mining solutions that will be used in the tutorial are *ProM* (Process Mining framework)⁶ and *Disco*⁷.

- *ProM* (Process Mining framework) is an Open Source framework for process mining algorithms [10] based on JAVA. It provides a plug-in based integration platform [4] that users and developers can exploit to run and deploy process techniques. This pluggable architecture currently hosts a very large amount of plug-ins covering all the different aspects of process mining, from data import to discovery, conformance checking, and enhancement along different perspectives [2].
- *Disco* is a commercial, stand-alone and lightweight process mining tool. It supports various file formats as input, in particular providing native support for importing CSV files, which can be annotated with case and event information prior to the import. Disco has usability, fidelity, and performance as design priorities, and makes process mining easy and fast [7].

2.2 The XES Standard

In recent years, the *XES* (*eXtensible Event Stream*) format emerged as the main reference format for the storage, interchange, and analysis of event logs. XES, which is based on XML, appeared for the first time in 2009 [6] as the successor of the MXML format [5]. It quickly became the de-facto standard adopted by the IEEE Task Force on Process Mining⁸, eventually becoming an official IEEE standard in 2016 [8]. This standard will be introduced in the tutorial.

⁵ <http://tinyurl.com/ovedwx4>

⁶ <http://www.processmining.org/prom/>

⁷ <https://fluxicon.com/disco/>

⁸ <http://www.win.tue.nl/ieeetfpm/doku.php>

3 Agenda

The tutorial is structured as follows:

- Introduction to the process mining framework
- Designing process models and collecting event logs
 - Process modeling: basics of Petri nets and BPMN
 - The XES standard and the OpenXES reference implementation
- Mining event logs, discovering processes
 - ProM, the open-source jack of all trades
 - Choosing a mining algorithm: Alpha, Heuristics miner, Fuzzy miner, or Multi-phase miner?
 - Disco, the user friendly tool
- Interpreting the mined models: discovery, conformance checking, and enhancement
 - Deciding from which perspective to look at the data: choosing a case ID
 - On the representational bias of process mining
- Hands-on session
 - Process mining success stories for business processes and software processes
 - Walk-through of two process mining examples examining a business process and a software process, discussing strategies to collect, filter, analyze, and interpret data
 - Discussion on the differences between mining business and software processes

References

1. van der Aalst, W., et al.: Process Mining Manifesto. In: Proc. of BPM Int. Workshops. vol. 99, pp. 169–194 (2012)
2. van der Aalst, W.M.P.: Process Mining - Data Science in Action. 2nd edn. (2016)
3. Astromskis, S., Janes, A., Mairegger, M.: A Process Mining Approach to Measure How Users Interact with Software: An Industrial Case Study. In: ICSSP '15 (2015)
4. van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M., van der Aalst, W.M.P.: The ProM Framework: A New Era in Process Mining Tool Support, pp. 444–454 (2005)
5. van Dongen, B.F., van der Aalst, W.M.P.: A meta model for process mining data. In: Proc. of EMOI - INTEROP. vol. 160. CEUR-WS.org (2005)
6. Günther, C.W.: XES Standard Definition Version 1.0. Tech. rep., Fluxicon Process Laboratories (Nov 2009), available at <http://www.xes-standard.org>
7. Günther, C.W., Rozinat, A.: Disco: Discover Your Processes. In: Proc. of the Demo Track of BPM. vol. 940, pp. 40–44 (2012)
8. IEEE Computational Intelligence Society: IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams. IEEE Std 1849-2016 pp. i – 50 (2016)
9. Rubin, V.A., Mitsyuk, A.A., Lomazova, I.A., van der Aalst, W.M.P.: Process Mining Can Be Applied to Software Tool. In: Proc. of the ESEM '14. ACM (2014)
10. Verbeek, H.M.W., Buijs, J.C.A.M., van Dongen, B.F., van der Aalst, W.M.P.: XES, XESame, and ProM 6. In: Proc. of the CAiSE Forum. vol. 72, pp. 60–75 (2010)