Wil M. P. van der Aalst
Josep Carmona (Eds.)

Tutorial

LNBIP 448

Process Mining Handbook

Springer

OPEN ACCESS

Editors
Wil M. P. van der Aalst
RWTH Aachen
Aachen, Germany

Josep Carmona
Universitat Politècnica de Catalunya
Barcelona, Spain

ISSN 1865-1348 ISSN 1865-1356 (electronic)
Lecture Notes in Business Information Processing
ISBN 978-3-031-08847-6 ISBN 978-3-031-08848-3 (eBook)
https://doi.org/10.1007/978-3-031-08848-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Process mining emerged as a new discipline around the turn of the century. The combination of event data and process models poses interesting scientific problems. Initially, the focus was on the discovery of process models (e.g., Petri nets) from example traces. However, over time the scope of process mining broadened in several directions. Next to process discovery, topics such as conformance checking and performance analysis were added. Different perspectives were added (e.g., time, resources, roles, costs, and case types) to move beyond control-flow models. Along with directly-follows graph (DFGs) and Petri nets, a wide range of process model notations has been explored in the context of event data. Examples include declarative process models, process trees, artifact-centric and object-centric process models, UML activity models, and BPMN models. In recent years, the focus also shifted from backward-looking to forward-looking, connecting process mining to neighboring disciplines such as simulation, machine learning, and automation.

Over the past two decades, the discipline did not only expand in terms of scope but also in terms of adoption and tool support. The first commercial process mining tools emerged 15 years ago (Futura Process Intelligence, Disco, etc.). Now there are over 40 commercial products next to open-source process mining tools such as ProM, PM4Py, and bupaR. The adoption in industry has accelerated in the last five years. In several regions of the world, most of the larger companies are already using process mining, and the process mining market is expected to double every 18 months in the coming years.

Given the amazing developments in the last two decades, a comprehensive process mining summer school is long overdue. This book contains the core material of the first Summer School on Process Mining organized by the IEEE Task Force on Process Mining. The Task Force on Process Mining was established in October 2009 as part of the IEEE Computational Intelligence Society. Its activities led to the International Process Mining Conference (ICPM) series, a range of successful workshops (BPI, ATAED, PODS4H, etc.), the Process Mining Manifesto (translated into 15+ languages), the XES standard, publicly available datasets, online courses, and case studies. However, a dedicated summer school on process mining was missing. Therefore, we started the preparations for this in 2020. Due to the COVID-19 pandemic, this was delayed by one year, but this gave us more time to carefully prepare this handbook on process mining.

The summer school took place in Aachen, Germany, during July 4–8, 2022. The location of the summer school was the scenic SuperC building with nice views of the city center and close to the cathedral of Aachen, which was the first UNESCO World Heritage site in Germany.

The local organization was undertaken by the Process and Data Science (PADS) group at RWTH Aachen University. The event was financially supported by Wil M. P. van der Aalst's Alexander von Humboldt (AvH) professorship. The event was also supported by the RWTH Center for Artificial Intelligence, the Center of Excellence Internet of Production (IoP), Celonis, and Springer.

The book starts with a 360-degree overview of the field of process mining (Chapter 1). This first chapter introduces the basic concepts, the different types of process mining, process modeling notations, and storage formats for events.

Chapter 2 presents the foundations of process discovery. It starts with discovering directly-follows graphs from simple event logs and highlighting the challenges. Then basic bottom-up and top-down process discovery techniques are presented that produce Petri nets and BPMN models.

Chapter 3 presents four additional process discovery techniques: an approach based on state-based regions, an approach based on language-based regions, the split mining approach, and the log skeleton-based approach.

Techniques to discover declarative process models are presented in Chapter 4. The chapter focuses on discovering declarative specifications from event logs, monitoring declarative specifications against running process executions to promptly detect violations, and reasoning on declarative process specifications.

Chapter 5 presents techniques for conformance checking. An overview of the applications of conformance checking and a general framework are presented. The goal is to compare modeled and observed behavior.

Chapter 6 discusses event data in more detail, also describing the data-preprocessing pipeline, standards like XES, and data quality problems.

Chapter 7 takes a more applied view and discusses how process mining is used in different industries and the efforts involved in creating an event log. The chapter also lists best practices, illustrated using the order-to-cash (O2C) process in an SAP system.

Chapter 8 introduces a number of techniques for process enhancement, including process extension and process improvement. For example, it is shown how to add additional perspectives to a process model.

Chapter 9 introduces event knowledge graphs as a means to model multiple entities distributed over different perspectives. It is shown how to construct, query, and aggregate event knowledge graphs to get insights into complex behaviors.

Predictive process monitoring techniques are introduced in Chapter 10. This is the branch of process mining that aims at predicting the future of ongoing (uncompleted) process executions.

Streaming process mining refers to the set of techniques and tools which have the goal of processing a stream of data (as opposed to a fixed event log). Chapter 11 presents such techniques.

The topic of responsible process mining is addressed in Chapter 12. The chapter summarizes and discusses current approaches that aim to make process mining responsible by design, using the well-known FACT criteria (Fairness, Accuracy, Confidentiality, and Transparency).

Chapter 13 discusses the evolution of the field of process mining, i.e., the transition from process discovery to process execution management. The focus is on driving business value.

Chapter 14 makes the case that healthcare is a very promising application domain for process mining with a great societal value. An overview of healthcare processes and healthcare process data is given, followed by a discussion of common use cases.

Chapter 15 shows that process mining is a valuable tool for financial auditing. Both internal and external audits are introduced, along with the connection between the two audits and the application of process mining.

Chapter 16 introduces a family of techniques, called robotic process mining, that discover repetitive routines that can be automated using robotic process automation (RPA) technology.

Chapter 17 concludes the book with an analysis of the current state of the process mining discipline and outlook on future developments and challenges. Pointers to the lecture material will be made available via www.process-mining-summer-school.org, www.processmining.org, and www.tf-pm.org. These complement this book.

Finally, we thank all the participants, authors, speakers, and the organizations supporting this once-in-a-lifetime event. In particular, we thank the Alexander von Humboldt Foundation. Enjoy reading!

April 2022

Wil M. P. van der Aalst
Josep Carmona

Contents