# Preface

This book encompasses a revised version of the PhD dissertation of Jorge Munoz-Gama written at the Computer Science Department of the Universitat Politècnica de Catalunya (Spain). In 2015, the dissertation won the "Best Process Mining Dissertation Award," assigned by the IEEE Task Force on Process Mining to the most outstanding PhD thesis, discussed between 2013 and 2014, focused on the area of business process intelligence.

In the past few decades, the capability of information systems to generate and record overwhelming amounts of event data has witnessed an exponential growth in several domains, and in particular in industrial scenarios. Devices connected to the Internet (Internet of Things), social interaction, mobile computing, and cloud computing provide new sources of event data and this trend will continue in the next decades. The omnipresence of large amounts of event data stored in logs is an important enabler for process mining, a novel discipline for addressing challenges related to business process management, process modeling, and business intelligence. Process mining techniques can be used to discover, analyze, and improve real processes, by extracting models from observed behavior. The capability of these models to represent the reality determines the quality of the results obtained from them.

The aim of this book is conformance checking, one of the main areas of process mining. In conformance checking, existing process models are compared with actual observations of the process in order to assess their quality. These models are typically the result of a hand-made analysis influenced by the bias of the analysts and the process owners, modeling a possibly outdated representation of the process. Conformance checking techniques are a way to visualize the differences between the assumed process represented in the model and the real process in the event log, pinpointing possible problems to be addressed, and the business process management results that rely on these models.

Conformance checking is a complex multidimensional analysis, where orthogonal dimensions such as fitness (measuring and ensuring that models capture all the behavior in the log) and precision (not including unnecessary behavior) determine the quality of the models. Moreover, a conformance analysis of real-life processes may overcome additional challenges such as the presence of noise or the size of the models.

The first part of the book focuses on analyzing and measuring the precision dimension of conformance, where models describing precisely the reality are preferred to overly general models. The book includes a novel technique based on detecting *escaping arcs*, i.e., points where the modeled behavior deviates from the one in the log. The detected escaping arcs are used to determine the precision between log and model, and to locate possible actuation points in order to achieve a more precise model. The book also presents a confidence interval on the provided precision metric, and a multi-factor measure to assess the severity of the detected imprecisions. These techniques open the door to noise-robust analysis of real-life processes and the possibility

of ranking the misconformances detected regarding the importance and potential impact in the process.

Checking conformance can be time consuming for real-life scenarios, and understanding the reasons behind the conformance mismatches can be an effort-demanding task. The second part of the book changes the focus from the precision dimension to the fitness dimension, and proposes the use of decomposed techniques to aid in checking and diagnosing fitness. The proposed approach is based on decomposing the model in single-entry single-exit components. The resulting fragments represent subprocesses within the main process with a simple interface with the rest of the model. Fitness checking per component provides well-localized conformance information, aiding the causes behind the mismatches. Moreover, the relations between components can be explored to improve the diagnosis capabilities of the analysis, identifying areas with a high degree of mismatches, or providing a hierarchy for a zoom-in zoom-out analysis.

Finally, the book proposes two main applications of the decomposed approach. First, the theory is extended to incorporate data information for fitness checking in a decomposed manner. Second, a real-time event-based framework is presented for monitoring fitness in an on-line setting.

This book combines both application and research perspectives. It provides concrete use cases that illustrate the problems addressed by the techniques in the book, but at the same time, it contains complete conceptualization and formalization of the problem and the techniques, and through evaluations on the quality and the performance of the proposed techniques. Hence, this book is aimed at business analysts willing to improve their organization processes, and also data scientists interested in the topic of process-oriented data science.

April 2016 Jorge Munoz-Gama

Springer

http://www.springer.com/978-3-319-49450-0

Conformance Checking and Diagnosis in Process Mining
Comparing Observed and Modeled Processes
Munoz-Gama, J.
2016, XIV, 202 p. 90 illus., Softcover
ISBN: 978-3-319-49450-0