Process Mining

Wil M.P. van der Aalst

Process Mining

Discovery, Conformance and Enhancement of Business Processes

Springer

Wil M.P. van der Aalst
Department Mathematics & Computer Science
Eindhoven University of Technology
Den Dolech 2
5612 AZ Eindhoven
The Netherlands
w.m.p.v.d.aalst@tue.nl

ISBN 978-3-642-19344-6 e-ISBN 978-3-642-19345-3
DOI 10.1007/978-3-642-19345-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011926240

ACM Computing Classification (1998): H.4.1, H.2.8, I.2.6, F.3.2, D.2.2, J.1

Cover design: deblik

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Thanks to Karin for understanding that science is more rewarding than running errands

Thanks to all people that contributed to ProM; the fruits of their efforts demonstrate that sharing a common goal is more meaningful than "cashing in the next publon"[1]

In remembrance of Gerry Straatman-Beelen (1932–2010)

[1] publon = smallest publishable unit

Preface

Process mining provides a new means to improve processes in a variety of application domains. There are two main drivers for this new technology. On the one hand, more and more events are being recorded thus providing detailed information about the history of processes. Despite the omnipresence of event data, most organizations diagnose problems based on fiction rather than facts. On the other hand, vendors of Business Process Management (BPM) and Business Intelligence (BI) software have been promising miracles. Although BPM and BI technologies received lots of attention, they did not live up to the expectations raised by academics, consultants, and software vendors.

Process mining is an emerging discipline providing comprehensive sets of tools to provide fact-based insights and to support process improvements. This new discipline builds on process model-driven approaches and data mining. However, process mining is much more than an amalgamation of existing approaches. For example, existing data mining techniques are too data-centric to provide a comprehensive understanding of the end-to-end processes in an organization. BI tools focus on simple dashboards and reporting rather than clear-cut business process insights. BPM suites heavily rely on experts modeling idealized to-be processes and do not help the stakeholders to understand the as-is processes.

This book presents a range of process mining techniques that help organizations to uncover their actual business processes. Process mining is not limited to process discovery. By tightly coupling event data and process models, it is possible to check conformance, detect deviations, predict delays, support decision making, and recommend process redesigns. Process mining breathes life into otherwise static process models and puts today's massive data volumes in a process context. Hence, managements trends related to process improvement (e.g., Six Sigma, TQM, CPI, and CPM) and compliance (SOX, BAM, etc.) can benefit from process mining.

Process mining, as described in this book, emerged in the last decade [102, 106]. However, the roots date back about half a century. For example, Anil Nerode presented an approach to synthesize finite-state machines from example traces in 1958 [71], Carl Adam Petri introduced the first modeling language adequately capturing concurrency in 1962 [73], and Mark Gold was the first to systematically explore

different notions of learnability in 1967 [45]. When data mining started to flourish in the nineties, little attention was given to processes. Moreover, only recently event logs have become omnipresent thus enabling end-to-end process discovery. Since the first survey on process mining in 2003 [102], progress has been spectacular. Process mining techniques have become mature and supported by various tools. Moreover, whereas initially the primary focus was on process discovery, the process mining spectrum has broadened markedly. For instance, conformance checking, multi-perspective process mining, and operational support have become integral parts of ProM, one of the leading process mining tools.

This is the first book on process mining. Therefore, the intended audience is quite broad. The book provides a comprehensive overview of the state-of-the-art in process mining. It is intended as an introduction to the topic for practitioners, students, and academics. On the one hand, the book is accessible for people that are new to the topic. On the other hand, the book does not avoid explaining important concepts on a rigorous manner. The book aims to be self-contained while covering the entire process mining spectrum from process discovery to operational support. Therefore, it also serves as a reference handbook for people dealing with BPM or BI on a day-to-day basis.

The reader can immediately put process mining into practice due to the applicability of the techniques, the availability of (open-source) process mining software, and the abundance of event data in today's information systems. I sincerely hope that you enjoy reading this book and start using some of the amazing process mining techniques available today.

Schleiden, Germany Wil M.P. van der Aalst
December 2010

Acknowledgements

Many individuals and organizations contributed to the techniques and tools described in this book. Therefore, it is important to acknowledge their support, efforts, and contributions.

All of this started in 1999 with a research project named "Process Design by Discovery: Harvesting Workflow Knowledge from Ad-hoc Executions" initiated by Ton Weijters and myself. At that time, I was still working as a visiting professor at the University of Colorado in Boulder. However, the research school BETA had encouraged me to start collaborating with existing staff in my new research group at TU/e (Eindhoven University of Technology). After talking to Ton it was clear that we could benefit from combining his knowledge of machine learning with my knowledge of workflow management and Petri nets. Process mining (at that time we called it workflow mining) was the obvious topic for which we could combine our expertise. This was the start of a very successful collaboration. Thanks Ton!

Since then many PhD students have been working on the topic: Laura Maruster, Ana Karla Alves de Medeiros, Boudewijn van Dongen, Minseok Song, Christian Günther, Anne Rozinat, Carmen Bratosin, R.P. Jagadeesh Chandra (JC) Bose, Ronny Mans, Maja Pesic, Joyce Nakatumba, Helen Schonenberg, Arya Adriansyah, and Joos Buijs. I'm extremely grateful for their efforts.

Ana Karla Alves de Medeiros was the first PhD student to work on the topic under my supervision (genetic process mining). She did a wonderful job; her thesis on genetic process mining was awarded with the prestigious ASML 2007 Promotion Prize and was selected as the best thesis by the KNAW research school BETA. Also Boudewijn van Dongen has been involved in the development of ProM right from the start. As a Master student he already developed the process mining tool EMiT, i.e., the predecessor of ProM. He turned out to be a brilliant PhD student and developed a variety of process mining techniques. Eric Verbeek did a PhD on workflow verification, but over time he got more and more involved in process mining research and the development of ProM. Many people underestimate the importance of a scientific programmer like Eric. Tool development and continuity are essential for scientific progress! Boudewijn and Eric have been the driving force behind

ProM and their contributions have been crucial for process mining research at TU/e. Moreover, they are always willing to help others. Thanks guys!

Christian Günther and Anne Rozinat joined the team in 2005. Their contributions have been of crucial importance for extending the scope of process mining and lifting the ambition level. Christian managed to make ProM look beautiful while significantly improving its performance. Moreover, his Fuzzy miner facilitated dealing with Spaghetti processes. Anne managed to widen the process mining spectrum by adding conformance checking and multi-perspective process mining to ProM. It is great that they succeeded in founding a process mining company (Fluxicon). Another person crucial for the development of ProM is Peter van den Brand. He set up the initial framework and played an important role in the development of the architecture of ProM 6. Based on his experiences with ProM, he set up a process mining company (Futura Process Intelligence). It is great to work with people like Peter, Christian, and Anne; they are essential for turning research results into commercial products. I sincerely hope that Fluxicon and Futura Process Intelligence continue to be successful (not only because of prospective sports cars ...).

Academics of various universities contributed to ProM and supported our process mining research. We are grateful to the Technical University of Lisbon, Katholieke Universiteit Leuven, Universitat Politècnica de Catalunya, Universität Paderborn, University of Rostock, Humboldt-Universität zu Berlin, University of Calabria, Queensland University of Technology, Tsinghua University, Universität Innsbruck, Ulsan National Institute of Science and Technology, Università di Bologna, Zhejiang University, Vienna University of Technology, Universität Ulm, Open University, Jilin University, University of Padua, and University of Nancy for their help. I would also like to thank the members of the IEEE Task Force on Process Mining for promoting the topic. We are grateful to all other organizations that supported process mining research at TU/e: NWO, STW, EU, IOP, LOIS, BETA, SIKS, Stichting EIT Informatica Onderwijs, Pallas Athena, IBM, LaQuSo, Philips Healthcare, ESI, Jacquard, Nuffic, BPM Usergroup, and WWTF. Special thanks go to Pallas Athena for promoting the topic of process mining and their collaboration in a variety of projects. More than 100 organizations provided event logs that helped us to improve our process mining techniques. Here, I would like to explicitly mention the AMC hospital, Philips Healthcare, ASML, Ricoh, Vestia, Catharina hospital, Thales, Océ, Rijkswaterstaat, Heusden, Harderwijk, Deloitte, and all organizations involved in the SUPER, ACSI, PoSecCo, and CoSeLoG projects. We are grateful for allowing us to use their data and for providing feedback.

It is impossible to name all of the individuals that contributed to ProM or helped to advance process mining. Nevertheless, I would like to make a modest attempt. Besides the people mentioned earlier, I would like to thank Piet Bakker, Huub de Beer, Tobias Blickle, Andrea Burattin, Riet van Buul, Toon Calders, Jorge Cardoso, Josep Carmona, Alina Chipaila, Francisco Curbera, Marlon Dumas, Schahram Dustdar, Paul Eertink, Dyon Egberts, Dirk Fahland, Diogo Ferreira, Walid Gaaloul, Stijn Goedertier, Adela Grando, Gianluigi Greco, Dolf Grünbauer, Antonella Guzzo, Kees van Hee, Joachim Herbst, Arthur ter Hofstede, John Hoogland, Ivo de Jong, Ivan Khodyrev, Thom Langerwerf, Massimiliano de Leoni, Jiafei Li, Ine van der

Ligt, Zheng Liu, Niels Lohmann, Peter Hornix, Fabrizio Maggi, Jan Mendling, Frits Minderhoud, Arnold Moleman, Marco Montali, Michael zur Muehlen, Jorge Munoz-Gama, Mariska Netjes, Andriy Nikolov, Mykola Pechenizkiy, Carlos Pedrinaci, Viara Popova, Silvana Quaglini, Manfred Reichert, Hajo Reijers, Remmert Remmerts de Vries, Stefanie Rinderle-Ma, Marcello La Rosa, Michael Rosemann, Vladimir Rubin, Stefania Rusu, Eduardo Portela Santos, Natalia Sidorova, Alessandro Sperduti, Christian Stahl, Keith Swenson, Nikola Trcka, Kenny van Uden, Irene Vanderfeesten, George Varvaressos, Marc Verdonk, Sicco Verwer, Jan Vogelaar, Hans Vrins, Jianmin Wang, Teun Wagemakers, Barbara Weber, Lijie Wen, Jan Martijn van der Werf, Mathias Weske, Michael Westergaard, Moe Wynn, Bart Ydo, and Marco Zapletal for their support. Thanks to all that read earlier drafts of this book (special thanks go to Christian, Eric, and Ton for their detailed comments).

Thanks to Springer-Verlag for publishing this book. Ralf Gerstner encouraged me to write this book and handled things in a truly excellent manner. Thanks Ralf!

More than 95% of book was written in beautiful Schleiden. Despite my sabbatical, there were many other tasks competing for attention. Thanks to my weekly visits to Schleiden (without Internet access!), it was possible to write this book in a three month period. The excellent coffee of Serafin helped when proofreading the individual chapters, the scenery did the rest.

As always, acknowledgements end with thanking the people most precious. Lion's share of credits should go to Karin, Anne, Willem, Sjaak, and Loes. They often had to manage without me under difficult circumstances. Without their continuing support, this book would have taken ages.

Schleiden, Germany Wil M.P. van der Aalst
December 2010

Contents