ProM: The Process Mining Toolkit

W.M.P. van der Aalst[1,2], B.F. van Dongen[1], C. Günther[1], A. Rozinat[2], H.M.W. Verbeek[1], and A.J.M.M. Weijters[2]

[1] Department of Mathematics and Computer Science,
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
{w.m.p.v.d.aalst,b.f.v.dongen,c.guenther,h.m.w.verbeek}@tue.nl
[2] Department of Industrial Engineering and Innovation Sciences,
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
{a.rozinat,a.j.m.m.weijters}@tue.nl

Abstract. Nowadays, all kinds of information systems store detailed information in logs. Process mining has emerged as a way to analyze these systems based on these detailed logs. Unlike classical data mining, the focus of process mining is on processes. First, process mining allows us to extract a process model from an event log. Second, it allows us to detect discrepancies between a modeled process (as it was envisioned to be) and an event log (as it actually is). Third, it can enrich an existing model with knowledge derived from an event log. This paper presents our tool ProM, which is the world-leading tool in the area of process mining.

1 Process Mining

The goal of process mining is to extract information (like process models) from event logs. Typically, process mining assumes that it is possible to record events such that each event refers to an activity (a step in the process) and is related to a particular case (a process instance). Furthermore, additional data stored in the log (like the performer of the event, the timestamp of the event, or data elements recorded with the event) can be used.

The omnipresence of event logs is an important enabler of process mining: Analysis of run-time behavior is only possible if events are recorded. Fortunately, all kinds of information systems provide the necessary detailed logs, like classical workflow management systems (Staffware), ERP systems (SAP), case handling systems (FLOWer), PDM systems (Windchill), CRM systems (Microsoft Dynamics CRM), middleware (IBM WebSphere), and hospital information systems (Chipsoft). Also, all kinds of embedded systems increasingly log events, like medical systems (X-ray machines), mobile phones, car entertainment systems, production systems (e.g., wafer steppers), copiers, and sensor networks.

Process mining has emerged as a way to analyze systems and their actual use based on the event logs they produce [1,2,3,4,5,8,9]. Unlike classical data

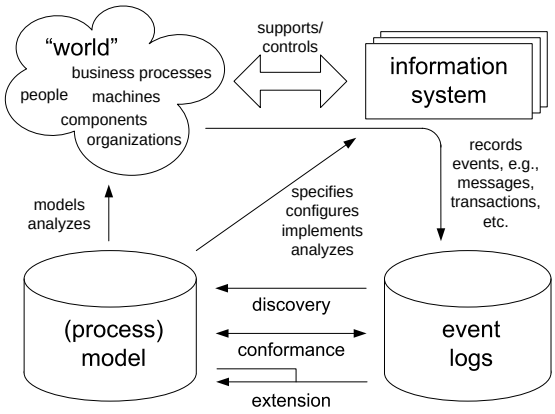

Fig. 1. Process mining aims at extracting knowledge from event logs.

mining, the focus of process mining is on concurrent processes instead of on static or mainly sequential structures. Note that commercial "Business Intelligence" (BI) tools are not doing any process mining: They typically look at aggregate data (frequencies, averages, utilization, service levels). Unlike BI tools, process mining looks "inside the process" (causal dependencies, bottlenecks) and at a very refined level. In a hospital context, BI tools focus on performance indicators such as the number of knee operations, the length of waiting lists, and the success rate of surgery, where process mining is more concerned with the paths followed by individual patients and whether certain procedures are followed or not.

Using process mining, typical manager questions that can be answered include:

- What is the most frequent path in my process?
- To what extend do my cases comply with my process model?
- What are the routing probabilities in my process?
- What are the throughput times of my cases?
- What are the service times for my tasks?
- When will a case be completed?
- How much time was spent between any two tasks in my process?
- What are the business rules in my process, and are they being obeyed?
- How many of my people are typically involved in a case?
- Which people are central in my organization?

2 ProM

ProM is the world-leading process mining toolkit. It is an extensible framework that supports a wide variety of process mining techniques in the form of plug-ins. It is platform independent as it is implemented in Java, and can be downloaded free of charge from `www.processmining.org`. ProM is issued under

an open source license and we invite researchers and developers to contribute in the form of new plug-ins. The development of ProM is not restricted to the Eindhoven University of Technology: The current version of ProM includes work from researchers from all over the world, including for example Australia, Austria, China, Germany, and Italy.

Currently, there are already more than 230 plug-ins available, and we support the import of (and the conversion between) several process modeling languages, like Petri nets (PNML, TPN), EPCs/EPKs (Aris graph format, EPML), YAWL, and many more. There are mining plug-ins, such as plug-ins supporting control-flow mining techniques (Alpha algorithm, Genetic mining, Heuristics Miner, Multi-phase mining), plug-ins analyzing the organizational perspective (Social Network miner, Staff Assignment miner), plug-ins dealing with the data perspective (Decision miner), plug-ins for mining less-structured, flexible processes (Fuzzy Miner), elaborate data visualization plug-ins (Cloud Chamber Miner), and many more. Furthermore, there are analysis plug-ins dealing with the verification of process models (Woflan analysis), verification of Linear Temporal Logic (LTL) formulas on a log, checking the conformance between a given process model and a log, and performance analysis (Basic statistical analysis, and Performance Analysis with a given process model). Finally, ProM offers a large array of log filters, which are a valuable tool for cleaning logs from undesired, or unimportant, artefacts.

3 Case studies

Thus far, ProM has been applied in a wide variety of organizations, which include municipalities (Alkmaar, Heusden, Harderwijk, etc.), government agencies (*Rijkswaterstaat*, Centraal Justitieel Incasso Bureau, Justice department), insurance related agencies (UWV), banks (ING Bank), hospitals (*AMC hospital*, Catharina hospital), multinationals (DSM, Deloitte),high-tech system manufacturers and their customers (Philips Healthcare, *ASML*, Thales), and media companies (e.g. Winkwaves). To give some insights in the results we obtained so far, we provide some details on the three italicized organizations.

For a provincial office of *Rijkswaterstaat* (the Dutch National Public Works Department), we have conducted a case study on its invoice process, which has shown that the bad performance of this process was mainly due to the fact that some of the employees often work at remote sites. Furthermore, the case study showed that it is worthwhile to combine different mining perspectives to reach a richer understanding of the process. In this case, for example, the process model revealed the problems (loops), but it took an organizational model to identify the key players, and a case-oriented analysis to understand the impact of these loops on the process performance. Please see [1] for more information on this case study.

For *ASML* (the leading manufacturer of wafer scanners in the world), we have conducted a case study on its test process, which has yielded concrete suggestions for process improvement. These suggestions included reordering of

tasks to prevent feedback loops and using idle time for scheduling. However, this case study has also shown that further research is needed to develop process mining techniques that are particularly suitable for analyzing less structured processes like the highly dynamic test process of ASML. Please see [7] for details.

For the Dutch *AMC hospital*, we have conducted a case study which has shown that we were able to derive understandable models for large groups of patients, which was confirmed by people of the hospital. Nevertheless, this case study has also shown that traditional process mining approaches have problems dealing with unstructured processes as, for example, can be found in a hospital environment. Please see [6] for more information.

4 Conclusion

Process mining is a fertile field of research, and the ProM toolkit is the leading tool to open up this field. Using ProM, we can answer questions that are very relevant to managers, and case studies have shown that we are also able to do so in a real world setting.

References

1. W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, and H.M.W. Verbeek. Business Process Mining: An Industrial Application. *Information Systems*, 32(5):713–732, 2007.
2. W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142, 2004.
3. R. Agrawal, D. Gunopulos, and F. Leymann. Mining Process Models from Workflow Logs. In *Sixth International Conference on Extending Database Technology*, pages 469–483, 1998.
4. A. Datta. Automating the Discovery of As-Is Business Process Models: Probabilistic and Algorithmic Approaches. *Information Systems Research*, 9(3):275–301, 1998.
5. J. Herbst. A Machine Learning Approach to Workflow Management. In *Proceedings 11th European Conference on Machine Learning*, volume 1810 of *Lecture Notes in Computer Science*, pages 183–194. Springer-Verlag, Berlin, 2000.
6. R. S. Mans, M.H. Schonenberg, M. Song, W. M. P. van der Aalst, and P. J. M. Bakker. Application of process mining in healthcare - a case study in a Dutch hospital. In A. Fred, J. Filipe, and H. Gamboa, editors, *Biomedical Engineering Systems and Technologies*, volume 25 of *Communications in Computer and Information Science*, pages 425–438. Springer Berlin Heidelberg, 2009.
7. A. Rozinat, I. S. M. de Jong, C. W. Günther, and W. M. P. van der Aalst. Process mining applied to the test process of wafer steppers in ASML. In *IEEE Transactions on Systems, Man, and CyberneticsPart C: Applications and Reviews*. 2009. To appear.
8. A. Rozinat and W. M. P. van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems (IS)*, 33(1):64–95.
9. A.J.M.M. Weijters and W.M.P. van der Aalst. Rediscovering Workflow Models from Event-Based Data using Little Thumb. *Integrated Computer-Aided Engineering*, 10(2):151–162, 2003.