Process Mining with ProM

W.M.P. van der Aalst B.F. van Dongen C.W. Günther R.S. Mans
A.K. Alves de Medeiros A. Rozinat M. Song H.M.W. Verbeek
A.J.M.M. Weijters

Eindhoven University of Technology, Eindhoven, The Netherlands,
E-mail: a.j.m.m.weijters@tue.nl

Abstract

This demonstration paper describes version 4.1 of the *ProM* process mining tool. *Process mining* techniques attempt to extract non-trivial and useful process information from so-called "event logs". ProM allows for the discovery of different process perspectives (e.g., control-flow, time, resources, and data) and supports related techniques such as control-flow mining, performance analysis, resource analysis, conformance checking, verification, etc. This makes ProM a practical and versatile tool for process analysis and discovering.

Within organizations there has been a shift from *data* orientation to *process* orientation. By process we mean the way an organization arranges there work and recourses, for instance the order in which tasks are performed and which group of people are allowed to perform specific tasks. Sometimes, organizations have very explicit process descriptions of the way the work is organized and this description is supported by a process aware information system like, for instance, a workflow management system (WFM). But even if there are explicit descriptions of the way the work should be done, the practical way of working can differ considerably from the prescribed way of working. Other times, there is no, or only a very immature process description available. However, in many situations it is possible to gather information about the processes as they take place. For instance, in many hospitals, information about the different treatments of a patient are registered (date, time, treatment, medical staff) for, reasons like financial administration. This kind of information in combination with appropriate mining techniques can also be used to get more insight in the health care process. We use the term *process mining* for the method of distilling process knowledge from a set of real executions.

Figure 1 shows a fragment of the log in MXML format, the format used by ProM. Event logs are used as the starting point for mining. We distinguish three different mining perspectives: (1) the process perspective, (2) the organizational perspective and (3) the case perspective. The *process perspective* focuses on the control-flow, i.e., the ordering of activities. The goal of mining this perspective is to find a good characterization of all possible paths, expressed in terms of, for instance, a Petri net. The *organizational perspective* focuses on the originator field, i.e., which performers are involved in performing the activities and how they are related. The goal is to either structure the organization by classifying people in terms of roles and organizational units or to show relations between individual performers. The *case perspective* focuses on properties of cases. Cases can be characterized by their path in the process or by the originators working on a case. However, cases can also be characterized by the values of the corresponding data elements. For example, in the phone repair log it may be interesting to know the differences in throughput times between different telephone types.

To address the three perspectives and the logical and performance issues a set of plug-ins has been developed for the ProM framework [2]. ProM is open source and uses a plug-able architecture, e.g., people can add new process mining techniques by adding plug-ins without spending any efforts on the loading and filtering of event logs and the visualization of the resulting models. An example is the plug-in implementing the α-algorithm [1], i.e., a technique to automatically derive Petri nets from event logs. ProM version 4.1 provides six different types of plug-ins, in total 142 plug-ins are available. ProM is open source and can be downloaded from www.processmining.org. For more details about the ProMframework, its plug-ins, and the common XML-format, we refer to the same URL.

```
</ProcessInstance>
      <ProcessInstance id="266" description="">
        <AuditTrailEntry>
        <Data>
         <Attribute name = "phoneT">T1</Attribute>
        <Attribute name = "countryT">B </Attribute>
         </Data>
         <WorkflowModelElement>Bregistration
                  </WorkflowModelElement>
        <EventType >complete</EventType>
        <Timestamp>2006-01-02T08:01:00</Timestamp>
        <Originator>BLadmin</Originator>
        </AuditTrailEntry>
        <AuditTrailEntry>
        <WorkflowModelElement>Banalyse
                  </WorkflowModelElement>
        <EventType >complete</EventType>
        <Timestamp>2006-01-02T08:03:00</Timestamp>
        <Originator>BT3</Originator>
        </AuditTrailEntry>
        <AuditTrailEntry>
        <Data><Attribute name = "repn">0 </Attribute>
        <Attribute name = "err">8 </Attribute>
         </Data><WorkflowModelElement>Brepair
                  </WorkflowModelElement>
        <EventType >start</EventType>
        <Timestamp>2006-01-02T10:04:00</Timestamp>
        <Originator>Ties</Originator>
        </AuditTrailEntry>
```

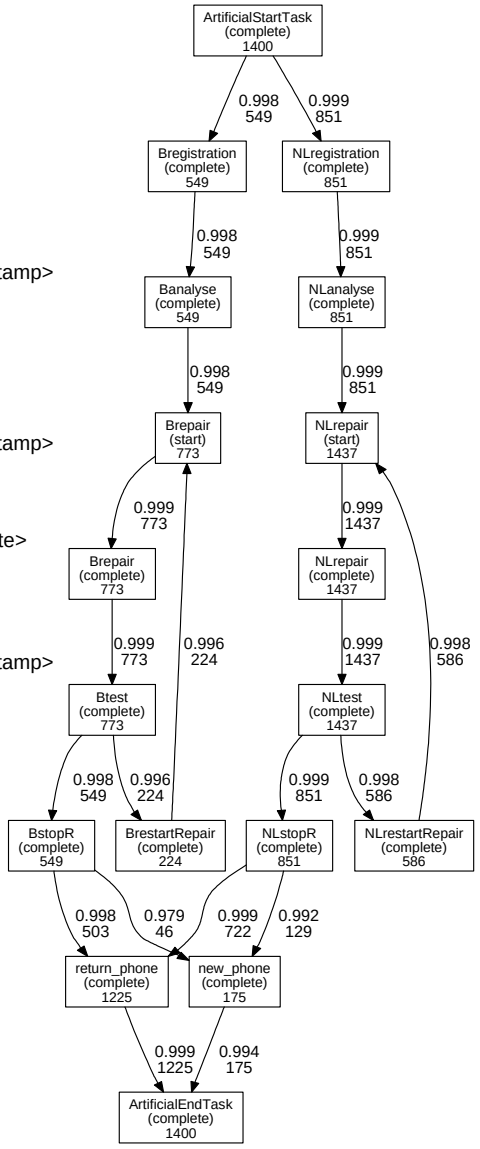

Figure 1: Fragment of an event log in MXML-format from a telephone repair process and the mined control-flow model.

References

[1] 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.

[2] B. van Dongen, A.K. Alves de Medeiros, H.M.W. Verbeek, A.J.M.M. Weijters, and W.M.P. van der Aalst. The ProM framework: A New Era in Process Mining Tool Support. In G. Ciardo and P. Darondeau, editors, *Application and Theory of Petri Nets 2005*, volume 3536 of *Lecture Notes in Computer Science*, pages 444–454. Springer-Verlag, Berlin, 2005.