LTAT.05.025

Business Process Mining

Lecture 3: Concepts & Capabilities of Process Mining

Anti Alman

Junior Research Fellow @ University of Tartu
Structure of the Course

Week 2: Performance Measurement & Dashboards

- Process mining
  - Automated discovery
  - Conformance Checking
  - Performance Mining
  - What-If Process Mining
  - Predictive monitoring

Week 3: Concepts and Capabilities

Week 4: Week 8: Extract-Transform-Load for Process Mining
Week 5: Week 9: Declarative Process Mining
Week 6: Week 11: Process Mining Algorithms
Week 7: Week 12: Process Mining In Practice
Week 10: Week 13: Course Recap and Future Trends
How does Process Mining work?

01. Process Logging
   The systems supporting the execution of the business process are configured to record events capturing the execution of tasks in the process.

02. Data Collection
   This business process execution data needs to be transformed into "Event Logs". Correct time-stamps and activity names are essential.

03. Data Preparation
   The data is extracted and prepared by data engineers (IT side) based on business requirements.

04. Process Mining
   Automated process discovery and other types of analysis can then be used to discover and analyze the process.
Process Mining

event log
### Structure of a Business Process Event Log

**Required data**
- Case identifier
- Time of the activity
  - At least one timestamp
- Name of the activity

**Optional data**
- Resource identifier
- Process specific data
  - Case attributes
  - Event attributes

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Timestamp</th>
<th>Activity</th>
<th>Resource</th>
<th>Loan goal</th>
<th>Requested amt</th>
<th>Offered amt</th>
</tr>
</thead>
<tbody>
<tr>
<td>C001</td>
<td>18-10-2016</td>
<td>Check completeness</td>
<td>Sue</td>
<td>Mortgage</td>
<td>100 000</td>
<td>-</td>
</tr>
<tr>
<td>C001</td>
<td>19-10-2016</td>
<td>Check credit history</td>
<td>Sue</td>
<td>Mortgage</td>
<td>100 000</td>
<td>-</td>
</tr>
<tr>
<td>C001</td>
<td>19-10-2016</td>
<td>Calculate risk score</td>
<td>Bob</td>
<td>Mortgage</td>
<td>100 000</td>
<td>-</td>
</tr>
<tr>
<td>C001</td>
<td>20-10-2016</td>
<td>Make offer</td>
<td>Mike</td>
<td>Mortgage</td>
<td>100 000</td>
<td>70 000</td>
</tr>
<tr>
<td>C001</td>
<td>25-10-2016</td>
<td>Make offer</td>
<td>Mike</td>
<td>Mortgage</td>
<td>100 000</td>
<td>80 000</td>
</tr>
<tr>
<td>C002</td>
<td>20-10-2016</td>
<td>Check completeness</td>
<td>Sue</td>
<td>Car</td>
<td>15 000</td>
<td>-</td>
</tr>
<tr>
<td>C002</td>
<td>20-10-2016</td>
<td>Check credit history</td>
<td>Sue</td>
<td>Car</td>
<td>15 000</td>
<td>-</td>
</tr>
<tr>
<td>C002</td>
<td>22-10-2016</td>
<td>Calculate risk score</td>
<td>Elsa</td>
<td>Car</td>
<td>15 000</td>
<td>-</td>
</tr>
<tr>
<td>C002</td>
<td>24-10-2016</td>
<td>Reject application</td>
<td>Elsa</td>
<td>Car</td>
<td>15 000</td>
<td>-</td>
</tr>
<tr>
<td>C003</td>
<td>02-11-2016</td>
<td>Check completeness</td>
<td>Maria</td>
<td>Mortgage</td>
<td>30 000</td>
<td>-</td>
</tr>
<tr>
<td>C003</td>
<td>04-11-2016</td>
<td>Ask for additional data</td>
<td>Maria</td>
<td>Mortgage</td>
<td>30 000</td>
<td>-</td>
</tr>
<tr>
<td>C003</td>
<td>10-11-2016</td>
<td>Check credit history</td>
<td>Maria</td>
<td>Mortgage</td>
<td>30 000</td>
<td>-</td>
</tr>
</tbody>
</table>

...
Demo Time!
Automated Process Discovery

<table>
<thead>
<tr>
<th>CID</th>
<th>Task</th>
<th>Time Stamp</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13219</td>
<td>Enter Loan Application</td>
<td>2007-11-09 11:20:10</td>
<td></td>
</tr>
<tr>
<td>13219</td>
<td>Retrieve Applicant Data</td>
<td>2007-11-09 11:22:15</td>
<td></td>
</tr>
<tr>
<td>13220</td>
<td>Enter Loan Application</td>
<td>2007-11-09 11:22:40</td>
<td></td>
</tr>
<tr>
<td>13219</td>
<td>Compute Installments</td>
<td>2007-11-09 11:22:45</td>
<td></td>
</tr>
<tr>
<td>13219</td>
<td>Notify Eligibility</td>
<td>2007-11-09 11:23:00</td>
<td></td>
</tr>
<tr>
<td>13219</td>
<td>Approve Simple Application</td>
<td>2007-11-09 11:24:30</td>
<td></td>
</tr>
<tr>
<td>13220</td>
<td>Compute Installments</td>
<td>2007-11-09 11:24:35</td>
<td></td>
</tr>
</tbody>
</table>

Process Map
(directly follows graph)

BPMN process model
A process map of an event log is a graph where:

- Each activity is represented by one node
- An arc from activity A to activity B means that B is directly followed by A in at least one case in the log
Anatomy of a Process map

- **Activity (and its frequency)**
  - Directly-follows relation between two activities (and its frequency)
  - Eventually-follows relation between two activities (not visualized)
  - Activity self-loop (rework)
  - Short loop (possible rework)
  - Final activity (multiple possible)

- **Initial activity (multiple possible)**

- **Activity (and its frequency)**

11
Process maps: overlays

Nodes in a process map can be coloured, and arcs’ thickness can be changed, to capture:

- **Frequency**: How often a given activity or directly-follows relation occurs?
- **Duration**: processing times and cycle times for activities, waiting times for directly-follow relations
- **Other attributes**: some tools support enhancement by other attributes, e.g. cost, revenue, sales volume etc., if these data attributes are available
Demo Time!
Process Map Complexity

• Process mining tools use process maps as the main visualization technique for event logs
• However, full process maps can be very large and complicated, thus requiring simplification
  • Process map abstraction
  • Event log filtering
Process map: Abstraction

Common process mining tools provide abstraction capabilities to simplify process maps:

- Show only **most frequent** activities
- Show only **most frequent** arcs

Some tools offer further abstraction capabilities to:

- Show only **least frequent** activities
- Show only **least frequent** arcs
- Show only **fastest/slowest** activities
- Show only **fastest/slowest** arcs
Log filters

Attribute filters
- Retain only cases such that at least one event fulfils a condition over one or more attributes, for example: *at least one event where Activity = “Close order” and Resource = “Anti Alman”*

Performance filter
- Retain only cases that have a duration above or below a given value (more than 3 days)

Path filter
- Retain traces where there is a pair of events that fulfil a given condition (e.g. “Create invoice” eventually followed by “Create purchase order”)

Repetition filter
- Retain only cases where activity “Review order” is performed three times or more
Demo Time!
Process Mining

Automated Process Discovery

event log

Conformance Checking

Business rules / normative model

✓ / ✗

discovered process model
Conformance checking

Modeled process
(Expected: 8 hours)

Actual process
(In reality: 18 hours)
Rule-Based Conformance Checking

• Oftentimes, we don’t have a full process model available

• Instead, we have a set of compliance rules, e.g.
  • Certain tasks MUST be executed in every case
  • Certain tasks can be executed at most once
  • When a task A occurs, another task B must also occur (within e.g. 2 hours)
  • Two tasks A and B CANNOT be executed by the same person (four-eyes principle)

• In Apromore (and similar tools) we mostly use log filters
  • Some other tools allow for declarative process rules (e.g. RuM)
Trace Alignment

Automatically detect differences between a process instance and the process model

- Skipped/unforeseen activities
- Control flow differences
Demo Time!
Process Mining

Automated Process Discovery

Performance Mining

Conformance Checking

Business rules / normative model

Enhanced process model

discovered process model

event log

✓ / ✗
Performance Mining

• Process mining tools typically support performance mining in three ways:
  • Performance-enhanced process maps
  • Dashboards
  • Log animation
Performance-Enhanced Process Map

Activity label (and its duration)

Create Purchase Requisition
30.4 mins

Analyze Purchase Requisition
6.5 mins

Create Request for Quotation Requester
10.4 mins

Create Request for Quotation Requester Manager
2 mins

Analyze Request for Quotation
23 mins

Amend Request for Quotation Requester
9.8 mins

Send Request for Quotation to Supplier
22.7 mins

Amend Request for Quotation Requester Manager
19.3 mins

Directly-follows relation between two labels (with waiting time)

Bottleneck node (all incoming arcs are slow)

4.2 hrs  17.3 hrs

21.7 d

5.7 mins

14 d

50.4 mins  14.9 d  15.5 hrs  16 hrs  14.9 d
Demo Time!
Process Mining

- Automated Process Discovery
- Performance Mining
- Conformance Checking
- Variant Analysis

- Event log
- Event log'
- Discovered process model
- Enhanced process model
- Difference diagnostics
- Business rules / normative model
Variant Analysis

Find differences in performance and control-flow between multiple variants of a process

A process variant is a subset of the executions of a process corresponding to a given product, customer, etc.
In process mining tools, such as Apromore, variant analysis can be approached as follows:

1. First, use filtering to “slice” the event log into multiple logs (one per variant).
2. Then compare the multiple logs using:
   ▪ Side-by-side comparison of process maps or process models
   ▪ A comparative multi-log dashboard
   ▪ A multi-log animation
Demo Time!
The bulk of process mining techniques can be divided into four capabilities.

- **Automated Process Discovery**
  By using the automated process discovery, an actual process model is created from the event log. This represents the current sequence of work steps and does not necessarily have to correspond to the target process.

- **Conformance Checking**
  Conformance checking provides a comparison of the actual and the target process and thus enables the detection of differences and deviations.

- **Variant Analysis**
  Process outcomes often differ in practice. But why? The variant analysis enables a simple comparison of several process variants and serves as the basis for process optimization.

- **Performance Mining**
  By applying performance mining techniques, e.g. nodes or cycle times can be highlighted in color to simplify the subsequent analysis.
The Next Practice Session

• We will shift from “pen-and-paper” to using Apromore
  • Address: https://apromore.cloud.ut.ee/
  • You should have access with the university username and password
    • Same credentials as for the study information system (ÕIS)
  • Please check your access before the practice session
    • In case of problems notify us via Slack
Structure of the Course

Week 2: Performance Measurement & Dashboards

Week 4: Automated discovery
Week 5: Conformance Checking
Week 6: Performance Mining
Week 7: What-If Process Mining
Week 10: Predictive monitoring

Week 3: Concepts and Capabilities

Week 8: Extract-Transform-Load for Process Mining
Week 9: Declarative Process Mining
Week 11: Process Mining Algorithms
Week 12: Process Mining In Practice
Week 13: Course Recap and Future Trends