Lecture 12: Business Process Monitoring & Mining

Marlon Dumas
University of Tartu

marlon.dumas@ut.ee
3 months later
Business Process Monitoring

Event stream → DB logs → Event log → Process mining → Dashboards & reports

Event stream
DB logs
Event log
Process mining
Dashboards & reports
Types of process dashboards

- Operational dashboards (runtime)
- Tactical dashboards (historical)
- Strategic dashboards (historical)
Operational process dashboards

• Aimed at process workers & operational managers
• Emphasis on monitoring (detect-and-respond), e.g.:
  - Work-in-progress
  - Problematic cases - e.g. overdue/at risk cases
  - Resource load
Tactical dashboards

• Aimed at process owners / managers
• Emphasis on analysis and management
  • E.g. detecting bottlenecks
• Typical process performance indicators
  • Cycle times
  • Error rates
  • Resource utilization
Tactical Performance Dashboard @ Australian Insurer
Strategic dashboards

- Aimed at executives & managers
- Emphasis on linking process performance to strategic objectives
## Strategic Performance Dashboard @ Australian Utilities Provider

<table>
<thead>
<tr>
<th>Key Performance</th>
<th>Manage Unplanned Outages</th>
<th>Manage Emergencies &amp; Disasters</th>
<th>Manage Work Programming &amp; Resourcing</th>
<th>Manage Procurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer Satisfaction</td>
<td>0.5</td>
<td>0.55</td>
<td>-</td>
<td><strong>0.2</strong></td>
</tr>
<tr>
<td>Customer Complaint</td>
<td>0.6</td>
<td>-</td>
<td>-</td>
<td><strong>0.5</strong></td>
</tr>
<tr>
<td>Customer Feedback</td>
<td><strong>0.4</strong></td>
<td>-</td>
<td>-</td>
<td><strong>0.8</strong></td>
</tr>
<tr>
<td>Connection Less Than Agreed Time</td>
<td><strong>0.3</strong></td>
<td>0.6</td>
<td>0.7</td>
<td>-</td>
</tr>
</tbody>
</table>
Process: Manage Procurement

Process: Manage Emergencies & Disasters

Process: Manage Unplanned Outages

Overall Process Performance

Financial
People
Customer Excellence
Operational Excellence
Risk Management
Health & Safety

Customer Complaint
Customer Satisfaction

Customer Rating (%)
Customer Loyalty Index
Average Time Spent on Plan

Satisfied Customer Index
Market Share (%)
Exercise

Sketch operational and tactical process monitoring dashboards for CVS Pharmacy’s prescription fulfillment process (See Chapter 1, Exercise 1.6). Consider the viewpoints of each stakeholder in the process.

• Customer
• Pharmacist
• Technician
• Process owner (overseeing 500+ pharmacies distributed geographically)
Business Process Monitoring

Event stream → DB logs → Event log → Dashboards & reports → Process mining

Dashboards & reports

Process mining

UNIVERSITY of TARTU
INSTITUTE OF COMPUTER SCIENCE
Process Mining

- **Discovery**
  - Discovered model

- **Performance**
  - Enhanced model

- **Variance**
  - Event log

- **Conformance**
  - Event log'
  - Input model
  - Diagnostics

- **Difference**

---

**Diagram Details**

- Process Mining workflow:
  - Discovery to Performance
  - Performance to Variance
  - Variance to Conformance
  - Conformance with validation output (√, ×)

---

**Source Information**

- University of Tartu
- Institute of Computer Science
Event logs structure: minimum requirements

Concrete formats:
- Comma-Separated Values (CSV)
- XES (XML format)
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 T 11:20:10</td>
<td>-</td>
</tr>
<tr>
<td>13219</td>
<td>Retrieve Applicant Data</td>
<td>2007-11-09 T 11:22:15</td>
<td>-</td>
</tr>
<tr>
<td>13220</td>
<td>Enter Loan Application</td>
<td>2007-11-09 T 11:22:40</td>
<td>-</td>
</tr>
<tr>
<td>13219</td>
<td>Compute Installments</td>
<td>2007-11-09 T 11:22:45</td>
<td>-</td>
</tr>
<tr>
<td>13219</td>
<td>Notify Eligibility</td>
<td>2007-11-09 T 11:23:00</td>
<td>-</td>
</tr>
<tr>
<td>13219</td>
<td>Approve Simple Application</td>
<td>2007-11-09 T 11:24:30</td>
<td>-</td>
</tr>
<tr>
<td>13220</td>
<td>Compute Installments</td>
<td>2007-11-09 T 11:24:35</td>
<td>-</td>
</tr>
</tbody>
</table>

Diagram: 
- Enter Loan Application
- Retrieve Applicant Data
- Compute Installments
- Notify Eligibility
- Approve Simple Application
- Approve Complex Application
- Notify Rejection
# Process Mining Tools

<table>
<thead>
<tr>
<th>Open-source</th>
<th>Lightweight</th>
<th>Mid-range</th>
<th>Heavyweight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apromore</td>
<td>Disco</td>
<td>Minit</td>
<td>ARIS Process Performance Manager</td>
</tr>
<tr>
<td>ProM</td>
<td></td>
<td>myInvenio</td>
<td>Celonis Process Mining</td>
</tr>
<tr>
<td>bupaR</td>
<td></td>
<td>ProcessGold</td>
<td>Perceptive Process Mining (Lexmark)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QPR Process Analyzer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Signavio Process Intelligence</td>
<td>Interstage Process Discovery (Fujitsu)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>StereoLOGIC Discovery Analyst</td>
<td></td>
</tr>
</tbody>
</table>
Fluxicon Disco
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 trace in the log

Arcs in a process map can be annotated with:
- Absolute frequency: how many times B directly follows A?
- Relative frequency: in what percentage of times when A is executed, is directly followed by B?
- Time: What is the average time between the occurrence of A and the occurrence of B?
Process Maps – Example

**Event log:**
10: a,b,c,g,e,h
10: a,b,c,f,g,h
10: a,b,d,g,e,h
10: a,b,d,e,g,h
10: a,b,e,c,g,h
10: a,b,e,d,g,h
10: a,c,b,e,g,h
10: a,c,b,f,g,h
10: a,d,b,e,g,h
10: a,d,b,f,g,h
## Process Maps – Exercise

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Task Name</th>
<th>Originator</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>File Fine</td>
<td>Anne</td>
<td>20-07-2004 14:00:00</td>
</tr>
<tr>
<td>2</td>
<td>File Fine</td>
<td>Anne</td>
<td>20-07-2004 15:00:00</td>
</tr>
<tr>
<td>1</td>
<td>Send Bill</td>
<td>system</td>
<td>20-07-2004 15:05:00</td>
</tr>
<tr>
<td>2</td>
<td>Send Bill</td>
<td>system</td>
<td>20-07-2004 15:07:00</td>
</tr>
<tr>
<td>3</td>
<td>File Fine</td>
<td>Anne</td>
<td>21-07-2004 10:00:00</td>
</tr>
<tr>
<td>3</td>
<td>Send Bill</td>
<td>system</td>
<td>21-07-2004 14:00:00</td>
</tr>
<tr>
<td>4</td>
<td>File Fine</td>
<td>Anne</td>
<td>22-07-2004 11:00:00</td>
</tr>
<tr>
<td>4</td>
<td>Send Bill</td>
<td>system</td>
<td>22-07-2004 11:10:00</td>
</tr>
<tr>
<td>1</td>
<td>Process Payment</td>
<td>system</td>
<td>24-07-2004 15:05:00</td>
</tr>
<tr>
<td>1</td>
<td>Close Case</td>
<td>system</td>
<td>24-07-2004 15:06:00</td>
</tr>
<tr>
<td>2</td>
<td>Reminder</td>
<td>Mary</td>
<td>20-08-2004 10:00:00</td>
</tr>
<tr>
<td>3</td>
<td>Reminder</td>
<td>John</td>
<td>21-08-2004 10:00:00</td>
</tr>
<tr>
<td>2</td>
<td>Process Payment</td>
<td>system</td>
<td>22-08-2004 09:05:00</td>
</tr>
<tr>
<td>2</td>
<td>Close case</td>
<td>system</td>
<td>22-08-2004 09:06:00</td>
</tr>
<tr>
<td>4</td>
<td>Reminder</td>
<td>John</td>
<td>22-08-2004 15:10:00</td>
</tr>
<tr>
<td>4</td>
<td>Reminder</td>
<td>Mary</td>
<td>22-08-2004 17:10:00</td>
</tr>
<tr>
<td>4</td>
<td>Process Payment</td>
<td>system</td>
<td>29-08-2004 14:01:00</td>
</tr>
<tr>
<td>4</td>
<td>Close Case</td>
<td>system</td>
<td>29-08-2004 17:30:00</td>
</tr>
<tr>
<td>3</td>
<td>Reminder</td>
<td>John</td>
<td>21-09-2004 10:00:00</td>
</tr>
<tr>
<td>3</td>
<td>Reminder</td>
<td>John</td>
<td>21-10-2004 10:00:00</td>
</tr>
<tr>
<td>3</td>
<td>Process Payment</td>
<td>system</td>
<td>25-10-2004 14:00:00</td>
</tr>
<tr>
<td>3</td>
<td>Close Case</td>
<td>system</td>
<td>25-10-2004 14:01:00</td>
</tr>
</tbody>
</table>
Process Maps in Disco

• Disco (and other commercial process mining tools) use process maps as the main visualization technique for event logs

• These tools also provide three types of operations:
  1. **Abstract** the process map:
     • Show only most frequent activities
     • Show only most frequent arcs
  2. **Filter** the traces in the event log...
Types of filters

• Event filters
  • Retain only events that fulfil a given condition (e.g. all events of type “Create purchase order”)

• Performance filter
  • Retain traces that have a duration above or below a given value

• Event pair filter (a.k.a. “follower” filter)
  • Retain traces where there is a pair of events that fulfil a given condition (e.g. “Create invoice” followed by “Create purchase order”)

• Endpoint filter
  • Retain traces that start with or finish with an event that fulfils a given condition
Process Maps in Disco

• Disco (and other commercial process mining tools) use process maps as the main visualization technique for event logs

• These tools also provide three types of operations:
  1. **Abstract** the process map:
     • Show only most frequent activities
     • Show only most frequent arcs
  2. **Filter** the traces in the event log
  3. Enhance the process map
Process Map Enhancement

• Nodes and arcs in a process map can be color-coded or thickness-coded to capture:
  • Frequency: How often a given task or a given directly-follows relation occurs?
  • Time performance: processing times, waiting times, cycles times of tasks
  • More advanced tools support enhancement by other attributes, e.g. cost, revenue, etc. if the data is available.
Disco tutorial
Exercise

Using Disco, answer the following questions on the PurchasingExample log:

• How many cases had to settle a dispute with the purchasing agent?

• Is there a difference in cycle time for the cases that had to settle a dispute with the purchasing agent, compared to the ones that did not? Make sure you only compare cases that actually reach the endpoint ‘Pay invoice’

• Are there any cases where the invoice is released and authorized by the same resource? And if so, who is doing this most often?
Process Maps - Limitations

• Process maps over-generalize: some paths of a process map might not exist and might not make sense
  • Example: Draw the process map of [ abc, adc, afce, afec ] and check which traces it can recognize, for which there is no support in the event log.

• Process maps make it difficult to distinguish conditional branching, parallelism, and loops.
  • See previous example... or a simpler one: [abcd, acbd]

• Solution: automated BPMN process discovery
  • More on this next week...
Example

• Let us consider the following event log of a telephone repair process?
  • [http://tinyurl.com/repairLogs](http://tinyurl.com/repairLogs)

• Open up this log and check if you can understand how this process is performed based on the process map only...
Process Performance Mining

• Dotted charts
  • One line per trace, each line contains points, one point per event
  • Each event type is mapped to a colour
  • Position of the point denotes its occurrence time (in a relative scale)
  • Birds-eye view of the timing of different events (e.g. activity end times), but does not allow one to see the “processing” times of activities

• Timeline diagrams
  • One line per trace, each line contains segments capturing the start and end of tasks
  • Captures process time (unlike dotted charts)
  • Not scalable for large event logs – good to show “representative” traces

• Performance-enhanced process maps
  • Process maps where nodes are colour-coded w.r.t a performance measure. Nodes may represent activities (default option)
  • But they may represent resources and then arcs denote hands-offs between resources
Dotted chart
Timeline diagram

See: http://timelines.nirdizati.org
Performance-enhanced process map
Nodes are activities (default)
Performance-enhanced process map
Nodes are tasks (handoff graph)

Screenshot of MyInvenio
Exercise

• Consider the following event log of a telephone repair process: [http://tinyurl.com/repairLogs](http://tinyurl.com/repairLogs)
  • What are the bottlenecks in this process?
  • Which task has the longest waiting time and which one has the longest processing time?
Given two logs, find the *differences* and *root causes* for variation or deviance between the two logs.
Case Study: Variants Analysis at Suncorp

Time elapsed (days)

Payout estimate ($)

- Good
- OK
- Bad

Expected Performance Line

+ (simple and quick) - expected
- (simple and slow) - needs improvement
- (complex and slow) - expected
- (complex and quick) - good

NOTE: x and y axis are logarithmically-scaled
Variants Analysis via Process Map Comparison

Simple claims and quick

Simple claims and slow
Variants analysis - Exercise

We consider a process for handling health insurance claims, for which we have extracted two event logs, namely L1 and L2. Log L1 contains all the cases executed in 2011, while L2 contains all cases executed in 2012. The logs are available in the book’s companion website or directly at: http://tinyurl.com/InsuranceLogs

Based on these logs, answer the following questions using a process mining tool:

1. What is the cycle time of each log?
2. Where are the bottlenecks (highest waiting times) in each of the two logs and how do these bottlenecks differ?
3. Describe the differences between the frequency of tasks and the order in which tasks are executed in 2011 (L1) versus 2012 (L2). Hint: If you are using process maps, you should consider using the abstraction slider in your tool to hide some of the most infrequent arcs so as to make the maps more readable.
Recap

• Process monitoring is about analysing events produced during the execution of a process to understand the performance and conformance of the process.

• Two major types of approaches to monitoring are:
  • Performance dashboards
  • Process mining

• Process mining is centered around diagrammatic representations of a process
  • Process maps or BPMN process models

• Process mining tools such as Disco allow us to apply abstraction, filtering, and enhancement to process maps and event logs.