Multidimensional Process Mining using Process Cubes

Alfredo Bolt
TU/e
Process Complexity

Too much behavior!
Process Complexity

Current focus on mining the “big picture”
All-in-one → “Where’s Waldo?” effect
Alternative: Manual filtering ?!!

You will miss some details!
Not clearly visible in the big picture
Discarded because of frequency
Then some questions emerge…

How do behaviors differ from each other?
Process Variations

What if behavior can be related to data?
Multidimensional Process Mining

Where/When else can I see the same behavior?
Clustering
Process Variations

Imagine a process of taking a course:
Students are instances
Lectures are activities
Exam is an activity too

Now, imagine an event log containing many courses:
Each course has its own set of lectures
Each course has many students

If we discover a process model, how would it look like?
“Big Picture” issues...
We need a better way to...

Analyze processes from different data perspectives
Multidimensional Analysis

Group, study and compare different behaviors
Process Cubes
Multidimensional analysis: OLAP

On Line Analytical Processing (Cube)
MOLAP / ROLAP / HOLAP

Fact = Value + Dimensional attributes

On Line → Real Time → Pre-calculations
Fancy filtering / clustering with a nice presentation
OLAP $\Rightarrow$ Process Mining

Facts v/s Events
Both have *dimensional* attributes and a *value*
But they are represented in *different* ways

OLAP makes an assumption to make aggregation and summarization possible:
All groupings of facts must be disjoint

Grouping of Events are not guaranteed to be disjoint:
Traces can share a common event
Events can have more than one value for a dimensional attribute
   Person working for two different departments
Process Cubes

Not On Line
No pre-calculations

Same functionalities as OLAP
But different representations
How does Process Cubes (PC) work?

1. Data Import
2. PC Structure Configuration
3. Mapping
4. PC View Configuration
5. Results representation
# Data Import

## Event Base

Just a collection of events!

No traces, no tags, just properties

<table>
<thead>
<tr>
<th>event_id</th>
<th>sales_order</th>
<th>timestamp</th>
<th>action</th>
<th>resource</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>1</td>
<td>28-12-2014:06.30</td>
<td>A</td>
<td>John</td>
<td>100</td>
</tr>
<tr>
<td>0002</td>
<td>1</td>
<td>28-12-2014:07.15</td>
<td>B</td>
<td>Anna</td>
<td></td>
</tr>
<tr>
<td>0003</td>
<td>1</td>
<td>28-12-2014:08.45</td>
<td>C</td>
<td>John</td>
<td></td>
</tr>
<tr>
<td>0004</td>
<td>2</td>
<td>28-12-2014:12.20</td>
<td>A</td>
<td>Peter</td>
<td>150</td>
</tr>
<tr>
<td>0005</td>
<td>1</td>
<td>28-12-2014:20.28</td>
<td>D</td>
<td>Mike</td>
<td></td>
</tr>
<tr>
<td>0006</td>
<td>2</td>
<td>28-12-2014:23.30</td>
<td>C</td>
<td>Anna</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
PC Structure Configuration

Process Cube Structure
Set of Dimensions

Dimension
Directed Acyclic Graph + Value sets
Nodes are attributes
Edges are hierarchical relations
Each attribute has a defined set of possible values
Mapping

Attributes (in PCS) mapped to properties (in EB)
To associate events with the Process Cube Structure

Each attribute has a calculation function
Attribute values can be calculated for all events based on their property values

Examples:

\[(\text{att})\text{ name} = (\text{prop})\text{ name}\]

If (\text{prop}) purchase amount > 1000 \(\text{€}\) then (\text{att}) customer type = Silver
PC View Configuration

Visible dimensions
Used for calculating the cube cells *(DICE)*

Non-visible dimensions
Used only for filtering *(SLICE)*
PC View Configuration

For each attribute, some values are selected
Fancy Filtering

For each (visible) dimension, a granularity is selected
Fancy clustering
Results representation

Grid of Cells
Cell: Collection of events that matches some criteria
(intersection of attribute values)

How do we represent this visually?

Event Log “Mask” from attributes (many possibilities):
Define traces
Define activities
Define timestamps, etc…

ProM plugins ➔ Visual representation
Summing up

Now we can easily compare different process variations using different representations

Start from the “Big Picture” and drill down to details

Subjected to the analyst’s knowledge of the data
LIVE DEMO TIME!
Questions?