Chapter 6
Advanced Process Discovery Techniques

prof.dr.ir. Wil van der Aalst
www.processmining.org
# Overview

## Part I: Preliminaries

<table>
<thead>
<tr>
<th>Chapter 2</th>
<th>Chapter 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Modeling and Analysis</td>
<td>Data Mining</td>
</tr>
</tbody>
</table>

## Part II: From Event Logs to Process Models

<table>
<thead>
<tr>
<th>Chapter 4</th>
<th>Chapter 5</th>
<th>Chapter 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting the Data</td>
<td>Process Discovery: An Introduction</td>
<td>Advanced Process Discovery Techniques</td>
</tr>
</tbody>
</table>

## Part III: Beyond Process Discovery

<table>
<thead>
<tr>
<th>Chapter 7</th>
<th>Chapter 8</th>
<th>Chapter 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conformance Checking</td>
<td>Mining Additional Perspectives</td>
<td>Operational Support</td>
</tr>
</tbody>
</table>

## Part IV: Putting Process Mining to Work

<table>
<thead>
<tr>
<th>Chapter 10</th>
<th>Chapter 11</th>
<th>Chapter 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tool Support</td>
<td>Analyzing “Lasagna Processes”</td>
<td>Analyzing “Spaghetti Processes”</td>
</tr>
</tbody>
</table>

## Part V: Reflection

<table>
<thead>
<tr>
<th>Chapter 13</th>
<th>Chapter 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartography and Navigation</td>
<td>Epilogue</td>
</tr>
</tbody>
</table>
Process discovery

A process discovery software system records events, e.g., messages, transactions, etc. It supports/controls the ‘world’ of people, machines, components, and organizations. The system specifies, configures, implements, and analyzes models that are used to discover business processes. The event logs enhance conformance and support discovery of processes.
Challenge

“able to replay event log”

fitness

generalization

“not overfitting the log”

“Occam’s razor”

simplicity

precision

“not underfitting the log”

process discovery
Observing a stable process infinitely long

- frequent behavior
- trace in event log
- all behavior (including noise)
Target model
Non-fitting model
Overfitting model
Underfitting model
Characteristics of process discovery algorithms

• Representational bias
  - Inability to represent concurrency
  - Inability to deal with (arbitrary) loops
  - Inability to represent silent actions
  - Inability to represent duplicate actions
  - Inability to model OR-splits/joins
  - Inability to represent non-free-choice behavior
  - Inability to represent hierarchy

• Ability to deal with noise

• Completeness notion assumed

• Approach used (direct algorithmic approaches, two-phase approaches, computational intelligence approaches, partial approaches, etc.)
Examples

- Algorithmic techniques
  - Alpha miner
  - Alpha+, Alpha++, Alpha#
  - FSM miner
  - Fuzzy miner
  - Heuristic miner
  - Multi phase miner
- Genetic process mining
  - Single/duplicate tasks
  - Distributed GM
- Region-based process mining
  - State-based regions
  - Language based regions
- Classical approaches not dealing with concurrency
  - Inductive inference (Mark Gold, Dana Angluin et al.)
  - Sequence mining
Heuristic mining

- To deal with noise and incompleteness.
- To have a better representational bias than the \( \alpha \) algorithm (AND/XOR/OR/skip).
- Uses C-nets.
Example log; problem α algorithm

\[ L = [\langle a, e \rangle^5, \langle a, b, c, e \rangle^{10}, \langle a, c, b, e \rangle^{10}, \langle a, b, e \rangle, \langle a, c, e \rangle, \langle a, d, e \rangle^{10}, \langle a, d, d, e \rangle^2, \langle a, d, d, d, e \rangle^1] \]
Taking into account frequencies

\[ L = \left[ \langle a, e \rangle^5, \langle a, b, c, e \rangle^{10}, \langle a, c, b, e \rangle^{10}, \langle a, b, e \rangle^1, \langle a, c, e \rangle^1, \right. \\
\left. \langle a, d, e \rangle^{10}, \langle a, d, d, e \rangle^2, \langle a, d, d, d, e \rangle^1 \right] \\
\]

\[ |a >_L b| = \sum_{\sigma \in L} L(\sigma) \times |\{1 \leq i < |\sigma| \mid \sigma(i) = a \land \sigma(i+1) = b\}| \]

| \(|L| \) | \(a\) | \(b\) | \(c\) | \(d\) | \(e\) |
|------|------|------|------|------|------|
| \(a\) | 0    | 11   | 11   | 13   | 5    |
| \(b\) | 0    | 0    | 10   | 0    | 11   |
| \(c\) | 0    | 10   | 0    | 0    | 11   |
| \(d\) | 0    | 0    | 0    | 4    | 13   |
| \(e\) | 0    | 0    | 0    | 0    | 0    |
Dependency measure

\[ |a >_L b| = \sum_{\sigma \in L} L(\sigma) \times |\{1 \leq i < |\sigma| \mid \sigma(i) = a \land \sigma(i+1) = b\}| \]

\[ |a \Rightarrow_L b| \text{ is the value of the dependency relation between } a \text{ and } b: \]

\[ |a \Rightarrow_L b| = \begin{cases} \frac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\ \frac{|a >_L a|}{|a >_L a| + 1} & \text{if } a = b \end{cases} \]
Example

<table>
<thead>
<tr>
<th>$\Rightarrow_L$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$\frac{0}{0+1} = 0$</td>
<td>$\frac{11-0}{11+0+1} = 0.92$</td>
<td>$\frac{11-0}{11+0+1} = 0.92$</td>
<td>$\frac{13-0}{13+0+1} = 0.93$</td>
<td>$\frac{5-0}{5+0+1} = 0.83$</td>
</tr>
<tr>
<td>$b$</td>
<td>$\frac{0-11}{0+11+1} = -0.92$</td>
<td>$\frac{0}{0+1} = 0$</td>
<td>$\frac{10-10}{10+10+1} = 0$</td>
<td>$\frac{0-0}{0+0+1} = 0$</td>
<td>$\frac{11-0}{11+0+1} = 0.92$</td>
</tr>
<tr>
<td>$c$</td>
<td>$\frac{0-11}{0+11+1} = -0.92$</td>
<td>$\frac{10-10}{10+10+1} = 0$</td>
<td>$\frac{0}{0+1} = 0$</td>
<td>$\frac{0-0}{0+0+1} = 0$</td>
<td>$\frac{11-0}{11+0+1} = 0.92$</td>
</tr>
<tr>
<td>$d$</td>
<td>$\frac{0-13}{0+13+1} = -0.93$</td>
<td>$\frac{0-0}{0+0+1} = 0$</td>
<td>$\frac{0-0}{0+0+1} = 0$</td>
<td>$\frac{4}{4+1} = 0.80$</td>
<td>$\frac{13-0}{13+0+1} = 0.93$</td>
</tr>
<tr>
<td>$e$</td>
<td>$\frac{0-5}{0+5+1} = -0.83$</td>
<td>$\frac{0-11}{0+11+1} = -0.92$</td>
<td>$\frac{0-11}{0+11+1} = -0.92$</td>
<td>$\frac{0-13}{0+13+1} = -0.93$</td>
<td>$\frac{0}{0+1} = 0$</td>
</tr>
</tbody>
</table>

$|a \Rightarrow_L b|$ is the value of the dependency relation between $a$ and $b$:

$$|a \Rightarrow_L b| = \begin{cases} 
\frac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\
\frac{|a >_L a|}{|a >_L a| + 1} & \text{if } a = b
\end{cases}$$

<table>
<thead>
<tr>
<th>$&gt;_L$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>$b$</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>$c$</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>$d$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>$e$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Lower threshold (2 direct successions and a dependency of at least 0.7)

<table>
<thead>
<tr>
<th>&gt;L</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b'L</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>11</td>
<td>0.92</td>
<td>11</td>
<td>0.92</td>
</tr>
<tr>
<td>b</td>
<td>11</td>
<td>0.92</td>
<td>0</td>
<td>11</td>
<td>0.92</td>
</tr>
<tr>
<td>c</td>
<td>11</td>
<td>0.92</td>
<td>0</td>
<td>11</td>
<td>0.92</td>
</tr>
<tr>
<td>d</td>
<td>11</td>
<td>0.92</td>
<td>0</td>
<td>11</td>
<td>0.92</td>
</tr>
<tr>
<td>e</td>
<td>11</td>
<td>0.92</td>
<td>0</td>
<td>11</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Higher threshold (5 direct successions and a dependency of at least 0.9)

```
<table>
<thead>
<tr>
<th>&gt;L</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```

```
| a   | \( \frac{0}{0+1} = 0 \) | \( \frac{11-0}{11+0+1} = 0.92 \) | \( \frac{11-0}{11+0+1} = 0.92 \) | \( \frac{13-0}{13+0+1} = 0.93 \) | \( \frac{5-0}{5+0+1} = 0.83 \) |
| b   | \( \frac{0-11}{0+1+11} = -0.92 \) | \( \frac{0}{0+1} = 0 \) | \( \frac{10-10}{10+10+1} = 0 \) | \( \frac{0-0}{0+0+1} = 0 \) | \( \frac{11-0}{11+0+1} = 0.92 \) |
| c   | \( \frac{0-11}{0+1+11} = -0.92 \) | \( \frac{10-10}{10+10+1} = 0 \) | \( \frac{0}{0+1} = 0 \) | \( \frac{0-0}{0+0+1} = 0 \) | \( \frac{11-0}{11+0+1} = 0.92 \) |
| d   | \( \frac{0-13}{0+1+13} = -0.93 \) | \( \frac{0-0}{0+0+1} = 0 \) | \( \frac{0-0}{0+0+1} = 0 \) | \( \frac{4}{4+1} = 0.80 \) | \( \frac{13-0}{13+0+1} = 0.93 \) |
| e   | \( \frac{0-5}{0+5+1} = -0.83 \) | \( \frac{0-11}{0+1+11} = -0.92 \) | \( \frac{0-11}{0+1+11} = -0.92 \) | \( \frac{0-13}{0+1+13} = -0.93 \) | \( \frac{0}{0+1} = 0 \) |
```
Learning splits and joins

$L = [\langle a, e \rangle^5, \langle a, b, c, e \rangle^{10}, \langle a, c, b, e \rangle^{10}, \langle a, b, e \rangle^1, \langle a, c, e \rangle^1, \langle a, d, e \rangle^{10}, \langle a, d, d, e \rangle^2, \langle a, d, d, d, e \rangle^1]$
Alternative visualization

\[ L = [\langle a, e \rangle^5, \langle a, b, c, e \rangle^{10}, \langle a, c, b, e \rangle^{10}, \langle a, b, e \rangle^1, \langle a, c, e \rangle^1, \langle a, d, e \rangle^{10}, \langle a, d, d, e \rangle^2, \langle a, d, d, d, e \rangle^1] \]
Characteristics of heuristic mining

- Can deal with noise and therefore quite robust.
- Improved representational bias.
- Split and join rules are only considered locally (therefore most of the discovered model are not sound and require repair actions).
Genetic process mining

- event log
- create initial population
- compute fitness
- tournament
- select best individual
- elitism
- parents
- next generation
- crossover
- children
- mutation
- "dead" individuals
- tournament
- termination
- "dead" individuals
- tournament
- create initial population
- compute fitness
Design decisions

- Representation of individuals
- Initialization
- Fitness function
- Selection strategy (tournament and elitism)
- Crossover
- Mutation
Example: crossover
Example: mutation

- Start register request
- Examine thoroughly
- Examine casually
- Decide
- Pay compensation
- Reject request
- Reinitiate request

- Start
- Register request
- Examine casually
- Check ticket
- Reinitiate request

Remove place

Added arc
Characteristics of genetic process mining

• Requires a lot of computing power.
• Can be distributed easily.
• Can deal with noise, infrequent behavior, duplicate tasks, invisible tasks, etc.
• Allows for incremental improvement and combinations with other approaches (heuristics post-optimization, etc.).
Region-based mining

- Two types of regions theory:
  - State-based regions
  - Language-based regions
- All about discovering places (like in the $\alpha$ algorithm!)

$A=\{a_1, a_2, \ldots, a_m\}$

$B=\{b_1, b_2, \ldots, b_n\}$
State-based regions

Two steps:
1. Discover a transition system (different abstractions are possible)
2. Convert transition system into an “equivalent” Petri net.
Step 1: learning a transition system

- past, future, past+future
- sequence, multiset, set abstraction
- limited horizon to abstract further
- filtering e.g. based on transaction type, names, etc.
- labels based on activity name or other features
Past without abstraction (full sequence)

\[ L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle] \]
Future without abstraction

\[ L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle] \]
Past with multiset abstraction

\[ L_1 = \left[ \langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle \right] \]
Only last event matters for state

\[ L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle] \]
Step 2: constructing a Petri net using regions

- a = enter
- b = enter
- c = exit
- d = exit
- e = do not cross
- f = do not cross

R
Example

\[ L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle] \]
Region $R = (X,Y,c)$ corresponding to place $p_R$: $X = \{a_1,a_2,c_1\} = \text{transitions producing a token for } p_R$, $Y = \{b_1,b_2,c_1\} = \text{transitions consuming a token from } p_R$, and $c$ is the initial marking of $p_R$. 
Based idea: enough tokens should be present when consuming

A place is feasible if it can be added without disabling any of the traces in the event log.

for any $\sigma \in L, k \in \{1, \ldots, |\sigma|\}$, $\sigma_1 = hd^{k-1}(\sigma)$, $a = \sigma(k)$, $\sigma_2 = hd^k(\sigma) = \sigma_1 \oplus a$:

$$c + \sum_{t \in X} \partial_{\text{multiset}}(\sigma_1)(t) - \sum_{t \in Y} \partial_{\text{multiset}}(\sigma_2)(t) \geq 0.$$
Example

\[
L_9 = [\langle a, c, d \rangle^{45}, \langle b, c, e \rangle^{42}]
\]

\[
c - y_a \geq 0
\]

\[
c + x_a - (y_a + y_c) \geq 0
\]

\[
c + x_a + x_c - (y_a + y_c + y_d) \geq 0
\]

\[
c - y_b \geq 0
\]

\[
c + x_b - (y_b + y_c) \geq 0
\]

\[
c + x_b + x_c - (y_b + y_c + y_e) \geq 0
\]

\[
c, x_a, \ldots, x_e, y_a, \ldots, y_e \in \{0, 1\}
\]
$$L_9 = [(a, c, d)^{45}, (b, c, e)^{42}]$$

$$R_1 = (\emptyset, \{a, b\}, 1)$$

\[c = y_a = y_b = 1, \quad x_a = x_b = x_c = x_d = x_e = y_c = y_d = y_e = 0\]

$$R_2 = (\{a, b\}, \{c\}, 0)$$

\[x_a = x_b = y_c = 1, \quad c = x_c = x_d = x_e = y_a = y_b = y_d = y_e = 0\]

$$R_3 = (\{c\}, \{d, e\}, 0)$$

\[x_c = y_d = y_e = 1, \quad c = x_a = x_b = x_d = x_e = y_a = y_b = y_c = 0\]

$$R_4 = (\{d, e\}, \emptyset, 0)$$

\[x_d = x_e = 1, \quad c = x_a = x_b = x_c = y_a = y_b = y_c = y_d = y_e = 0\]

$$R_5 = (\{a\}, \{d\}, 0)$$

\[x_a = y_d = 1, \quad c = x_b = x_c = x_d = x_e = y_a = y_b = y_c = y_e = 0\]

$$R_6 = (\{b\}, \{e\}, 0)$$

\[x_b = y_e = 1, \quad c = x_a = x_c = x_d = x_e = y_a = y_b = y_c = y_d = 0\]
Model

$L_9 = [\langle a, c, d \rangle^{45}, \langle b, c, e \rangle^{42}]$

$R_1 = (\emptyset, \{a, b\}, 1)$
$c = y_a = y_b = 1, x_a = x_b = x_c = x_d = x_e = y_c = y_d = y_e = 0$

$R_2 = (\{a, b\}, \{c\}, 0)$
$x_a = x_b = y_c = 1, c = x_c = x_d = x_e = y_a = y_b = y_d = y_e = 0$

$R_3 = (\{c\}, \{d, e\}, 0)$
$x_c = y_d = y_e = 1, c = x_a = x_b = x_d = x_e = y_a = y_b = y_c = 0$

$R_4 = (\{d, e\}, \emptyset, 0)$
$x_d = x_e = 1, c = x_a = x_b = x_c = y_a = y_b = y_c = y_d = y_e = 0$

$R_5 = (\{a\}, \{d\}, 0)$
$x_a = y_d = 1, c = x_b = x_c = x_d = x_e = y_a = y_b = y_c = y_e = 0$

$R_6 = (\{b\}, \{e\}, 0)$
$x_b = y_e = 1, c = x_a = x_c = x_d = x_e = y_a = y_b = y_c = y_d = 0$
Characteristics of region-based mining

- Can be used to discover more complex control-flow structures.
- Classical approaches need to be adapted (overfitting!).
- Representational bias can be parameterized (e.g., free-choice nets, label splitting, etc.).
- Problems dealing with noise.
Other approaches, e.g. fuzzy mining
Evaluating the discovered process

**Structure:** Is this the simplest model (Occam's Razor)?

**Fitness:** Is the event log possible according to the model?

**Precision:** Is the model not underfitting (allow for too much)?

**Generalization:** Is the model not overfitting (only allow for the “accidental” examples)?

**Generalization:** Is the model not overfitting (only allow for the “accidental” examples)?