Process Mining

Lesson 8 – Advanced Techniques

Doc. Cognini Riccardo

Characteristic of Techniques

- Represent BIAS (problems in notation)

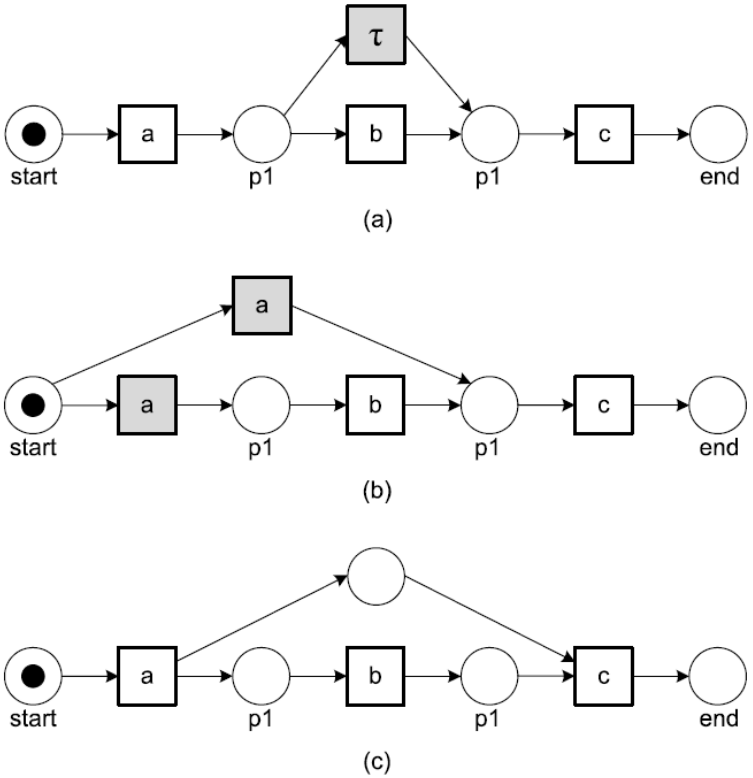
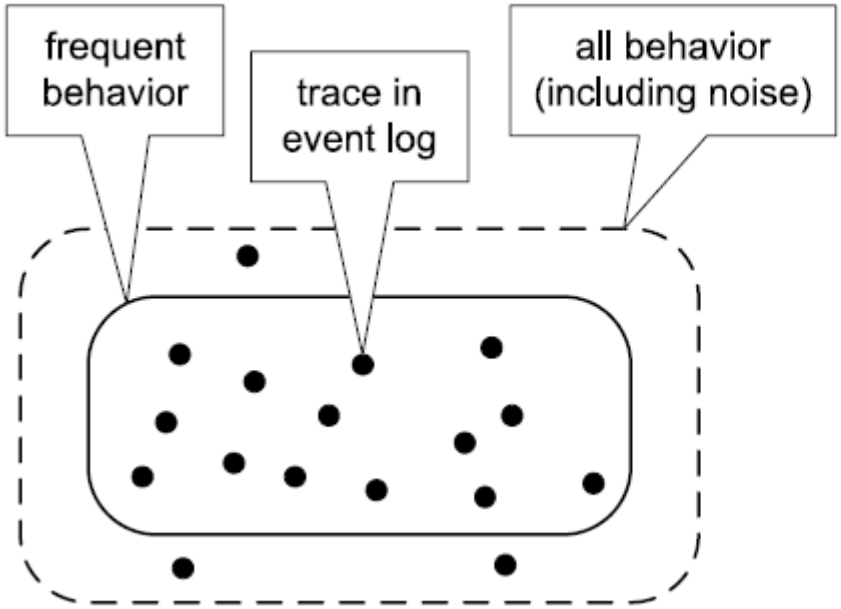
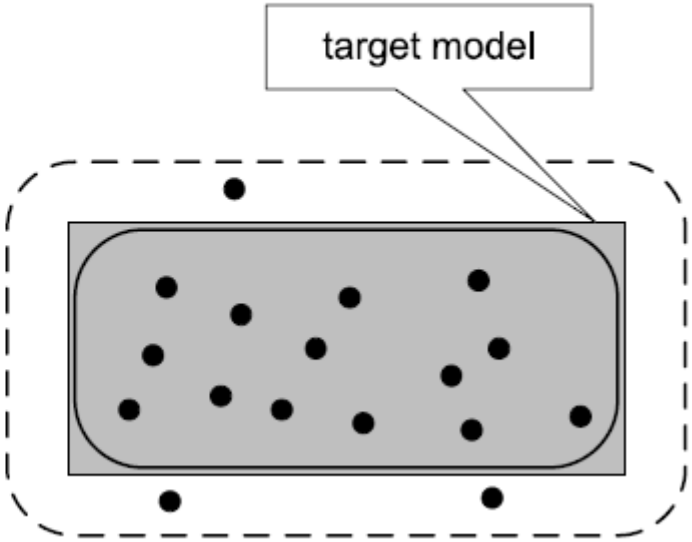


Fig. 5.21 Three WF-nets for the event log $L_{11} = [\langle a, b, c\rangle^{20}, \langle a, c\rangle^{30}]$

Characteristic of Techniques

- Ability to deal with noise





Characteristic of Techniques

- Completeness – of the logs – Assumpition

- Approach Used:
  - ❑ **Direct Algorithm**
  - ❑ **Two-Phases Approach** (low and high level model)
  - ❑ **Computational Intelligence Approaches** (genetic programming, machine learning, neural networks) – evolutionary approach with iterations
  - ❑ **Partial Approaches**, discovery rules, not complete models!

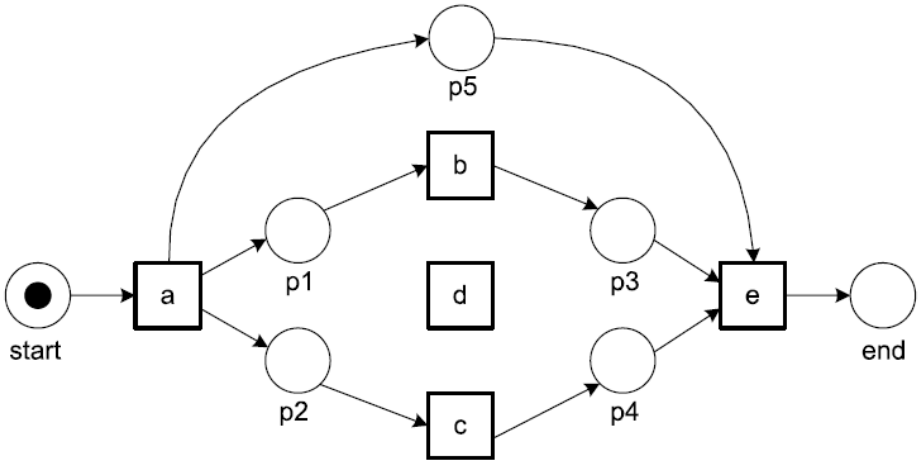
Some Approaches

- Heuristic Mining
- Genetic Process Mining
- Region-Based Mining

Heuristic Mining

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

If we assume the three traces with frequency one to be noise, then the remaining 37 traces in the log correspond to valid sequences of the C-net in Fig. 6.2. Before explaining how to derive such a C-net, we first apply the α-algorithm to event log L. The result is shown in Fig. 6.3.



Alpha Algoritm

Heuristic Mining

Heuristic Mining

Basic Concept of the Algorithm

Table 6.1 Frequency of the "directly follows" relation in event log L: $|x >_L y|$ is the number of times x is directly followed by y in L

$\|>_L\|$	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

Table 6.2 Dependency measures between the five activities based on event log L

$\|\Rightarrow_L\|$	a	b	c	d	e
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	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0}{0+1} = 0$	$\frac{10-10}{10+10+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
c	$\frac{0-11}{0+11+1} = -0.92$	$\frac{10-10}{10+10+1} = 0$	$\frac{0}{0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
d	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0-0}{0+0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{4}{4+1} = 0.80$	$\frac{13-0}{13+0+1} = 0.93$
e	$\frac{0-5}{0+5+1} = -0.83$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0}{0+1} = 0$

Heuristic Mining

Basic Concept of the Algorithm

Fig. 6.4 Dependency graph using a threshold of 2 for $|>_L|$ and 0.7 for $|\Rightarrow_L|$: *each arc* shows the $|>_L|$ value and the $|\Rightarrow_L|$ value between brackets. For example, $|a >_L d| = 13$ and $|a \Rightarrow_L d| = 0.93$

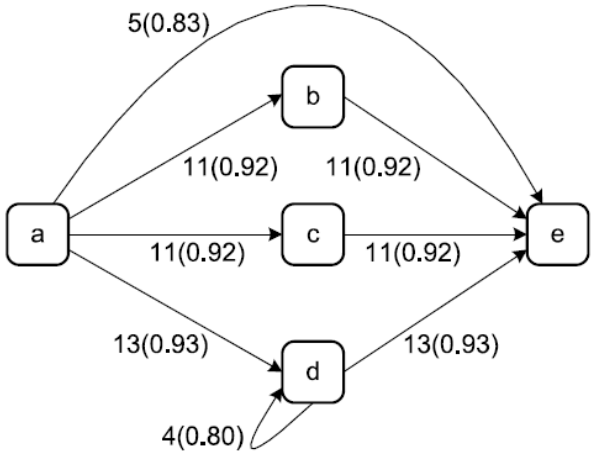
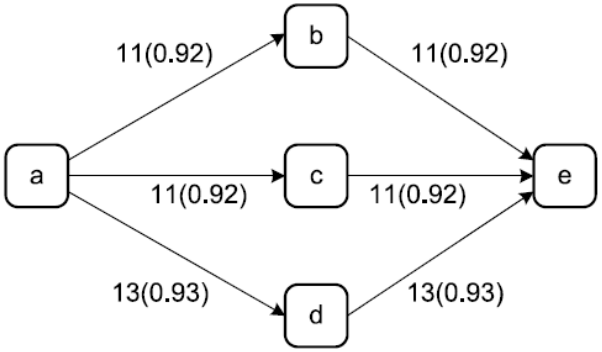


Fig. 6.5 Dependency graph using a threshold of 5 for $|>_L|$ and 0.9 for $|\Rightarrow_L|$. The self loop involving d disappeared because $|d >_L d| = 4 < 5$ and $|d \Rightarrow_L d| = 0.80 < 0.9$. The connection between a and e disappeared because $|a \Rightarrow_L e| = 0.83 < 0.9$

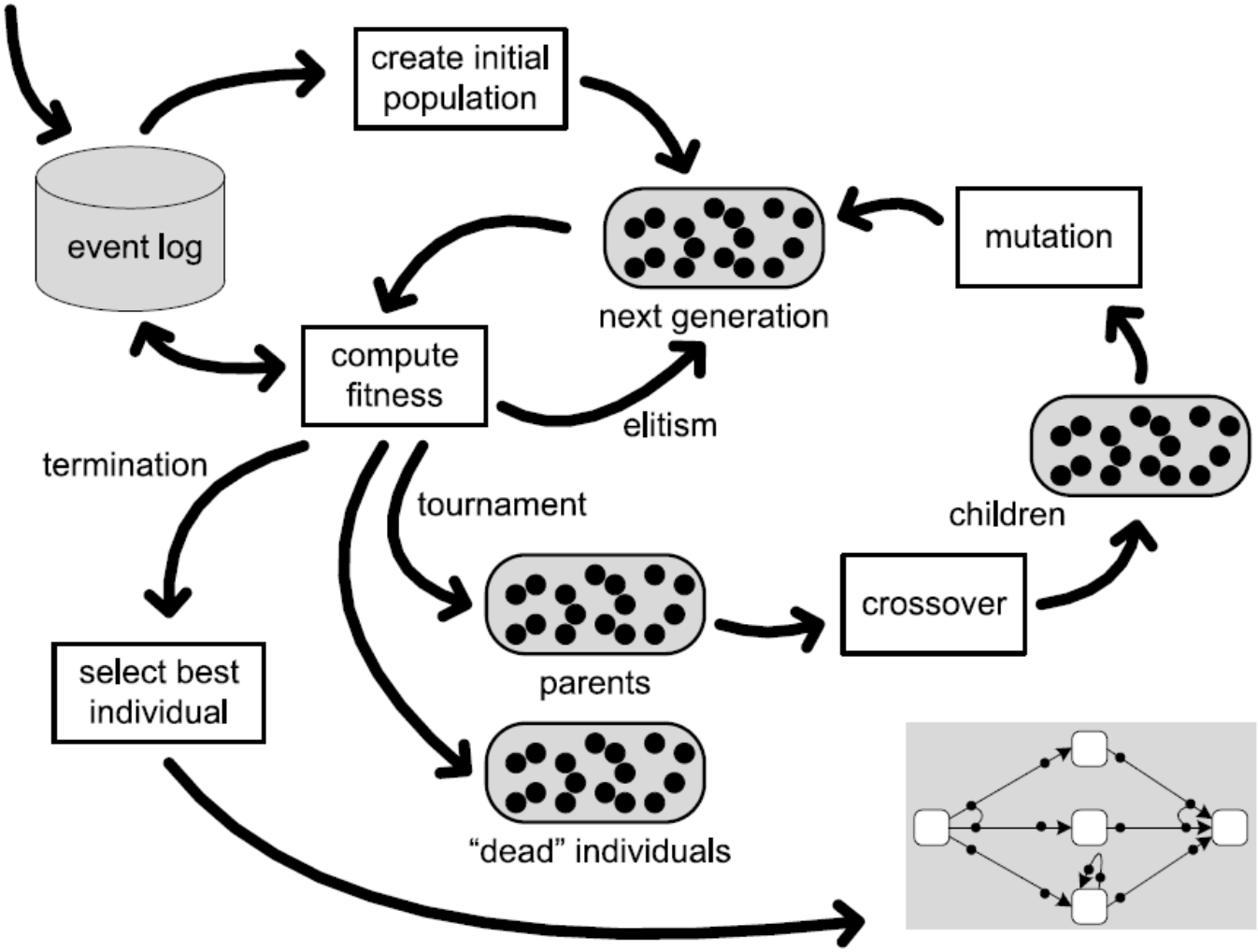


Heuristic Mining

Basic Concept of the Algorithm

- Activities should not be excluded by the threshold (but just if the trace is not considered)
- More than one model can be obtained changing the threshold

Genetic Process Mining



Region Based Mining

First Step aims to discover the transition system from the logs

Fig. 6.12 Transition system $TS_{L_1, l_1^{state}()}$ derived from $L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$ using $l_1^{state}(\sigma, k) = hd^k(\sigma)$

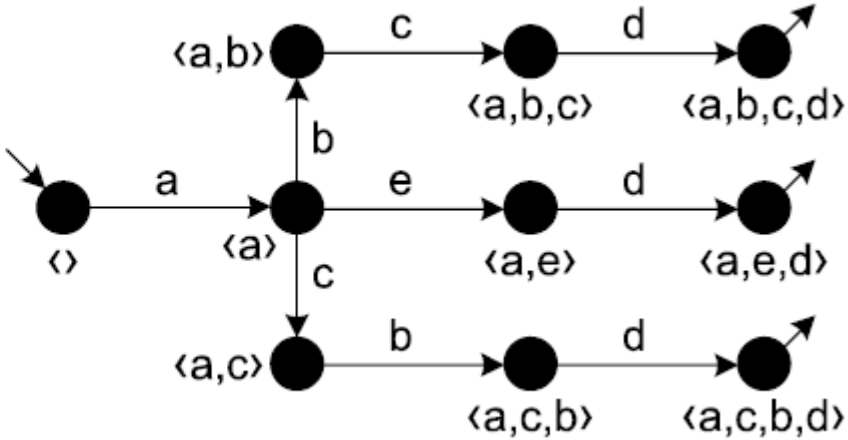
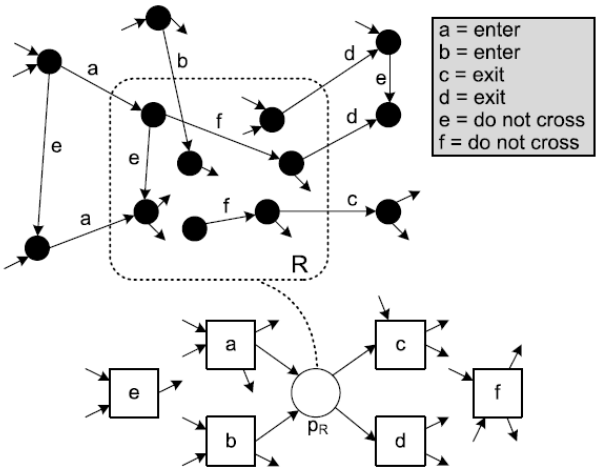


Fig. 6.16 Region R corresponding to place p_R. All activities can be classified into *entering* the region (a and b), *leaving* the region (c and d), and *non-crossing* (e and f)



QUESTIONS?