Exercise 4

Business Intelligence *

Raphael Mitsch (a1006529)

June 18, 2019

1 Task 4-1: Process Discovery - α-algorithm (3 points)

Since we start with workflow traces already, we apply the second step - analyzing order relations between tasks per trace. This yields

$$Case_1 = A >_W B, B > C, C > E, E > H \tag{1}$$
$$Case_2 = A > B, B > D, D > E, E > H \tag{2}$$
$$Case_3 = A > B, B > D, D > F, F > I, I > H \tag{3}$$
$$Case_4 = A > B, B > D, D > F, F > I, I > H \tag{4}$$
$$Case_5 = A > B, B > C, C > G, G > H \tag{5}$$
$$Case_7 = A > B, B > C, C > E, E > H \tag{6}$$

We continue with the analysis of order relations over the entire workflow W:

*University of Vienna, fall semester 2018/2019.

$$A \rightarrow_W B \tag{7}$$
$$B \rightarrow_W C, B \rightarrow_W D \tag{8}$$
$$C \rightarrow_W E, C \rightarrow_W G \tag{9}$$
$$D \rightarrow_W E, D \rightarrow_W F \tag{10}$$
$$E \rightarrow_W H \tag{11}$$
$$F \rightarrow_W I \tag{12}$$
$$I \rightarrow_W H \tag{13}$$
$$G \rightarrow_W H \tag{14}$$
$$A\#_W C, A\#_W D, A\#_W E, A\#_W F, A\#_W I, A\#_W H, A\#_W G \tag{15}$$
$$B\#_W E, B\#_W H, B\#_W I, B\#_W F, B\#_W G \tag{16}$$
$$C\#_W D, C\#_W F, C\#_W I, C\#_W H \tag{17}$$
$$D\#_W I, D\#_W H, D\#_W G \tag{18}$$
$$E\#_W F, E\#_W I, E\#_W G \tag{19}$$
$$F\#_W G, F\#_W H \tag{20}$$
$$G\#_W I \tag{21}$$

Note that we have no occurrences of $a||_W b$, i. e. there are no two tasks a, b to which both $a >_W b$ and $b >_W a$ applies.

Finally, we derive the associated Petri net. We use the `dot` language for defining and visualizing the net in Figure 2. The corresponding code is attached in the uploaded archive.

2 Task 4-2: Process Discovery – Heuristic Miner (3 points)

We are using the same log as in exercise 1 (see equation set 1) and apply the Heuristic Mining concept as introduced in the lecture. In the following we determine the number of connections $|a >_W b|$ for all pairs a, b and the corresponding $a \Rightarrow_W b = \frac{|a>_W b| - |b>_W a|}{|a>_W b| + |b>_W a| + 1}$. Note that, as already mentioned, there are no two tasks a, b to which both $a >_W b$ and $b >_W a$ applies. Hence we can simplify to $a \Rightarrow_W b = \frac{|a>_W b|}{|a>_W b| + 1}$.

$$|A >_W B| = 6 = \frac{6}{7} \tag{22}$$

$$|B >_W C| = 3 = \frac{3}{4} \tag{23}$$

$$|B >_W D| = 3 = \frac{3}{4} \tag{24}$$

$$|C >_W E| = 2 = \frac{2}{3} \tag{25}$$

$$|C >_W G| = 1 = \frac{1}{2} \tag{26}$$

$$|D >_W E| = 1 = \frac{1}{2} \tag{27}$$

$$|D >_W F| = 2 = \frac{2}{3} \tag{28}$$

$$|E >_W H| = 3 = \frac{3}{4} \tag{29}$$

$$|F >_W I| = 2 = \frac{2}{3} \tag{30}$$

$$|I >_W H| = 2 = \frac{2}{3} \tag{31}$$

$$|G >_W H| = 1 = \frac{1}{2} \tag{32}$$

All pairs not mentioned in equation set are of strength zero, i. e. these pairs of tasks didn't occur in the workflow log. Once more we visualize the resulting graph with `dot` - see Figure 3.

3 Task 4-3: Conformance Checking (3 points))

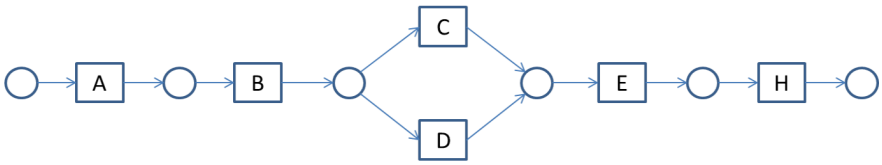

Figure 1: Petri net to be used for conformance checking in task 3.

We consider the following traces:

Instance	Trace
120	<A,B,C,E,H >
120	<A,B,D,E,H >
20	<A,B,D,H >
20	<A,B,C,H >

These traces are to be checked against the Petri net seen in Figure 3. To do so, we count the numbers of

- missing,
- remaining,
- consumed and
- produced

tokens (respectively referred to as m_i, r_i, c_i, and p_i) and compute the conformance fitness $f = \frac{1}{2}(1 - \frac{\sum_{i=1}^{k} n_i m_i}{\sum_{i=1}^{k} n_i c_i}) + \frac{1}{2}(1 - \frac{\sum_{i=1}^{k} n_i r_i}{\sum_{i=1}^{k} n_i p_i})$.

1. The first trace, $< A, B, C, E, H >$ is in full conformance with the specified Petri net and thus $m_1 = 0, r_1 = 0, c_1 = 7, p_1 = 7$.
2. $< A, B, D, E, H >$ yields the same result (the only difference being task D instead of C, which is an acceptable sequence). Hence $m_2 = 0, r_2 = 0, c_2 = 7, p_2 = 7$.
3. $< A, B, D, H >$ is in violation, since $\neg(H >_W D)$. This implies one missing token, since we proceed from E to H and therefore have no remaining tokens. $m_3 = 1, r_3 = 0, c_3 = 5, p_3 = 6$.
4. $< A, B, C, H >$ violates the given Petri net in the same way as sequence 3 does, hence we can assign the same metrics. $m_4 = 1, r_4 = 0, c_4 = 5, p_4 = 6$.

We compute our conformance score f as

$$\frac{1}{2}(1 - \frac{120 \cdot (0) \cdot 2 + 20 \cdot (1) \cdot 2}{120 \cdot (7) \cdot 2 + 20 \cdot (6) \cdot 2}) + \tag{33}$$

$$\frac{1}{2}(1 - \frac{120 \cdot (0) \cdot 2 + 20 \cdot (0) \cdot 2}{120 \cdot (7) \cdot 2 + 20 \cdot (7) \cdot 2}) \tag{34}$$

$$= \frac{1}{2}(1 - \frac{40}{1920}) + \frac{1}{2} \tag{35}$$

$$= 0.489 + 0.5 = 0.989 \tag{36}$$

Although two of our four traces are not valid, our overall conformance score is relatively high due to the two correct traces being correct and having a drastically higher number of instances and therefore more weight in the computation of the conformance score f.

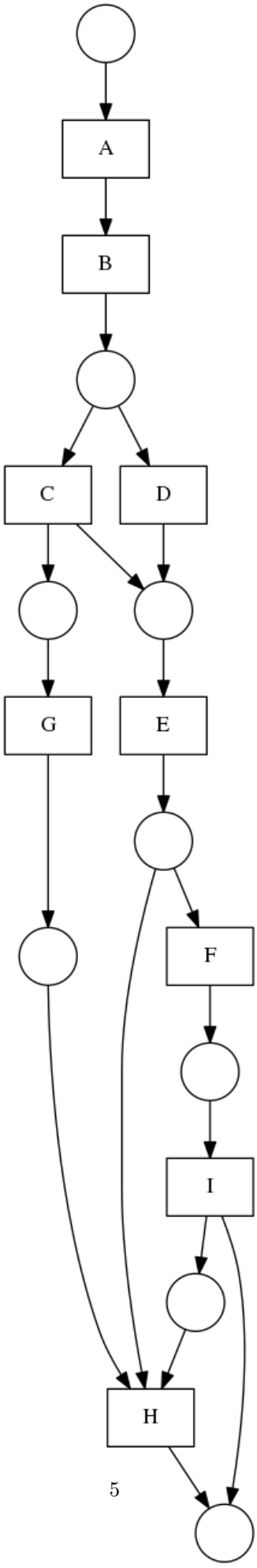

Figure 2: Petri net resulting from α-algorithm for task 1.

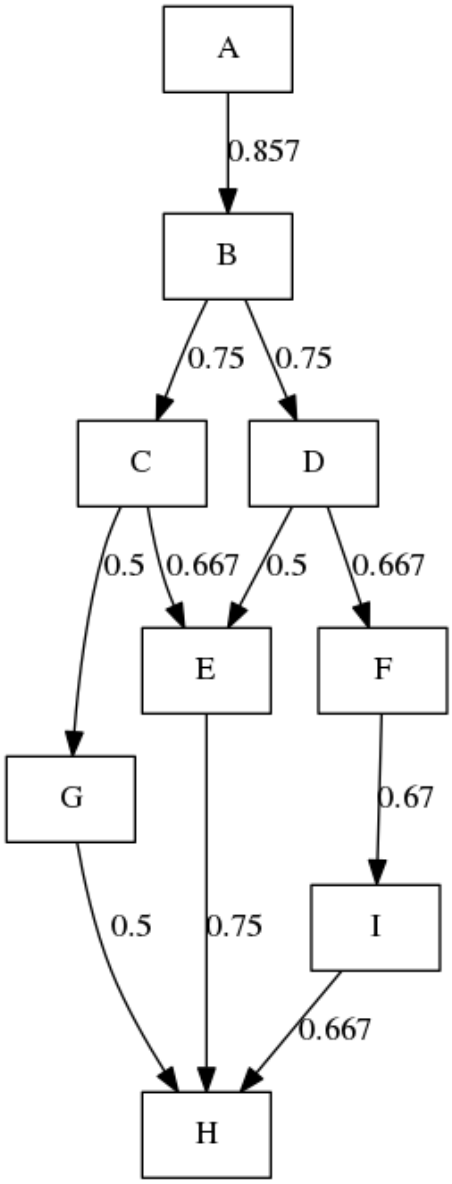

Figure 3: Graph resulting from heuristic mining for task 2.