

Process Mining

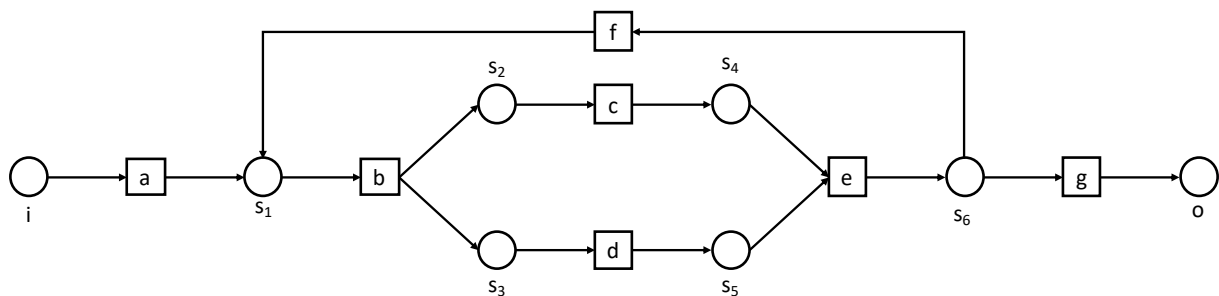
Wintersemester 2020 / 2021

Übung 8

α -Algorithmus • Heuristic Miner

Aufgabe 1: Footprint-Matrix

Gegeben sei das folgende Workflow-Netz:



Erstellen Sie eine Footprint-Matrix zu diesem Petri-Netz.

Lösung

	a	b	c	d	e	f	g
a	#	→	#	#	#	#	#
b	←	#	→	→	#	←	#
c	#	←	#		→	#	#
d	#	←		#	→	#	#
e	#	#	←	←	#	→	→
f	#	→	#	#	←	#	#
g	#	#	#	#	←	#	#

Aufgabe 2: α -Algorithmus

Bestimmen Sie für das unten abgebildete vereinfachte Ereignislog das zugehörige Workflow-Netz mit Hilfe des α -Algorithmus.

t	Fallnummer	Aktivität
1	2	a
2	3	a
3	2	c
4	2	b
5	1	a
6	3	e
7	3	d
8	1	b
9	1	c
10	2	d
11	1	d

Lösung

1. Menge aller Aktivitäten in dem Ereignislog: $T_L = \{a, b, c, d, e\}$
2. Erstelle die Footprint-Matrix:

	a	b	c	d	e
a	#	\rightarrow	\rightarrow	#	\rightarrow
b	\leftarrow	#	\parallel	\rightarrow	#
c	\leftarrow	\parallel	#	\rightarrow	#
d	#	\leftarrow	\leftarrow	#	\leftarrow
e	\leftarrow	#	#	\rightarrow	#

3. Menge aller Aktivitäten in dem Ereignislog, die in einem Fall zuerst auftreten:
 $T_I = \{a\}$.
4. Menge aller Aktivitäten in dem Ereignislog, die in einem Fall zuletzt auftreten:
 $T_O = \{d\}$.

5. Menge aller potenziellen Verbindungen:

i) Füge alle Sequenzen zu X_L^i hinzu:

$$X_L^i = \{(\{a\}, \{b\}), (\{a\}, \{c\}), (\{a\}, \{e\}), (\{b\}, \{d\}), (\{c\}, \{d\}), (\{e\}, \{d\})\}$$

ii) Füge die exklusive Auswahl zu X_L^{ii} hinzu:

$$X_L^{ii} = \{(\{a\}, \{b, e\}), (\{a\}, \{c, e\})\}$$

iii) Füge die einfache Zusammenführung zu X_L^{iii} hinzu:

$$X_L^{iii} = \{(\{b, e\}, \{d\}), (\{c, e\}, \{d\})\}$$

iv) Vereinige die Teillösungen $X_L^i, X_L^{ii}, X_L^{iii}$:

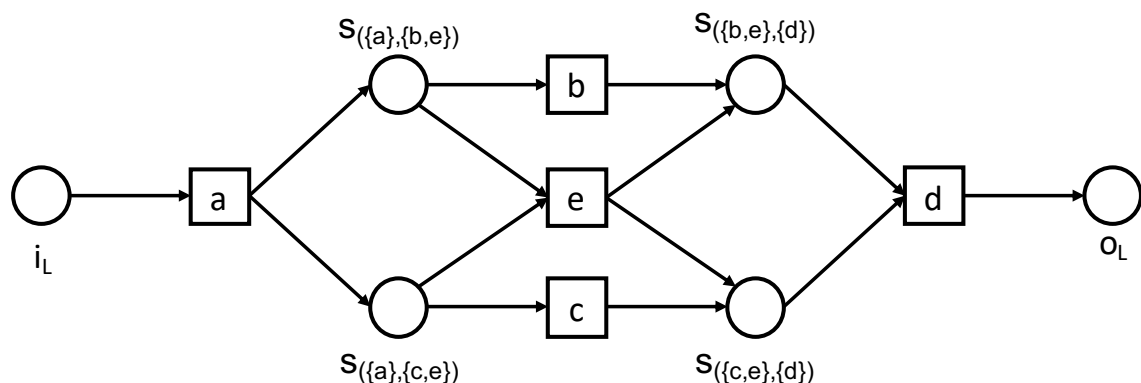
$$X_L = \{(\{a\}, \{b\}), (\{a\}, \{c\}), (\{a\}, \{e\}), (\{b\}, \{d\}), (\{c\}, \{d\}), (\{e\}, \{d\}), (\{a\}, \{b, e\}), (\{a\}, \{c, e\}), (\{b, e\}, \{d\}), (\{c, e\}, \{d\})\}$$

6. Eliminiere alle überflüssigen Elemente:

i) Entferne die Tupel $(\{a\}, \{b\})$ und $(\{a\}, \{e\})$, da $(\{a\}, \{b, e\})$ enthalten ist sowie zusätzlich $(\{a\}, \{c\})$ durch $(\{a\}, \{c, e\})$.ii) Entferne die Tupel $(\{b\}, \{d\})$ und $(\{e\}, \{d\})$, da $(\{b, e\}, \{d\})$ enthalten ist sowie zusätzlich $(\{c\}, \{d\})$ durch $(\{c, e\}, \{d\})$.Somit ist $Y_L = \{(\{a\}, \{b, e\}), (\{a\}, \{c, e\}), (\{b, e\}, \{d\}), (\{c, e\}, \{d\})\}$.7. Erzeuge alle Stellen des Workflow-Netzes aus Y_L inklusive der Start- und Endstelle: $S_L = \{s(\{a\}, \{b, e\}), s(\{a\}, \{c, e\}), s(\{b, e\}, \{d\}), s(\{c, e\}, \{d\}), i_L, o_L\}$

8. Erzeuge die Flussrelation zwischen den Stellen und Transitionen:

$$F_L = \{(a, s(\{a\}, \{b, e\})), (s(\{a\}, \{b, e\}), b), (s(\{a\}, \{b, e\}), e), (a, s(\{a\}, \{c, e\})), (s(\{a\}, \{c, e\}), c), (s(\{a\}, \{c, e\}), e), (b, s(\{b, e\}, \{d\})), (e, s(\{b, e\}, \{d\})), (s(\{b, e\}, \{d\}), d), (c, s(\{c, e\}, \{d\})), (e, s(\{c, e\}, \{d\})), (s(\{c, e\}, \{d\}), d), (i_L, a), (d, o_L)\}$$

9. Somit ergibt sich das Workflow-Netz $\alpha(L) = (S_L, T_L, F_L)$:

Aufgabe 3: Heuristic Miner

Bestimmen Sie für das unten abgebildete vereinfachte Ereignislog das zugehörige Workflow-Netz mit Hilfe des Heuristic Miner.

t	Fallnummer	Aktivität
1	2	a
2	3	a
3	2	c
4	2	b
5	1	a
6	3	e
7	3	d
8	1	b
9	1	c
10	2	d
11	1	d

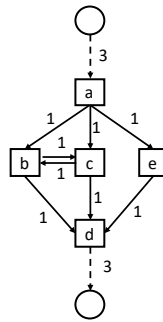
Lösung

1. Directly-Follows Graph erstellen:

i) Häufigkeitsbeziehungen identifizieren:

Relation	Häufigkeit
(-,a)	3
(a,b)	1
(a,c)	1
(a,e)	1
(b,c)	1
(b,d)	1
(c,b)	1
(c,d)	1
(e,d)	1
(d,-)	3

ii) Zugehörigen Directly-Follows-Graph (DFG) zeichnen:



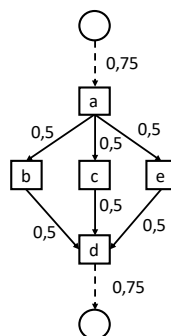
2. Abhängigkeitsmaß berechnen:

Relation	Häufigkeit	Abhängigkeitsmaß
(-,a)	3	0,75
(a,b)	1	0,5
(a,c)	1	0,5
(a,e)	1	0,5
(b,c)	1	0
(b,d)	1	0,5
(c,b)	1	0
(c,d)	1	0,5
(e,d)	1	0,5
(d,-)	3	0,75

3. Eventuell Filter anwenden:

$$|\Rightarrow_L \geq 0,5|$$

4. Abhängigkeitsgraph zeichnen:



Aufgabe 4: Wahr oder Falsch

Beurteilen Sie welche der folgenden Aussagen wahr bzw. falsch sind:

- a) Der α -Algorithmus, wird so genannt, da es der zur Zeit beste Process-Discovery Algorithmus ist und kaum Schwächen besitzt.
- b) Je genauer ein Process-Discovery Modell die Ereignisse abbildet, desto besser.

Lösung

- a) **Falsch.** Der α -Algorithmus verdankt seinen Namen der Tatsache, dass es einer der ersten Verfahren für die Entdeckung eines Ablaufmodells in Form eines Prozessmodells war.
Schwächen des α -Algorithmus sind unter anderem:
 - Kurze Schleifen können nicht erkannt werden
 - Häufigkeitsbeziehungen werden nicht beachtet
 - Schwierigkeiten mit Rauschen
- b) **Falsch.** Ein Modell, das Traces zu genau abbildet, erlaubt es nicht leicht abweichende Traces vom Modell darzustellen. Die Traces werden vom Modell auswendig gelernt und lassen keine Generalisierung zu. Es kommt zu einem *Overfitting*. Neue Traces sind möglicherweise nicht von einem solchen Modell abbildbar. Eine genaue Abbildung aller Traces würde außerdem dazu führen, dass selten auftretende Variationen (möglicherweise Fehlaufzeichnungen) vom Modell abgebildet werden.