Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

Department of Computer Science

Process Analytics

Prof. Dr. Agnes Koschmider
Dominik Janssen

Advanced Process Mining

Summer term 2020

Exercise sheet 2

4D of Quality • DFG • Heuristic Miner

Exercise 1: Four dimensions of quality

Given the following event log, discover three process models that fulfil the following quality properties:

a) **fitness**: high **precision**: high **generalisation**: low **simplicity**: low

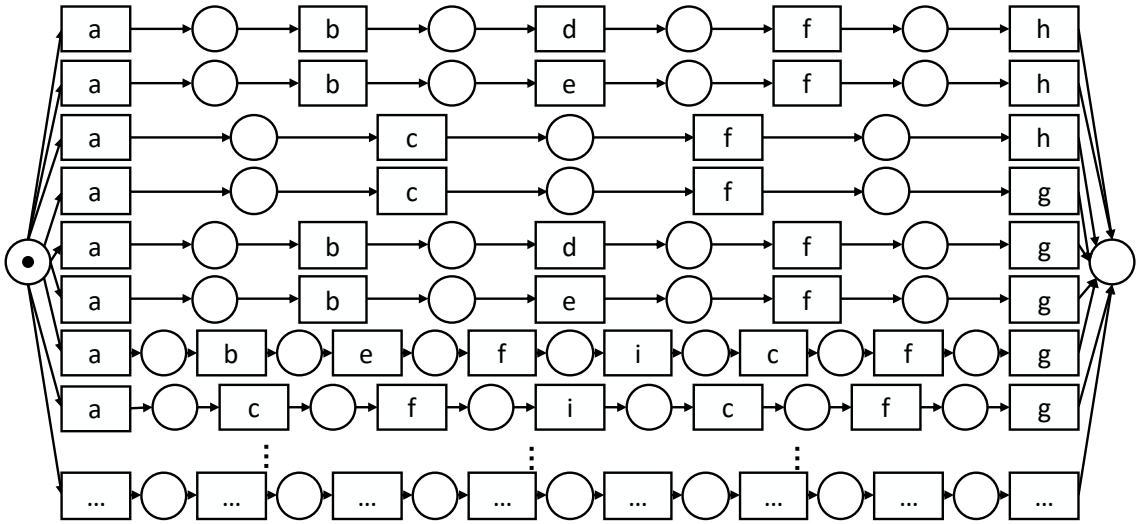
b) **fitness**: low **precision**: high **generalisation**: low **simplicity**: high

c) **fitness**: high **precision**: low **generalisation**: high **simplicity**: high

#	Trace
342	abdfh
200	abefh
101	acfh
62	acfg
55	abdfg
17	abefg
16	abeficfg
13	acficfg
8	abefibdfibcfh
7	acficficficficfg

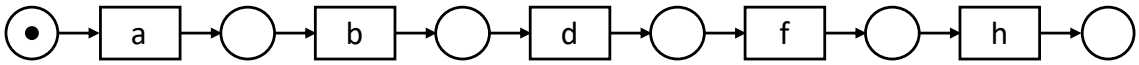
Solution

a) **fitness**: high **precision**: high **generalisation**: low **simplicity**: low



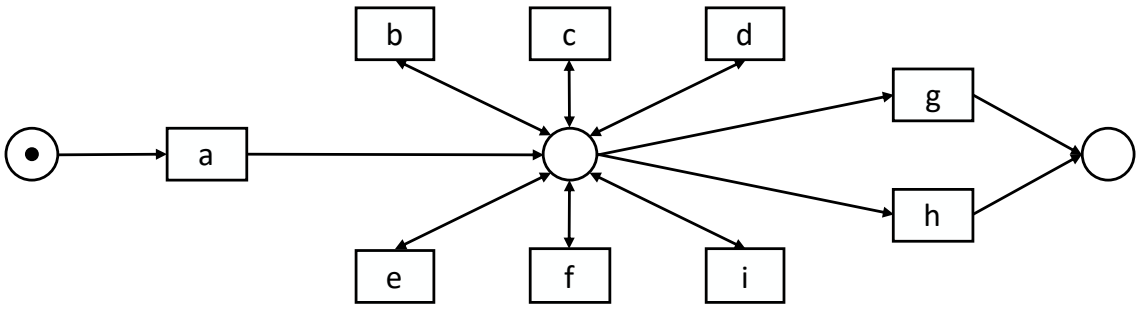
This model is a so called *enumerating model*. It includes all the recorded sequences. Each trace has its own sequential process fragment.

b) **fitness**: low **precision**: high **generalisation**: low **simplicity**: high



This model only covers the most frequent trace. None of the other recorded traces are represented by this model.

c) **fitness**: high **precision**: low **generalisation**: high **simplicity**: high



The fitness is perfect, since all recorded traces can be replayed by the model. However the precision is very bad, because this model allows much more behaviour than was recorded. The model shown above is a variant of the so called *flower model*.

Exercise 2: Directly Follows Graph

#	Trace
342	abdfh
200	abefh
101	acfh
62	acfg
55	abdfg
17	abefg
16	abeficfg
13	acficfg
8	abefibdfibcfh
7	acficficficficfg
1	bcdh

a) What appropriate filtering criteria might be applied on this log?

b) Create a Directly-Follows Graph on the following event log

Solution

a) These are some filtering criteria that can be appropriate:

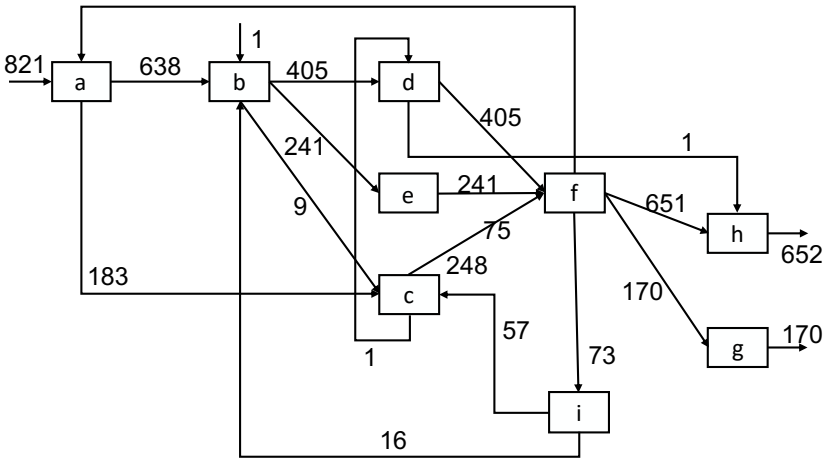
- Remove all events that have more than 5 activities
- Remove all events that do not start with activity 'a'
- Remove all traces that occurred less than 10 times

b) In order to get a Directly-Follows Graph, first an overview has to be made about the activities that follow one another directly. Their quantity has to be counted and finally a graph can be drawn from the data:

- Create a table with pairs of consecutive activities and their quantity:

$>_L$	a	b	c	d	e	f	g	h	i
a		638	183						
b			9	405	241				
c				1		248			
d						405		1	
e						241			
f							170	651	73
g									
h									
i		16	57						

- Draw the Directly-Follows Graph from the collected information in the table:



Exercise 3: Heuristic Miner

#	Trace
342	abdfh
200	abefh
101	acfh
62	acfg
55	abdfg
17	abefg
16	abeficfg
13	acficfg
8	abefibdfibcfh
7	acficficficficfg
1	bcdh

a) Discover a process model from this event log with the Heuristics Miner algorithm

b) Is the dependency matrix always symmetrical?

c) If you multiply the occurrence of all observed traces by 1000 and the threshold will remain the same at 0.1 $>\Rightarrow_L>$ 0.95, does this change the discovered Petri net?

Solution

a) 1. Construct Directly-Follows Matrix:

$>_L$	a	b	c	d	e	f	g	h	i
a		638	183						
b			9	405	241				
c				1		248			
d						405		1	
e						241			
f							170	651	73
g									
h									
i		16	57						

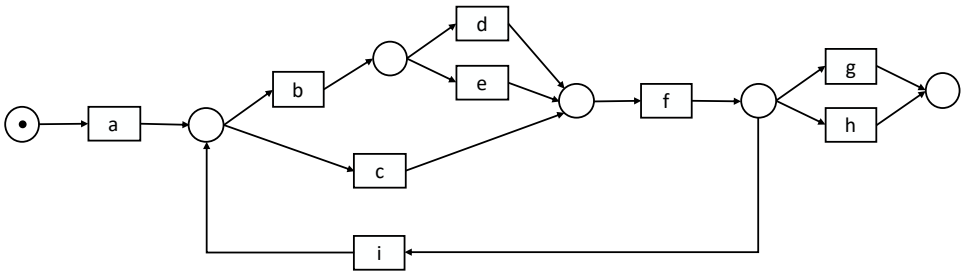
2. Construct Dependency Matrix:

$\Rightarrow_L$	a	b	c	d	e	f	g	h	i
a		0.998	0.995						
b	-0.998		0.900	0.998	0.996				-0.941
c	-0.995	-0.900		0.500		0.996			-0.983
d		-0.998	-0.500			0.998		0.500	
e		-0.996				0.996			
f			-0.996	-0.998	-0.996		0.994	0.998	0.986
g						-0.994			
h				-0.500		-0.998			
i		0.941	0.983			-0.986			

3. Define a threshold and filter the dependency matrix.
Set the threshold to 0.1 $>\Rightarrow_L>$ 0.95

$\Rightarrow_L$	a	b	c	d	e	f	g	h	i
a		0.998	0.995						
b	-0.998		0.900	0.998	0.996				-0.941
c	-0.995	-0.900		0.500		0.996			-0.983
d		-0.998	-0.500			0.998		0.500	
e		-0.996				0.996			
f			-0.996	-0.998	-0.996		0.994	0.998	0.986
g						-0.994			
h				-0.500		-0.998			
i		0.941	0.983			-0.986			

4. Construct the Petri net:



b) Mirroring the values on the diagonal of the matrix will result in the same values with the opposite sign.

c) Yes, the resulting Petri net will now also include arcs that were filtered in the dependency matrix before, due to a value below the threshold. Although the relative frequency stayed exactly the same, the former excluded arcs are now part of the discovered model.