

Instruction 3 – Assignment 3

Note: The answers for this assignment should be posted at the Study web in the folder "1BM45_Process mining -> assignment 3 -> Submitted"

Deadline: 27-2-2008 at 11 am

Question 1:

Have a look at the event log and the mined models in Figure 1. Rank the models based on how well, in your opinion, they describe and summarize the behavior in the log. Explain why you have placed a certain model in a given position in the rank.

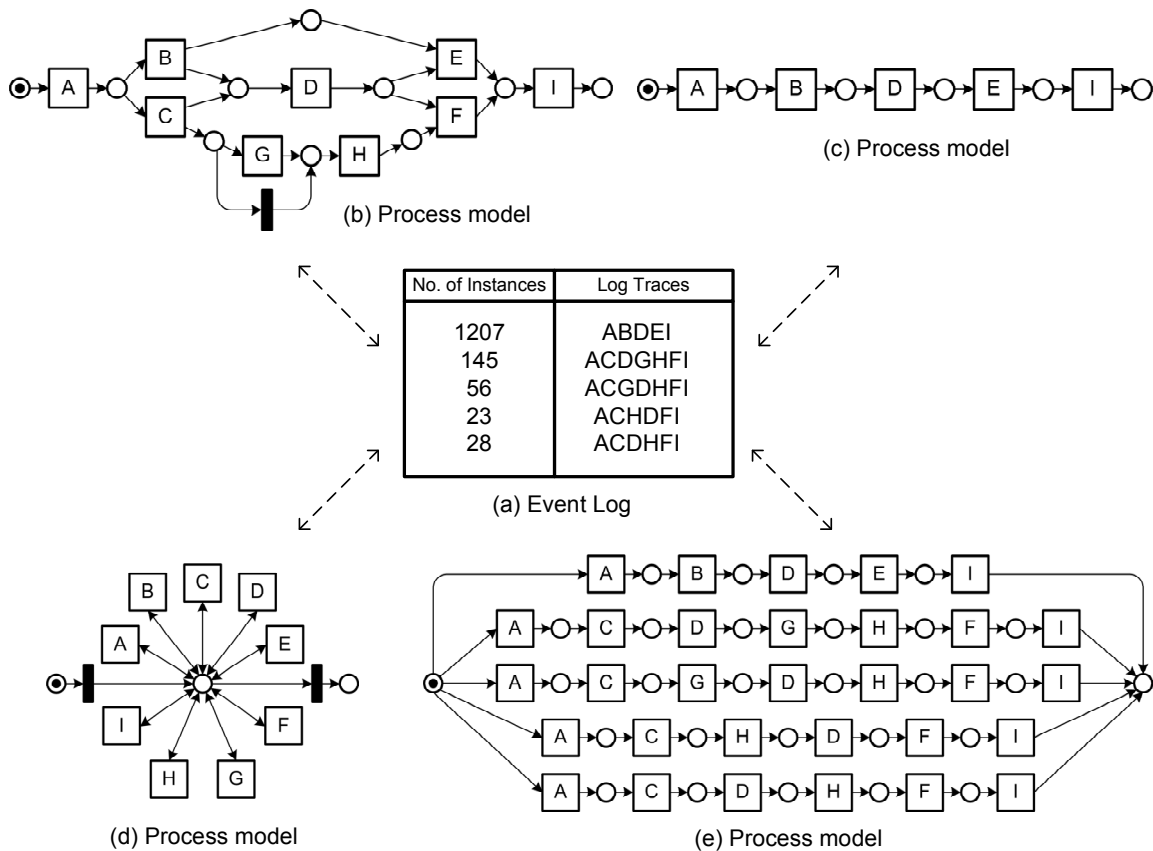


Figure 1 - An event log and some mined models.

Question 2:

For this question, do the following:

1. Fill in Table 1 below by indicating which constructs are supported by each control-flow mining algorithm in the table.
2. Explain why you think that a given algorithm does not support a given construct. Do this for every control-flow mining algorithm in Table 1.

	Control-Flow Mining Algorithm					
Construct Type	Alpha Algorithm	Heuristics Miner	Genetic Algorithm	Duplicate Tasks GA	Fuzzy Miner	Multi-Phase Macro
Sequence						
XOR-split/join						
AND-split/join						
OR-split/join						
Loops						
Non-Free Choice						
Invisible Tasks						
Duplicate Tasks						
Can handle noisy logs?						

Table 1 Overview of the structural constructs the mining algorithms support plus their robustness to noisy logs. (✓= yes and ✗= no)

Question 3:

For this question, do the following:

(Preparation steps)

- Download the log at [http://tabu.tm.tue.nl/wiki/ media/tutorial/repairexample.zip?id=tutorials&cache=cache](http://tabu.tm.tue.nl/wiki/media/tutorial/repairexample.zip?id=tutorials&cache=cache)
(This is the log used by the ProM Tutorial at www.processmining.org -> Tutorials)
- Filter the log to contain only:
 - a. Complete event types;
 - b. The process instances that start with the task "Register (complete)" and end with the task "Archive Repair (complete)".
 Save this filtered log.

(Actual sub-questions to which the answers should be included)

1. Mine the filtered log obtained after executing preparation steps above. Perform the mining with each of the following algorithms:
 - a. Alpha Algorithm
 - b. Heuristics Miner
 - c. Fuzzy Miner
 - d. Multi-Phase Macro
 Include a screenshot of the mined model for each of these algorithms.
2. Based on the models mined in sub-question (1), discuss the differences in the mined models. The discussion should include:
 - a. The notation used by the algorithm to show a mined model;
 - b. The constructs present in the mined model;
 - c. One or two sentences that summarize the main difference between a given model and the other ones.

Question 4:

Many of the mining algorithms implemented in the ProM framework aim at mining a *workflow net*. These nets have the characteristic that they have a single *start* place and a single *end* place. Furthermore, the start place must not have any input (or incoming) arcs and the end place, no output (or outgoing) arcs.

When you get a log to mine, in many cases you do not know if the underlying model is a workflow net. If this is the case, the algorithm will not be able to find the net. So, the question here is, how could you pre-process the log such that you can make sure you will always have a workflow net as output? Explain why your solution would work.