Technische Universiteit
Eindhoven
University of Technology

DrFurby Classifier

Eric Verbeek
Felix Mannhardt

Where innovation starts

- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

- The training logs contain no noise
  - All traces in the training logs are to be classified as positive
- The winner is the one that classifies the most traces correctly
  - The readability of the model is not relevant

- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

- A collection of discovered accepting Petri nets
  - Initial marking
  - Set of final markings
- Semantics:
  - Replay the trace-to-classify on every discovered net and accumulate the replay costs
    - Costs for move-on-log: 10
    - Costs for move-on-model: 4
    - Other costs: 0
  - Classify the trace-to-classify as positive if and only if the accumulated replay costs are 0
  - Use decomposed replay to speed up the replay
    - Decomposed replay preserves perfect fitness, that is, costs 0

- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

- Use decomposed discovery
  - Preserves perfect fitness
- Use only discovery algorithms that guarantee perfect fitness
  - ILP Miner
  - Hybrid ILP Miner
  - Inductive Miner (variant that guarantees perfect fitness)
  - …?

- As a result, all traces of the training logs will be classified as positive

- Use as many decomposed discovery algorithms as are useful
  - An algorithm is considered to be useful if adding it results in additional negatives

- Note that we already have guaranteed that all traces from the training logs will be classified as positive. With adding additional discovery algorithms, we try to squeeze out as many negatives as possible.

- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

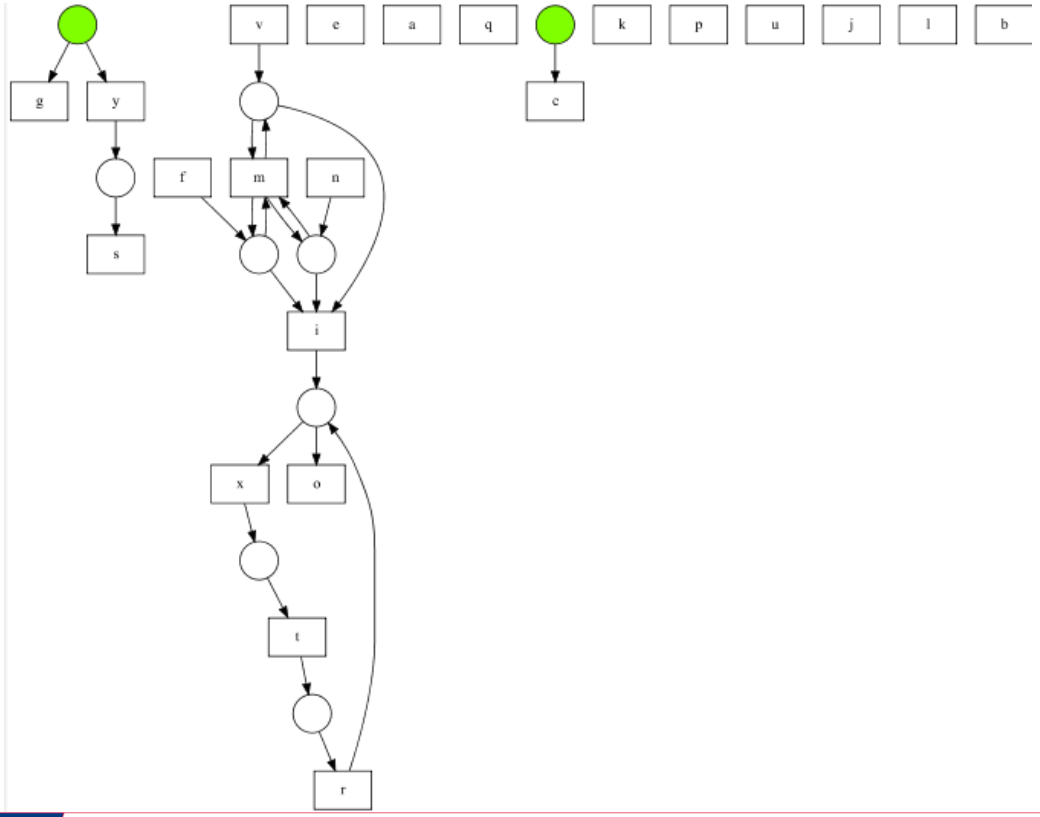
- We used the April and May logs to select the best set of useful decomposed discovery algorithms:
  - As few algorithms as needed to achieve the best result

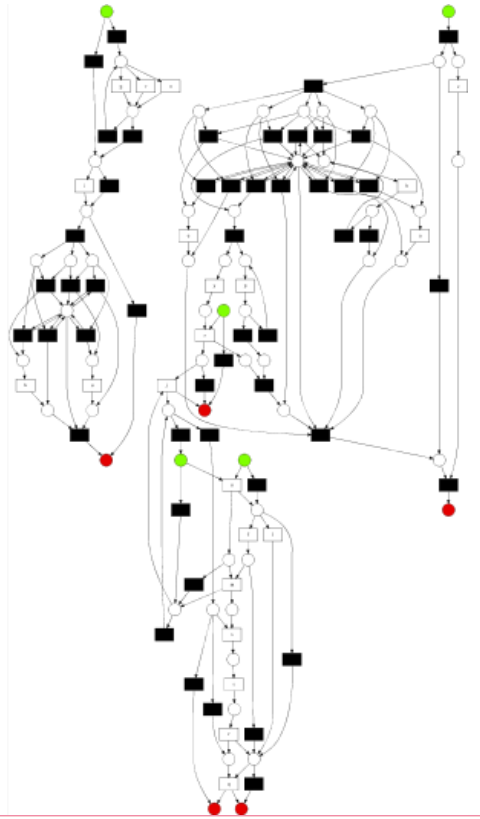
- The best result (showing numbers of traces classified as negative):

Training log	F	1	2	3	4	5	6	7	8	9	10
April log	1	10	10	10	10	10	10	10	9	10	10
May log	5	10	10	8	10	9	10	10	8	10	10

- As an example, we classified in the May 3 log only 8 traces as negatives. Hence, we know that we have at least 2 false positives.

- The best set of useful decomposed discovery algorithms:
  - Non-decomposed Hybrid ILP Miner (HIM-0)
  - Maximal-decomposed Inductive Miner (IM-100)





- Result from the April and May logs, as confirmed by organizers:
  - 1 misclassification for the April logs
  - 5 misclassifications for the May logs
  - These number match the numbers of ‘known’ false positives

Training log	**F**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
April log	1	10	10	10	10	10	10	10	9	10	10
May log	5	10	10	8	10	9	10	10	8	10	10

  - No false negatives!

- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

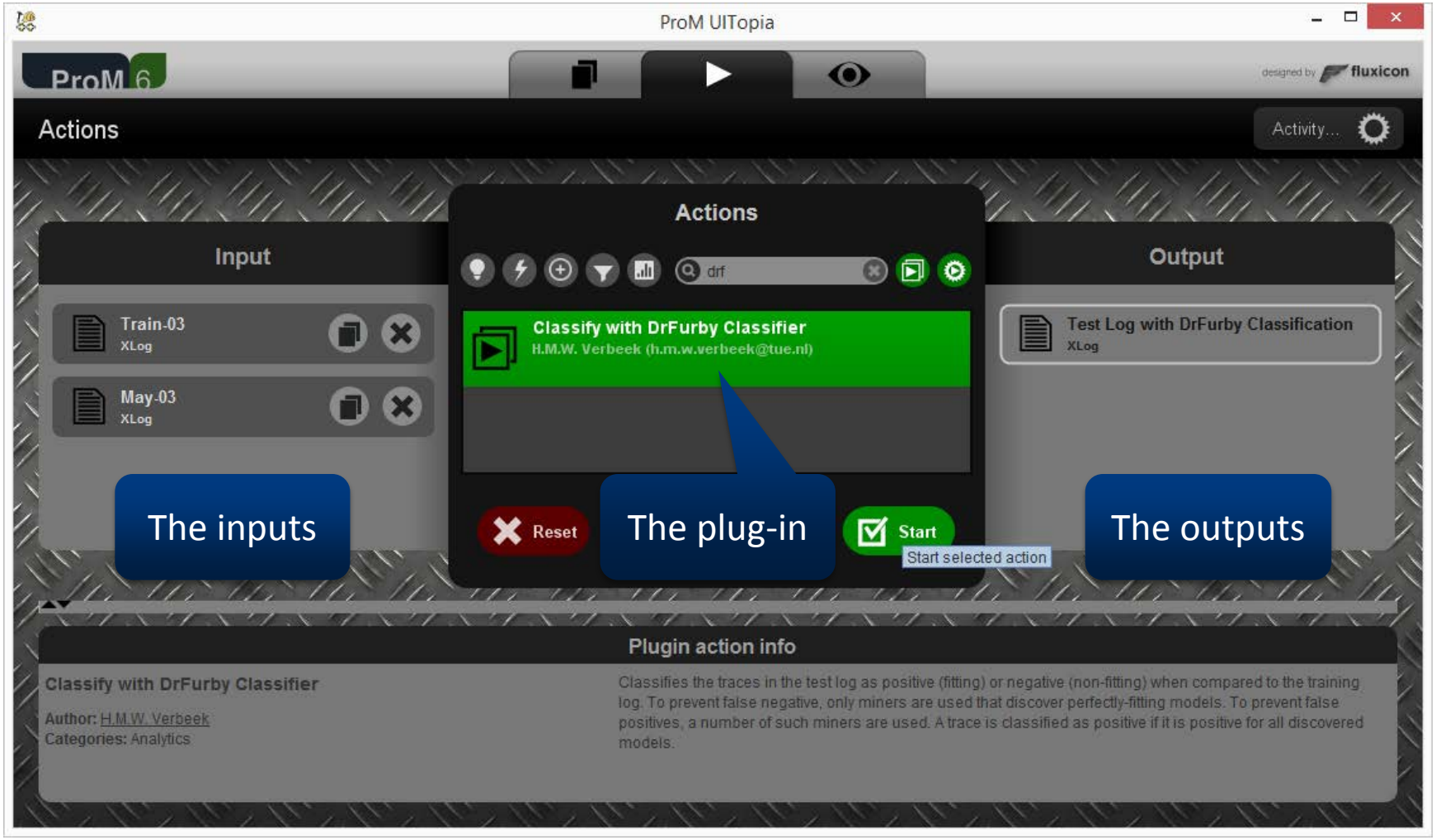
- The DrFurby Classifier enriches the log-to-classify, using a “drfurby” extension:
  - Log attributes:
    - name: Name of the log-to-classify
    - positive: Number of traces classified as positive
    - negative: Number of traces classified as negative
    - millis: Number of milliseconds it took to classify the log
  - Trace attributes:
    - classification: “positive” or “negative”
    - him0Costs: Costs of replaying this trace on the net as discovered by IM-100
    - im100Costs: Costs of replaying this trace on the net as discovered by HIM-0
    - totalCosts: Accumulated costs of replaying this trace on all discovered nets

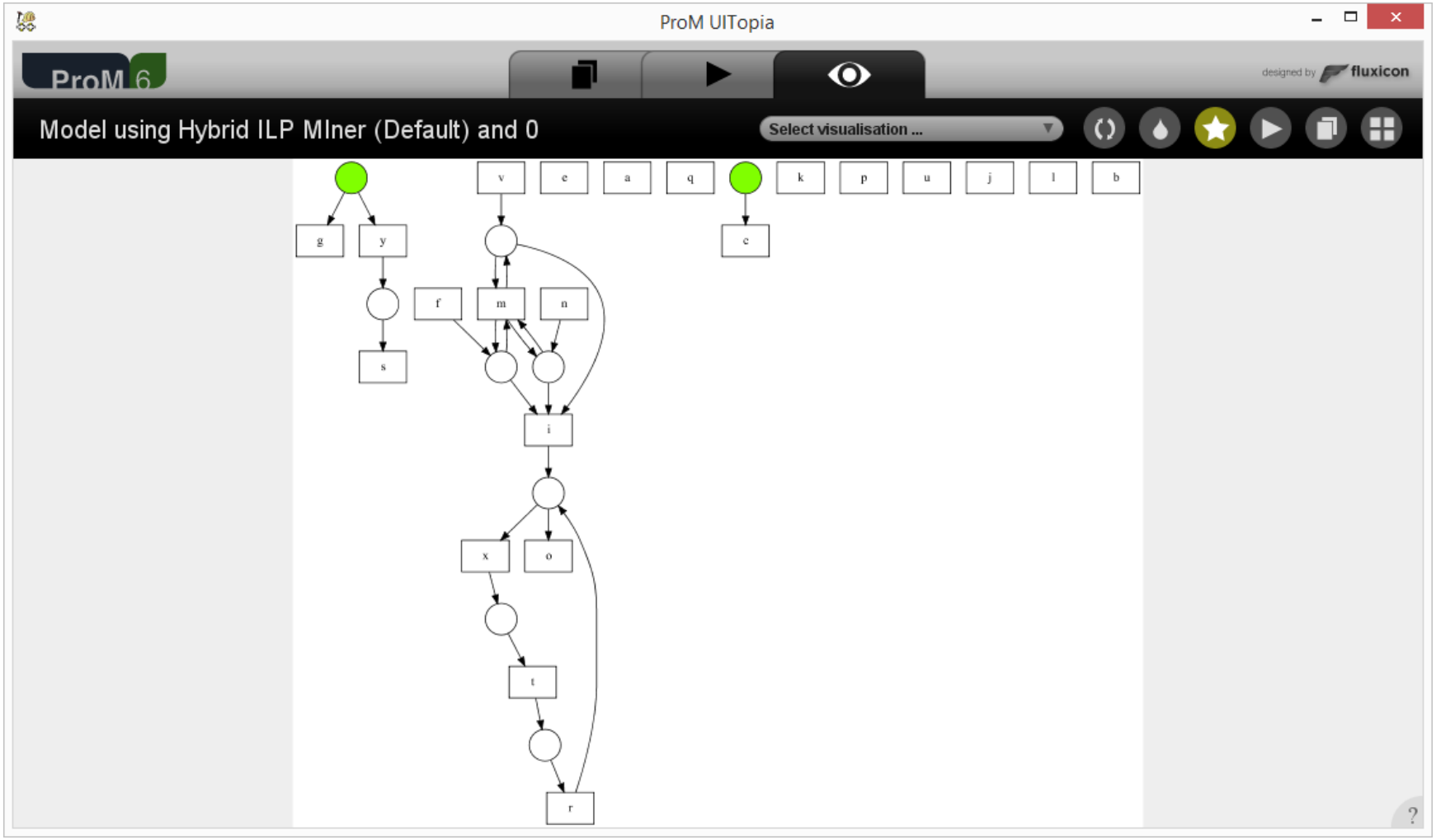
- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

The Process Mining Toolkit

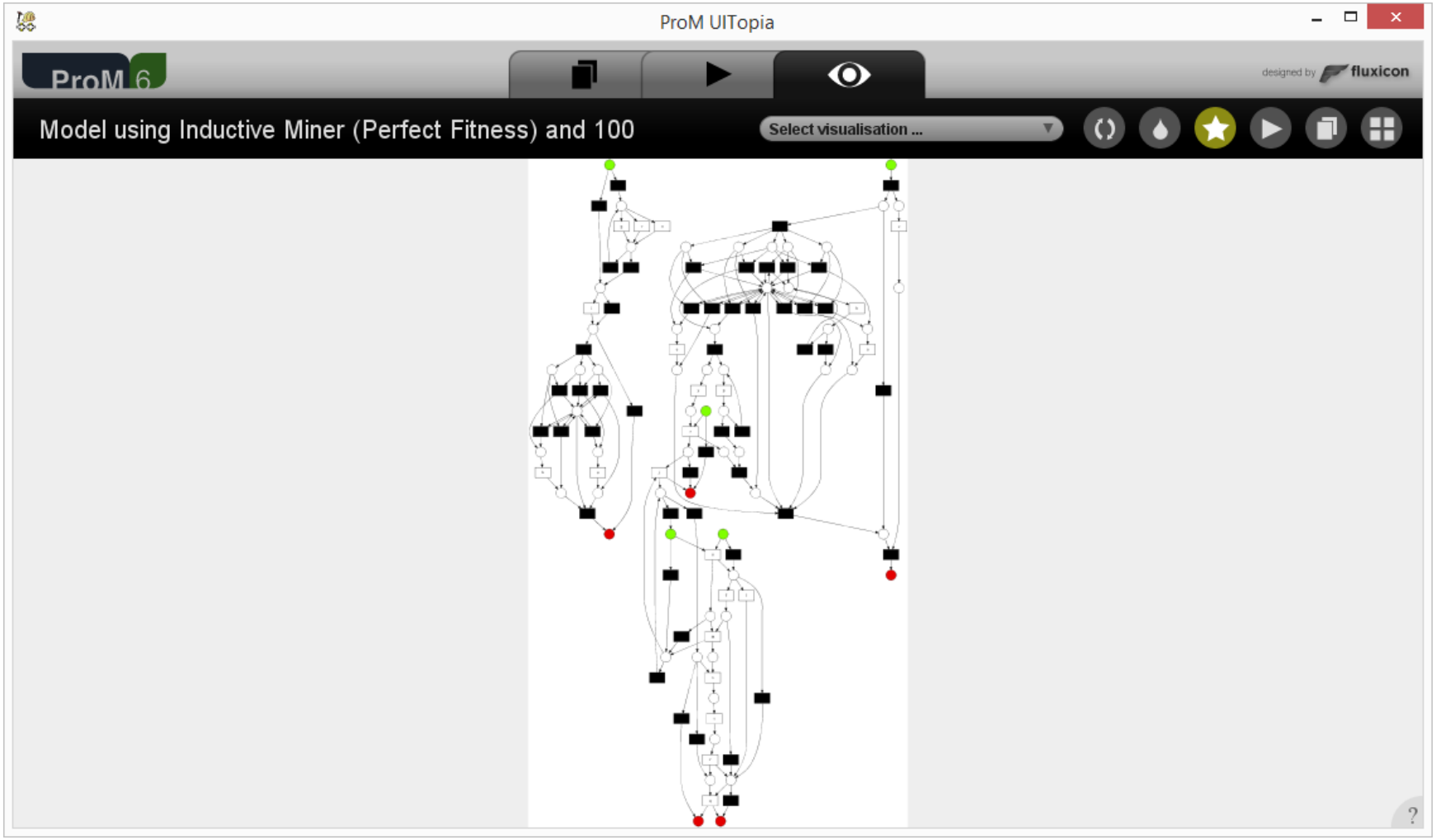
/ Wil van der Aalst / Peter van den Brand / Massimiliano de Leoni / Boudewijn van Dongen / Dirk Fahland / Christian Günther / Bart Hompes / Maikel Leemans / Sander Leemans / Xixi Lu / Felix Mannhardt / Eric Verbeek / Michael Westergaard

TU/e Technische Universiteit Eindhoven University of Technology
Where innovation starts

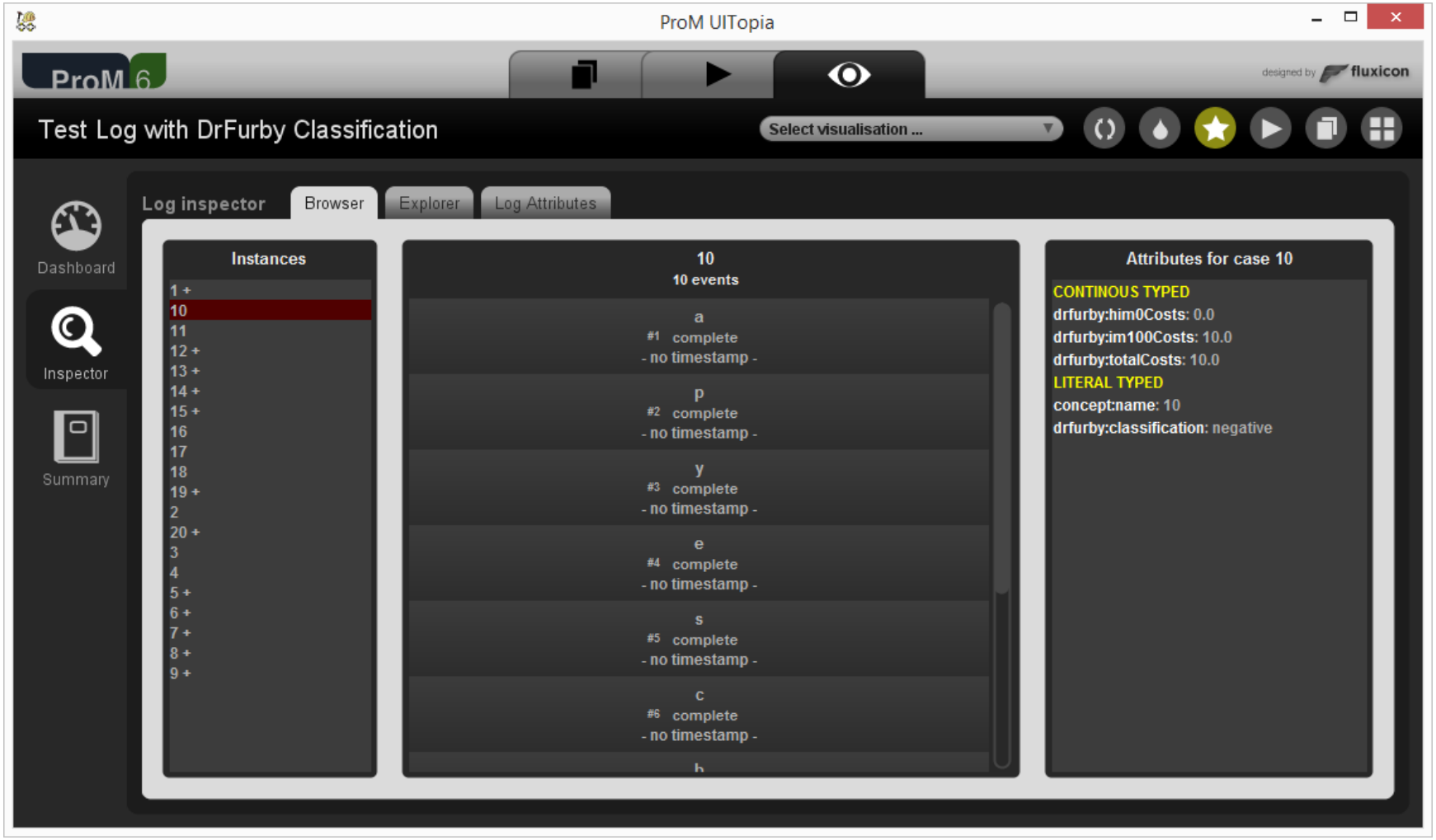

ProM UITopia
ProM 6
designed by fluxicon
Actions
Activity...
Input
Train-03
XLog
May-03
XLog
Actions
drf
Classify with DrFurby Classifier
H.M.W. Verbeek (h.m.w.verbeek@tue.nl)
Reset
Start
Start selected action
Output
Test Log with DrFurby Classification
XLog
The inputs
The plug-in
The outputs
Plugin action info
Classify with DrFurby Classifier
Author: H.M.W. Verbeek
Categories: Analytics
Classifies the traces in the test log as positive (fitting) or negative (non-fitting) when compared to the training log. To prevent false negative, only miners are used that discover perfectly-fitting models. To prevent false positives, a number of such miners are used. A trace is classified as positive if it is positive for all discovered models.

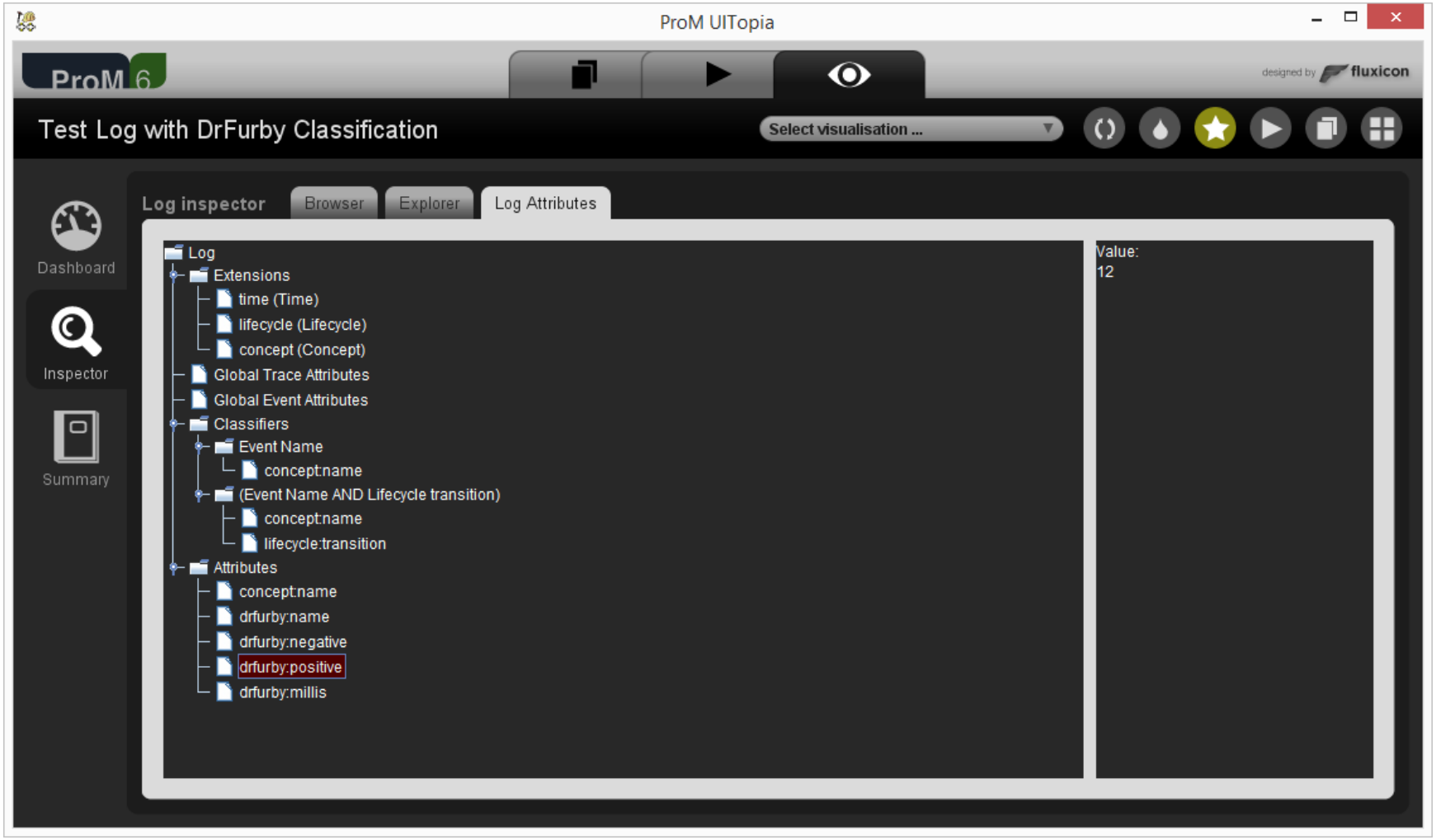


Implementation: result of HIM0 on log 3



Implementation: result of IM100 on log 3


ProM UITopia
ProM 6
designed by fluxicon
Test Log with DrFurby Classification
Select visualisation ...
Dashboard
Inspector
Summary
Log inspector
Browser
Explorer
Log Attributes
Instances
1 +
10
11
12 +
13 +
14 +
15 +
16
17
18
19 +
2
20 +
3
4
5 +
6 +
7 +
8 +
9 +
10
10 events
a
#1 complete
- no timestamp -
p
#2 complete
- no timestamp -
y
#3 complete
- no timestamp -
e
#4 complete
- no timestamp -
s
#5 complete
- no timestamp -
c
#6 complete
- no timestamp -
Attributes for case 10
CONTINOUS TYPED
drfurby:him0Costs: 0.0
drfurby:im100Costs: 10.0
drfurby:totalCosts: 10.0
LITERAL TYPED
concept:name: 10
drfurby:classification: negative



Implementation: result of IM100 on log 3

- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

- Results on the June logs:
  - Results for April and May logs added to allow comparison

Training log	F	1	2	3	4	5	6	7	8	9	10
April log	1	10	10	10	10	10	10	10	9	10	10
May log	5	10	10	8	10	9	10	10	8	10	10
June log	7	10	10	7	10	10	9	9	8	10	10

- 7 misclassifications
  - 7 false positives
  - No false negatives

- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Contest
- Conclusions

- DrFurby Classifier
  - Takes a training log and a test log
    - Assumes training log to be free of noise
  - Classifies the traces in the test log using the training log
- Some problems with some logs
  - Like 3 and 8
- No false negatives, only false positives
  - No guarantee on the former, however

Questions?

- Backup

- "DrFurby"
  - Is pronounced almost identically as "dr. Verbeek" ☺
  - Happened years ago, Boudewijn knows the story, he was there.

- Why decomposition?
  - Results on the April and May logs:

Training log	F	1	2	3	4	5	6	7	8	9	10
April log	1	10	10	10	10	10	10	10	9	10	10
May log	5	10	10	8	10	9	10	10	8	10	10
June log	7	10	10	7	10	10	9	9	8	10	10

  - Results on the same logs without using decomposition:

Training log	F	1	2	3	4	5	6	7	8	9	10
April log	4	10	10	7	10	10	10	10	9	10	10
May log	8	10	10	5	10	9	10	10	8	10	10
June log	9	10	10	5	10	10	9	9	8	10	10

  - No change for 9 out of 10 logs, but a significant improvement for log 3