

DrFurby Classifier

Eric Verbeek
Felix Mannhardt



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

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

- 10 Models, **not disclosed**
- For every model:
 - March log containing 1000 positive traces
 - ‘Training’ log
 - April/May logs containing 10 positive and 10 negative traces
 - ‘Calibration’ logs
 - June log containing 10 positive and 10 negative traces
 - ‘Test’ log
- 200 classifications
 - Contribution with most correctly classified traces wins
 - In case of a tie, the fastest wins

- Contest
- **Observations**
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Results
- 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



- Contest
- Observations
- **Model**
- Discovery
- Algorithm selection
- Extension
- Implementation
- Results
- 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

- Contest
- Observations
- Model
- **Discovery**
- Algorithm selection
- Extension
- Implementation
- Results
- 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.

- Contest
- Observations
- Model
- Discovery
- **Algorithm selection**
- Extension
- Implementation
- Results
- 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

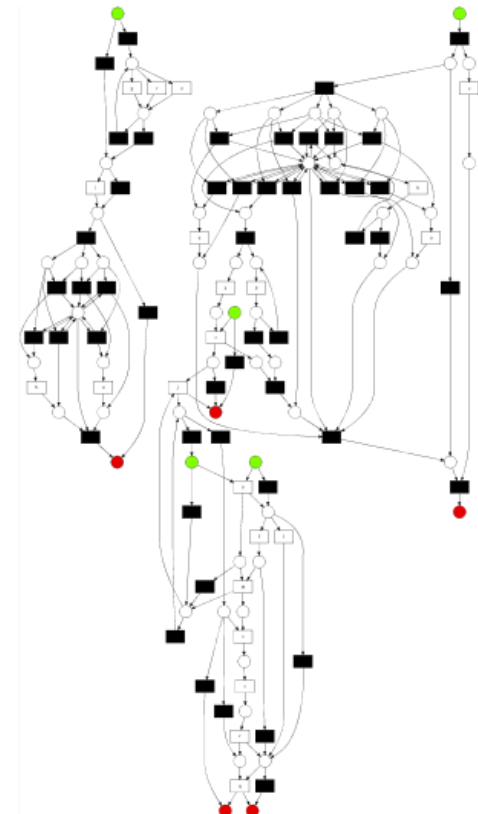
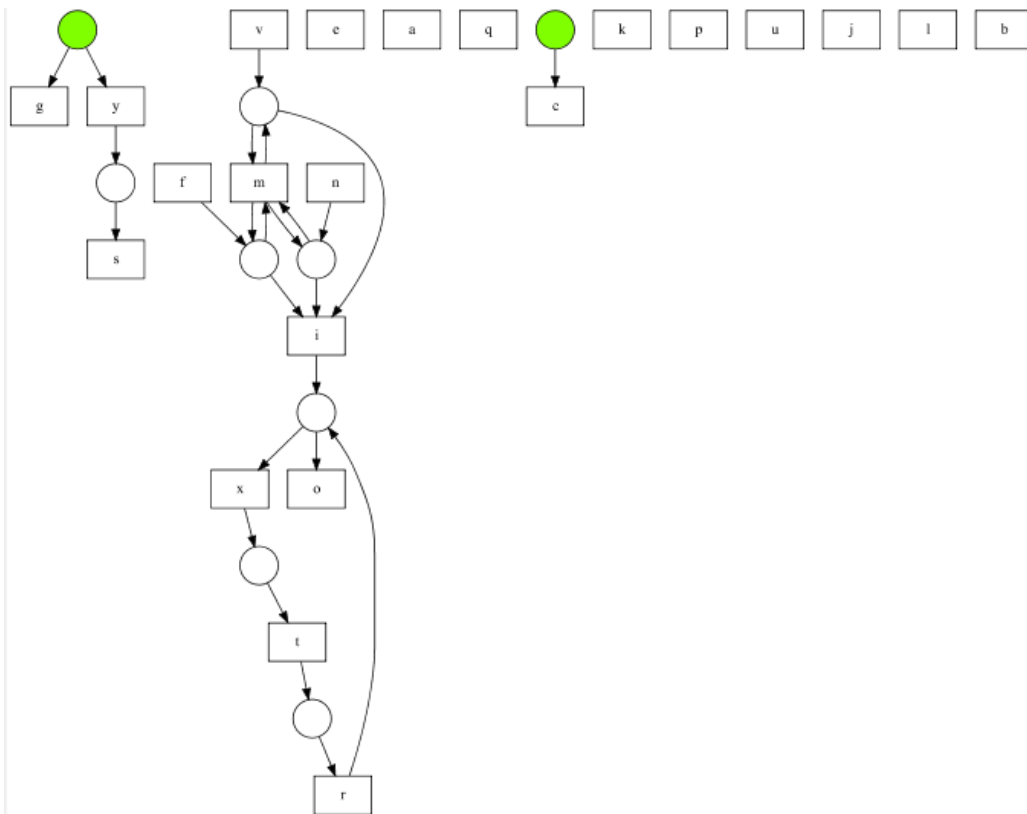


- The best result (showing numbers of traces classified as negative):

March 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:
 - Hybrid ILP Miner without decomposition (HIM-0)
 - Inductive Miner with maximal decomposition (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!



- Contest
- Observations
- Model
- Discovery
- Algorithm selection
- **Extension**
- Implementation
- Results
- 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

Prom

Revision 28643

6.6

/ 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

The screenshot displays the ProM 6.6 interface with the 'Actions' panel open. The panel is divided into three main sections: 'Input', 'Actions', and 'Output'. The 'Input' section contains two XLog files: 'Train-03 XLog' and 'May-03 XLog'. The 'Actions' section features a search bar with 'drf' and a list of actions. The 'Classify with DrFurby Classifier' action by H.M.W. Verbeek is highlighted in green. Below the action list are 'Reset' and 'Start' buttons. The 'Output' section shows a 'Test Log with DrFurby Classification XLog'. A 'Plugin action info' panel at the bottom provides details for the selected action.

The inputs

The plug-in

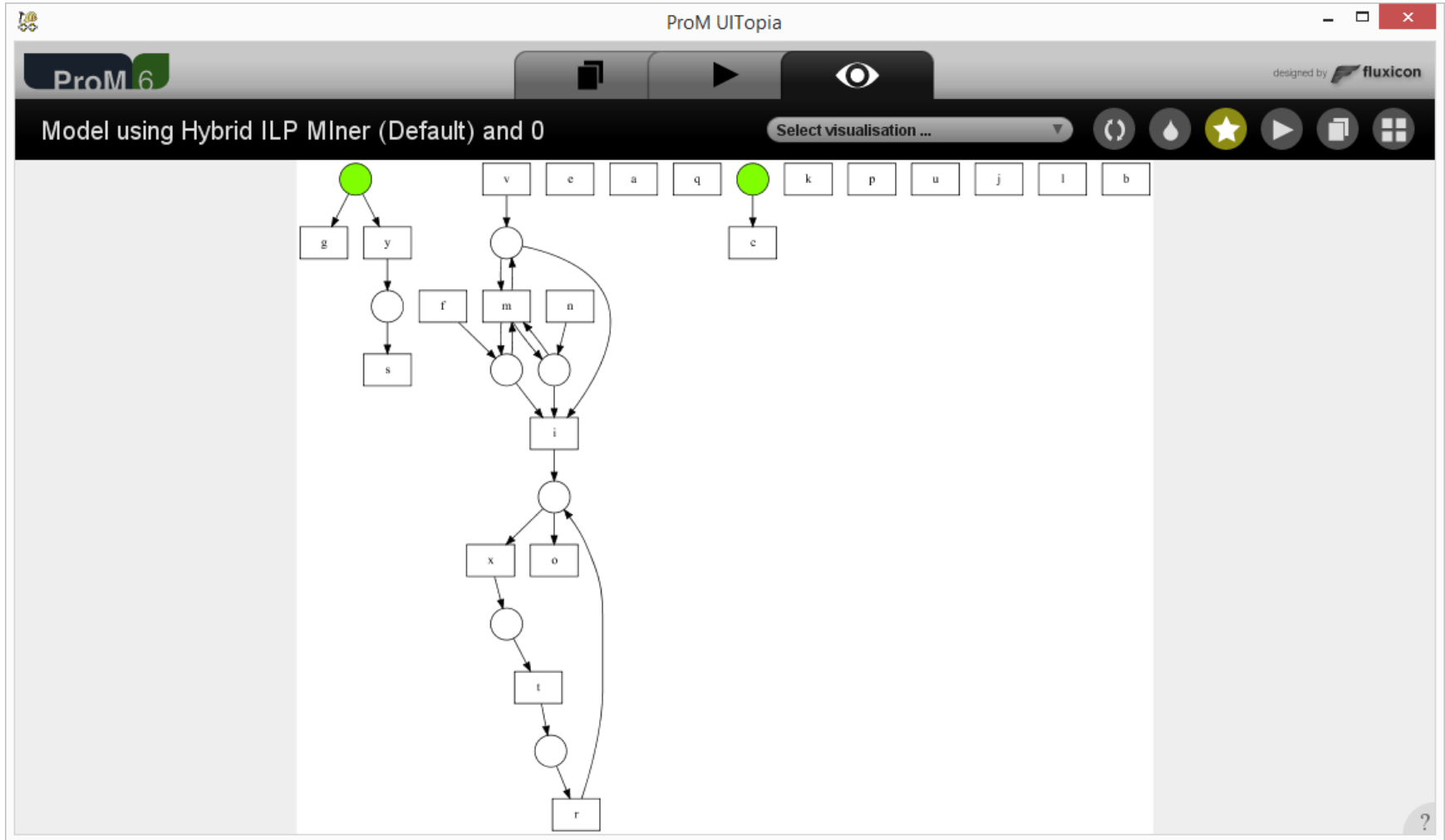
The outputs

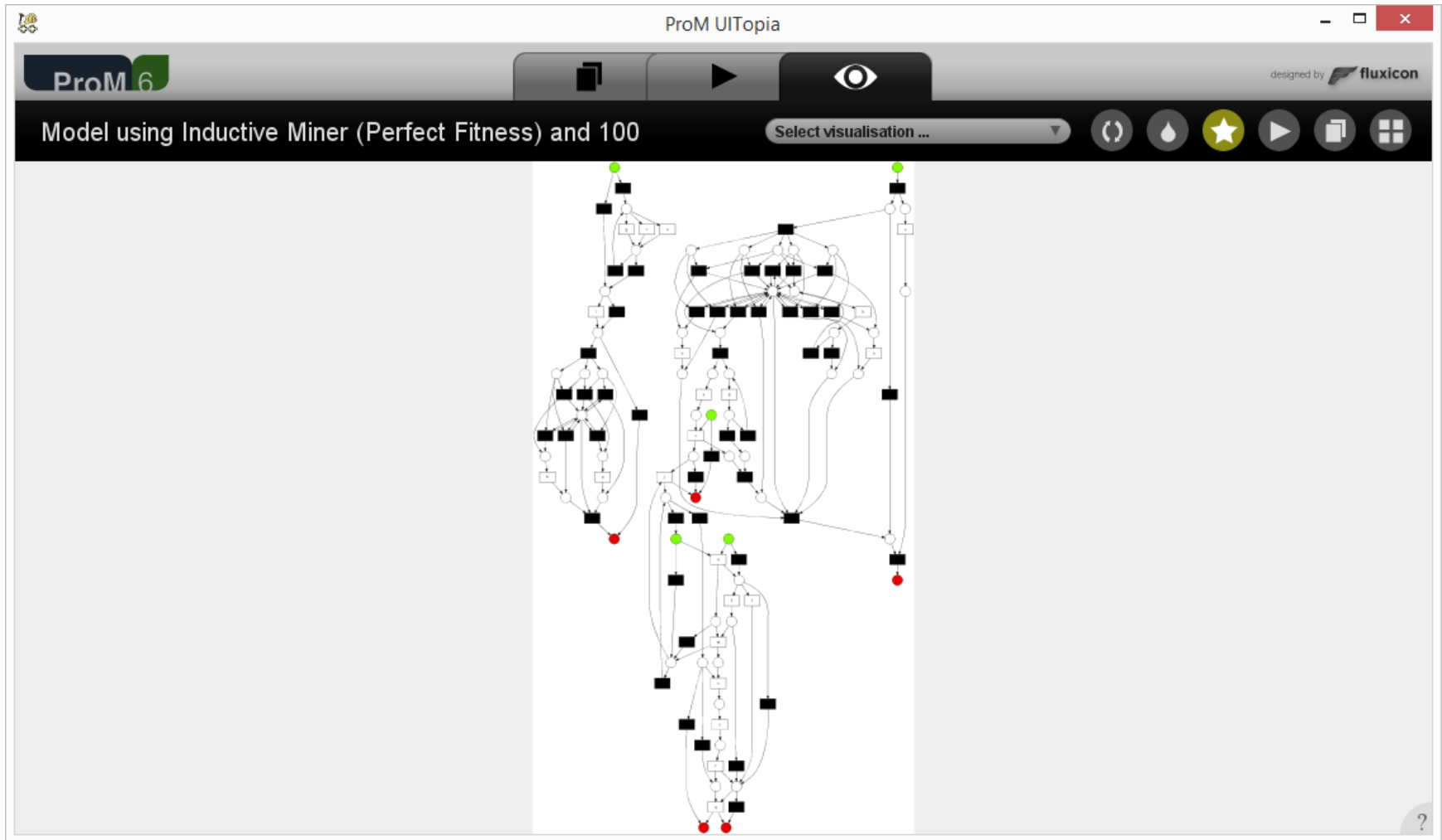
Start selected action

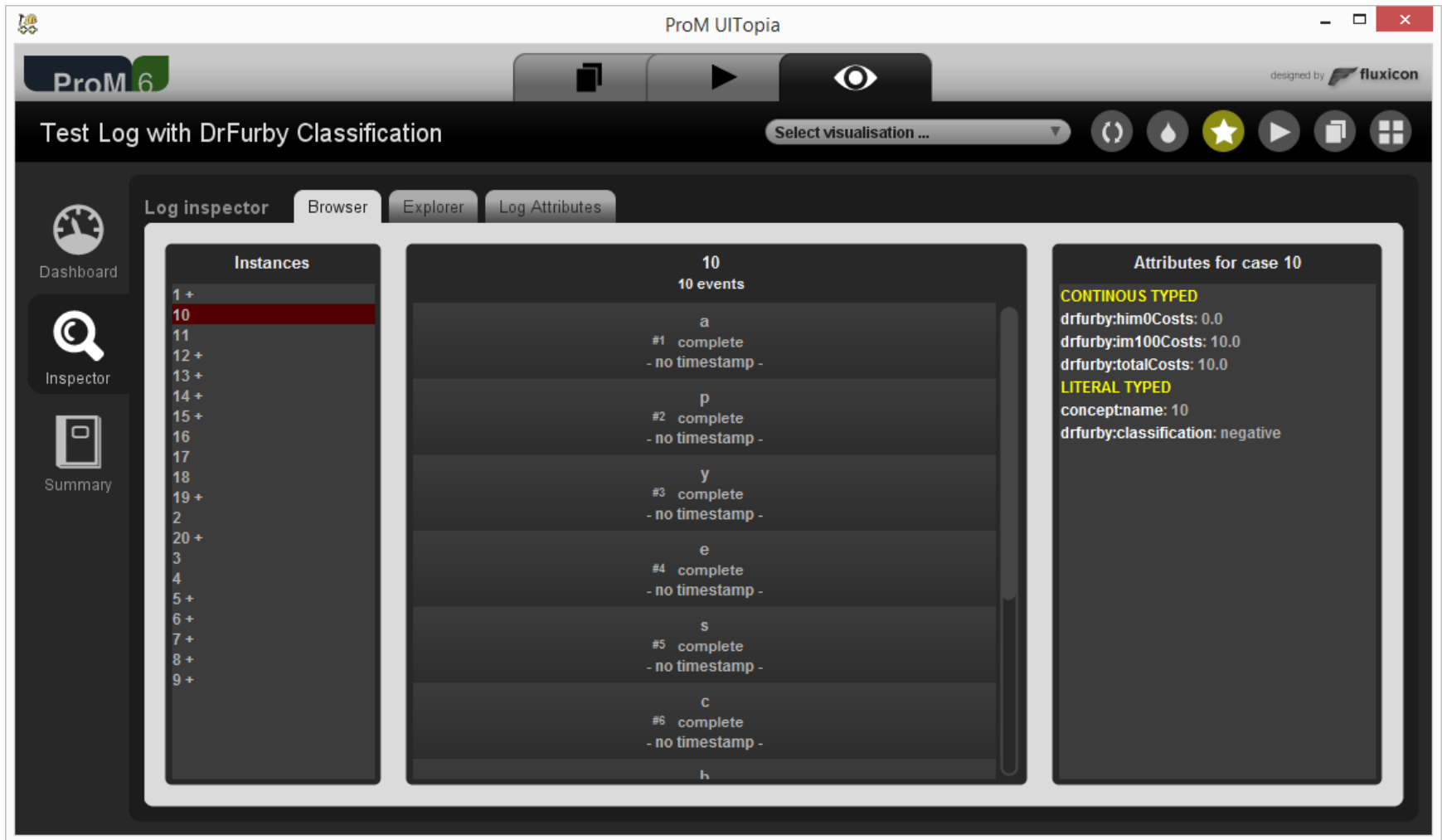
Plugin action info

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.







ProM 6

Test Log with DrFurby Classification

Log inspector | Browser | Explorer | Log Attributes

Dashboard

Inspector

Summary

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 -
- h

Attributes for case 10

CONTINUOUS TYPED
drfurby:him0Costs: 0.0
drfurby:im100Costs: 10.0
drfurby:totalCosts: 10.0

LITERAL TYPED
concept:name: 10
drfurby:classification: negative

The screenshot shows the ProM UITopia interface. The title bar reads "ProM UITopia". Below the title bar, there is a navigation bar with "ProM 6" on the left and "designed by fluxicon" on the right. The main content area is titled "Test Log with DrFurby Classification". Below this title, there is a "Select visualisation ..." dropdown menu and several icons. The interface is divided into a left sidebar and a main workspace. The sidebar contains "Dashboard", "Inspector", and "Summary" sections. The main workspace is titled "Log inspector" and has three tabs: "Browser", "Explorer", and "Log Attributes". The "Log Attributes" tab is active, showing a tree view of log data. The tree structure is as follows:

- Log
 - Extensions
 - time (Time)
 - lifecycle (Lifecycle)
 - concept (Concept)
 - Global Trace Attributes
 - Global Event Attributes
 - Classifiers
 - Event Name
 - concept.name
 - (Event Name AND Lifecycle transition)
 - concept.name
 - lifecycle.transition
 - Attributes
 - concept.name
 - drfurby.name
 - drfurby.negative
 - drfurby.positive
 - drfurby.millis

The "drfurby.positive" attribute is highlighted with a red box. To the right of the tree view, the value "12" is displayed under the heading "Value:".

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

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

March 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

- With 193 correctly classified traces, we won the contest.
- As a result, we received:
 - a free flight to Rio and back (Felix),
 - 4 free nights in a hotel in Rio (Felix),
 - free admission to the conference (Felix), and
 - a trophy, **which never materialized** (Eric, ☹).



- Contest
- Observations
- Model
- Discovery
- Algorithm selection
- Extension
- Implementation
- Results
- **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?

- “DrFurby”
- Runner-ups
- Decomposition

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



- Rahi Ghawi: 192 (7 FP, 1 FN)
 - Either (depends on log at hand):
 - Inductive Miner without decomposition, or
 - ILP Miner with maximal decomposition
- Moshe Steiner and Liat Bodaker: 192 (7 FP, 1 FN)
 - Alpha+
 - Model-specific postprocessing
 - Model should classify all traces from March log as positive

- Why decomposition?

- Results on the April and May logs:

March 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:

March 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