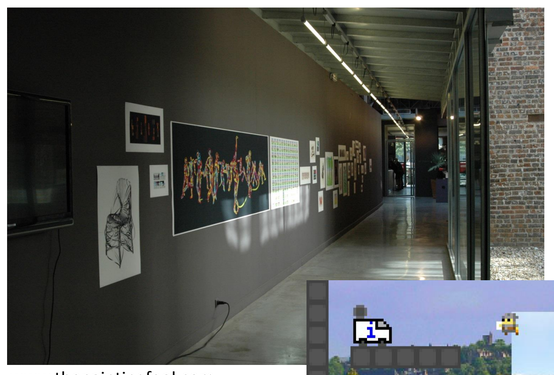
Machine learning for computational creativity evaluation

Martin Žnidaršič
Jožef Stefan Institute, Ljubljana, Slovenia

TASK
Development of evaluation models for computer generated creative artefacts , such as fictional ideas, poems, paintings, slogans, scientific hypotheses, music, games, …

Current focus: fictional (What-if…) ideas.



www.thepaintingfool.com

www.gamesbyangelina.org

- What if birds are not singing, but screaming, because they are afraid of heights?
- What if rocks are actually soft but tense up when we touch them?
- What if there would be a little bee that couldn't make honey?
- What if all animals speak a universal language and we are the only odd ones out?

PROBLEM
It is much easier to generate, than to evaluate:

- typicality
- novelty
- quality
- *fictionality*
- *narrative potential*
- …

A million monkeys typing

RELATED CONSIDERATIONS

system autonomy ⟺ output quality

curration coefficient

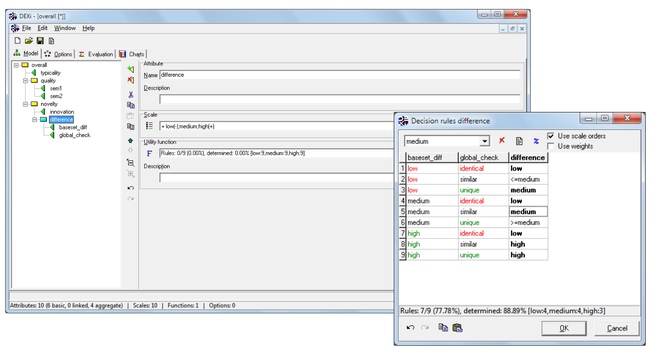
DATA

artefacts → evaluations

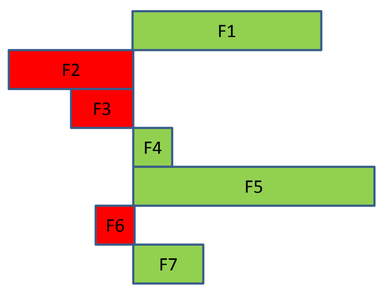
BoW	*length*	*sentiment*	*semantic ftrs.*	*actuality*	*rhyming*	*ambiguity*	…	gender	age	interests	…	$score_i$

SETTING

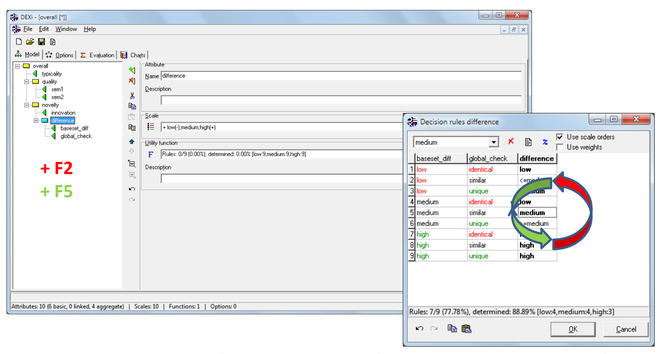
baseline
(hierarchical rule-based expert model)



ML



baseline + ML



Research question: How much better results (if at all) than baseline can we get by employing machine learning and by including machine learned features in expert models?

FIRST IMPRESSIONS

- Features are very artefact specific
- Evaluation is laborious
- Low inter-evaluator agreement

ACKNOWLEDGMENTS

This work is supported by the Slovenian Research Agency and through EC funding for the project WHIM 611560 supported by FP7, the ICT theme, and the Future Emerging Technologies FET program.