

# Current Applied Research in Machine Learning: Medical Abstracts and Digital Games

Stan Matwin

School of Information Technology and Engineering,  
University of Ottawa

[stan@site.uottawa.ca](mailto:stan@site.uottawa.ca)



joint work with, D. Inkpen, J. Sayyad Shirabad, K. White, J. Su, J. Huang, Y. Ma,  
W. Elazmeh, O. Frunza

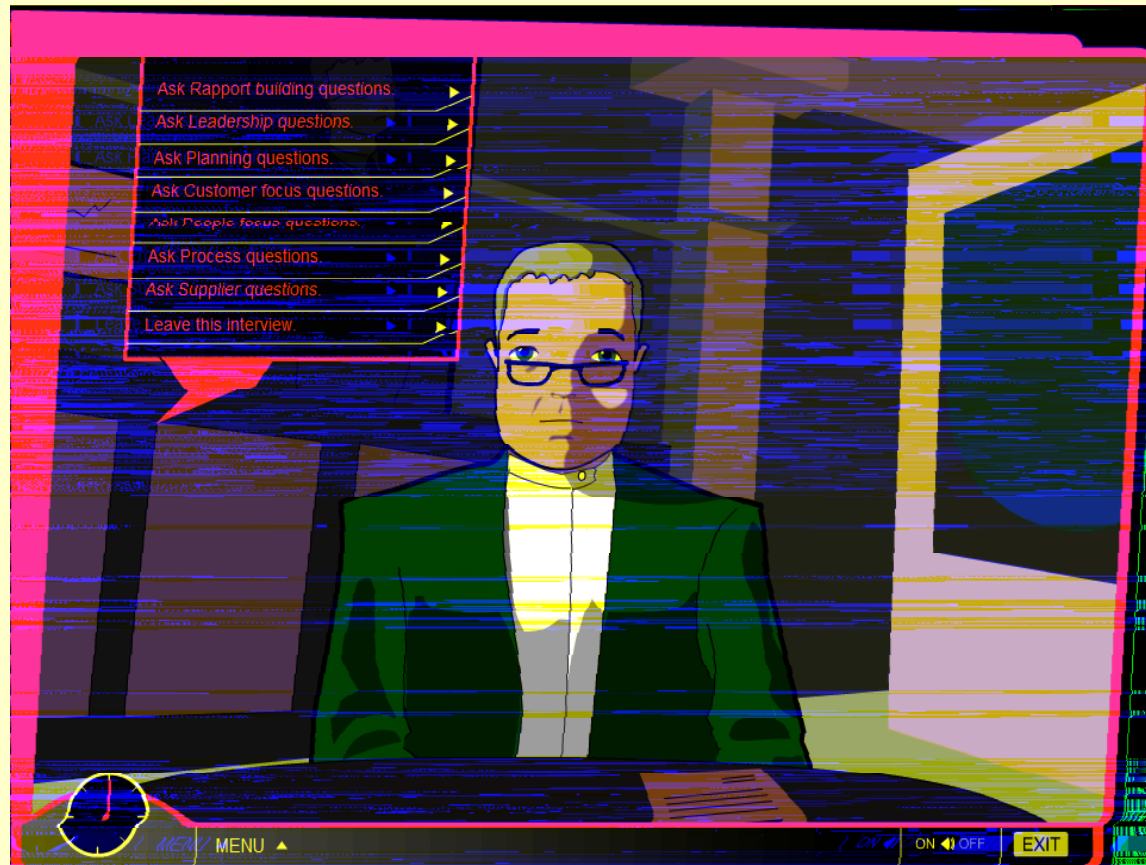
# Overview

- Digital-games based learning
- Use of machine learning
- Challenges
- Classification of medical abstract
- Use of machine learning
- Challenges
- Commonalities in challenges

# Digital Game Based Learning

- Intersection of Digital Games and E-learning.
- Uses techniques from the interactive entertainment industry to make computer-based training appealing to the end-learner.
- Chances are four to one that an employee under the age of 34 has been playing video games since their teenage years [Beck & Wade 04].

# An Auditor Certification Game



# Player Decision Through the Game

- What question to ask based on:
  - Given answers
  - Auditee character's body language
  - Environmental observations
  - General experience in auditing task

# The problem

- We are interested in the use of digital games for the purpose of testing knowledge of an examinee
- What makes a person an expert is not a crisply defined concept.
- In a typical DGBL scenario there are many there are many gameplay patterns that could be considered expert behaviour.

# Solution

- Cast the problem as a **classifier learning problem**
- Initial training phase:
  - Experts and non-experts will play games and submit their game logs (GL).
  - The gameplay analysis system will learn profiles of expert and non-expert (auditors) by learning models from the repository of stored GLs.

# Solution (Continued)

- Production phase:
  - Submit a GL to the gameplay analysis system.
  - The system will classify the GL (expert/non-expert) and generate a report
  - The instructional designer will receive the report and can accept it or further review it.
  - The GL, with its corresponding expert or non-expert label, is then added to the repository of GLs.



# Learning the Profiles

- Deciding as to what information in the game logs should be present in an example e.g. what questions were asked
- Encoding the example and creating other supporting input needed by the algorithm e.g. a Boolean variable for each question.
- Generating the model from the example using a classifier learning algorithm e.g. Decision Tree learner

# Generating the Feedback

- The learned model (profile) needs to be explainable e.g. c4.5 decision tree (Quinlan 1993), or rules generated by a rule learner such as Ripper (Cohen 1995).
- The feedback depends on the algorithm used

# Observation

- There are many fewer experts in real world than non-experts
  - Generate synthetic expert gameplays
- Not all events e.g. questions are created equally
  - We will increase expressiveness and generalization power of the learned models by using additional meta data about the events
- Meta data examples:
  - Open, closed, challenging, factual

# Challenges

- very little data to learn from
- data is imbalanced
- explainability of results a must

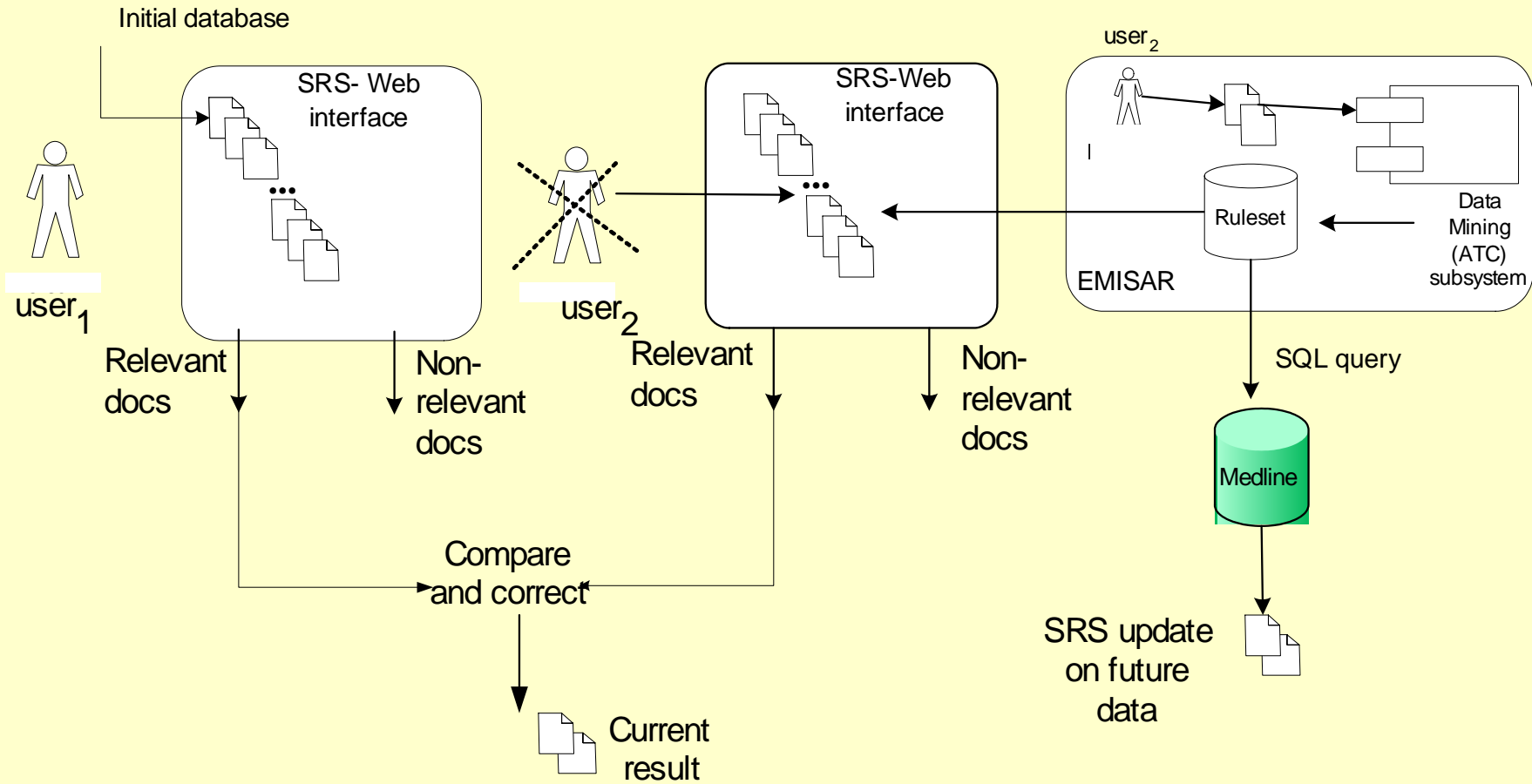
# Classifying medical abstracts

- Systematic Reviews -Trialstat
- Task fits with text classification, so classifier learning is a possible solution
- Embedding Machine Intelligence for Systematic Reviews (EMISAR)

# Systematic reviews

- A big industry
- Time and personnel consuming, expensive
- A knowledge-based process
- Needs tools

# Our solution



# Challenges

- Imbalanced data
- Noise in the data
- Reviews are heterogeneous
- Very high recall required
- What makes satisfactory precision?
- Cohen's WSS measure



# Commonalities

- Need to understand the application well
- Imbalanced data
- Embedded solution
- What is the right performance measure?

# Future Work

- Alternative algorithms including the capturing unique features of the system we embed in
- Finding the right classifier
- Use of transduction/co-training
- Active learning

*Thank you*