Comparing two models for selecting the next exercise

Papers:


&


Presented by: Diederik Roijers
Contents

- How is the next exercise selected?
- How is the first exercise selected?
- How do the models handle irregular behaviour?
- How do the models handle learning?
- Domains
- Experiments
- Discussion: strategies in these models
Faulhaber:

Semantics aware student model (SLM): concept k, procedure p, mastery m
SLM node: m(k,p)

Exercise:
Interpretation by Item Response Theory
Processed to the competencies by Transferable Belief Model
The next exercise: Faulhaber

Relations between exercises, concepts and competency metadata determines to which nodes evidences are attributed.

\[ IRT(m(k,p)) \]

Lastest 6 k,p evidences per student, \( \theta = m \)
The next exercise: Stacey

Stacey:

Misconceptions. A Bayesian network giving the chance that the student has a certain misconception.

Types of exercises, corresponding to the misconceptions.
System architecture:

Fig. 1. Intelligent Tutoring System Architecture
The next exercise: Stacey

Bayesian network architecture:

Fig. 2. BN representation of the student model
The next exercise: Stacey

After each set of entered evidence, the priors for the fineClass nodes are updated. No evidence “lingers” eternally.

The model for the misconceptions and how these misconceptions correspond to the response to exercises from each type was modeled together with experts.

If the experts do not know how a student of a certain type would do the initial probability is 0.5 chance of a correct answer.
Selecting the next exercise

Is based on a teaching model.

In [S03] it is made explicit that you might sometimes want to give very hard exercises to *show that a misconception is still present*.

In [FM08] *no such information* is given. This article focuses on the model for mastery.

In this respect could one replace the BN in the system architecture of [S03] by [FM08]'s model of masteries?
How is the first exercise selected?

In [S03] this is done by an *entrance test.*

In [FM08], where the model is used for *activeMath,* I imagine one could select the first 6 exercises ad random, and proceed from there.

Any other ideas on this?
Irregular behaviour

[FM08]:

after 6 exercises it's gone

[S03]:

models a small probability of careless mistake
Contents

- How is the next exercise selected?
- How is the first exercise selected?
- How do the models handle irregular behaviour?
- How do the models handle learning?
- Domains
- Experiments
- Discussion: strategies in these models
Learning

[FM08]:
- last six relevant items for mastery

[S03]:
- “resetting” the network with current priors
Domains

[FM08]:
- activeMath

[S03]:
- comparing two decimal numbers
- games: flying photographer, hidden numbers
Domains

[FM08]:
Very broad
Mathematics, high school, university
Not all errors known
Not all misconceptions known

[S03]:
Very small
Only decimal number comparisons
Teachers have a lot of information
Common misconceptions known
Domains (discussion)

[FM08]:
Applies not so well to decimals
Procedures non-existent
Only knows how it should be
Not what misconceptions exist
Anyone?

[S03]:
Does not apply to all of ActiveMath
Not every domain is so well known
However...
If we exclude the games,

In a procedure,
for every rule/concept
(mis)conception $\rightarrow$ (in)ability to apply

Buggy rule applications as misconceptions?
(Per rule or procedures, a set of buggy rules)

But how to estimate the initial priors?
- How is the next exercise selected?
- How is the first exercise selected?
- How do the models handle irregular behaviour?
- How do the models handle learning?
- Domains
- Experiments
- Discussion: strategies in these models
Experiments

[FM08] compares the *accuracy* of the prediction of the correctness of the student answers to other models.

Does not define when an answer is correct, and has no “gray zone” for partially correct answers.

No test on whether this model helps the students in their learning process.

How could this be done?
[S03]:
Field test with actual students playing games.

Pairs of students, introducing peer-review and the problem that not all feedback was good for both students.

Period of 3 weeks to measure long term effects. Teachers intervened, instruction got the overhand over experimenting.
Experiments (critique)

In my opinion both experimental verifications contain some problems.

For [FM08], I think the accuracy measure is a very blunt estimator which does not take a lot of available information into account. Which type of errors are made? Partial correctness

The model is compared only to one other model, and not against much simpler models which do not take the student into account (null measurement).
Experiments (critique)

There is no explanation for the shape of the graph in [FM08]; there is a strange drop after the fourth exercise which ought to be explained.

Nor is there a reference to which exact exercises have been used and no explanation about how well this result holds for all kinds of domains.

The setting of the parameters seem to be rather arbitrary.
Experiments (critique)

In [S03] it appears as though the enormous amount of environment variables have throttled the experimental verification completely. As they say themselves, the desire to instruct has won against the experimental verification while testing the system.

[FM08] does not contain an evaluation of whether the model improves the performance of the students.
[FM08] assigns evidence to the right mastery afterwards, but...

How about the order of solving a problem?

How about alternative paths?

Maybe processing while solving the exercise is an option? (Per step)

In the SLM/IRT/TBM model?

In the BN model?

Collections of BNs per rule?