Evaluating Intelligent Tutoring Systems with Learning Curves

Brent Martin
Antonija Mitrovic
Intelligent Computer Tutoring Group
Computer Science Department
University of Canterbury
Christchurch, New Zealand

Overview

- SQL-Tutor and CBM
- 2001 study – problem generation
- Inadequacy of learning curves
- New measure: initial slope
- Conclusions

Constraint-based Modeling

- The space of incorrect knowledge is vast - therefore abstractions are needed
- Represent only the basic principles
- *IF <relevance condition> is true, THEN <satisfaction condition> must also be true*

Advantages of CBM

- Very efficient computationally
- No need for an expert module
- No need for a bug library
- Insensitive to the radical strategy variability phenomenon
- Neutral with respect to pedagogy
- Suited to complex, open-ended domains

Extending CBM

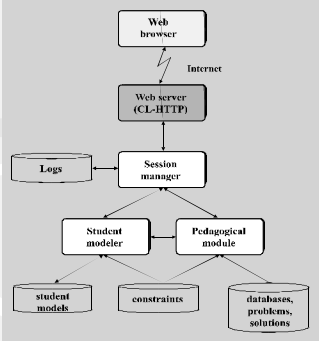
- Long-term student model
  - Overlay model
  - Probabilistic model
- Problem selection
- Problem generation
- Tailoring hints
- Open student models
- Supporting metacognitive skills

SQL-Tutor

- Solaris & Windows version 1998
- Web version since 1999
- Also available on Addison-Wesley's DatabasePlace Web portal

  http://www.aw-bc.com/databaseplace/
- More than 650 constraints

Architecture of SQLT-Web



An example constraint from SQL-Tutor

(147
"You have used some names in the WHERE clause that are not from this database."
(match SS WHERE (?* (^name ?n) ?*))
(or (test SS (^valid-table (?n ?t))
(test SS (^attribute-p (?n ?a ?t))))
"WHERE")

Generating new problems

- Many problems necessary
- A function of the authoring system – problem generation
- New problem generation algorithm
- Criteria for the algorithm
  - Must be effective
  - Must require less human involvement
  - Results in good problems
- 800 problems generated, 200 selected

2001 Study

- Three versions of the system
  - Control version
  - Open student model
  - Problem generation
- 4 lectures and 2 labs on SQL prior to the study
- During the study 4 additional lectures and 4 labs
- Pre-test on 10.9.2001
- System open on September 12
- Student logs
- Post-test on 9.10.2001

Results 2001

- 159 students enrolled in the course
- 100 sat the pre-test (mean 5, out of 9)
- 80 students logged on to SQL-Tutor
- 9 logs excluded
- 101 sat the post-test: overall mean 5.99 (1.60)

Group	Stud.	Pre-test	Logged	Valid	Post-test
Control	34	4.82 (1.44)	29	24	6.42 (1.38)
OSM	33	5.12 (1.41)	23	21	6.67 (1.56)
Prob.Gen	33	5.06 (1.25)	28	26	6.78 (1.37)

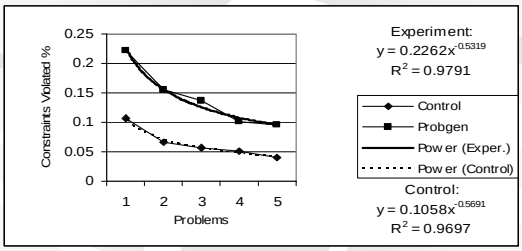
Analyses

- No significant difference in pre/post test performance
- No significant difference in the number of attempts per problem, time per attempt
- Students appeared to find the level of difficulty of problems about the same for the two versions of the system

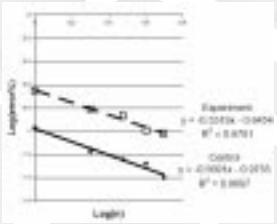
Learning curves

- Plot the reduction in error rate as students solve problems
- Probability of violating constraints at each problem
- Power law
- Slope of the curve is an indication of learning speed

Learning curves for the 2001 study



Learning curves from 2001 study



Why are learning curves different?

- Groups of different abilities?
- Different constraint sets?
- Different times at which constraints are presented?
- …

Testing the effects of various constraint sets

Measuring method	Control Slope	Experiment Slope
Original constraint set	0.57	0.53
Exclude trivially true	0.31	0.53
Experimental constraint set	0.44	0.53

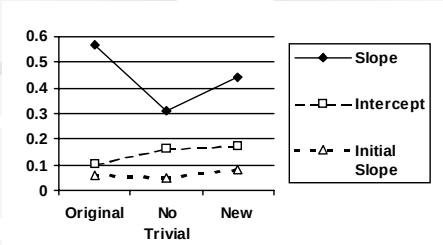
Problem with learning rate

- Sensitive to how the students' performance is being measured
- Y intercept indicates the amount of learning
- New measure: initial slope (at x=1)
- Indicates raw improvement in performance after receiving feedback on a constraint once
- Takes into account both the student's learning rate and the amount of the unlearned material

Measuring initial slope

Measuring method	Control	Experiment
Original constraint set	0.06	0.12
Exclude trivially true	0.05	0.12
Experimental constraint set	0.08	0.12

Comparing the measures



Conclusions

- Evaluating different version of the system is difficult
- Learning rate measures how fast the student is learning, but does not take into account the amount of material learnt
- Y intercept represents the amount of learning being undertaken, but does not say anything about the speed
- Initial slope combines the speed and the amount learnt, and is relatively invariant to how students performance is measured