Small Improvement for the Model Accuracy – Big Difference for the Students

Michael V. Yudelson, Steve Ritter

Carnegie Learning, Inc.

1 Company Profile

Carnegie Leaning began as a start-up from Carnegie Mellon University in 1998. An independent company now, it produces and sells math tutoring software – the Cognitive Tutor, math textbooks, and professional development in the United States and abroad. It is estimated that some 650,000 students are using Cognitive Tutor products every year. Carnegie Learning's Cognitive Tutor has always been a product tightly connected to cutting-edge research in the area of cognitive psychology and human-computer interaction.

2 Cognitive Tutor Innovation

The core of Cognitive Tutor is the skill model of the math curriculum. The model is used to track students' progress and to drive personalized problem selection. Over the years, the parameters of the skill model were revised several times by cognitive scientists. In general, when models of student learning are considered, primary attention is given to the accuracy. The practical effects of using an alternative model or a differently parameterized instance of a model are often overlooked. When student model defines how the system functions (e.g., problem sequencing), these effects should, arguably, come first.

In this report, we talk about comparing three parameter instances of Cognitive Tutor cognitive model. We introduce an algorithm that, given the data and the model parameters, computes the work students might save (in terms of number of problems and time) and extra work students would need to master the material. We show that, although in terms of accuracy the three model instances do not differ much, the projected student workload differs significantly.

3 Method

We use a dataset collected by Carnegie Learning's Cognitive Tutors in 2010. Included in this dataset, are student transactions resulting in the skill updates. After retaining students who completed at least two units of content, we have got 86,361,054 records of 56,046 students from 326 schools practicing 3,760 skills.

The algorithm for computing time savings is similar to the one used in [1]. Given the data and the cognitive model parameters, we simulate the computation of the cognitive model values. The central value is the probability that

Table 1. Model comparison. Time (hh:mm) and problems are given per student.

Model	Accuracy	Time saved	Time extra	Problems saved	Problems extra
BKT Shipped	0.771687	00:03	00:50	2.15	19.13
BKT EM	0.817828	01:09	00:50	41.53	19.17
BKT GD	0.807074	01:43	00:24	50.87	10.67
BKT GD Lag	0.819358	00:38	01:02	22.44	25.55

the skill is mastered. Once it reaches 0.95 the Cognitive Tutor stops issuing practice problems for that skill. Once all of the skills in the curriculum section are mastered, student is allowed to proceed to the next section. When all of the problems of the section are exhausted, but not all skills are mastered, the student is promoted to the next section.

In the simulation, we consider student activity by blocks corresponding to problems. If, after solving a problem, all the skills are considered mastered but there are more problems in the data, all of them are considered *saved* problems under the model in question. The time student spent solving them is also considered as *saved*. If we ran out of student data but, as per the model, not all the section skills were mastered, we computed projected number of extra problems and the time necessary using a heuristic. For all unmastered skills, we computed the number of successful attempts it would require to take it to the mastered level and selected the highest value. We used this value as the number of extra problems necessary to complete the section. We multiply the number of extra problems by the average problem length in this section to get the extra time.

4 Results

Table 1 is a summary of our model comparisons. Here, we list both the model accuracy as well as number of problems and the time. Together with the fit models, we list results for the parameters that were shipped with the product. Projections could show savings and extra work simultaneously: students can save time on some sections, but require more work on others. Also, due to teacher ability to move students ahead in the curriculum and occasional adjustments of the initial masteries as a result of the pre-test, the extra time could be inflated.

As we can see, the differences in accuracy between the fit models BKT EM, BKT GD, and BKT GD Lag are quite small – all close to 1%. In this set, BKT GD fit model has the lowest accuracy. At the same time, it's projected saving are the highest and extra requirements are the lowest.

References

1. Yudelson, M., Koedinger, K., Gordon, G. (2013) Individualized Bayesian Knowledge Tracing Models. In: Lane, H.C., and Yacef, K., Mostow, J., Pavlik, P.I. (eds.) Proceedings of 16th International Conference on Artificial Intelligence in Education (AIED 2013), Memphis, TN. LNCS vol. 7926, (pp. 171–180).