Modeling Student Retention in an Environment with Delayed Testing

Shoujing Li
Worcester Polytechnic Institute
100 Institute RD; Worcester MA
sli@wpi.edu

Xiaolu Xiong
Worcester Polytechnic Institute
100 Institute RD; Worcester MA
xxiong@wpi.edu

Joseph E. Beck
Worcester Polytechnic Institute
100 Institute RD; Worcester MA
josephbeck@wpi.edu

ABSTRACT

Student modeling has been widely used in the prediction of student correctness behavior on the immediate next action. Some researchers have been working on student modeling to predict delayed performance, that is, *retention*. Prior work has found that the factors influencing retention differ from those that influence short-term performance. However, this prior research did not use data which were specially targeted to measure retention. In this study, we describe our experiments of using dedicated retention performance data to test the students' ability to retain, and experiment with a new feature called *mastery speed*, indicates how many problems the students need to attain initial mastery. We found that this new feature is the most useful of our features. It's not only a helpful predictor for 7-day retention tests, but also a long-term factor that influences students' later retention tests even after 105 days. We also found that, although statistically reliable, most features are not *useful predictors*, such as the number of students' previous correct and incorrect responses which are not as helpful in predicting students' retention performance as in PFA.

Keywords

Educational data mining, Knowledge retention, Robust learning, Feature selection, Intelligent tutoring system.

1. INTRODUCTION

Automatic Reassessment and Relearning System (ARRS) is an extension of the mastery learning problem sets in the ASSISTments system (www.assistments.org), a non-profit web-based tutoring system for 4th through 10th grade mathematics. Mastery Learning is a pedagogical strategy which, in most ITS, indicates that a student is presented with problems to solve until he masters the skill. The exact definition of "mastery" varies from tutor to tutor: some tutors consider a student to have mastered the skill if his estimated knowledge is very high, for example over 0.95 (e.g., [3]), while ASSISTments uses a heuristic of three correct responses in a row. The idea of ARRS is if a student masters a problem set, such mastery is not necessarily an indication of long-term retention. Therefore, ARRS will present the student with a reassessment test on the same skill at expanding intervals: first 7 days after the initial mastery is due, then 14 days after the prior test, than 28 days later, and finally 56 days later. Thus, the retention tests are spread over an interval of at least 105 (7+14+28+56) days. In this study, we defined retention performance as the reassessment test performance one week after a student was assigned a skill (i.e., the first reassessment test). Note, that if a student fails the reassessment test, ASSISTments will give him an opportunity to relearn the skill. Once a student relearns (demonstrates mastery) a skill, he will receive another reassessment test at the same delay at which he previously responded incorrectly. In other words, if the student failed the second reassessment test, he would have to relearn the skill and achieve 3 correct answers in a row, before receiving another reassessment test 14 days later.

In our previous study, we identified *mastery speed* as a useful construct in prediction of retention performance. Mastery speed refers to the number of attempted problems during the process of achieving mastery. Mastery speed represents a combination of how well the student knew this skill initially, and how quickly he can learn the skill.

2. MODELS AND RESULTS

2.1 Data set

For this study, we used data from the ARRS system, specifically students' 7-day test performance and other features about their previous learning on that particular skill. We had 48,873 questions answered by 4054 students, from 91 different skills. Then we calculated the following features which were used in our regression models:

- *mastery_speed*: the number of problems needed to master a certain skill. We binned this feature into 6 categories ('<3 attempts', '3-4 attempts', '5-8 attempts', '>8 attempts', 'not mastered', 'skipped initial mastery'). Students could master a skill in less than 3 attempts if their teachers overrode ASSISTments mastery criterion.
- *n_correct* (*n_incorrect*): the number of students' prior correct (incorrect) responses on that skill before the retention test.
- *n_day_seen*: the number of distinct days that the students have practiced this skill.
- *g_mean_performance*: the exponential moving average of students' performance before the reassessment test. We used the same formula as in Wang and Beck's previous work [2]: *g_mean_performance (opp) = g_mean_performance (opp-1) * 0.7 + correctness (opp) * 0.3* using opp to represent the opportunity count and a decay of 0.7.
- *g_mean_time*: the exponential moving average of students' response time on that skill before the reassessment test [2]. The formula is: *g_mean_time (opp) = g_mean_time (opp-1) * 0.7 + response_time (opp) * 0.3*.
- *problem_easiness*: percentage correct for this problem.

2.2 Separate Model with each Feature

In our binary logistic regression models, we used correctness as the dependent variable. We first tested a base model with just three features: *user_id*, *skill_id*, and *problem_easiness*, which showed as reliable predictors in a model we created with all features in our feature set. The base model provided an R^2 of 0.373. The next step we took was to test each feature one at a time added to the base model. Table 1 shows the Beta coefficient, p-values and R^2 gain for each regression model.

Each row in the table represents one regression model, with the feature listed and other three features in the basic model. The last column, R^2 gain, shows the increase in R^2 from adding that feature to the base model. Given even the modest (by EDM standards) data set we have for this study, circa 50,000 rows, even trivially small effects can show up as statistically "significant." Therefore, we compute how much improvement the feature actually provides us with. From the table, it's clear that *mastery_speed* is the most powerful predictor for students' retention performance. And also the students' previous performance on that skill (*g_mean_performance*) has a clear influence on prediction. The other variables have a trivial impact on performance. Note that even the best two features have a small impact on retention.

Compared with prior work [2], we found that *n_day_seen* did not replicate as being a useful feature. Strangely, a student's raw number of correct and incorrect response has little impact on retention. But *g_mean_performance* which measures students' previous performance on correctness has a clear influence on students' retention, which indicates that simply counting the raw number of correct or incorrect responses does not seem that helpful. Using exponential moving average which weights recent attempts more heavily as we did to compute *g_mean_performance*, is a helpful way to use students' previous correctness information.

Table 1. Parameters table for separate Models

Feature	R^2	B	p-value	R^2 gain
mastery_speed	0.379	---	0.000	0.006
n_correct	0.374	0.010	0.000	0.001
n_incorrect	0.373	-0.007	0.004	0.000
n_day_seen	0.373	0.026	0.002	0.000
g_mean_performance	0.378	1.130	0.000	0.005
g_mean_time	0.373	0.000	0.649	0.000

2.3 Impact of Mastery Speed

From the previous models we presented, we found that mastery speed has a clear influence on students' 7-day reassessment tests. However, what about the 14 day test, 28 day test, and even the 56 day tests? We collected all student performances on all four reassessment tests. As shown in Figure 1, we calculated the percentage of correct answers on each retention test, disaggregated by initial mastery speed.

Students get better as they move to the later retention tests. This is expected since they must get the previous tests correct in order to move on, and some weaker students are forced to repeat and so are systematically oversampled on the left side of the graph. On the 7-day retention test, students who mastered a skill quickly with 3 or 4 attempts (blue line) have a 24% higher chance of responding correctly than those students who required more than 8 attempts to master a skill (green line). Such a difference is perhaps not surprising. More interesting is the persistence of this differential performance: the 56 day level tests, the group who mastered quickly are still performing about 15% better than the students who mastered slowly. This difference persists in spite of weaker students being screened out on earlier retention tests. This result tells us that the initial mastery speed is of importance in terms of students' retention performance even after 105 days.

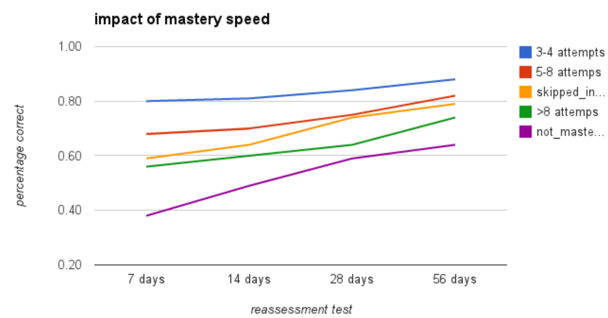

Figure 1. Impact of mastery speed on retention tests

3. CONCLUSIONS AND FUTURE WORK

This paper represents our attempt to model student retention performance in the context of a computer tutor. The two most interesting results were mastery speed being the best predictor, and the effects of performance on initial mastery persisting across such a lengthy interval. We did not anticipate this effect, and were therefore surprised by it.

There are several interesting open questions that might be further explored in the future. First, we have noticed anecdotally and through preliminary analysis that students sometimes get confused among similar skills during problem solving, an example of proactive interference [1]. Computer tutors would seem to be a strong research vehicle for better understanding of such effects in an authentic learning context, and over longer time than typical psychology lab studies. Another question is that we have found that slow mastery speed results in poor performance on delayed tests. An open question is whether a stronger mastery criterion, such as 4 or 5 correct in a row, would be helpful.

ACKNOWLEDGMENTS

We want to acknowledge the funding on NSF grant DRL-1109483 as well as funding of ASSISTments. See here (http://www.webcitation.org/67MTL3EIs) for the funding sources for ASSISTments.

REFERENCES

[1] Anderson, J.R., Rules of the Mind. Lawrence Erlbaum (1993).

[2] Wang, Y., & Beck, J. E. (2012). Using Student Modeling to Estimate Student Knowledge Retention. In Proceedings of the 5th International Conference on Educational Data Mining, 176-179.

[3] Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction, 4(4), 253-27