Improving Retention Performance Prediction with Prerequisite Skill Features

1st Author
1st author's affiliation
1st line of address
2nd line of address
1st author's E-mail address

2nd Author
2nd author's affiliation
1st line of address
2nd line of address
2nd E-mail

3rd Author
3rd author's affiliation
1st line of address
2nd line of address
3rd E-mail

ABSTRACT
This paper describes our experiment and analysis of utilizing prerequisite skill features to improve the predicting student retention performance. There are two aspects that make this paper interesting. First, instead of focusing on short-term performance, we investigated the student retention performance after a delay of 7 days. Previous studies showed that, relative to predicting immediate performance, different features were useful in predicting delayed performance. Therefore, our second goal is to discover new features other than these were used on accessing students’ local performance. We explored several prerequisite skill features that can be captured in intelligent tutoring system; in our particular case, these prerequisite skill features were acquired from Common Core standard skills and student data while working on these skills. We showed that some of them have encouraging predictive power. Our analysis confirmed the value of prerequisite skill features in predicting retention performance, the prediction results showed an improvement from an R² of 0.182 with a baseline feature set to an R² value of 0.192.

Keywords
Educational data mining, feature selection, knowledge retention, intelligent tutoring system

1. INTRODUCTION
For decades, researchers in the field of educational data mining (EDM) have been developing various methods of predicting the correctness of the next student response to the question (e.g., [2, 3]), in other words, predicting student short-term performance. Student modeling has been widely used for making such inferences. Although performing well on the immediate next problem is an indicator of mastery, it is by far not the only criteria. For example, the Pittsburgh Science of Learning Center’s theoretic framework focuses on robust learning (e.g., [4, 11]), which includes the ability to transfer knowledge to new contexts, preparation for future learning of related skills, and retention - the ability of students to remember the knowledge they learned over a long time period. Especially for a cumulative subject such as mathematics, robust learning, particularly retention, is more important than short-term indicators of mastery, because we are more concerned with students’ capability to remember the knowledge that they acquired over a long period of time.

1.1 ARRS
Inspired by the notion of robust learning [6] and the design of the enhanced ITS mastery cycle proposed by Wang and Beck [10], we developed and deployed a system called the Automatic Reassessment and Relearning System (ARRS) to make decisions about when to review each skill the student mastered. The ASSISTments is a non-profit, web-based tutoring project (www.assistments.org) for 4th through 10th grade mathematics tutoring (approximately 9 through 16 years of age). In the school year of 2012 to 2013, it served approximately 20,000 active students nationwide. One of the important compounds of ASSISTments is the mastery learning problem set, which simplifies the notion of skill mastery to three consecutive correct responses with the number of attempted problems before students achieve mastery. Note that three problems for a skill represent the lower boundary for the amount of practice students require. However, if students make mistakes, they are required to obtain three correct answers in a row to additional problems. In fact, some students require over 20 practice attempts to reach mastery. ASSISTments, by default, limits the daily practice number for a skill to 10 attempts, so these students need multiple days to master a skill. The current workflow of ARRS is relatively simple: after classroom teaching of a certain skill, teachers use ASSISTments to assign a mastery learning problem set of that skill to students, and students are required to first master the skill by completing the Mastery learning problem set; ARRS will then automatically reassess students on the same skill 7 days later with a retention test, also called the reassessment test in ASSISTments, built from the same sets of problems the student already mastered. If students answer the problem correctly, we treat them as if they are still retaining this skill, and ARRS will test them 14 days later, 28 days later, and then finally 56 days after that. If a student fails the retention test, ARRS will give him an opportunity to relearn the skill. Once a student relearns (demonstrates mastery) a skill, he will receive another retention test at the same delay at which he previously responded incorrectly. In other words, if the student failed the second retention test, he would have to relearn the skill and achieve 3 correct answers in a row, before receiving another reassessment test 14 days later. It is important to note that ARRS will not assign retention tests to students who did not start or finish the mastery learning since assigning such tests to these students misses the purpose of reassessing.

1.2 Prerequisite skill system
Cognitive domains usually have a model that represents the relationship between knowledge components. Each of these knowledge components is a major skill in the domain that students are expected to have. Knowledge components are grouped based on the grade at which students are expected to have those skills. The relationship between these knowledge components is presented as either prerequisite or post-prerequisite. A
prerequisite skill is a skill that students are expected to have to be able to succeed in that given skill. Without knowledge of the prerequisite skill(s) of a given skill, a student is not expected to be responding correctly to questions from that given skill. Postrequisites of a skill are skills that can be taught to students once they have the skill under consideration. The strengths of these relationships vary, but this is not the subject of this paper. The map in Figure 1 is representation of a subset of the prerequisite skill model used by a number of features in ASSISTments. The ovals represent the skills and the arrows linking the ovals show the prerequisite and postrequisite relationship between the skills. The codes are the Massachusetts Common Core State standards for the Math skills [8]. ASSISTments started adopting the Common Core standards since fall 2013.

Cognitive models, together with their skills maps, have been used to determine students’ cognitive levels in a given domain. This determination is done through the use of adaptive assessment systems that traverse the model, presenting problems to students from skills depending on how well they perform at any given point in the test. For example, when a student answers a problem from a given skill incorrectly, problems are presented from the prerequisite skill to determine how well they know the prerequisite skills. This process is followed recursively until the student bottoms out. A student is said to have bottomed out once he gets the last item in the sequence of prerequisites incorrect.

2. INTUITION AND APPROACH
Inheriting the general approach of student modeling means using data about a student’s performance in order to assess his degree of knowledge. But consider a situation where a student has very high performance in general but performed poorly in prerequisite skills to a particular skill, when this student encounters the postrequisite skill, we would not expect him to have robust mastery, therefore, his performance on retention tests to that postrequisite skill could be poor. However, most models only focused student’s general performance or most recent performance. Hence we formed a hypothesis that the prerequisite skill performance can be independent from student local performance and can be used to enhance our models of predicting retention performance.

2.1 Modeling Retention
The first step toward developing a model to predict student retention performance is to create an operational definition of performance that can be used as a training label. As we employed data from the ARRS system, we operationalized retention performance as the reassessment test performance one week after student mastered a skill (i.e., the first retention test). We selected these data as they are the most numerous, and investigating later reassessment tests requires finding a way to represent the student performance on prior reassessment tests and relearning, if any. Thus, for simplicity, we start by modeling performance on the first test. Although since the deployment of ARRS, students have answered over 300,000 retention tests. However, only a small portion of problem sets have been linked with Common Core skills and only 1775 tests from these skills were answered. We separate these data into 1420 instances for the training set and 355 for the testing set; the training set was selected by randomly selecting 20% of the dataset.

Given this definition of retention, we built such models by a data driven approach: rather than attempting to predict every next student performance, instead we focus on student performances that occur on his first attempt at the first retention test for each skill. In this way, even though we are not explicitly modeling the forgetting and retention, our straightforward modeling approach captures aspects of performance that relate to student long-term retention of the skill. Wang and Beck [10] already suggested that features like number of student correct and incorrect responses were not reliable predictors of long-term performance, which is in contrast to most student-modeling efforts on predicting performance on the next response. So naturally, we interest in what factors and features influence the prediction of long-term performance in our ARRS data.

In the study of modeling student delayed performance, we initially noticed [9] that the number of problems required to achieve mastery has great influence on the delayed performance. We refer this number as mastery speed, it represents a combination of how well the student knew this skill originally and how quickly he can learn the skill within ASSISTments. In our previous studies, we observed students performed very differently on the mastery speed so we separated the possible mastery speed into interpretable bins. We observed that, in general, the slower the mastery speed, the lower the probability that the student can answer the problems in the retention test correctly. Students who mastered a skill in 3 or 4 problems had an 82% chance of responding correctly on the first retention test, while students who took 5 to 7 problems to master a skill only had a 71% chance of responding correctly on the first retention test. Finally, there is a group of students who took over 7 problems to master a skill, and who, predictably, did the worst at 62%. Why this result is of interest is that most ITS would credit both students with having mastered the skill, as three correct responses in a row is sufficient for, in most circumstances, Bayesian knowledge tracing to have a P(knows) over 0.95, its criterion for mastery. Thus, not all mastery is alike, with a 10% decrease in percentage correctness between each level of mastery speed. It is surprising that such stark differences should appear given our method of disaggregating mastery was straightforward. It is also important to note that mastery speed does necessarily refer to properties about the student, as is a property of a particular student-skill pair. For example, for a particularly difficult skill, most students could have a slow mastery speed of 7 or 8 attempts. Therefore, a student with such a mastery speed is not below average.

2.2 Modeling Prerequisite Skill Effects
We employed the mastery speed, as well as two other basic features to establish a baseline for our modeling work. These features forced on item and skill information, including: (1) item easiness and (2) skill ID. Note that because we are not using the
identifier of students in the modelling work, thus our models can test our ability to generalize to new students.

We fit this base model using multinomial logistic regression; we got a $R^2$ of 0.182 on testing data set. The p-values of the model show that master speed, and item easiness and skill ID are all reliable predictors. We were driven to find more predictive power from our ARRS dataset.

When building this model, we assumed that student performance is relevant to features of the skill, student and a global item easiness feature. We have the intuition that there might be some other factors between the local performance of a student and the whole population can help improve our prediction. As ASSISTments starting to adopt the Common Core Standards, especially adding the skills system in problem sets meta data, we considered the following couple scenarios: To predict the one-week retention test performance for a certain student on a given skill, if we knew that all of his answers on the prerequisite skill were incorrectly, how would we predict the correctness of the student’s answer? Similarly, how would knowing that all of his answers on the prerequisite skills were correctly affect our prediction? Therefore, we made a hypothesis that the prerequisite skills performance and student retention performance are not independent and, to enhance our model, added features that represent the prerequisite skill performance.

To test our hypothesis, the next step was to gather a set of prerequisite skill features and identify which features can be used as predictors. Towards this end, we selected the following three features to capture different prerequisite skill information:

1. **prerequisite skill ID**: the unique identifier of each prerequisite skill. By modeling skill ID as a factor, we are estimating an overall effect of these skills;

2. **student prerequisite skill performance**: measure a student’s performance on a direct prerequisite skill of the retention test skill, this number is presented by the percentage of correctness of all the problems that answered by the students for this prerequisite skill;

3. **prerequisite skill easiness**: the percentage of correctness for this prerequisite skill across all answers and all students. The higher this value is, the more likely the problems of this prerequisite skill can be answered correctly.

### 3. PREDICTION RESULT ANALYSIS

To investigate how our prerequisite skill features could impact our prediction of student retention test performance, we started from our base model, described previously, and extended it with a representation of the prerequisite skill features. We experimented with using prerequisite skill ID as a factor, student prerequisite skill performance, prerequisite skill easiness, so there are three models to be calculated other than the baseline model. Table 1 provides the results for each of these models, the prediction performance were measured in terms of $R^2$ on the testing set.

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>base model + student prerequisite skill performance</td>
<td>0.189</td>
</tr>
<tr>
<td>base model + prerequisite skill ID</td>
<td>0.185</td>
</tr>
<tr>
<td>base model + prerequisite skill easiness</td>
<td>0.182</td>
</tr>
<tr>
<td>base model</td>
<td>0.182</td>
</tr>
</tbody>
</table>

From above results, we can see that improved models were obtained both on prerequisite skill ID and student prerequisite skill performance. The results from using student prerequisite skill performance clearly indicate that a student’s performance on prerequisite skills is helpful for improving predictions. These results confirmed our intuition about using prerequisite skill information as predictor of student retention test performance. The predictive power of prerequisite skill ID may suggest that there seems to be an overall skill effect differs from the average performance on prerequisite skills, which is modeled by prerequisite skill easiness. However, the performance of prerequisite skill ID may deceptive as instances from the same skill are in both the test and the training set. Although this procedure overstates the results, creating a training set without instances of a particular skill would leave the model unable to generalize at all. For those rare instances when a skill only appeared in the training data, we imputed the value of the various skill parameters by using the mean value observed in the training data.

In terms of whether combining the two features would be fruitful in improving accuracy? While this might not be terribly surprising, a model using both prerequisite skill ID and student prerequisite skill achieved an $R^2$ value of 0.192 and the result is statistically reliable ($p \approx 4.5 \times 10^{-7}$). This led us to believe that these two features are largely independent predictors and whatever prerequisite skill ID represents, it is relatively distinct from student prerequisite skill performance as the $R^2$ increases noticeably when both are modeled. The Beta coefficient values and p-values for each covariate are shown in Table 2.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Beta</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>problem easiness</td>
<td>6.306</td>
<td>.00</td>
</tr>
<tr>
<td>prerequisite skill</td>
<td>2.24</td>
<td>.00</td>
</tr>
</tbody>
</table>

The positive Beta values indicate that the larger the covariate is, the more likely the student responded to this problem correctly. So we see that the easiness of retention test problem is still more like to affect students’ performance compare to their prerequisite skill performance.

### 4. CONTRIBUTIONS

This paper makes two contributions. First, this paper confirms speed of mastery as a useful construct. Bayesian Knowledge Tracing [3] and its extensions [6] have what is known as the Markov assumption that past history is unimportant, as the
required knowledge can be represented with the current state. For example, if a student has an estimated knowledge of 0.8, it does not matter whether he achieved that level with 3 problems or with 30. We have found that such longer term trends are important, and thus researchers should either use a modeling approach that does not have the Markov assumption, or enrich their state representation to better account for longer-term performance. Speed of mastery is unlikely to be considered in the course of more normal uses of student modeling, as predicting student performance post-mastery is an atypical problem, as frequently systems aren’t as concerned with topics the student has already mastered. Thus, this work has confirmed mastery speed as a new feature relevant to robust learning.

Secondly, this paper explored and identified prerequisite skill features as being worth modeling. Most prior works have focused on features at local skills and performance information, such as item difficulty in PFA [7], or prior knowledge as being associated with all students on a particular skill in Bayesian knowledge tracing [3]. We suspect this bias is from the KDD Cup 2010 on Educational Data Mining, as that dataset did not include indicators about skill hierarchy and relationships. However, our analysis benefited from the adopting of Common Core skills in ASSISTments, it accounted for influences that will affect postrequisite knowledge.

5. FUTURE WORK AND CONCLUSIONS
In this work attempt to model prerequisite skill features to better predict student retention performance in intelligent tutoring system on a small dataset. We need to further investigate our model with larger dataset and other data sources like Cognitive Tutor.

In this paper we only investigate the direct prerequisite skill of test skills. We have not yet looked into the skill system as a hierarchy of complete knowledge components, shall we consider the notion of the student's all prerequisite skills performance as being relevant, and what is more, there are other potions. For example, we could measure how well a student did on the retention tests of prerequisite skills. Also, it is possible that skill interference is also affecting the retention performance. Exploring these avenues to discover prerequisite skill impacts on performance is an interesting future direction.

On the other hand, the cognitive meaning mastery speed is still unclear. Do students who achieve mastery quickly already understand the skill, and have retained it from prior instruction, or are they simply learning quickly, and quick learners also retain better. Understanding what speed of mastery means is a difficult problem. Further, we knew that slow mastery speed doesn’t lead to good retention performance, but would a stronger mastery criterion, such as answering 4 or 5 problems currently in a row, be helpful for students? Will it cause more wheel-spinning? We need to tell the difference between learning versus screening out high knowledge students.

6. ACKNOWLEDGMENTS
This part is intentionally left blank for blind reviewing.

7. REFERENCES