Class distinctions: Leveraging class-level features to predict student retention performance

Xiaolu Xiong, Joseph E. Beck, Shoujing Li

Department of Computer Science
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA

xxiong@wpi.edu, josephbeck@wpi.edu, sli@wpi.edu

Abstract. This paper describes our experiments and analysis of utilizing class-level features to predict student performance for retention tests. There are two aspects that make this paper interesting. First, instead of focusing on short-team performance, we investigated student performance after a delay of at least 7 days. Second, we explored several class-level features that can be captured in intelligent tutoring systems (ITS), and we showed that some of them have encouraging predictive power. With the help of class-level features, the prediction result indicated an improvement from an R^2 of 0.183 with a normal feature set to an R^2 value of 0.224.

Keywords: Educational data mining, Feature selection, Knowledge retention, Intelligent tutoring system

1 Introduction

Currently, most ITS present a sequence of problems and, if the student performs well, decide that the student has mastered the skill. Similarly, researchers of educational data mining have investigated the prediction of student behavior on the immediate next action, in other words, student short-term performance [3]. Although performing well on a group of problems is an indicator of mastery, it is by far not the only criteria.

Inspired by the notion of robust learning [1] and the design of the enhanced ITS mastery cycle proposed by Wang and Beck [4], 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. ARRS is an extension of the ASSISTments system (www.assistments.org). The idea of ARRS is if a student masters a problem set with three correct responses in a row, 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: firstly 7 days after mastery, then 14 days, 28 days and 56 days after the very first test. If a student fails the reassessment test, ASSISTments will give him an opportunity to relearn the

adfa, p. 1, 2011.

skill. Relearning means that the student must again demonstrate mastery by responding correctly to three items in a row. Once a student relearns a skill, he will receive another reassessment test at the same time delay at which he previously responded incorrectly.

2 Intuition and approach

In general, student modeling uses data about a student's performance in order to assess his degree of knowledge. However, consider a situation where all of a student's classmates respond incorrectly to a particular item. When this student encounters the item, we would not expect him to respond correctly based on his peers' performance. Strangely, most student modeling approaches would not take advantage of this information, even though it is presumably relevant to understanding this student's knowledge. We formed a hypothesis that the class performance and student individual performance are not independent and can be used to enhance our models. However, in the study of ARRS data, we initially noticed that the number of attempted problems before students achieve mastery has great influence on the one-week delayed performance [5].

2.1 Modeling retention

At a minimum, students require 3 correct attempts to master a skill. If a student gets the first item wrong, he could master the skill in 4 attempts. We refer to the number of problems required as the *mastery speed* that represents a combination of how well the student knew this skill originally, and how quickly he can learn the skill. 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 over 8 attempts to master a skill only had a 59% chance of responding correctly on the first retention test. Finally, there is a group of students who tried but failed to master the skill, and who, predictably, did the worst.

2.2 Modeling class-level effects

To test our hypothesis of class-level features, we selected the following three features to capture different class-level information: (1) *class_id*: classes were created by teachers who are using the ASSISTments, and represent each distinct class a teacher has. By modeling *class_id* as a factor, we are estimating an overall effect of the classroom. (2) *class_prior_performance*: measures the class' performance on prior reassessment tests on same skill. For each reassessment test, the performance is represented by using the percentage of correctness of tests that have been answered in the same class, on the same skill, and have been answered before the student attempts this retention item. (3) *class_other_skill_performance*: measures the class' performance on all reassessment tests on all other skills. This feature is permitted to use data from the future, and is thus not realistic in an actual system, but provides an upper bound for how well such information could work.

3 Model results

To train our model, we used 42,332 instances of a student using the ARRS system and attempting the first retention test for each skill. We separated these pieces of data into 33,866 instances for the training set and 8,466 for the testing set. The testing set was selected by randomly choosing 20% of the dataset, so there is an issue of non-independence as the same student appears in both sets. We first employed the *mastery speed*, as well as three other basic features, to establish a baseline for our modeling work. These features forced on item and skill information, including: (1) *on_grade*, whether this skill is typically taught in the same grade-level of the student. (2) *grade_diff*, the binned value of grade difference and (3) *item_easiness*. We fitted this base model using multinomial logistic regression; we got an R^2 of 0.183.

To investigate how our class-level features could impact our predictions on student retention test performance, we started from our base model, described previously, and added to it a representation of the class' performance. We experimented with using the *class_id* as a factor, prior performance on this skill's retention test, and all performance on all retention tests that did not involve this skill. Table 1 provides the results for each of these models. We provide both the classic R^2 metric, as well as the Nagelkerke (pseudo) R^2 for comparison purposes as other logistic regression results reported have used Nagelkerke [2].

Table 1. Class-level model performance

Model	R^2 on training set	R^2 on testing set
Base model + *class_id*	0.158 (Nagelkerke: 0.215)	0.159
Base model + *class_prior_performance*	0.155 (Nagelkerke 0.204)	0.153
Base model + *class_other_skill_performance*	0.145 (Nagelkerke 0.185)	0.142
Base model	0.143 (Nagelkerke 0.183)	0.142

From the above results, we can see that new model with *class_id* and *class_prior_performance* performed slightly better than the base model. The importance of *class_id* in the prediction may suggest that there seems to be an overall class effect that differs from average performance on other skills, which is modeled by *class_other_skill_performance*. One question is whether combining the two features would be fruitful in improving accuracy? Somewhat surprisingly, a model using both *class_id* and *class_prior_performance* achieved an R^2 value of 0.165 (Nagelkerke 0.224). Thus, whatever *class_id* represents, it is relatively distinct from *class_prior_performance* as the R^2 increases noticeably when both are modeled.

4 Contributions, Future work and conclusions

This paper makes three contributions. Firstly, this paper identifies speed of mastery as a useful new feature relevant to robust learning. Secondly, this paper explored and identified class-level effects as being worth modeling. Our analysis adopted class-level features in order to account for influences that will affect all members of the class. The third contribution of this paper is by employing class id in our prediction; we adopted a generic approach for intuitively “clustering” students. Our approach of clustering requires little additional information, no complex processing, and it is easy to understand our clusters and the semantics behind them.

For examining class-level effects and predicting retention, we used a classifier with features that were known to be predictive, such as *mastery speed*. There are many follow-up problems that we are interested in: Are there better ways of using the class-level data? How well has this teacher’s classes done in preceding years? Does this teacher’s students systematically under- or over-perform on retention tests? Exploring these avenues to discover class-level impacts on performance is an interesting future direction.

This paper has presented a problem of predicting whether students will retain information after a delay of 7 days. We found that mastery alone is insufficient to predict retention, and the ease with which students achieve mastery is critical. However, the cognitive meaning of this statement is 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. One other clear conclusion is that class matters, and the performance of the students’ peers is useful for predicting his performance.

References

1. Baker, R., Gowda, S., Corbett, A., & Ocumpaugh, J. (2012). Towards automatically detecting whether student learning is shallow. In Intelligent Tutoring Systems, 444-453. Springer Berlin/Heidelberg.
2. Gong, Y., Beck, J. & Heffernan, N. (2010). Comparing Knowledge Tracing and Performance Factor Analysis by Using Multiple Model Fitting. Proceedings of the 10th International Conference on Intelligent Tutoring Systems, 35-44.
3. Pardos, Z. A., & Heffernan, N. T. (2010). Using HMMs and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset. Journal of Machine Learning Research W & CP.
4. 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.
5. Xiong, X., Li, S., Beck, J. (Accepted) Will you get it right next week: Predict delayed performance in enhanced ITS mastery cycle. The 26th International FLAIRS Conference.