Applying Learning Factors Analysis to Build Stereotypic Student Models

Anna N. RAFFERTY[A] and Michael YUDELSON[B]
[A]: *Stanford University, rafferty@stanford.edu*
[B]: *University of Pittsburgh*

Abstract. This paper demonstrates how stereotypic student groups can be created to enhance cognitive models in computer tutors. Computer tutors use cognitive models to track what skills students are learning and what practice attempts are most needed; often, these models contain relatively few individualized student parameters due to computational concerns. A hybrid approach emphasizing tractability and customization is used to balance the need for computable cognitive models and more flexibility to reflect learner characteristics. We use learning factors analysis to incorporate these learner characteristics and demonstrate that the resulting groups do require different cognitive models. In particular, learning rates and the actual skills that students seem to be learning based on mathematical models differ between groups.

Keywords. Cognitive Models, Cognitive Tutors, LFA, Educational Data Mining

Introduction

A number of intelligent computer tutoring systems exist to assist with student instruction or practice (for a review see [1]). These systems have the potential to greatly enhance education by allowing students to work at their own pace on individualized material and access the teacher more readily when additional assistance is necessary, as the teacher will no longer need to spend as much time addressing basic questions with which the tutoring program can assist. However, this vision cannot be realized if tutors cannot respond to each individual student in a customized manner.

1. Literature Review

1.1. Student Models

In this paper, we examine the issue of increasing intelligent tutoring systems ability to consider individual student learning patterns. Currently, such customization is hampered by the large increase in complexity that results from adding many student specific modifications to the software; most intelligent tutoring systems use some type of mathematical model to track what students are learning, and the model increases in complexity as more parameters are added [2]. A larger number of parameters makes programming the software more difficult and poses quandaries for determining what the initial values of these student parameters should be. While [3] notes the promise of Bayesian networks for working with complex student parameters, their approach requires considerable information about higher level parameters and fails to discover commonalities among student learning characteristics. Additionally, since students must be modeled from the first problem they attempt, it is difficult to use only data from the current student to set model parameters because data are initially sparse. We seek a solution that balances the complexity of the overall model and the need to mediate responses based on students' performance. Rich [4] suggests that students are more likely to be better modeled by a model fitting a subset of the student population, chosen based on student characteristics, than by a model for the entire set of students; we follow this strategy by creating multiple group models in which each has the same number of parameters as the original model.

One component of creating an educational software program is forming the model that monitors student behavior. Authors have created cognitive models to represent the skills that students are using, often programmed as production rules, that can then be used in intelligent tutoring systems [2]. For example, one skill in a geometry tutor might be determining the perimeter of a trapezoid; this would be a separate skill from finding the area of the trapezoid or the perimeter of another shape. As students work through the tutoring program, the model tracks their progress and chooses what problems will de displayed next [2]. When models are not student-specific, data from previous student tests can be used to determine such factors as the probability that a skill will be known prior to practice and the rate at which learning may occur on a specific skill [5]. By using stereotypic student groups we gain the benefits of using parameter estimates derived from previous tests of the program. This simplifies modeling calculations and allows student modeling to be implemented more quickly than might be possible if individual estimates were calculated during tutoring.

1.2. Learning Factors Analysis

Currently, many cognitive models are manually created by experts or are learned from student behavior, which may be difficult to interpret or incomplete. The former is problematic as well, both because of the time necessary to build a full model of a domain and such problems as the expert blind spot. However, algorithms exist to refine existing cognitive models, lessening the burden on experts and reducing bias in the model. These algorithms generally redefine the skills in the model, determining when one skill may be better defined as two or when two skills might be better combined into one. One such algorithm is Learning Factors Analysis (LFA) [2]. LFA is an extension of the power law of learning [6], which represents the exponential decrease in error rate that occurs with increasing number of opportunities to practice a skill. While this power law is defined only over one skill and one particular student, LFA models multiple students and multiple skills by adding student and skill intercepts and skill learning rates, as represented by:

$$\ln\left(\frac{p}{1-p}\right) = \sum \alpha_i X_i + \sum \beta_j Y_j + \sum \gamma_j Y_j T_j$$

p = probability of getting an item right; X, T, Y, YT = covariates for students and skills

i = number of students, j = number of skills

This increases the space of models to allow for students with different initial skills levels, variations in the probability that each skill has already been learned, and variations skill learning rates. Notably, however, this does not allow for variations in the rate at which each student learns; [2] comments that this was justified in their application because the authors sought to improve the model globally rather than improve individual student models. We seek to extend the applications for LFA.

Once the space of models has been defined, LFA searches this space using A* search, guided by the Bayesian Information Criterion (BIC) [2]. BIC balances model complexity with the likelihood of the model given the data to avoid overfitting:

$$BIC = 2 * \log\text{-likelihood} + \text{number of parameters} * \text{number of observations}$$

To move through the space of possible models, LFA splits skills based on difficulty factors. Difficulty factors must be defined a priori on the problem space and are used to indicate certain shared features across problems that are not currently considered in the skill model. For example, one difficulty factor in the data set we use is the number of times that the skill had to be applied to solve a problem. Intuitively, applying the same skill more than once may add difficulty; by defining a difficulty factor over this parameter, one can empirically determine if these problems seem to be using the same skill (as defined in the original cognitive model) or two different skills. LFA allows one additional skill to be added for each value of each difficulty factor [2].

2. Methods

The first portion of this research involved separating students into groups based on learning and knowledge characteristics to determine if different groups of students have qualitatively different patterns for learning different skills. The dataset consisted of over 4000 data points generated by 24 students while using the area unit of the Geometry Cognitive Tutor. The original data model contained 15 knowledge components and 5 difficulty factors defined on these components. These knowledge components focused on finding area and side/height lengths for various shapes.

Students were first categorized into groups by calculating average success rate on each student's first encounter with a problem of each knowledge component type. The first attempt for each skill was used to avoid conflating greater success based on learning from the tutor and greater success based on greater initial knowledge; this approach assumes an overall initial knowledge factor, independent of skills.

Students were also split into groups based on learning gain. Learning gain was considered to be the difference between the student knowledge parameter when calculated on the first half of problem attempts and the student knowledge parameter when calculated on the second half of problem attempts based on when problems were encountered. Each of these halves was treated as an assessment period in which no knowledge was gained, dropping the time component from the model.

The log of the ratio of the probability that the skill is known to the probability it is not known (logit) is equal to the sum of the student initial knowledge probability and the skill initial knowledge probability; the learning rate is irrelevant because we assume no learning occurs. These initial probabilities are determined by fitting a modified LFA regression equation with the learning parameter removed on each half of the problem attempts. In order to assess how student knowledge changed between the two halves, we perform a least squares fit to find student intercepts. The beta parameters are held constant, so changes in success rate are accounted for solely by student parameters.

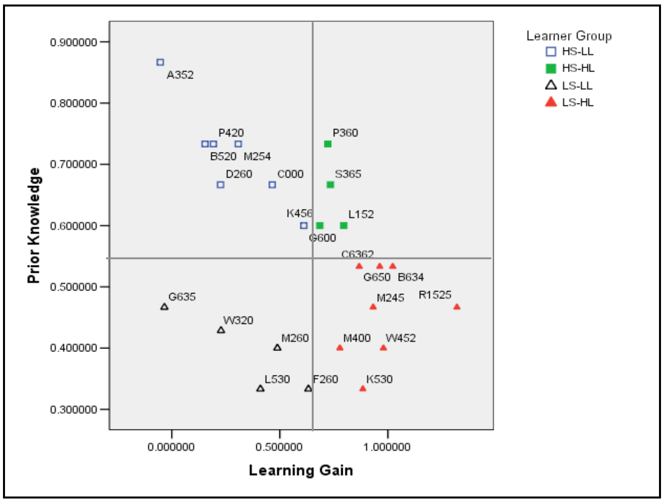

Figure 1. Student groups created by splitting students based on learning gain and initial knowledge; HS/LS correspond to High/Low Starting Skill Level, HL/LL correspond to High/Low Learning Gain

Using these two types of splits, students were broken into four groups (Figure 1). Students' initial knowledge was only weakly correlated with learning gain (r^2=.17); thus, these splits create groups with distinct learning and knowledge characteristics.

3. Results

3.1. Differences in Model Fit

No differences in model fit were found when applying LFA to subgroups split on only one of learning gain and initial knowledge. However, when LFA was applied to the four groups of students, significantly different learning rates were found between groups (see Figure 2); the original model could not express these differences in learning rates (shown by curve slopes)**.** This result thus empirically challenges the assumption by [2] that learning rates can be held constant across students. While this result might be expected given that students were split on learning gain, it was not entirely predicted by this split. Particularly pronounced effects were found when comparing students with low initial knowledge who had high versus low learning (Figure 2); we also found that of high learning students, those with low initial knowledge showed more efficient learning than those with high initial knowledge.

3.2. Differences in Knowledge Components Between Models

Another compelling result was found by adjusting the skills in the cognitive model. We compounded the skills into seven basic skills, each of which represented problems involving a single shape type or were more general arithmetic skills unrelated to any shape type; for example, problems involving the skills "trapezoid-area," "trapezoid-base," and "trapezoid-height" were all reclassified as involving the "trapezoid" skill in the new model. As empirically suggested by [7] and shown theoretically by the learning curve equations, a compact model allows for faster mastery of skills if

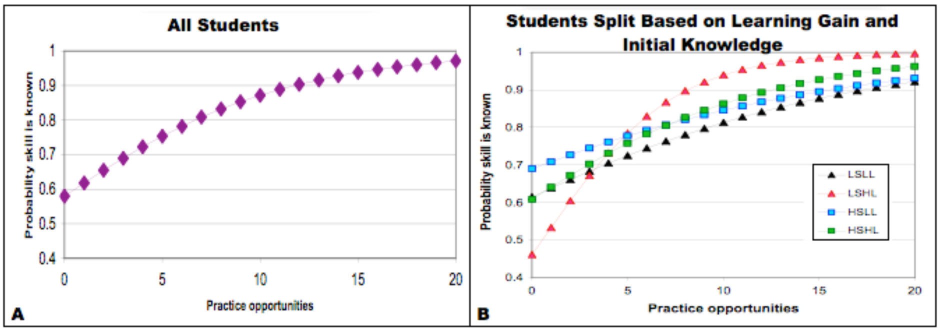

Figure 2. Comparison of median learning curves (calculated by taking the intercept of a median student in the group and a median learning rate and intercept from the fitted skills) based on fitted model. A) Median learning curve without student groups. B) Median learning curves for each of the four student groups.

learning rates do not decrease sufficiently to compensate for the greater number of practice attempts for each skill. For instance, consider a model with four skills and five practice attempts for each skill. If the model can be refactored into a more compact state in which one of the original four skills is combined with another, then the new skill will have twice as many attempts as the original skill. Given the learning curve equation $Y = aX^{-b}$, we see that if the learning rate (X) and the original difficulty of the skill (a) do not increase too much, then all 10 attempts may not be needed to achieve mastery; if the number of attempts necessary for mastery is decreased, the speed of learning is increased. Thus, by fitting a compact model to the data from each of the groups, we investigated the efficiency of the domain model that each group learned.

We found the compact model fit the students with low initial knowledge and high learning gain best. When LFA was applied to these data using the compact model, no reliable skill splits were found among the top three models, and the average number of skill splits for the top three models was less than one; the compact model provided the best fit to the data of the models in the LFA search space. In contrast, when LFA was applied to the data from each of the other groups, one or more reliable skill splits were found, and the average number of skill splits for the top three models was greater than 1 for each group. This provides one suggestion for why the low initial knowledge and high learning gain shows a much steeper learning curve than the other groups; these students are learning a domain model in which each skill attempt is equivalent to multiple attempts on different skills for the other groups.

4. Discussion

The data presented in this paper show that it is not enough to use a uniform learning rate to model students; students clearly differ not only in their initial knowledge but also in the number of practice attempts they require to learn new skills. This suggests that existing cognitive models still have room for improvement, and future revisions of such models might benefit from a hybrid of student and general model parameters, as shown by the stereotypic student group approach used here. Additionally, future research should examine whether the differences in learning rates between students are constant across units or other divisions of subject material. If this is the case, then models might benefit from estimating which group a student belongs in for the next unit based on her or his performance in the current unit. This suggests a method by which student may be categorized into a group while using the tutor: initial knowledge can be estimated for the previous unit, while learning gain can be predicted from the previous unit. This assumes that the student's ability to learn from the tutor will not change dramatically over short time spans. Additionally, while we found differences between the group with the highest characteristic learning rate and the other groups, both in the average number of skill splits and the reliability between models of the skill splits, the differences among the other groups were less well defined. In particular, results were mixed for the groups with high levels of prior knowledge. For these groups, there was not a great deal of difference in the number of skill splits reliably found, and the average number of skill splits was actually higher for the group with greater learning gain, a result that was the opposite of what we expected given [7].

Despite the need for further investigation, this study provides valuable information about what assumptions can be made about learners. By suggesting a hybrid approach to individual student characteristics, we hope to encourage creators of cognitive models to incorporate traits that vary among students. Additionally, we seek to increase interest in understanding precisely what students are learning and whether student domain models are consistent across students and match domain models stored by the tutor.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. SLC-0354453. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We would like to acknowledge the Pittsburgh Science of Learning Center Summer School for providing the training and opportunity to conduct much of this research, and to thank Daniel Schwartz, Joseph Beck, and Kenneth Koedinger for their insightful comments on this research as it progressed, and Hao Cen for his insight and assistance with LFA.

References

[1] Mitrovic, A., Koedinger, K. R., & Martin, B. (2003). A comparative analysis of cognitive tutoring and constraint-based modeling. In P. Brusilovsky, A. Corbett, & F. de Rosis (Eds.), *Proceedings of the Ninth International Conference on User Modeling,* 313-322. Berlin: Springer-Verlag.

[2] Cen, H., Koedinger, K. R., & Junker, B. (to appear). Learning Factors Analysis – A general method of cognitive model evaluation and improvement. In Proceedings of Intelligent Tutoring Systems, 2006.

[3] Mislevy, R.J., Almond, R. G., Yan, D., Steinberg L.S. Bayes nets in educational assessment: where the numbers come from. In K.B. Laskey & H. Prade (Eds.), *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. 437-446. San Francisco: Moran Kaufmann.

[4] Rich, E. (1979) User Modeling via Stereotypes. *Cognitive Science: A Multidisciplinary Journal*, 3(4), 329-354.

[5] Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: modeling the acquisition of procedural knowledge. *User Modeling and User Adapted Interaction,* 4(4), 253-278.

[6] Newell, A, & Rosenbloom, P.S. (1993). Mechanisms of skill acquisition and the law of practice. In P. S. Rosenbloom, J. E. Laird, & A. Newell (Eds.), *The Soar Papers: Research on integrated intelligence*. Cambridge, MA: MIT Press.

[7] Koedinger, K. R., & Mathan, S. (2004). Distinguishing qualitatively different kinds of learning using log files and learning curves. In the Working Notes of the ITS2004 Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes.