The Student Skill Model

Yutao Wang, Neil T. Heffernan

Worcester Polytechnic Institute
Department of Computer Science
yutaowang@wpi.edu, nth@wpi.edu

Abstract. One of the most popular methods for modeling students' knowledge is Corbett and Anderson's[1] Bayesian Knowledge Tracing (KT) model. The original Knowledge Tracing model does not allow for individualization. Recently, Pardos and Heffernan [4] showed that more information about students' prior knowledge can help build a better fitting model and provide a more accurate prediction of student data. Our goal was to further explore the individualization of student parameters in order to allow the Bayesian network to keep track of each of the four parameters per student: prior knowledge, guess, slip and learning. We proposed a new Bayesian network model called the Student Skill model (SS), and evaluated it in comparison with the traditional knowledge tracing model in both simulated and realword experiments. The new model predicts student responses better than the standard knowledge tracing model when the number of students and the number of skills are large.

Keywords: Knowledge Tracing, Individualization, Bayesian Networks, Data Mining, Prediction, Intelligent Tutoring Systems

1 Introduction

One of the most popular methods for modeling students' knowledge is Corbett and Anderson's[1] Bayesian Knowledge Tracing model. The original Knowledge Tracing model does not allow for individualization. Several researchers have tried to show the power of individualization. Corbett and Andersen presented a method to individualize students' parameters with a two phase process and reported mixed results[2]. Recently, Pardos and Heffernan [4] showed that by a single process Bayesian network model: the prior per student model, more information about students' prior knowledge can help better fit model and provide more accurate prediction of student data. The result is inspiring; however, the author only looked into the students' prior knowledge and didn't extend the individualization to the other aspects of student knowledge, such as guess rate or learning rate. Pardos and Heffernan [5] also tried a method where they trained all four parameters per student in a pre-process, then took those values and put them into a per skill model to learn how the user parameters interacted with the skill. This method requires a two phase data process, which is complicated to use in real-world.

Our goal was to further explore the individualization of student parameters in order to allow the Bayesian network to keep track of all our parameters per student as well as skill specific parameters simultaneously. We proposed a new Bayesian network model called the Student Skill model (SS), and evaluated it in comparion to the traditional Knowledge Tracing model (KT) in both simulation and real data experiments. The new model predicts student responses better than standard knowledge tracing model when the number of students and the number of skills are large.

2 The Student Skill Model

The Knowledge Tracing model assumes that all students have the same probability of knowing a particular skill at their first opportunity, or guess/slip in one skill, or learning a particular skill even though students seem likely differ in these aspects. Our goal was to add individualization into the original Knowledge Tracing model.

The new model we proposed in this paper is called the Student Skill model. It can learn four student parameters and four skill parameters simultaneously in a single phase process. The model is shown in Fig.1.

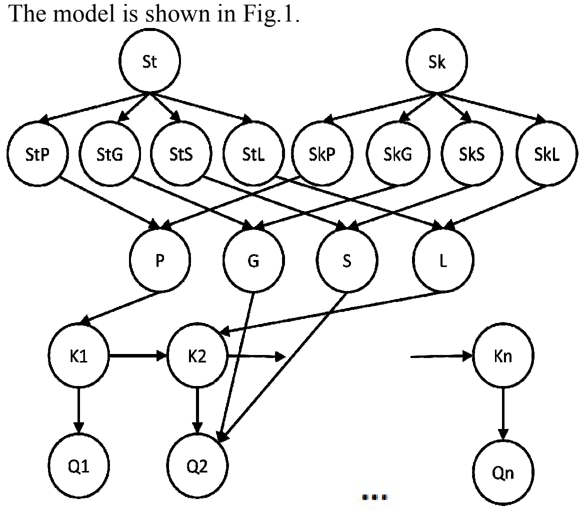

Fig. 1. The Student Skill model

The lowest two levels of this model are the same as the original Knowledge Tracing model (nodes K1~Kn and Q1~Qn in Fig.1). The Student Skill model adds upper levels to represent the student and skill information and their interaction. We used two multinomial nodes to represent the identity of each student (node St in Fig.1) and each skill (node Sk in Fig.1). Instead of pointing the student identity and the skill identity nodes directly to the knowledge nodes, which would result in a huge number of parameters, we added a level of nodes to represent the four student parameters (node StP, StG, StS and StL in Fig.1) and the four skill parameters (node SkP, SkG, SkS and SkL in Fig.1). Those parameter nodes are binary nodes that represent the high/low level of the corresponding parameters. For example, if the StP node is 1 for a student, then the student has high level of prior knowledge, and if the StP node is 0 for a student, means the student has low level of prior knowledge. The next level uses conditional probability tables to combines the influence of the student parameters and the skill parameters and generates the four standard Knowledge Tracing parameters (node P, G, S and L in Fig.1) to be used in the lowest two levels.

The number of parameters in this model for *n* students and *m* skills can be computed as: $4n + 4m + 16$, while the number of parameters in the Knowledge Tracing model is: $4m$. The cost of individualization is the additional $4n + 16$ parameters.

3 Model Evaluation

The model is evaluated in both simulated and real data experiments. In our experiments, we used the Bayes Net Toolbox for Matlab developed by Murphy [3] to implement the Bayesian network student models and the Expectation Maximization (EM) algorithm to fit the model parameters to the dataset. We choose initial parameters for each skill in Knowledge Tracing as follows: initial knowledge = 0.5, learning = 0.1, guess = 0.1, slip = 0.1.

3.1 Simulation Experiments

Methodology.

To evaluate the ability of the Student Skill model to function properly, in this experiment, we generated data from the Student Skill model and compared the prediction accuracy with the Knowledge Tracing model. The data records generated in the simulation represent student performances, with 1 representing correct and 0 representing incorrect. To simulate the random noise in the real data, we randomly flipped over 1% of the student performance data.

To split the training and testing data set, for each student, we randomly selected half of the skills data and put them into a training set. The remaining data went to the testing set. Both the Knowledge Tracing model and Student Skill model were trained and tested on the same dataset. A sequence of performances of given students and skills were predicted by both of these models.

Results.

Prediction accuracy is the selected metric for evaluating the results. In one simulation, the number of skills was set at 30 while the number of students was changed from 5 to 100 to observe the influence the number of student had on SS and KT respectively. Similarly, in another simulation, the number of students was set to be 30 while the number of skills was changed.

We observed that, in situations with a small number of students as well as those with a small number of skills, the Knowledge Tracing model outperformed the Student Skill model. However, when the number of students and the number of skills were increased, the performance of the Student Skill model improved and eventually exceeded the Knowledge Tracing model. The reason for this trend could be the fact that the Student Skill model contains more parameters than the Knowledge Tracing model, and with fewer data points, the model behaves less reliably.

We also compared the Student Skill model and the Knowledge Tracing model under different student parameter variance. The number of students and the number of skills were both set to 40, and the number of data points per student per skill was set to 10. The student variance was controlled by the real parameters used to generate simulated data. When the student variance was 0, all students shared the same parameters. We observed that the Student Skill model performs worse when there is no variance in student parameters and when the students are highly variant, the Student Skill model outperformed the Knowledge Tracing model.

3.2 Real Data Experiments

One of the dangers of relying on simulation experiments is that the dataset may not reflect real-world conditions. Without evaluation using real data, the success of the new model during simulation could simply be caused by the data being generated from this model. To further evaluate the Student Skill model, we applied it to real datasets and again compared its performance with the Knowledge Tracing model.

Dataset.
The data used in the analysis presented here came from the ASSISTments platform, a freely available web-based tutoring system for 4th through 10th grade mathematics. We randomly pulled out the data of one hundred 12-14 year old 8^{th} grade students and fifty skills from September 2010 to September 2011 school year. There are 53,450 total problem logs in the dataset.

Methodology.
The dataset was randomly split into four bins by student and skill in order to perform a four-fold cross-validation of the predictions and increase the reliability of the results. For each student, we made a list of the skills the student had seen and split that list randomly into four bins, placing all data for that student and that skill into the respective bin. There were four rounds of training and testing, during each round a different bin served as the test set, and the data from the remaining three bins served as the training set. Again, both the Knowledge Tracing model and the Student Skill model were trained and tested on the same dataset. A sequence of performances of the given students and skills were predicted by both of these models.

Results.
The accuracy of the prediction was evaluated in terms of the Root Mean Squared Error (RMSE). A lower value means higher accuracy. The cross-validation results are shown in Table 1.

Table 1. RMSE results of KT vs SS.

Fold ID	*SS*	*KT*	*P value*	*Student Level P value*
Fold1	0.4017	0.4055	0.0432	0.0404
Fold2	0.4194	0.4385	0.0459	0.0365
Fold3	0.4144	0.4348	0.0477	0.0451
Fold4	0.4441	0.4538	0.0420	0.0406
average	0.4199	0.4331	-------	--------

To test the reliability of the four folds experiment, we did a paired T test for each fold as well as the result of all the folds. The P value that compares the final RMSE of the SS model and the KT model of the four folds is 0.0439. The P value for each individual fold is shown in the fourth column. Our experiment shows that the difference between SS and KT is statistically significant, and the average RMSE shows that SS is more accurate than KT under our experimental conditions. We also did reliability analysis by computing RMSE for each student to account for the non-independence of actions within each student's dataset, and then compared each pair of models using a two tailed paired t-test. The Student Level P values are reported in the last column. All the results are statistically reliable.

4 Discussion and Future work

In this paper, we built a new Bayesian network model for modeling individual student parameters called the Student Skill model and compared it with the knowledge tracing model in both simulation and real data experiments.

In our experiments, we found that the Student Skill model is not always better than the Knowledge Tracing model. Under simulatied conditions, we found that the new model is generally more accurate when the amount of students and skills are large. We are interested in other features that can indicate which model works batter under what situations, in the hope that these two models can be combined in order to utilize both models' advantages.

5 Contribution

Several researchers have tried to show the power of individualization. Corbett and Andersen's presented a method to individualize students' parameters with a two phase process: first run Knowledge Tracing on all the students and then run a separate regression to learn a set of slip, guess, learning and prior parameters per students. Pardos and Heffernan [4] explored the individualized student prior, but did not learn all of the student parameters and skill parameters in one single model. We presented the SS model, which is elegant in accounting for individual differences (of learning rate, prior knowledge and guess and slip rates). Our simulation showed that we could reliably fit such a model. The simulation showed plausible results, such as that the SS model is better if more variation per student.

Our contribution is in presenting a model that allows us to use EM to learn parameters individualized to each student, while at the same time learn parameters for each skill. We presented simulation and real data experiments that showed this method can provide meaningful results. Knowledge Tracing is a special case of this model and can be derived by fixing the student parameters of the Student Skill model to the same values. In a practical sense, researchers need to figure out when the SS model can start to be used, as our simulation showed that SS is better than KT when 1) the number of skills a student has learned is high, and 2) the number of students is high.

Acknowledgements

This research was supported by the National Science foundation via grant "Graduates in K-12 Education" (GK-12) Fellowship, award number DGE0742503 and Neil Heffernan's CAREER grant. We also acknowledge the many additional funders of ASSISTments Platform found here: http://www.webcitation.org/5ym157Yfr
All of the opinions expressed in this paper are those solely of the authors and not those of our funding organizations.

References

1. Corbett, A., Anderson, J. (1995). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction* 4:253-278.
2. Corbett A. and Bhatnagar A. (1997). Student Modeling in the ACT Programming Tutor: Adjusting a Procedural Learning Model with Declarative Knowledge. In User Modeling: Proceedings of the 6th International Conference, pp. 243-254.
3. Murphy, K.P. 2001, The Bayes Net Toolbox for Matlab, Computing Science and Statistics: *Proceedings of Interface*, 33.
4. Pardos, Z.A., Heffernan, N.T. 2010. Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing. *In Proceedings of the 18th International Conference on User Modeling, Adaptation and Personalization*. pp. 225-266.
5. Pardos, Z.A., Heffernan, N. T.: Using HMMs and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset. To appear in the Journal of Machine Learning Research W & CP, In Press.