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. 255-266. Big Island, Hawaii.

Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing

Zachary A. Pardos[1], Neil T. Heffernan

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

Abstract. The field of intelligent tutoring systems has been using the well known knowledge tracing model, popularized by Corbett and Anderson (1995), to track student knowledge for over a decade. Surprisingly, models currently in use do not allow for individual learning rates nor individualized estimates of student initial knowledge. Corbett and Anderson, in their original articles, were interested in trying to add individualization to their model which they accomplished but with mixed results. Since their original work, the field has not made significant progress towards individualization of knowledge tracing models in fitting data. In this work, we introduce an elegant way of formulating the individualization problem entirely within a Bayesian networks framework that fits individualized as well as skill specific parameters simultaneously, in a single step. With this new individualization technique we are able to show a reliable improvement in prediction of real world data by individualizing the initial knowledge parameter. We explore three difference strategies for setting the initial individualized knowledge parameters and report that the best strategy is one in which information from multiple skills is used to inform each student's prior. Using this strategy we achieved lower prediction error in 33 of the 42 problem sets evaluated. The implication of this work is the ability to enhance existing intelligent tutoring systems to more accurately estimate when a student has reached mastery of a skill. Adaptation of instruction based on individualized knowledge and learning speed is discussed as well as open research questions facing those that wish to exploit student and skill information in their user models.

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

1 Introduction

Our initial goal was simple; to show that with more data about students' prior knowledge, we should be able to achieve a better fitting model and more accurate prediction of student data. The problem to solve was that there existed no Bayesian network model to exploit per user prior knowledge information. Knowledge tracing

[1] National Science Foundation funded GK-12 Fellow

(KT) is the predominant method used to model student knowledge and learning over time. This model, however, assumes that all students share the same initial prior knowledge and does not allow for per student prior information to be incorporated. The model we have engineered is a modification to knowledge tracing that increases its generality by allowing for multiple prior knowledge parameters to be specified and lets the Bayesian network determine which prior parameter value a student belongs to if that information is not known before hand. The improvements we see in predicting real world data sets are palpable, with the new model predicting student responses better than standard knowledge tracing in 33 out of the 42 problem sets with the use of information from other skills to inform a prior per student that applied to all problem sets. Equally encouraging was that the individualized model predicted better than knowledge tracing in 30 out of 42 problem sets without the use of any external data. Correlation between actual and predicted responses also improved significantly with the individualized model.

1.1 Inception of knowledge tracing

Knowledge tracing has become the dominant method of modeling student knowledge. It is a variation on a model of learning first introduced by Atkinson in 1972 [1]. Knowledge tracing assumes that each skill has 4 parameters; two knowledge parameters and two performance parameters. The two knowledge parameters are: initial (or prior) knowledge and learn rate. The initial knowledge parameter is the probability that a particular skill was known by the student before interacting with the tutor. The learn rate is the probability that a student will transition between the unlearned and the learned state after each learning opportunity (or question). The two performance parameters are: guess rate and slip rate. The guess rate is the probability that a student will answer correctly even if she does not know the skill associated with the question. The slip rate is the probability that a student will answer incorrectly even if she knows the required skill. Corbett and Anderson introduced this method to the intelligent tutoring field in 1995 [2]. It is currently employed by the cognitive tutor, used by hundreds of thousands of students, and many other intelligent tutoring systems to predict performance and determine when a student has mastered a particular skill.

It might strike the uninitiated as a surprise that the dominant method of modeling student knowledge in intelligent tutoring systems, knowledge tracing, does not allow for students to have different learn rates even though it seems likely that students differ in this regard. Similarly, knowledge tracing assumes that all students have the same probability of knowing a particular skill at their first opportunity.

In this paper we hope to reinvigorate the field to further explore and adopt models that explicitly represent the assumption that students differ in their individual initial knowledge, learning rate and possibly their propensity to guess or slip.

1.2 Previous approaches to predicting student data using knowledge tracing

Corbett and Anderson were interested in implementing the learning rate and prior knowledge individualization that was originally described as part of Atkinson's model of learning. They accomplished this but with limited success. They created a two step process for learning the parameters of their model where the four KT parameters were learned for each skill in the first step and the individual weights were applied to those parameters for each student in the second step. The second step used a form of regression to fit student specific weights to the parameters of each skill. Various factors were also identified for influencing the individual priors and learn rates [3]. The results [2] of their work showed that while the individualized model's predictions correlated better with the actual test results than the non-individualized model, their individualized model did not show an improvement in the overall accuracy of the predictions.

More recent work by Baker et al [4] has found utility in the contextualization of the guess and slip parameters using a multi-staged machine-learning processes that also uses regression to fine tune parameter values. Baker's work has shown an improvement in the internal fit of their model versus other knowledge tracing approaches when correlating inferred knowledge at a learning opportunity with the actual student response at that opportunity but has yet to validate the model with an external validity test.

One of the knowledge tracing approaches compared to the contextual guess and slip method was the Dirichlet approach introduced by Beck et al [5]. The goal of this method was not individualization or contextualization but rather to learn plausible knowledge tracing model parameters by biasing the values of the initial knowledge parameter. The investigators of this work engaged in predicting student data from a reading tutor but found only a 1% increase in performance over standard knowledge tracing (0.006 on the AUC scale). This improvement was achieved by setting model parameters manually based on the authors understanding of the domain and not by learning the parameters from data.

1.3 The ASSISTment System

Our dataset consisted of student responses from The ASSISTment System, a web based math tutoring system for 7th-12th grade students that provides preparation for the state standardized test by using released math problems from previous tests as questions on the system. Tutorial help is given if a student answers the question wrong or asks for help. The tutorial help assists the student learn the required knowledge by breaking the problem into sub questions called scaffolding or giving the student hints on how to solve the question.

2 The Model

Our model uses Bayesian networks to learn the parameters of the model and predict performance. Reye [6] showed that the formulas used by Corbett and Anderson in their knowledge tracing work could be derived from a Hidden Markov Model or Dynamic Bayesian Network (DBN). Corbett and colleagues later released a toolkit [7] using non-individualized Bayesian knowledge tracing to allow researchers to fit their own data and student models with DBNs.

2.1 The Prior Per Student model vs. standard Knowledge Tracing

The model we present in this paper focuses only on individualizing the prior knowledge parameter. We call it the Prior Per Student (PPS) model. The difference between PPS and Knowledge Tracing (KT) is the ability to represent a different prior knowledge parameter for each student. Knowledge Tracing is a special case of this prior per student model and can be derived by fixing all the priors of the PPS model to the same values or by specifying that there is only one shared student ID. This equivalence was confirmed empirically.

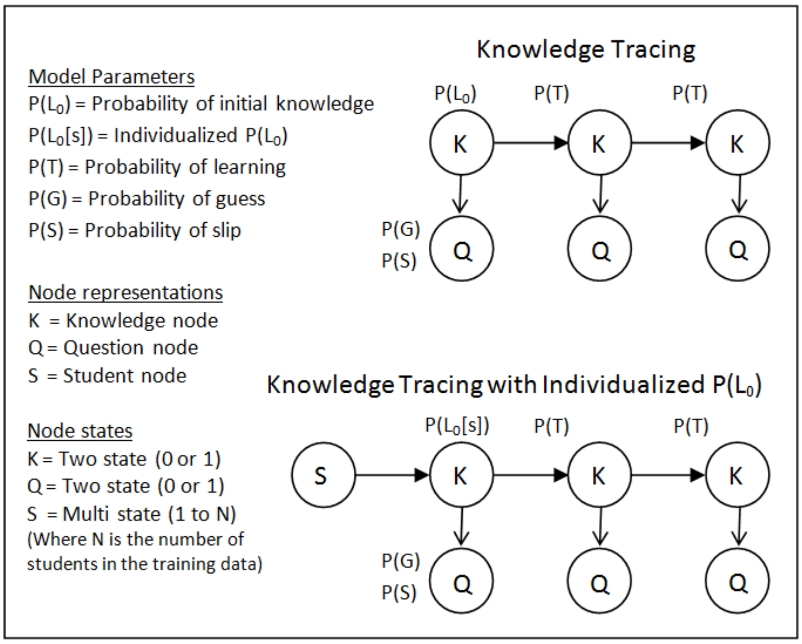

Fig. 1. The topology and parameter description of Knowledge Tracing and PPS

The two model designs are shown in Figure 1. Initial knowledge and prior knowledge are synonymous. The individualization of the prior is achieved by adding a student node. The student node can take on values that range from one to the number of students being considered. The conditional probability table of the initial knowledge node is therefore conditioned upon the student node value. The student node itself also has a conditional probability table associated with it which determines the probability that a student will be of a particular ID. The parameters for this node are fixed to be 1/N where N is the number of students. The parameter values set for this node are not relevant since the student node is an observed node that corresponds to the student ID and need never be inferred.

This model can be easily changed to individualize learning rates instead of prior knowledge by connecting the student node to the subsequent knowledge nodes thus training an individualized P(T) conditioned upon student as shown in Figure 2.

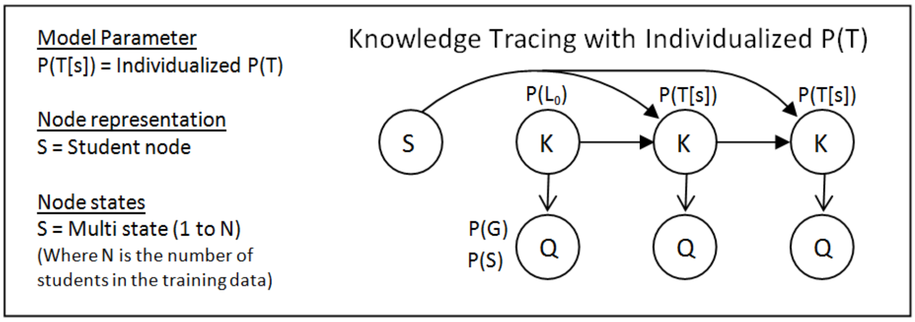

Fig. 2. Graphical depiction of our individualization modeling technique applied to the probability of learning parameter. This model is not evaluated in this paper but is presented to demonstrate the simplicity in adapting our model to other parameters.

2.2 Parameter Learning and Inference

There are two distinct steps in knowledge tracing models. The first step is learning the parameters of the model from all student data. The second step is tracing an individual student's knowledge given their respective data. All knowledge tracing models allow for initial knowledge to be inferred per student in the second step. The original KT work [2] that individualized parameters added an additional step in between 1 and 2 to fit individual weights to the general parameters learned in step one. The PPS model allows for the individualized parameters to be learned along with the non-individualized parameters of the model in a single step. Assuming there is variance worth modeling in the individualization parameter, we believe that a single step procedure allows for more accurate parameters to be learned since a global best fit to the data can now be searched for instead of a best fit of the individual parameters after the skill specific parameters are already learned.

In our model each student has a student ID represented in the student node. This number is presented during step one to associate a student with his or her prior parameter. In step two, the individual student knowledge tracing, this number is again presented along with the student's respective data in order to again associate that student with the individualized parameters learned for that student in the first step.

3 External Validity: Student Performance Prediction

In order to test the real world utility of the prior per student model, we used the last question of each of our problem sets as the test question. For each problem set we trained two separate models: the prior per student model and the standard knowledge tracing model. Both models then made predictions of each student's last question responses which could then be compared to the students' actual responses.

3.1 Dataset description

Our dataset consisted of student responses to problem sets that satisfied the following constraints:

- Items in the problem set must have been given in a random order
- A student must have answered all items in the problem set in one day
- The problem set must have data from at least 100 students
- There are at least four items in the problem set of the exact same skill
- Data is from Fall of 2008 to Spring of 2010

Forty-two problem sets matched these constraints. Only the items within the problem set with the exact same skill tagging were used. 70% of the items in the 42 problem sets were multiple choice, 30% were fill in the blank (numeric). The size of our resulting problem sets ranged from 4 items to 13. There were 4,354 unique students in total with each problem set having an average of 312 students ($\sigma = 201$) and each student completing an average of three problem sets ($\sigma = 3.1$).

Table 1. Sample of the data from a five item problem set

Student ID	1st response	2nd response	3rd response	4th response	5th response
750	0	1	1	1	1
751	0	1	1	1	0
752	1	1	0	1	0

In Table 1, each response represents either a correct or incorrect answer to the original question of the item. Scaffold responses are ignored in our analysis and requests for help are marked as incorrect responses by the system.

3.2 Prediction procedure

Each problem set was evaluated individually by first constructing the appropriate sized Bayesian network for that problem set. In the case of the individualized model, the size of the constructed student node corresponded to the number of students with data for that problem set. All the data for that problem set, except for responses to the last question, was organized into an array to be used to train the parameters of the network using the Expectation Maximization (EM) algorithm. The initial values for the learn rate, guess and slip parameters were set to different values between 0.05 and 0.90 chosen at random. After EM had learned parameters for the network, student performance was predicted. The prediction was done one student at a time by entering ,as evidence to the network, the responses of the particular student except for the response to the last question. A static unrolled dynamic Bayesian network was used. This enabled individual inferences of knowledge and performance to be made about the student at each question including the last question. The probability of the student answering the last question correctly was computed and saved to later be compared to the actual response.

3.3 Approaches to setting the individualized initial knowledge values

In the prediction procedure, due to the number of parameters in the model, care had to be given to how the individualized priors would be set before the parameters of the network were learned with EM. There were two decisions we focused on: a) what initial values should the individualized priors be set to and b) whether or not those values should be fixed or adjustable during the EM parameter learning process. Since it was impossible to know the ground truth prior knowledge for each student for each problem set, we generated three heuristic strategies for setting these values, each of which will be evaluated in the results section.

3.3.1 Setting initial individualized knowledge to random values

One strategy was to treat the individualized priors exactly like the learn, guess and slip parameters by setting them to random values to then be adjusted by EM during the parameter learning process. This strategy effectively learns a prior per student per skill. This is perhaps the most naïve strategy that assumes there is no means of estimating a prior from other sources of information and no better heuristic for setting prior values. To further clarify, if there are 600 students there will be 600 random values between 0 and 1 set for for each skill. EM will then have 600 parameters to learn in addition to the learn, guess and slip parameters of each skill. For the non-individualized model, the singular prior was set to a random value and was allowed to be adjusted by EM.

3.3.2 Setting initial individualized knowledge based on 1st response heuristic

This strategy was based on the idea that a student's prior is largely a reflection of their performance on the first question with guess and slip probabilities taken into account. If a student answered the first question correctly, their prior was set to one minus an *ad-hoc* guess value. If they answered the first question incorrectly, their prior was set to an *ad-hoc* slip value. *Ad-hoc* guess and slip values are used because ground truth guess and slip values cannot be known and because these values must be used before parameters are learned. The accuracy of these values could largely impact the effectiveness of this strategy. An *ad-hoc* guess value of 0.15 and slip value of 0.10 were used for this heuristic. Note that these guess and slip values are not learned by EM and are separate from the performance parameters. The non-individualized prior was set to the mean of the first responses and was allowed to be adjusted while the individualized priors were fixed. This strategy will be referred to as the "cold start heuristic" due to its bootstrapping approach.

3.3.3 Setting initial individualized knowledge based on global percent correct

This last strategy was based on the assumption that there is a correlation between student performance on one problem set to the next, or from one skill to the next. This is also the closest strategy to a model that assumes there is a single prior per student that is the same across all skills. For each student, a percent correct was computed, averaged over each problem set they completed. This was calculated using data from all of the problem sets they completed except the problem set being predicted. If a student had only completed the problem set being predicted then her prior was set to the average of the other student priors. The single KT prior was also set to the average of the individualized priors for this strategy. The individualized priors were fixed while the non-individualized prior was adjustable.

3.4 Performance prediction results

The prediction performance of the models was calculated in terms of mean absolute error (MAE). The mean absolute error for a problem set was calculated by taking the mean of the absolute difference between the predicted probability of correct on the last question and the actual response for each student. This was calculated for each model's prediction of correct on the last question. The model with the lowest mean absolute error for a problem set was deemed to be the more accurate predictor of that problem set. Correlation was also calculated between actual and predicted responses.

Table 2. Prediction accuracy and correlation of each model and initial prior strategy

	Most accurate predictor (of 42)		**Avg. Correlation**	
$P(L_0)$ Strategy	**PPS**	**KT**	**PPS**	**KT**
Percent correct heuristic	33	8	0.3515	0.1933
Cold start heuristic	30	12	0.3014	0.1726
Random parameter values	26	16	0.2518	0.1726

Table 2 shows the number of problem sets that PPS predicted more accurately than KT and vice versa in terms of MAE for each prior strategy. This metric was used instead of average MAE to avoid taking an average of averages. With the percent correct heuristic, the PPS model was able to better predict student data in 33 of the 42 problem sets. The binomial with $p = 0.50$ tells us that the probability of 33 success or more in 42 trials is $<< 0.05$ (cutoff is 27 to achieve statistical significance), indicating a result that was not the product of random chance. In one problem set the MAE of PPS and KT were equal resulting in a total other than 42 (33 + 8 = 41). The cold start heuristic, which used the 1^{st} response from the problem set and two *ad-hoc* parameter values, also performed well; better predicting 30 of the 42 problem sets which was also statistically significantly reliable. We recalculated MAE for PPS and KT for the percent correct heuristic this time taking the mean absolute difference between the rounded probability of correct on the last question and actual response for each student. The result was that PPS predicted better than KT in 28 out of the 42 problem sets and tied KT in MAE in 10 of the problem sets leaving KT with 4 problem sets predicted more accurately than PPS with the recalculated MAE. This demonstrates a meaningful difference between PPS and KT in predicting actual student responses.

The correlation between the predicted probability of last response and actual last response using the percent correct strategy was also evaluated for each problem set. The PPS model had a higher correlation coefficient than the KT model in 32 out of 39 problem sets. A correlation coefficient was not able to be calculated for the KT model in three of the problem sets due to a lack of variation in prediction across students.

This occurred in one problem set for the PPS model. The average correlation coefficient across all problem sets was 0.1933 for KT and 0.3515 for PPS using the percent correct heuristic. The MAE and correlation of the random parameter strategy using PPS was better than KT. This was surprising since the PPS random parameter strategy represents a prior per student per skill which could be considered an over parameterization of the model. This is evidence to us that the PPS model may outperform KT in prediction under a wide variety of conditions.

3.4.1 Response sequence analysis of results

We wanted to further inspect our models to see under what circumstances they correctly and incorrectly predicted the data. To do this we looked at response sequences and counted how many times their prediction of the last question was right or wrong (rounding predicted probability of correct). For example: student response sequence [0 1 1 1] means that the student answered incorrectly on the first question but then answered correctly on the following three. The PPS (using percent correct heuristic) and KT models were given the first three responses in addition to the parameters of the model to predict the fourth. If PPS predicted 0.68 and KT predicted 0.72 probability of correct for the last question, they would both be counted as predicting that instance correctly. We conducted this analysis on the 11 problem sets of length four. There were 4,448 total student response sequence instances among the 11 problem sets. Tables 3 and 4 show the top sequences in terms of number of instances where both models predicted the last question correctly (Table 3) and incorrectly (Table 4). Tables 5-6 show the top instances of sequences where one model predicted the last question correctly but the other did not.

Table 3. Predicted correctly by both

# of Instances	Response sequence
1167	1 1 1 1
340	0 1 1 1
253	1 0 1 1
252	1 1 0 1

Table 4. Predicted incorrectly by both

# of Instances	Response sequence
251	1 1 1 0
154	0 1 1 0
135	1 1 0 0
106	1 0 1 0

Table 5. Predicted correctly by PPS only

# of Instances	Response sequence
175	0 0 0 0
84	0 1 0 0
72	0 0 1 0
61	1 0 0 0

Table 6. Predicted correctly by KT only

# of Instances	Response sequence
75	0 0 0 1
54	1 0 0 1
51	0 0 1 1
47	0 1 0 1

Table 3 shows the sequences most frequently predicted correctly by both models. These happen to also be among the top 5 occurring sequences overall. The top occurring sequence [1 1 1 1] accounts for more than 1/3 of the instances. Table 4 shows that the sequence where students answer all questions correctly except the last question is most often predicted incorrectly by both models. Table 5 shows that PPS is able to predict the sequence where no problems are answered correctly. In no instances does KT predict sequences [0 1 1 0] or [1 1 1 0] correctly. This sequence analysis may not generalize to other datasets but it provides a means to identify areas the model can improve in and where it is most strong. Figure 3 shows a graphical representation of the distribution of sequences predicted by KT and PPS versus the actual distribution of sequences. This distribution combines the predicted sequences from all 11 of the four item problem sets. The response sequences are sorted by frequency of actual response sequences from left to right in descending order.

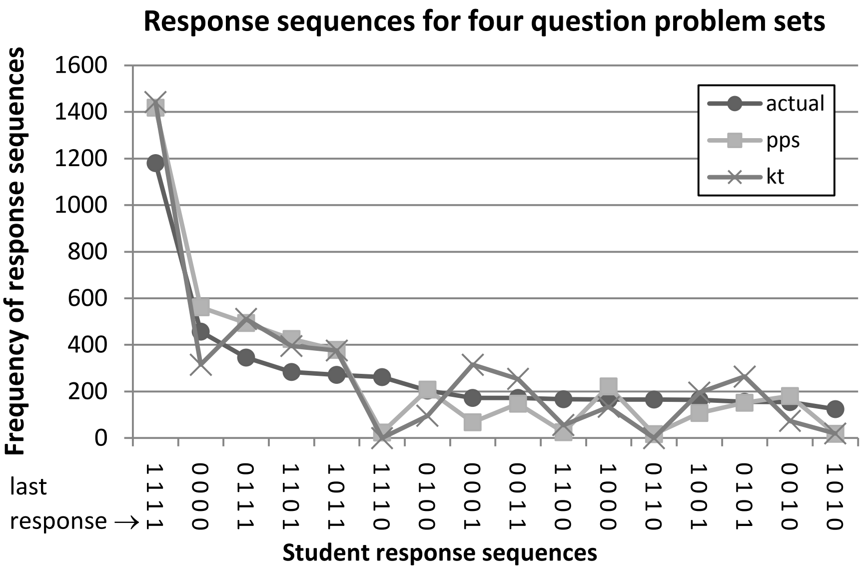

Fig. 3. Actual and predicted sequence distributions of PPS (percent correct heuristic) and KT

The average residual of PPS is smaller than KT but as the chart shows, it is not by much. This suggests that while PPS has been shown to provide reliably better predictions, the increase in performance prediction accuracy may not be substantial.

4 Contribution

In this work we have shown how any Bayesian knowledge tracing model can easily be extended to support individualization of any or all of the four KT parameters using the simple technique of creating a student node and connecting it to the parameter node or nodes to be individualized. The model we have presented allows for individualized and skill specific parameters of the model to be learned simultaneously in a single step thus enabling global best fit parameters to potentially be learned, a potential that is prohibitive with multi step parameter learning methods [2,4].

We have also shown the utility of using this technique to individualize the prior parameter by demonstrating reliable improvement over standard knowledge tracing in predicting real world student responses. The superior performance of the model that uses PPS based on the student's percent correct across all skills makes a significant scientific suggestion that it may be more important to model a single prior per student across skills rather than a single prior per skill across students, as is the norm.

5 Discussion and Future Work

We hope this paper is the beginning of a resurgence in attempting to better individualize and thereby personalize students' learning experiences in intelligent tutoring systems.

We would like to know when using a prior per student is not beneficial. Certainly if in reality all students had the same prior per skill then there would be no utility in modeling an individualized prior. On the other hand, if student priors for a skill are highly varied, which appears to be the case, then individualized priors will lead to a better fitting model by allowing the variation in that parameter to be captured.

Is an individual parameter per student necessary or can the same or better performance be achieved by grouping individual parameters into clusters? The relatively high performance of our cold start heuristic model suggests that much can be gained by grouping students into one of two priors based on their first response to a given skill. While this heuristic worked, we suspect there are superior representations and ones that allow for the value of the cluster prior to be learned rather than set *ad-hoc* as we did. Ritter et al [8] recently showed that clustering of similar skills can drastically reduce the number of parameters that need to be learned when fitting hundreds of skills while still maintaining a high degree of fit to the data. Perhaps a similar approach can be employed to find clusters of students and learning their parameters instead of learning individualized parameters for every student.

Our work here has focused on just one of the four parameters in knowledge tracing. We are particularly excited to see if by explicitly modeling the fact that students have different rates of learning we can achieve higher levels of prediction accuracy. The questions and tutorial feedback a student receives could be adapted to his or learning rate. Student learning rates could also be reported to teachers allowing them to more precisely or more quickly understand their classes of students. Guess and slip individualization is also possible and a direct comparison to Baker's contextual guess and slip method would be an informative piece of future work.

We have shown that choosing a prior per student representation over the prior per skill representation of knowledge tracing is beneficial in fitting our dataset; however, a superior model is likely one that combines the attributes of the student with the attributes of a skill. How to design this model that properly treats the interaction of these two pieces of information is an open research question for the field. We believe that in order to extend the benefit of individualization to new users of a system, multiple problem sets must be linked in a single Bayesian network that uses evidence from the multiple problem sets to help trace individual student knowledge and more fully reap the benefits suggested by the percent correct heuristic.

This work has concentrated on knowledge tracing, however, we recognize there are alternatives. Draney, Wilson and Pirolli [9] have introduced a model they argue is more parsimonious than knowledge tracing due to having fewer parameters.

Additionally, Pavlik et al [10] have reported using different algorithms, as well as brute force, for fitting the parameters of their models. We also point out that more standard models that do not track knowledge such as item response theory that have had large uses in and outside of the ITS field for estimating individual student and question parameters. We know there is value in these other approaches and strive as a field to learn how best to exploit information about students, questions and skills towards the goal of a truly effective, adaptive and intelligent tutoring system.

Acknowledgements

We would like to thank all of the people associated with creating the ASSISTment system listed at www.ASSISTment.org. We would also like to acknowledge funding from the US Department of Education, the National Science Foundation, the Office of Naval Research and the Spencer Foundation. All of the opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders.

References

1. Atkinson, R. C., Paulson, J. A. An approach to the psychology of instruction. *Psychological Bulletin*, 1972, 78, 49-61.
2. Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 253–278.
3. 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.
4. Baker, R.S.J.d., Corbett, A.T., Aleven, V.: More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. In: Wolf, B., Aimeur, E., Nkambou, R., Lajoie, S. (Eds.) Intelligent Tutoring Systems. LNCS, vol. 5091/2008, pp. 406-415. Springer Berlin (2008)
5. Beck, J.E., Chang, K.M.: Identifiability: A Fundamental Problem of Student Modeling. In: Conati, C., McCoy, K., Paliouras, G. (Eds.) User Modeling 2007. LNCS, vol. 4511/2009, pp. 137-146. Springer Berlin (2007)
6. Reye, J. (2004). Student modelling based on belief networks. *International Journal of Artificial Intelligence in Education*: Vol. 14, 63-96.
7. Chang, K.M., Beck, J.E., Mostow, J., & Corbett, A.: A Bayes Net Toolkit for Student Modeling in Intelligent Tutoring Systems. In: Ikeda, M., Ashley, K., Chan, T.W. (Eds.) Intelligent Tutoring Systems. LNCS, vol. 4053/2006, pp. 104-113. Springer Berlin (2006)
8. Ritter, S., Harris, T., Nixon, T., Dickison, D., Murray, C., Towle, B.(2009). Reducing the knowledge tracing space. In *Proceedings of the 2nd International Conference on Educational Data Mining*. pp. 151-160. Cordoba, Spain.
9. Draney, K. L., Pirolli, P., & Wilson, M. (1995). A measurement model for a complex cognitive skill. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 103–125). Hillsdale, NJ: Erlbaum.
10. Pavlik, P.I., Cen, H., Koedinger, K.R. (2009). Performance Factors Analysis - A New Alternative to Knowledge Tracing. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education*. Brighton, UK, 531-538.