The Prediction of Student First Response Using Prerequisite Skills

Anthony Botelho
Worcester Polytechnic Institute
abotelho@wpi.edu

Hao Wan
Worcester Polytechnic Institute
hale@wpi.edu

Neil Heffernan
Worcester Polytechnic Institute
nth@wpi.edu

ABSTRACT
A large amount of research in the field of educational data analytics has focused primarily on student next problem correctness. Although the prediction of such information is useful in assessing current student performance, it is better for teachers and instructors to place attention on student knowledge over a longer period of time. Several researchers have articulated that it is important to predict aspects that are more meaningful, inspiring our work here to utilize the large amounts of student data available to derive more substantial predictions over student knowledge. Our goal in this paper is to utilize prerequisite information to better predict student knowledge quantitatively as a subsequent skill is begun. Learning systems like ASSISTments and Khan Academy already record such prerequisite information, and can therefore be used to construct a method of prediction as described in this paper. Using these inter-skill relationships, our method estimates students' initial knowledge based on performance on each prerequisite skill. We compare our method with the standard Knowledge Tracing (KT) model and majority class in terms of the predictive accuracy of students' first responses on subsequent skills. Our results support our method as a viable means of representing student prerequisite knowledge in a subsequent skill, leading to results that outperform the majority class and that are comparably superior to KT by providing more definitive student knowledge estimates without sacrificing predictive accuracy.

Author Keywords
predicting student knowledge; mastery speed; initial knowledge; prerequisite; subsequent skills; first response prediction; Knowledge Tracing

INTRODUCTION
A large amount of research in the field of educational data analytics has focused primarily on student next problem correctness. Events such as the Knowledge Discover and Data Mining Competition held in 2010 ($www.kdd.org$), more commonly referred to as the KDD Cup, directs the attention of the field to the prediction of next problem correctness; while perhaps useful in performance evaluation, the ability to predict next problem correctness has certain limitations in regards to utility especially when assessing student knowledge over larger periods of time. Others in the field have begun raising other meaningful questions[3] [8][10], realizing the importance of predicting or observing aspects that are much more substantial. Intelligent tutoring systems (ITS) provide a wealth of student data from which more meaningful predictions and observations can be derived. Our work here aims to utilize this data to provide more significant information pertaining to student knowledge to teachers and instructors.

For our research, we emphasize constructing a more precise prediction on students' initial knowledge approaching a new skill. In the general case, students move gradually from an initial state of knowledge toward mastery, and student models should capture this change. Thus, a more accurate estimation of this initial knowledge could lead to a better understanding of a student's knowledge state at any observable time, and consequently, we could use the model's results to develop more precise predictions of future performance.

In this paper, we utilize prerequisite information to predict student initial knowledge on subsequent skills. If a skill 'A', is a prerequisite of skill 'B,' students should have mastered 'A' before proceeding to 'B.' The prerequisite relationships used in this work are defined by domain experts. Due to human effect, some skill relationships might be overestimated, or they may not exist in other applications. As such, we are seeking to answer the following two questions in this paper:

1. Does prerequisite information really help to improve the estimation of initial knowledge on subsequent skills?

2. Are all prerequisite relationships reliable?

We address these questions through three experiments to first observe trends of distribution across our proposed binning method, and then to compare the predictive accuracy of that method to that of KT and majority class across all skills and at an individual skill level.

The next section will introduce a background of our comparative model, KT, after which we will described the dataset used in our trials. The following section will discuss our proposed binning methodology before illustrating the results of

L@S 2015, March 14–18, 2015, Vancouver, BC, Canada.

http://dx.doi.org/10.1145/2724660.2724675

our experiments, and, finally, we state our contributions, conclusions and intended future work in this field of research.

BACKGROUND

The knowledge tracing (KT) model [4] developed by Corbett and Anderson has long been successful in the field of student assessment. Its implementation and use in tutoring systems and use in performance analysis systems exemplifies its practical applications, scalability, and appropriation across many fields of study. The KT model is widely used in these tutoring systems and the field of educational data analytics due to its accuracy in predicting student correctness by utilizing only a small amount of data.

The KT model gains its accuracy through the training of four parameters representing students' prior knowledge, learning rate, probability of guessing, or answering correctly while not knowing a skill, and chance of slipping, or answering incorrectly while in a supposed "learned" state. Knowledge tracing relies heavily on the successful training of these parameters to properly model a student beginning a new skill, and then to build upon that model at a student level given a sequence of responses. For this reason, each student beginning a particular skill receives the same base model. Therefore, each student within a skill will be given the same prediction for the first response. The model could be greatly improved if another prediction procedure, such as the method described in this work, could use a more intelligent approach to predict first response.

In the standard KT model, initial knowledge is represented by a parameter P(L_0), the probability of mastering the skill [4]. As such, KT is often used to estimate each student's initial knowledge [8]. In the standard KT model, the parameter P(L_0) is trained on all students' records in the a training set, and assumes that all students have the same initial state of knowledge. However, this assumption is too strong to use the model to predict each individual student's first response. To overcome this drawback, Pardos and Heffernan use three heuristic functions to model individualization in KT [9], and find that the method, setting initial individualized knowledge based on individual students' performance over all skills, yields superior results. This approach, however, overestimates the relationships between skills. If learning a skill does not promote, or even hinder [5], learning another skill, then it is not appropriate to use knowledge in one skill to estimate another.

Baker et al. uses another method [6] that compares a student's overall performance and all other students' performance on a skill to build an individualized model. Like the standard KT model, this method suffers two major problems: falling into local maxima and the existence of multiple global maxima[2]. Thus, we cannot know if the value of P(L_0) obtained by the model represents true student initial knowledge.

Knowledge Tracing's many strengths have made it a kind of comparative model in many works and is used again here as such. Knowledge Tracing builds upon the performance history of each student to calculate a probability that the student will answer the next problem correct. For this reason, it often fails to accurately predict students' first responses as there is less information for KT to accurately calculate a prediction. The method of prediction proposed in this work focuses entirely on first responses of students undertaking a new skill by observing student performance in prerequisite skills. Using knowledge tracing as a comparative model, our method of prediction aims to outperform KT in terms of accuracy in regard to students' first responses while providing a more definitive measure of initial knowledge.

DATASET

The dataset used in our work is comprised of real-world algebra and geometry-based student data from the 2009-2010 academic year taken from the ASSISTments tutoring system. This system is administered by teachers to students through assigned problem sets that track student performance in addition to many other features to be used for better assessing each student's knowledge and understanding of each topic, or skill. It is intended that each student completes problems pertaining to the assigned skill until a status of mastery is reached, which by default is defined as three consecutive correct answers. Each problem, or opportunity as it will be referred to in this work, is recorded by the system and is used to evaluate that student's overall performance.

Within ASSISTments, skills are arranged in an intended prerequisite-to-subsequent skill structure defined by domain experts as a recommended sequence of topics for instructors. It is the teacher's choice which skills and problems to assign as well as the order in which to assign them. As will be discussed later, the relationships between skills in this predefined prerequisite structure is worth further inspection, but are trusted for our initial experiments.

It was found that of the 230 skills listed as subsequent skills in our ASSISTments dataset, 28 contained usable prerequisite data; we define student data as usable if the sequence in which students complete skills matches the prerequisite structure defined within the system. The usable dataset consisted of 983 unique students across all skills, providing 3466 rows of response data. We acknowledge that our results may provide more reliable conclusions with a larger dataset, but our work here is intended to be used as initial work from which further research may expand upon and is therefore viewed as sufficient for this paper.

From the student performance, we also calculate each student's individual speed of mastery, defined as the number of opportunities, or individual problems, completed in order to gain mastery status as described above. We use this mastery speed as a measure of student knowledge and aptitude across skills and is used to calculate predicted responses as described in the next section. For this work, only problem correctness, expressed as binary values in the system, is used to calculate mastery speed and overall student performance, neglecting other features such as time between problems and skills and also partial credit evaluations. Other methods of determining mastery, discussed briefly in a later section, may lead to improved accuracy in our method, but are not the focus of this work; we use the simple "three right in a row" method of determining mastery for all of our experiments.

Student	Prerequisite	Mastery Speed	Skill	First Response Correctness
Tom	Adding	4	Division	Correct
Tom	Mult.	8	Division	
Bill	Adding	3	Division	Incorrect
Bill	Mult.	6	Division	
Joe	Adding	3	Division	Correct
Joe	Mult.n	3	Division	
Sue	Adding	5 LPC	Division	Incorrect
Sue	Mult.	DNF LPC	Division	

$\Downarrow$

Attempts	Prediction	Number of Students
3-4 incl.	1.0	1
4-8 excl.	0.5	2
8+	0.0	0
DNF High % Cor.	0.0	0
DNF Low % Cor.	0.0	1

Figure 1. The hypothetical students and data shown, fabricated to show our methodology, exemplifies the table creation process. Using a training set, a probability table is created for each skill by categorizing students with similar performance history

Methodology

The method described in this work attempts to better predict student first problem correctness on a subsequent skill by categorizing, or binning, students with similar mastery speed in a prerequisite skill; for purpose of clarification, the terms bin and category will be used interchangeably throughout this paper. Such a method has shown success in the past [12] when making other predictions such as next problem correctness using different features from a similar dataset. This method, labeled as "Prerequisite Binning" (PB), involves categorizing students based on a set of features, such as mastery speed, and inferring a relationship between them. For example, we binned students with similar ranges of mastery speed in order to create a prediction for any student that also could be placed in the same bin. If successfully identified, certain trends may appear within the bins, which are addressed in a later section.

The method of binning, as mentioned, groups students based on prerequisite mastery speed. For this, we used a 5 fold cross-validation on our dataset, using 80% as a training set to predict performance on the remaining 20%. The training set was used to construct the bins, which splits students based, again, on the number of opportunities needed to master each prerequisite skill. An average mastery speed across all prerequisite skills was calculated, placing students into one of five bins. The first bin contains those who averaged three to four opportunities inclusively $(3 \leq x \leq 4)$ to master all prerequisite skills; as three opportunities is the lowest possible mastery speed and four opportunities indicates an incorrect response on only the first problem, this bin presumably represents the highest knowledge students. The second bin, following the first in terms of mastery range, contains students who require, on average, between four and eight opportunities exclusively $(4 < x < 8)$. The third bin encompasses students with an average mastery speed of eight or more $(8 \leq x)$ across all prerequisite skills. Following this categorizing strategy, a fourth bin would contain those students that did not reach mastery status on prerequisite skills before proceeding to the subsequent skill. However, our dataset shows that a large percentage of students fall into this category, many of which respond to only a small number of problems; the reason for neglecting to finish a particular skill could be explained by boredom, simple negligence, poor time management, or a lack of knowledge. For these reasons, the "did not finish" (DNF) category, describing students that did not master all prerequisite skills, is represented by two bins. Our fourth bin contains students that did not master at least one of the prerequisite skills with a high percent correctness (HPC) across those skills (greater than or equal to 66.67% correctness), while the fifth and final bin contains such students with a low percent correctness (LPC) across all prerequisite skills (less than 66.67% correctness). The fourth and fifth bins handle the case where a student began a prerequisite skill, but did not reach mastery status; this means that at least one problem was attempted, but the student either completed less than three or failed to answer correctly on three consecutive opportunities. Bin four is therefore meant to represent students that failed to complete the prerequisite skills for reasons other than lack of knowledge, while the fifth contains students genuinely struggling and are perhaps experiencing wheel spinning[1].

With students from the training set categorized based on performance in prerequisite skills, a prediction value was calculated for each bin by finding the percentage of students in each category to respond correctly on the first opportunity of the subsequent skill. The reasoning for this method of binning, again, stems from the theory that particular trends exist for students in each bin and will extend to other students that also fall into that category. Therefore, it was expected that the prediction value of each bin constructed by the training set would apply to similar students in the test set.

Bin	Num. of Students	Num. of Students with First Response Correctness	Bin Prediction
1	29	24	0.828
2	53	26	0.491
3	3	0	0.000
4	2	1	0.500
5	3	0	0.000

Table 1. The bin student distribution and prediction values calculated for Fold 1 of Skill 47 of our dataset.

Figure 1 exemplifies the bin creation process using a set of hypothetical students (the names and values do not reflect any real person/dataset and are purely exemplary). In that example, prerequisite information from four students is used to construct the five bins. As Tom averaged a mastery speed of 6 opportunities across all his prerequisite skills, he is placed into the second bin with Bill, who averaged a mastery speed of 4.5 opportunities. Since Tom answered correctly on the first problem of the subsequent skill and Bill did not, the prediction for the second bin becomes 0.5, as half of the students in that bin answered the first problem of the subsequent skill correctly. Joe mastered each prerequisite skill with the minimum three attempts and is therefore placed into the first bin. That bin is given a probabilistic prediction of 1.0 due to the fact that all students in that bin answered correctly on the first question of the subsequent skill. Sue is placed into the fifth bin, as she did not master one of the prerequisite skills, and had a low percent correctness (less than 66.67%) across both prerequisites. She did not answer the first problem correct on the subsequent skill, leading to a prediction of 0.0, as no student in that bin answered correctly on the first question of the subsequent skill.

The values depicted in Table 1 were generated from our actual dataset. This table illustrates the prediction calculation methodology using skill 47 of our dataset corresponding in the ASSISTments tutoring system to the Conversion of Fraction Decimals Percents. As described in the earlier example, students in a training set are placed into each bin based on estimated student knowledge. Using this categorization, a prediction is calculated by observing the number of students in each bin to answer the first problem of the subsequent skill correctly.

RESULTS

The results of our work are exemplified through several metrics. Before comparing the predictive accuracy of our binning method to any other model, we must verify that each bin represents the intended level of knowledge within our dataset. Our method is able to illustrate this representation by observing the percentage of students within each bin to answer correctly on the first problem of a subsequent skill.

Bin	Number of Students	Percent Correct on First Response
1	806	61.79%
2	1170	60.00%
3	172	54.65%
4	732	52.59%
5	586	50.51%

Table 2. The overall percent correctness on the first response of all subsequent skills for each of the five bins.

Table 2 shows the distribution of knowledge within each bin across all skills in the observed dataset. The values show a distribution of higher knowledge students in the lower bins and lower knowledge students in the higher bins. This result supports the claim that our method is properly representing the intended level of knowledge.

The distribution of the number of students in each bin, particularly the fourth and fifth bin, indicates that our dataset contains a large percentage of students that did not complete prerequisites before attempting a subsequent skill. This was the reasoning behind splitting this ”DNF” bin into a high knowledge and low knowledge bin based on percent correct in the prerequisite skills. Further splitting these bins may lead to better predictive accuracy of our method, but is sufficient for our work in its current state and avoids over-complicating what is meant to be a simple categorization method.

Comparison of Overall Performance

The results of our method, entitled ”Prerequisite Binning” in Table 3, was compared to knowledge tracing as well as a majority class (MC) prediction to act as a control in our experiment. We chose knowledge tracing as it is widely used and studied in the field of educational data analytics and attempts to learn student initial knowledge for use in its calculation. Through this experiment we are first observing the effectiveness of our model by comparing it to the majority class, a prediction made for all students using the average correctness of the dataset, and then observing the differences in error between our method and KT; results illustrating a comparable error between the two methods supports the use of our binning method over KT, as it provides more definitive estimates of student knowledge without sacrificing predictive accuracy. Knowledge tracing was run using Kevin Murphy’s Bayes Net Toolbox for MATLAB [7] with initial parameters of 0.30, 0.14, 0.20, and 0.08 for prior, learn, guess, and slip respectively. For our experiment we ran a five fold cross validation on our dataset, using 80% of the data from each skill as a training set to predict the remaining 20%. The results in Table 3 represent the averages of all folds for each method.

Each of the three prediction methods are compared using RMSE and AUC two common measurements of error. A low RMSE indicates a more accurate prediction method while a larger AUC indicates higher accuracy. As observed in Table 3, the prerequisite binning method outperforms the majority class in both metrics indicating that it is a successful prediction method. When compared to knowledge tracing, however, the results show nearly the same RMSE value, but a superior AUC value.

While the binning method may not outperform knowledge tracing in all metrics, the predictive accuracy is comparable. The purpose of this work, again, is not to provide a method that outperforms KT, but rather to construct a modeling method that can provide teachers with more meaningful information regarding student knowledge. Unlike KT, where the learned parameters such as prior/initial knowledge are unusable metrics in describing true student knowledge due to the identifiability problem [2], our binning method provides an initial knowledge estimate based on previously observed performance; this initial knowledge estimate, represented as the probabilistic prediction calculated for each bin, is shown to be just as reliable as KT in predictive accuracy, while also providing a more definitive metric to describe a bin-wide initial knowledge that avoids problems of identifiability.

	RMSE	AUC
Majority Class	.496	.570
KT	.472	.626
Prerequisite Binning	.473	.651

Table 3. Results of our trials over all skills

Based on the results of our trial, we can conclude that prerequisite information can be used to predict student performance on subsequent skills in regards to first response. This supports our argument that knowledge and learning can be observed between prerequisite and subsequent skills.

Comparison Over Individual Skills

We also compare our method with KT on each individual skill. Figure 2 shows the difference of RMSE for these two models, that is: RMSE(KT) − RMSE(Bin); each positive difference value, therefore, indicates that our binning method outperforms KT in that skill, while negative difference values indicate KT outperforms binning in that particular skill. Each bar in the figure has an accompanying p-value above. This p-value is computed by applying a statistical T-test on the five-fold cross validation results. From this figure, we observe that our method outperforms KT in 14 of the 28 observed skills. Looking at the T-test results, there is a significant difference (p-value ≤ 0.05) between the two models on only 3 skills. This statistic further supports the comparability of the two models in terms of accuracy.

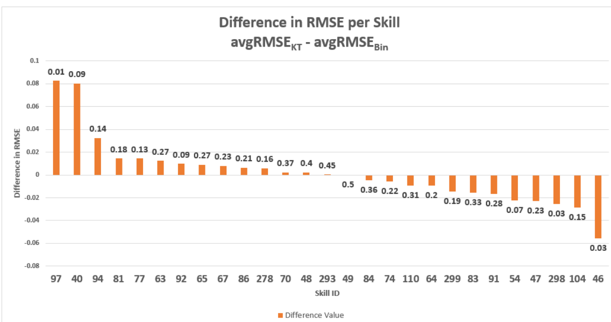

Figure 2. The difference of RMSE per skill when comparing our method of binning to standard knowledge tracing, ordered from highest to lowest difference. The number above each skill indicates the p-value of the difference.

A similar histogram illustrating the difference of RMSE for the majority class and our binning method, RMSE(MC) − RMSE(Bin), can be seen in Figure 3. The majority class represents a prediction for each student that is equal to the percent correctness of the training set of students. Again, as we used a five fold cross validation, 80% of the data from each skill is used as a training set to predict the remaining 20%. Comparing our binning method to the majority class should provide results that take into account the difficulty of each skill, defined by the average correctness calculated in majority class predictions.

This result attempts to answer the second question in introduction pertaining to the reliability of the prerequisite skill relationships. In accordance with our initial thoughts, the stronger the relationship between a prerequisite and subsequent skill, the better we can predict the performance of the subsequent skill from the knowledge of the prerequisite skill. Using Figure 3, we can observe significant differences (p-value ≤ 0.05) in terms of RMSE on a total of 5 individual skills. Therefore, at least on skills 97 and 49, the skills with better statistically significant results, we have strong confidence that the prerequisite relationships are reliable. For those skills with significantly lower results, skills 54, 298, and 46, the causal relation of the prerequisite skills may not be strong as expected by domain experts. All other skills, however, do not illustrate results significant enough to make a claim. These particular inconclusive results may be explained by inspecting our dataset. As indicated in our first observations pertaining to the distribution of students in each bin, a large percentage of students are categorized into bins four and five. Many of those students, as indicated by our dataset, attempt less than three problems, preventing mastery and also making it more difficult to properly estimate knowledge.

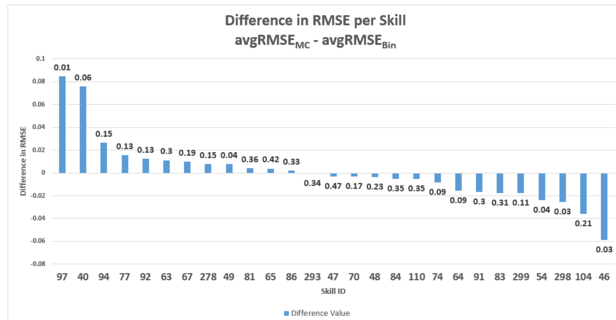

Figure 3. The difference of RMSE per skill when comparing our method of binning to majority class predictions, ordered from highest to lowest difference. The number above each skill indicates the p-value of the difference.

There may be two reasons for this occurrence. First, the prerequisite skills may too hard for the students to master. This may result from the teacher's decision not to assign particular prerequisite skills, or the skill relationship graph is incomplete. A second possibility may allude to a case where a teacher does not assign enough questions for students to master the prerequisite skills. As a teacher has control over the administering of skill problems, a number of such scenarios could lead to such results. In summary, these findings potentially indicate a need to further inspect the causal relationships defined by domain experts as they appear in the observed systems.

CONTRIBUTION

Our goal in this paper was to utilize the prerequisite information that many systems record to infer aspects of the students in our data. The current predominantly used knowledge tracing model employed in many learning systems assumes that all the skills are independent of each other. In this work, however, prerequisite information is used to better understand the relationship between the prerequisite and subsequent skills. The added consideration of this relationship in a model can be used to make better statements and inferences about not only the students, but also in the manner that such skills are presented to students.

We have shown here, to our knowledge, the first model that attempts to use the relationships between prerequisite skills to predict subsequent knowledge. This is on its way to make a larger contribution to better personalizing and individualizing student models by acknowledging and utilizing more of the data. We will make note that there are many other researchers that have used aggregate information, but have not paid attention to the prerequisite structure. Many psychometricians have found, for instance, that if students who do well on a topic A tend to do well on a topic B, that information can be used to better predict performance on topic B. In this context, however, we prefer to view such information differently. Our ultimate goal is to be able to make statements to teachers regarding information that is more causally related, and we do not want to influence predictions of future performance for unrelated tasks where there is little knowledge overlap. By imposing this constraint upon us, it will reduce our ability to make predictions, but will increase the significance of our statements to teachers.

The goal of this paper extends beyond the intent to develop a more accurate prediction methodology. We wish to look at the causal effects from which our results derive. It is more of a question of why using this data from prerequisite skills produces the accurate predictions across some skills and not in others. Our findings support the intuitive claim that certain skills are related, while others are not. Our trials provide a means of visualizing aspects of such skills to show that, as in Figure 3, prerequisite information does not have the same effect for all subsequent skills. Observing little difference in some skills between a method utilizing prerequisite information and a method, such as KT or majority class, that does not use such information may point to several issues in either our dataset or the prerequisite graph of the system. It is an interesting observation that some skills, while listed as a prerequisite, may not have as strong a relationship to a subsequent skill, which is vitally important information to teachers who need to consider a sequence to introduce new skills.

CONCLUSION AND FUTURE WORKS

The results and observations presented in this paper open new research opportunities in student assessment. Through our results we have observed several factors that help to better model student knowledge and aptitude across skills. The trials of this paper certainly raise some curiosities as to the extent subsequent skills are affected by prerequisite performance. In this paper, we focus exclusively on first responses of subsequent skills and, as the results were successful, we can now look beyond the first response to observe trends in prerequisite influence over an entire subsequent skill response sequence.

With these findings, our method can be adapted and/or appropriated to benefit other models like KT. Implementing our method into a modified KT model could lead to more accurate representations of student initial knowledge. As the method we propose here requires little in terms of processing time while providing more definitive student knowledge estimates than other models like KT, we aim to, through similar methods, represent other aspects of student learning such as aptitude and knowledge retention.

The accuracy of this method of binning is largely impacted by the reliability of the method of determining mastery. In our experiments, as it is in ASSISTments, mastery is defined simply as a student answering correctly on three consecutive opportunities; this method, while simple, may not be the best means of representing such a status universally for all students. Further work in exploring more precise methods of determining mastery speed may prove to benefit our method; such a method may include the individualization of mastery speed requirements for each bin, as it is likely that not all student levels of knowledge can be confidently labeled as having mastered a skill with the same number of sequential correct responses.

In this work, we concern ourselves with and direct our attention to the concept of student growth and knowledge over time. We believe that such information identifies aspects of the student more definitively than next problem correctness. In the future, we hope to continue similar work, looking into the influences that prerequisite skills exhibit in the other student models, like the wheel spinning model [1]. We would also like to make further observations and inferences on prerequisite skills, such as their impact on the student learning process itself, or the retention performance [11] of this prior.

ACKNOWLEDGMENT

We acknowledge funding from NSF (grant #'s 1316736, 1252297, 1109483, 1031398, 0742503), ONR's STEM Grand Challenges, and IES (grant #'s R305A120125, R305C100024).

REFERENCES

1. Beck, J., and Rodrigo, M. M. T. Understanding wheel spinning in the context of affective factors. In *Intelligent Tutoring Systems*, Springer International Publishing (2014), 162–167.
2. Beck, J. E., and min Chang, K. Identifiability: A fundamental problem of student modeling. In *User Modeling 2007*, Springer Berlin Heidelberg (2007), 137–146.
3. Beck, J. E., min Chang, K., Mostow, J., and Corbett, A. Does help help? introducing the bayesian evaluation and assessment methodology. In *Intelligent Tutoring Systems*, Springer Berlin Heidelberg (2008), 383–394.
4. Corbett, A., and Anderson, J. Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction 4*, 4 (1994), 253–278.
5. Cree, V. E., and Macaulay, C., Eds. *Transfer of Learning in Professional & Vocational Education*.
6. d. Baker, R. S. J., Corbett, A. T., Gowda, S. M., Wagner, A. Z., MacLaren, B. A., Kauffman, L. R., Mitchell, A. P., and Giguere, S. Contextual slip and prediction of student performance after use of an intelligent tutor. In *User Modeling, Adaptation, and Personalization*, Springer Berlin Heidelberg (2010), 52–63.
7. Murphy, K. Bayes net toolbox for matlab. `https://github.com/bayesnet/bnt/`.
8. Pardos, Z. A., and Heffernan, N. T. Navigating the parameter space of bayesian knowledge tracing models:

Visualizations of the convergence of the expectation maximization algorithm.

9. Pardos, Z. A., and Heffernan, N. T. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *User Modeling, Adaptation, and Personalization*, Springer Berlin Heidelberg (2010), 255–266.
10. Wang, Y., and Heffernan, N. Extending knowledge tracing to allow partial credit: Using continuous versus binary nodes. In *Artificial Intelligence in Education*, Springer Berlin Heidelberg (2013), 181–188.
11. Xiong, X., Beck, J., and Li, S. Class distinctions: Leveraging class-level features to predict student retention performance. In *Artificial Intelligence in Education*, H. Lane, K. Yacef, J. Mostow, and P. Pavlik, Eds., vol. 7926 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013, 820–823.
12. Xiong, X., Li, S., and Beck, J. E. Will you get it right next week: Predict delayed performance in enhanced its mastery cycle. In *FLAIRS Conference* (2013).