Leveraging First Response Time into the Knowledge Tracing Model

Yutao Wang
Worcester Polytechnic Institute
100 Institute RD.
Worcester, MA, USA
yutaowang@wpi.edu

Neil T. Heffernan
Worcester Polytechnic Institute
100 Institute RD.
Worcester, MA, USA
nth@wpi.edu

ABSTRACT

The field of educational data mining has been using the Knowledge Tracing model, which only look at the correctness of student first response, for tracking student knowledge. Recently, lots of other features are studied to extend the Knowledge Tracing model to better model student knowledge. The goal of this paper is to analyze whether or not the information of student first response time of a question can be leveraged into Knowledge Tracing model and improve Knowledge Tracing's prediction accuracy. In our experiments, we used discretized first response time data to predict students' correctness of the next question, and leveraged the result into a Knowledge Tracing model. Our analysis confirmed the value of student first response time in modeling student knowledge.

Keywords

Educational data mining, intelligent tutoring systems, student modeling, first response time.

1. INTRODUCTION

Modeling student behavior is crucial for education. For decades, researchers in the field of educational data mining (EDM) have been developing various methods of modeling student behavior using their performance as observations. One example is one of the dominant student model called Knowledge Tracing (KT) model built by Corbett and Anderson in 1995[1], which uses a dynamic Bayesian network to model student learning. Recently, lots of other features are studied in the framework of the Knowledge Tracing model to extend the Knowledge Tracing model to better model student knowledge. These features include the difficulty of problems [2], if it is a new day since a student last saw a problem [3], the assistance students require in answering a problem [4], etc. This paper analyses another piece of information: student first response time. We want to find out if students' first response time of a question can be used for improve KT's prediction accuracy.

Student response time, as an important feature that characterizes student behavior, is studied in the field of Intelligent Tutoring Systems in various models either due to its subjective importance or after some data analysis.

Some of these models use response time for understanding students' behaviors during problem solving in tutoring systems. Beck J.E. 2005 [5] used response times to model student disengagement; Shih B. et al. 2008 [6] built a response time model for bottom-out hints as worked examples; Arroyo I. et al. 2010 [7] used time required to solve a problem to model student effort.

Some models use different time information as one of many features in their models to indicate student knowledge. Such as Rai and Beck 2011[8] used the average time spent on each attempt in modeling their game-like math tutor.

Those works did not focus on using student first respond time as a direct indicator of student knowledge.

1.1 The Tutoring System

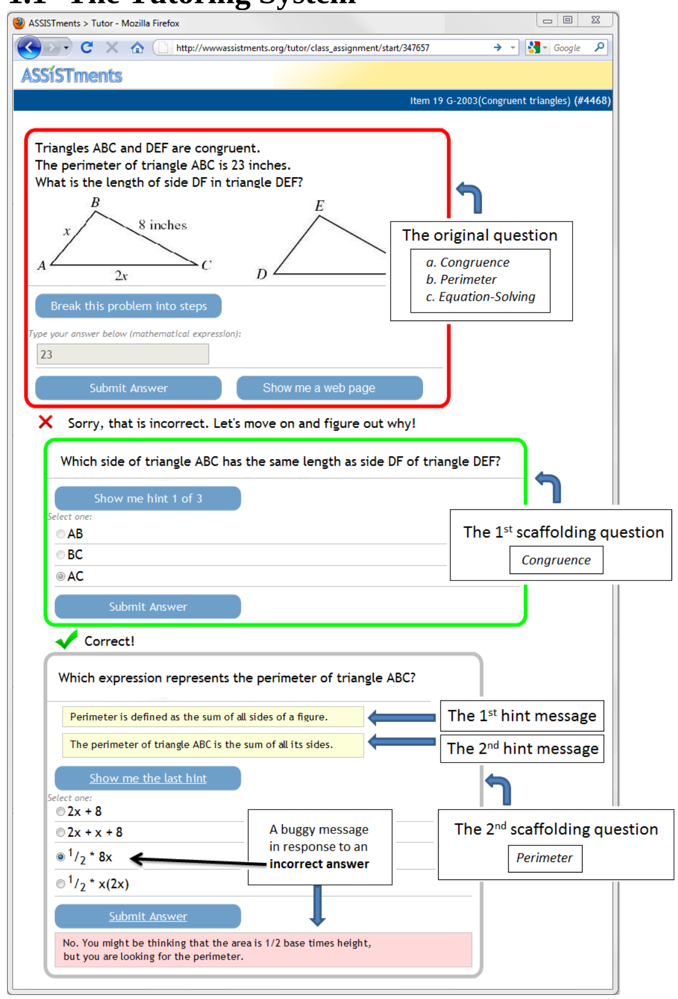

Figure 1 A typical senerio in ASSISTments system.

The data used in the analysis came from the ASSISTments system, a freely available web-based tutoring system for 4th through 10th grade mathematics (approximately 9 through 16 years of age). The system is mainly used in urban school districts of the Northeast United States. Students use it in lab classes that they attend periodically, or for doing homework at night.

The system provides tutorial assistance as buggy messages or scaffolding questions if a student makes a wrong attempt, and hint messages if a student asks for help. Figure 1 shows an example scenario in the ASSISTments system.

1.2 The KT Model

The Knowledge Tracing model shown in Figure 2 has been widely used in ITS and many variants have been developed to improve its performance (Baker et al. 2010, Pardos and Heffernan 2010). It uses 4 parameters for each skill, with two for student knowledge and the other two for student performance. The parameters prior knowledge and learning are called learning parameters. Prior knowledge is the likelihood the student knows the skill when he/she first uses the tutor. Learning is the probability a student will acquire a skill as a result of an opportunity to practice it. The parameters slip and guess are called the performance parameters in the model. An assumption of this model is that even if a student knows a skill, there is a chance he/she might still respond incorrectly to a question of that skill. This probability is the slip parameter. Conversely, a student who does not know the skill might be able to generate a correct response. This probability is referred to as the guess parameter.

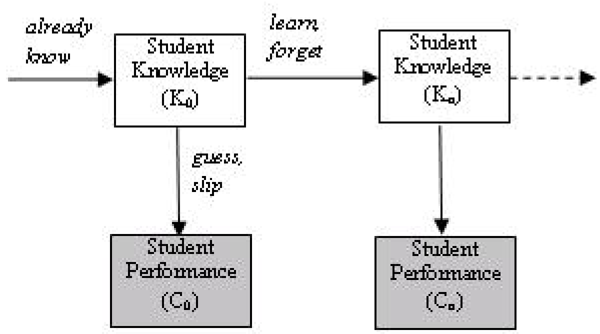

Figure 2. Knowledge Tracing model

Prior Knowledge = Pr (K0=True)
Guess = Pr (Cn=True | Kn=False)
Slip = Pr (Cn=False | Kn =True)
Learning rate = Pr (Kn =True | Kn−1=False)

In our experiment, we used the Bayes Net Toolbox for Matlab developed by Murphy (2001) to implement Knowledge Tracing, and the Expectation Maximization (EM) algorithm to fit the model to the dataset. The EM algorithm finds a set of parameters that maximize the likelihood of the data by iteratively running an expectation step to calculate expected likelihood given student performance data and a maximization step to compute the parameters that maximize that expected likelihood. There have been reported issues of local maxima when using the EM algorithm. Pardos and Heffernan (2001) concluded, based on a simulation study, that with the initial parameters of this algorithm in a reasonable range (the sum of initial guess and slip value is smaller than 0.5), the algorithm will always converge to a point near the true parameter value. In our experiments, we choose initial parameters for each skill as follows*: initial knowledge* = 0.5, *learning* = 0.1, *guess* = 0.1, *slip* = 0.1. These initial parameters are set to be similar with the results of previous experiments that estimated the Knowledge Tracing model parameters on some other datasets from the ASSISTments system.

2. PROBLEM AND APPROACH

Although there has been study done in both student response time and student knowledge, there is no research in using student response time to indicates student knowledge. In this paper, we focus on leveraging student first response time into the Knowledge Tracing model to see whether or not student first response time is valuable in modeling student knowledge and enhance KT model's prediction accuracy of student performance.

There are various explanations in different student first response time. For example, a short first response time could either mean the student is proficiency on the skill or the student is guessing the result or gaming the system; also, a long first response time could either mean the student is thinking about the given problem or he/she is just doing some off task behavior. As a result, the connection between student first response time and student knowledge could be blurred by many other factors. However, since student response time is one of the most important information of student behavior that could be easily gathered by Intelligent Tutoring Systems, analyses on its ability of modeling student knowledge and improving performance prediction is still meaningful to this field. To handle the other factors that could influence the result, we discretized the first response time data to eliminate unnecessary details of the information, and aim for finding the general indication of this information towards student knowledge and future performance.

2.1 Data

The data we analysed are from school year September 2010 to September 2011, which consisted of 15931 students who solved at least 20 problems within ASSISTments. We filtered out skills that have fewer than 50 students and randomly selected 2015 student users. As a result, we have 498 ,988 data records. Each data record is recorded right after a student answered a problem, and logged relevant information including the identity of the student, the problem identity and skills required to solve it, the correctness of the student's first response to this problem, the first response time the student spent on this problem, and the timestamp when the student start and finish solving this problem.

2.2 Discretization of First Response Time

As we discussed before, since student first response time includes information other than student knowledge. To eliminate unnecessary details of the information, which could be relevant to other factors, we discretized student first response time data into several bins.

Our goal is to find out if the main character of student first response time contains unique information about student knowledge in compare with other features. We discretized student first response time data into four categories. The way we define these categories are based on the follows assumptions.

The first assumption is, in general, students that need more time to first respond to a problem have lower knowledge than students that need less first response time, because the former require more time to answer the question.

The second assumption is, in general, the data records that show extremely little time of student first response time are likely to indicate some special behaviors such as gaming, thus, the first response time in those data records may not be as useful in indicating students knowledge.

The third assumption is, in general, the data records that show extremely long time of student first response time are also likely to indicate special behaviors such as off task behaviors, thus, the first response time in those data records may also not be useful in indicating student knowledge.

According to these assumptions, the four categories of student first response time are: extremely short, short, long, extremely long.

Also, considering student first response time highly varies by problem, we computed different cut points of these four categories for each problem.

In our experiments, for each problem, we put all of the corresponding first response time that are in the shortest 5% range for that problem into the first bin: the extremely short time bin; the student first response time within 5% to 50% range went into the second bin: the short time bin; the 50% to 95% range went into the third bin: the long time bin; and the top 5% went into the forth bin: the extremely long time bin. These four bins are denoted as bin1 to bin4 in our training dataset. This numbers 5%, 50% and 95% are selected based on experimenting with a few different sets of values. We did not try more sophisticated criteria, such as standard deviation, which might be able to further improve the result.

This method allows us to consider the main trend of the student response time per problem, without being affected by rare and extreme situations or data.

2.3 Predicting Student Performance

In this section, the purpose of our analysis is to find out if student first response time is valuable in modeling student knowledge and predicting student performance. We want to model only student first response time in this step, so that the result won't be affected by other additional features. Also, we want the model to be very simple so that it can be easily computed and leveraged into other existing student models that using other features for modeling student knowledge.

We choose to use a purely data driven tabling model that is similar to our previous work [4], which makes no assumptions about how the new information reflects student knowledge. To do so, we simply built a one by four parameter table, in which column index represents the category of student first response time in the previous question, and each cell contains the probability that the student will answer the current question correctly. For that value, we simply use the percentage of students who answered the current question correct when the previous question fell into the corresponding category.

Table 1 shows the parameter table we computed from the training data.

Bin 1	Bin 2	Bin 3	Bin 4
0.3829	0.7103	0.6428	0.5389

Table 1. Parameter table computed from the training dataset.

This model is very simple and easy to compute. But also, it is very limited. The only information it takes into account is the student first response time and the difficulty or the type of question. The information of the question is included in the model for when we discretized the first response time, we choose different bin cut points for different questions.

To evaluate how well this simple model fits the data compare to a baseline of always guessing the mean value of the data as a prediction. We used Root Mean Squared Error (RMSE) as a metric to examine the predictive performance on an unseen test set. The RMSE of the baseline prediction is 0.4589 and the RMSE of the student first response time model is 0.4552, which indicates this value is indeed contain some predictive power, although the benefit of this information is not obvious.

2.4 Leveraging First Response Time into KT

In this section, our goal is to find out whether or not leveraging the result of the simple model above into an existing student model which does not take into account student first response time information could help improve the existing student model, and thus result in better prediction accuracy. We choose the KT model in our experiments.

By combining the student first response time model with the KT model, we leverage new information into the KT model. To find out the result of this method, we used a linear regression model to combine the simple model we built with the traditional KT model by making the student performance as the dependent variable in the regression model, and the prediction results from the student first response model and the KT model as independent variables.

We again used the RMSE to examine the predictive performance of the KT model and the combination of these two models. The result is shown in Table 2. The FRT in Table 2 represents the first response time model, KT represents the Knowledge Tracing model, and the Comb represents the linear regression combination of these two models. This table also provides the comparison of the number of parameters of each model. Since the data set has 220 skills, KT generated in total 4*220 parameters.

	FRT	KT	Comb
RMSE	0.4552	0.4251	0.4213
#of params	4	880	886

Table 2. Comparision of the RMSE result of different models.

The linear regression formula for combining two models tells us the information about the weight of each model in regarding with their impact to the final model. The formula generated from our training process of the linear regression is:

-0.1227 + 0.1928 * FRT_prediction + 0.9821 * KT_prediction.

from which we can tell that the influence of the student first response time model to the final result is small. However, the RMSE shows an improvement from the KT model.

To find out if this improvement is statistically reliable, we did reliability analysis by computing the student level RMSE to account for the non-independence of each student and their actions and then compared the KT and the Comb model using a two tailed paired t-test. The p value is 0.0389, which indicates that although the improvement is small, it is reliable.

3. CONTRIBUTIONS

This paper makes two main contributions. First, we analysed the predicting power of student first response time on student performance. In compare with other work on the student response time, which focus on explaining student in task or off task behavior, this work shows that student first response time contains certain information about student knowledge.

The second contribution this paper makes is to show that by leveraging the student first response time information, we can improve the prediction accuracy of the traditional KT model. In compare with other more complicated and time consuming methods, this model is very flexible and easy to apply to any existing student modeling techniques to incorporate into them the new information of student first response time.

4. FUTURE WORK AND CONCLUSIONS

The model we proposed for using student first response time to improve KT model is a simple and fast way of utilizing additional information. However, experiments show using student first response time alone did not provide a good performance prediction. There are several questions that we are interested in exploring.

One question is if the prediction accuracy of using student first response time can be improved by taking into account student and skill information. Currently we use only four parameters for all of the data. This can be easily extended to deal with individualization and separate skill by computing parameter tables for each skill or each student separately.

Another question we want to explore is a way to combine the response time and other information that gathered when a student answers a question, such as the number of hints and attempts a student need to answer the question. We are interested in combine these features because they seem to be highly related. We built a tabling model using the assistance student needs for answering a question in 2010[4], and searching for a method to merge these two models together is a reasonable next step.

In conclusion, in this paper, we use a method that is easy to compute and apply to leverage discretized student first response time information into the KT model to improve the prediction accuracy of the KT model. The result shows a clear value of student first response time in indicating student knowledge.

ACKNOWLEDGMENTS

This research was supported by the National Science foundation via grant “Graduates in K-12 Education” (GK-12) Fellowship, award number DGE0742503 and Neil Hef-fernan's CAREER grant. We also acknowledge the many additional funders of AS-SISTments 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] Pardos, Z. A., Heffernan, N. T. (2011) KT-IDEM: Introducing Item Difficulty to the Knowledge Tracing Model. In Proceedings of the 19th International Conference on User Modeling, Adaptation and Personalization. Girona, Spain.

[3] Qiu, Y., Qi, Y., Lv, H., Pardos, Z.A, Herffernan, N. T.: Does Time Matter? Modeling the Effect of Time with Bayesian Knowledge Tracing, In Proceedings of the 4th International Conference on Educational Data Mining. 2011.

[4] Wang Y., Heffernan N. T., The “Assistance” Model: Leveraging How Many Hints and At-tempts a Student Needs, In Proceedings of the 24th International FLAIRS Conference. Palm Beach, Florida.

[5] Beck, J. E.: Using response times to model student disengagement, In Proceedings of AIED 2005.

[6] Shih, B., Koedinger K. R., Scheines, R.: A Response Time Model for Bottom-Out Hints as Worked Examples, In Proceedings of the 1st International Conference on Educational Data Mining, 2008, p. 117-126.

[7] Ivon Arroyo, Hasmik Meheranian, Beverly Park Woolf: Effort-based Tutoring: An Empirical Approach to Intelligent Tutoring. EDM 2010: 1-10

[8] Dovan Rai, Joseph Beck: Exploring User Data From a Game-like Math Tutor: A Case Study in Causal Modeling. EDM 2011: 307-313