

Do students really learn an equal amount independent of whether they get an item correct or wrong?

Seth Adjei¹, S. M. A. Salehizadeh², Yutao Wang³, Neil Heffernan⁴

Department of Computer Science

Worcester Polytechnic Institute

100 Institute Road, Worcester, MA

[saadjei¹, ssalehizadeh², yutaowang³, nth⁴]@wpi.edu

ABSTRACT

The field of educational data mining has been paying attention to Knowledge Tracing (KT) for a long time. Corbett and Anderson assumed the amount of learning that students do does not depend on whether students get items right or wrong. Ohlsson and others argued that the student should learn more from a previous incorrect performance. We decided to investigate a Bayes Network similar to KT but that allows us to have learning rates that are different according to whether students get items correct or not. While the idea of allowing learning rates from previous incorrect performances to be higher seems intuitive, our experiments showed that this way does not always lead to better predictions. Of course reasoning from a null result is dangerous, our contribution is that this intuitive idea is not one that other researchers should waste time in working on, unless they come up with a different model from the model we used (which is the naïve way of modifying KT).

Keywords:

Knowledge Tracing, Bayesian Networks, Learn from performance, Tutoring Strategies

1. INTRODUCTION

The field of Education Data mining has depended to a large extent on the model that was developed by Corbett and Anderson [2] and enhanced by a number of authors for predicting student performance. Over the years many new models have been built to improve upon the prediction accuracy of KT.

Wang and Heffernan have also shown that better predictions are achieved with the inclusion of additional parameters relating to the skills and groups to which a student belongs. [4]

Standard KT makes a number of assumptions, including the fact that the rate of learning is constant and that the transition from one knowledge state to the other is not dependent of previous performance. [2] Other researchers have introduced different models that seem to deal with this anomaly with the KT model [7], while some have compared these different models to determine which best predicts student

performances considering prior performance. [6] Ohlsson theorized that humans in general are able to learn from their previous error performance, especially in situations where there is an explanation for the cause of the error. [1] Ohlsson further reiterated that in order to avoid repeating an incorrect action the knowledge behind the action must be changed. It is therefore intuitive that the student will learn from the previous performance. In this paper we present yet another modification of the KT model which considers the previous performance of a student on a particular item.

2. LP Model

We considered a new assumption for the KT model as follows: “Students can learn from their previous observed performance once there is some tutoring associated with the wrong performance.” This resulted in a new model which we call “Learn from Performance” (LP) model and present in Fig.1b).

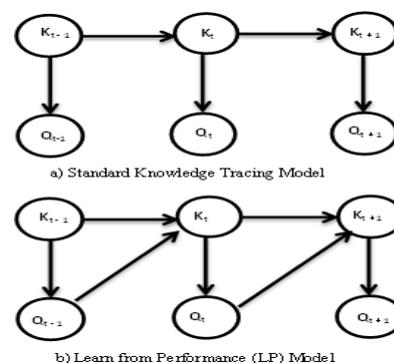


Figure.1 Student Performance Models (a) KT (b) LP

To account for learning after performance, our new model, LP, introduces one link from performance at time $t-1$ (Q_{t-1}) to knowledge at time t (K_t). This modification therefore introduces two (2) additional parameters to the model. These parameters are *learn_from_correct* (LC), and *learn_from_incorrect* ($LinC$). LC is the probability that the student gained some additional knowledge from having answered the most previous question in the skill correct. This is especially so in the case where the student sees a different variation of questions relating to a given skill. $LinC$ is the probability that the student has learnt something new from performing poorly from the most previous question. There is also the probability that the student had the knowledge prior to the previous performance but answered the previous question wrong and hence he lost it. In other words this is the probability of shallow learning or *confusion*.

3. METHODOLOGY AND RESULTS

To evaluate our model we used both real data from the ASSISTments Tutoring System [3] and simulated data. The Bayes Net Toolbox for Matlab [5] was used to implement the standard KT as well as the proposed LP model.

3.1 Simulated Data

In order to evaluate the LP model and to ensure that we were not over-fitting the model, we decided to test the model's strength using simulated data from both KT and LP. The networks' parameter values were set to ground truth values.

To evaluate the ability of the LP model to see if it is effective and can learn back its parameters, we generated data for 200 typical students from the LP Bayesian Network and used both models to predict the performance of students. In the simulated experiment we performed a 5-fold cross validation on both models and computed the performance metrics (MAE, RMSE and AUC).

The results indicated that LP learned most of its parameters: *prior*, *learn_from_incorrect* (*LinC*), *guess* and *slip* back to an appreciable degree. The *learn_from_correct* (*LC*) parameter, however, does not seem to get close to the ground truth values that were used in generating the simulation data for LP. Table 1 shows that LP outperformed KT reliably using LP generated data. However there were mixed results when KT generated data was used.

Table 1. Simulation Performance Comparison Results with LP Generated Data (1000 samples)

Fold	MAE		RMSE		AUC	
	KT	LP	KT	LP	KT	LP
1	0.326	0.270	0.399	0.380	0.707	0.832
2	0.335	0.271	0.409	0.382	0.779	0.820
3	0.336	0.284	0.409	0.391	0.798	0.816
4	0.351	0.285	0.426	0.391	0.798	0.827
5	0.334	0.275	0.405	0.381	0.762	0.834
Mean	0.336	0.277	0.410	0.385	0.794	0.826
P-value	<0.05		<0.05		<0.05	

3.2 Real Data Experiments

Given mixed results in simulation experiments, we further tested the model using real data to confirm our observation. The data set we used is from the ASSISTments Tutoring System. We randomly chose twenty (20) skills that have a minimum of 1000 rows of problem logs. Each row represented the student's performance of a given Skill Builder problem and the number of times the student has encountered problems of that nature including the current opportunity.

We employed the Expectation Maximization function that comes with the Bayes Net Toolbox within Matlab. We split the data randomly into five equal folds for each skill. We then performed a five-fold cross validation of the predictions for each skill and for each model.

Table 2 displays the MAE, RMSE and AUC values for each skill and for each model.

Table 2. Skill Prediction Performance Comparison Results

SKILL (#)Name	MAE		RMSE		AUC	
	KT	LP	KT	LP	KT	LP
(1) Box and Whisker	0.349	0.268	0.423	0.487	0.681	0.500
(9) Stem and Leaf Plot	0.394	0.321	0.447	0.490	0.599	0.583
(10) Table	0.294	0.187	0.386	0.424	0.539	0.453
⋮	⋮	⋮	⋮	⋮	⋮	⋮
(58) Addition Whole Numbers	0.195	0.116	0.313	0.330	0.557	0.500
Mean	0.350	0.287	0.415	0.462	0.611	0.569
p-values	<0.05		<0.05		>0.05	

The results in Table 2 indicated that LP has better MAE values than KT. However, KT reliably outperformed LP with RMSE and not reliably so with AUC. These results show that the results do not indicate any better performances than

4. CONCLUSION

Ohlsson theorized that students do learn from their previous error performance, especially in the case where explanation of the reasons for the error is provided. On the basis of this, we developed a new naïve Bayes Network model that allows the amount of learning to increase when users get an item wrong. Our experiments with real and simulated data showed that we do not get better prediction of student performance with this proposed LP model than the standard KT model. Hence we conclude from our experiments that the assumption that Corbett and Anderson made was justifiable even though not intuitive according to Ohlsson. Our contribution is that this intuitive idea is not one that other researchers should waste time in working on, unless they come up with a different model from the model we used.

The code and data for these experiments are available at: <http://users.wpi.edu/~saadjei/>

REFERENCES

- [1]. Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, 103(2), 241–262
- [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]. ASSISTments Tutoring System, www.assistments.org
- [4]. Wang, Y., Heffernan, N.T. (2012). "The Student Skill Model". In *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, pp. 399-404, 2012
- [5]. Murphy, K (2006). *Bayes Net Tool Kit for Matlab*
- [6]. Yue Gong, Joseph E. Beck, Neil T. Heffernan (2011): How to Construct More Accurate Student Models: Comparing and Optimizing Knowledge Tracing and Performance Factor Analysis. *I. J. Artificial Intelligence in Education* 21(1-2): 27-46
- [7]. Pavlik Philip, Cen Hao, Koedinger K. R. (2009) Performance Factors Analysis--A New Alternative to Knowledge Tracing In *Proceeding of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (2009), pp. 531-538