Modeling students' problem solving performance in the computer-based mathematics learning environment

Young-Jin Lee
University of Kansas, Lawrence, Kansas, USA

Received 15 May 2017
Accepted 24 August 2017

Abstract

Purpose – The purpose of this paper is to develop a quantitative model of problem solving performance of students in the computer-based mathematics learning environment.

Design/methodology/approach – Regularized logistic regression was used to create a quantitative model of problem solving performance of students that predicts whether students can solve a mathematics problem correctly based on how well they solved other problems in the past. The usefulness of the model was evaluated by comparing the predicted probability of correct problem solving to the actual problem solving performance on the data set that was not used in the model building process.

Findings – The regularized logistic regression model showed a better predictive power than the standard Bayesian Knowledge Tracing model, the most frequently used quantitative model of student learning in the Educational Data Mining research.

Originality/value – Providing instructional scaffolding is critical in order to facilitate student learning. However, most computer-based learning environments use heuristics or rely on the discretion of students when they determine whether instructional scaffolding needs be provided. The predictive model of problem solving performance of students can be used as a quantitative guideline that can help make a better decision on when to provide instructional supports and guidance in the computer-based learning environment, which can potentially maximize the learning outcome of students.

Keywords Problem solving, User modeling, Educational Data Mining (EDM), Log file analysis

Paper type Research paper

Introduction

Recently, computer-based learning environments such as Massive Open Online Course and Intelligent Tutoring System become more prevalent. One important characteristic of these computer-based learning environments is that they can capture in their log files what students are doing while trying to learn new knowledge without interrupting their learning processes. Since log files allow for re-constructing how students used computer-based learning contents, analyzing the log files can enable us to quantitatively study the learning behaviors of students and the usefulness of computer-based learning contents. As a result, Learning Analytics and Educational Data Mining (EDM) are emerging as new exciting research fields.

One active research topic in EDM is predicting whether students can correctly solve a problem because many computer-based learning environments, especially for mathematics and science, use problem solving as a primary means for assessing student learning. As Koedinger and Aleven (2007) pointed out, in order to maximize student learning outcomes, it is critical to balance giving and withholding instructional supports and guidance in the computer-based learning environment. Students may not exert enough cognitive efforts and fail to acquire a schema from learning tasks if they receive instructional supports prematurely (Kapur, 2008; Schmidt and Bjork, 1992). Academically weaker students, on the other hand, will fail to learn unless they are provided with appropriate instructional supports and guidance in time. This issue would become more important in the computer-based learning environment where students learn

The International Journal of Information and Learning Technology
Vol. 34 No. 5, 2017
pp. 385-395

2056-4880
DOI 10.1108/IJILT-05-2017-0031

primarily on their own and there is no teacher who can regulate the learning process of students. Most computer-based learning environments rely on either simple heuristics (e.g. giving hints or feedback after students fail to solve a problem a certain number of times) or discretion of students in determining when instructional supports and guidance are going to be provided. Obviously, simple heuristics would not be able to maximize the learning outcome of students because it does not take into account the difficulty of learning tasks and the ability of students. Similarly, providing instructional supports on the demand of students may not improve their learning outcome because novice students do not possess metacognitive abilities and prior knowledge required for determining the right moment to ask for help (Clark and Mayer, 2003; Lawless and Brown, 1997). To maximize the learning outcome of students, computer-based learning environments should be able to make a more intelligent decision based on the difficulty of learning tasks and the ability of students, which requires a quantitative model of student performance.

This study investigates whether a predictive model of problem solving performance of students can be built from the log files of a computer-based mathematics learning environment. In particular, this study uses regularized logistic regression in estimating how likely middle school students are to correctly solve a problem solving step of mathematics problems, given their performance on the problems they tried to solve in the past. This study compares the performance of the regularized logistic regression model to Bayesian Knowledge Tracing (BKT), the most frequently used quantitative model of student learning in EDM (Beck and Chang, 2007; Corbett and Anderson, 1995; Pardos and Heffernan, 2011; Pardos *et al.*, 2012, 2013).

Data

Original data

This study analyzed the log files obtained from the Pittsburgh Science of Learning Center (PSLC) (http://learnlab.org) whose DataShop Web service (https://pslcdatashop.web.cmu.edu) provides log files of various computer-based learning environments capturing the learning processes of students trying to learn different subject matters, from foreign language to mathematics and physics (Koedinger *et al.*, 2010). This study used the "Assistments Math 2004-2005" data set that captured how 895 middle school students used a computer-based mathematics learning environment called ASSISTments for over 3,400 student hours. The original data set obtained from PSLC includes 580,786 database transactions where each transaction record contains information about students and their problem solving activities such as anonymized student ID, problem name, step name, problem solving time, and whether or not students were able to solve each problem solving step correctly (see Figure 1 and Table I). KC in the transaction record can be considered a concept or principle required to resolve the corresponding problem solving step (e.g. Pythagorean theorem), and is used by PSLC researchers in categorizing problem solving steps. For further exploration of problems that students tried to solve in the computer-based learning environment, visit www.assistments.org

Pre-processed data

Using these data, we wanted to build a predictive model that mimics how expert human teachers would do when they estimate whether their students can solve a problem correctly. When a student is about to solve a problem on a certain mathematics concept (e.g. Pythagorean theorem), expert human teachers would examine how well this student did on the problems requiring an understanding of the same mathematics concept (e.g. Pythagorean theorem) in the past. In order to create predictor variables that can capture this behavior, the original PSLC data set had to be pre-processed because each transaction record in the PSLC data does not provide information on the past problem

	A	B	C	D	E	F	G	H	I
1	Anonymized student ID	Problem name	Step name	Problem time	Step time	Number of problem views	Attempt At Step	KC	Outcome
2	Stu_00411460f7c92d2124a	358	Step0:358:	11:36:48 AM	11:38:54 AM	1	1	Sum-of-Interior-Angles-Triangle	INCORRECT
3	Stu_00411460f7c92d2124a	358	Step1:359:	11:36:48 AM	11:39:17 AM	1	1	Supplementary-Angles	HINT
4	Stu_00411460f7c92d2124a	358	Step1:359:	11:36:48 AM	11:39:38 AM	1	2	Supplementary-Angles	HINT
5	Stu_00411460f7c92d2124a	358	Step1:359:	11:36:48 AM	11:40:38 AM	1	3	Supplementary-Angles	CORRECT
6	Stu_00411460f7c92d2124a	358	Step2:360:	11:36:48 AM	11:41:03 AM	1	1	Sum-of-Interior-Angles-Triangle	HINT
7	Stu_00411460f7c92d2124a	358	Step2:360:	11:36:48 AM	11:42:02 AM	1	2	Sum-of-Interior-Angles-Triangle	HINT
8	Stu_00411460f7c92d2124a	358	Step2:360:	11:36:48 AM	11:42:58 AM	1	3	Sum-of-Interior-Angles-Triangle	CORRECT
9	Stu_00411460f7c92d2124a	358	Step3:361:	11:36:48 AM	11:43:49 AM	1	1	Sum-of-Interior-Angles-Triangle	CORRECT
10	Stu_00411460f7c92d2124a	106	Step0:106:	11:44:08 AM	11:44:52 AM	1	1	Interpreting-Linear-Equations	HINT
11	Stu_00411460f7c92d2124a	106	Step1:107:	11:44:08 AM	11:45:58 AM	1	1	Interpreting-Linear-Equations	CORRECT
12	Stu_00411460f7c92d2124a	106	Step2:108:	11:44:08 AM	11:46:53 AM	1	1	Interpreting-Linear-Equations	CORRECT
13	Stu_00411460f7c92d2124a	106	Step3:109:	11:44:08 AM	11:49:05 AM	1	1	Interpreting-Linear-Equations	HINT
14	Stu_00411460f7c92d2124a	106	Step3:109:	11:44:08 AM	11:50:16 AM	1	2	Interpreting-Linear-Equations	INCORRECT
15	Stu_00411460f7c92d2124a	106	Step3:109:	11:44:08 AM	11:50:47 AM	1	3	Interpreting-Linear-Equations	CORRECT
16	Stu_00411460f7c92d2124a	106	Step4:111:	11:44:08 AM	11:51:35 AM	1	1	Interpreting-Linear-Equations	HINT
17	Stu_00411460f7c92d2124a	106	Step4:111:	11:44:08 AM	11:52:50 AM	1	2	Interpreting-Linear-Equations	INCORRECT
18	Stu_00411460f7c92d2124a	106	Step0:106:	12:01:12 PM	12:01:22 PM	2	1	Interpreting-Linear-Equations	CORRECT
19	Stu_00411460f7c92d2124a	90	Step0:90:W	12:01:25 PM	12:03:15 PM	1	1	Point-Plotting	CORRECT
20	Stu_00411460f7c92d2124a	95	Step0:95:Tl	12:07:54 PM	12:08:28 PM	1	1	Square-Root	INCORRECT
21	Stu_00411460f7c92d2124a	95	Step1:96:Fi	12:07:54 PM	12:08:50 PM	1	1	Square-Root	HINT
22	Stu_00411460f7c92d2124a	95	Step1:96:Fi	12:07:54 PM	12:08:58 PM	1	2	Square-Root	INCORRECT
23	Stu_00411460f7c92d2124a	95	Step1:96:Fi	12:07:54 PM	12:09:14 PM	1	3	Square-Root	CORRECT
24	Stu_00411460f7c92d2124a	95	Step2:97:G	12:07:54 PM	12:09:42 PM	1	1	Square-Root	CORRECT
25	Stu_00411460f7c92d2124a	95	Step3:98:W	12:07:54 PM	12:09:47 PM	1	1	Square-Root	INCORRECT
26	Stu_00411460f7c92d2124a	95	Step3:98:W	12:07:54 PM	12:09:53 PM	1	2	Square-Root	CORRECT
27	Stu_00411460f7c92d2124a	95	Step4:99:W	12:07:54 PM	12:09:59 PM	1	1	Square-Root	CORRECT
28	Stu_00411460f7c92d2124a	95	Step5:100:	12:07:54 PM	12:10:04 PM	1	1	Square-Root	CORRECT
29	Stu_00411460f7c92d2124a	95	Step6:101:	12:07:54 PM	12:10:09 PM	1	1	Square-Root	CORRECT
30	Stu_00411460f7c92d2124a	95	Step7:102:	12:07:54 PM	12:10:27 PM	1	1	Square-Root	INCORRECT

Figure 1.
An example of ASSISTments log files obtained from PSLC

Column in PSLC data set	Description
Anonymized student ID	Anonymized student ID generated by DataShop
Problem name	Name of the problem associated with a current transaction
Step name	Name of the problem solving step associated with a current transaction
Problem time	Time at which students started solving a problem
Step time	Time at which students started working on a particular problem solving step in the problem
Number of problem views	Number of times students tried to solve a current problem
Number of attempts at step	Number of times students submitted an answer to a current problem solving step
KC	Knowledge Component associated with a current transaction
Outcome	Result of a current problem solving attempt (correct, incorrect or hint)

Table I.
Problem solving information available in the PSLC data set

solving performance. Thus, for each transaction record in the PSLC data, its anonymized student ID, KC, and step time were identified first. This information is then used to compile all transaction records with the same anonymized student ID, KC, and earlier step time. From these records of past problem solving performance on the same KC by the same student, fraction of correct problem solving steps, fraction of incorrect problem solving steps, fraction of problem solving steps associated with hint request(s), and streak of correct answers were computed.

The first two variables, fraction of correct problem solving steps and fraction of incorrect problem solving steps, try to capture the ability of students solving a problem at the moment. Fraction of problem solving steps associated with hint request(s) is a predictor variable that tries to capture the ability of students in relation to the use of hints provided by the computer-based learning environment; if students had to use many hints to solve problems in the past, these students can be considered less capable than students who needed fewer hints. Since students may be able to get a problem correct without fully understanding the target mathematics concept, streak of correct answers to the problems on the same KC is also considered. When students found that they failed to solve a problem at their first attempt, they could learn from the feedback message and hints provided by the learning environment or they could use an external source of information (e.g. textbook).

In order to capture the effect of these activities, a binary variable indicating whether the transaction is a first attempt to solve a problem was used. Also, certain KCs and problem solving steps are more difficult than others. In order to account for the difficulty of KCs and problem solving steps, dummy variables for KCs and step names were created. In addition, dummy variables for outcomes from the last two problem solving attempts (Correct, Incorrect, Hint or N/A) were created because the most recent problem solving performance may be able to represent the ability of students more accurately. These variables were then used as predictors of the regularized logistic regression model described below. Finally, a new outcome variable (Correct or Wrong) was created by combining Incorrect and Hint cases because requesting a hint was considered a wrong attempt in this study. Using these pre-processed predictor and outcome variables, predictive models of problem solving performance of students replicating the behaviors of expert human teachers were developed, and their predictive powers were examined.

Method

Univariate analysis

Once data pre-processing was complete, a series of univariate analyses was conducted in order to examine the usefulness of the information obtained from the data pre-processing. First of all, the data were very sparse, which makes it difficult to build an accurate model of problem solving performance of students. The data set contained 895 students, 1,389 unique problem solving steps, and 66 KCs. But, the median of number of unique problem solving steps students tried to solve was mere 181. Moreover, several students tried to solve just one problem solving step, which resulted in what is called complete separation of data in which a predictor is associated with only one outcome value (Albert and Anderson, 1984). Since the standard logistic regression model cannot handle the data with complete separation, this study used regularized logistic regression as explained below.

Of the four continuous predictors, fraction of correct problem solving steps and fraction of hint requests showed an interesting relationship with the outcome variable. Figure 2 shows a histogram of correct/wrong answers vs fraction of correct answers students submitted in the past. When the fraction of correct answers from the past problem solving steps is greater than about 0.55, more students got the current problem solving step correct except for the cases where the fraction of correct answers is 1.0. Interestingly, when students got all past problem solving steps on the same KC correct, they appeared to have a roughly

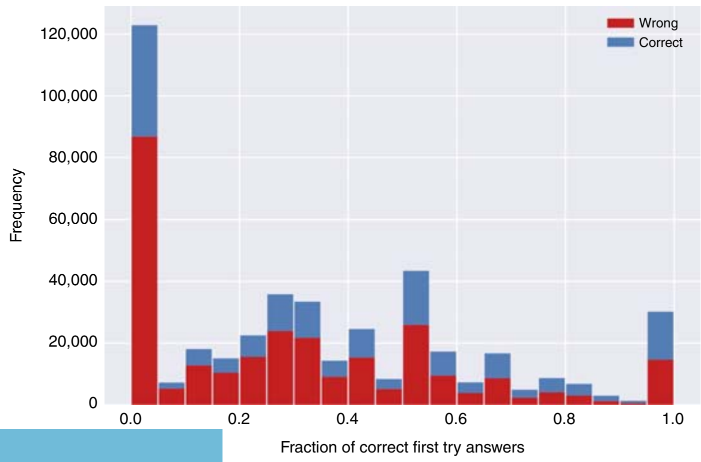

Figure 2.
Histogram of correct/wrong answers vs fraction of correct answers submitted in the past

50-50 chance to solve the current problem solving step correctly. Although more in-depth analyses are warranted to fully understand why it happened, this result appears to be related to the fact that the correct fraction of 1.0 occurs much more frequently when students solved a small number of problems in the past, which would not provide enough information about the ability of students being modeled.

Also, more hint requests are associated with more wrong answers (see Figure 3). When students requested hints more than about 60 percent of the problem solving steps they tried to solve in the past, they are more likely to fail to solve the current one. Similar to the fraction of correct answers, extreme fraction values (0.0 and 1.0) appear to contain more noise than intermediate values.

One important characteristic of the computer-based mathematics learning environment used in this study is that it can provide hints in response to wrong answers or student requests. Students can submit an answer to the same problem solving step more than once after using hint(s) provided by the computer-based learning environment. This study hypothesizes that the problem solving performance of students can be different, depending on whether they are trying to solve a problem for the first time or not. The fraction of correct answers was found to be larger at the first attempt (correct fraction = 0.43; 91,788 correct answers out of 214,455 attempts), compared to the subsequent attempts (correct fraction = 0.31; 70,148 correct answers out of 226,599 attempts), because of the selection effect.

When students were able to get the most recent problem solving step correct, they are more likely to get the current problem solving step correct (correct fraction = 0.45; 66,225 correct answers out of 147,848 attempts), compared to the students who either got the most recent problem solving step wrong or requested a hint. Interestingly, getting the most recent problem solving step wrong is associated with a little higher correct probability (correct fraction = 0.35; 57,374 correct answers out of 161,720 attempts) than hint request (correct fraction = 0.29; 38,337 correct answers out of 131,486 attempts). When students were able to get the second most recent problem solving step correct, they showed a similar correct fraction (correct fraction = 0.45; 57,847 correct answers out of 129,392 attempts). However, when students got two most recent problem solving steps correct in a row, the fraction of correct answers on the current problem solving step increased to 0.51 (32,223 correct answers out of 63,024 attempts) as shown in Table II.

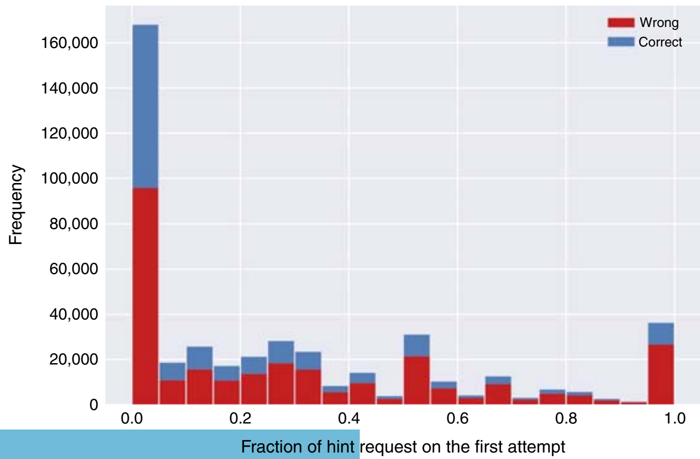

Figure 3. Histogram of correct/wrong answers vs fraction of hint requests made in the past

Table II. Frequency of correct/wrong answers at the current problem solving attempt vs last two problem solving performance

Outcomes from the last two problem solving steps	Outcome of the current problem solving step Correct	Wrong
2nd most recent outcomes only		
Correct	57,847	71,545
Hint	32,609	77,368
Incorrect	50,516	91,118
Most recent outcomes only		
Correct	66,225	81,623
Hint	38,337	93,149
Incorrect	57,374	104,346
Last two outcomes		
Correct, Correct	32,223	30,801
Correct, Hint	6,599	12,423
Correct, Incorrect	19,025	28,321

Regularized logistic regression model of problem solving performance of students

When the outcome variable is binary ($y_i = 0$ for wrong answers, and $y_i = 1$ for correct answers), the logistic regression model represents a conditional probability through a linear function of predictors x_i:

$$\Pr(y_i = 1 \mid x_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_j^T x_i)}} \tag{1}$$

Equation (1) allows for estimating a probability that a student can successfully solve a current problem solving step (y_i), given his or her problem solving performance observed in the past (x_i), and regression coefficients (β_0 and β_j). In the ordinary logistic regression model, regression coefficients can be obtained by maximizing a binomial log likelihood function (Hastie *et al.*, 2013).

However, as briefly mentioned above, due to the complete separation issue, the solution to the binomial log likelihood function did not exist for the data set analyzed in this study. To get around this problem, this study used a regularized logistic regression model which has an additional penalty term in the objective function to be maximized as shown in the following equation (Hastie *et al.*, 2013). Note that $P(\beta_j)$ selects the type, and λ determines the amount of regularization. When $\lambda = 0$, the following equation reduces to the objective function of the ordinary logistic regression model:

$$\max_{\beta_0, \beta_j} \left\{ \sum_{i=1}^{N} \left[y_i \left(\beta_0 + \beta_j^T x_i \right) - \log \left(1 + e^{\beta_0 + \beta_j^T x_i} \right) \right] - \lambda P(\beta_j) \right\} \tag{2}$$

When L_1 norm is used for the penalty term, $P(\beta_j)$, the regularized logistic regression model can exclude some predictors in maximizing Equation (2). When L_2 norm is used, on the other hand, the regularized logistic regression model includes all predictors, but the magnitude of regression coefficients, β_j, is reduced instead. We can think of regularization as a penalty against complexity. The ordinary logistic regression model is always more complex than the regularized logistic regression model because it includes all regression coefficients whose magnitude is not controlled. Since the complex model can pick up "peculiarities" or "noise" in the current data being modeled, it does not generalize well to new unseen data that do not have such peculiarities or noise (Hastie *et al.*, 2013). Since the regularized logistic regression model penalizes a complex model by way of its penalty term, the effect of peculiarities or noise in the data being modeled can be minimized, which can result in a better performance on the new unseen data. The regularized logistic regression model can address the issue of complete separation of data because data points causing complete separation can be treated as peculiarities or noise in the data. This study used the LIBLINEAR library (Fan *et al.*, 2008) in building four regularized logistic regression models using L_1 and L_2 penalties with (RLR-L1 and RLR-L2) and without KCs (RLR-L1-No-KC and RLR-L2-No-KC) as predictors.

In order to estimate the predictive power of regularized logistic regression models without bias, the pre-processed data set was divided into training and test sets. When creating a test set, which consists of 20 percent of the pre-processed data, stratified random sampling was used to ensure that the ratio of positive ($y_i = 1$) to negative ($y_i = 0$) instances in both sets are similar. Fivefold cross-validation was used to find the best value for the tuning parameter λ of the regularized logistic regression models. For each λ value in $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$, the training set was randomly divided into five sets of roughly equal size with similar proportions of positive and negative instances. Then, a regularized logistic regression model was built using all samples in the training set except for one subset. An unbiased predictive power of the regularized logistic regression model with a specific λ value was estimated from the samples in the held-out training set that played a role of unseen future data. For the same λ value, these processes were repeated five times with a different subset of the training set being used as a held-out set. The average of the five estimates of predictive power was used to represent how well a regularized logistic regression model with a specific value of λ could predict successful problem solving performance of students in the future. The final prediction models were built by fitting the entire training set with the best λ value determined from the fivefold cross-validation procedure. Finally, these models were used to predict whether students in the test set, which was not used in the model building process, would be able to resolve problem solving steps successfully, and their predictions were compared to the observed problem solving performance recorded in the test set.

Results

Figure 4 compares the two sets of receiver operator characteristic (ROC) curves of the two regularized logistic regression and the standard BKT models plotted for the test set.

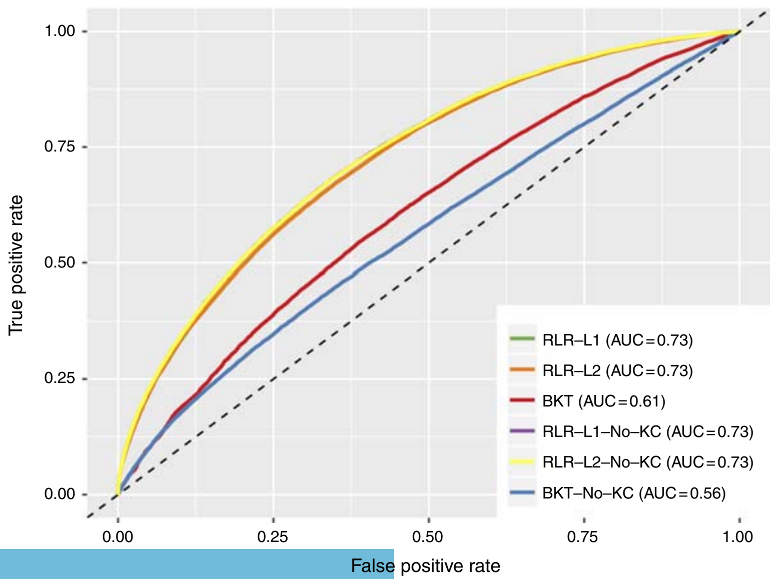

Figure 4. Comparison of ROC curves and AUC values of predictive models of problem solving performance of students

While the first set of predictive models (RLR-L1, RLR-L2 and BKT) included KCs as predictor variables, the second set of predictive models (RLR-L1-No-KC, RLR-L2-No-KC, BKT-No-KC) did not, allowing for examining the importance of KCs as predictor variables in the model. Publicly available C++ codes (https://github.com/IEDMS/standard-bkt) were used in building the standard BKT model reported in this study. When applied to a binary classification problem like this study, an ROC curve is a plot of false positive (incorrectly predict that students will be able to solve a current problem solving step when they failed) vs true positive (correctly predict that students will be able to solve a current problem solving step) rates. Area under an ROC curve (AUC), which varies from 0.5 (predictive power no better than simple guessing) to 1.0 (perfect prediction), represents a quality of a binary classification model because a good classification model has a small false positive rate and a large true positive rate. AUC computed on the test set represents a probability that a binary classification model can correctly predict an outcome variable when it is provided with a new data set in the future (Fawcett, 2006). The ROC curves from four regularized logistic regression models (RLR-L1, RLR-L2, RLR-L1-No-KC, and RLR-L2-No-KC) are completely overlapping, and their AUC values (0.73) are about 20 percent larger than that of BKT with KCs (0.61), indicating that the regularized logistic regression models would be able to perform better than the standard BKT model in predicting future problem solving performance of students. When KCs are included in the predictive model, the standard BKT model's predictive power increased by approximately 9 percent. However, KCs seem to have no impact on the predictive power of the regularized logistic regression model, suggesting that predictor variables capturing past problem solving performance of students contain more information than KCs do. Since the AUC values of the four regularized logistic regression models are essentially same, only the regularized logistic regression model with KCs and L_1 penalty (LRL-L1) is discussed in the subsequent sections.

The confusion matrices of prediction models reveal that the regularized logistic regression model shows a more balanced performance (see Figure 5). Since the BKT model predicted that just about 14 percent of answers would be correct, its recall for correct responses is very small (0.20 = 6,300/(6,300 + 25,827)), compared to wrong responses (0.89 = 50,019/(50,019 + 6,064)). On the other hand, the regularized logistic regression model predicted about 45 percent correct responses, and showed comparable recall values: 0.66 (= 21,052/(21,052 + 11,075)) for correct responses and 0.67 (= 37,314/(37,314 + 18,769)) for wrong responses. As a result, the BKT model had a much higher false positive rate (0.80 = 25,827/(25,827 + 6,300)), compared to the regularized logistic regression model (0.34 = 11,075/(11,075 + 21,052)).

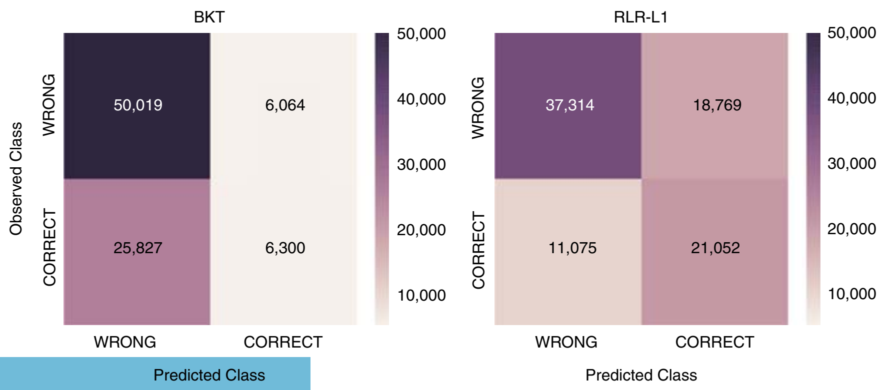

Figure 5.
Comparison of confusion matrices with a cut-off probability of 0.5

Table III shows a breakdown of AUC values by the difficulty of problem solving steps. The BKT model showed a larger AUC value for easy problem solving steps (step difficulty < 25th percentile), but its AUC value decreased by about 17 percent as the difficulty of problem solving steps increases (25th percentile ⩽ step difficulty < 50th percentile). AUC values of the BKT model are not much better than simple guessing when the step difficulty is greater than 25th percentile. Compared to the BKT model, AUC values of the regularized logistic regression model were less sensitive to step difficulty. It showed larger AUC values for both easy and difficult problem solving steps, and its AUC value did not decrease as much as the BKT model when the problem solving steps had medium difficulty (25th percentile ⩽ step difficulty < 75th percentile).

Discussion

This study found that regularized logistic regression models with L_1 and L_2 penalties made better predictions on the problem solving performance of students than the standard BKT model, the most frequently used quantitative model of student learning in EDM. One possible explanation for this result is that regularized logistic regression models incorporated more information about how students solved relevant problems and what kinds of problems they were. While the standard BKT model had only one predictor, correctness of the most recent problem solving attempt, regularized logistic regression models developed in this study included predictor variables capturing various aspects of problems and problem solving performance of students. These additional predictors seem to have enabled regularized logistic regression models to have a better predictive power than the standard BKT model does. This interpretation is in line with what Pardos and Heffernan (2011) found in their study; they were able to improve the predictive power of their learning model by including problem difficulty as an additional parameter to the standard BKT model.

It is not easy to develop KCs that can improve the performance of predictive models; KCs created by PSLC researchers increased the predictive power of standard BKT model only by about 9 percent. Moreover, once KCs are developed, they have to be tagged to all problems manually, which can be time consuming as the number of problems in the computer-based learning environment increases. Considering the fact that predictive power of regularized logistic regression models did not deteriorate when KCs were not used as predictor variables, regularized logistic regression model can be considered more efficient than standard BKT model.

One thing that makes it difficult to build an accurate predictive model of learning performance of students in the computer-based learning environment is gaming behavior. It has been found that some learners do not exert enough cognitive efforts to learn from learning activities and instructional supports provided in the computer-based learning environment (Baker *et al.*, 2008; Lee, 2015). For instance, students showing gaming behaviors would keep asking for hints, without reflecting on the information presented in the hint, in order to get to the last hint which typically provides an answer. Since these students are not going to learn from hints, using hints would be negatively correlated with the learning outcomes of these students, which could diminish the overall benefit of using hints. Thus, it

Step difficulty	AUC BKT	AUC RLR-L1
< 25th percentile	0.65	0.72
25th-50th percentile	0.55	0.64
50th-75th percentile	0.53	0.63
⩾ 75th percentile	0.55	0.70

Table III. AUC values of prediction models vs difficulty of problem solving steps would be important to include a predictor variable that can capture the gaming behavior of students in order to build a more accurate model of learning performance of students. One way to address this issue might be to incorporate problem solving time into the model. Examining how much time students spent before attempting to answer the same problem again or requesting additional hints might allow us to encode the existence of gaming behavior of students, which can in turn enhance the accuracy of the model.

Providing instructional supports and guidance is essential to facilitate student learning especially in the computer-based learning environment in which students are mostly learning on their own. Since most computer-based learning environments rely on simple heuristics or student demand, they are not able to provide instructional supports and guidance when students need them the most. By incorporating a predictive model of student performance into computer-based learning environments, we may be able to provide more personalized instructional supports and guidance tailored to the ability of students and the difficulty of learning tasks, which can potentially increase the efficiency of student learning.

Limitations of study

Learning is a complex phenomenon which can take many different forms. Since this study focuses on one specific form of learning, namely solving mathematics problems while using hints from a computer-based learning environment or other external resources, the findings from this study may not be generalized to other forms of learning. Also, although log files of computer-based learning environments provide rich information about learning behaviors of students, there are many things missing in the log files. For instance, the log files analyzed in this study do not provide information on student gender and affection, which may have an impact on how students used the computer-based learning environment. Similarly, these log files do not provide all learning activities students experienced; it is possible that students may have used an external source of information such as textbook or peer collaboration. It would be meaningful to examine whether such information can improve the predictive power when they are incorporated into the model.

Conclusion

The goal of this study was to develop a quantitative model that can predict whether middle school students can solve a mathematics problem without using any hints provided in the computer-based learning environment, based on how well they solved relevant problems in the past. Although providing instructional scaffolding is critical in facilitating student learning (Koedinger and Aleven, 2007), most computer-based learning environments are using simple heuristics or relying on students when they determine whether or not instructional scaffolding needs to be provided, which is unlikely to maximize the learning outcome of students. The findings from this study may suggest that the regularized logistic regression can be used in building a quantitative model of problem solving performance of students that can help determine when to provide instructional supports and guidance to students with different abilities in the computer-based learning environment.

References

Albert, A. and Anderson, J.A. (1984), "On the existence of maximum likelihood estimates in logistic regression models", *Biometrika*, Vol. 71 No. 1, pp. 1-10.

Baker, R., Walonoski, J., Heffernan, N.T., Roll, I., Corbett, A. and Koedinger, K. (2008), "Why students engage in 'gaming the system' behavior in interactive learning environments", *Journal of Interactive Learning Research*, Vol. 19 No. 2, pp. 185-224.

Beck, J.E. and Chang, K.M. (2007), "Identifiability: a fundamental problem of student modeling", in Conati, C., McCoy, K. and Pallouras, G. (Eds), *Advances in Neural Information Processing Systems*, Springer-Verlag, Berlin, pp. 137-146.

Clark, R.C. and Mayer, R.E. (2003), *E-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, Pfeiffer, San Francisco, CA.

Corbett, A. and Anderson, J. (1995), "Knowledge tracing: modeling the acquisition of procedural knowledge", *User Modeling and User-Adapted Interaction*, Vol. 4 No. 4, pp. 253-278.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. and Lin, C.-J. (2008), "LIBLINEAR: a library for large linear classification", *Journal of Machine Learning Research*, Vol. 9 No. 1, pp. 1871-1874.

Fawcett, T. (2006), "An introduction to ROC analysis", *Pattern Recognition Letters*, Vol. 27 No. 8, pp. 861-874.

Hastie, T., Tibshirani, R. and Friedman, J. (2013), *The Elements of Statistical Learning*, Springer Science & Business Media, New York, NY.

Kapur, M. (2008), "Productive failure", *Cognition and Instruction*, Vol. 26 No. 3, pp. 379-424.

Koedinger, K.R. and Aleven, V. (2007), "Exploring the assistance dilemma in experiments with cognitive tutors", *Educational Psychology Review*, Vol. 19 No. 3, pp. 239-264.

Koedinger, K.R., Baker, R., Cunningham, K., Skogsholm, A., Leber, B. and Stamper, J. (2010), "A data repository for the EDM community: the PSLC DataShop", in Romero, C. *et al.* (Eds), *Handbook of Educational Data Mining*, CRC Press, Boca Raton, FL, pp. 43-55.

Lawless, K.A. and Brown, S.W. (1997), "Multimedia learning environments: issues of learner control and navigation", *Instructional Science*, Vol. 25 No. 2, pp. 117-131.

Lee, Y.-J. (2015), "Analyzing log files to predict students' problem solving performance in a computer-based physics tutor", *Educational Technology & Society*, Vol. 18 No. 2, pp. 225-236.

Pardos, Z.A. and Heffernan, N.T. (2011), "KT-IDEM: introducing item difficulty to the knowledge tracing model", in Konstan, J. *et al.* (Eds), *Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization*, Springer-Verlag, Berlin, pp. 243-254.

Pardos, Z.A., Gowda, S.M., Baker, R. and Heffernan, N.T. (2012), "The sum is greater than the parts: ensembling models of student knowledge in educational software", *SIGKDD Explorations Newsletter*, Vol. 13 No. 2, pp. 37-44.

Pardos, Z., Bergner, Y. and Seaton, D.T. (2013), "Adapting Bayesian Knowledge Tracing to a Massive Open Online Course in edX", in D'Mello, S.K. *et al.* (Eds), *6th International Conference on Educational Data Mining*, International Educational Data Mining Society, Memphis, TN, pp. 137-144.

Schmidt, R.A. and Bjork, R.A. (1992), "New conceptualizations of practice: common principles in three paradigms suggest new concepts for training", *Psychological Science*, Vol. 3 No. 4, pp. 207-217.

Corresponding author
Young-Jin Lee can be contacted at: yjlee@ku.edu