Building Models to Predict Hint-or-Attempt Actions of Students

Francisco Enrique Vicente Castro **Seth Adjei** **Tyler Colombo** **Neil Heffernan**
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609, USA
{fgcastro, saadjei, tjcolombo nth}@wpi.edu

ABSTRACT
A great deal of research in educational data mining is geared towards predicting student performance. Bayesian Knowledge Tracing, Performance Factors Analysis, and the different variations of these have been introduced and have had some success at predicting student knowledge. It is worth noting, however, that very little has been done to determine what a student's first course of action will be when dealing with a problem, which may include attempting the problem or asking for help. Even though learner "course of actions" have been studied, it has mostly been used to predict correctness in succeeding problems. In this study, we present initial attempts at building models that utilize student action information: (a) the number of attempts taken and hints requested, and (b) history backtracks of hint request behavior, both of these are used to predict a student's first course of action when working with problems in the ASSISTments tutoring system. Experimental results show that the models have reliable predictive accuracy when predicting students' first course of action on the next problem.

Author Keywords
Educational data mining; intelligent tutoring systems; student modeling; student behavior.

1. INTRODUCTION
Most educational data mining (EDM) research focus on modeling student behavior and performance. Algorithms such as Bayesian Knowledge Tracing [1], and Performance Factors Analysis [4] have been used to achieve this end. In intelligent tutoring systems, it is crucial to be able to understand student behavior to provide better tutoring practices and improved content selection for these systems. Student behavior may provide another means to identify low-knowledge or low-performing students and determine when to proactively intervene. Previous works show that students who are more likely to ask for help on problems learn less and perform less. A study on students' help-seeking behavior in an SQL tutoring system [3] suggests that students who used help very frequently had the lowest learning rate and had shallow learning. A study that used the sequence of attempts and hint requests to predict student correctness found that students who first made attempts on problems performed better than those who requested for help first [2]. The Assistance Model [6] used the number of hints and attempts a student needed to answer a previous question to predict student performance. Gaining the capability to recognize students' need for assistance ahead of time by looking at students' pattern of actions could lead to more proactive interventions, such as identifying prerequisite skills, adapting pedagogical methodologies, or gaining insight on student problem solving methodologies.

With these in mind, we then ask: how do we determine when students will ask for help when using an ITS? On the exploratory level of model development, what information may be useful for developing models that forecast students' need for assistance? In this work, we define two models that use information on problem attempts and help requests used by students in the ASSISTments tutoring system: (1) *Attempt/Hint Count model* (AHC) makes use of information on the number of attempts and hints used by students on a question to predict the occurrence of a help request as the first action on the next problem, and (2) *Hint History model* (HH) makes use of the history of hint request as the first action in preceding questions to predict the occurrence of a help request as the first action on the next problem.

We utilized tabling methods to generate prediction values from the information used by each model. Tabling methods have been found to be effective alternatives for performing predictions using datasets and offer the advantage of being computationally inexpensive and easily expandable to leverage more features into simple models [2, 7].

2. DATASET
The data used in the analysis is from ASSISTments, an online tutoring system maintained at the Worcester Polytechnic Institute that provides tutorial assistance if students make incorrect attempts or ask for help [5]. The dataset is from released ASSISTments data that spans about five months within the 2012-2013 school year, containing

599,368 student log entries. More details about ASSISTments data can be accessed from: https://sites.google.com/site/assistmentsdata/how-to-interpret.

Analysis for the AHC model was done on problem logs with 1 to 5 attempts taken in answering problems, accounting for 98% of all data entries (585,926 rows). Problem entries with 3, 4, and 5 available hints (AvH) were used and these accounted for 70% of the data (415,895 rows). The resulting dataset contains 420 problem sets and 12,966 students, totaling to 299,968 entries. The resulting dataset was separated into problem groups that differed in the number of available hints to avoid comparing the hint request behavior of students who had more opportunities to hint against students with fewer opportunities to do so.

Problem Group	Problem Sets	Students	Dataset entries
3 AvH	285	11,402	169,100
4 AvH	224	10,282	111,754
5 AvH	60	4,724	19,114

Table 1. AHC dataset for each of the problem groups

For the HH model, we selected entries in the dataset where each student sequence had at least 4 rows. The student sequence is the sequence of problems that a student answered. Sequences had to at least have 4 rows for the HH model which looks at the history of hint use, 3 problems prior the next problem. The resulting dataset contained 279,925 entries with 555 problem sets and 12,429 students.

3. STUDENT ACTION MODELS

In ASSISTments, students exhibit varying behaviors when encountering problems: submitting an answer to a problem first ("attempting the problem"), asking for help (hint) first, asking for hints after an initial attempt, alternating between attempts and requests for hints, or continuously attempting a problem until a correct answer has been submitted. These behaviors have likewise been observed in [2].

3.1 Initial Experiments: AHC

The AHC prediction table maps the number of attempts and hints used to the probability that the student attempted or asked for a hint on the next problem. The probability is the percentage of students who asked for a hint on the next problem. Table 2 shows a sample prediction table from training data. Table 3 shows a matching scenario using Table 2. A value under *Hints Taken* in Table 2 such as 2/3 indicates that a student used 2 out of 3 available hints for the problem and values on the first column indicate the count of attempts. Five-fold cross validation was used to train and test the AHC model on the three problem groups. Problem set and student-level analyses were done to see whether the model generalizes across unseen problem sets and students.

3.2 Secondary Experiment: HH

For HH analysis, the prediction table was generated by using the percentage of hint use as first action in three previous problems. Table 4 shows a prediction table from training data. Column labels correspond to the number of times the first action was an attempt on the problem or a hint request. For example, 1H/2A indicates that in three prior problems, a total of 1 hint as first action and 2 attempts as first action were used. Counts of attempts and hints as first action were then generated for each column. In the table, for those who used a total of 2 hints and 1 attempt in three previous problems, there are 3330 instances of attempts and 1833 instances of hint requests as first action on the next problem. *% Hint* is the percentage of instances of hint use within the bin. Problem set and student-level five-fold cross validation was used to train and test the HH model.

Attempts Taken	Hints Taken 0 / 3	1 / 3	2 / 3	3 / 3
1	0.0211	0.1001	0.2213	0.4025
2	0.0261	0.0558	0.0747	0.1105
3	0.0237	0.0447	0.0737	0.0916
4	0.0363	0.0287	0.0743	0.0949
5	0.0132	0.0263	0.0857	0.0912

Table 2. AHC Prediction Table

Student	A_C	H_C	H_T	FANP
92677	1	0	3	0.0211
92680	2	3	3	0.1105

Table 3. Matching scenario using Table 2 (Note: A_C = Attempt Count, H_C = Hint Count, H_T = Hint Total, FANP = First Action Next Problem)

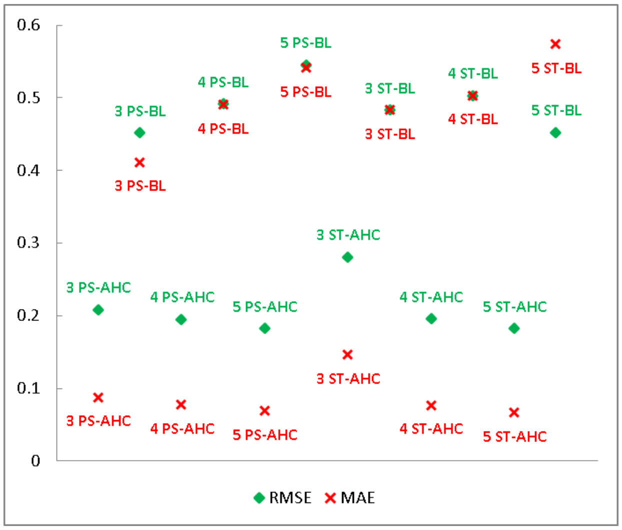
	Previous 3 First Action Hints / Attempts			
	0H / 3A	**1H / 2A**	**2H / 1A**	**3H / 0A**
# Attempt	111017	17219	3330	683
# Hint	5859	3254	1833	1663
% Hint	0.0501	0.1589	0.3550	0.7089

Table 4. HH Prediction Table

To analyze whether the number of history points affected the predictive power of HH, an additional analysis with four problems prior the next problem was done.

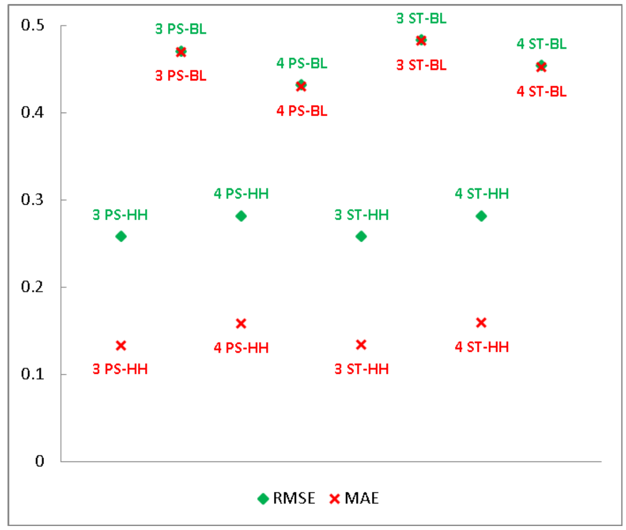
4. RESULTS AND DISCUSSION

The predictive performance of the AHC and HH models were evaluated using root mean squared error (RMSE), mean absolute error (MAE), and area under the ROC curve (AUC). Additionally, a naïve baseline (BL) model was generated for comparison, as we have found no other gold standard model for first-course-of-action prediction to compare our work with. The BL model uses the percentage of hint instances on the students' second action on all problems in the dataset. Table 5 shows a scenario for BL prediction. *Hint %* is the percentage of hint instances in the problem entries, which translates to a prediction on the students' first action on the next problem. If a student's second action on the current problem is a hint, the prediction for FANP is *Hint %*, otherwise, use *Attempt %*. The intuition for this is the hypothesis that students who have greater tendency to ask for hints on succeeding actions may most likely ask for hints in succeeding problems.



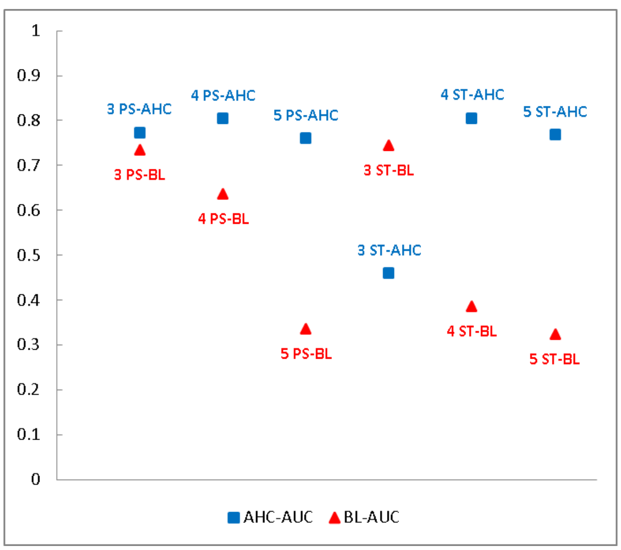
PS	3 AHC	3 BL	4 AHC	4 BL	5 AHC	5 BL
RMSE	0.2075	0.4506	0.1942	0.4910	0.1813	0.5445
MAE	0.0866	0.4104	0.0763	0.4899	0.0677	0.5403
ST	**3 AHC**	**3 BL**	**4 AHC**	**4 BL**	**5 AHC**	**5 BL**
RMSE	0.2799	0.4826	0.1945	0.5023	0.1811	0.4514
MAE	0.1452	0.4821	0.0758	0.5022	0.0653	0.5729

a. RMSE and MAE performance for AHC vs. BL across three problem groups (3, 4, and 5 available hints)



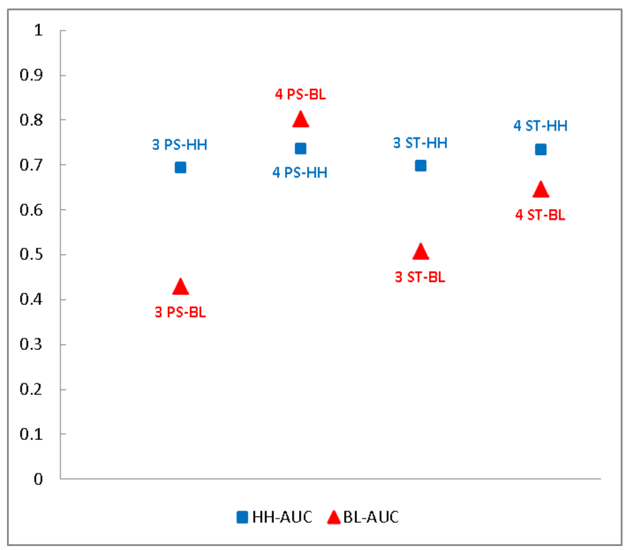
PS	3 HH	3 BL	4 HH	4 BL
RMSE	0.2574	0.4697	0.2809	0.4307
MAE	0.1327	0.4687	0.1572	0.4291
ST	**3 HH**	**3 BL**	**4 HH**	**4 BL**
RMSE	0.2573	0.4821	0.2808	0.4528
MAE	0.1328	0.4810	0.1580	0.4513

b. RMSE and MAE performance for HH vs. BL for 3 and 4 prior problems



PS	3 AHC	3 BL	4 AHC	4 BL	5 AHC	5 BL
AUC	0.7737	0.7332	0.8043	0.6338	0.7602	0.3338
ST	**3 AHC**	**3 BL**	**4 AHC**	**4 BL**	**5 AHC**	**5 BL**
AUC	0.4599	0.7419	0.8056	0.3841	0.7689	0.3223

c. AUC performance for AHC vs. BL across three problem groups (3, 4, and 5 available hints)



PS	3 HH	3 BL	4 HH	4 BL
AUC	0.6936	0.4298	0.7357	0.8026
ST	**3 HH**	**3 BL**	**4 HH**	**4 BL**
AUC	0.6989	0.5071	0.7355	0.6458

d. AUC performance for HH vs. BL for 3 and 4 prior problems

Figure 1. Problem set (PS) and student (ST) level RMSE and MAE performance for AHC, HH, and BL (a and b); Problem set and student level AUC performance for AHC, HH, and BL (c and d).

Problem entries	Hint Count: 2nd Action	Hint % (BL)	Attempt %
2200	852	0.3872	0.6127

Table 5. Sample scenario for BL prediction values

4.1 AHC Analysis

Problem set level findings for both AHC and BL are presented in Figure 1a. AHC consistently outperforms BL across all problem groups in both RMSE and MAE. Lower values for both metrics indicate better model fit. A reliability analysis to compare AHC with BL using a two-tailed paired t-test indicates that the findings are reliably different across all problem groups (p=0). The effectiveness of the model is likewise seen using the AUC metric (Figure 1c). AUC values closer to 1 indicate better model fit. It can be noted that AHC performance in all metrics are closely consistent, suggesting that the model is fairly generalizable across problems with varying numbers of hint availability. Predictive performance using student level analysis for problems with 4 and 5 available hints is fairly consistent across all three metrics; however, the model does not perform as well for problems with 3 available hints, suggesting that AHC may be used to predict the hint request behavior of unseen students, provided there is a high number of opportunities to ask for help. BL performance fails to improve as the number of available hints increase for both problem set and student-level analyses.

4.2 HH Analysis

A problem set level analysis of the HH model across the number of prior history points demonstrates that the HH model maintains a fairly consistent level of predictive performance across all three metrics. While HH significantly outperforms BL in MAE and RMSE, it is outperformed by the latter in AUC for 4 history points. This may be because the ordering of values in BL's predictions is not as close to the actual as those of HH. This situation rarely happens; we may have to try another dataset to confirm this behavior. On a student level analysis, HH outperforms BL across all values of first action prior history points (Figures 1b and 1d). A reliability analysis to compare HH with BL using a two-tailed paired t-test indicates that the findings are reliably different across all prior hint history with p=0. There is a consistency of results for all performance metrics for HH, while BL exhibits more prominent fluctuation in its results, suggesting that the HH model can be feasibly used to predict student hint request behavior for both unseen skills and unseen students, as well as across the number of first action history points with fair reliability.

5. CONTRIBUTION AND FUTURE WORK

Results of the experiments suggest that students' help request behavior can be feasibly predicted from data that are descriptive of student action information. While the methods in this study are a starting point in using action information, we feel that such initiatives are worth discussing for building up further studies in the field. The models provide utility for predicting when students will ask for help, using dataset information on problem attempts and help requests. Both models predicted students' first course of action when answering problems from an ITS with fairly consistent predictive performance and generalizability.

Future improvements to these models may include the accounting of patterns in student actions which may provide a rich source of information for possible prediction of need for assistance by students (partly explored here with the BL model). The dataset used contained other information including student response times and skill difficulty and exploiting these may provide further insight into factors of assistance need to aid in developing a proactive and effective early intervention framework. These models should be tested on other ITS datasets to determine whether these models are consistent across different datasets.

REFERENCES

[1] Corbett, A.T. and Anderson, J.R. Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction 4, 4 (1994), 253–278.

[2] Duong, H., Zhu, L., Wang, Y. and Heffernan, N.T. (2013). A prediction model that uses the sequence of attempts and hints to better predict knowledge: "Better to attempt the problem first, rather than ask for a hint". In S. D'Mello, R. Calvo, & A. Olney (Eds.) Proceedings of the 6th International Conference on Educational Data Mining (EDM2013). Memphis, TN. pp. 316-317.

[3] Mathews, M. and Mitrović, T. How Does Students' Help-Seeking Behaviour Affect Learning? In B.P. Woolf, E. Aïmeur, R. Nkambou and S. Lajoie, eds., Intelligent Tutoring Systems. Springer Berlin Heidelberg, 2008, 363–372.

[4] Pavlik, P. I., Cen, H., and Koedinger, K. (2009) Performance Factors Analysis - A New Alternative to Knowledge Tracing. Proceedings of the 14th International Conference on Artificial Intelligence in Education, Brighton, UK, pp. 531-538.

[5] Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N.T., Koedinger, K.R., Junker, B., Ritter, S., Knight, A., Aniszczyk, C., Choksey, S., Livak, T., Mercado, E., Turner, T.E., Upalekar. R, Walonoski, J.A., Macasek, M.A., and Rasmussen, K.P. (2005). The Assistment project: Blending assessment and assisting. In C.K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.) Proceedings of the 12th Artificial Intelligence in Education, Amsterdam: ISO Press. pp. 555-562.

[6] Wang, Y. and Heffernan, N. (2011). The "Assistance" Model: Leveraging How Many Hints and Attempts a Student Needs. The 24th International FLAIRS Conference.

[7] Wang, Q.Y., Kehrer, P., Pardos, Z. and Heffernan, N. Response Tabling– A simple and practical complement to Knowledge Tracing. KDD 2011 Workshop: Knowledge Discovery in Educational Data.