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

Francisco Enrique Vicente Castro
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609, USA
{fgcastro, saadjei, tjcolombo, nth}@wpi.edu

Seth Adjei
Tyler Colombo
Neil Heffernan

ABSTRACT
A great deal of research in Educational Data Mining (EDM) are geared towards predicting student performance. Models including Bayesian Knowledge Tracing (BKT), Performance Factors Analysis (PFA), and different variations of BKT and PFA 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 answering a problem, which may include immediately attempting the problem or asking for help or hints to work on a problem. Even though learner “course of actions” have been studied, it has mostly been used to predict their next performance or probabilities of correctness in succeeding problems. In this study, we present a number of Action Behavior (AB) models – initial attempts at models that utilize student action information, namely: (1) Action Behavior – Attempt/Hint Count (AB-AHC): the number of attempts taken and number of hints requested, and (2) Action Behavior – Hint History (AB-HH) a history backtrack of hint request behavior, both used to predict a student’s first course of action when answering problems in the ASSISTments intelligent tutoring system. Experimental results show that the Action Behavior 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 research in educational data mining in intelligent tutoring systems focus on modeling student behavior by looking at correctness of student responses to track student performance, where performance is usually taken to mean correctness on a student’s first attempt on a problem. Algorithms such as Bayesian Knowledge Tracing (BKT), which uses dynamic Bayesian networks to model student learning [1], and Performance Factors Analysis (PFA) [6] have been used to achieve this end. In fact, a good number of modifications have been introduced to the BKT model to improve predictive performance, such as including additional features in the model like item difficulty, using partial credit, and individualization of model parameters [4, 5, 8], but these efforts give little significance to determining when a student would need help based on a student’s behavior in working on problems.

In Intelligent Tutoring Systems (ITS), it is crucial to be able to understand student behavior to improve and provide better tutoring practices as well as improved content selection for these systems. Considering the behavior of students when dealing with problems may provide another means to identify low-knowledge or low-performing students and determine when to proactively intervene and help out these students. In previous works, it has been shown that students who are more likely to ask for help on problems learn less, and thus, perform less. A study on students’ help-seeking behavior in an SQL tutoring system by Mathews and Mitrovic [11] has suggested that students who used help very frequently had the lowest learning rate and had shallow learning. Another study that used a Sequence of Action (SOA) model, where the sequence of attempts and hint requests were used to predict student correctness on succeeding problems found that students who first made attempts on problems performed better on the next question than those who requested for help first, such as asking for a hint [2]. A similar work [9] called the Assistance Model (AM) used information about the number of hints and attempts a student needed to answer a previous question to predict student performance. Ensembling AM with BKT and PFA showed a reliable increase in predictive accuracy. Gaining the ability to recognize students’ need for assistance ahead of time by looking at students’ course and pattern of actions could lead to more effective and proactive interventions, such as identifying necessary prerequisite skills, revising possibly ill-formed tests or problem items in tutoring systems, adapting pedagogical methodologies, or even gaining insight on problem solving methodologies employed by students. As an application of
We thus identify and seek to answer an important research question: can we determine students’ need for assistance ahead of time? On the exploratory level of model development, we would also like to ask: what pieces of information may be useful for developing or engineering models that forecast students’ need for assistance?

In this work, we define two Action Behavior (AB) models, which make use of information on problem attempts and help requests employed by students in working on problems in the ASSISTments tutoring system:

1. Action Behavior – Attempt/Hint Count (AB-AHC) makes use of information on the number of attempts and hints used by students on a current question to predict the occurrence of a help request as the first course of action on the next problem.

2. Action Behavior – Hint History (AB-HH) makes use of history information on the number of hints requested as the first course of action in immediately preceding questions to predict the occurrence of a help request as the first course of action on the next problem.

For both AB models, we utilized tabling methods to generate our 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, 10]. The succeeding sections discuss the implementations of the AB models, the datasets that were used for each, and the results of the models.

2. DATASET
The data used in action behavior analysis comes from ASSISTments, an online tutoring system maintained at the Worcester Polytechnic Institute that provides tutorial assistance if students make incorrect attempts or asks for help [7]. The dataset was taken from released ASSISTments data that spans about five months within the 2012-2013 school year, which contained 599,368 rows of student log entries and 4163 unique students. A log entry/row represents a single problem of a student, which contains information such as student id, the first action employed by a student when answering a problem (either attempt to answer the problem or request for a hint), hint counts, which is the number of hints a student has requested for in solving a problem, and attempt counts, which is the number of times a student has attempted or submitted an answer for a problem. More detail about information in the ASSISTments data can be accessed from: https://sites.google.com/site/assistmentsdata/how-to-interpret.

2.1 Initial AB-AHC Analysis Dataset
We started with an initial analysis of the data using the AB-AHC model. We used problem logs with at least one attempt and at most 5 attempts. Each of the problems we analyzed from the logs had between 3 and 5 available hints (AvH). We chose to do the initial analysis on entries with one to five attempts taken in answering problems because these were the top five attempt counts that cover a significant number of row entries, accounting for 98% of all the entries (585,926 rows) in the data. We used problem entries with three, four, and five available hints and these accounted for 70% of the data (415,895 rows). Further, we included data logs on main problems and excluded scaffolding problems. Scaffolding in ASSISTments is a feedback mechanism in the system that breaks a problem down into steps. We excluded scaffolding from our data because once a student opted to break the problem down, ASSISTments then marks the problem as incorrect. The resulting dataset contains 420 unique problem sets and 12,966 students, totaling to 299,968 problem entries. Table 2 presents a breakdown of the dataset.

<table>
<thead>
<tr>
<th>Problem Group</th>
<th>Problem Sets</th>
<th>Students</th>
<th>Dataset entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 AvH</td>
<td>285</td>
<td>11,402</td>
<td>169,100</td>
</tr>
<tr>
<td>4 AvH</td>
<td>224</td>
<td>10,282</td>
<td>111,754</td>
</tr>
<tr>
<td>5 AvH</td>
<td>60</td>
<td>4,724</td>
<td>19,114</td>
</tr>
</tbody>
</table>

Table 2. Number of problem sets and dataset entries (rows) for each of the problem groups

The main reason why the resulting dataset was separated into problem groups that differed in the number of available hints for the problems was to avoid biasing the predictions against those problems that presented fewer opportunities to...
2.2 Secondary Analysis Dataset: AB-HH
For the secondary analysis using AB-HH, we selected problem entries in the dataset where each student sequence had more than four rows. The student sequence is the sequence of problems that a student answered. The sequences had to at least have four rows for the purpose of AB-HH where we will be looking at the history of hint use, three problems prior the next problem. Table 3 shows a sample scenario of a student sequence with the minimum of four rows. Student actions are denoted by either a 0 (attempt) or a 1 (hint). To get the first action on the next problem (FANP), we shift the values under FA one row up and copy the shifted values to the column FANP. Doing this, we have the FANP for Row 1 to be the FA in Row 2, FANP for Row 2 is the FA in Row 3, and so on, leaving FANP for the last row empty. The first previous question (P1) relative to the FANP is simply the FA. For the second previous question (P2), we do the reverse of how FANP was selected by shifting FA one row down. Doing this, we have P2 for Row 4 to be the FA in Row 3; P2 for Row 3 is the FA in Row 2, and so on, leaving P2 for the first row empty. Lastly, P3 is generated by shifting FA two rows down, leaving the first two rows under P3 empty. The rows with blanks are removed from analysis, thus, sequences with the minimum of four rows have one complete row for analysis (highlighted).

<table>
<thead>
<tr>
<th>Row #</th>
<th>FA</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>FANP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Sample student sequence scenario with four rows minimum (Note: FA = First action on the current problem, P3 = Third previous question, P2 = Second previous question, P1 = First previous question, FANP = First course of action on the next problem)

As was done in the initial analysis, we excluded scaffolding in the working dataset. The resulting dataset contained 279,925 problem entries with 555 unique problem sets and 12,429 unique students.

3. ACTION BEHAVIOR MODEL
Students take different initial courses of action in addressing tutor problems. In the case of the ASSISTments data, some students attempt by submitting an answer to a problem first, some ask for help (hint) first, some ask for hints after an initial attempt or alternate between attempting the problem and requesting for problem hints, and some continuously attempt a problem until a correct answer has been submitted or request for hints until the answer has been provided. This behavior has previously been observed in the work utilizing the SOA model [2]. Likewise, in ASSISTments, users cannot proceed to the next question without submitting a correct answer, so exhausting all available hints for a problem provides the student with the correct answer for the last hint. In ASSISTments terminology, this is referred to as bottom-out hinting.

Building upon this idea, we present two models that utilize different dataset features that leverage attempt and help-request behavior. For the dataset at hand, help-request is equivalent to requesting for a hint on the problem. The Action Behavior – Attempt/Hint Count (AB-AHC) model uses the number of attempts and number of hint requests employed by a student in working on a problem to predict the probability of hint request as the first course of action on the next problem (versus attempting to submit an answer for the problem first). The Action Behavior – Hint History (AB-HH) model makes use of students’ history of hint use in a number of problems prior the next problem to predict hint request as first action on the next problem. For the empirical procedures of this study, we arbitrarily selected to use the three previous problems before the next problem for hint use history, though our techniques can easily be extended to any number of previous problems as needed.

3.1 Initial Experiments: AB-AHC
We built a prediction table in which the number of attempts taken and the number of hints used by the student on a question were mapped to the probability that the student attempted or asked for a hint on the succeeding problem. The probability is computed by using the percentage of students who asked for a hint on the next problem. Matching a student’s Attempts Taken to Hints Taken maps the probability value of attempt/hint usage on the next problem to that student. Table 4 shows an example of a prediction table computed from training data. Table 5 shows a matching scenario snippet with AB-AHC using the prediction table in Table 4.

Tables 4 and 5 exhibits a scenario for problems with three available hints, that is, students have a maximum of three opportunities to ask for a hint on the current problem. A value in the Hints Taken row of Table 4 such as 2/3 indicate that a student used two out of the three available hints for the problem and values on the first column indicate the count of attempts taken.

To illustrate how prediction values in Table 4 are used, take, as example, the first row in Table 5. The row is a problem entry for student 92677. Because the student has an Attempt Count (A_C) of 1 and a Hint Count (H_C) of 0, we match the value under A_C to the corresponding value in the Attempts Taken column of Table 4, as well as match
the value under H_C to the corresponding value in Hints Taken; the value at the intersection of Attempts Taken and Hints Taken is the prediction of hint usage on the next problem, in this case, 0.0211. The same process of parameter table generation and prediction matching are employed for problems with 4 and 5 available hints.

<table>
<thead>
<tr>
<th>Attempts Taken</th>
<th>Hints Taken</th>
<th>0 / 3</th>
<th>1 / 3</th>
<th>2 / 3</th>
<th>3 / 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0211</td>
<td>0.1001</td>
<td>0.2213</td>
<td>0.4025</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0261</td>
<td>0.0558</td>
<td>0.0747</td>
<td>0.1105</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0237</td>
<td>0.0447</td>
<td>0.0737</td>
<td>0.0916</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0363</td>
<td>0.0287</td>
<td>0.0743</td>
<td>0.0949</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0132</td>
<td>0.0263</td>
<td>0.0857</td>
<td>0.0912</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. AB-AHC Prediction Table

To analyze whether the number of history points affected the predictive power of AB-HH, an additional analysis on history of hint use was done, with four problems prior the next problem. This analysis required selecting problem entries in the dataset where each student sequence had more than five rows, resulting with a dataset of 481 unique problem sets and 8128 unique students.

For example, 1H / 2A indicates that in the three prior problems, a total of 1 hint as first course of action and 2 attempts as first course of action were employed by the student. The counts of attempts and hints as first action were then generated for each column bin. In the table, we can see that for those who used a total of 2 hints and 1 attempt (2H / 1A) in the three previous problems, there are 3330 instances of attempts as first course of action on the next problem and 1833 instances of hint requests as first course of action on the next problem. The prediction values for each bin are found in the row % Hint, which indicates the percentage of instances of hint usage, within the bin.

A five-fold cross validation was likewise used to train and test the AB-HH model where the dataset was divided into five folds by problem set and student, for a problem set level and student level analysis.

<table>
<thead>
<tr>
<th>First Action Hints / Attempts in Previous 3 Problems</th>
<th>0H / 3A</th>
<th>1H / 2A</th>
<th>2H / 1A</th>
<th>3H / 0A</th>
</tr>
</thead>
<tbody>
<tr>
<td># Attempt</td>
<td>111017</td>
<td>17219</td>
<td>3330</td>
<td>683</td>
</tr>
<tr>
<td># Hint</td>
<td>5859</td>
<td>3254</td>
<td>1833</td>
<td>1663</td>
</tr>
<tr>
<td>% Hint</td>
<td>0.0501</td>
<td>0.1589</td>
<td>0.3550</td>
<td>0.7089</td>
</tr>
</tbody>
</table>

Table 6. AB-HH Prediction Table

For analysis using AB-HH, the prediction table was generated by using the percentage of hint use as first action in the three previous problems. Table 6 shows a prediction table computed from training data on hint history. Recalling Table 3, a row contains values for P3, P2, and P1, which will contain either a 0 (attempt) or 1 (hint). The column labels in Table 6 correspond to the number of times the first action was an attempt on the problem or a request for a hint.

3.2 Secondary Experiment: AB-HH

For analysis using AB-HH, the prediction table was generated by using the percentage of hint use as first action in the three previous problems. Table 6 shows a prediction table computed from training data on hint history. Recalling Table 3, a row contains values for P3, P2, and P1, which will contain either a 0 (attempt) or 1 (hint). The column labels in Table 6 correspond to the number of times the first action was an attempt on the problem or a request for a hint.

For example, 1H / 2A indicates that in the three prior problems, a total of 1 hint as first course of action and 2 attempts as first course of action were employed by the student. The counts of attempts and hints as first action were then generated for each column bin. In the table, we can see that for those who used a total of 2 hints and 1 attempt (2H / 1A) in the three previous problems, there are 3330 instances of attempts as first course of action on the next problem and 1833 instances of hint requests as first course of action on the next problem. The prediction values for each bin are found in the row % Hint, which indicates the percentage of instances of hint usage, within the bin.

A five-fold cross validation was likewise used to train and test the AB-HH model where the dataset was divided into five folds by problem set and student, for a problem set level and student level analysis.

To analyze whether the number of history points affected the predictive power of AB-HH, an additional analysis on history of hint use was done, with four problems prior the next problem. This analysis required selecting problem entries in the dataset where each student sequence had more than five rows, resulting with a dataset of 481 unique problem sets and 8128 unique students.

4. RESULTS AND DISCUSSION

The predictive performance of the AB-AHC and AB-HH models were evaluated using root mean squared error (RMSE), mean absolute error (MAE), and area under the ROC curve (AUC) metrics. Additionally, a more naïve majority class (MC) model was generated as an additional model for comparison, as we have found no other, gold standard model for first-course-of-action prediction to compare our work with. The majority class model is generated using the percentage of hint instances on the students’ second action on all problems in the dataset. Table 7 provides a sample scenario for majority class prediction computation.

As can be seen from Table 7, Hint % is the percentage of hint instances in the problem entries, which translates to a prediction on the students’ first course of action on the next problem. If a student’s second action on the current problem is a hint (1), the prediction for FANP is Hint %, otherwise, the prediction takes the value of Attempt %.
The simple intuition behind this is the hypothesis that students who have a greater tendency to ask for hints on succeeding actions may most likely ask for a hint in succeeding problems. Problem set-level five-fold cross validation was similarly performed to train and test the majority class model, for each dataset used. Thus, the number of unique problem sets in the datasets for the majority class is the same as the number of problem sets in the datasets used in the AB models. The RMSE, MAE, and AUC were likewise computed for majority class results.

<table>
<thead>
<tr>
<th>Problem entries</th>
<th>Hint Instances: 2nd Action</th>
<th>Hint % (MC)</th>
<th>Attempt %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2200</td>
<td>852</td>
<td>0.3872</td>
<td>0.6127</td>
</tr>
</tbody>
</table>

Table 7. Sample scenario for computing majority class prediction values (Note: MC = Majority Class)

4.1 AB-AHC Analysis

The problem set level findings for the AB-AHC model and the majority class model on the three problem groups (3, 4, and 5 available hints) are presented in Figure 1. It is clear from Figure 1 that the AB-AHC model consistently outperforms the majority class model across all three problem groups in predicting hint as first course of action in the next problem in both the RMSE and MAE metrics. Lower values for both metrics indicate better model fit. A reliability analysis was done to compare AB-AHC with the majority class model using a two-tailed paired t-test on the results from the problem set level five-fold cross validation and results indicate that the findings are reliably different across all problem groups with p=0.

The effectiveness of the AB-AHC model can likewise be gleaned from the results using the AUC metric (Figure 2). AUC values closer to one indicate better model fit.

In addition, it can be noted that the performance of AB-AHC in terms of RMSE, MAE, and AUC are closely consistent, suggesting that the model is fairly generalizable across problems with varying numbers of hint availability. Likewise, for all of the metrics, the predictive performance of the majority class model decreases as the number of available hints for problems increases.

Performing a student level analysis on both models yields the results found in Figures 3 (RMSE, MAE) and 4 (AUC). As observed, predictive performance over a student level analysis of the AB-AHC model for problems with 4 and 5 available hints seem fairly consistent across all three metrics, however, the model does not perform as well for problems with 3 available hints. This suggests that AB-AHC may be used to predict the hint request behavior of unseen students, provided that there is a high number of opportunities to request for help. Similar to the results in the problem set level analysis, MC model performance fails to improve as the number of available hints increase.
4.2 AB-HH Analysis

This section discusses results for the AB-HH model and the majority class model, for both 3 and 4 prior problem history of first action. Comparisons of model performance for problem set level and student level analysis between each number of prior problem history points, as well as problem set and student level within number of prior problem history points will be discussed in the following subsections.

4.2.1 AB-HH 3 vs. AB-HH 4

A problem set level analysis of the AB-HH model and the MC model across the number of prior history points (Figures 5a and 5b) demonstrates that the MC model performance improves as more first action history points is considered. The AB-HH model, however, maintains a fairly consistent level of predictive performance across all three metrics, and while AB-HH significantly outperforms MC in MAE and RMSE, it is outperformed by the latter in AUC for number of history points = 4. This may be because the ordering of the values in MC’s predictions is not as close to the actual as those of AB-HH. This situation rarely happens, and hence we may have to try another dataset to confirm this behavior.

Shifting to a student level analysis (Figures 6a and 6b), the predictive performance for both models exhibit a similar behavior as with the problem set level analysis, with MC improving in performance with an increased number of first action prior history points and AB-HH maintaining fairly stable and good predictive performance and, additionally, outperforming MC, across all values of first action prior history points. A reliability analysis was done to compare AB-HH with the MC model using a two-tailed paired t-test on the results from the problem set and student level five-fold cross validation and results indicate that the findings are reliably different across all prior hint history with p=0.

4.2.2 AB-HH/MC Problem Set Level vs. AB-HH/MC Student Level Within Prior History Points

Comparing within prior history points (i.e. 3 and 4 prior history problem set level vs. 3 and 4 prior history student level), it can be observed from Figure 7 that AB-HH manifests consistency of results for all performance metrics, while the MC exhibits more prominent fluctuation in its results across the three metrics. This suggests that the AB-HH model can feasibly be used to predict student hint request behavior for both unseen skills and unseen students with fair reliability. The data points presented in Figure 7 are the same values that can be found in the bar graphs in Figures 5 and 6, represented as combined scatter charts to exemplify the consistency in performance of the AB-HH model across problem set level and student level analyses.

The results for both analyses in Sections 4.2.1 and 4.2.2 are indicative of the features employed in AB-HH as reasonably good features that can be used for forecasting the first course of action of students in succeeding problems, as evidenced by its consistent performance across problem set level and student level analysis, as well as across number of first action history points.
a. Problem set level RMSE and MAE for AB-HH and MC across number of prior problem history points

b. Problem set level AUC for AB-HH and MC across number of prior problem history points

Figure 5. Problem set level performance metrics for AB-HH and MC across number of prior problem history points
(Note: Sk = Problem set level, St = Student level)

a. Student level RMSE and MAE for AB-HH and MC across number of prior problem history points

b. Student level AUC for AB-HH and MC across number of prior problem history points

Figure 6. Student level performance metrics for AB-HH and MC across number of prior problem history points
(Note: Sk = Problem set level, St = Student level)
5. CONTRIBUTION AND FUTURE WORK

The analyses of the results of our empirical studies suggest that students’ help request behavior can indeed be feasibly predicted from data features that are descriptive of student action information, in this case, data on student attempt and hint requests. While the methods employed in this study represent a starting point in utilizing action information, we feel that such initiatives are worth discussing for building up further studies in the field. The study addresses the research questions we have presented:

1. Can we determine students’ need for assistance ahead of time?

The methods provide utility for predicting student assistance need, defined here as a request for hints to aid in addressing tutoring system problems, using dataset features of action information. Both AB models were able to predict students’ first course of action when answering problems from an intelligent tutoring system with fairly consistent predictive performance and generalizability across problems with varying numbers of hint availability (for AB-AHC) and across unseen problem sets and students (AB-AHC and AB-HH).

2. What dataset information may be useful for developing or engineering models that forecast student assistance need?

The methods we have presented utilized the dataset features of information on problem attempts and help requests employed by students. AB-AHC makes use of information on the number of attempts and hints used by students on a question to predict the instance of help request as the first course of action on the next problem with successful results for problems with varying hint availability. AB-HH makes use of history information on the number of hints requested as the first course of action in immediately preceding questions. Both models were able to predict students’ first course of action on succeeding problems with reasonable accuracy and consistent performance.

Future work and improvements to these models may include the accounting of patterns in action behavior for problems, which has been partially explored in [2]. While our models made use of similar information, these only made use of the number of instances of when attempts or help requests were performed on problems. The pattern of actions used by students provides a rich source of information for possible prediction of when a need for assistance is exhibited by students (likewise partly explored here with the majority class model used for comparison). A similarly interesting extension is the possible consideration of learner affect when attempting or requesting for help. The dataset used for analysis contained other pieces of information including student response times and skill difficulty. As what has been employed in our models, exploiting these features may provide further insight into factors that facilitate assistance need and thus aid in developing a proactive and effective early intervention framework.

An additional future work would be to try these models on datasets from ITSs other than ASSISTments. This way we can determine whether these models are consistent across different datasets.
REFERENCES