Modeling learning patterns of students with a tutoring system using Hidden Markov Models.

Conference Paper · January 2007

Source: DBLP

3 authors:

Carole R. Beal
University of Pittsburgh
141 PUBLICATIONS 2,450 CITATIONS

SEE PROFILE

Sinjini Mitra
California State University, Fullerton
66 PUBLICATIONS 424 CITATIONS

SEE PROFILE

Paul R. Cohen
The University of Arizona
388 PUBLICATIONS 6,853 CITATIONS

SEE PROFILE

Modeling learning patterns of students with a tutoring system using Hidden Markov Models

Carole Beal [a,1], Sinjini Mitra [a] and Paul R. Cohen [a]
[a] *USC Information Sciences Institute*

Abstract. The current paper focuses on modeling actions of high school students with a mathematics tutoring system with Hidden Markov Models. The results indicated that including a hidden state estimate of learner engagement increased the accuracy and predictive power of the models, both within and across tutoring sessions. Groups of students with distinct engagement trajectories were identified, and findings were replicated in two independent samples. These results suggest that modeling learner engagement may help to increase the effectiveness of intelligent tutoring systems since it was observed that engagement trajectories were not predicted by prior math achievement of students.

Keywords. clustering, Hidden Markov Model, inference, intelligent tutoring system, learning action patterns, learner engagement, prediction, transition matrices

1. Introduction

Intelligent tutoring systems (ITS) research is one of the "success stories" of artificial intelligence: A number of studies have demonstrated that a detailed model of the learner's domain knowledge can be used to provide individualized instruction, and to accelerate student learning well above what would be expected in a whole-class context (Corbett & Anderson, 1995; Koedinger et al., 2000). Traditionally, tutoring systems researchers have focused primarily on modeling the learner's cognitive processes while solving problems, i.e., the "model tracing" approach. However, there is growing recognition that learning involves more than cognition, and that students' actions with an ITS also reflect "engagement", meaning the transient shifts in attention and the emotions that are often associated with learning (Qu & Johnson, 2005; Vicente & Pain, 2002). For example, a student may become bored or fatigued over the course of a session and deliberately enter incorrect answers in order to elicit the correct answer from the ITS (Baker et al., 2005).

Including estimates of engagement in learner models could be used to improve the effectiveness of tutoring systems, for example, by shifting the type of problem being delivered if boredom is detected, or by providing supportive feedback if the student seems frustrated. However, relatively little is known about modeling learners' engagement while using an ITS. One approach is to use sensors to capture physiological indices

[1]Correspondence to: Carole Beal, USC Information Sciences Institute, 4676 Admiralty Way, Suite 1001, Marina del Rey, CA 92647. Tel.: +1 310 448 8755; Fax: +1 310 822 0751; E-mail: cbeal@isi.edu.

of emotional states such as interest, boredom, and frustration (D'Mello et al., 2004). Although very promising in the laboratory, this approach is difficult to scale for use by large numbers of students in real classroom situations in which intrusive and fragile equipment cannot easily be used. Here, we investigate an alternative approach: model students' action traces using a hidden variable denoting their engagement level as they work with an ITS (Beck, 2005).

The rest of the paper is organized as follows. Section 2 describes the data that we use for our experiments. The statistical model is introduced in Section 3, which also includes results and analysis. Conclusions and future work are presented in Section 4.

2. Data Sources

Our analyses focused on records of student actions with a tutoring system for high school mathematics. Data from two independent samples were available: the "CA" sample included 91 students who completed an average of 20 problems, and the "BA" sample included 122 students who completed an average of 30 math problems. The BA sample was more heavily weighted towards lower-achieving students, whereas the CA sample included a broader range of achievement levels. CA students attempted a maximum of 5 sessions and BA students a maximum of 3 sessions.

In both samples, the data were automatically logged as students worked with the mathematics tutoring system during their regular school instruction. Students worked on a series of math problems, each including a figure, graph or table; the problem text and equation; and five answer options. On each problem, the student could choose answers and receive feedback on accuracy (e.g., a click on an incorrect answer elicited a red "X" whereas the correct answer elicited a green checkmark); request a series of multimedia hints leading to the solution; or move on to the next problem without answering.

Prior work had shown that a student's actions associated with a math problem could be machine-classified into five categories: *Guessing/Help Abuse*; *Independent-Accurate problem solving*; *Independent-Inaccurate problem solving*; *Learning with Help*; and *Skipping*, with 95% accuracy (Beal et al., 2006). In the present study, the action and latency data recorded on each math problem were similarly machine classified into 5 action patterns. Thus, each student's data record included an ordered sequence of action patterns representing her or his behavior on the math problems over the course of the tutoring session. An example of a student's data is shown in Table 1. The corresponding action sequence is [AAECBD...].

Table 1. Sample action pattern sequence.

Problems	1	2	3	4	5	6	...
Action Patterns	Guess	Guess	Skip	Ind-accurate	Learn	Ind-inaccurate	...
Code	A	A	E	C	B	D	...

3. Statistical Model and Results

We model the performance of students with *Hidden Markov Models* (HMM). Although HMMs have been applied extensively in speech processing (Rabiner, 1989) and video-based face identification (Liu & Chen, 2003), the application of HMMs to tutoring systems is relatively new, particularly with regard to modeling the behavior of individual students rather than groups (Stevens et al., 2005). We fit HMMs to the sequence of actions emitted by individual students, with 3 hypothesized hidden states corresponding to a student's "engagement level" $-$ *low, medium* and *high*, that we assume influences student interactions with the ITS. Having fit an HMM to students' action pattern data, the transition matrices help us determine their probabilities of moving from one level of engagement to another, as well as of remaining in a particular engagement state while attempting a series of problems. In other words, if a student's engagement level is known at time t, the transition matrix will tell us his/her likely engagement level at time $t + 1$.

The average transition matrices over all students in each sample from fitting HMMs to the actions patterns of students in Session 1 only are shown in Table 2. The rows represent time t, and the columns represent time $t+1$, so that a particular entry represents the probability of transitioning from a particular engagement level at time t to another one at time $t + 1$. For example, CA students who are in a low engagement state have a 45.63% chance of moving into a high engagement state at the next time step. The diagonal entries represent the probabilities of remaining in the same engagement level from one time step to the next. For example, a CA student has a 30.14% chance of remaining in the low engagement state at time $t + 1$ if he is in such a state at time t.

Table 2. Average transition matrices for CA and the BA samples.

Sample	Hidden state	Low	Medium	High
CA	Low	0.3014	0.2423	0.4563
	Medium	0.1982	0.4721	0.4176
	High	0.1517	0.3312	0.6050
BA	Low	0.4727	0.1966	0.3307
	Medium	0.1720	0.5583	0.2696
	High	0.1470	0.2772	0.5758

Both transition matrices suggest that there is a degree of "inertia" in students' actions, i.e., the probabilities for persisting in the same state are generally higher than those of shifting to another state. For example, in both samples, the highest probabilities are observed for students who are likely to persist in that state. However, there is considerable individual variation in the engagement trajectories (not shown due to space constraints). We address this issue in the next section.

3.1. Clustering Students based on HMM

In order to account for the variance in the individual transition matrices, we clustered students in each dataset based on the Kullback-Leibler (KL) distance between the individual transition matrices (Ramoni, Sebastiani & Cohen, 2001). Such groups represent students who follow similar trajectories through the different levels of engagement during an ITS session. We obtained 3 groups for the CA sample, and 4 groups for the BA sample. The average transition matrix for each group appear in Tables 3 and 4 respectively.

We see that there was considerable similarity in the clustered HMM groups across the two samples. More specifically, both samples include one fairly large cluster of students who tended to persist in medium and high engagement states $(50 - 65\%)$ ("steady

Table 3. Group transition matrices for 91 CA students. N denotes the number in each group.

Group	Hidden state	Low	Medium	High
1	Low	0.3641	0.1008	0.5351
N=51	Medium	0.2105	0.5801	0.2094
56%	High	0.1957	0.2928	0.5115
2	Low	0.2606	0.4899	0.2495
N=34	Medium	0.0384	0.3933	0.8036
37%	High	0.1124	0.4086	0.7143
3	Low	0.0000	0.0417	0.9583
N=6	Medium	1.0000	0.0000	0.0000
7%	High	0.0000	0.2189	0.7811

Table 4. Group transition matrices for 122 BA students. N denotes the number in each group.

Group	Hidden state	Low	Medium	High
1	Low	0.5056	0.1089	0.3855
N=81	Medium	0.1392	0.6637	0.1971
66%	High	0.0970	0.2201	0.6830
2	Low	0.2864	0.6809	0.0327
N=20	Medium	0.0260	0.2839	0.6901
17%	High	0.2970	0.1455	0.5574
3	Low	0.1735	0.1429	0.6837
N=7	Medium	0.9429	0.0000	0.0571
6%	High	0.0000	0.1805	0.8105
4	Low	0.7105	0.0344	0.2551
N=14	Medium	0.1648	0.6152	0.2201
11%	High	0.3064	0.6936	0.0000

state"). The probability of moving to the low engagement state is only about 20% or less for these students. A second group with a distinctly different profile is also observed in both samples: these students tend to become more engaged over time ("increasing engagement"), e.g., from medium to high engagement state (80% for CA, 69% for BA), and are not likely to move to a low engagement state. Both samples also include a small third group characterized by strong shifts in engagement ("fluctuating engagement"). One shift is from low to high engagement (95% for CA, 68% for BA), and another from medium to low engagement (100% for CA, 94% for BA). Finally, the BA sample includes a fourth group ("declining engagement") characterized by persisting in a low engagement state (71%) and moving from high to medium state (69%). Recall that the BA sample included more students with low achievement.

3.1.1. Behavior of students in HMM groups

In order to compare the behavior of students in the HMM clusters, we looked at the mean proportion scores for the 4 primary action patterns (rates for "Problem Skipping" were very low and are not analyzed here). The results were again fairly consistent across the two samples. For the CA sample, an analysis of variance (ANOVA) with HMM group as the grouping factor and proportion of actions classified as Guessing as the dependent measure showed a main effect of HMM Group, $F(2, 97) = 4.346$, $p < .05$, which implied that the amount of guessing varied significantly among the different groups. Tukey's pairwise comparison tests further indicated that Groups 1 and 3 had a significantly higher proportion of Guessing than Group 2. For the BA sample we observed $F(3, 109) = 9.436,\ p < .01$. Group 4 has significantly high Guessing scores than the other three groups and Groups 2 and 3 had significantly lower scores than the other two.

With regard to the effective use of multimedia help (Learn), there were no significant differences between the HMM clusters for the CA sample. For the BA sample, ANOVA showed that the 4 groups of BA students varied considerably with respect to the amount of learning ($F(3, 109) = 4.968,\ p < .01$). Tukey's tests indicated that the means for Groups 1, 2, 3 were similar to each other, but significantly higher than the Group 4 mean.

With regard to Independent-Inaccurate problem solving, there was a main effect of HMM Group for the CA sample, $F(2, 97) = 3.784, p < .05$ (although Tukey's test did not reveal any significant differences in the group means), but none for the BA sample.

ANOVA conducted on the proportion scores for Independent-Accurate problem solving showed no difference for the HMM groups for either sample. However, as shown in Table 5, mean scores for the CA sample were generally higher than for the BA sample, consistent with the fact that the BA sample included more low-achieving students.

Table 5. Mean Proportion scores for the HMM groups in the 2 samples

HMM Group	CA1	CA2	CA3	BA1	BA2	BA3	BA4
Guess	0.21	0.10	0.19	0.22	0.08	0.09	0.46
Learn	0.20	0.23	0.30	0.25	0.39	0.42	0.13
Ind. Accurate	0.27	0.39	0.39	0.19	0.20	0.22	0.10
Ind. Inaccurate	0.24	0.18	0.09	0.20	0.21	0.16	0.18

3.1.2. HMM clusters and math achievement

It may be possible that the HMM groups simply capture aspects of student engagement with the ITS which can be directly traced to their mathematics proficiency. For example, students who do not have very good math skills might be more likely to view the multimedia help features; students with strong math ability might be more likely to solve problems accurately and independently. This plausible interpretation would be consistent with the model-tracing view of student learning, i.e., that students' actions result from their cognitive understanding of the domain. However, in the present study we did not find a relation between students' math achievement and their HMM group membership, as explained below.

In the CA sample, the students' classroom teachers had provided independent ratings of each student's math proficiency before the start of the study, based on their knowledge of the students' performance on homework, tests and exams. A chi-square analysis indicated that there was no significant association between the students' HMM cluster membership and their ratings on math achievement. In the BA sample, students had completed a 42-item pre-test of math skills before starting to work with the ITS. A logistic regression analysis showed that there was no association between students' pre-test scores and HMM group membership. This suggests that math proficiency does not completely explain students' behavior with the ITS, as observed in two independent samples with different types of measurement. This is consistent with the view that ITS student models may be enhanced with estimates of engagement as well as domain knowledge.

3.2. Prediction based on HMM

In this section, we describe additional work designed to predict a student's likely engagement trajectory, with the goal of diagnosing the student's behavior early enough that interventions could be deployed to increase or sustain engagement. First of all, we used the fitted HMM for each individual student estimated from the action patterns on the first 15 problems in Session 1 to predict the action pattern at the next step for problems $M = 16, \ldots, S$ (where S is the number of problems attempted by each student). The results in Table 6 show that the prediction accuracy was 42.13% for the CA sample, and 48.35% for the BA sample. This is well above chance or random guessing (25%), which denotes that a student is likely to take any of the 4 primary actions at each time (i.e., assuming no prior knowledge about the system). This indicates that HMMs can be useful in predicting students' future behavior with the tutoring system beyond mere random guessing. Next we used the group HMMs to perform prediction by using the group transition matrix and group emission matrix corresponding to each student. Not surprisingly, the group HMM accuracy was poorer than the individual HMMs, yet they were more accurate than chance.

On comparing the prediction accuracies with those obtained from simple Markov Chain (MC) models which do not assume the existence of a hidden state, we find that the individual MCs performed less well than the individual HMMs, although still better than chance. Moreover, the group HMM models performed better than the individual MCs for the BA sample and did not differ significantly for the CA sample. Noticeably, the predictive power of the group MCs is the poorest. Thus, we conclude that the HMMs with the hidden state can model the variability in students' behavior during a problem-solving session more effectively than the MCs.

Table 6. Prediction accuracies for HMMs and MCs on Session 1.

	Ind. HMM	Ind. MC	Gr. HMM	Gr. MC
CA	42.13%	33.43%	34.40%	18.73%
BA	48.35%	32.48%	26.00%	15.52%

3.2.1. Predicting Future Sessions

The next set of experiments are based on students who have two sessions with the ITS – 10 CA students and 25 BA students. Here, for each student, we train the HMM based on all the problems he or she attempts in Session 1. We then use the second sessions of these students for validating our models using the fitted HMMs (both individual and group models) from the first sessions. We generate predictions for all the problems attempted in Session 2 for each of these students. We again compare our prediction results from the HMMs to those from MC models that do not take into account any hidden variable information. Table 7 shows these prediction accuracies. The highest accuracies for multiple sessions are obtained with the individual HMMs, followed by the group HMMs. Both are significantly better than pure random guessing or chance (25% for 4 possible actions), whereas the accuracies of the MCs are worse than chance. Note that, predictive abilities on future session are lower than that on the same session, as expected. This is because often the dynamics of student behavior are more stable over a session than across sessions.

Table 7. Prediction accuracies for Session 2.

	Ind. HMM	Ind. MC	Group HMM
CA	45.93%	10.15%	30.12%
BA	36.50%	12.53%	25.22%

3.2.2. Qualitative Assessment of Prediction Accuracies

The above results indicate that HMMs fitted to students' actions in one session will predict their actions in subsequent sessions with about 40% accuracy. This is certainly not perfect, but it is well above knowing nothing at all about the student in terms of designing interventions to sustain engagement, especially if we could distinguish those students for whom we could predict engagement on future sessions from those where our predictions are not likely to be accurate. Therefore, we next investigated if the prediction accuracy on the first session could signal the prediction accuracy on a future session using the individual HMMs that had the best accuracies. Both samples show significant positive correlations (0.8851 for CA and 0.7984 for BA) between the prediction accuracies from the 2 sessions. Thus, we conclude that if the HMMs are good at predicting the first session, they are likely to have high prediction accuracies on subsequent sessions as well. On the other hand, if an individual's HMM is not good at predicting the first session, it is unlikely to be very useful for predicting engagement trajectory in future sessions.

Comparing these individual prediction accuracies with the distributions of the action patterns on the two sessions, we find that students with similar prediction accuracies for both sessions typically have low entropy between the two distributions (measured by KL distance), and vice versa for students who have poor accuracies on both sessions. There are, however, a few students with low accuracies on Session 2 despite having medium accuracies on Session 1. This happens because these students miss certain action patterns on Session 1 that appear frequently on Session 2, thus leading to poor predictions.

4. Conclusions and Future Work

Our primary goal in this study was to use a student's actions with a tutoring system to estimate his or her engagement: an unobservable state that is hypothesized to influence the observed actions emitted by the student. Hidden Markov Models are an ideal analytic approach because these models allow for explicitly modeling unobservable influences. Here, we fit HMMs to students' actions with a math tutoring system, with a three-level hidden state corresponding to "engagement". The results indicated that students' action sequences could be modeled well with HMMs, that their actions at one point in time could be successfully used to predict the subsequent action, and that the prediction accuracy of these models were superior to simple Markov Chain models that did not include a hidden state. We also found that, in many cases, the HMM models from Session 1 could be used to predict students' future engagement trajectories on Session 2.

By clustering the HMMs, we also identified groups of students with distinct trajectories through the hidden state space. Similar groups were observed in two independent samples from studies conducted in different school systems and geographical locations, suggesting that the patterns may be fairly general and consistent. The largest group included students who were generally engaged with the tutoring system, and whose engagement tended to remain stable over the course of the session. A second group was also quite engaged but showed a stronger tendency to become more highly engaged over time. One interpretation is that students in these groups do not really require any intervention from the ITS to support their learning. Thus this facilitates evaluation of the impact of interventions by looking for shifts in a student's HMM cluster, and hence design more effective ITS. As a next step, we will use more refined models, such as non-stationary HMMs to explicitly model the amount of time a student spends on a problem.

The primary limitation of the study is that we do not have direct evidence that the hidden state in the HMMs actually corresponds to any of the processes that are known to reflect "engagement" in the learner, including transient shifts in the learner's attention, emotions and cognitive effort. But we still chose to term the hidden state "engagement" because we believe it to influence a student's behavior with a math tutoring session the most. However, regardless of the true nature of the hidden state, HMMs have proved to be an efficient modeling tool for students' action patterns.

Acknowledgment: The research was supported by the National Science Foundation (REC 0411886). The views expressed are not necessarily those of the funding agency. We would also like to thank the participating schools and project members Erin Shaw and Lei Qu.

References

[1] R.S. Baker, A.T. Corbett, K.R. Koedinger and I. Roll (2005). Detecting when students game the system, across tutor subjects and classroom cohorts, In L. Ardissono, P. Brna, & A. Mitrovic (Eds.), *User Modeling*, 220-224. Berlin: Springer-Verlag.

[2] C.R. Beal, L. Qu and H. Lee (2006). Classifying learner engagement through integration of multiple data sources. *Proceedings of AAAI'06*, Boston, MA.

[3] J.E. Beck (2005). Engagement tracing: Using response times to model student disengagement. In C-K. Looi et al. (Eds.), *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology*, 88-95. Amsterdam: IOS Press.

[4] A.T. Corbett and J.R. Anderson (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User Adapted Interaction*, **4**: 253-278.

[5] S.K. D'Mello, S.D. Craig, B. Gholson, S. Franklin, R. Picard and A.C. Graesser (2004). Integrating affect sensors into an intelligent tutoring system. In *Affective Interactions: The Computer in the Affective Loop. Proceedings of the 2005 International Conference on Intelligent User Interfaces*. 7-13. New York.

[6] K.R. Koedinger, A.T. Corbett, S. Ritter and L.J. Shapiro (2000). Carnegie Learning's Cognitive Tutor: Summary of research results. *White Paper, Carnegie Learning*. Pittsburgh PA.

[7] X. Liu and T. Chen (2003). Video-based face recognition using adaptive Hidden Markov Models. *Proceedings of CVPR'03*, Madison, Wisconsin.

[8] L.R. Rabiner (1989). A tutorial on HMM and selected applications in speech recognition. *In Proceedings of IEEE*, **77**(2):257-286.

[9] L. Qu and W.L. Johnson (2005). Detecting the learner's motivational states in an interactive learning environment. In C-K. Looi et al. (Eds.), *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology*, 547-554. IOS Press.

[10] M. Ramoni, P. Sebastiani and P.R. Cohen (2001). Bayesian clustering by dynamics. *Machine Learning*, **47**:91-121.

[11] R. Stevens, D. Johnson and A. Soller (2005). Probabilities and prediction: Modeling the development of scientific problem solving skills. *Cell Biology Education*, **4**:42-57. Springer.

[12] A. Vicente and H. Pain (2002). Informing the detection of the student's motivational state: An empirical study. *In 6th International Conference on Intelligent Tutoring Systems*, 933-943. Biarritz, France.