A Dynamic Mixture Model to Detect Student Motivation and Proficiency

Jeff Johns and **Beverly Woolf**
Computer Science Department
University of Massachusetts Amherst
Amherst, Massachusetts 01003
{johns, bev}@cs.umass.edu

Abstract

Unmotivated students do not reap the full rewards of using a computer-based intelligent tutoring system. Detection of improper behavior is thus an important component of an online student model. To meet this challenge, we present a dynamic mixture model based on Item Response Theory. This model, which simultaneously estimates a student's proficiency and changing motivation level, was tested with data of high school students using a geometry tutoring system. By accounting for student motivation, the dynamic mixture model can more accurately estimate proficiency and the probability of a correct response. The model's generality is an added benefit, making it applicable to many intelligent tutoring systems as well as other domains.

Introduction

An important aspect of any computer-based intelligent tutoring system (ITS) is the ability to determine a student's skill set and to tailor its pedagogical actions to address the student's deficiencies. Tutoring systems have demonstrated this ability in the classroom (VanLehn *et al.* 2005). However, even the most effective tutoring system will fail if the student is not receptive to the material being presented. Lack of motivation has been shown empirically to correlate with a decrease in learning rate (Baker, Corbett, & Koedinger 2004). While attempts to motivate a student by using multimedia and/or by couching the material as a game have proved partially successful, there is still significant room for improvement. In fact, these motivation tools can themselves cause undesirable behavior, where students uncover ways to game the system. This issue of motivation and performance is particularly relevant given the weight assigned to high stakes achievement tests, such as the Scholastic Aptitude Test (SAT), as well as other exams that can be required for graduation. Students use tutoring systems to practice for high-stakes tests but typically are not graded based on their performance, which leads to low effort. The concern of low motivation affecting performance is also being addressed by the educational assessment community (Wise & DeMars 2005).

Automated diagnosis is the first step in addressing a student's level of motivation. Several models have been proposed to infer a student's engagement using variables such as observed system use (time to respond, number of hints requested), general computer use (opening an Internet browser, mouse activity), and visual and auditory clues (talking to the person at the nearby computer). The purpose of this paper is not to point out new ways in which students display unmotivated behavior, but rather to provide a general, statistical framework for estimating both motivation and proficiency in a unified model provided that unmotivated behavior can be identified. We propose combining an Item Response Theory (IRT) model to gauge student proficiency and a hidden Markov model (HMM) to infer a student's motivation. The result is a very general, dynamic mixture model whose parameters can be estimated from student log data and can run online by an ITS. We validate the model using data from a tutoring system, but indicate the model can be applied to other types of data sets.

Background

The next two sections describe previous work in estimating motivation and provide a brief introduction to Item Response Theory.

Relevant Literature

Several models have been proposed to infer student motivation from behavioral measures. de Vicente and Pain (2000; 2002) designed a study where participants were asked to watch prerecorded screen interaction of students using their tutoring system. Participants in the study created over eighty inference rules linking motivation to variables that could be observed on the screen. The fact that so many rules could be derived purely from screen shots suggests that simple, inexpensive methods for estimating motivation should be useful for a tutoring system.

Conati (2002) presented a dynamic decision network to measure a student's emotional state based on variables such as heart rate, skin conductance, and eyebrow position (in contrast to the more easily attained data used by de Vicente and Pain). The structure and parameters of the model, in the form of prior and conditional probabilities, were set by hand and not estimated from data. The probabilistic model applies decision theory to choose the optimal tutor action to balance motivation and the student's learning.

A latent response model (Baker, Corbett, & Koedinger 2004) was learned to classify student actions as either gaming or not gaming the system. Furthermore, instances of gaming the system were divided into two cases: gaming with no impact on pretest-posttest gain and gaming with a negative impact on pretest-posttest gain. The features used in the latent response model were a student's actions in the tutor, such as response time, and probabilistic information regarding a student's latent skills.

Arroyo and Woolf (2005) developed a Bayesian network using a student's observed problem-solving behavior and unobserved attitude toward the tutor. The unobserved variables were estimated from a survey that students filled out after using the tutor. Correlation between pairs of variables was used to determine the network's connectivity.

Beck (2005) proposed a function relating response time to the probability of a correct response to model student disengagement in a reading tutor. He adapted the item characteristic curve from IRT to include a student's speed, proficiency, response time, and other problem-specific parameters. The learned model showed that disengagement negatively correlated with performance gain.

These models embody different assumptions about the variables required to estimate student motivation (e.g. static versus dynamic models, complex versus simple features, user specified versus learned model parameters, generic versus domain specific models). The model proposed in this paper is different because it encompasses the following four principles which do not all exist in any one of the previous models. First, the model should estimate both student motivation and proficiency. These variables need to be jointly estimated because poor performance could be due to either low motivation or insufficient ability. Only one of the aforementioned models performs this function (Beck 2005). Second, the proposed model should run in real time. There exists a tradeoff between model complexity and expressiveness to ensure tutoring systems can take action at the appropriate time. Ideally, the model parameters should also be estimable from a reasonable amount of data. Third, the model should be flexible enough to easily include other forms of unmotivated behavior as researchers identify them. Fourth, motivation needs to be treated as a dynamic variable in the model. Empirical evidence suggests that a student's motivation level tends to go in spurts. For example, Table 1 shows actual performance data (initial response time, total time to click the correct answer, and number of incorrect guesses) of a single student doing multiple-choice geometry problems. The problems are not arranged according to difficulty; therefore, the obvious shift in the student's behavior after the seventh problem could be attributed to a change in motivation.

Problem	Initial Time (s)	Total Time (s)	Number Incorrect
1	40	40	0
2	44	44	0
3	13	13	0
4	19	19	0
5	7	9	4
6	22	22	0
7	35	35	0
8	2	3	2
9	2	2	0
10	3	4	1
11	2	4	4
12	2	3	3

Table 1: Data from a single student using the geometry tutor. Notice the change in behavior after the first seven problems

Item Response Theory

IRT models were developed by psychometricians to examine test behavior at the problem level (van der Linden & Hambleton 1997). This granularity is in contrast to previous work that examined behavior at the aggregate level of test scores. While IRT models encompass a wide variety of test formats, we focus in this paper on IRT models for dichotomous user responses (correct or incorrect).

Item Response Theory posits a static, generative model that relates a student's ability, θ, to his/her performance on a given problem, U_i, via a nonlinear characteristic curve, $f(U_i|\theta)$. IRT models are data-centric models (Mayo & Mitrovic 2001) because they do not presuppose a decomposition of problems into separate, required skills. Each problem in an IRT model is assumed independent of the other problems. The random variable θ is drawn from a normal distribution with a specified mean and variance. The random variables associated with each problem, U_i, come from a Bernoulli distribution with the probability of a correct response given by the following parameterized function (Equation 1, Figure 1).

$$P(U_i = correct|\theta) = c_i + \frac{1 - c_i}{1 + \exp(-a_i(\theta - b_i))} \quad (1)$$

This is referred to as the three-parameter logistic equation, where a_i is the discrimination parameter that affects the slope of the curve, b_i is the difficulty parameter that affects the location, and c_i is the pseudo-guessing parameter that affects the lower asymptote. Note that the two-parameter logistic equation is a special case of the three-parameter equation where c_i is set to zero. Consistent and efficient methods exist for estimating these parameters. A more thorough description of the IRT model, it's properties, and the role of each of the parameters can be found in any text on the subject (Baker & Kim 2004; van der Linden & Hambleton 1997).

Model

We propose a dynamic mixture model based on Item Response Theory (DMM-IRT). The probabilistic model consists of four types of random variables: student proficiency, motivation, evidence of motivation, and a student's response to a problem.

The latent variables in the student model correspond to proficiency (θ) and motivation (M_i). Proficiency is defined to be a static variable (note, if statistical estimates of proficiency are made online while a student uses the tutor, then each new data point causes the estimate to change, but the