Cognitive Modeling to Represent Growth (Learning) Using Markov Decision Processes

RUSSELL G. ALMOND*

Educational Testing Service, Princeton, NJ, USA

Over time, teachers collect a great deal of information about each student. Integrating that information intelligently requires models for how the students’ proficiency changes over time. Armed with such models, teachers can filter the data—more accurately estimate the student’s current proficiency levels—and forecast the student’s future proficiency level. Furthermore, partially observed Markov decision processes (POMDPs) and recently developed computer algorithms can help tutors create strategies for student achievement and identify at-risk students. Implementing this vision requires models for how instructional actions change student proficiencies. This paper introduces a general action model (also called a bowtie model) that separately models the factors contributing to the success or effectiveness of an action and the proficiency growth under success and failure. The paper shows how the general action model expresses prerequisites and changes in action effectiveness due to Vygotsky’s zone of proximal development.

Keywords: Markov Decision Processes, Growth Models, Prerequisites, Zone of Proximal Development, Stochastic Processes.

1. INTRODUCTION

Black and Wiliam (1998) note that in many studies where assessment was used to inform instructional decision making there was a measurable improvement in student learning. To reason about the impact of assessment on instruction choice—or the impact of instruction on assessing student
proficiency—a computer requires a mathematical model that supports reasoning about the evidence provided by assessment for a given proficiency state and the effects of instruction on that proficiency state. While there are many mathematical models for assessments, and some of them even have parameters for the effects of instruction, few of them support planning: defining a set of rules, a policy, for selecting the next instructional action.

This paper proposes a general framework based on decision analysis (Section 2) that supports planning (MDP; Section 2.2). The framework can be divided into a cross-sectional piece which describes how observations from assessments inform current proficiency estimates and a longitudinal piece which describes the effect of instruction. This paper relies on the evidence-centered assessment design (ECD; Mislevy, Steinberg, & Almond, 2003) framework for the cross-sectional piece and Section 3 introduces a general action or bowtie model for the longitudinal piece.

In addition to its use in planning—using assessment to inform instructional decision making—this framework also allows the computer to sensibly integrate information from past observations about the student. In a typical tutoring regimen, the tutor gives a series of short benchmark assessments to track the students’ progress. The framework described here allows that information to be used for filtering—using past observations to improve estimation of current proficiency levels—and forecasting—predicting future proficiency levels.

2. A GENERAL FRAMEWORK FOR TEMPORAL MODELS

In a typical tutoring regimen, tutor and student meet at regular intervals. At each meeting, the tutor assesses the current state of the student and then chooses an action to take, usually an activity for the student to do between now and the next meeting. Section 1 casts the decision made at a single meeting in a decision analytic framework. Replicating this decision over many time points produces a Markov decision process (MDP; Section 2.2). Section 3 compares the framework proposed here to others previously studied in the literature.

2.1 The value of assessment

In a typical educational setting, the payoff for both the student and the tutor comes not from the assessment or the instruction, but rather from the student’s proficiency at the end of the course. Figure 1 shows an influence diagram (Howard & Matheson, 1981) which illustrates this concept. Influence diagrams have three types of variables represented by three different node shapes:
• Square boxes are *decision variables*. Arrows going into decision variables represent information available at the time when the decision is to be made.

• Circles are *chance nodes* (random variables). Arrows going into the circles indicate that the distribution for a given variable is specified as conditional on its parents in the graph.

• Hexagonal boxes are *utilities*. Utilities are a way of comparing possible outcomes from the system by assigning them a value on a common scale (often with monetary units). Costs are negative utilities.

![Influence diagram for skill training decision.](image)

On the right of Figure 1, there is a node representing the utility associated with a student knowing the skill at the end of the course. The student’s probability of knowing the skill at the end of the course will depend on both the student’s skill level at the beginning of the course and what kind of instruction the student receives. The instruction has certain costs associated with it. We do not know the student’s ability at the beginning of the course, but we can give the student a pretest whose outcome will depend on the student’s ability. This pretest also has a cost associated with it. We can observe the outcome of the pretest when we make the decision about what instruction to give.

The decision of what instruction to provide depends not only on whether or not the student seems to have the skill from the pretest, but also the value of the skill and the cost of the instruction. If the instruction is very expensive and the skill not very valuable, it may not be cost-effective to give the instruction. Similarly, the decision about whether or not to test will depend on the cost of the test and the cost of the instruction. If the instruction is very inexpensive, it may be more cost-effective to just give the instruction and not bother with the pretest.

A concept frequently associated with decision analysis problems of this type is
the value of information (Matheson, 1990). Compare the utility of the best action we can take without observing the pretest score, to the expected results of taking an action chosen based on the pretest score. The difference between those two is the expected value of information. Note that the assessment is only worthwhile when the value of information from the test exceeds the cost of the test.

2.2 Markov decision processes

At each time point the tutor gives a benchmark assessment and selects an action, some instructional activity. Repeating this simple decision model at multiple time points produces Figure 2. Each time slice, $t = 1, 2, 3$ represents a short period of time in which the proficiencies of the learner are assumed to be constant. The variables $S_t$ (vector valued to reflect multiple skills) represent the proficiency of the student at each time slice. The variables $O_t$ (vector valued to represent multiple tasks or multiple aspects on which a task result may be judged) represent the observed outcomes from assessment tasks assigned during that time slice. The activities are chosen by the tutor and occur between time slices.

If we assume that $S_{t-1}$ is independent of $S_{t+1}$ given $S_t$, then Figure 2 represents a Markov decision process (MDP). In general, the proficiency variables $S$ are not observable; therefore, Figure 2 represents a partially observable Markov decision process (POMDP; Boutilier, Dean, & Hanks, 1999).

---

1 Technically, there is one other restriction on the MDP, and that is that the utility must also factor across time. In the educational setting, this is easily achieved by putting the utility on the final state of the proficiency variables.
There is a close correspondence between the POMDP model of Figure 2 and ECD objects (Mislevy et al., 2003). The nodes $S_t$ represent the proficiency model, and the edges between $S_t$ and $O_t$ represent the evidence models. This framework adds *action models* (Section 3) representing the change between $S_{t_1}$ and $S_{t_2}$.

### 2.3 Similarity to other temporal models

Removing the activity nodes from Figure 2 produces a figure that looks like a hidden Markov model. Langheine and Pol (1990) describe a general framework for using Markov models in psychological measurement. Eid (2002) uses hidden Markov models to model change in consumer preference, and Rijmen, Vansteelandt, & De Boeck (in press) uses these models to model change in patient state at a psychiatric clinic. Generally, in education the states of the proficiency variables are ordered and the student moves from lower states to higher states, where in the other applications the subject can move readily between states. Reye (2004) looks at Markov models in the context of education and uses a model based on dynamic Bayesian networks (Dean & Kanazawa, 1989), an important special case of the framework described here.

Similar models for change in student ability have been constructed using latent growth curves, structural equation models, and hierarchical linear models (Raudenbush, 2001). One complication addressed in many of these models, but not in this paper, is the hierarchy of student, classroom, teacher, and school effects, which can all affect the rate at which students learn.

Unlike other approaches, the MDP framework treats instruction as an action variable to be manipulated, rather than an observed covariate. This requires an explicit model for how an action affects student proficiency.

### 3. GENERAL ACTION MODEL

Let $S_t = (S_{i,t}, \ldots, S_{K,t})$ be the student’s proficiencies on $K$ different skills at Time $t$. Under the assumptions of the MDP, only the current value of the proficiency variables, $S_{i,t}$, and the action chosen at Time $t$, $a_t$, are relevant for predicting the state of the proficiency variables at the next time point, $S_{i,t+1}$. Assume that each Action $a$ targets a focal skill, $S_{i,a}$. For each action, the MDP requires a set of transition probabilities $P_a(S_{i,t+1} | S_{i,t}, a)$ describing the possible states of the focal skill at the next time point given all of the skills at the current time point.

One factor that affects the transition probabilities is how successfully the student completes the assigned work. The *general action model* (Figure 3) assumes that the outcome from each action can be represented with a binary
variable indicating successful completion. Mathias et al. (2006) call this model a bowtie model because the left and right halves are separated by the knot of the Success variable. The bowtie model contains a key simplification: The value of each focal skill at Time $t+1$ is conditionally independent from the prerequisites given the Success of the action and the value of the skill at Time $t$.

![Bowtie Model Diagram](image)

FIGURE 3
A General Action Model.

Under the bowtie model, experts need only worry about interactions among the skills when describing the prerequisites for Success of an action. When defining the effects of the action, the experts can work one skill at a time. Mathias et al. (2006) found that domain experts had difficulty specifying a complete set of transition probabilities for an action, but did feel comfortable supplying a list of prerequisites and consequences of an action along with a relative indication of the strength and direction for each variable.

The general action model makes a number of simplifying assumptions, but it also has clean graphical structure and can be parameterized to contain only a few parameters related to concepts with which experts will resonate. More importantly, it splits the problem of modeling an action into two pieces: modeling the factors contributing to success (Section 3.1) and modeling the effects of the action given success (Section 3.2).

### 3.1 The noisy-and model of success

Let $X_t$ represent the value of the Success variable for the action assigned at Time $t$, coded 1 for success and 0 for failure. Here “success” lumps together motivation—did the student do the lesson; mastery of the proficiency addressed in the lesson—did the student do it right; and appropriateness—
was this lesson an effective use of the student’s practice time. Although the size of the conditional probability table for Success could grow exponentially with the number of prerequisites, a number of attractive parameterizations with fewer parameters exist.

Usually, for an action to be successful, all the prerequisites for that action must be met. For each prerequisite, $S_k$, let $s_{k,a}^{-}$ be the minimum level of that proficiency required for Action $a$ to have a high probability of success and let $\delta(S_{k,a}^{-} \geq s_{k,a}^{-}) = 1$ if $S_{k,a}^{-} \geq s_{k,a}^{-}$ is true (student meets prerequisite for Proficiency $k$) and $\delta(S_{k,a}^{-} \geq s_{k,a}^{-}) = 0$ otherwise. Even given the current proficiency level of the student, there remains some uncertainty about the success of the action, $X_t$. In particular, let $q_{k,a}$ be the probability that a student assigned Action $a$ compensates for missing the prerequisite, $S_k$, and let $q_{0,a}$ be the probability that a student who meets the prerequisites successfully completes the activity. Then the noisy-and distribution (Junker & Sijtsma, 2001; Pearl, 1988) can model the relationship between $S$ and $X$:

$$P(X_t = 1 | S_t, a) = q_{0,a} \prod_k q_{k,a}^{1-\delta(S_{k,a}^{-} \geq s_{k,a}^{-})}.$$  

Vygotsky (1978)’s zone of proximal development can be modeled by making the interval two-sided, such as $\delta(S_{k,a}^{+} \geq s_{k,a}^{-})$. This states that the action would usually be unsuccessful (ineffective) if the prerequisite skill was either too high (task is too easy) or too low (too hard).

The number of parameters in this model is linear in the number of prerequisite skills. The experts must specify only the lower and upper bound for each prerequisite and the probability that the student can work around the missing prerequisite, $q_{k,a}$, as well as the probability of success when the prerequisites are met, $q_{0,a}$. In practice, experts will specify prior distributions for the parameters, which can be refined with appropriate longitudinal data. The boundaries should be close to the natural way that experts think about instructional activities (i.e., this activity is appropriate for persons whose skill is in this range and who meet these prerequisites).

Note that the value of the Success probability can be observed or unobserved. When Success is observed at Time $t$, it also provides information about the proficiency of the student at Time $t$. When Success is not observed, there may be some indirect measure of success, such as the student’s self-report. The Success variable could be generalized to take on multiple values such as the grade received on an assignment. However, the number of parameters increases as the number of states of Success increases, on both the right and left sides of the bowtie model.
3.2. Linear growth model

At this point, we make a fundamental assumption: Skills deteriorate over time unless that decrease is offset by practice. Practice causes the skill to increase; however, the effectiveness of the practice depends on how well the selected assignment (the action) is matched to the student’s current ability levels.

We assume that in the absence of instruction Proficiency $k$ deteriorates at a fixed rate $\mu$. Instruction produces an increase in the skill at the rate $\lambda(a, S_t, X_t)$, which is a function of the current proficiency state, $S_t$, the action taken at Time $t$, $a$, and the success of that action, $X_t$. Thus, the proficiency value at Time $t_2$ will be:

$$S_{k,t_2} = S_{k,t_1} + \lambda_k(a_t, S_t, X_t) \Delta t - \mu_k \Delta t_1 + \epsilon_{t,k},$$

where $\epsilon_{t,k} \sim N(0, \Delta \sigma_k^2)$. Making the variance depend on the elapsed time is a usual assumption of Brownian motion processes (Ross, 1989).

Both $\lambda_k(a_t, S_t, X_t)$ and $\mu$ are not identifiable from data without further constraints. One possible constraint is to set a zero growth rate for all unsuccessful actions, $\lambda(a_t, S_t, X_t = 0) = 0$. This model assumes common rate of skill deterioration, $\mu$, for failure across all actions. Another possibility is to set $\mu$ to a fixed value and model the proficiency change when the action fails individually for each action.

Although Equation 1 captures the prerequisite relationships, it does not do a good job of capturing the zone of proximal development (Vygotsky, 1978). We expect that an assignment that is a little bit too easy will have some value, just not as much. The same will be true for an assignment that is a little bit too hard. On the other hand, an assignment that is much too easy or much too hard should produce little benefit. Let $s_{k,a}$ be the target difficulty for Action $a$. One possible parameterization for $\lambda_k(a_t, S_t, X_t)$ is:

$$\lambda_k^*(S_t, a, X_t) = \eta_{k,a,X_t} - \gamma_{k,a,X_t} (S_{k,t} - s_{k,a}^*)^2.$$

The first term, $\eta_{k,a,X_t}$, is an action-specific constant effect. The second term is a penalty for the action being at an inappropriate level; the parameter $\gamma_{k,a,X_t}$ describes the strength of this penalty. Finally, note that one set of values is needed for a successful action, $X_t = 1$, and one for an unsuccessful action, $X_t = 0$.

4. USING THE MODEL FOR FILTERING AND PLANNING

Almond (2007) describes a simple simulation experiment using the POMDP framework. The paper describes two series of meetings between a music tutor...
and student, one monthly and one weekly. In both cases, the proficiency level at
each time point was estimated using a particle filter (see Doucet, Freitas, & Gordon,
2001; Liu, 2001) that follows the model closely and a simple exponential filter
(Brockwell & Davis, 2002) that estimates the current time point as a weighted
average between the current observation and the previous estimate.

The particle filter performed well in the artificial setting of Almond (2007).
At each time point, the true simulated proficiency fell within the 95%
forecasting interval returned by the filter. The exponential filter did not perform
as well. Under different values of its tuning parameter, the estimates were either
too sensitive to the fluctuations of the noisy benchmark test or too sensitive to
the initial conditions. Another possibility not explored in Almond is to use the
Kalman filter (Kalman & Bucy, 1961) in place of the more computationally
intensive particle filter. The Kalman filter will generally have a larger
forecasting error than the particle filter, but may be sufficient in situations where
the computational resources are limited.

Finally, the results from the simulation study do not take into account the
uncertainty in the parameters due to estimation. The particle filter used the same
parameters as the data generation process; thus, it was operating under nearly
ideal conditions. It is still unknown how easy it will be to estimate parameters
for this model. Fortunately, the problem can be divided into two pieces: estimating
the parameters for the evidence models (for the benchmark tests) and
estimating the parameters for the action models (for student growth). The former
(evidence models) can be done with cross-sectional data, for which it is
relatively inexpensive to get large sample sizes. Estimating growth requires
longitudinal data, which are always more expensive to collect.

The ultimate goal is to be able to use the POMDP framework to help the
teacher form a plan—a series of actions or strategies for selecting the actions at
each lesson—that has the best probability of getting the student to the goal state.
The solution to a MDP is a policy—a mapping from the proficiency states of the
student to actions (assignments given to the student).

Finding solutions to POMDPs is hard (Mundhenk, Lusena, Goldsmith, &
Allender, 2000), but a number of approximate algorithms for solving POMDPs
exist. Heuristic reasoning about prerequisites can reduce the number of
candidate actions.

We can use the POMDP model to calculate the probability that the student
will achieve a specified level of proficiency by a specified time point. In
particular, the model can identify students who are at-risk for not meeting the
proficiency goals. Additional resources from outside the system would need to
be brought to bear for these students.
5. DISCUSSION

The general action model described above captures many of the salient features of proficiency acquisition using only a small number of parameters. In particular, the number of parameters grows linearly in the number of prerequisites for each action, rather than exponentially. The bowtie model is a compromise between producing a parsimonious model and a model faithful to the cognitive science. As such, it may not satisfy either pure statisticians or cognitive scientists, but it should have sufficient ability to forecast student ability to be useful, especially for the purposes of planning.

Even though the models are parsimonious, the ability to estimate the parameters from data is still largely untested. Markov chain Monte Carlo and similar techniques provide an approach to the problem, but weak identifiability of the parameters could cause convergence problems. These models need to be tested both through simulation experiments and applications to real data.

In typical classroom assessments, periodic benchmark assessments must be short to minimize the amount of time taken away from instruction. As test length plays a strong role in determining reliability, such tests will have modest reliability at best. The MDP framework provides a sensible way to improve the current estimate of student proficiency with observations from previous benchmark assessments.

There is a point of diminishing returns. If the test adds almost no information to our current picture of the student’s proficiency, there is little point in testing. Likewise, there is little point in testing if the action will be the same no matter the outcome of the test. Actually, the lack of meaningful choices of actions may prove to be the largest problem that this framework poses. No matter how good the diagnostic assessment system is, it will have little value if the teacher does not have a rich set of viable instructional strategies available to ensure that the educational goals are reached. On the other hand, a diagnostic assessment system that integrates explicit models of student growth together with a well-aligned set of instructional actions could add value for both students and teachers.

ACKNOWLEDGMENTS

The general action model (Section 3) stems from some consulting work I did with Judy Goldsmith and her colleagues at the University of Kentucky (Mathias et al., 2006) and is really joint work of that collaboration.
REFERENCES


