

Modeling Student Learning: Binary or Continuous Skill?

Radek Pelánek
Masaryk University Brno
pelanek@fi.muni.cz

ABSTRACT

Student learning is usually modeled by one of two main approaches: using binary skill, with Bayesian Knowledge Tracing being the standard model, or using continuous skill, with models based on logistic function (e.g., Performance Factor Analysis). We use simulated data to analyze relations between these two approaches in the basic setting of student learning of a single skill. The analysis shows that although different models often provide very similar predictions, they differ in the impact on student practice and in the meaningfulness of parameter values.

Keywords

student modeling; learning; Bayesian Knowledge Tracing; simulated data

1. INTRODUCTION

In this work we focus on modeling of student learning in the basic setting: we assume that for each student we have a sequence of answers related to a single skill and we consider only correctness of these answers, i.e., we do not take into account additional information like response times or partial correctness due to the use of hints. We work only with basic models and focus on experiments with simulated data. This setting is of course a coarse simplification, since in a real application we typically have some additional information on student answers, questions are related to multiple skills, and model extensions are used. But in order to successfully use complex models, it is necessary to have deep understanding of the base case and this understanding is still lacking. There are many feasible modeling approaches, but they are usually proposed and studied independently and their relations, similarities, and differences have not been well studied. The use of simulated data allows us to analyze behaviour of models in detail thanks to the knowledge of “ground truth” values; moreover, we can manipulate in controlled way generation of data and thus easily evaluate behaviour of models under different assumptions.

2. MODELING STUDENT LEARNING

Most approaches to modeling of student learning can be viewed as hidden Markov models (also called latent process models, state-space models). We assume a hidden (latent) state variable (called “skill”) and two types of equations. Observation equation describes the dependence of observed variables (correctness of answers) on the hidden variable (skill). State equation describes the change of the hidden variable (i.e., learning). There are two main types of models depending on whether the latent skill is binary or continuous. It is in principle possible to consider discrete skill with more than two states, but such models are not commonly used. The standard form of a binary skill model is Bayesian Knowledge Tracing (BKT) [1]. Models based on continuous latent skill typically use logistic function for observation equation, they differ in their approach to skill estimation.

Bayesian Knowledge Tracing assumes a sudden change in knowledge. It is a hidden Markov model where skill is a binary latent variable (either learned or unlearned). Figure 1 illustrates the model; the illustration is done in a non-standard way to stress the relation of the model to the model with continuous skill. The estimated skill is updated using a Bayes rule based on the observed answers; the prediction of student response is then done based on the estimated skill. Note that although the model is based on the assumption of binary skill, the skill estimate is actually continuous number (in the $[0, 1]$ interval).

Models which utilize the assumption of continuous latent skill consider skill in the $(-\infty, \infty)$ interval and for the relation between the skill and the probability of correct answer use the logistic function $\sigma(x) = \frac{1}{1+e^{-x}}$. Although it would be possible to consider also other functions, the logistic function is currently the standard choice. As a simple model of learning we consider a simple linear growth of the skill (Figure 1). More specifically, for the initial skill θ_0 we assume normally distributed skill $\theta_0 \sim N(\mu, \sigma^2)$ and for the change in learning we consider linear learning $\theta_k = \theta_0 + k \cdot \Delta$, where Δ is either a global parameter or individualized learning parameter (in that case we assume a normal distribution of its values). This model is a simplified version of the Additive Factors Model [3]; the original additive factor model uses multiple skills. A principled way of estimating continuous skills is the Bayesian approach, which computes not just a point estimate of skill, but a distribution over skill capturing also the uncertainty of the estimate. This approach be implemented for example using particle filter, i.e., discretized

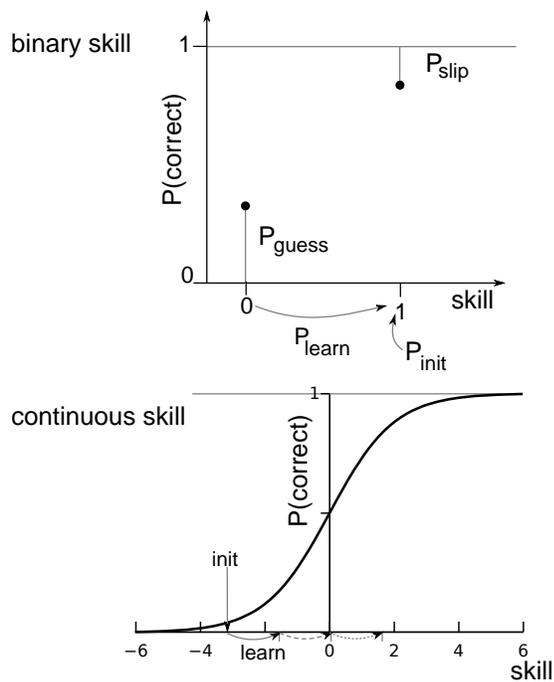


Figure 1: Binary and continuous skill models of student learning – high level overview.

representation of posterior distribution. A more pragmatic approach to skill estimation is Performance Factor Analysis [5], which computes the skill estimate as a linear combination of the number successes and failures of a student. This approach can be extended to take into account ordering of attempts and time intervals between them [2, 4].

Which type of model is better depends on the learning situation. Binary skill models assume a sudden switch from unlearned to learned state. Such assumption is appropriate mainly for fine-grained skills which require understanding or insight (such as “addition of simple fractions”). Models with continuous skill assume gradual increase of skill. This is appropriate either for modeling coarse-grained skills (e.g., “fractions” as a single skill) or for situations where gradual strengthening happens (e.g., memorizing facts).

3. EXPERIMENTS

To analyze the described models and relations between them we performed experiments with simulated data. We generated simulated data by one of the models and then analyzed the generated data using both models with binary and continuous skills. For generating data we used 10 scenarios with different parameter settings.

With respect to accuracy of predictions the results show that both types of models bring consistent improvement over baselines like moving average and time decay models [6]. The basic comparison of binary and continuous skill models is also not surprising: each approach dominates in scenarios which correspond to its assumptions. Nevertheless, in many cases the differences are small and the predictions are actually highly correlated.

Models are not used only for predictions, but they may be useful in themselves for system developers and researchers. Plausible and explainable model parameters may be used to get insight into behaviour of tutoring systems and also for “discovery with models” (higher level modeling). Results of our analysis show that in the case when there is a mismatch between source of the data and a model, interpretation of parameters may be misleading. As a specific example consider simulated students behaving according to the continuous model with $\theta_0 \sim N(-1, 1)$, $\Delta = 0.2$. Here the fitted BKT guess and slip parameters are 0.24 and 0.16. Intuitive interpretation of BKT parameters would thus suggest high chance of guessing an answer. In the ground truth model, however, chance of guessing converges to zero for unskilled students.

One of the main applications of student models is to guide the behaviour of adaptive educational systems. A typical example is the use of student models for mastery learning – students have to practice certain skill until they reach mastery, the attainment of mastery is decided by a student model. Mastery is declared when a skill estimate is higher than a given threshold. How does the choice of student model and a threshold impact student practice? Our results show that the BKT model is relatively insensitive to the choice of the threshold and that the model provides weak decisions for scenarios with continuous learning, specifically when the learning rate is low. Continuous skill models can provide good decision for all scenarios if used with a good threshold. However, optimal thresholds differ significantly for scenarios with binary skill and continuous skill.

To summarize, our study with simulated data suggests that the choice between models with binary and continuous skill does not seem a key concern as long as we are interested only in predictions of students’ answers, but it can have significant impact on parameter interpretation and mastery learning.

4. REFERENCES

- [1] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.
- [2] A. Galyardt and I. Goldin. Recent-performance factors analysis. In *Educational Data Mining*, 2014.
- [3] Tanja Käser, Kenneth R Koedinger, and Markus Gross. Different parameters - same prediction: An analysis of learning curves. In *Educational Data Mining*, pages 52–59, 2014.
- [4] J. Papoušek, R. Pelánek, and V. Stanislav. Adaptive practice of facts in domains with varied prior knowledge. In *Educational Data Mining*, pages 6–13, 2014.
- [5] P. I. Pavlik, H. Cen, and K. R. Koedinger. Performance factors analysis-a new alternative to knowledge tracing. In *Artificial Intelligence in Education*, volume 200, pages 531–538. IOS Press, 2009.
- [6] R. Pelánek. Time decay functions and Elo system in student modeling. In *Educational Data Mining (EDM)*, pages 21–27, 2014.