

Logistic Regression in a Dynamic Bayes Net Models Multiple Subskills Better!

YANBO XU

JACK MOSTOW

Carnegie Mellon University, United States

A single student step in an intelligent tutor may involve multiple subskills. Conventional approaches either sidestep this problem, model the step as using only its least known subskill, or treat the subskills as necessary and probabilistically independent. In contrast, we use logistic regression in a Dynamic Bayes Net (LR-DBN) to trace the multiple subskills. We compare these three types of models on a published data set from a cognitive tutor. LR-DBN fits the data significantly better, with only half as many prediction errors on unseen data.

Key Words and Phrases: Knowledge Tracing, multiple subskills, conjunctive models, Dynamic Bayes Net

1. INTRODUCTION

Knowledge tracing [Corbett and Anderson, 1995] is widely used to estimate from the observable steps that require a skill the probability that the student knows the skill. However, steps that require multiple subskills are problematic.

One approach [Cen et al., 2006] tries to sidestep the problem by modeling each set of subskills as a distinct individual skill, e.g., computing the area of a circle embedded in a figure vs. by itself. However, modeling them as individual skills ignores transfer of learning between them. Another solution [Cen et al., 2008; Gong et al., 2010] models the step as applying each subskill independently, and assigns it full credit or blame for getting the step right or wrong. Then the subskills can be traced independently. This solution assumes that the probability of having the knowledge needed for the step is the product of the probabilities of knowing the subskills. A third solution assigns all the credit or blame to the least known subskill. This solution approximates the probability of having the knowledge required for the step as the minimum of those probabilities instead of their product.

Recently, Xu and Mostow [2011] described a method to trace multiple subskills by using logistic regression in a Dynamic Bayes Net (LR-DBN). It models the transition probabilities from the knowledge state $K^{(n-1)}$ at step $n-1$ to the knowledge state $K^{(n)}$ at step t as logistic regressions over all m subskills:

$$P(K^{(0)} = false) = \frac{\exp(-\sum_{j=1}^m \beta_j^{(0)} s_j^{(0)})}{1 + \exp(-\sum_{j=1}^m \beta_j^{(0)} s_j^{(0)})}, \quad (1)$$

$$P(K^{(n)} = false \mid K^{(n-1)} = false) = \frac{\exp(-\sum_{j=1}^m \beta_{1,j} s_j^{(n)})}{1 + \exp(-\sum_{j=1}^m \beta_{1,j} s_j^{(n)})}, \quad (2)$$

$$P(K^{(n)} = false \mid K^{(n-1)} = true) = \frac{\exp(-\sum_{j=1}^m \beta_{2,j} s_j^{(n)})}{1 + \exp(-\sum_{j=1}^m \beta_{2,j} s_j^{(n)})}. \quad (3)$$

Here the indicator variable s_j is 1 if the step requires subskill j , 0 otherwise; $\beta_j^{(0)}$ is the coefficient for subskill j fit at the initial state; and $\beta_{1,j}$ and $\beta_{2,j}$ are other coefficients.

We now compare the prediction accuracy and complexity of the three types of models: conjunctive product, conjunctive minimum, and LR-DBN.

2. EXPERIMENTS AND RESULTS

We compare the three models on a published [Koedinger et al., 2010] data set from 123 students working on a geometry area unit of the Bridge to Algebra Cognitive Tutor[®]. All three models use the same 50 subskills. We fit the models separately for each student on the first half of all of that student's steps, test on the second half, and average the results.

Table I shows the mean fit of each student's data to the three types of model. The values in parentheses show 95% confidence intervals based on standard error calculated from the unbiased weighted sample variance of individual students' accuracies. LR-DBN does best with 92.7% ($\pm 2.0\%$) accuracy. It makes only half as many prediction errors as the other models, neither of which predicts significantly better than the majority class.

Since the data is unbalanced (84.7% of all steps are correct), we also report within-class accuracy. All three models exceed 96% accuracy within the positive class, with no significant differences. In contrast, accuracy within the negative class is highest by far for LR-DBN (72.3%), while the other two models do much worse than (50-50) chance. Confidence intervals are looser within the negative class, both because it is smaller (15.3% of the data vs. 84.7%), and because its accuracy varies more across students.

LR-DBN is less complex than the other two models in that it has fewer parameters for each student. For each student, it has 50 times 3 logistic regression coefficients for the initial state and two transition probabilities, plus two more parameters for *guess* and *slip*, for a total of 152. In contrast, the other two models have 200 = 50 times 4 parameters for *already know*, *guess*, *slip*, and *learn* since we fit knowledge tracing per subskill per student. Table II compares their total AIC and BIC scores for model complexity.

Future work should compare on other data sets to see if LR-DBN fares best there too, and to alternative models of multiple subskills to see if LR-DBN beats them as well.

Table I. Mean Per-student Fit of Each Type of Model

	Accuracy	Accuracy Within Positive Class	Accuracy Within Negative Class
Conjunctive product	85.0% ($\pm 1.0\%$)	96.5% ($\pm 0.7\%$)	21.2% ($\pm 3.8\%$)
Conjunctive minimum	85.8% ($\pm 1.0\%$)	98.7% ($\pm 1.0\%$)	14.6% ($\pm 2.3\%$)
LR-DBN	92.7% ($\pm 2.0\%$)	96.5% ($\pm 1.3\%$)	72.3% ($\pm 7.8\%$)

Table II. Summed Complexity Measures of Each Type of Model

	AIC	BIC
Conjunctive product	66,402.1	135,581.9
Conjunctive minimum	66,402.1	135,581.9
LR-DBN	60,545.2	114,159.5

REFERENCES

- CEN, H., KOEDINGER, K. and JUNKER, B. 2006. Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, Jhongli, Taiwan, June 26-30, 2006, 164-175.
- CEN, H., KOEDINGER, K.R. and JUNKER, B. 2008. Comparing Two IRT Models for Conjunctive Skills. In *Ninth International Conference on Intelligent Tutoring Systems*, Montreal, 2008, 796-798.
- CORBETT, A. and ANDERSON, J. 1995. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4, 253-278.
- GONG, Y., BECK, J. and HEFFERNAN, N.T. 2010. Comparing Knowledge Tracing and Performance Factor Analysis by Using Multiple Model Fitting Procedures. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, Pittsburgh, PA, 2010, V. ALEVEN, J. KAY and J. MOSTOW, Eds. Springer Berlin / Heidelberg, 35-44.
- KOEDINGER, K.R., BAKER, R.S.J.D., CUNNINGHAM, K., SKOGSHOLM, A., LEBER, B. and STAMPER, J. 2010. A Data Repository for the EDM community: The PSLC DataShop. In *Handbook of Educational Data Mining*, C. ROMERO, S. VENTURA, M. PECHENIZKIY and R.S.J.D. BAKER, Eds. CRC Press, Boca Raton, FL, 43-55.
- XU, Y. and MOSTOW, J. 2011. Using Logistic Regression to Trace Multiple Subskills in a Dynamic Bayes Net. In *4th International Conference on Educational Data Mining* Eindhoven, Netherlands, July 6-8, 2011.