Comparing Two IRT Models for Conjunctive Skills

Hao Cen, Kenneth Koedinger, and Brian Junker

Carnegie Mellon University
5000 Forbes, Pittsburgh, PA, U.S.A.
hcen@andrew.cmu.edu, koedinger@cmu.edu, brian@stat.cmu.edu

Abstract. A step in ITS often involve multiple skills. Thus a step requiring a conjunction of skills is harder than steps that require requiring each individual skill only. We developed two Item-Response Models – Additive Factor Model (AFM) and Conjunctive Factor Model (CFM) – to model the conjunctive skills in the student data sets. Both models are compared on simulated data sets and a real assessment data set. We showed that CFM was as good as or better than AFM in the mean cross validation errors on the simulated data. In the real data set CFM is not clearly better. However, AFM is essentially performing as a conjunctive model.

1 Introduction

A step in ITS is "the smallest possible correct entry that a student can make. It connects the transaction-level representation to the theoretically-derived knowledge component level" [1]. Often times in ITS, students face steps with conjunctive skill requirements -- The student needs multiple skills to solve the whole step. Thus, a step requiring a conjunction of skills is harder than steps that require requiring each individual skill only. In this paper, we want to answer the following questions:

1. Are the skills in the student log data set conjunctive in character?
2. What is a better method to model this conjunctivity?

These questions are of importance to the ITS community who is interested building quality cognitive model. Two popular classes of ITS – the Model Tracing Tutor, which builds explicitly on the cognitive model, and the Constraint-based Tutor [2] – represents domain knowledge as state constraints The effectiveness of these tutors is largely determined by the quality of written skills and constraints. A conjunctive model can be useful in evaluating the cognitive model against empirical student data.

2 Two IRT Models

To test the skill conjunctivity in the data, we developed two parametric IRT models. The first model – the Additive Factor Model (AFM) –, originated from [3], is depicted by Equation (1). The term "Additive" comes from the fact that a linear combination of skill parameters determines $\text{logit}(p_{ij})$ in the equation

$$p_{ij} = \text{Pr}(Y_{ij} = 1 \mid \theta_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{\exp(\theta_i + \sum\limits_{k=1}^{K} q_{jk} \beta_k + \sum\limits_{k=1}^{K} q_{jk} \gamma_k T_{ik})}{1 + \exp(\theta_i + \sum\limits_{k=1}^{K} q_{jk} \beta_k + \sum\limits_{k=1}^{K} q_{jk} \gamma_k T_{ik})} \tag{1}$$

Where

Y_{ij} = the response of student i on item j

θ_i = coefficient for proficiency of student i

β_k = coefficient for difficulty of skill k

γ_k = coefficient for the learning rate of skill k

T_{ik} = the number of practice opportunities student i has had on the skill k

q_{jk} = 1 if item j uses skill k; 0 otherwise

K = the total number of skills in the Q-matrix

The intuition of this model is that the probability of a student getting a step correct is proportional to the amount of required knowledge the student knows, plus the "easiness" of that skill, plus the amount of learning gained for each practice opportunity.

The second model -- the Conjunctive Factor Model (CFM) -- depicted by Equation (2) can be thought as modeling the conjunctivity as a multiplication of skill parameters. CFM is a special case of Embretson's multicomponent latent trait model [4] and is customized for the high dimensional feature of ITS, as there are many more skills in a cognitive model than the number of cognitive attributes in a traditional assessment.

$$p_{ij} = \prod_{k=1}^{K} \left(\frac{e^{\theta_i + \beta_k + \gamma_k T_{ik}}}{1 + e^{\theta_i + \beta_k + \gamma_k T_{jk}}} \right)^{q_{jk}} \tag{2}$$

To fit the parameters from the data, we designed a Penalized Maximum Likelihood Estimation method (PMLE) depicted in equation (3) to overcome over fitting. This PMLE penalizes the oversized student parameters in the joint estimation of the student and the skill parameters. Maximizing Equation (3) is equivalent to finding a posterior mode for a Bayesian model, with a normal prior on θ and flat priors on β and γ. A higher value for λ below corresponds to lower prior variance. The BFGS optimization algorithm is used in computing PMLE.

$$ll_{PMLE} = ll_{MLE} - \frac{1}{2} \lambda \sum_{i=1}^{I} \theta_i^2, \ \lambda = 1 \text{ by default} \tag{3}$$

where I = the total number of students.

3   Simulation Results, Real Data Results and Discussion

To compare CFM with AFM, we used both a simulated data set and a real assessment data set. Since students in the assessment data set were not exposed to repetitive learning opportunities, we removed the learning term from both models. Cross validation errors and the interpretability of the actual parameter fits are used to evaluate the models.

The simulated data is used to answer the question "If the data is conjunctive, which model is better?" We simulated data drawn from a CFM model with 100 student parameters, 3 skill parameters, and 7 items. We explored for different sets of the three skill probability values (.1, .5, .9), (.1, .1, .1), (.4, .5, .6) and (.9, .9, .9). In nearly all cases, CFM-PMLE was as good as or better than AFM-PMLE in cross validation. The biggest difference was from the skill set (.9, .9, .9) because the skill parameter values are so high that AFM-P cannot behave in a conjunctive form (which it can if the logic values are negative). Table 1 shows the results from one of the above skill sets.

Table 1. Model comparison of the simulated data. β = (.1, .5, .9).in probability.

	CVMean	CVSd	$\hat{\beta}$ in probability	$\hat{\beta}$ in logit
AFM-P	0.120	0.281	(0.03, 0.34, 0.73)	(-3.4, -0.67, 0.97)
CFM-P	0.111	0.174	(0.07, 0.5, 0.89)	(-2.54, 0.02, 2.07)

The real data is used to answer the questions "Is the data conjunctive? If so, which model is better" The real data set EAPS is taken from a difficulty factor study of 247 U.S. algebra students [5]. There are 1976 observations and 96 distinctive items. A simplification of their skill coding involves 3 skills.

Table 2. Model comparison of the EAPS data

	CVMean	CVSd	$\hat{\beta}$ in probability	$\hat{\beta}$ in logit
AFM-P	0.202	0.142	(0.35, 0.47, 0.43)	(-0.63, -0.14, -0.3)
CFM-P	0.187	0.221	(0.61, 0.7, 0.67)	(0.43, 0.85, 0.7)

In the real data set, CFM-P is not clearly better. However, AFM-P is essentially performing as a conjunctive model because the estimates are all negative. We saw this in the simulated data set with lower probability skill estimates. A goal of future research is to test CFM-P on real data sets where AFM-P cannot act conjunctively as in Table 2.

References

1. Kurt VanLehn, K.K., Skogsholm, A., Nwaigwe, A., Hausmann, R.G.M., Weinstein, A., Billings, B.: Whats in a step? Toward general, abstract representations of tutoring system log data. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511. Springer, Heidelberg (2007)
2. Mitrovic, A., Koedinger, K.R., Martin, B.: A Comparative Analysis of Cognitive Tutoring and Constraint-Based Modeling. In: User Modeling 2003. Springer, Heidelberg (2003)
3. Cen, H., Koedinger, K., Junker, B.: Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement. In: 8th International Conference on Intelligent Tutoring Systems (2006)
4. Embretson, S.: Multicomponent Response Models. In: Linden, W.V.D., Hambleton, R.K. (eds.) Handbook of Modern Item Response Theory. Springer, Heidelberg (1997)
5. Koedinger, K.R., Nathan, M.J.: The real story behind story problems: Effects of representations on quantitative reasoning. The Journal of the Learning Sciences (2003)