Generalized Learning Factors Analysis:
Improving Cognitive Models with Machine Learning
Thesis Proposal to the Machine Learning Department

Hao Cen
Committee: Ken Koedinger (HCII &Psy)
Brian Junker (Stat), Geoff Gordon (ML), Noel Walkington (Math)

I. The Challenge

Of all the initiatives to improve the math level of U.S. students, vastly improving K-12 math education has been a top priority. One major development toward this end is Intelligent Tutoring Systems. The technology that drives intelligent tutoring systems is grounded in research into artificial intelligence and cognitive psychology, which seeks to understand the mechanisms that underlie human thought, including language processing, mathematical reasoning, learning, and memory. As students attempt to solve problems using these tutoring systems, the programs analyze their strengths and weaknesses and on that basis provide individualized instruction. Intelligent tutoring systems do not replace teachers. Rather, they allow teachers to devote more one-on-one time to each student, and to work with students of varying abilities simultaneously. They allow teachers to design assignments targeted to individual student needs, thereby increasing student advancement at a better rate.

A primary example of Intelligent Tutoring Systems helping U.S. children learn math is *Cognitive Tutors*, an award-winning computer-based math program that grows out of the extensive research in human learning and artificial intelligence at Carnegie Mellon. Evidence indicates that students using the *Cognitive Tutors* program perform 30% better on questions from the TIMSS assessment, 85% better on assessments of complex mathematical problem solving and thinking, and attain 15-25% higher scores on the SAT and Iowa Algebra Aptitude Test. The equivalent learning results hold for both minority and non-minority students (Kenneth R. Koedinger & Anderson, 1997; Kenneth R. Koedinger, Corbett, Ritter, & Shapiro, 2002; Sarkis, 2004). By 2007, more than 500,000 middle school students began using *Cognitive Tutors* across the United States.

The full potential of ITS has not yet been reached, though. The issues mainly concern the efficiency level of the cognitive models used, which is at the heart of most tutoring programs. These models describe a set of math skills that represent how students solve math problems.

With cognitive models, ITS assesses student knowledge systematically and presents curricula tailored to individual skill levels and generates appropriate feedback for students and teachers. An incorrect representation of the domain skills may lead to erroneous curriculum design and negatively affect student motivation. An inaccurate model may waste limited student learning time, and teacher instructional energies, both of which are vital to full achievement. According to Carnegie Learning, teachers reported that many students could not complete the tutor curriculum on time. This issue is serious. First, if students cannot complete the cognitive tutor curriculum, they are likely to fall behind their peers. Second, schools today are calling for increased instruction time to ensure adequate yearly progress. The reality, however, is that students have a limited amount of total available learning time, and teachers have a restricted amount of instructional time. Saving one hour of instructional time can be far more productive than increasing instruction by the same amount. This saved time does not reduce student or teacher workloads, but simply makes better use of the energy and attention given to this subject, thus allowing for greater devotion to other academic areas, thus increasing performance in those subjects. The learning gain may be remarkable.

Getting the appropriate cognitive model is challenging because:

1) There are hundreds of skills involved in a single sub-domain of math. For example, the middle school geometry curriculum is estimated to have over 200 individual skills.
2) Many math skills are not explicitly stated in textbooks, and textbook authors often expect students to acquire those skills via problem solving.
3) Skill is not directly observable. The mastering of a skill can only be inferred from student performance on tasks that require those skills.
4) Initial cognitive models were written by math experts. Many prior studies in cognitive psychology have shown that experts often make false predictions about what causes difficulty for students due to "expert blind spots" (K. Koedinger & Nathan, 1997; K. R. Koedinger & Nathan, 2003; M. J. Nathan & Koedinger, 2000, 2003; M. J. Nathan, Koedinger, & Alibali, 2001; M. J. Nathan, Long, & Alibali, 2002; Mitchell J. Nathan & Petrosino, 2003; M. J. Nathan, Stephens, Masarik, Alibali, & Koedinger, 2002)

The existing cognitive models are usually an incomplete representation of student knowledge, resulting in both less accurate assessment of student knowledge and lower student learning efficiency than desired. Improving the existing cognitive models, given the rate at which the Cognitive Tutor is used across the U.S., has an immediate and significant impact on student learning, and has a long-term impact on transforming math curriculum design. Now, an increasing number of student learning data becomes available. Within Pittsburgh Science of Learning Center, a central education data warehouse has hosted over 50 student learning data sets ranging from the domain of algebra to foreign language learning. The challenge, then, is how do we get a better cognitive model, using student learning data?

II. Grand Vision of Learning Factors Analysis (LFA)

This challenge can be formulated as a machine learning problem -- How do we characterize the process or the features of the process that generates the student responses on items? Figure 1 shows a visual representation of the problem. On the right hand side are the items. Each item has responses as 1 for correct and 0 for incorrect.

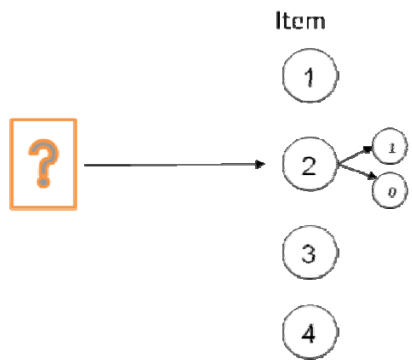

Figure 1 A student response model

Two extreme solutions to this problem are 1) to model the student responses with all the items (Figure 2), and 2) to model the student responses with a single latent factor, such as student intelligence, aptitude (Figure 3). These two approaches represent the way that modern educational tests are built upon.

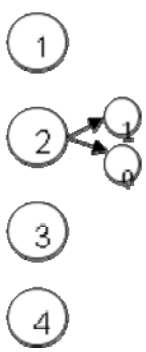

Figure 2 Item response model

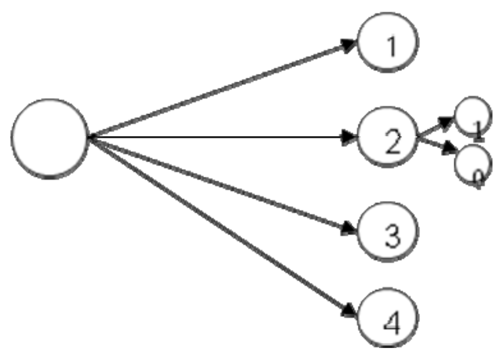

Figure 3 Single latent variable response model

Today's classrooms increasingly need Cognitively Based Assessment for Learning to support instruction and learning. Knowing student performance at the item level or at the single latent factor level alone offers limited information for students, teachers, parents and other stake takers for improving instruction and learning.

III. Generalized Learning Factors Analysis

Learning Factors Analysis is the method I have worked on with Dr. Ken Koedinger and Dr. Brian Junker over the past three years to address this challenge (Cen, Koedinger, & Junker, 2005, 2006; Cen, Koedinger, & Junker, 2007). One distinction between LFA and the previous two extremes is that LFA characterizes student responses on items in terms of the skills students uses, i.e. the cognitive model, seen in Figure 4.

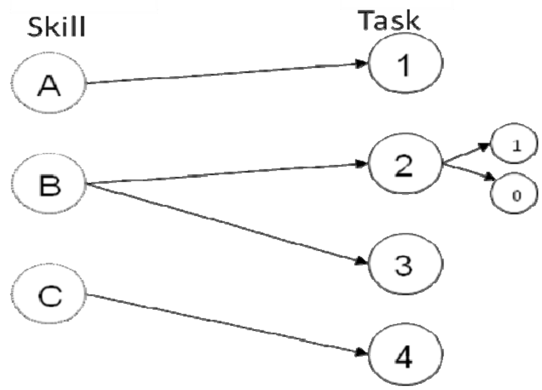

Figure 4 Skill response model

This method combines human learning theory, machine learning technology, and psychometrics. It can semi-automatically evaluate existing cognitive models against student learning data and search for better models. In this thesis, I generalize the method to handle a larger class of models and handle a larger scale of data

LFA draws strengths from different fields. In AI, it uses combinatorial search (Russell & Norvig, 2003). In machine learning, it builds upon latent factor model (Gordon, 2002). In data mining, it borrows the idea of improving Q-matrix from (Barnes, 2005). In statistics, particularly a branch of statistics called psychometrics, it shares strength with Q-matrix (Tatsuoka, 1983) and item response models (DiBello, Stout, & Roussos, 1995; Embretson, 1997; von Davier, 2005). In cognitive psychology, it extends the early work in learning curve analysis (Corbett & Anderson, 1995). LFA seamlessly putting the ideas from different field into one framework and in some of those fields, LFA makes a unique contribution and extension. In the following parts, I explain each component in details.

1. The Q-Matrix

Q-matrix is a boolean matrix describing the relationship between items and skills (Tatsuoka, 1983) . A cell value of 1 at the row i, column j means that the item i requires the use of skill j. A cell value of 0 means otherwise. Table 1 shows such a relationship between two testing items and four associated skills. Notice the first item requires only one skill and the second item requires two skills simultaneously. Requiring multiples skills simultaneously is called conjunctivity of skills.

Table 1 A sample Q-matrix

Item | Skill	Add	Sub	Mul	Div
2*8	0	0	1	0
2*8 - 3	0	1	1	0

2. The Additive Factor Model (AFM) (Draney, Pirolli, & Wilson, 1995)

Modeling the item response given the skills requires a statistical model. The first model we developed, depicted by equation (1), capture that the probability for student i to get item j right is proportional to required knowledge the student knows plus "easiness" of that skill and the learning acquired through practice,

$$p_{ij} = \Pr(U_{ij} = 1 \mid \theta_i, \beta_k, \gamma_k) = \frac{\exp(\theta_i + \sum_{k=1}^{K} q_{jk}\beta_k + \gamma_k T_{ik})}{1 + \exp(\theta_i + \sum_{k=1}^{K} q_{jk}\beta_k + \gamma_k T_{ik})} \qquad (1)$$

where

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

θ_i = coefficient for student i

β_k = coefficient for 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

$$q_{jk} = \begin{cases} 1 & \text{item j uses skill k} \\ 0 & \text{otherwise} \end{cases}$$

The term "Additive" comes from the linear combination of skill ks in item j in the exponent. That is, if an item requires multiple skills, this model will use the linear combination of the item parameters to predict the overall response.

The model has a connection with Logistic Regression by modeling success as a Bernoulli distribution with the probability of p, and student intelligence, skill easiness, and learning as predictors.

This model also has a connection with Item Response Theory. The additive factor model without the learning term reduces to the Linear Logistic Test Model (Fischer, 1977) with skills as the item attributes.

3. The conjunctive factor model (CFM)

One problem with AFM is the way it handles conjunctive skills. Suppose there is an item requiring two skills, as shown in Table 2. Assume a student has a theta value of 0; two skills above have beta values logit(.8) and logit(.5) (which is 0); and there are no learning (gamma = 0). In a conjunctive sense, we need a prediction of .4 (= 1/(1 + exp(-(logit(.8))) * 1/(1 + exp(-(logit(.5))) = .8 * .5). AFM will predict the third item with probability of .8 (=1/(1 + exp(-(logit(.8) + logit(.5)))), making a harder item easier.

Table 2 Skills and predicted probability for three algebra items

Item	Skill	P
2*8	mult	.8
7 - 3	sub	.5
2*8 - 3	mult, sub	.5 * .8 = .4

The conjunctive factor model (CFM), depicted by equation (2), captures the idea that when an item requires multiple skills present, the item is harder than the items requiring only one of those skills. The parameters in CFM have same meaning as those in AFM. CFM and AFM reduces to the same form when there is only one skill per item.

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

The conjunctive IRT model in equation 2 builds upon Embretson's multicomponent latent trait model (MLTM) (1997), Dibello's Unified Model (UM)(1995), and Davier's General Diagnostic Model (GDM) (2005).

4. Parameter estimation

Maximum Likelihood Estimation (MLE) has good asymptotic properties for estimators. Thus, we used a method called Joint Maximum Likelihood Estimation (JML) to estimate the student, skill, learning parameters all together. We also compared JML with a variation MLE called Penalized Joint Maximum Likelihood Estimation (PJML) (Harrell, 2001), which estimates all parameters together and penalizes the likelihood from having extreme student parameters, to avoid over fitting.

5. Assessment of the Statistical Models

Good statistical models balance between model fit & complexity minimizing prediction risk. They captures sufficient variation in data but is not overly complicated (Wasserman, 2004).

We choose two measures for model assessment -- Stratified K-Fold Cross Validation, shown in equation (3), which is time-consuming and more accurate estimate of prediction errors, and BIC, shown in equation (4) , which can be fast to compute but may be a crude approximate of the prediction errors. The stratification is taken on the student side.

$$CV = \sum_{i=1}^{n} (Y_i - \hat{f}^{-\kappa(i)}(x_i))^2 \qquad (3)$$

$$BIC = -2LogLikelihood + numParemeter * numObservation \quad (4)$$

To compare the performance on the two models and the two parameter estimation methods, we simulated data generated from a CFM model with 100 student parameters drawn from a standard normal distribution, 7 items, and 3 skills. The skills have values (-2.2, 0, 2.2), which can be converted to a probability scale as (.1, .5, .9). The Q-matrix for this data set and the model comparisons are shown in Table 3 and Table 4.

Table 3 The Q-matrix for the simulated data

	A	B	C
T100	1		
T010		1	
T001			1

T110	1	1	
T101	1		1
T011		1	1
T111	1	1	1

Table 4 The model statistics for the simulated data. AFM and CFM use maximum likelihood estimation. PAFM and PCFM stand for AFM and CFM with penalized maximum likelihood estimation.

	CVMean	CVSd	BIC	A	B	C
AFM	15.87	3.93	1018.30	-5.91	-2.47	0.34
CFM	15.84	4.30	988.09	-5.52	-1.68	3.16
PAFM	12.03	2.81	1156.83	-3.40	-0.67	0.97
PCFM	11.11	1.74	1127.78	-2.54	0.02	2.07

Among the combination of the statistical models and the parameter estimation methods, PCFM works best by having the smallest CV, the smallest standard deviation of the CVs in the validation folds, and the closest parameter estimates. With the penalized maximum likelihood estimation, BIC indicates the conjunctive model is better than the additive model.

We also compared AFM, CFM and the two parameter estimation methods over a real assessment data set, taken from the EAPS study (Kenneth R. Koedinger & MacLaren, 2002). This data set has 246 student counts, 96 items and 3 skills with known conjunctive structure. A partial Q-matrix for this data set and the model comparisons are given in Table 5 and Table 6. Still, PCFM wins in terms of the CV scores. This gives us some confidence on using PCFM for data with conjunctive skills and BIC for a fast measure of model quality.

Table 5 The partial Q-matrix for the real data

Item	S	H	U
bball-equation-result-easy-div	1		
bball-equation-result-easy-mult	1		
bball-equation-result-hard-div	1	1	
bball-equation-result-hard-mult	1	1	
bball-equation-start-easy-div	1		1

Table 6 The model statistics for the real data

	CVMean	CVSd	BIC	A	B	C

AFM	48.70	2.58	3199.88	-1.74	-1.29	-1.51
CFM	48.39	5.33	3225.26	2.18	2.85	2.69
PAFM	48.34	3.03	3695.06	-0.52	0.03	-0.13
PCFM	44.51	1.45	3544.37	0.43	0.85	0.70

6. Assessment of the Cognitive models

With a chosen statistical model, we can then proceed to compare various cognitive models. In the EAPS data, we compared the EAP with known conjunctive skills. The results show that the three-skill cognitive model is better than the one-skill model and the item skill model in terms of BIC. This justifies the use of skills to predict item responses.

Table 7 The model statistics for comparing cogntive models given the conjunctive factor model and penalized maximum likelihood estimation

	NumStudents	NumSkills	LL	BIC
One skill	247	1	-932.54	3696.77
Three skills	247	3	-848.95	3544.37
96 Items	247	96	-961.319	4525.607

7. The P-Matrix

Section 7,8, 9 are the innovative parts of Generalized Learning Factors Analysis to create a better cognitive model. First, corresponding to the Q-Matrix, we propose a new concept call P-Matrix (the Problem Matrix). A Q-matrix is pre-labeled by domain experts before it is put to use by students. A P-matrix is post labeled by domain experts. After domain experts reviewed the student responses data, they may find some items labeled with the same set of skills have various degrees of difficulties. As seen in Table 8, the second and the third item are labeled with the same set of skills. However, the third item is associated with a higher error rate. A further investigation of the item shows that the third item deals with negative numbers, imposing more difficulty for students. Thus, we can create a P-matrix with item as the row and hypothetical difficulty factors as the columns. In this example, we can put "Dealing with negative number" as one difficulty factor. The first two items have zero as the factor value and the third item has 1 as the factor value.

Table 8 A Q-matrix

Item | Skill	Add	Sub	Mul	Div
2*8	0	0	1	0
2*8 – 3	0	1	1	0
2*8 - 30	0	1	1	0

Table 9 A P-matrix

Item | Skill	Dealing with negative number	…

2*8	0	
2*8 - 3	0	
2*8 - 30	1	

How do I find those factors and formulate a sensible P-matrix? One way to do that is to compare the student error rates on the items with the same set of skills. The visual approach is illustrated in Figure 5, where two items labeled with the same skills “Triangle-side” have different success rates. If the difference is large, there may be difficulty factors in the items. A statistical test can be done along with the visual inspection. The p-value of a proportion test, as shown in Figure 6, tests whether the difference of the success rates is significant. In this example, the p-value is close to .05, which suggests the existence of different skills. A further look at the item descriptions reveals the story: although both items are associated with a skill called “Triangle-side” – given the area and the length of one side of the triangle, calculate the length of the other side. Item 1 asks to calculate the length of the base while item 2 asks to calculate the length of the height. Thus, a factor “Base/Height” can be put down into a P-matrix.

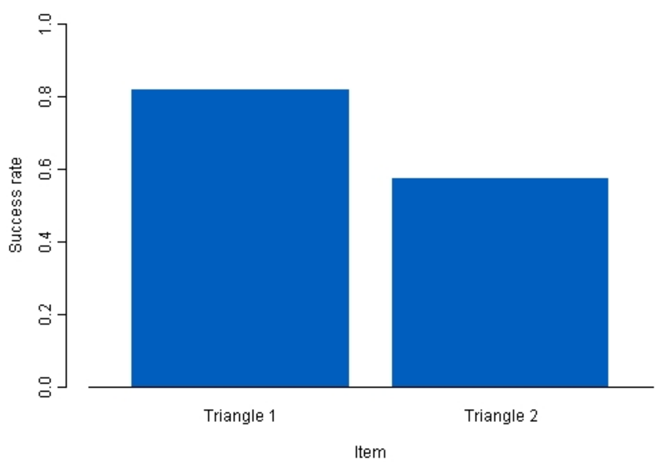

Figure 5 Success rates over two items

```
> x <- c(27, 20)
> n <- c(33, 35)
> prop.test(x, n)

      2-sample test for equality of proportions with continuity
correction

data:  x out of n
X-square = 3.758, df = 1, p-value = 0.0526
alternative hypothesis: two.sided
95 percent confidence interval:
 0.007087852 0.486418641
sample estimates:
 prop'n in Group 1 prop'n in Group 2
         0.8181818         0.5714286
```

Figure 6 Proportion test output from Splus

8. Model operators

The second step to create a better cognitive model is to turn an existing cognitive model, a Q-matrix, and a P-matrix into new models. Model operators perform that function. One model operator is "Split". In the example above, we have a cognitive model with 15 skills. The last of the ten skills is "triangle-side". "Split" splits "triangle-side" into "triangle-side-base" and "triangle-side-height". Now the new model has 16 skills. Since this data set has only single-skilled items, we can run either AFM or CFM, both without the learning term with PMLE on the model and compare their BICs. Shown in Table 10, the new model is not better than the original cognitive model in terms of BIC, suggesting the factor is not necessary.

Table 10 Model statistics after the split

	LL	BIC
Original	-2,003	4,330
After split	-2,000	4,333

The "Split" operator does not change the conjunctivity of the Q-matrix. The other operator "Add" changes the conjunctivity of the Q-matrix. In triangle example above, suppose we have two items – one calculating the area of the triangle given the sides, and the other calculating one side of the triangle, given the area and the other side. Instead of having these two separate skills for different scenarios, researchers can formulate one skill for the triangle area and the other skill as the algebra manipulation, which is to turn the triangle area formula into triangle side formula. To solve the first item, a student only needs the first skill. To solve second item, a student needs to use the first skill and the second skill.

9. Model search

If we have several difficulty factors in a P-matrix, we can do the third step in improving a cognitive model. A distinguishing feature of the LFA method is its semi-automatic model search process. We formulated finding a better cognitive model as a combinatorial search problem. Given an existing cognitive model, a Q-matrix, a P-matrix, the LFA method automatically incorporates those factors into models, and finds new models that researchers may wish to investigate further.

The search algorithm in LFA is a best first search (Russell & Norvig, 2003). It starts from an initial node, iteratively creates new adjoining nodes, and explores them to reach a goal node. Difficulty factors are incorporated into an existing IRT model through splits or adds. To limit the search space, it employs a heuristic to rank each node and visits the nodes in the order of this heuristic estimate. Measuring both the model fit and the model complexity, Bayesian Information Criterion (BIC) (Wasserman, 2004) are the heuristics used in the search. As shown in Figure 7, at the beginning of a search with BIC as the heuristic, the original model is evaluated and BIC is computed. Then the model is split into a few new models by incorporating the factors. BICs are computed from each of the new models. The search algorithm chooses the best one (the shaded node with value 4301) for the next model generation. The search algorithm does not always move to a lower level in the search hierarchy. It may go up to select a model (the shaded node with value 4212) to expand if all the new models have worse heuristic scores than the previous model had. After several expansions, it finds a best model with the lowest BIC value within all the models searched.

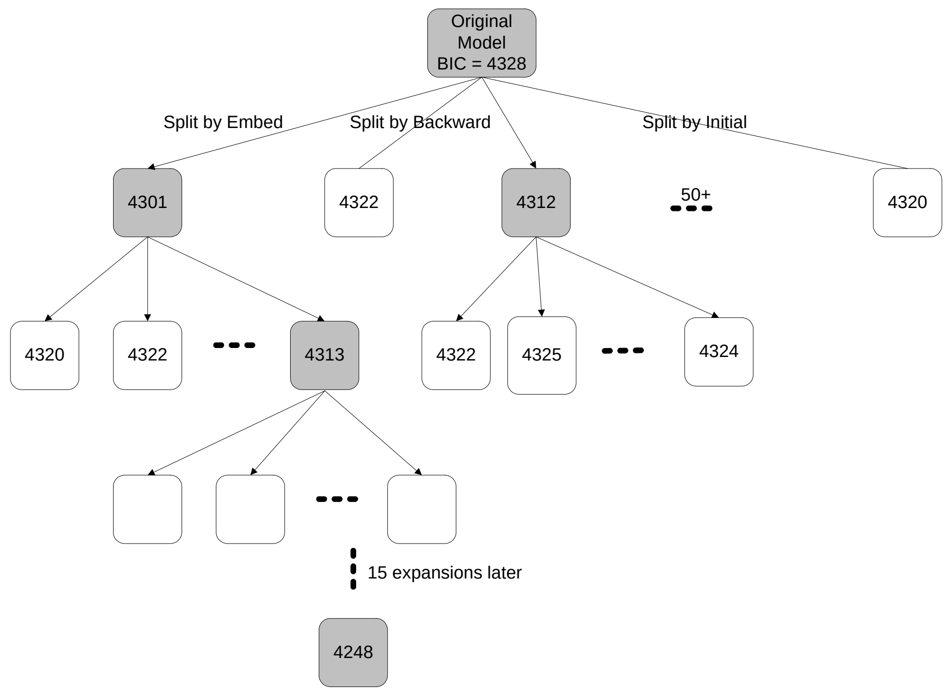

Figure 7 A best-first search through the cognitive model space

IV. Contributions

Generalized Learning Factors Analysis draws strengths from different fields and contributes to different fields.

In machine learning, this method combines statistical modeling and AI search and provides a fast parameter fitting and model search algorithm over large data sets.

In statistics, it develops a set of unified psychometric models for student learning and assessment, and demonstrates robust measures for those models.

In education, it provide a powerful tool for education researchers to estimate the amount of practice needed (Cen et al., 2007), test learning transfer (Leszczenski, 2007), and model student learning behaviors (Nwaigwe, Koedinger, VanLehn, Hausman, & Weinstein, 2007).

Table 11 compares the LFA method with earlier approaches.

Table 11 Comparing LFA to other methods.

Task |Features	IRT-based, account for student differences	Handle conjunctive skills	Automatic search for better models	Applicable to learning data	Discover factors that are directly interpretable

DiBello et al.'s models	Yes	Yes			
Q-Matrix	Yes	Yes	Yes		
Draney, et al.'s model	Yes			Yes	
Gordon's EPCA		Additive		Yes	Depends
LFA	Yes	Yes	Yes	Yes	Yes

V. What has Been Done

This thesis is built upon my early research since I entered CMU. Following is what I have finished.

1) built the parameter estimation procedures for the statistical models and the combinatorial search algorithm
2) Shown AFM well predicting student performance on the geometry learning data and applied to reduce over practice in the real curriculum
3) Shown CFM perform better than AFM on simulated conjunctive-skilled data and real EAP assessment data

VI. What has Not Been Done

I plan to finish the following:

1) Search for a convincingly better and interpretable cognitive model by combining CFM with the search on assessment data.

   In this step, I will use the EAPS assessment data. This data set has a known conjunctive structure and we have a list of hypothetical factors to search through.
2) Use Exponential PCA to explore latent factors

   EPCA can identify a linear combination of items that explain the variance in the data. These combinations may well serve as a set of factors. I plan to first to find those factors and compare their performance with the expert-labeled factors.
3) Implementing a faster parameter estimation and model searching algorithm for its coming uses with large data sets.

   I plan to fix student parameter constant throughout the search or fix student parameters in each node expansion.

   The other approach is to use Exponential PCA to guide the search. For example, the factors in the P matrix can be chosen in preference to other factors if the chosen factors have higher correlation with the EPCA factors.
4) Combine CFM with the search on learning data (optional if data available)

VII. Timeline

Dec. 2007

1) Combine CFM with the search on the EAP data

2) Write up a conference paper on CFM
3) Use Exponential PCA to explore latent factors
4) Compare the performance of EPCA factors with expert prelabeled cognitive model and with LFA improved cognitive model
5) Implement faster parameter estimation and model searching

Jan – Feb. 2008 Write up the thesis draft and get feedback. Refine early work.

Mar. - May. 2008 Revise and defend the thesis.

.

VIII. Personal Qualifications

Several external factors and internal factors lead me to a good candidate for this piece of work.

1) Strong and diversified committee members
2) Finished the ML core curriculum & took/audited 80% of the master level statistics courses
3) Written half a million lines of Java code and command over 5 core libraries on data structure, numerical, ML
4) Genuine interest in contributing to the boundary between machine learning and human learning

References

Barnes, T. (2005). *The Q-matrix Method: Mining Student Response Data for Knowledge*. Paper presented at the American Association for Artificial Intelligence 2005 Educational Data Mining Workshop.

Cen, H., Koedinger, K., & Junker, B. (2005). *Automating cognitive model improvement by a* search and logistic regression*. Paper presented at the American Association for Artificial Intelligence 2005 Workshop on Educational Datamining.

Cen, H., Koedinger, K., & Junker, B. (2006). *Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement*. Paper presented at the 8th International Conference on Intelligent Tutoring Systems.

Cen, H., Koedinger, K., & Junker, B. W. (2007). *Is Over Practice Necessary? – Improving Learning Efficiency with the Cognitive Tutor through Educational Data Mining (submitted)*. Paper presented at the 13th International Conference on Artificial Intelligence in Education

Corbett, A. T., & Anderson, J. R. (1995). *Knowledge tracing: Modeling the acquisition of procedural knowledge*. Paper presented at the User Modeling and User-Adapted Interaction.

DiBello, L., Stout, W., & Roussos, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. Nichols, S. Chipman & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale , NJ: Erlbaum.

Draney, K., Pirolli, P., & Wilson, M. (1995). A Measurement Model for a Complex Cognitive Skill. In *Cognitively Diagnostic Assessment*. Hillsdale, NJ Erlbaum.

Embretson, S. (1997). Multicomponent Response Models. In W. V. D. Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* Springer.

Fischer, G. H. (1977). Linear logistic test models: Theory and application. In H. Spada & W. F. Kempf (Eds.), *Structural Models of Thinking and Learning* (pp. 203-225): Bern: Huber.

Gordon, G. (2002). *Generalized2 Linear2 Models*. Paper presented at the Annual Conference on Neural Information Processing Systems 2002.

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*: Springer.

Koedinger, K., & Nathan, M. J. (1997). *Teachers' notions of students' algebra problem-solving difficulties.* Paper presented at the the annual meeting of the James S. McDonnell Foundation Program for Cognitive Studies for Educational Practice.

Koedinger, K. R., & Anderson, J. R. (1997). Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education*(8), 30-43.

Koedinger, K. R., Corbett, A. T., Ritter, S., & Shapiro, L. J. (2002). *Carnegie Learning's Cognitive Tutor™: Summary Research Results*: Available from Carnegie Learning, Inc., http://www.carnegielearning.com/approach_research_reports.cfm.

Koedinger, K. R., & MacLaren, B. A. (2002). *Developing a Pedagogical Domain Theory of Early Algebra Problem Solving*: Human Computer Interaction Institute, Carnegie Mellon University.

Koedinger, K. R., & Nathan, M. J. (2003). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*.

Leszczenski, J. M. (2007). *Learning Factors Analysis Learns to Read (Thesis).* Carnegie Mellon University, Pittsburgh PA, 15213.

Nathan, M. J., & Koedinger, K. R. (2000). Teachers' and researchers' beliefs about the development of algebraic reasoning. *Journal for Research in Mathematics Education, 31*, 168–190.

Nathan, M. J., & Koedinger, K. R. (2003). An investigation of teachers' beliefs of students' algebra development. *Cognition and Instruction, 18*(2), 207–235.

Nathan, M. J., Koedinger, K. R., & Alibali, M. W. (2001). *Expert blind spot: When content knowledge eclipses pedagogical content knowledge.* Paper presented at the the Third International Conference on Cognitive Science, Beijing, China.

Nathan, M. J., Long, S. D., & Alibali, M. W. (2002). The symbol precedence view of mathematical development: A corpus analysis of the rhetorical structure of algebra textbooks. *Discourse Processes, 33*(1), 1–21.

Nathan, M. J., & Petrosino, A. (2003). Expert Blind Spot Among Preservice Teachers. *American Educational Research Journal, 40*(4), 905–928.

Nathan, M. J., Stephens, A. C., Masarik, D. K., Alibali, M. W., & Koedinger, K. R. (2002). *Representational fluency in middle school: A classroom-based study.* Paper presented at the the 24th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.

Nwaigwe, A., Koedinger, K., VanLehn, K., Hausman, R., & Weinstein, A. (2007). *Exploring alternative methods for error attribution in learning curves analyses in intelligent tutoring systems*. Paper presented at the 13th International Conference on Artificial Intelligence in Education

Russell, S., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2nd ed.): Prentice Hall

Sarkis, H. (2004). *Cognitive Tutor Algebra 1 Program Evaluation*: Available from Carnegie Learning, Inc., 1200 Penn Avenue, Suite 150, Pittsburgh, PA 15222.
Tatsuoka, K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*(4), 345-354.
von Davier, M. (2005). *A General Diagnostic Model Applied to Language Testing Data*: Educational Testing Service.
Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*: Springer