Using Data Mining Findings to Aid Searching For Better Skill Models

Mingyu Feng[1], Neil T. Heffernan[2], Kenneth Koedinger[3],

[1] SRI International, Menlo Park, CA 94025
[2] Worcester Polytechnic Institute, Worcester, MA 01609
[3] Carnegie Mellon University, Pittsburgh, PA 15213

{mingyu.feng@sri.com, nth@wpi.edu, koedinger@cmu.edu}

Abstract. One key component of creating an intelligent tutoring system is forming a model that monitors student behavior. Researchers in machine learning area have been using automatic/semi-automatic techniques to search for skill models. One of the semi-automatic approach is learning factor analysis (LFA, Cen, Koedinger & Junker, 2006), which involves human making hypothesis and identifying difficulty factors in the related items. In this paper, we propose a hybrid approach in which we leverage findings from our previous educational data mining work to aid the search for a better skill model and thus, improve the efficiency of LFA. Preliminary results suggest that our approach can lead to significantly better fitted skill models fast.

Keywords: Data mining, transfer model, learning factor analysis.

1 Introduction

One key component of creating an intelligent tutoring system (ITS) is forming a model that monitors student behavior. An ITS needs the construction of complex models to represent the skills that students are using and their knowledge states. As students work through the program, the model tracks their progress and chooses what problems will be displayed next. By using a better skill model, a system should be able to do a better job of predicting which items students will get correct in real-time. That means the system can do a better job of selecting the next best item for students to work on. For instance, one criterion of the next "best" item could be the one that has the largest ratio of expected test-score gain to expected time to complete the problem where expected test score gain will be a function that depends upon both the expected rise in skills from doing that item at that time, as well as the weight of those skills on the test. A better model would also help to address the issues as we mentioned above to help teachers adjust their instruction in a data-driven manner. Such a model will allow a teacher who has one week before a high-stakes statewide test know what topics to review to maximize the class average. We can make a calculation averaging the whole class to suggest what will give the teacher the biggest "bang for the buck."

Given the importance of transfer models, it is not surprising that their construction and improvement has been a major focus in the community. Researchers in machine learning area have been using automatic/semi-automatic techniques to search for skill models. Tatsuoka and colleagues developed the rule space method (Tatsuoka, 1990, 1993) in which hypothesized expert rules and actual student errors in fraction addition can be mapped and compared. The expert point that is closest to the student response is assumed to be the rule that the student is using. Barnes has done considerable work with trying to induce transfer models, in this work called Q-matrices (Birenbaum, Kelly, & Tatsuoka, 1993), from data (Barnes, 2005, 2006). Koedinger and colleagues (Koedinger & Junker, 1999; Cen, Koedinger & Junker, 2005, 2006) proposed a semi-automatic approach called Learning Factor Analysis (LFA) as a generic solution to evaluate, compare, and search through potential cognitive models of learning. Pavlik, Cen, & Koedinger (2009) proposed a method called learning factors transfer analysis to automatically generate domain models. Ferguson, Woolf, & Mahadevan (2009) developed a method to use transfer learning to guide the improvement of skill models. They hand-coded the transfer features in problems and thus constructed a hierarchical transfer learning model as an improvement of existing flat skill model. Ritter et al. (2009) addressed the issue of model improvement by investigating the nature of the space of parameters from knowledge tracing. They used a k-means clustering to drastically reduce the parameter space used to model students from 2,400 skills to 23 clusters without compromising the behavior of the system (the Cognitive Tutors).

In this paper, we propose a hybrid approach in which we leverage findings from our previous educational data mining work to aid the search for a better skill model using LFA.

2 Methods

2.1 A Generic Method: Learning Factor Analysis (LFA)

Cen, Koedinger & Junker (2005, 2006) proposed a generic, computation intensive method called learning factor analysis (LFA) for cognitive model evaluation and refinement. LFA was initially conceptualized by Koedinger and Junker (1999). It aims to "combine statistics, human expertise and combinatorial search to evaluate and improve a cognitive model". LFA has three parts: a statistical model that evaluates how a cognitive model fit the data; difficulty factors associated with problems; and a group of operators that could be applied to manipulate current cognitive models based on the difficulty factors to generate a combinatorial search space of models.

The statistical model is an extension of the power law of learning (Newell & Rosenbloom, 1993), which describes the error rates decrease exponentially according to a power function as the number of opportunities to practice a skill increase. The power law applies to one particular student and over only one skill while LFA models multiple students and multiple skills by adding in student and skill intercepts and skill learning rates, typically using multiple logistic regression models (e.g., Cen, Koedinger & Junker, 2006).

In LFA, a difficulty factor is a hidden feature in a problem that makes the problem easier or harder to solve. It is usually identified by subject exerts based up instruction theory and task analysis. An example factor in math with two possible values is using a rule CIRCLE-AREA (e.g. $S = \pi * r^2$) *forward* (to calculate circle area given radius) or *backward* (to calculate radius given circle area). Here *forward* and *backward* are values of a difficulty factor.

LFA performs heuristic search over a search space where each state is a new cognitive model to locate the best one. Given the difficulty factor, LFA could apply one of the three operators “split”, “add”, and “merge” on skills in current based model to generate sub-models. By applying operator “add” to existing cognitive model, it is hypothesized that there is an unrepresented skill required by the items that are associated with a difficulty factor. Therefore, a new skill shall be added in and tagged to the items. The “merge” operator assumes students only need one representation for multiple skills; yet the “split” operator on the contrary hypothesizes multiple representations be used to represent the variation in one piece of knowledge component.

Various heuristics such as AIC, BIC, R-square and Log likelihood, have been considered as model evaluation and selection measures.

2.2 Findings from Previous Research

In our previous work (Feng, Heffernan & Beck, 2009), we conducted a focused item-level analysis of a subset of items to track how student performance on these items changed during the same ASSISTment session. We wanted to see if we can tell which item in a group is the most effective at causing learning.

The hypothesis was that students learn from groups of items that share the same background knowledge requirement. Our subject manner expert picked 181 items out of the 300 8^{th} grade (approximately 13 to 14 years old) math items in ASSISTments. Items that have same deep features or knowledge requirements, such as approximating square roots, but have different surface features, such as cover stories, were organized into a **Group of Learning OPportunity (GLOP).** The selected 181 items falled into 38 GLOPs with the number of items in each GLOP varied from 2 to 11. The items were a fair sampling of the curriculum and cover knowledge from all of the five major content strands identified by the Massachusetts Mathematics Curriculum Framework. Items in the same group were collected into the same section of ASSISTments, and seen in random order by students. Each student potentially saw 38 different GLOPs that involve different 8^{th} grade math skills (e.g. fraction-multiplication, inducing-functions, symbolization articulation) in random order. Figure 1 shows four items in one GLOP that were about the concept “Area” where all these problems asked students to compute the area of the shaded part in the figures.

We collected data for this analysis during Oct. 31, 2006 to Oct. 11, 2007 from 2,502 8^{th} grade students mostly from Worcester, Massachusetts area. Each student on average worked on 22 items.

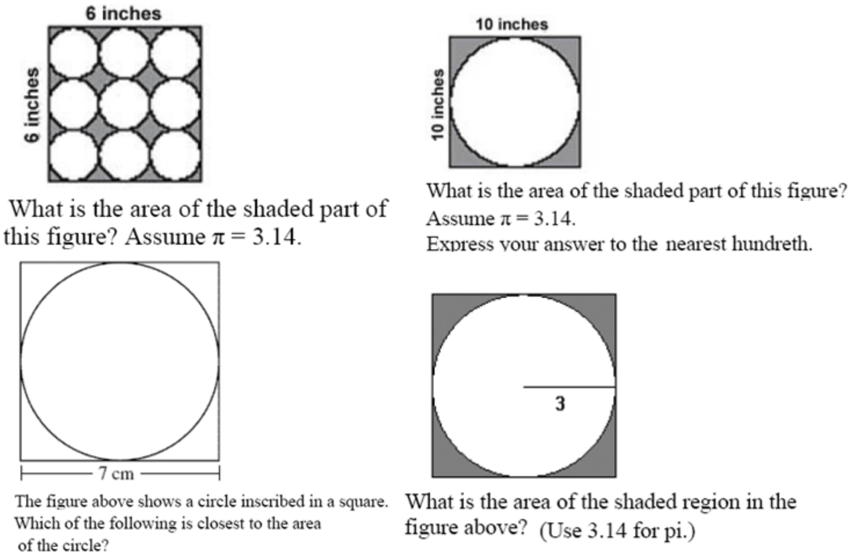

Fig. 1. Items from a sample GLOP that addresses the knowledge about area

We first attempted to determine whether the system effectively teaches (Feng, Heffernan, Beck, & Koedinger, 2008). Learning is assessed by comparing student performance the first time they were given one item from a GLOP with their performance when they were given more items (also more opportunities) from the same GLOP in the same day. In Feng et al. 2009, we reported how we could reliably tell which item is most effective at causing learning using learning decomposition (Beck, 2006). We found out that the items in ASSISTments vary in their instructional effectiveness in helping student learn the skill(s) associated with a GLOP. Some items in ASSISTments caused significant learning while some other items were not as useful at promoting learning.

2.3 Aiding LFA Search Using Data Mining Findings

Now that we could reliably tell difference of learning among items, we wanted to employ this information to improve existing cognitive models which improve the overall predictive power of the system and potentially better understand and increase student learning.

As a basis of LFA, the identification of difficulty factors needs human expertise. They have always been found by subject experts through a process of “difficulty factor assessment” (DFA, Koedinger, 2000). Based upon theory or task analysis, researchers hypothesized the likely factors that cause student difficulties, and by assessing performance difference on pairs of problems that vary by only one factor, the experts identified the hidden knowledge component that could be used to improve a skill model. Because of this phase of human making hypothesis and identification, LFA becomes a semi-automatic approach, although intuitively it is appealing as a fully automated method.

Can we raise efficiency of LFA by suggesting difficulty factors automatically yet still get better models? Feng et al. (2009) showed certain items in a random sequence cause significantly less learning than others. Intuitively, it is highly possible that there is certain factor inherited in the items, which make it harder for the learning from this item to transfer to later items. This could be either because later items demand more skills than the current one, or because what a student learns from a current item does not help later items. In both conditions, there is probably “mis-tagging” with this item. Presumably, such a factor can be utilized by LFA to manipulate the original skill model to search for the best-fit model. What we really hope to see is that having a human expert sitting in front of a computer, with the help of our educational data mining results, she can quickly determine what factors each item may have. Before doing that, we want to check to see if our results could be used to make suggestions on factors and whether there was some validity in the approach.

In order to test this idea, we create factor tables for all the GLOPs. In each table, we use one factor with two values “High” and “Low” indicating the effectiveness of the items. The item that has caused least learning is associated with “Low” while all other items are associated with “High”. Table 1 shows the factor table we created for GLOP 1 together with the skill that is currently tagged with the items in the GLOP and the learning coefficient associated with each item. Noticing that all items are tagged with the same skill “Interpreting-Circle-graph” as they all belong to one GLOP.

Table 1. Assigning factor to GLOP 1 based on learning coefficients

GLOP ID	*Item ID*	*Skill*	*Factor*	*Coefficient learned from previous work (higher for better learning effect)*
1	1022	Interpreting-Circle-graph	High	0.464
1	1660	Interpreting-Circle-graph	High	0.414
1	1045	Interpreting-Circle-graph	High	0.127
1	1649	Interpreting-Circle-graph	Low	-0.176

2.4 Results

Results on the GLOPs. Given the factor tables, we ran LFA search over all the GLOPs. Following Cen, Koedinger & Junker (2005, 2006), BIC was used as the heuristic to evaluate the models in that it balances simplicity and predictive power of models. Among the 38 GLOPs we have examined, LFA was able to find statistically significantly better models (a difference of 10 points or more on BIC) for 12 of them, using the factors as assigned in the factor tables. Among the 12 GLOPs, 5 of them included 2 items; 3 included 4 items; the rest 4 GLOPs had 5, 6, 8, and 9 items respectively. For 11 out of the 12 GLOPs, the application of the “add” operator leaded to a better fitted model, which suggests that there were more knowledge other than the current skills that needed to be represented in the cognitive model in order to better track student learning.

A sanity check with randomization assignment of learning factors. To get a feeling of how well the suggested factors do, we conducted a simple sanity check where WE randomly assign one item each GLOP with the “Low” value of the factor, and then run the same searching process as before. Obviously, for the 2-item GLOPs, the results will be the same as before. But for GLOPs with more items, the search process using randomly assigned factor values only find better models for 2 out of the 27 GLOPs, which makes our previous results of 7 out of 27 somewhat impressive.

3 Discussion and Future Work

We admit that there are other ways of assigning values in the factor table. Yet, we are also glad to see that the results show some validity for the very simple way of suggesting factors. We are especially happy to see that when there are more than 5 items in a GLOP, this method can still help find better models for 4 out of the 15 GLOPs.

This work is preliminary despite of the inspiring results, in that the amount of data we have applied this method to is very limited. We would like to apply this approach on data collected from other tutoring systems to verify the generality as well. Moreover, we do realize that using human experts’ suggested factors would be another control condition to compare to. But considering the amount of efforts and time that needed to be spent on difficulty factor assessment, maybe it would not be a totally “fair” comparison. Another reasonable study would be to run a randomized controlled study to compare two conditions where in one condition, human experts use solely DFA to identify factors, while in the other condition, human experts are provided the item level tutoring effectiveness results as we show in Section 2. The study should be controlled for time, and then controlled for groups to examine on what aspects the results can be helpful.

4. Conclusion

This paper describes one practice on how to use the educational data mining findings to help improve cognitive modeling in ASSISTments, following our previous effort on detecting item effectiveness. A semi-automatic approach, i.e. learning factor analysis, is considered. Preliminary results show our findings in item effectiveness from educational data mining can be used to assign difficulty factors for, and thus, automate LFA so that it can efficiently search for superior cognitive models.

References

1. Barnes, T., (2005). Q-matrix Method: Mining Student Response Data for Knowledge. In Beck. J (Eds). *Educational Data Mining: Papers from the 2005 AAAI Workshop*.

2. Barnes, T. (2006). Evaluation of the q-matrix method in understanding student logic proofs. *Proceedings of the 19th International Conference of the Florida Artificial Intelligence Research Society* (FLAIRS 2006), Melbourne Beach, FL, May 11-13, 2006.

3. Beck, J.E. (2006). Using learning decomposition to analyze student fluency development. *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems*. Jhongli, Taiwan. pp. 21-28.

4. Birenbaum, M., Kelly, A., & Tatsuoka, K. (1993). Diagnosing knowledge states in algebra using the rule-space model. *Journal for Research in Mathematics Education*, 24(5), 442-459.

5. Cen. H., Koedinger K., & Junker B. (2005). Automating Cognitive Model Improvement by A*Search and Logistic Regression. In Beck. J (Eds). *Educational Data Mining: Papers from the 2005 AAAI Workshop*.

6. Cen, H., K. Koedinger, and B. Junker. (2006). *Learning factors analysis: a general method for cognitive model evaluation and improvement.* Presented at the Eighth International Conference on Intelligent Tutoring Systems (ITS 2006), Jhongli, Taiwan.

7. Ferguson, K., Woolf, B., & Mahadevan,S. (2009). Transfer learning and representation discovery in intelligent tutoring systems. In Dimitrova, Mizoguchi, du Boulay, & Graesser (Eds), *Artificial Intelligence in Education*. Amsterdam, Netherlands: IOS Press. pp. 605-607.

8. Feng, M., Heffernan, N., & Beck, J. (2009). Using learning decomposition to analyze instructional effectiveness in the ASSISTment system. In Dimitrova, Mizoguchi, du Boulay, & Graesser (Eds), *Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED-2009).* Amsterdam, Netherlands: IOS Press. Brighton, UK.

9. Feng, M., Heffernan, N., Beck, J, & Koedinger, K. (2008). Can we predict which groups of questions students will learn from? In Beck & Baker (Eds.). *Proceedings of the 1st International Conference on Education Data Mining*. Montreal, 2008.

10. Koedinger, K. & Junker, B. (1999). Learning factor analysis: Mining student-tutor interactions to optimize instruction. Presented at Social Science Data Infrastructure Conference. New York University. November, 12-13, 1999.

11. Koedinger, K. (2000). Research statement for Dr. Kenneth R. Koedinger, June, 2000. Retrieved from http://pact.cs.cmu.edu/koedinger/koedingerReserach.html.

12. Nathan, M. J., and Koedinger, K. R. (2000). An investigation of teachers' beliefs of students' algebra development. *Cognition and Instruction*, 18(2), 209-237.

13. Newell, A, & Rosenbloom, P.S. (1993). Mechanisms of skill acquisition and the law of practice. In P. S. Rosenbloom, J. E. Laird, & A. Newell (Eds.), *The Soar Papers: Research on integrated intelligence*. Cambridge, MA: MIT Press.

14. Pavlik, P.I., Cen, H. & Koedinger, K.R. (2009a). Performance factors analysis – A new alternative to knowledge tracing. In Dimitrova, Mizoguchi, du Boulay, & Graesser (Eds), *Artificial Intelligence in Education*. Amsterdam, Netherlands: IOS Press. pp.531-538.

15. Pavlik, P.I., Cen, H., & Koedinger, K.R. (2009b). Learning factors transfer analysis: Using learning curve analysis to automatically generate domain models. In Barnes, Desmarais, Romero, & Ventura (Eds.). *Proceedings of the 2nd International Conference on Educational Data Mining*. pp.121-130. Cordoba, Spain.

16. Ritter, S., Harris, T., Nixon, T., Dickison, D., Murray, C., Towle, B. (2009). Reducing the knowledge tracing space. In Barnes, Desmarais, Romero, & Ventura (Eds.). *Proceedings of the 2nd International Conference on Educational Data Mining*. pp. 151-160. Cordoba, Spain.

17. Tatsuoka, K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Education Measurement*, 20(4), 345-354.

18. Tatsuoka, K.K. (1990). Toward an Integration of Item Response Theory and Cognitive Error Diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto, (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). Hillsdale, NJ: Lawrence Erlbaum Associates.