Towards IRT-based student modeling from problem solving steps

Manuel Hernando
Universidad de Málaga
Bulevar Louis Pasteur, 35.
29071 Málaga, Spain
(+34) 952 132 863
mhernando@lcc.uma.es

Sergey Sosnovsky
German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
(+49) 681 85775 5367
sergey.sosnovsky@dfki.de

Eduardo Guzmán
Universidad de Málaga
Bulevar Louis Pasteur, 35.
29071 Málaga, Spain
(+34) 952 137 146
guzman@lcc.uma.es

Eric Andres
German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
(+49) 681 85775 5367
eric.andres@gmail.com

Susanne Narciss
Technische Universität Dresden
Psychologie des Lehrens und Lernens
01062 Dresden, Germany
(+49) 351 4633 6059
susanne.narciss@tu-dresden.de

ABSTRACT

In this research, we use Item Response Theory based model for computing procedural knowledge of a sample of primary school children solving fraction addition exercises. For each exercise, the model needs to automatically construct a solution graph. We have explored different strategies for building such graphs and the effects they have on the quality of the model predictions. The results obtained shed light on the applicability of Item Response Theory for the task of measuring procedural skills and provide recommendations on the choice of IRT model adjustment.

Keywords

Student Modeling, Item Response Theory, Problem Solving, Procedural Knowledge.

1. INTRODUCTION

Intelligent tutoring systems (ITS) are designed to provide individualized computer-supported learning. One of the most important characteristics of a good ITS is a high-quality student modeling component, that maintains representation of student knowledge and helps the ITS to support personalized tutoring helping each student improve her knowledge in the optimal way.

High-quality student modeling starts with accurate knowledge assessment. The classical approach to infer procedural knowledge is based on exposing students to problem solving, as it is the most natural way for a student to demonstrate procedural skills.

In the field of testing, Item Response Theory (IRT) [2]is known to provide accurate and invariant measurement of declarative knowledge. In [5], we have proposed a model that employs IRT for procedural knowledge assessment and can be used in problem solving environments. As a part of this approach, dynamic problem solution graphs are automatically constructed from student logs. Such graphs are updated and improved every time a new student interaction with a target exercise has been registered. The work presented in this paper explores different alternatives for constructing the graphs, and analyses how various evidence aggregation techniques influence the quality of the resulting IRT models and the accuracy of knowledge assessment they support.

2. PROCEDURAL ITEM RESPONSE THEORY

There are three types of IRT-based models, according to how they score student responses to the test items (questions) and update student knowledge [3]: dichotomous, polytomous, and quasipolytomous models. Dichotomous models consider only two scores per item (correct/incorrect); polytomous models assume different scores for different answers, thus, being more informative than dichotomous models, but requiring more data to calibrate [2]; quasipolytomous models [4] are halfway between dichotomous and polytomous: some possible answers have their own scores and others are clustered into aggregate options.

The process of solving a multistep learning problem can be represented as a graph that contains all the steps and actions a student could perform, where nodes correspond to the states of the solution process and the arcs to the actions of a student transitioning her from one state another. In this work, instead of using pre-constructed graphs, we data-mine individual problem solution graphs from the student activity logs.

The procedural IRT mode makes an analogy between problem solving and testing by considering a students' path through the process of solving a multistep learning problem as a testing sequence. Each node could be understood as an item and each step as an item response.

3. PROBLEM SOLVING ENVIRONMENT AND DATA USED

The data used in this study comes from the controlled experiment conducted in Spring of 2012 in Dresden (Germany) with 6th- and 7th-grade pupils. Students had to solve simple fraction problems in the computer-based learning environment ActiveMath [6][7]. The overall experiment contained several phases and covered several topics of fraction arithmetic. In this paper, we have focused on multistep problems on "*Adding Fractions with Unlike Denominators*" that students were solving during the posttest phase of the experiment. After filtering out subjects who did not manage to try the target set of problems, we have 61 students (25 males and 36 females) contributing to the final dataset.

The problems were based on an interface allowing students to construct individual solution paths by providing structured templates for intermediate steps [1]. While solving a problem, a student could choose a type of the operation to perform on the next step and then fill in the corresponding template. Only students defined the number and sequence of steps that they needed to reach the final solution.

4. EXPERIMENTAL MODELS

We have explored three different strategies to generate problem solution graph from the log data. First, we have applied our approach in a straightforward way – by generating one graph per problem without any aggregation and applying the IRT to this graph. We have called this model *Direct Application* (DA). The second model seeks to increase the supporting evidence per single steps by merging the states that represent the same semantic operation in a problem solution graph. We call this model *Semantic Operation* (SO). Finally, the *Common Graph* (CG) model logically develops the approach of the SO model by aggregating semantically equivalent operations across problems. As a result, a single graph is constructed to represent the entire subset of isomorphic problems related to "*Adding Fractions with Unlike Denominators*".

5. EVALUATION

We have used two sets of problems in this research: the *target set* consists of three multi-step problems on adding fractions with unlike denominators; the *assessment set* contains 13 one-step problems on fraction expansion, fraction reduction and adding fractions with a common denominator.

In order to evaluate the quality of each model in terms of its predictive validity, we compare the obtained estimates with the knowledge scores students achieve on the *assessment problem set*. These scores are also computed using the IRT approach. Each of the 13 *assessment set* problems is a single-step problem, therefore it corresponds to a single test item. We have looked into which model produces better predictions of student knowledge assuming that a better model will be closer to the control assessment.

We have used different quasipolytomous models depending on the supporting threshold of arcs (understanding threshold as the minimum acceptable support of steps) being a threshold = 1 a pure polytomous model and the maximum threshold a pure dichotomous.

Table 1 shows the results of our experiments, the two columns contain the maximum and the minimum values for Pearson's correlation (*r*). The values depend on the support threshold chosen for a particular quasipolytomous IRT setup as described above. Essentially, all models produce knowledge predictions that are significantly positively correlated with the controlled assessment. In all three cases, the maximum correlation effect size is rather high; however, the difference between the straightforward *DA* model and the SO/CG models semantically aggregating students' results is considerable.

Table 1. Correlations of the scores on the assessment test and the target tests produced by the experimental models

Model / Test	r_{max} (threshold)	r_{min} (threshold)
DA	.42 (9)	.27 (15)
SO	.54 (5)	.34 (21)
CG	.51 (3)	.35 (90)

6. CONCLUSION

In this paper we have studied different strategies to elicit the problem solving graph for assessing the student procedural knowledge with an IRT-based model. We have distinguished three different strategies: building a graph directly from student behavior graph, building the graph grouping states by semantic operations, and building a graph that represents more than a single problem. Results suggest that all of the strategies could be valid to infer procedural knowledge but we get better results when we group some states. However, when we use the same graph for more than a problem we have not obtained any advantage, even SO model obtains better results.

The use of IRT in a problem-solving environment for assessing procedural ensures that the results obtained are invariant and well-founded, since they are computed using data-driven statistical procedures. Results of our work are promising but we should to test them for larger student samples.

REFERENCES

[1] Andres, E., Sosnovsky, S., Schnaubert, L., Narciss, S. 2013. Using Fine-Grained Interaction Data to Improve Models of Problem Solving. In proceedings of *DAILE Workshop*, 4th STELLAR Alpine Rendez-Vous.

[2] Embretson, S.E., Reise, S.P. 2000. Item response theory for psychologists. Lawrence Erlbaum, Mahwah.

[3] Guzmán, E., Conejo, R., de-la Cruz, J.L.P. 2007.Adaptive testing for hierarchical student models. *User Model. and User-Adapt. Interact.*, 17, 119-157.

[4] Hernando, M., Guzmán, E., Conejo, R. 2013. Validating Item Response Theory Models in Simulated Environments. In *Proceedings of the AIED Workshop on Simulated Learners*, 41-50. Memphis, TN, U.S.A.

[5] Hernando, M., Guzmán, E., Conejo, R. 2013. Measuring Procedural Knowledge in Problem Solving Environments with Item Response Theory. In *Proceedings of the 16th Int. Conf. on AI in Education*, 653-656. Memphis, TN, U.S.A.

[6] Melis, E., Goguadze, G., Homik, M., Libbrecht, P., Ullrich, C., Winterstein, S. 2006. Semantic-aware components and services of ActiveMath. *Brit. J. Educ. Technol.*, 37(3), 405–423.

[7] Sosnovsky, S., Dietrich, M., Andrès, E.,Goguadze, G., Winterstein, S., Libbrecht, P., et al. 2013. Math-bridge: bridging the gaps in European remedial mathematics with technology-enhanced learning. In Using tools for learning mathematics and statistics, 437–451. Berlin: Springer.