Student Knowledge Diagnosis Using Item Response Theory and Constraint-Based Modeling

Jaime GALVEZ [a,1], Eduardo GUZMAN [a], Ricardo CONEJO [a], Eva MILLAN [a]
[a] *Dpto. de Lenguajes y Ciencias de la Computación,*
Universidad de Málaga, Spain

Abstract. One of the most popular student modeling techniques currently available is Constraint Based Modeling (CBM), which is based on Ohlsson's theory of learning from performance errors. It focuses on the domain principles to correct faulty knowledge and assumes that a student will reach a correct solution without violating these fundamental domain concepts. However, even though this is a powerful and computationally simple technique, most student models of CBM-based tutors handle simple long-term models or based on heuristics to quantitatively estimate the knowledge measured. In this paper we propose a student knowledge diagnosis model which combines CBM with the Item Response Theory (IRT). IRT is a probabilistic and data-driven theory which guarantees accurate and invariant student knowledge estimations. By means of this synergy between CBM and IRT we suggest the construction of long-term student models composed of the estimations of their knowledge. This paper also includes an experiment we have carried out with real students, which explores the validity of the diagnoses made with our model.

Keywords. Constraint Based Modeling, Item Response Theory, Student Modeling

1. Introduction

In general, Intelligent Learning Systems include mechanisms to diagnose the student's knowledge level in order to suggest the most appropriate action and to facilitate subsequent learning. Thus, the learning process is personalized to each student's needs and preferences. To achieve this, every system needs a student model with an estimation of the learner's knowledge. Self [1] pointed out that this task is intractable, incomplete and, for this reason, inaccurate models are generally accepted.

In this sense, one of the approaches which reduces the complexity of the modeling task is the Constraint Based Modeling (CBM) technique. This is one of the most successful approaches for student modeling, a fact that has been demonstrated through its many tutor implementations [2]. CBM reduces the complexity of modeling by using the domain principles as a tool to detect student faulty knowledge. Besides this, the low computational requirements and the easy application of this technique, in comparison

[1] Corresponding Author. Email addresses: jgalvez@lcc.uma.es (J. Gálvez), guzman@lcc.uma.es (E. Guzmán), conejo@lcc.uma.es (R. Conejo), eva@lcc.uma.es (E. Millán)

to other approaches such as Model Tracing [3], makes it a very powerful technique for building learning tutors.

CBM-based student model implementations rely on short-term models composed of all the student errors, that is, they represent qualitative knowledge estimation. Although estimation mechanisms of the student knowledge level in CBM-based tutors [4] exist, most of them are usually based on heuristics (excluding some proposals which use Bayesian Networks [5]). We consider CBM could be extended in this sense, to improve the model with the Item Response Theory (IRT), a probabilistic and well-founded theory usually applied in testing systems to determine accurate measurements of student knowledge. In this way, IRT strengths could be used to improve the precision of the CBM student model. Here, we propose a model for student knowledge diagnosis through problem solving activities, which combines a CBM-based domain model with the assessment mechanisms supplied by the IRT. Unlike CBM-based tutors, we do not focus on the tutorial use of CBM. Rather, we use it for diagnosis purposes. As a result, our model can diagnose the students' performance while they solve procedural tasks using well-founded techniques.

In the next section the fundamental principles of CBM and IRT are described. Section 3 describes the features of our diagnosis model. Subsequently, we describe the experiment we have conducted with real students, which explores the validity of the results generated by our model. Finally conclusions and future work are outlined.

2. Background

2.1. Constraint Based Modeling Fundamentals

The use of CBM in a learning system allows improving the students' learning by making them learn from their own errors when solving a problem for a given domain. This is Olsson's theory [6] of learning from performance errors, which is the main basis of CBM. This theory defines the learning as a two phase process where, first, an error is detected and, afterwards, it is corrected.

Errors occur when students try to solve a problem and do not have the necessary declarative knowledge transformed into procedural knowledge. To detect errors, the system generates a representation of the solution being built, which is updated according to the actions being performed by students in the system interface. This representation is checked against a set of principles that form the domain model of a CBM-based tutor. These principles, which are the main unit of knowledge in CBM are represented as state constraints and must be satisfied by every correct solution for a given problem. That is, no correct solution can be achieved by generating a problem state that violates any of the state constraints in the domain.

According to CBM, each state constraint is defined by an ordered pair of conditions: R_c, S_c. R_c is the relevance condition and determines the kinds of problems and the state for which this constraint is relevant, i.e., where it could ever be applied. S_c is the satisfaction condition and contains the error condition that causes a problem to infringe the associated principle. When the R_c of a constraint is true for a given state of a problem solution, this constraint is pedagogically significant and then, the S_c must also be true. Otherwise, the constraint is violated and a mistake has been detected.

After error detection the student model is updated and the system responds to correct the student's misconceptions. Ohlsson postulates that this remediation occurs when the students try to apply their procedural knowledge by solving problems and they are warned about the violation of principles pertaining to the domain.

The learning process also depends on the model the system keeps about the student knowledge. The short-term student model consists of all violated constraints and those that have been satisfied. This model is used to build a long-term student model with an estimation of the knowledge and can be used to select the problem to be posed in the best possible way to overcome the individual's misconceptions. Consequently, this element is fundamental in a CBM-based tutoring system at the time of adapting the learning process. The more accurate the model, the better the remediation and adaptation.

2.2. *Item Response Theory*

The *Item Response Theory* (IRT), conceived by Thurstone [7], is the most popular well-founded discipline in charge of quantitatively measuring certain traits, such as intelligence, abilities, an individual's mastery in a particular concept, personality characteristics, etc. This theory is based on two main principles [8]: First, the student's knowledge for a test item can be explained by a series of factors called *knowledge level*. The second principle states that the relationship between the probability of answering an item correctly and the student's knowledge level can be described by means of a monotonically increasing function called *Item Characteristic Curve* (ICC) where the greater the student knowledge level, the higher the probability of answering correctly. This function is the central concept of IRT. One of the functions used to model the ICC is the 3-parameter logistic function (3PL):

$$P(u = 1|\theta) = c_i + (1 - c_i)\frac{1}{1 + e^{-1.7a_i(\theta - b_i)}} \qquad (1)$$

where $P(u_i = 1| \theta)$ represents the probability of answering the item i correctly given the student's knowledge level θ , which is usually measured in a continuous scale between *[-3.0, .., 3.0]*. The other three parameters of this curve depend on the item and mean the following: a_i is a *discrimination factor*, which is a value proportional to the slope of the curve. The bigger this value, the higher the distinction of the knowledge levels over the item; b_i or *difficulty index*, which matches the θ value for which the probability of a correct response is the same as that of a wrong response (without taking into account the c_i); finally, c_i or *guessing index*, represents the probability of a student with no knowledge at all answering an item correctly.

The popularity of IRT comes from the consistency of its results: In other theories, such as the Classical Test Theory, the student knowledge estimation results depend on the population where a certain test is carried out, and the test score cannot be compared to others obtained in different tests. In contrast, IRT has invariance, i.e., the knowledge level inferred using this theory does not depend on the test taken. As a result, if two different tests measuring the same concept are administered to the same student, we should obtain a similar estimation of his/her knowledge level.

In order to apply the IRT, it is necessary to have the ICC values corresponding to each item of the domain available. To this end, a prior statistical data-driven phase of *calibration* is needed. In this procedure the parameters describing the ICC are inferred. The calibration entry is formed by the set of students' performances who took the items (whose characteristic curves need to be inferred). The only information needed from these performances is the answer to each item.

The main use of IRT is in adaptive tests [9], which are tests where the presentation of each item and the decision to finish it are decided dynamically, based on students' answers. The final goal of an adaptive test is to quantitatively estimate the student knowledge level using the fewest number of items possible.

3. The Student Knowledge Diagnosis Model

Mitrovic and Martin [4] have demonstrated empirically the validity of the CBM strategy and also that it is more suitable than other approaches. However Ohlsson and Mitrovic [10] pointed out that, to support a wide range of pedagogical decisions, it is necessary to model long-term student knowledge. In this sense, most of the CBM-based tutors compute the student knowledge level as the proportion of constraints he/she knows. This heuristic has none of the desirable features of knowledge diagnosis such as invariance. This means that estimations made in this way depend greatly on the problems the students took.

Consequently, it would be desirable to have well-founded student knowledge inference mechanisms, which could guarantee the independence of the estimations. Test Theories, more specifically the IRT, provide these desirable features for assessing the student declarative knowledge. Nevertheless, at first sight, these theories are difficult to apply when procedural activities are being assessed. In fact, to evaluate the student's performance in a problem using IRT, the same way as a teacher would do, would require too many questions.

Our proposal tries to overcome the limitations of both techniques, i.e. CBM and IRT by combining them. We think that the heuristics commonly used in CBM for long-term student modeling could be improved by means of a model developed using the fundamentals of IRT. Accordingly, we propose a student knowledge diagnosis model where knowledge proofs are the actions performed by the student while he/she is taking a problem, which in turn, will lead to violating (or not) constraints in a CBM-based domain model.

If in an IRT-based diagnosis model the elements used to determine the student knowledge level are the items (i.e. test questions), in our proposal we use the constraints. Therefore, each constraint will have a characteristic curve assigned which represents the student's probability of firing it. We will call this curve *Constraint Characteristic Curve (CCC)*. Observe that the CCC will have exactly the opposite shape of an ICC, since ICCs represent the probability of answering correctly (knowledge), whereas the CCCs represent the probability of violating a constraint in a problem (detection of faulty knowledge). In other words, when a constraint is violated, this means that the student lacks knowledge, and therefore, the curve must decrease monotonically. The greater the student knowledge level, the lower the probability of firing the constraint. In IRT, this could be considered equivalent to a wrong response to an item.

Our diagnosis model is composed of the following elements:

- *The domain model*, which merges the domain requirements of a CBM-based domain model with those needed by an IRT diagnosis model. Thus it should contain the set of problems, the constraints and their relevance to each problem (i.e. whether or not the constraint could be violated in the problem), and finally the CCCs.

- *The short-term student model*: It contains, for each problem solved by the student, its relevant constraints, their state, i.e. violated or satisfied, and the number of times a constraint has been violated.
- *The long-term student model*, which takes the information provided by the short-term model and using the CCCs of the domain model infers the student knowledge probability distributions in the concepts assessed.

The student knowledge distribution $P(\theta \mid \pi, \tau)$ can be computed as the product of the CCCs of those constraints which have been violated, and the opposite of those constraints relevant for the problem which have been satisfied. This equation is inspired by the knowledge inference used in IRT:

$$P(\theta|\phi,\tau) = \prod_{i=1}^{m}\prod_{j=1}^{n}\left[P(c_j|\theta)^{f_{ij}}(1-P(c_j|\theta))\right]^{r_{ij}} \qquad (2)$$

where □ $=p_1, p_2, ..., p_m$ represents the set of problems solved by the student and $\tau = c_1, c_2, ..., c_n$ the set of constraints for this domain. $P(c_j \mid \theta)$ is the characteristic curve of the constraint c_j; r_{ij} is a binary value indicating whether or not the constraint c_j is relevant for the problem p_i; and f_{ij} is 1 when the student's actions in the problem p_i have violated the constraint c_j, 0 otherwise.

In our previous works [11], we use discrete characteristic curves, whose values are pairs of knowledge level / probability which can be obtained from reduced amount of data. In this proposal, the CCCs are also discrete, where each curve value indicates the probability that a student with certain knowledge level has of firing this constraint.

The result of applying the equation 2 is a probability distribution where we have all the knowledge level scale values and for each one, the probability of the student of having this value. The knowledge level can be easily inferred using, for example, the expected value of such distribution (in IRT it is called Expectation A Posteriori), or the knowledge level corresponding to the higher probability (i.e. Maximum a Posteriori).

The use of IRT provides our model with some advantages: The student knowledge estimations are invariant, that is, they do not depend on the problems the student took; the degree of estimation accuracy can be controlled; knowledge inference procedure is data-driven and well-founded, since CCCs can be inferred from prior student performance.

4. Experimentation

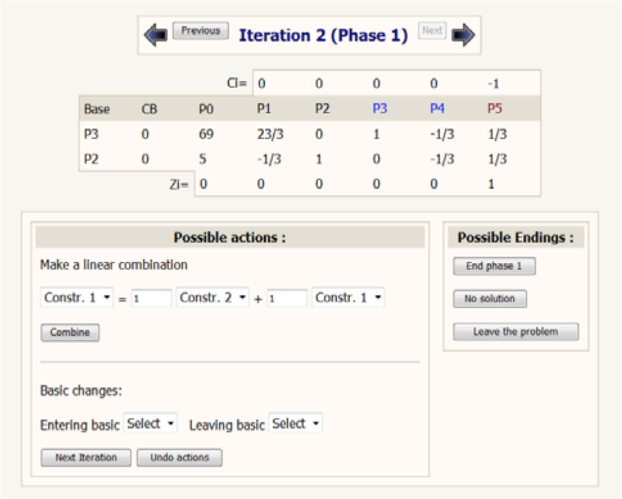
We have explored whether or not we obtain similar knowledge level estimation results when we use an IRT-based test versus a problem solving environment which implements our model. To make sure the results of both experiments were comparable; we tried to design them in such a way that both systems would evaluate the same concepts. For this reason, we chose a well-defined and limited procedural domain: the Simplex and the Two-Phase algorithms (the latter being a variant of the first one) [12]. In order to make the diagnoses, we used two different web-based tools: the Siette system [13], i.e. an application for student knowledge diagnosis using tests; and a problem solving environment of Linear Programming which implements our model.

4.1. *Experimental Design*

The experiment was conducted in November of 2008 with M.Sc. Computer Science students from the University of Málaga (Spain). It was incorporated into a six month course of Operative Research which involves Linear Programming techniques such as the two algorithms mentioned previously.

Students attended a two hour lecture on these algorithms. The following week they were assessed during two different sessions in the laboratories of the School. In the first session three problems were administered using our problem solving environment which took half an hour. Immediately afterwards, students took a test of 56 items using Siette with a time limit of an hour.

The problem solving environment was designed in collaboration with the course teacher and using our experience with former tools for the same domain [14]. The environment has an inference engine (implemented using JBoss Rules) where the representation of the solution given by the student is checked against the constraints. Figure 2 shows the interface provided to the student for solving a problem.



Iteration 2 (Phase 1)

Base	CB	P0	P1	P2	P3	P4	P5
		Ci=	0	0	0	0	-1
P3	0	69	23/3	0	1	-1/3	1/3
P2	0	5	-1/3	1	0	-1/3	1/3
	Zi=	0	0	0	0	0	1

Possible actions :

Make a linear combination

Constr. 1 = 1 Constr. 2 + 1 Constr. 1

Basic changes:

Entering basic Select Leaving basic Select

Possible Endings :

Figure 1. Interface of the Problem Solving Environment.

The teacher also contributed to the identification of the domain constraints. We grouped the constraints into three different categories, according to the phase where they are applied, i.e. transforming the problem into the Augmented Form, algorithm iterations and algorithm endings. A total of eighteen constraints were identified. Although this is a relatively low number in comparison to most of the CBM tutors, we consider that it is because the knowledge required to apply the Simplex method is very specific, well-defined, and has a reduced set of principles.

The problems presented to students in the first session were designed with the goal of covering as many constraints as possible. Accordingly, the teacher suggested three problems: one for applying the Simplex algorithm with only one solution, another for the same algorithm but with infinite solutions, and finally one for applying the Two-Phase algorithm. After each problem, the students were only shown a feedback indicating whether or not the solution was correct.

The Siette test was comprised of 56 items. In the construction of this test we tried to elaborate items which were appropriate to evaluate the same knowledge as the problem constraints. Most of them (52 items) were multiple-choice with four choices where only one was correct. The rest were open answer items corrected automatically by means of regular expression-based patterns.

Initially, 23 students participated in the experiment. However, due to several problems (for instance, some students left the session before finishing it), the data from seven of them were discarded. With the performances of these students we performed the following steps: First of all, we calibrated the test ICCs. To this end, we needed the students' performance in the test, i.e., we needed to know, for each student which items were answered correctly and which ones incorrectly. With this information we carried out the calibration process with MULTILOG [15], which is one of the most popular tools for this purpose. We also indicated that curves should be calibrated according to the 3PL model. Once obtained the calibrated ICCs, we inferred the students' knowledge level in the test.

Next, we accomplished an analogous process with the data obtained from the problem solving environment. First we calibrated the constraint curves. As mentioned before, in our model, we assume each constraint is equivalent to an item. Therefore, to perform the calibration, we needed to know, for each student, the constraint which they violated and which they did not. Once again we used MULTILOG for this purpose. The calibration process results were the CCCs with which we computed the students' knowledge level using equation 2.

4.1. Results

The goal of this experiment was to compare the student's knowledge diagnosis provided by Siette with the student's knowledge level inferred by our model. Consequently, we compared both values using a paired t-test with a *95%* confidence level, which is an appropriate technique for cases such as this where the sample size is small. This test compares two paired sets to determine whether they differ from each other significantly. The null hypothesis of paired t-test is that the mean of the differences is equal to zero. The result, *p=0.2091*, clearly suggests we cannot reject the hypothesis that students' knowledge estimations made by Siette are similar to those made by our model.

5. Conclusions and Future Work

In this paper we have introduced a model for assessing knowledge. This proposal combines the fundamentals of CBM with perhaps the most popular well-founded theory for test-based assessments, i.e. the IRT. With this combination we have tried to overcome some of the problems each of the techniques have separately. From the student diagnosis perspective, the incorporation of IRT-based assessment techniques could improve the CBM diagnosis, thereby guaranteeing the reliability and the validity of the results. Moreover, using our model, IRT can be applied to procedural domains, alleviating the overload that a long test could cause the students.

The experiment described has explored whether or not the student knowledge estimations using only an IRT-based testing system are comparable with our proposal, which combines IRT with CBM. Statistical analysis suggests that our model could diagnose in the same way as an IRT-based test does. However, whereas our proposal assessment session is composed of three problems, the IRT-based test contained 56 items. In this sense, we should point out that we selected a domain with a reduced set of constraints, since otherwise the test would have required a huge number of test items to be able to carry out a fair comparison between models. In our proposal, CBM constraints serve as student knowledge evidence providers with a lower cost than test items.

Regarding future lines of research, the work presented here is the first stage of a framework we are developing for constructing student models and procedural diagnosis tools based on this proposal [16]. Moreover, we think that new features could be added to our model by taking advantage of some characteristics of IRT, such as adaptive problem selection, which could allow the most appropriate problem to be selected dynamically; adaptation of the length of a problem solving session in terms of the required diagnosis accuracy; or problem difficulty inference.

References

[1] Self, J. A., Bypassing the Intractable Problem of Student Modeling, *Intelligent Tutoring Systems: at the Crossroads of Artificial Intelligence and Education*, pp. 107-123, 1990.

[2] Mitrovic, A., Martin, B., and Suraweera, P., Intelligent Tutors for All: The Constraint-Based Approach, *IEEE Intelligent Systems*, vol. 22, pp. 38-45, 2007.

[3] Mitrovic, A., Koedinger, K. R., and Martin, B., A comparative analysis of cognitive tutoring and constraint-based modeling, *User Modeling*, pp. 313-322, 2003.

[4] Mitrovic, A. and Martin, B., Evaluating Adaptive Problem Selection, *Adaptive Hypermedia 2004*, pp. 185-194, 2004.

[5] Mayo, M. and Mitrovic, A., Optimising ITS behaviour with Bayesian networks and decision theory, *International Journal of Artificial Intelligence in Education*, 12, pp. 124-153, 2001.

[6] Ohlsson, S., Constraint-based student modeling, *Student Modelling: the Key to Individualized Knowledge-based Instruction*, pp. 167-189, 1994.

[7] Thurstone, L. L., A method of scaling psychological and educational tests, *Journal of Educational Psychology*, vol. 16, pp. 433-451, 1925.

[8] Hambleton, R. K., Swaminathan, H., and Rogers, J. H., *Fundamentals of Item Response Theory (Measurement Methods for the Social Science)*, Sage Publications, Inc, July 1991.

[9] Van der Linden, W. J., and Glas, C. A. W., *Computerized Adaptive Testing: Theory and Practice*, Netherlands: Kluwer Academic Publishers; 2000.

[10] Ohlsson, S. and Mitrovic, A., Constraint-based knowledge representation for individualized instruction, *Computer Science and Information Systems*, vol. 3, pp. 1-22, 2006.

[11] Guzmán, E., Conejo, R., and J.L. Pérez-de-la-Cruz, Adaptive Testing for Hierarchical Student Models, *User Modeling and User-Adapted Interaction*, vol. 17, pp. 119-157, 2007.

[12] G. B. Dantzig, On the non-existence of tests of Student's hypothesis having power functions independent of σ, *Annals of Mathematical Statistics*, vol. 11 (2), pp. 186-192, 1940.

[13] Guzmán, E., Conejo, R., and Pérez-de-la-Cruz, J.L. Improving Student Performance using Self-Assessment Tests, *IEEE Intelligent Systems*, vol. 22, pp. 46-52, 2007.

[14] Millán, E., García-Hervás, E., Guzmán, E., Rueda, A., and Pérez-de-la-Cruz, J.L., TAPLI: An adaptive web-based learning environment for linear programming, *Lecture Notes in Artificial Intelligence*, 3040, pp. 676-685, 2003.

[15] Thissen, D., Multilog: Multiple, categorical item analysis and test scoring using item response theory (version 5.1). Mooresville, IN: Scientific Software, 1988.

[16] Gálvez, J., Guzmán, E., and Conejo, R., A SOA-Based Framework for Constructing Problem Solving Environments, *International Conference on Advanced Learning Technologies*, pp. 126-127, 2008.