Qualitative and Quantitative Student Models

Jose-Luis Perez-de-la-Cruz [1], Ricardo Conejo and Eduardo Guzmán
Dpt. LCC, ETSI Informática, Universidad de Málaga

Abstract. This paper is a first attempt to relate quantitative, unidimensional models to the fine-grained models usually found in the AI-ED community. More concretely, we define a certain type of qualitative student models that take into account the strict prerequisite relation, and show how a quantitative model arises from it in a natural way.

1. Introduction

In AI-ED literature, we can find proposals to model a student by means of comprehensive, fine-grained structures taking into account, for example, bug libraries, mental models, episodic memory, or learning preferences and styles. These rich, *qualitative* structures are usually difficult to initialize and update for a given student.

The very opposite approach is to model the student by just a real number θ (*performance measure*, in the terminology of [4]). In many real situations (for example, assigning students to groups), students are ranked in function of the results of a test and then the tutorial action is selected. At least as a first approximation, some systems use such an approach, directly or defining fuzzy labels on θ (the system KNOME[1] could be conceptualized in this way). Needless to say, the advantages of such *quantitative* models arise from the existence of well-founded mathematical techniques that allow their easy computation and updating.

A richer model makes feasible a better ITS. However, a more careful consideration shows that this is not always the case [4], [5], [8]. To cite J. Self, "it is not essential that ITSs possess precise student models, containing detailed representations of all the component mentioned above, in order to tutor students satisfactorily" [5]. In fact, "a student model is what enables a system to care about a student" [6], so "there is no practical benefit to be gained from incorporating in our student models features which the tutoring component makes no use of" [5]. On the other hand, it is clear that just a real number will be seldom a powerful model for tutoring; even for assessment tasks, the increasing interest in formative assessment creates the "…challenge of converting each examinee's test response pattern into a *multidimensional* student profile score report detailing the examinee's skills learned and skills needing study" (our emphasis) [7].

[1] Correspondence to: J. L. Perez-de-la-Cruz Dpt. LCC, ETSI Informática, Universidad de Málaga, Bulevar Luis Pasteur s/n, 29071 Málaga, Spain. Tel.: +34 952 132801; Fax: +34 952 131397; E-mail: perez@lcc.uma.es

So a trade-off is needed between the expressive richness of a model and the easiness of its creation and maintenance; and this trade-off will be governed by the gains in "tutoring power" vs. the losses in "creation and updating costs."

The research here presented addresses some of these problems. To this end, we will define a fine-grained structure for modeling student's knowledge and show how a quantitative unidimensional model can be suitably defined from it (section 2). Then we apply this theoretical framework to certain simple cases (section 3) that are amenable to explicit analytical techniques and to more complex cases whose study demands simulation tools (section 4). Finally, the conclusions drawn are summarized and future lines of research are sketched.

2. Theoretical Framework

A domain D is a directed acyclical graph $D(K, A)$ where K —the set of nodes— is the set of *knowledge atoms* and A —the set of arcs— is the *prerequisite relation*, i. e., $k_i \rightarrow k_j$ when the knowledge atom k_j cannot be mastered without mastering the atom k_i. Notice that, in this way, we are considering only *conjunctive* prerequisites. We will denote by N the cardinality of K, i. e., the number of knowledge atoms in the domain.

Given a domain D, a *qualitative student model* C (in the following, a *model*) is a subset of K such that, if $k_i \in C$ and $(k_j, k_i) \in A$, then $k_j \in C$, i. e., a subset of K that satisfies the constraints posed by the prerequisite relation. Notice that we are considering only *binary* valued for the mastering of a knowledge atom, i. e., for each k_i, the student knows totally/does not know the atom.

Let C_1, C_2 be two models. C_1 is a father of C_2 (or, alternatively, C_2 is a son of C_1) when $C_1 \subseteq C_2$ and $card(C_1) = card(C_2) - 1$, i. e., C_1 is a father of C_2 when C_2 can be generated by adding just an atom to C_1 in a way allowed by the prerequisite constraints. We will denote by $\sigma(C)$ the number of sons of C and by $F(C)$ the set of fathers of C.

The weight $w(C)$ of a a model C is defined recursively as follows:

$$w(C) = \begin{cases} 1 & \text{if } C = \emptyset \\ \sum_{C_i \in F(C)} \frac{w(C_i)}{\sigma(C_i)} & \text{otherwise} \end{cases}$$

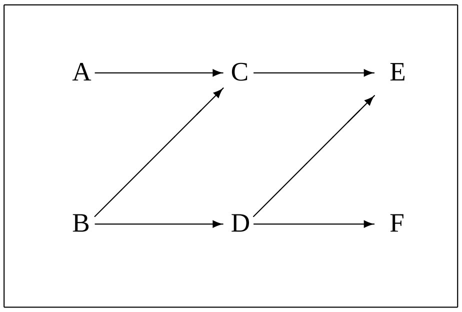
Notice that for each model C, $0 \leq w(C) \leq 1$, and that for each $m, 0 \leq m \leq N$,

$$\sum_{card(C)=m} w(C) = 1.$$

Perhaps an example will clarify the meaning of these definitions. Let us consider the domain of the figure 1(a). There are 6 atoms. Atoms A and B are prerequisites of C; atom B is prerequisite of D; atoms C and D are prerequisites of E; and atom D is prerequisite of F. There are 13 possible models. Their cardinalities and weights are summarized in figure 1(b).

Given a domain D, a *quantitative unidimensional model* P is a real number. It can be termed the student's *knowledge level*.

Now we want to define a function from models into knowledge levels, i. e., a function $f : 2^K \rightarrow \Re$. Some properties are intuitively desirable for the intended function f. For example, given a domain, f must be strictly monotonic, i. e, if $C_1 \subset C_2$, then



C	atoms in C	$card(C)$	$w(C)$
C_1		0	1
C_2	A	1	1/2
C_3	B	1	1/2
C_4	A, B	2	3/4
C_5	B, D	2	1/4
C_6	A, B, C	3	3/8
C_7	A, B, D	3	4/8
C_8	B, D, F	3	1/8
C_9	A, B, C, D	4	11/16
C_{10}	A, B, D, F	4	5/16
C_{11}	A, B, C, D, E	5	11/32
C_{12}	A, B, C, D, F	5	21/32
C_{13}	A, B, C, D, E, F	6	1

(a) (b)

Figure 1. A toy domain (a) and its models (b).

$f(C_1) < f(C_2)$, i. e, if the student knows more atoms, then his knowledge level is greater. The most obvious way is defining f as the count of known atoms $card(C)$, normalized into the common interval $[0, 1]$ and spread along all the real line, for example by means of the antilogistic function:

$$f(C) = \theta_C = log\frac{\frac{card(C)}{N}}{1 - \frac{card(C)}{N}}; \qquad card(C) = n(\theta) = N\frac{1}{1 + e^{-\theta}}$$

Notice that f takes a finite number of values, namely, $N + 1$. When $C = \emptyset$, $\theta_C = -\infty$; when $C = K$, $\theta_C = \infty$.

Let us assume that observable behavior consists of answers to certain questions, called *test items*. The relationship between θ_C and each test item T_i is given by an *Item Characteristic Curve*, *ICC*, such that $ICC_i(\theta)$ is the probability of giving a right answer to T_i if the student's knowledge is θ. To simplify the exposition, let us assume that every test item T_i depends just on one knowledge atom k_j. Let us also assume that there are neither slips nor guesses, i. e., that a student S answers correctly T_i if and only if $k_j \in C_S$, where C_S is the model corresponding to S's present knowledge. Then $ICC_i(\theta)$ is simply the probability of mastering the knowledge atom k_j given that the knowledge level is θ. The usual expression for an ICC whit no slip nor guess is the logistic function (see, for example, [2])

$$ICC(\theta) = \frac{1}{1 + e^{-a(\theta - b)}}$$

where b is the item difficulty level, such that when $\theta = b$, then $ICC(\theta) = 1/2$; and a is the item discrimination factor, such that when $\theta = b$, $dICC/d\theta = a/4$. Obviously, every $ICC_i(\theta)$ is monotonic.

For our models, a very naive approach would be to define $ICC_i(\theta)$ as follows: (i) count the number $N(\theta)$ of models C whose cardinality is $n(\theta)$; (ii) count the num-

θ	$-\infty$	-1.609	-0.697	0.000	0.693	1.609	∞
ICC_1	0.000	0.500	0.750	0.875	1.000	1.000	1.000
ICC_2	0.000	0.500	1.000	1.000	1.000	1.000	1.000
ICC_3	0.000	0.000	0.000	0.375	0.625	1.000	1.000
ICC_4	0.000	0.000	0.250	0.625	1.000	1.000	1.000
ICC_5	0.000	0.000	0.000	0.000	0.000	0.312	1.000
ICC_6	0.000	0.000	0.000	0.125	0.375	0.687	1.000

Table 1. Values of the ICCs for the domain of figure 1.

ber $N_1(\theta, k_i)$ of models C whose cardinality is $n(\theta)$ and $k_i \in C$; then, $ICC_i(\theta) = N_1(\theta, k_i)/N(\theta)$. However, this definition leads to nonmonotonic functions, i. e., it is possible that $\theta_1 \leq \theta_2$ and $N_1(\theta_1, k_i)/N(\theta_1) > N_1(\theta_2, k_i)/N(\theta_2)$. consider for example a domain with atoms $\{A, B, C, D\}$ and arcs $\{(B, C), (B, D)\}$. There are two models of cardinality 1: $C_1 = \{A\}$ and $C_2 = \{B\}$. $A \in C_1$ but $A \notin C_2$, hence $N_1(\theta_1, A)/N(\theta_1) = 1/2$. However, there are three models of cardinality 2: $C_3 = \{A, B\}$; $C_4 = \{B, C\}$; and $C_5 = \{B, D\}$. $A \in C_3$ but $A \notin C_3$ and $A \notin C_4$. Therefore, $N_1(\theta_2, A)/N(\theta_2) = 1/3$.

In fact, the real definition must take into account the different "likelihood" of every model C. We will adopt the following definition: let Θ_i be the set of models C such that $card(C) = n(\theta)$ and $k_i \in C$. Then $ICC_i(\theta) = \sum_{C \in \Theta_i} w(C)$. In this way, the "likelihood" of a model C is given by the relative number of paths of learning that can lead from the empty state of knowledge to the state represented by C. It is easy to show that $0 \leq ICC_i(\theta) \leq 1$ and that the function so defined is monotonic.

For example, let us show the values of $ICC_i(\theta)$ for the atoms in the domain of figure 1(a). Let us consider atom 1. For $\theta = -\infty$, i. e., $n(\theta) = 0$, there is just a model (the empty one, C_1 in table 1(b)) and $1 \notin C_1$, hence $ICC_1(-\infty) = 0$. For $n(\theta) = 1$, i. e., $\theta = -1.609$, there are two models, C_2 and C_3, with equal weight 1/2. Since $1 \in C_2$ but $1 \notin C_3$, $ICC_1(-1.609) = 0.5$. For $n(\theta) = 2$, i. e., $\theta = -0.697$, there are two models, C_4 and C_5, $w(C_4) = 3/4$, $w(C_5) = 1/4$. Since $1 \in C_4$ but $1 \notin C_5$, $ICC_1(-0.697) = 0.75$. In this way we can compute the values given in table 1.

3. Some Simple Cases

3.1. Lineal Domains

In the simplest cases, it is possible to derive analytically $ICC(\theta)$ and study its relationship to the features of the qualitative underlying model. For example, let us assume that the domain is lineal, i. e., that knowledge atoms are totally ordered,

$k_1 \rightarrow k_2 \rightarrow k_3 \rightarrow \ldots \rightarrow k_p$

In this case, there is exactly one model C_j for each possible cardinality j (therefore, its weight is 1) and $k_i \in C_j$ if and only if $i \leq j$. Therefore,

$$ICC_i(\theta) = \begin{cases} 0 \text{ if } \theta \leq \log \frac{i}{p-i} \\ 1 \text{ otherwise} \end{cases}$$

This is a degenerated logistic function with $a = \infty$ and $b = \log \frac{i}{p-i}$. In other words, the difficulty of k_i is $\log \frac{i}{p-i}$ and its discrimination is ∞. Let us assume now that a test item T_j requires the knowledge of *several* knowledge atoms $k_{j_1}, \ldots, k_{j_m}$. Then ICC_{T_j} is just ICC_{j_m}, i. e., the shape of the function is the same and the parameters are those of the most difficult knowledge atom.

Notice that in such domains given the knowledge level θ, for every knowledge atom k_j we can decide if k_j is known by the student. In this case, if we represent in the model the concrete atoms known by the student there is no gain of information; the quantitative model is an exact representation of the fine-grained one.

3.2. Flat Domains

Let us assume now that the domain is totally flat, i. e., there are no prerequisites. In this case, there are exactly $\binom{N}{j}$ models for each possible cardinality j. Obviously, their weights are equal to $1/\binom{N}{j}$. From these models, $\binom{N-1}{j-1}$ contain a certain atom i. Therefore, all ICCs are the same ICC and

$$ICC(\theta) = \frac{\binom{N-1}{n(\theta)-1}}{\binom{N}{n(\theta)}} = \frac{n(\theta)}{N} = \frac{1}{1+e^{-\theta}}$$

This is a logistic function with $a = 1$ and $b = 0$. In other words, the difficulty of every item is 0 and the discrimination is 1 (or 1/1.7, depending on the normalization adopted). On the other hand, let us assume now that a test item T_j requires the knowledge of several knowledge atoms $k_{j1}, \ldots, k_{jm}$. Analogously we can prove that $ICC(\theta) = \frac{n(\theta)(n(\theta)-1)\ldots(n(\theta)-m+1)}{N(N-1)\ldots(N-m+1)}$ and when $N \to \infty$, $ICC(\theta) \to \frac{1}{(1+e^{-\theta})^m}$. This is not the usual logistic function; however, if we define the difficulty level b as the value of θ such that $ICC(\theta) = 1/2$, then $b = \log \frac{1}{\sqrt[m]{2}-1}$; and, if we define the discrimination factor a as 1/4 times the slope at that point, then $a = m(2 - \sqrt[m]{2})$.

Notice that “the IRT model, in and of itself, simply does not address the question of why some items might be more or less difficult than others” ([3], p. 30); and the same could be asserted about the differences in the discriminating power between different items. However, in flat domains, our approach explain the real nature of these parameters: both difficulty and discrimination are monotone functions of the number m of atoms required to answer the test item. On the other hand, both parameters are assumed independent in IRT theory. If our analysis is correct, it is not the case for flat domains.

4. Some Simulations

For more realistic domains, it becomes impossible to explicitly obtain expressions for response curves. We have developed a simulation tool in order to study empirically the quantitative approximations in those models. With this tool we can define domains structured in levels. Each level contains a number of knowledge atoms. For each atom at a level i, its direct prerequisites are placed at the level $i-1$. Every atom (for level $i > 1$) has at least one prerequisite.

Different possibilities are allowed by the tool. For example, we can input a given domain with all its nodes and arcs. On the other hand, we can generate a domain at ran-

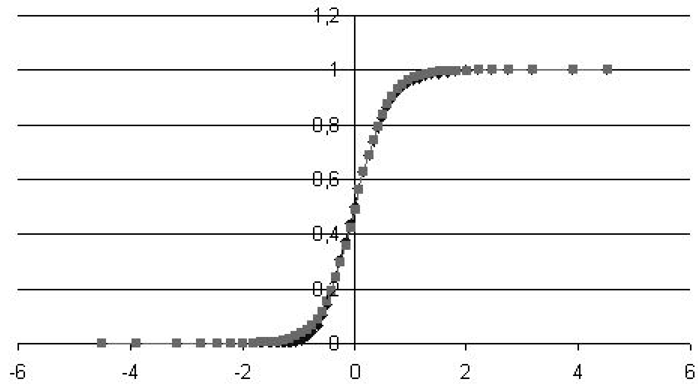


Figure 2. Real and logistic ICC.

dom, giving as input (i) the number of levels; (ii) for each level, the number of atoms; and (3) for each level, the expected number of prerequisites of an atom. In any case, the domain is processed by (i) computing all possible models and their weights; (ii) counting the presence/absence of each atom in each model; (iii) compiling the corresponding ICCs for each knowledge atom. Since the number of domains grows —in general— in an exponential way, this process can be very expensive in space and time. For example, for the domain used to generate the plots shown in this section, there are 50 atoms but 62515 domains (a big number, but distant from 2^{50}, the total number of subsets.) The domain consist of 50 knowledge atoms structured in 5 levels of 10 atoms. The number of prerequisites for each atom is at least 1 and its expected value is 3.

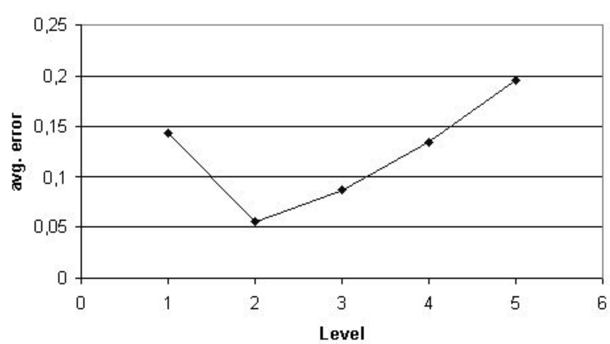


Figure 3. Average error vs. atom level.

The graphics in this section display the relation between some magnitudes in this domain. The aim of the graphics is just showing the kind of problems we are addressing and the kind of answers we are looking for. No claims of generality are made about the hints or tendencies shown by the figures. Not even a statistical analysis of the significance of the data has been performed; in fact, it must wait until a more exhaustive battery of simulations had been performed.

The first issue we want to study is the adequacy of usual logistic ICCs to response curves empirically found. Since we are considering that the response to a test item is deterministically given by the mastery of one knowledge atom, there are 50 response curves, one for each knowledge level. In figure 2 a real ICC is shown and compared to the its best (2 parameter) logistic approximation. The fitness seems good. More formally, the mean value of the quadratic error for the 50 curves is 0,1233.

However, the error is not the same for all atoms. The atom displayed in figure 2 lies "at the middle" of the domain. It can be studied, too, the relation between the level of the atom and the mean error. The results are shown in figure 3. The error is greater for the levels placed at the beginning or at the end of the domain.

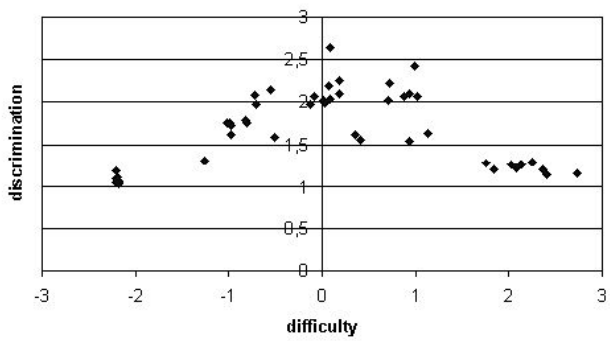


Figure 4. Discrimination vs. difficulty.

Another issue is the study of the correlation between the difficulty and the discrimination of an item. As said in section 3, both parameters are assumed independent. However, figure 4 shows that perhaps it is not the case in real domains.

5. Conclusions and Future Work

We have defined a certain family of qualitative, fine-grained student models. These models, simple as they are, take into account the prerequisite relation. We have derived a quantitative model from the qualitative one and shown how the response curves can be derived. The derivations have been done analytically for some simple cases and by means of simulations in more complex cases.

A lot of work must be done along these lines, with the final aim of determining in which cases quantitative models could be a sensible choice.

References

[1] D. N. Chin, KNOME: Modeling What the user Knows in UC. In Kobsa, A. and Wahlster, W. (eds.): *User Models in Dialog Systems*, Springer, Berlin (1988) 74–107.

[2] S. E. Embretson, S. Reise, *Item Response Theory for Psychologists*. Lawrence Erlbaum, 2000.
[3] Mislevy, R. L.: *Test Theory Reconceived*. CSE Technical Report **376** (1994), National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.
[4] S. Ohlsson, Some principles of intelligent tutoring. *Instructional Science* **14** (1986) 293–326.
[5] J. Self, Bypassing the Intractable Problem of Student Modelling. In C. Frasson and G. Gauthier (eds.): *Proc. ITS'88*, Ablex, Norwood, N. J. (1988) 107–123.
[6] J. Self, The defining characteristics of intelligent tutoring system research: ITSs care, precisely. *Intl. J. of Artificial Intelligence in Education* **10** (1999) 350–364.
[7] W. Stout, Psychometrics: From Practice to Theory and Back. *Psychometrika* **67** (2002) 485–518.
[8] B. P. Woolf, T. Murray, Using Machine Learning to Advise a Student Model. *Intl. J. of Artificial Intelligence in Education* **3(4)** (1992) 401–416.