Student evaluation through membership functions in CAT systems

J. Suárez
Centro de Investigación en Tecnologías de Información y Sistemas Universidad Autónoma del Estado de Hidalgo,(CITIS), Universidad Autónoma del Estado de Hidalgo,
Carretera Pachuca/Tulancingo, Km. 4.5, Cd. Universitaria, C.P. 42184, Pachuca de Soto, Hidalgo, México,
e-mail: jsuarez@uaeh.reduaeh.mx

Recibido el 26 de noviembre de 2002; aceptado el 23 de mayo de 2003

Generally speaking, the process of student evaluation is based on a procedure where we assume that the student belongs just to one set in a completely specified way, for example the set of excellent students, or the set of regular students. In this paper we use fuzzy sets concepts just to propose a different procedure which can be useful to manage the student's performance in a variation of a computerized adaptive testing administration process. This can be made by assuming that, for a given student, a membership function μ_A is assigned. This membership function gives the membership degree of the student to the fuzzy set A, which can be the set of excellent students, or the set of regular students, or the set of poor performance students. Furthermore, we assume that the item bank contains items belonging, with certain degree, to fuzzy sets describing the complexity of the given items. For example, we can talk about the fuzzy set of difficult questions or the fuzzy set of easy questions. By considering the evaluation process as a problem in the field of control theory, we establish a proper metaphor with a very simple, and very well studied, physical system with behavior described by variables such as the position, velocity and acceleration. Based in this model, we propose fuzzy rules just to control the item administration process as a function of the ability of the student and the item complexity.

Keywords: Fuzzy set; computer adaptive testing; testing administration; student performance.

Generalmente hablando, el proceso de evaluación estudiantil se basa en un procedimiento donde se supone que el estudiante pertenece a un conjunto de una manera completamente especificada; por ejemplo al conjunto de estudiantes excelentes o al conjunto de estudiantes regulares. En este trabajo utilizamos conceptos de lógica difusa con el objeto de proponer un procedimiento diferente que sea útil para administrar el desempeño estudiantil en la evaluación de un tópico perteneciente a algún área de conocimiento, es decir, como una variante de los métodos tradicionalmente usados para realizar evaluación adaptativa computarizada. Esto se puede lograr suponiendo que a un estudiante se le asigna un grado de pertenencia a un conjunto A a través de la correspodiente función de membresía μ_A. Esta función de membresía proporciona el grado de pertenencia del estudiante al conjunto A, el cual puede representar al conjunto de los estudiantes excelentes o al conjunto de estudiantes regulares o al conjunto de estudiantes deficientes. Además, suponemos que el banco de preguntas contiene ítems que pertenecen, con cierto grado, a conjuntos difusos que describen la complejidad de los mismos. Por ejemplo, podemos hablar del conjunto difuso de preguntas difíciles, o del conjunto difuso de preguntas fáciles. Considerando el proceso de evaluación como un problema de teoría de control, establecemos una metáfora apropiada con un sistema físico simple y bastante estudiado, cuyo comportamiento es descrito por variables tales como la posición, la velocidad y la aceleración. Basados en este modelo físico, proponemos reglas difusas sencillas para controlar el proceso de administración de preguntas como una función de la habilidad del estudiante y de la complejidad del ítem.

Descriptores: Conjunto difuso; evaluación adaptativa computarizada; administración de evaluación; desempeño del estudiante.

PACS: 01.40.Fk; 01.40.Gm; 01.50.Kw

1. Introduction

The item administration in computerized adaptive testing is traditionally based on random selection or on item response theory (IRT) which, at the same time, is based on statistical and probabilistic aspects [1](see Appendix B). The major questions in item administration process are how to start, how to continue, and how to finish the evaluation process [2]. The usual way of starting the evaluation process is through an estimate of the student's ability, which can be useful to select the proper complexity of the starting question. However, by doing so, it requires a previous knowledge of the estimate based on earlier performance of the student itself. In some cases, these previous data cannot be available.

One alternative way of solving this problem consists in giving to the student itself the chance of selecting the starting question and, therefore, of deciding what her/his level of knowledge is, leaving the problem of deciding what the actual level is to the item administration procedure, which is going to be described later on.

First of all, to solve a question with certain complexity (membership degree to some fuzzy set?) requires to define what we understand by *complexity*. There does exist a very common criterion to establish the difficulty degree of one item and it is based on timing aspects; in other words, if the solution requires a time long enough, then we can expect to assign to such a question a bigger complexity (there are other ways of assigning degrees of complexity to an item such as, for example, the number of previous topics that are required to find its corresponding solution). In the next sections this is what we will understand by item complexity. Therefore, in what follows, we will assume that there is an item bank where timing complexities have been assigned previously by some item administrator.

On the other hand, by considering the time of solution as a criterion to establish the complexity of one item, we are indirectly introducing some restrictions in the time to find one solution to every item in the evaluation. In this sense, there exists a strong contrast with traditional CAT evaluation process, where the time of evaluation does not play a main rôle at all. In the next sections we will see that this assumption is an important one in the definition of the fuzzy procedure for item administration.

2. Item and its membership functions

Assume that we have an item bank $\boldsymbol{R}$ with scheme $(R_1(\boldsymbol{X}_1), R_2(\boldsymbol{X}_2), \ldots, R_n(\boldsymbol{X}_n))$, where $R_i(\boldsymbol{X}_i)$ is a relation scheme with attributes $\boldsymbol{X}_i$, which represents a topic or subject i. Every relation in $\boldsymbol{R}$ contains different items, each related to different fields of knowledge. To simplify the analysis, we will assume that $\boldsymbol{R}$ contains questions with *high degree of complexity* and questions with *low degree of complexity*. In this way, we can define two fuzzy sets (see appendix A) H and E corresponding, respectively, to the set of questions with high degree of complexity and low degree of complexity.

As we said before in Sec. 1, this complexity degree is given by the time required to solve the problem. Perhaps, the number giving this complexity degree can be obtained through an statistical analysis of time solutions given by different experts, following what is called *horizontal method*, although several different procedures do exist, namely: vertical, comparison, inference, parametric estimation, and fuzzy clustering [3].

For example, in the horizontal technique, which is purely experimental, some elements of the universe of discourse of one concept A are selected, say $x_1, x_2, \ldots, x_n$, and a group of experts is questioned about the compatibility of x_i with the concept A. In our case, x_i can represent the time for finding an item's solution and A the fuzzy set of difficult items. The expert's answer takes only the values yes or no. The estimated value of the membership function in x_i is taken as the quotient of positives answers (yes) $P(x_i)$ and the total of experts questioned,

$$\mu_A(x_i) = \frac{P(x_i)}{N},$$

where $i = 1, \ldots, N$.

Going back to our main concern, if μ_H denotes the membership function of difficult questions, then we should expect that time solutions equal to zero correspond to easy questions and, therefore, that

$$\lim_{t \to 0} \mu_H(t = 0) = 0,$$

where t is the time required to solve the problem or item. On the other hand, very long time solutions imply very difficult questions and we should expect that

$$\lim_{t \to \infty} \mu_H(t) = 1.$$

The simplest way of defining the membership function μ_E of the set of easy questions is through the formula

$$\mu_E(t) = 1 - \mu_H(t),$$

where we see that

$$\lim_{t \to 0} \mu_E(t) = 1,$$

and

$$\lim_{t \to \infty} \mu_E(t) = 0,$$

which mean that items with very short time solutions completely belong to the set of easy questions, while items with very long time solutions does not belong at all to the set of easy questions, as we should expect. The following is an example of function $\mu_H(t)$:

$$\mu_H(t) = max\left\{0, tanh(\alpha(t - t_0))\right\},$$

where α and t_0 are positive constants, although more common membership functions, such as triangular and trapezoid, will be used in the next sections. In any case, the common sense and the experience dictate the shape of the membership functions [4].

3. Student and its membership functions

With the intention of simplifying the analysis, and based on the experience and common sense, we assume that an student may have poor, regular or excellent performance. Furthermore, these adjectives let us to identify the respective fuzzy set with the letters P, R, and B. It could be a finer partition, but this one is enough for our purposes. In the same way as we have assigned a membership function to every set of types of complexity, we can assign membership functions to the sets P, R, and B, which will be denoted as μ_P, μ_R and μ_B, respectively.

We have seen that time is the independent variable to compute the degree of membership in item complexity, however we require to define a different independent variable to compute the degree of membership in the sets of student performance. In this case, we will assume that the student performance is given by a grading scale, where the lower and upper ends of the scale correspond to poor performance and excellent performance, respectively. For example, the scale can be graded from 0 to 10 points or from 0 to 100 points.

At this stage of the discussion, we need to make clear that, at the start of the evaluation process, the examined student is who initially decides the membership degree to every fuzzy set P, R and B, and that the subsequent decisions are determined by the adaptive testing itself. The membership functions of the fuzzy sets related to the complexity of the item and to the class of students are very important in taking these subsequent decisions. The idea of all of it consists in going through proper modifications of the starting membership degree of complexity and performance fuzzy sets, until some given criteria are satisfied. At the end, we expect to obtain the actual student's ability, along with the complexity level of the items that the student can solve.

The previous description of the evaluation process contains an implicit dynamical system, which we are interested, therefore, in defining next. To do so, we will use a metaphorical analogy of this system with one very simple physical system based on the idea of uniformly accelerated motion (it could be non–uniformly accelerated, but we consider here the simplest case).

4. Simple fuzzy rules

As we said before, the model of test administration in an adaptive testing system is motivated by physical phenomena where the concept of uniformly accelerated motion is present. Every fuzzy set P, R and B can be interpreted as the actually existent 'distance' between the examinee and the tutor level (supposedly to be that of the teacher in charge of the student's learning). This distance will be given by the actual student's experience about the topic of the exam, this experience being represented by the present student's grading.

The very fact that the examinee tries to solve an item with time complexity t corresponds to the physical variable called *speed* or *velocity*. For a bigger time complexity of the item, then bigger will be the time required to find the solution. By making such a requirement, the student is really asking for decreasing the distance between her/him and her/his tutor.

Going back to the metaphor of the uniformly accelerated motion, the tutor represents the driver of a car to some speed, while the examinee is associated with the driver of a car going immediately behind the car of the first driver. Therefore, asking for an increase in the time complexity of an item is equivalent to increase the velocity in the second car of the metaphor.

Increasing or decreasing the speed implies the existence of an accelerated motion and, therefore, of a variable *acceleration*, which is useful to control changes in velocity. Analogously, in the adaptive testing model it should be possible to handle a variable *acceleration* which can be used to increase or decrease (or to hold) the time complexity of a given item (*i.e.* the degree of difficulty or simplicity of the item).

Clearly, to increase and decrease the velocity has an effect on the distance between the cars of the metaphor, which means that we should expect that the same should happen in the adaptive testing case when the performance of the examinee is considered as a function of the complexity of the items (remember that the examinee performance is the equivalent of the distance in the metaphor). In other words, it should be there also a modification step of the examinee's performance.

From the previous comments, we deduce that there are then three related variables; namely, the examinee performance, the time complexity of the items and their corresponding corrections or modifications. How to relate them? The answer to this question is given by the specification of inference rules based on fuzzy sets aspects. Although these rules are commonly obtained by experience and common sense, in what follows we propose the following ones which, for simplicity, assume only dichotomous items; in other words, items with answers yes or no:

1. If examinee's level is P and
   (a) The item complexity is H and
      i. The answer is incorrect, then decrease item complexity and hold examinee's level
      ii. The answer is correct, then hold item complexity and increase examinee's level
   (b) The item complexity is E and
      i. The answer is incorrect, then decrease item complexity and hold examinee's level
      ii. The answer is correct, then increase item complexity and examinee's level
2. If examinee's level is R and
   (a) The item complexity is H and
      i. The answer is incorrect, then decrease item complexity and hold examinee's level
      ii. The answer is correct, then hold item complexity and increase examinee's level
   (b) The item complexity is E and
      i. The answer is incorrect, then decrease item complexity and examinee's level
      ii. The answer is correct, then increase item complexity and examinee's level
3. If examinee's level is B and
   (a) The item complexity is H and
      i. The answer is incorrect, then decrease item complexity and examinee's level
      ii. The answer is correct, then increase item complexity and hold examinee's level
   (b) The item complexity is E and
      i. The answer is incorrect, then decrease item complexity and examinee's level
      ii. The answer is correct, then increase item complexity and hold examinee's level

The linguistic terms 'increase', 'decrease' and 'hold' require the definition of their corresponding fuzzy sets, which will be denoted by I, D and S, respectively. By taking one element of these sets is equivalent to increasing or decreasing the speed or distance in our already familiar accelerated motion [4, 5]. In our simple example, it represents an increment (negative, zero or positive) in the time complexity of the item or in the examinee's level (which is equivalent to modify the speed and the distance in our metaphor). In control theory, the common shape of their corresponding membership functions is as Fig. 1 shows [4, 5].

Table I resumes in a simple way the fuzzy rules previously given, and the first argument of the binary operator $\wedge$ refers to the item time complexity correction, while the second one to the examinee's performance. These rules define the behavior of a dynamical system with black box given in Fig. 2. The behavior of the dynamical system is completely defined by the set of eight membership functions of our example (three for examinee's performance, two for item time complexity and three for modifications), and the twelve inference rules relating the fuzzy sets represented by these membership functions.

TABLE I. Fuzzy rules.

Fuzzy rules	Item time complexity			
Examinee's level	E		H	
P	$D \wedge S$	$I \wedge I$	$D \wedge S$	$S \wedge I$
R	$D \wedge D$	$I \wedge I$	$D \wedge S$	$S \wedge I$
B	$D \wedge D$	$I \wedge S$	$D \wedge D$	$I \wedge S$
Answer type:	Wrong	Right	Wrong	Right

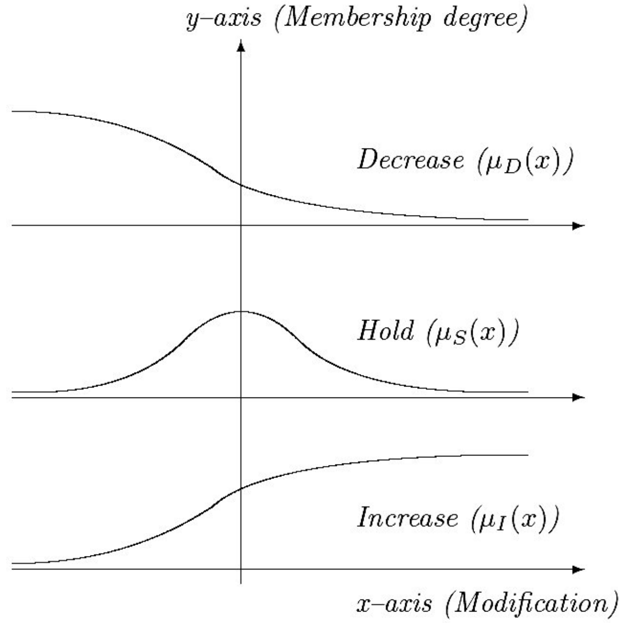

FIGURE 1. Membership functions to modify item time complexity or examinee's performance.

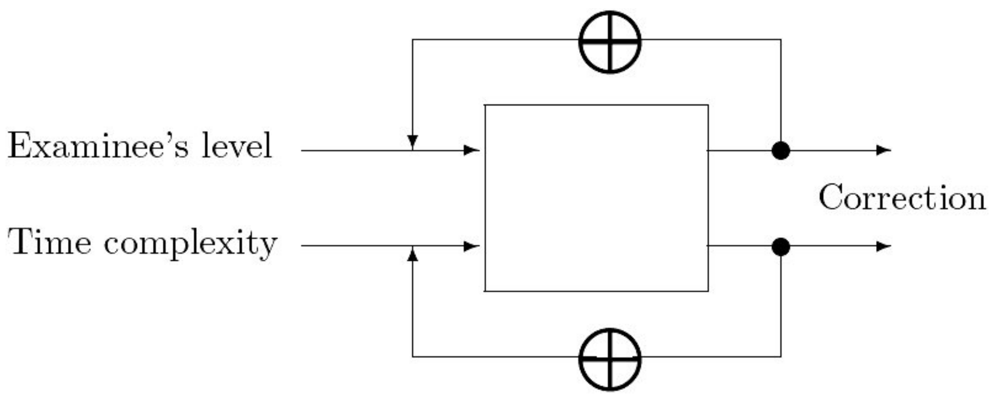

FIGURE 2. Black box.

As a concrete example, we consider the case where the fuzzy sets have trapezoid and triangular membership functions, as Fig. 3 shows. We should make clear that, in the example, the working of the system requires to use one of the 54 possible combinations of membership functions just to produce one single correction to complexity and examinee's performance. With these definitions at hands, a computer simulation of the system's behavior was realized under different conditions.

5. Simulation results

Simulation results were obtained by mean of a program whose structure is shown by Fig. 4, and the implementation was made with MatLab instructions [6]. The main module defines the different parameters of trapezoid membership functions, which are then send as parameters to Simpson function (which calculates the center of mass of complexity and performance correction). Furthermore, the main module specifies the function CMass to be used, the integration interval and the partition n.

On the other hand, a finer partition of the fussy sets increases the number of membership functions and the complexity of the internal structure of CMass modules. As we can see, these modules are in charge of the modifications to complexity and performance. The phase space that describes the dynamical behavior of the evaluation process is defined in terms of the examinee's performance (x) and item time complexity variables (y).

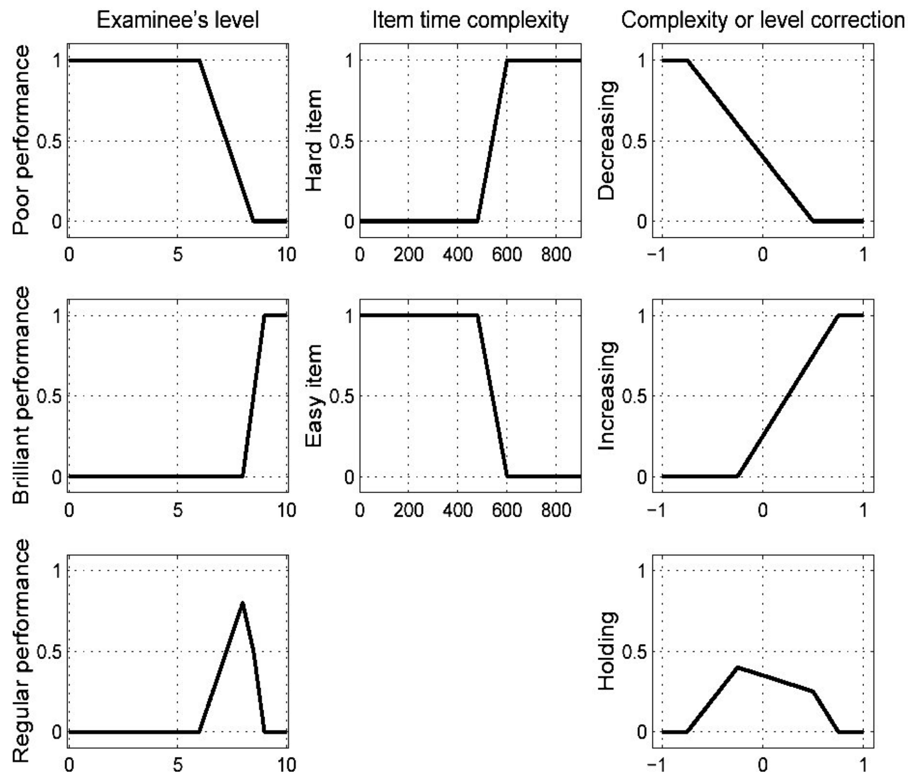

FIGURE 3. Fuzzy rules of the example.

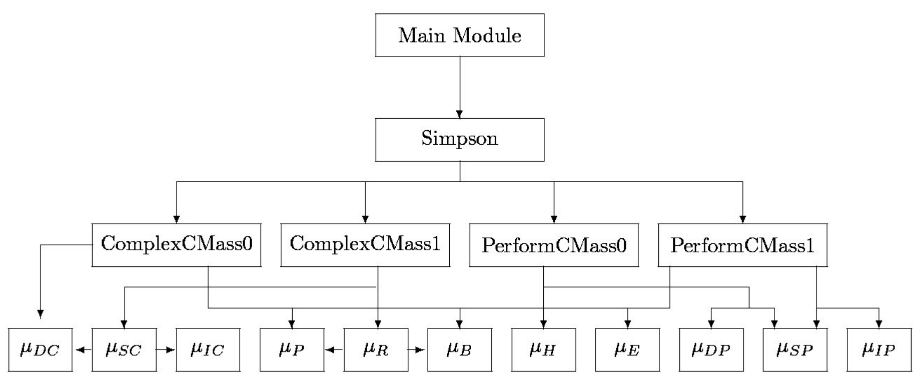

FIGURE 4. Block diagram.

Clearly, an excellent student is one who has answered the items in such a way that the mean value of (x, y) is located in highest regions, such as for example $[8, 10] \times [800, 900]$. Since there are three possible fuzzy sets in performance and two possible fuzzy sets in item time complexity, the phase space partition consists in six different regions such as Fig. 5 shows. The dynamical behavior of excellent students are then located in the intersection of column B and row H.

Table II shows the conditions used to run the simulation. There, it is shown that poor performance is obtained when the grading is a member of the closed interval [0,6], while a brilliant performance has a grading in the closed interval [8,10]. So that the regular performance is located in the interval [6,8]. On the other hand, Table II shows also the parameters of the correction functions.

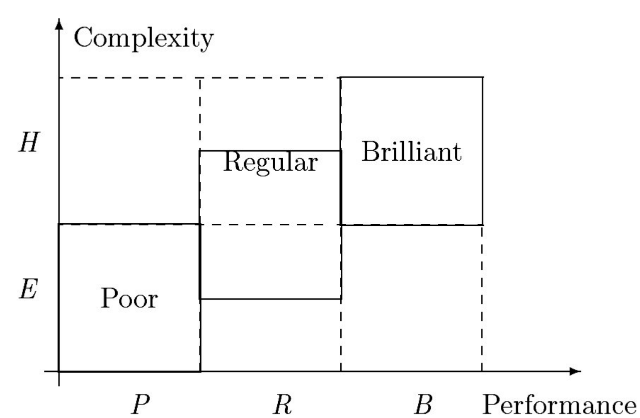

FIGURE 5. Phase space partition diagram.

TABLE II. Experimental conditions.

Function definitions			Parameters		—
Membership function	Interval	Units	x_0	x_1	Comments
μ_P	$[0, 10]$	grading	0	6	—
μ_R	$\surd$	$\surd$	—	—	Defined in terms of μ_P and μ_B
μ_B	$\surd$	$\surd$	8	10	—
μ_{DC}	$[-1, 1]$	grading	-0.25	0.0	—
(μ_{DP})		(seconds)	(-0.75)	(0.0)	
μ_{SC}	$\surd$	$\surd$	—	—	Defined in terms of $\mu_{DC}(\mu_{DP})$
(μ_{SP})					and $\mu_{IC}(\mu_{IP})$ (Watch the scales!)
μ_{IC}	$\surd$	$\surd$	0.0	0.25	—
(μ_{IP})			(0.0)	(0.75)	
μ_E	$[0, 900]$	seconds	300	600	—
μ_H	$\surd$	$\surd$	$\surd$	$\surd$	—

TABLE III. Experimental results.

Simulation results for honest students			Initial points	
Examinee's performance	Length of exam	% right answers	Performance	Complexity
Poor	20	30	3	300
Regular	20	35	6	600
Brilliant	20	90	9	850

Figure 6 and Table III show the simulation results when we assumed that the students are ’honest’ (in the sense that the given initial performance and complexity are very close to the actual ones), in every case the answer’s configurations for poor, regular and brilliant performance are, respectively, the following 00100100100100100010, 00101001001001001001, and 00111111111111111111 (remember that the exam consists of items accepting just one type of answer; namely, yes or no). To consider successful previous performance in the same examination, the corrections were weighted by the number of previous right answers; in other words, the correction term is taken as proportional to the number of previous right answers.

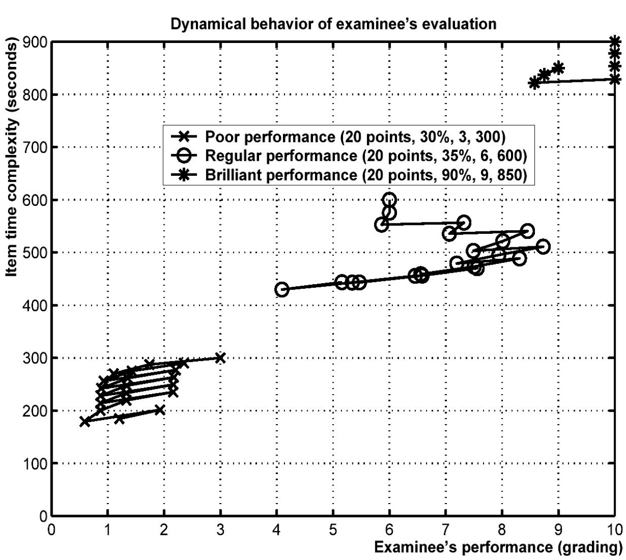

FIGURE 6. Simulation results for ’honest’ students.

We must make clear that the percentage of right answers does not necessarily represent the actual level of examinee’s knowledge. This is so, because the questions do not necessarily have the same item time complexity. In any case, the regions in phase space can be useful to decide where the level of knowledge is located, or at least what’s its behavior.

6. Conclusions

Although the simulation results seem very promising, there are still several questions that need to be addressed. First of all, we need to search for a procedure to properly weight the contributions of earlier right answers to the modifications in performance and complexity. Second, we need to consider also the effect that the shape of membership functions has on stability conditions. This includes to consider not only trapezoid and triangular functions, but differentiable ones such as, for example, sigmoidal and gaussian.

Third, we should notice that the very fact of working with dichotomous questions can be generalized to category of answers, where we assume that some fuzzy sets are also defined. In other words, the method can be also useful to deal with open questions, where the answer belongs, with some degree, to the defined fuzzy sets. Fourth, our main concern focuses on the fact of designing a fuzzy neural network, where the fuzzy rules shown in this paper can be embedded.

Fifth, interestingly results the fact of trying to define *in situ* the fuzzy rules. In other words, by realizing a sequence of examinations, it could be interesting to deduce the involved fuzzy rules, in the same way as some fuzzy control techniques do [7]. Finally, the development of a CAT system involving the concepts here presented must include a comparison of the advantages and disadvantages with those based on item response theory and its selection algorithms.

Appendix A. Basic concepts of fuzzy sets

In this work, the term *fuzzy set* refers to vagueness in a piece of information as in, for example, the set of tall people. This vagueness is measured through what is called *membership function*, μ. So that, if A is the fuzzy set and $x \in A$, $\mu_A(x)$ gives the degree of vagueness of $x \in A$. This should not be surprising, since in classical set theory the characteristic function μ_A of the set A is defined as follows:

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \notin A \\ 1 & \text{otherwise} \end{cases},$$

which means that x is an element of A or x is not an element of A. Fuzzy set theory extends this idea by taking the membership function $\mu_A : B \rightarrow [0,1]$, where $[0,1]$ is the real unit interval and $A \subset B$. In fuzzy set theory, like classical set theory, definitions of operations on fuzzy sets is still possible. Basics operations are therefore union and intersection, which are respectively defined as follows:

$$\mu_{A\cup B}(x) = max\{\mu_A(x), \mu_B(x)\}$$

and

$$\mu_{A\cap B}(x) = min\{\mu_A(x), \mu_B(x)\}.$$

On the other hand, there exist also what is called fuzzy reasoning, which is an extension of classical logic. Assuming that A and B are fuzzy propositions, we can respectively talk about fuzzy disjunction and conjunction as follows:

$$\mu_{A\wedge B}(x) = max\{\mu_A(x), \mu_B(x)\},$$

and

$$\mu_{A\vee B}(x) = min\{\mu_A(x), \mu_B(x)\}.$$

The definitions of disjunction and conjunction takes us to consider the possibility of defining the implication, where the antecedent of the rule is given by conjunctions of fuzzy propositions, and the consequent is just a single fuzzy proposition. There are more than one method to define or compute the membership function of the implication, and a popular one is called Mamdani's direct method [5] (it was proposed in 1975 by the researcher Ebrahim Mamdani) which can be described as follows.

To compute the membership function $\mu_{A_1\wedge\ldots\wedge A_n} \longrightarrow B$, where A_i and B are fuzzy sets, we start from the fact that

$$\mu_{A_1\wedge\ldots\wedge A_n}(x_1,\ldots,x_n) = min\{\mu_{A_1}(x_1),\ldots,\mu_{A_n}(x_n)\},$$

where x_i is an element of the universe of discourse of A_i. Next, by considering what is called the generalized *modus ponens*, define the desired membership function

$$\mu_{A_1\wedge\ldots\wedge A_n\longrightarrow B}(z) = min\{\mu_{A_1\wedge\ldots\wedge A_n}(x_1,\ldots,x_n), \mu_B(z)\},$$

for all z in the universe of discourse of B. If there are more than one implication, then apply the fuzzy disjunction operator as follows

$$\mu_{A\longrightarrow B}(z) = max\{\mu_{A_{1,1}\wedge\ldots\wedge A_{1,n}\longrightarrow B_1}(z),\ldots,\mu_{A_{m,1}\wedge\ldots\wedge A_{m,n}\longrightarrow B_m}(z)\} \quad (1)$$

where $A \longrightarrow B$ represents the set of implications shown as sub–indices in the right hand side of Eq. 1.

Appendix B. Basic concepts in computer adaptive testing

First of all, the study of adaptive testing requires the definition of the general case, which can be stated as follows: Given a set $\mathbb{E}$ of examinees, a collection $\mathbb{C}$ of sets of examinations $\mathbb{T}_i$, where $i = 0,\ldots,\infty$, the set $\mathbb{A}_{i,j}$ of the j–th examinee's answers for a given set of examinations $\mathbb{T}_i$, where the mark to item $q_{i,j,k} \in \mathbb{T}_i$ can take a value on the set $\mathbb{M}$, the set of possible marks for a given question, the main concern of adaptive testing is to find the best approximation to the real j-th examinee's m–dimensional proficiency $\boldsymbol{\theta}_j = (\theta_0, \theta_1, \ldots, \theta_{m-1})$, which intends to measure some previously defined constructs, through the answers to the set of questions $\mathbb{T}_i$.

To do this, adaptive testing makes use of initial proficiency estimates, item information and examinee's performance along the test. These considerations lead to a decision criterion which is useful to manage the item selection procedure, which is one of the main characteristics of adaptive testing. We consider here that $\mathbb{M}$ has elements in the real unit interval $[0,1]$, where the case of binary answer (yes–no, true–false) is just a particular one when *yes* and *true* are identified with 1, and *no* and *false* with 0.

Example B.1

As an example, consider the case of two examinees and an examination with three questions, and that the set of valid marks for the answers is

$$\mathbb{M} = \left\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\right\};$$

in other words, the mark for an answer of some particular question in the exam can take one of a total of four values. Table IV shows one of the $4^3 \times 4^3 = 4096$ possible results of the testing process.

TABLE IV. Sample of examination process.

	Questions		
Examinee	1	2	3
1	$\frac{1}{2}$	1	0
2	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
	Mark		

Example B.2

As another example, consider now that the set of valid marks for the answers is $\mathbb{M} = \{0, 1\}$. Table V shows one of the $2^3 \times 2^3 = 64$ possible results of the testing process.

TABLE V. Sample of examination process.

	Questions		
Examinee	1	2	3
1	1	0	1
2	0	1	1
	Mark		

In any case, there does exist a transformation $\mathcal{T} : \mathbb{A}_{i,j} \to \mathbb{M}$, which maps an answer to a single value in $\mathbb{M}$. Additionally, for a given element in the set $\mathbb{M}$, in some cases an *a priori* probability for every question is known, and this probability is a function of the m–dimensional proficiency variable $\mathbf{\Theta}_{i,j} = (\Theta_{i,j,0}, \Theta_{i,j,1}, \ldots, \Theta_{i,j,m-1})$, where j refers to the item j in exam i. The probability is denoted as $P_x(\mathbf{\Theta}_{i,j})$, and this expression represents the probability of answering the item j with a mark equal to x, by knowing that the examinee has a proficiency $\mathbf{\Theta}_{i,j}$.

Example B.3

In case of one–dimensional proficiency and binary answers, the proficiency becomes a single variable Θ, and the usual simplest way of defining the probability of giving a right answer to an item is through the idea of *one parameter logistic function* or, briefly, *1PL model*, which is defined as follows

$$P(\Theta) = \frac{1}{1 + e^{-(\Theta - \beta)}},$$

where the parameter β is known as the complexity of the question. In this case, a parameter β is assigned to every item in the exam. Since there are only two possible marks (0 or 1), the index in $P_1(\Theta)$ is removed for simplicity. The probability of giving a wrong answer is therefore $1 - P(\Theta)$.

The main point in Example B.3 is that the information that an item's answer provides about examinee's proficiency at any given point along the proficiency scale depends only on item parameters, which are explicitly included in a more general logistic function, and this function is called *item characteristic curve* or ICC for short. Therefore, by knowing an initial estimate of the examinee's proficiency, a procedure for selecting items with complexity similar to this proficiency can be implemented. The mathematical basis of the whole thing is given by what is called *item response theory* or IRT for short, which is supported by statistical tools.

Therefore, in and adaptive testing, the usual general algorithm to select the proper item for an examinee with a given proficiency proceeds as follows: At each step of the examination, and considering the statistics of the examinee's performance along the same test, select the more informative item [8]. The term CAT arises from the fact that the previous algorithm can be implemented in a computer. So, CAT means *Computer Adaptive Testing*.

1. Wim J. Van der Linden and R.K. Hambleton, *Handbook of Modern Item Response Theory* (Springer Verlag, New York, USA, 1997).
2. D. Thissen and R.J. Mislevy, *Testing Algorithms*, in Computerized Adaptive Testing: A Primer, Second Edition, Howard Wainer *et al.* (Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, 2000).
3. W. Pedrycz and F. Gomide, *An Introduction to Fuzzy Sets: Analysis and Design* (Bradford Book, The MIT Press, London, England, 1998).
4. S.T. Welstead, *Neural Network and Fuzzy Logic Applications* (John Wiley & Sons, New York, USA, 1994).
5. K. Tanaka, *An Introduction to Fuzzy Logic for Practical Applications* (Springer Verlag, New York, USA, 1997).
6. G. Lindfield and J. Penny, *Numerical Methods Using MatLab* (Ellis Horwood, London, England, 1995).
7. B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical System Approach to Machine Intelligence* (Englewood Cliffs, N.J., Prentice Hall, 1992).
8. Wim J. Van der Linden and P.J. Pashley, *Item Selection and Ability Estimation in Adaptive Testing*, in Computerized Adaptive Testing: Theory and Practice, Wim J. Van der Linden and Cees A.W. Glas (Eds.), ICO (Kluwer Academic Publishers, P.O. Box 17, 3300 AA Dordrecht, The Netherlands, 2000).