Adaptive Testing System Modeled Through Fuzzy Logic

J. Suarez-Cansino and R. A. Hernandez-Gomez
Universidad Autonoma del Estado de Hidalgo
Instituto de Ciencias Basicas e Ingenieria
Centro de Investigacion en Tecnologias de Informacion y Sistemas
Carretera Pachuca - Tulancingo, Km. 4.5, Pachuca de Soto, Hidalgo, Mexico,

Abstract.

This paper shows the main ideas behind a computerized adaptive testing system, where the item administration component is modelled through simple fuzzy logic concepts. The work introduces the idea of considering the examinees as members of fuzzy sets by defining some membership functions. These membership functions are defined by assuming the previous experience of the authors in several academic areas. In addition, it is also assumed that the item database contains questions belonging to fuzzy sets defined through the idea of item complexity. The *rationale* of the proposed item administration model considers the evaluation process as a control theory problem, by establishing a metaphor with a physical system from a kinematical point of view. In accordance with this physical model, fuzzy rules are proposed to the computer adaptive testing system, where some computer simulated results are also presented. Finally, this work suggests some useful ideas about the database design in a real computerized adaptive testing system.

Key-words: - adaptive, modeled, membership functions, CAT, item database

1 INTRODUCTION

In the engineering field, the definition of instrument is generally associated with a device for making some useful work or obtaining some expected, or unexpected, experimental results. If this instrument is used to qualify or quantify objects or concepts, then this instrument is called a measure instrument. As an example, a thermometer is used to measure temperatures on substances or environments, so a thermometer is considered a *measure instrument*; however, even though a HUB in a computer network is an instrument, it cannot be taken as a measure one.

Nevertheless, there exist not so obvious instruments than those mentioned above. For example, consider the case of a test applied to a number of students to be certified in some knowledge area. If the test can be properly combined with others instruments as, for example, paper and pencil, then the combination can be thought as a measure instrument. In fact, from the psychometric point of view, there exists an increasing interest in studying the well known field of Computer Adaptive Testing (CAT), which can be useful to define such a kind of measure instruments.

A CAT is a measure instrument essentially defined by an item database, an item administrator and an evaluator. The main idea of a CAT system is to evaluate a student on a given knowledge area by adapting the items to the student knowledge.

Traditionally, the construction of the item administrator is based on logistic models as, for example, those defined by Rash. Theoretically speaking, there is a well established mathematical tool for dealing with these models, which is called Item Response Theory, or IRT for short. As an alternative paradigm, this paper shows the main ideas behind a computerized adaptive testing system whose evaluation administrator is based in a quite simple fuzzy logic model.

One of the main assumptions made here is that the students are considered as belonging, with some membership degree, to a well defined fuzzy set whose membership functions are defined based on academic experience. Moreover, it is also assumed that the item database is made of items belonging to fuzzy sets determined by item's complexity.

The proposed model is mainly based by considering the students evaluation process as a control theory task, setting up a metaphoric analogy with a simple physical system aroused from some kinematics ideas. According to this physical model, adaptive testing fuzzy rules are proposed, showing a few simulated examples of the system evolution. Finally, the paper gives some ideas about the item database design for a real CAT.

2 FUZZY MODEL

By using a fuzzy logic model for constructing the item administrator of an adaptive testing system, instead of a probabilistic one, has a philosophic interpretation dealing with vagueness terms, which are connected with the daily duty of the academic evaluations. In the same way that a computerized adaptive test has an analogy with the form that an examiner applies an oral test to a student; namely, if the student answers correctly an item, then the examiner rises or at least holds the complexity degree of the item, there does also exist a similar
analogy of the fuzzy model in the item administration with the behavior of the examinee, when he chooses the next item from the corresponding item database.

Our assertion is that the search of new items to apply, although it could be based in a student historical record and the answer to previous questions, is made by using vague or fuzzy terms, instead of probabilistic assignations normally used in adaptive evaluators with 1PL, 2PL or 3PL models [4, 5]. In other words, this paper affirms that the adaptive testing acts in a fuzzy way instead of a probabilistic one. This means that, instead of determining the probability that an examinee belongs to a students set with a given performance, the fuzzy evaluation determines the membership degree of the examinee to the set.

The justification that this is so has a practical explanation, which is based on the fact that it is relatively easier for an examiner to estimate the membership degree of a student to a given set, than the probability of membership to the same set. Something similar can be said about an item complexity. It's an easier matter to talk vaguely about an item complexity than giving a probability of right answer to the same question.

2.1 Rationale

Debate about the novelties that fuzzy logic presents compared with probability theory still goes on. Defenders of the first model even assure that probability is a particular case of fuzzy logic. Our interest in this section is to propose a rationale about why to use fuzzy logic instead of probabilistic models in the computerized adaptive testing area. In fact, we tried to do so in the introduction. The next justification is mainly based in some examples previously handled by the fuzzy model defenders, but adapted to our interest area.

When we propose a fuzzy model for the item administrator we are emphasizing the fact that with a probabilistic model we loose the possibility of working on perception based information, which is the kind of information usually handled in evaluative processes. These perceptions are manifested in expressions such as
\X is a brilliant pupil", \X is a regular student", \Question X is quite difficult", etc.

For further exemplification, we adapt a Lotfi Zadeh argument based on the problem of spheres in a box to a situation having brilliant or regular students in a classroom [2]. The idea is to consider this problem as a version of getting information based on perceptions instead of getting it through measurements and operations on measurements as a probabilistic model does. The students in the classroom can be located in one of the several skill levels from poor to excellent, according to their experience in a previously given activity.

In a first probabilistic trial to determine the skills level of every student in the classroom, a classical paper and pencil evaluation is applied. In normal evaluation conditions, the result of the examination depends on the exam itself, no matter if the examination of every student has the same questions, or the examination is an individualized one. In other words, the result depends on the contents of the examination. This means that the student's ability will surely change if the contents of the examination are changed.

Briefly speaking, every student will be located in a skill level depending on the test content. With this in mind, can we realistically say how many poor performance pupils are in the classroom?, can we reliably give the probability that when choosing a pupil, this will be a regular performance pupil? Of course we can, but at the end, and this is what we want to remark, the answer will depends on the test contents we use to classify the students in the classroom.

So, has any sense to say that a pupil has a poor performance with one test and a regular performance with another, if both tests were meant to measure the same skills or knowledge? This situation recalls another where two persons are consulted to give an opinion of a third one. In general, their opinions can even be totally opposites. In our case, the tests are the persons consulted. The evaluation problem becomes then in an appreciation matter or, as we mentioned earlier, in a perception one.

Our intention is to remove the skill level dependence on the test contents putting forward test administration models which controls the test contents, in such a way that we get the same skill level for a given knowledge level. This is the essential idea of adaptive evaluators systems, which builds the examination in accordance to the examinee; in other words, depending on the student knowledge level. However, in the construction of these systems, information processes dealing with probabilistic methods are also involved, which return us again to the paper and pencil evaluation scene already explained.

An example of this situation is given by the fact of assigning complexity to questions, which are based in a statistic process with the participation of experts or the students themselves. Given a question X, how many experts or examinees give certain complexity degree to this question and how many think that the same question has a different complexity degree? Undoubtedly this is a perception problem as well. The problem in this case is not if this information must be obtained by means of statistical methods, but how to use this information to build possible answers to the previous questions, in such a way that appreciation matters will also be included.

2.2 Fuzzy Model

Looking for analysis simplification, and using experience and common sense, we suppose that, in an examination, students can have poor, regular or brilliant performance. Furthermore, we identify from these adjectives their related fuzzy sets with capital letters *P*, *R* and *B*, respectively. It could be made a finer partition, but this is good enough for our present purposes. In the same way that membership functions are assigned to each item complexity level, membership functions can also be assigned to the *P*, *R* and *B* fuzzy sets, and they are denoted as ${}^{1}{}_{P}$, ${}^{1}{}_{R}$ and ${}^{1}{}_{B}$, respectively.

So, we will assume that questions, with high and low complexity degree, exist on database, in such a way that we can define two fuzzy sets, *H* and *L*, corresponding to high and low complexity degree questions, respectively. As was indicated in a previous work [1], this complexity degree is determined by the time required to solve the problem, and the membership functions are denoted by μ_H and m_L, respectively.

The test administrator model in an adaptive testing system is motivated by physical phenomena such as uniform accelerated motion. Every fuzzy set *P*, *R* and *B* may be interpreted as the "distance" between the examinee and the tutor levels (the tutor is assumed to be in charge of the pupil's learning). The distance will be the result of the actual experience that the student has on the evaluation topics, represented by the student's mark.

Table 1 shows in a simple way the fuzzy rules used, and the first argument in the binary operator refers to the question complexity correction, while the second argument refers to the student performance correction. These rules determine the behavior of a dynamical system symbolized as a black box by Figure 1.

The dynamical system behavior is wholly defined by the eight membership functions set in our example (three for the student performance, two for the item complexity and three for corrections or modifications). In this case, the arguments of the binary operator, namely *D*, *I* and *S*, stand for decreasing, increasing and holding complexity or performance, respectively.

Fuzzy Rules	Question Temporal Complexity			
Examinee Level	***L***		***H***	
P	D ∧ S	I ∧ I	D ∧ S	S ∧ I
R	D ∧ D	I ∧ I	D ∧ S	S ∧ I
B	D ∧ D	I ∧ S	D ∧ D	I ∧ S
Answer Type	Wrong	Right	Wrong	Right

Table 1. Fuzzy rules

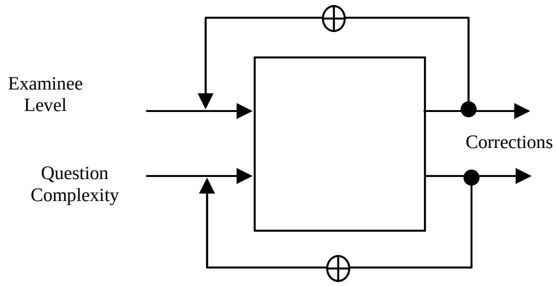
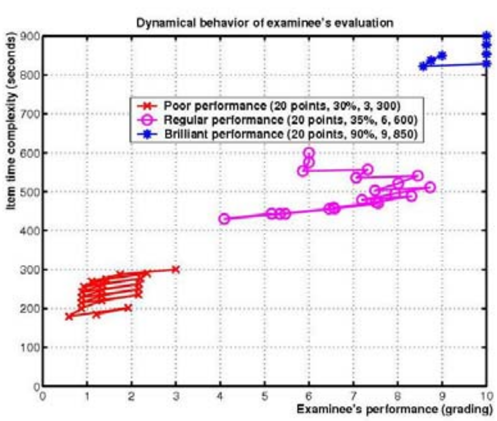


Figure 1: Black box

2.3 Simulation results

As a concrete example, we take the case where the fuzzy sets membership functions have trapezoidal and triangular geometry. It's worth to say that, in this example, the system operation requires one of the 54 possible combinations of the membership functions just to give only one complexity correction and only one correction to the examinee performance. With all these definitions and considerations at hand, a simulation of the system behavior was made on the computer under different conditions.



The simulation results were obtained in such a

Fig. 2. Simulation results for `honest' students

way that the phase space describes the dynamical behavior of the evaluation process. This phase space is defined by the examinee performance (x) and the question complexity (y). It's clear that a brilliant student answers the questions in such a way that the (x, y) mean value lies in the highest zone, for instance [8, 10] x [800, 900]. As long as there are three possible fuzzy sets for performance and two for item's complexity, the phase space partition consist of six different regions as Figure 2 suggests.

3 Item Database Design

Our interest goes further than the simple fact of developing and to instrument an adaptive evaluator for just a specific knowledge area. Interest is mainly focused in the development and instrumentation of an item database with questions about topics and sub-topics on different subjects, of different educational levels going from elementary through higher degrees.

In this way, the item database must take into account the fact that we have several different educational levels and, in general, the school calendar is not the same in all of them. Nevertheless with the main intention of working in a common structure, we consider from here on that these levels are defined by nine grades, denominated from minor to mayor as first, second, third, fourth, fifth, sixth, seventh, eighth and ninth.

Furthermore, every grade of each level will consist of a subjects list which, at the same time, defines a list of topics per subject, a list of subtopics per topic; i.e., the database structure is the following: Educational Level, Grade, Subject, Topic and Sub-topic.

3.1 Item Database Development

The complete item database is shown by Figure 3. There, we can observe the links between different n-uples, which give a special connection to the shown relations. Furthermore, it is important to point out that, on the item database instrumentation, we must be careful with reference and entity integrity.

With the item database diagram so defined, we find convenient to operate the database through a proper interface for capture, reading and production of information. In the Answers diagram, the value of the attribute Answers could be pointers to more complex structures than a simple register. In general, such answers could be given in a graphical form or in terms of variables containing the right values. In general, the system can generate real time answers.

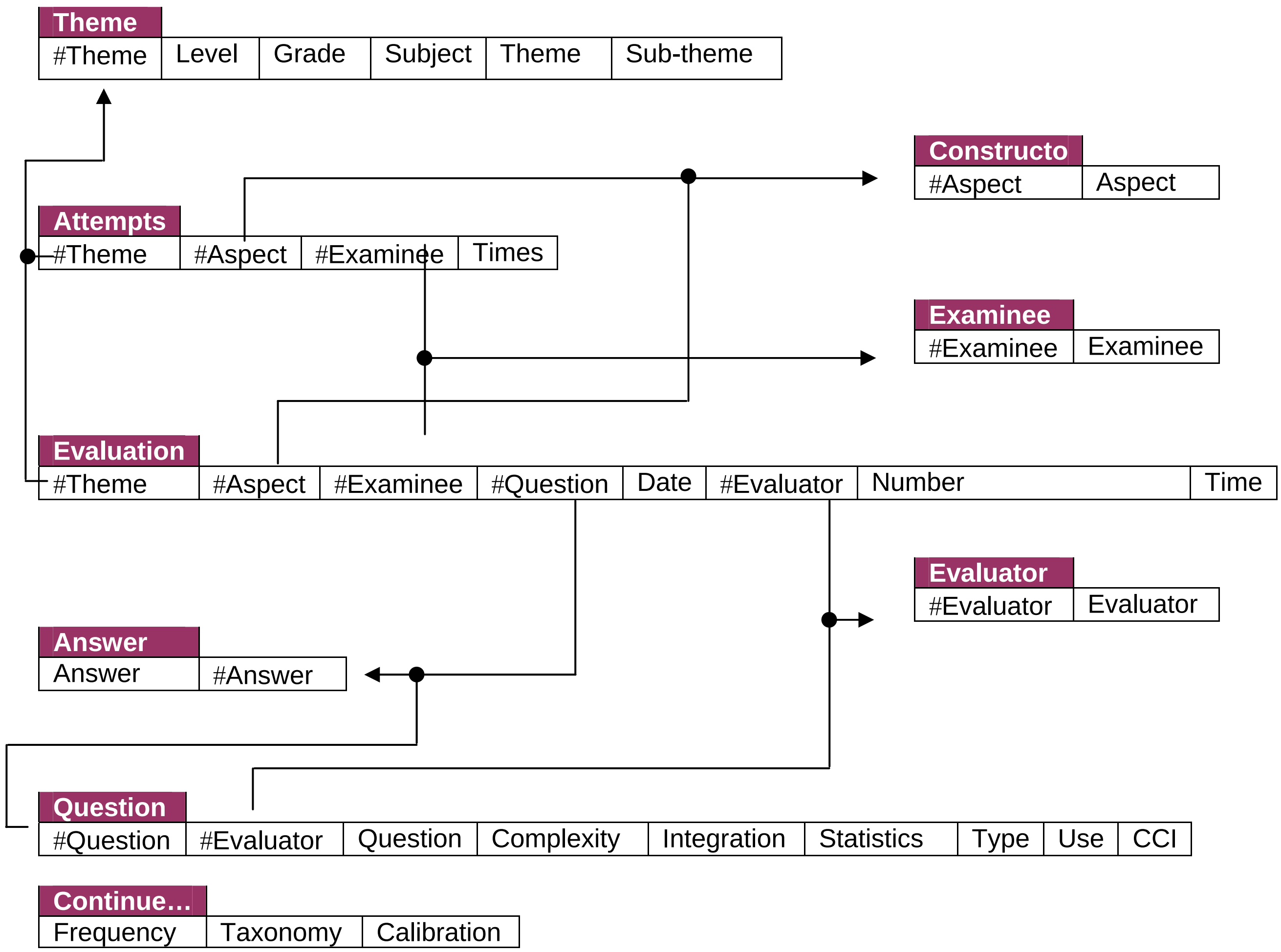


Figure 3. Relations in item data base scheme.

4 Conclusions

Nowadays, alternative models, different to probabilistic ones, which help to design the operation of an item administrator, in a computerized adaptive testing, are difficult to obtain in the literature. As it's seen in this paper, it seems possible to build a computerized adaptive testing by using a fuzzy model as an item administrator. Here we propose the minimum useful ideas to realize such a project.

On the other hand, it is quite interesting to focus the work to a lot of related research topics in the fuzzy logic field, neural networks and adaptive testing. For instance, once the fuzzy model operation has been explained, we figure out about the possibility of building an RBF artificial neural network because of the close relation with fuzzy models. Moreover, a lot of interesting topics can be proposed as, for example, membership functions and their formal definitions, item's complexity, item administration, etc.

References

1. Suarez Cansino Joel, Student evaluation through membership functions in CAT systems, Revista Mexicana de Física, 49, 4, 371-378, (2003).

2. Thissen D., R. J. Mislevy, Testing Algortihms, in Computerized Adaptive Testing: A Primer, Second Edition, Howard Wainer et al., Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, USA, (2000).

3. Van der Linden Wim J., R. K. Hambleton, Handbook of Modern Item Response Theorem, Springer - Verlag, New York, USA, (1997).

4. Zadeh Lot¯ A., Probability Theory and Fuzzy Logic, Research Report, Computer Science Division and the Electronics Research Laboratory, Department of EECS, University of California Berkeley, USA, pp. 1 - 5, December 9, 2002.