Neural Network-based Fuzzy Modeling of the Student in Intelligent Tutoring Systems

R. Stathacopoulou, G. D. Magoulas, M. Grigoriadou
Department of Informatics, University of Athens, TYPA Buildings, GR-15784 Athens, Greece
E-mail: {sreg, magoulas, gregor}@di.uoa.gr

Abstract

An empirical approach that makes use of neuro-fuzzy synergism to evaluate the students in the context of an intelligent tutoring system is presented. In this way, a qualitative model of the student is generated, which is able to evaluate information regarding student's knowledge and cognitive abilities in a domain area. The neuro-fuzzy model has been tested on a prototype tutoring system in the physics domain of the vertical projectory motions and the results have been very satisfactory.

Introduction

The use of artificial intelligence techniques to educational software design influenced the evolution from Computer Assisted Instruction (CAI) to Intelligent Tutoring Systems (ITS) or Intelligent Computer Assisted Instruction (ICAI). Developers have tried to incorporate "intelligence" in the areas of knowledge, problem solving, tutoring and communication with the student in order to create a system which is expert in a particular field and also able to provide individualized instruction. For the purposes of design and conceptualization, ITSs are described as having four major components [24]:

- The *domain knowledge*, which is aimed to store, manipulate and reason with knowledge of the domain being taught.
- The *pedagogical module*, which provides information about the teaching strategy that must be used to a specific student.
- The *student model*, that stores and analyzes information of student's current state of knowledge and personal characteristics.
- The *interface*, which handles the form of communication between the ITS and the student.

Since one of the most important features an ITS should provide is the capability to adapt its behavior to the specific traits of the student, student model seems to be the most important component and student modeling (also called learner modeling) focuses the interest of researchers from the areas of cognitive psychology, artificial intelligence and computer science.

A human teacher bases his pedagogical decisions on the information about student's learning performance obtained during the instruction, as well as by observing his problem solutions. The pedagogical module of an ITS uses the information collected by the student model during the interaction of the student with the ITS based on the actions performed by him or her. From this point of view the student model is analogous to an educational or psychological test instrument that attempts to measure student characteristics [8]. It can be seen as a representation in the system of some characteristics of a particular student [13]. Inferring a student model is called diagnosis because it uncovers the hidden cognitive state from observable behavior [21]. Student modeling is the process of creating and maintaining students models. Acquiring these models from observable behavior is a hard task because is based on guesses about the learner [12].

A lot of work has be done with artificial intelligence techniques to model student's reasoning [6] [4] [1] and with Bayesian networks to model student's behavior in a probabilistic way [3]. Fuzzy logic techniques have been used to improve the performance of an ITS due to their ability to handle imprecise information, such as student's actions, and to provide human descriptions of knowledge and of student's cognitive abilities. In the BSS1 tutoring system a general fuzzy logic engine was designed and implemented to support development of intelligent features which can better manage the student's learning [23]. Uncertainty of student's performance in Sherlock II and in the MDF tutor was managed with "fuzzy" distributions [5] [2]. For a tutoring system in the domain of physics the "Knowledge and Learning Student Model" was designed using fuzzy logic techniques, inferring about student's knowledge level and cognitive abilities from student's behavior [14] [15] . Neural networks have also been used in the design of ITSs, either simulating student's cognitive process [9] or for adaptive external control of student pacing [10].

In this work we propose a neuro-fuzzy synergism for student modeling. Fuzzy logic techniques are used to provide human-like approximate diagnosis of student's knowledge and cognitive abilities. Neural networks are trained to imitate human teacher's decisions regarding student's characteristics and fixed weight neural networks are used to evaluate and aggregate membership functions.

Evaluating student's behavior

Student's behavior is any observable response that is used as input to the student modeling process [20]. Communication channel between the student and the ITS is very restricted (usually a keyboard and a mouse) and observable responses of the student during the interaction with the ITS are limited and quite different from observable responses by a human teacher. In the neuro-fuzzy model several kinds of information, measured during the interaction, are evaluated. The number of correct or incorrect answers is evaluated with an overlay technique [24] comparing the answers with the domain knowledge, providing assessments of student's knowledge level. A specific group of questions and exercises is used to detect misconceptions. The time spent to read the theory and to find the correct answers is measured and evaluated, since the time of task is a very reliable and powerful predictor of learning [18]. The number of attempts to find the correct answer and the number of times needed to review the theory is also measured [14] [15]. Assessments of several characteristics of the student are finally created relevant to individual differences in learning performance, such as the knowledge status (knowledge level, mistakes, misconceptions, etc.), the cognitive abilities (learning speed, attention, memory limitations, etc.) [11].

The neuro-fuzzy model

Numeric data from the interaction with the ITS are measured and then transformed into linguistic terms. The process has four stages: a fuzzifier, a fuzzy relational system, a fuzzy aggregation network and a defuzzifier (see Figure 1). The student model is implemented as a set of connectionist networks, each one processing fuzzy information regarding student's behavior. The first stage of each connectionist network fuzzifies inputs that contribute to the evaluation of a specific characteristic of the student. The second stage consists of neural networks that are trained to realize fuzzy relations operated with the max-min composition. These fuzzy relations represent the estimation of human tutors to the degree of association between an observed response and a student characteristic. In the third stage a fuzzy aggregation network, utilizing the union operator, is applied for the generation of the final fuzzy set corresponding to each student characteristic. This fuzzy set is approximated by fuzzy singletons in order to feed the next stage. This last stage of each connectionist network consists of a backpropagation neural network trained to decide regarding the different characteristics, i.e. knowledge level, misconceptions, etc. of a student by classifying him to different levels (categories). The function of this network can be considered as a heuristic defuzzification procedure: the network is trained to imitate human teacher's decisions regarding student's characteristics and can be adapted to the teacher's personal view.

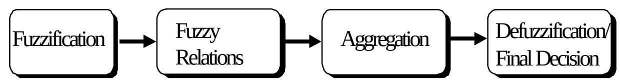

Figure 1. The four stages of the neuro-fuzzy model.

The Fuzzifier

A human teacher acquires knowledge about student's actions in an approximate subjective way (e.g. the time needed to solve the exercises was short or he answered enough questions during the lesson). The ITS collects information in numeric form that can been evaluated with sharply defined criteria. For this reason a fuzzifier is used to simulate fuzziness in human cognition calculating the membership grade of each measured value x of student's responses for every kind of information (number or answers, time, number of attempts, and so on) to each of three linguistic terms like the following: {Small, Medium, Large}. In our case, we have used regular shapes for the membership functions. Such membership functions can be calculated by fixed weight neural networks.

Depending on the kind of information k we determine different shapes based on the estimations of experts (teachers). Membership functions are subjective and context-dependent that means that it is hard for a computer system to automatically generate them in a concrete and formal way [22]. For this reason we use a second input to the fuzzifier to adjust the output, a parameter m, which allows the membership functions to be context-sensitive and adapted to teacher's personal view. In this way an input pair (xk,mk) of the kth kind of information is transformed to the triple $[y_{Sk}(x_k,m_k),\ y_{Mk}(x_k,m_k),\ y_{Lk}(x_k,m_k)]$ for the terms Small, Medium and Large respectively. Thus, we have used as membership function for the extreme terms, like Small and Large, the sigmoid function, for the intermediate term, like Medium, the pseudotrapezoidal function (composed of two sigmoid functions). The adjusting parameter m is the expected mean value of the kind of information estimated by human experts. The network of Figure 2 is used to calculate the membership grades of the three terms.

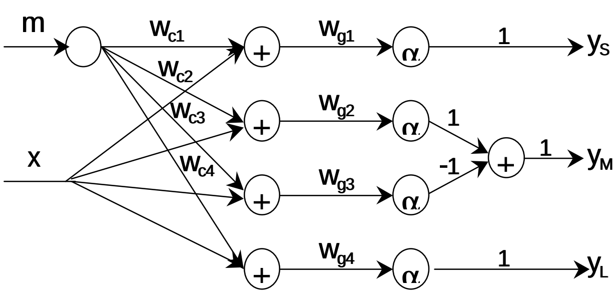

Figure 2. An instance of the fuzzifer stage.

In Figure 2, the notation indicates x the current measurement of the observed response; m the expected mean value of the kind of information according to human teachers opinions

$$y_S(x,m) = \frac{1}{1+\exp(-w_{g1}(x+w_{c1}m))}$$

$$y_M(x,m) = \frac{1}{1+\exp(-w_{g2}(x+w_{c2}m))} - \frac{1}{1+\exp(-w_{g3}(x+w_{c3}m))}$$

$$y_L(x,m) = \frac{1}{1+\exp(-w_{g4}(x+w_{c4}m))}$$

w_{ci} , i=1,..4 is the central position of the sigmoid function
w_{gi}, i=1,..4 is the gradient of the sigmoid function.
The exact shape of the membership functions is determined by the central position $m*w_{ci}$ and the gradients w_{gi}, where w_{ci}, w_{gi}, are also defined according to human teachers opinions.

The fuzzy relational system
A fuzzy relational system is used to transform the fuzzy measurements into student's characteristics (knowledge level, learning speed, attention and so on). Student's characteristics are expressed in five terms, using predicates depending on the characteristic and the predicate modifiers rather and almost. For example, for the learning speed we use : Slow, Rather Slow; Normal, Fast, Almost Fast. The fuzzy system consists of a set of connectionist networks each one realizing a fuzzy relation [16] [17] of the type

$$Y_k \circ R_{kc} = Z_c$$

where, *k* denotes the kind of information, and c the kind of student's characteristic. Y_k is the observed variable from the set $\{y_{Sk}, y_{Mk}, y_{Lk}\}$ and Z_c is the fuzzy output of the student's characteristics $\{z_{cl}, z_{c2}, z_{c3}, z_{c4}, z_{c5}\}$. R_{kc} is a 3×5 matrix representing the estimations of human teachers to the degree of association between an observed response and a student characteristic and ° is the max-min composition operator.

Since student's observable behavior is quite often inconsistent and contradictory and might be conditioned by factors such as distraction, tiredness and unintentional mistakes [4], the fuzzy relation through the min operation reduces the influence of noise in the observed responses. The fuzzy system is implemented by a set of single layer networks with 3 input nodes and 5 output nodes. The output nodes perform the max-min composition and the synaptic weights are the elements of the R matrix. The number of neural networks *n* ($n \le k*c$) is defined according to whether or not a kind of information is associated with a characteristic.

The fuzzy aggregation network
A fuzzy aggregation network is used to calculate the final fuzzy sets of student's characteristics from the several fuzzy outputs (according to the kind of information) produced by the fuzzy relational system. The network weights are evaluated using the Saaty's method [19] and determine the importance of each preliminary decision in evaluating a characteristic of the student. A preliminary decision is expressed by a fuzzy subset relating a measurement or answer to the possible qualitative characterizations of a characteristic. The union operator is used because it allows all the fuzzy outputs of a specific characteristic to contribute to the final fuzzy set [14] [15].

The defuzzifier
The defuzzifier is used to create nonfuzzy assessments of student's characteristics (for further processing by the pedagogical module). A backpropagation network is trained for this task, imitating the teachers evaluation procedure. The function of this network can be considered as a heuristic defuzzification procedure: the network is trained using the BPVS algorithm [7] to imitate the teachers evaluation procedure and can be adapted to a teacher's subjective evaluation procedure.

Discussion
The neuro-fuzzy model incorporates both general and subjective knowledge in a cognitive domain. General knowledge is incorporated in the definition of the fuzzy sets and in assigning weights of importance of each preliminary decision to evaluate a characteristic of the student. This knowledge represents the expertise of the teacher in defining the characteristics of the student. The backpropagation network represents the experience of the teacher in evaluating the students and can be adapted by training to a teacher's personal way of evaluation. Furthermore, this approach permits the representation and processing of incomplete, imprecise and vague information about the student, i.e. controversial answers and unstable behavior, as well as precise data. This evaluation of student's characteristics is further used for deciding about the appropriate teaching strategy.

Evaluation of the model

The neuro-fuzzy model has been tested on a prototype tutoring system in the physics domain of *the vertical projectory motions* [14]. The approach followed in [14] was a reconstruction of the microworld of the real lesson in the classroom. The distinguishable parts of knowledge are: free projectory motion, vertical projectory motion upward and vertical projectory motion downward. The educational material has been organized in lessons with questions/exercises that belong to several categories corresponding to particular theoretical subjects and investigating the student's characteristics. The experts (teachers) have defined the relative importance of each question, category of questions and measurement in the evaluation procedure. The expertise of the model is incorporated in the fuzzy sets, in the structure of the aggregation network and in the training phase of the backpropagation network. Depending on the input, a fuzzy subset is generated, appropriately weighted and combined with the fuzzy subsets of the other contributed measurements or answers. All this information is used to finally decide regarding a characteristic of the learner.

Experiments have been performed using a population of 300 simulated student cases to compare and evaluate the performance of the new approach with the decisions of 5 teachers. The overall average classification success has been 95%. Below, we exhibit results regarding the average success in evaluating student's learning speed, (learning speed is a cognitive ability of the student). In Figure 3 the average performance is analyzed with respect to the five possible characterizations for the learning speed. The system exhibits good performance in evaluating *slow*, *rather slow*, *normal* and *almost fast* students. However, *fast* students have been badly classified. This was expected because the system has been trained on the basis that teachers rarely classify a student in the best category.

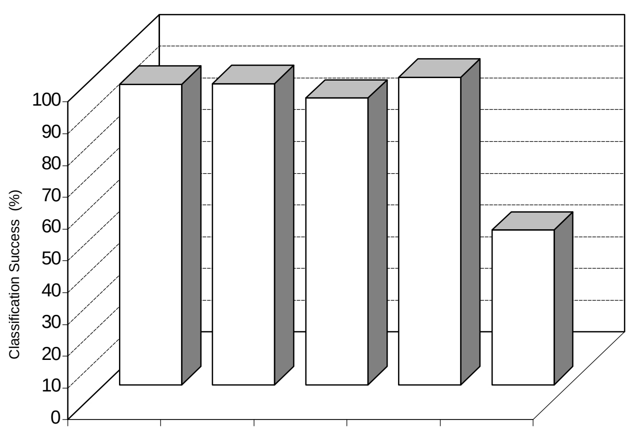

Figure 3. Average classification success in categorizing students with respect to their learning speed.

We have further analyzed the errors made by the system and we have identified two types. The type 1 error expresses that a student has been badly classified in a adjoining category. For example, this type of error occurs when a student is evaluated by a human expert as *normal* (regarding his learning speed), but the system classifies him in the category *almost fast* or *rather slow*. On the other hand, when this student is classified as *slow* or *fast* the type 2 error occurs. Note that type 2 errors have occurred only when *normal* students have been evaluated. Detailed results are presented in Figure 4. To evaluate the generalization performance three different rules have been used and indicated in the figure by THL=0, THL=0.7 and THL=0.5. THL=0 indicates the *max* rule, i.e. a student is considered correctly classified if the output neuron that corresponds to the correct category has the greatest value among the output neurons; THL=0.7 indicates that a threshold equal to 0.7 has been placed in the neurons output before applying the *max* rule. Thus, only neurons with activation value greater than 0.7 contribute to the evaluation. Output patterns that contain values less than 0.7 are now considered as "Do not know" cases and the corresponding characteristic is not evaluated for the student.

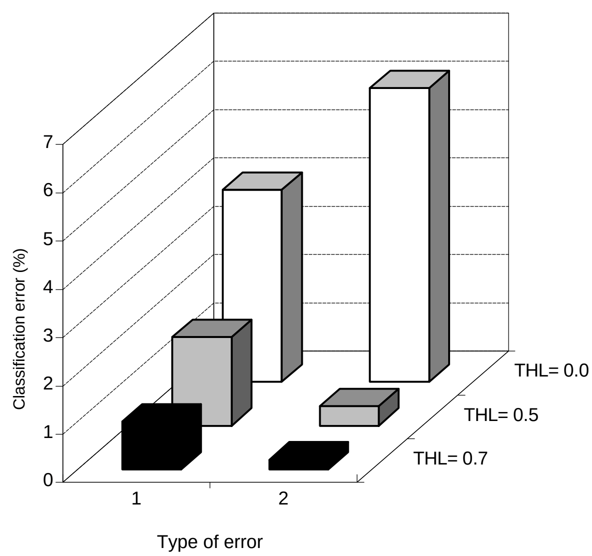

Figure 4. Classification error in categorizing *normal* students with respect to the type of the error.

In Figure 5 the classification error for *almost fast* students is exhibited (in this case only type 1 error occurred).

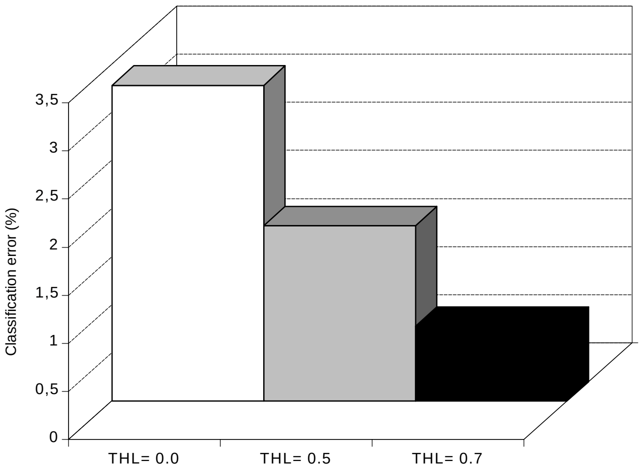

Figure 5. Type 1 classification error in categorizing *almost fast* students.

Conclusions

In the proposed approach student's evaluation depends on the designer's ability to analyze the cognitive domain suitably, define fuzzy sets and relate student response with appropriate knowledge and cognitive characteristics. The applicability of the evaluation procedure can further be extended by exploiting the training and generalization capabilities of the neural networks to extract information from existing student records. These records implicitly contain a true picture of the possible knowledge levels of the students and of the possible learning paths. We currently investigate techniques to enhance the generalization performance of our approach.

References

[1] Aiello, L. Cialdea, M. and Nardi, D. (1993). Reasoning about Student Knowledge and Reasoning. Journal of Artificial Intelligence in Education, 4(4), p. 397-413

[2] Beck,J Stern,M and Woolf, B,P. Using the Student Model to Control Problem Difficulty. World wide web document

[3] Conati, C. Gertner, A. VanLehn, K. Druzdzel, M.(1997). On-Line Student Modeling for Coached Problem Solving Using Bayesian Networks. User Modeling: Proceedings of the Sixth International Conference Available on-line from

[4] Giangrandi, P. and Tasso, C.(1995).Truth Maintenance Techniques for Modelling Student's Behaviour. Journal of Artificial Intelligence in Education,6(2/3), p. 153-202

[5] Katz,S. Lesgold, A Eggan, G Gordin, M. (1992)Modelling the student in Sherlock II, Journal of Artificial Intelligence in Education Vol 3 No 4 p 495

[6] Kono, Y. Ikeda, M. and Mizoguchi, R. (1994). THEMIS: A Nonmonotonic Inductive Student Modeling System. Journal of Artificial Intelligence in Education, 5(3), p. 371-413

[7] Magoulas, G.D., Vrahatis, M.N., and Androulakis, G.S.(1997), Effective back-propagation training with variable stepsize, Neural Networks, 10, 69-82.

[8] Mark, M. A. and Greer, J. E (1993). Evaluation Methodologies for Intelligent Tutoring Systems. Journal of Artificial Intelligence in Education 4(2/3),p. 129-153

[9] Mengel, S. and Lively, W. (1991). On the use of neural networks in intelligent tutoring systems, Journal of Artificial Intelligence in Education,2, 43-56.

[10] Montazemi, A.R. and Wang, F.(1995). On the Effectiveness of a Neural Network for Adaptive External Pacing, Journal of Artificial Intelligence in Education, 6(4), p.379-404

[11] Nawrocki (1987). Maintenance in training. In Greg Kearsley (Eds).Artificial Intelligence and Instruction. Application and methods. Addisson-Wesley

[12] Paiva,A. Self, J. (1994). A Learner Model Reason Maintenance System. ECAI 94. 11th European Conference on Artificial Intelligence. Edited by A. Cohn. Published in 1994 by John Wiley & Sons, Ltd.

[13] Paiva,A. Self, J. Hartley, R. (1994).On the Dynamics of Learner Models.. ECAI 94. 11th European Conference on Artificial Intelligence. Edited by A. Cohn. Published in 1994 by John Wiley & Sons, Ltd.

[14] Panagiotou, M. Grigoriadou, M. (1995). An Application of Fuzzy Logic to Student Modeling. Proceedings of the IFIP World conference on Computer in Education. (WCCE95). Birmigham

[15] Panagiotou, M. Grigoriadou, M. and Philokiprou G. (1994). Knowledge and Learning Student Model. In Vosniadou S. (Eds). Technology-based Learning Environments, NATO ASI series F, Springer, p. 243-249

[16] Pedrycz, W. (1984). Indentification in Fuzzy Systems. IEEE Transactions on Systems, Mans and Cybernetics, Vol.SMC-14, No 2, March-April 1984, p. 361-366

[17] Pedrycz, W. (1991).Neurocomputations in Relational Systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 13, No 3, March 1991, p. 289-296.

[18] Renkl, A.(1997) Learning from Worked-out Examples: A study on Individual Differences. Cognitive Science Vol. 21,No 1,January-March,1997, p.1-29

[19] Saaty, T. L.(1978) Exploring the interface between hierarchies, multiple objectives and fuzzy sets, Fuzzy sets and systems, 1, 57-68.

[20] Sison, R. and Shimura, M.(1996). The Application of Machine Learning to Student Modeling (A 1996 Perspective): Toward a Multistrategic Learning Student Modeling System. World wide web document

[21] VanLehn Kurt (1988). Student Modeling. In M.C.Polson and J.J. Richardson (Eds). Foundations of Intelligent Tutoring Systems. Lawrence Elbaum Associates Publishers

[22] Wang P. (1993). The Interpretation of Fuzziness. World wide web document.

[23] Warendorf K. and Tsao S. J. (1997). Application of Fuzzy Logic Techniques in the BSS1 Tutoring System. Journal of Artificial Intelligence in Education,8(1), p.113-146

[24] Wenger (1987).Artificial Intelligence and Tutoring Systems. Morgan Kaufmann Publishers, Inc.