

Fuzzy Inference for Student Diagnosis in Adaptive Educational Hypermedia

Maria Grigoriadou¹, Harry Kornilakis¹, Kyparisia A. Papanikolaou¹, George D. Magoulas²

¹Department of Informatics & Telecommunications, University of Athens,
Panepistimiopolis, GR-15784 Athens, Greece
{gregor, harryk, spap}@di.uoa.gr

²Department of Information Systems and Computing, Brunel University, UB8 3PH, U.K.
George.Magoulas@brunel.ac.uk

Abstract: In this paper we propose a method that implements student diagnosis in the context of the Adaptive Hypermedia Educational System INSPIRE - Intelligent System for Personalized Instruction in a Remote Environment. The method explores ideas from the fields of fuzzy logic and multicriteria decision-making in order to deal with uncertainty and incorporate in the system a more complete and accurate description of the expert's knowledge as well as flexibility in student's assessment. To be more precise, an inference system, using fuzzy logic and the Analytic Hierarchy Process to represent the knowledge of the teacher-expert on student's diagnosis, analyzes student's answers to questions of varying difficulty and importance, and estimates the student's knowledge level. Preliminary experiments with real students indicate that the method is characterized by effectiveness in handling the uncertainty of student diagnosis, and is found to be closer to the assessment performed by a human teacher, when compared to a more traditional method of assessment.

Keywords: Student Diagnosis, Fuzzy Logic, Analytic Hierarchy Process, Adaptive Educational Hypermedia Systems.

1 Introduction

Adaptive Educational Hypermedia Systems (AEHSs) (Brusilovsky, 1996; 1999) constitute a new generation of Educational Hypermedia (EH) systems, which possess the ability to make intelligent decisions about the interactions that take place during learning aiming to support students without being directive. Such systems build a model of the goals, preferences and knowledge of each individual student and use this model throughout the interaction with him/her for adapting the content and/or the navigation to the needs of the particular student. Thus, the quality of personalized instruction offered by an AEHS is largely determined by the coverage and accuracy of the information constructing the student model and by the ability of the system to dynamically update it. As the student model is used as a source of system's adaptation, in most cases it includes information regarding student's behavior and knowledge, which have

repercussions for his/her performance and learning. However, the construction of such a model is a research challenge from both the Instructional Design and Artificial Intelligence (AI) perspectives, involved in the design of student interaction with the system.

However an educational system, due to the restricted communication channel, is only able to directly obtain raw measurements, by monitoring the interaction with the student. The process of inferring students' internal characteristics from their observable behavior is called *student diagnosis* (VanLehn, 1988). Important issues outlining student diagnosis refer to: (i) the observable student's behavior that should be recorded in terms of specific measurements, (ii) the internal characteristics that can be inferred based on the recorded information and that are important to learning, and (iii) the method adopted for extracting this information through student monitoring and tracking. Thus, with regards to the AI perspective, the main demand is the development of a reliable method that will be able to analyze effectively, in a way a teacher would follow, measurements regarding student's behavior and make estimations on student's internal characteristics updating the student model accordingly. This model will be further used to guide system's adaptive behavior. The main obstacle in the diagnosis process is uncertainty coming partly from the communication among the teacher, the developer and the system and partly from inaccuracies in the measurements conducted.

In this paper we present the method for *student diagnosis* that is being used for supporting the adaptive capabilities of INSPIRE - INtelligent System for Personalized Instruction in a Remote Environment, which is a Web-based AEHS for distance learning, recently developed at the laboratory of "Educational & Language Technology" of the department of Informatics and Telecommunications, University of Athens. In Section 2 an overview of INSPIRE is presented. Section 3 examines the individualities of the student diagnosis problem and proposes several technologies in order to deal with them. In Section 4 the method used for student diagnosis combining the analytic hierarchy process and fuzzy logic is presented. Furthermore, the way this process exploits teacher's expertise in the assessment procedure and simulates his/her individual way of assessing students' knowledge level, is presented. In Section 5 an example of the diagnostic process is shown and the experimental results are discussed. The paper ends with concluded remarks on the advantages and disadvantages of the proposed method and further research.

2 An Overview of INSPIRE

INSPIRE, (Papanikolaou et al, 2001), is an AEHS that aims to assist distance students during their study by constructing and presenting lessons that correspond to specific learning outcomes, accommodating student's knowledge level and learning style. This process of content personalization requires, apart from the information of student's learning style, a thorough knowledge of the student's knowledge level. To this end a number of assessment tests have been developed for INSPIRE, each of them assessing the student's knowledge on the main topics of the domain that s/he studies. Based on

the performance of the student on these assessment tests, INSPIRE makes estimations on the knowledge level of the student on the various topics using the student diagnosis process that will be described below in Section 4.2. These estimations are then used to personalize the content that will be delivered to the student. In the following by student diagnosis we refer to the above process of deducing the students' knowledge level on each topic (internal characteristics) from their answers to assessment tests (observable behavior).

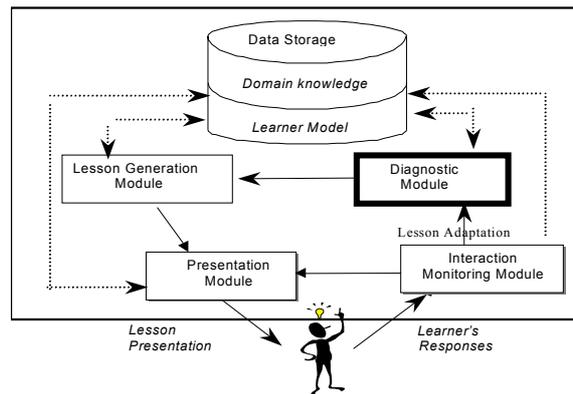


Fig. 1. Schematic of INSPIRE's architecture

INSPIRE is comprised of four modules and the data storage (Fig.1). The modules of the system are: (i) the *Interaction Monitoring Module (IMM)* that monitors and handles student's responses, including answers to tests, during his/her interaction with the system, (ii) the *Student's Diagnostic Module* that processes data recorded about the student and decides on how to classify the student's knowledge and learning style, (iii) the *Lesson Generation Module (LGM)* that generates personalized lessons following students' knowledge level and (iv) the *Presentation Module (PM)* whose function is to present the lessons created by the LGM to the student following his/her learning style. The data storage contains the domain knowledge, and the student model, which is the data structure that holds all the information that the system has gathered about the student, and upon which the Diagnostic Module acts. This information includes the number and type of questions the student tried to answer, the attempts s/he made to answer each question, the time s/he has spent on each page, the percentage of the study time that s/he has devoted to each type of material (presentations, examples, simulations etc.) and similar other measurable quantities.

In this paper, we will focus on the operation of the Diagnostic Module, which appears highlighted in Fig.1. The Diagnostic Module receives its input from the IMM, which gathers numeric information about the interaction of the student with the system. Especially, we shall concentrate on the output of the Diagnostic Module that provides an estimation of student's knowledge level in the domain of interest. The LGM uses this information further, in order to generate the personalized content that will be delivered to the student.

3 The Problem of Student Diagnosis

The presence of uncertainty is an important factor that often leads to errors in student diagnosis. This uncertainty appears partly due to errors and approximations involved when gathering data from measurements, and partly due to the abstract nature of human cognition and the loss of information resulting from its quantification. Uncertainty in measurements stems from several factors such as careless errors and lucky guesses in the student's responses. In an educational system where there is no direct interaction between the tutor and the student the collected data tend to be more haphazard, than those obtained through traditional face-to-face interaction. Furthermore, it is harder for these systems to rely upon background information and relevant experience, as human tutors can (Jameson, 1996). Especially in a web-based learning environment inaccurate measurements caused by technical difficulties, such as network congestion, cannot be ignored. On the other hand, when trying to explicitly represent the mental and emotional states and processes, an additional layer of approximation is introduced. The student's knowledge is constantly changing during the dynamic process of learning and it is therefore quite difficult to be certain about his/her current mental state. Considering the above attributes of the problem, it is obvious that the development of a reliable method for student diagnosis is based on successful handling of uncertainty.

However, the diagnostic process relates to the way a human-expert assesses students' knowledge level on a certain topic focussing on how the assessment tests are marked by him/her. To this end, different approaches are applied, such as the *norm-referenced assessment* that is traditionally used in end examinations and the *criterion-referenced assessment* that is associated with continuous (or intermittent) assessment so that many more of the lesson objectives and competences are assessed (Reece & Walker, 1997). In the first approach tests are marked so that the normal distribution is achieved while in the second one the assessment process is based on specific criteria that are defined in terms of objectives and competences which state what the students must achieve. Thus, the way that the teacher's expertise in assessment is incorporated in the system in order to guide the diagnostic process is an important issue influencing the efficiency of the provided estimations. Furthermore, the system should allow teachers that use the system to convey this knowledge in a simple and comprehensible manner without being forced to make complicated quantifications of abstract knowledge.

In our case, the diagnostic process is based on the criterion-referenced assessment, which is considered as a part of the developmental process of learning aiming to assess the students' quality of learning. This way the educational system is continuously provided with information on students' performance in order to be able to adapt its output accordingly. As a method of dealing with uncertainty and incorporating teacher's expertise and flexibility in student's assessment, we use a combination of fuzzy logic with a multicriteria decision-making approach.

Zadeh (1965) was the first to introduce fuzzy logic and fuzzy systems as a method to handle numerical uncertainty and express imprecision and subjectivity in human thinking. Use of fuzzy logic in numerous applications has shown that it offers high expressive powers, an enhanced ability to model real-world problems, and a method-

ology for building systems tolerant to imprecision and uncertainty (Lin & Lee, 1996). In multicriteria decision-making, Saaty's Analytical Hierarchy Process (AHP) is widely used to define the relative importance of a number of criteria (Saaty, 1980), which in our case emulate the criteria used by the expert-teacher in order to assess student's knowledge level.

The approach proposed in this paper builds on previous results reported in (Panagiotou & Grigoriadou, 1995; Magoulas et al., 2001) and enhances INSPIRE with the ability to consider multiple criteria simultaneously. Usually the process of assessing student's knowledge level is influenced by several conditions to which the expert adapts the assessment, such as the current knowledge level of the student, the topic being considered, etc. Thus, defining the relative importance of the criteria used according to several preconditions provides the system with knowledge coming from teacher's expertise in the assessment procedure and makes the system flexible enough to accommodate the personal way of assessment of each individual teacher.

4 Student Diagnosis in INSPIRE

Lessons in INSPIRE are generated so as to lead the student to the accomplishment of a specific learning goal, which corresponds to a topic of the domain knowledge. Each learning goal is associated with a subset of outcome topics, in which one must become proficient in order to accomplish the learning goal.

In order to measure the student's knowledge at each of the outcome topics, assessment tests have been developed. Each assessment test covers the material of one topic and it is available to the student while s/he is studying that topic. Questions of an assessment test are grouped in several categories that correspond to specific abilities that the student must exhibit and which are in accordance with the three levels of performance proposed by Merrill (1983): (i) *Remember Level*: questions that test the ability of students to recall the provided information, (ii) *Use Level*: questions that test the ability of students to apply the provided information to specific case(s), (iii) *Find Level*: questions that test the ability of students to propose and solve original problems.

When the student selects to take the assessment test, the questions appear one after the other with increasing difficulty, i.e. the easier questions of the Remember Level appear first, then the questions of the Use Level and finally the questions of the Find Level. At any point the student has the option to stop taking the test. Based on the answers given to the questions of a specific topic, we want to make an estimation of the knowledge level of the student on that topic. That estimation should be as close as possible to the way an experienced teacher evaluates a student. We use a qualitative model, which classifies knowledge on a topic to one of the four levels of proficiency {Insufficient, Rather Insufficient, Rather Sufficient, Sufficient}.

Ultimately, the goal of the diagnosis is to obtain information about the knowledge of the student in each topic, in terms of the four characterizations and in a way an expert teacher would. In order to achieve this we need to model the knowledge and experience of the expert and also to model the inference process used by the expert.

4.1 Modeling the Expert's Knowledge

As valuable resources in modeling teacher's expertise in assessing students' knowledge as well as in modeling teacher's personal way of assessing, are considered:

- the criteria that the teacher defines in order to assess student's knowledge level
- teacher's estimations of the importance of different types of assessment questions that correspond to the above criteria, with respect to the student's knowledge level at the time s/he asks to be assessed and the type of the topic under consideration, i.e. a theoretical concept or a procedure etc., and
- teacher's estimations of the relationship between student's correct answers and his/her proficiency of the topic.

Definition of Criteria. Criterion-referenced assessment is associated with the concept of mastery learning, which is important in cases where students need to master a topic prior to moving onto another one (Reece & Walker, 1997). The process of assessing students' knowledge on a certain topic is facilitated by the introduction of specific criteria given in terms of objectives and competences which state what the student must achieve on this topic. These criteria guide also the marking process, e.g. students can achieve full marks if they attain the required standard suggested, or, their marks differentiate according to the objectives of a topic that they achieved.

In our approach, the diagnostic process for assessing students' knowledge level uses three qualitative criteria. These criteria correspond to the three levels of performance *Remember*, *Use*, *Find* (see Section 4), aiming to assess the relative students' abilities.

Importance of Different Types of Questions. As we have mentioned assessment tests in INSPIRE consist of questions of three different categories (see Section 4). The importance of the questions of each category may vary depending on the topic that the questions assess and on the proficiency of the student at the time s/he takes the assessment test. For example, for the topic "The role of cache memory" knowledge of the theory (Remember Level) is more important, while for the topic "Mapping techniques" it is more important that the student is able to solve problems on that topic (Use Level). The importance of the questions of the different categories is one aspect of the knowledge of the teacher that should be modeled when performing student diagnosis. In order to assist the teacher to convey this knowledge to the system we use the AHP for assigning weights to the different criteria expressing their relative importance. These criteria correspond to the three different categories of questions and consequently their weights are also considered as weights of the corresponding questions. The weights of the various criteria on which the system is based in order to evaluate the knowledge level of the student, change depending on:

- *The knowledge level of the student at the time s/he asks to be assessed.* For example, when the student is a novice on a topic, the weights are specified so that the questions that assess the understanding of the theory (1st criterion) have a greater weight compared to those that assess application of the theory (2nd criterion). In other words, we assume that the student should initially study the theory - Remember Level - and then continue with the application of the theory. As the student progresses, his/her knowledge level changes from {Insufficient} to

{Rather Insufficient} (s/he has covered the theory and should move on to the application), the weights of the criteria change, and the weight of the questions of the Use Level increases. After this change, in order to reach a {Rather Sufficient} knowledge level the student should answer the questions about the application and so on.

- *The type of the topic that is being examined.* For example, if a topic is a procedure, then students should learn mainly how to apply it in different cases, and thus the application (2nd criterion) should have greater weight compared to the theory (1st criterion). Accordingly, for a more theoretical topic, understanding the theory is more important compared to its application.

In more detail, the AHP offers a framework that lets someone specify the importance of a number of different criteria, by giving linguistic comparisons expressing the relative importance between pairs of criteria. Suppose that we have n criteria c_1 through c_n and wish to specify the importance of these criteria. According to AHP we only need to give as input their relative importance. For each pair of criteria c_i and c_j , a value a_{ij} , between 1 and 9, is specified, declaring the relative importance of criterion c_i over c_j . For example $a_{ij}=1$, means ‘ c_i is as important as c_j ’, $a_{ij}=2$, means ‘ c_i is slightly more important than c_j ’, up to $a_{ij}=9$, which means ‘ c_i is extremely more important than c_j ’. Based on these values we generate the pairwise comparison matrix A as follows:

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix}, \quad a_{ij} \in \{1,2,\dots,9\}, \text{ and } a_{ji} = 1/a_{ij} \quad (1)$$

Then the weight w_i of each criterion c_i is calculated as:

$$w_i = \frac{\left(\prod_{j=1}^n a_{ij} \right)^{1/n}}{\sum_{i=1}^n \left(\prod_{j=1}^n a_{ij} \right)^{1/n}} \quad (2)$$

By letting the teacher specify different weights to the criteria assessing student’s knowledge for each topic, it is possible to take into account specific characteristics of the topic, such as if a topic is theory oriented or application oriented, etc. Furthermore, by using different weights for the novice and for the more advanced student it is possible to adapt the diagnosis to his/her current knowledge level. Therefore, knowledge of the theory will be more important when the student is a novice, while as s/he becomes more familiar with a topic, being able to apply the theory and solve problems becomes more important.

In our case we have three categories of questions (see Section 4), which correspond to three different criteria for assessing student’s knowledge level (these are in accordance with the three levels of performance Remember, Use, Find). For each

topic, the teacher specifies the relative importance of each criterion to the other for the case that the student is a novice, for the case that s/he is more advanced and so on. Therefore, for each topic, multiple 3×3 pairwise comparison matrices are specified, each one corresponding to the state of the learner before taking the test.

Relationship between Correct Answers and Proficiency. The marking of criterion-referenced assessment, as already mentioned, relates to the criteria (objectives/competences) defined by the teacher (Reece & Walker, 1997). The teacher designs questions that assess student's competences in terms of certain objectives and then s/he relates the percentage of questions that a student has answered correctly to the knowledge of the student on the specific topic. In INSPIRE we try to model this marking process through the use of fuzzy sets aiming to combine quantitative measurements (number of right answers in different types of questions) in order to get qualitative characterizations for the student's knowledge.

By the term fuzzy set we mean a function $f: U \rightarrow [0,1]$, where U is the universe of discourse of the function. The value $f(x)$ of the function for an input x , represents the degree of membership of x in the fuzzy set. For example let's suppose we have the fuzzy set "Insufficient knowledge of the Remember Level". The function f is equal to $f(x) = \{1, 0.6, 0.3, 0.1, 0, 0, 0, 0, 0, 0\}$, where $x = \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90\}$ is the percentage of the questions of the Remember level that the student has answered correctly. Then, for input equal to 10 the value of the function f is $f(10) = 0.6$. The interpretation of this is that the knowledge on the Remember level of a student who has answered 10% of the questions correctly, can be considered {Insufficient} to the degree of 0.6. The degrees of membership can be extracted from the teacher by asking questions such as, "How much do you consider that someone's knowledge on theory is Insufficient, if s/he has answered 10% of the questions on theory correctly?". Note that the universe of discourse is discretized, which results in working with fuzzy sets that have 10 elements. The actual percentage of correct answers is obviously a continuous value, but for practical purposes we make it discrete by rounding it to a multiple of ten.

In total, we need the teacher to provide us with twelve such fuzzy sets. One fuzzy set is required for each of the three different criteria {Remember, Use, Find} and for each of the four levels of proficiency {Insufficient, Rather Insufficient, Rather Sufficient, Sufficient}. So, for example, we will have fuzzy sets describing "Insufficient knowledge of the Remember Level", "Rather Insufficient knowledge of the Remember Level", "Sufficient knowledge of the Use Level" etc. We shall call these fuzzy sets f_L^P , where $L \in \{R, U, F\}$ and $P \in \{I, RI, RS, S\}$. The fuzzy set f_L^P will represent a proficiency level equal to P on the L level of performance.

4.2 The Diagnostic Process

After the student has answered a question of an assessment test, the diagnostic process begins. The diagnosis aims to estimate the knowledge level of the student on a specific topic, i.e. on the topic that the answered question refers to.

We *first* need to divide the number of correctly answered questions of each category by the total number of questions for that category, in order to calculate the percentage of correctly answered questions on each different category. Afterwards that value is rounded to the closest multiple of ten percent as the discrete fuzzy sets have 10 elements with values $\{0, 10, 20, \dots, 90\}$ (see previous section). Thus, we get three percentages of correct answers, one for each category of questions. Let us call these percentages x_R, x_U, x_F for the Remember, Use and Find level respectively.

Then, using these three values and the twelve fuzzy sets that the teacher specified (see previous section), we form the matrix D , containing the degrees of membership of the knowledge level of the student to each of the twelve fuzzy sets.

$$D = \begin{bmatrix} f_R^I(x_R) & f_R^{RI}(x_R) & f_R^{RS}(x_R) & f_R^S(x_R) \\ f_U^I(x_U) & f_U^{RI}(x_U) & f_U^{RS}(x_U) & f_U^S(x_U) \\ f_F^I(x_F) & f_F^{RI}(x_F) & f_F^{RS}(x_F) & f_F^S(x_F) \end{bmatrix} \quad (3)$$

At this *point* we need to consider the effect that each of the three criteria will have on the final diagnosis. As we mentioned in our discussion about the importance of the different types of questions, the teacher has specified their importance for each topic and for different levels of students' proficiency before taking the test in the weights w_i . So, based on the current topic and on the knowledge level of the student at the time s/he asks to be assessed, the appropriate vector $W=[w_R, w_U, w_F]$ is selected, where w_R is the weight for the questions assessing the Remember Level (1st criterion), w_U for the questions assessing the Use Level (2nd) and w_F for the questions assessing the Find Level (3rd). By multiplying the vector W by the matrix D , we calculate the vector $P=W \cdot D$, which is the degree of membership of the student's knowledge in each of the four proficiency levels, with respect to all three criteria. Thus, we get the vector $P=[p_1, p_2, p_3, p_4]$, where p_1 is the degree to which the student's knowledge is {Insufficient}, p_2 is the degree to which it is {Rather Insufficient}, etc.

Finally, as we have calculated the vector P it is possible for us to give a final estimation on the knowledge level of the student on the topic. The vector P contains the estimation on the knowledge level with respect to each of the four possible levels {Insufficient, Rather Insufficient, Rather Sufficient and Sufficient}. In order to reach a final result we need to combine the four elements of the vector P , so as to select one of the four alternative levels. This is performed using the Center of Gravity method, according to which we calculate the number v as follows (Lin & Lee, 1996):

$$v = \frac{\sum_{i=1}^4 p_i i}{\sum_{i=1}^4 p_i} \quad (4)$$

and then round v to the nearest integer. Depending on the result, we make the final estimation on the student's knowledge level. So, if $round(v) = 1$, we characterize the knowledge of the student on the topic as {Insufficient}, if $round(v) = 2$ as {Rather Insufficient} and so on.

5 Experimental Results

The data presented in this section come from an experiment, which performed as a part of the formative evaluation of the system, aiming to evaluate the adaptive dimension of INSPIRE. Specifically, the answers that the students gave to an assessment test, were used in order to check the validity of the performance of the diagnostic module of INSPIRE. To this end, the output of the diagnostic module concerning students' knowledge level was compared with the diagnosis of an expert-teacher and with the simple diagnostic process of calculating the percentage of right answers, a method adopted in many AEHSs. In the experiment participated 20 postgraduate students of the Department of Informatics and Telecommunications of the University of Athens, and the professor of the course who had the role of the expert-teacher. The students had already studied the handouts of the module "Computer Architecture" and they had been examined on the module. They worked independently, one on each computer. The students accessed INSPIRE through a common browser and they studied the learning goal "What is the role of cache memory and which are its basic operations" for a period of one hour.

All the tasks that the students had to perform were listed in following a usage scenario (Carroll & Rosson, 1990). Through the scenario and after the students had studied the educational material of the topic "Mapping Techniques", they were asked to submit the corresponding assessment test. The assessment test consisted of fifteen questions organized as follows: (i) seven of them tested the Remember Level, (ii) five of them tested the Use Level and (iii) three questions tested the Find Level.

As each question of the test was being submitted by the student, the diagnostic process of INSPIRE estimated his/her knowledge level on the topic and his/her model was then updated accordingly. The vector of weights of the three criteria, used for assessing student's knowledge (three types of questions) for the topic "Mapping Techniques", had been calculated using the AHP (see Relations (1)-(2)) based on the relative importance of the criteria as provided by the professor before the experiment and was equal to: (i) $W = [0.1775, 0.5190, 0.3035]$ for novice students, i.e. those whose knowledge level before answering the question was {Insufficient or Rather Insufficient}, and (ii) $W = [0.1047, 0.2583, 0.6370]$ for more experienced students, i.e. those with knowledge level {Rather Sufficient or Sufficient}. The final estimation of the student's knowledge level was the one made after the student had submitted every question of the test (see in Fig. 2 the row labeled "INSPIRE").

After the experiment was completed, the professor examined students' answers to the test and estimated their knowledge level on the topic, based on the number, the type and difficulty of the correctly answered questions and the general impression given by the test (see in Fig. 2 the row labeled "Expert"). Furthermore, we estimated the students' knowledge level based on the percentage of correct answers they had given, based on the heuristic rules that *if* the percentage of correct answers is (see Fig. 2 - in the row labeled "Percentage"): (i) between 0% and 25% *then* the knowledge level is considered as {Insufficient}, (ii) between 26% and 50% *then* the knowledge level is considered as {Rather Insufficient}, (iii) between 51% and 75% *then* the knowledge level is considered as {Rather Sufficient} and (iv) over 75% *then* the knowledge level is considered as {Sufficient}.

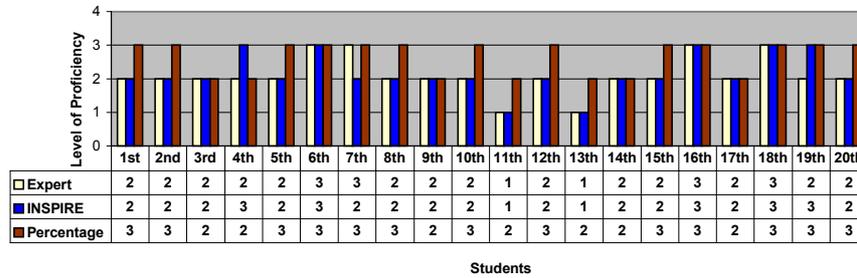


Fig. 2. The estimations of the knowledge level of 20 students on the topic "Mapping techniques" using different methods of assessment: an Expert-teacher, INSPIRE and the Percentage of students' correct answers. The vertical axis shows the level of proficiency, {Insufficient, Rather Insufficient, Rather Sufficient, Sufficient} which corresponds to {1,2,3,4}. These values are also summarized in the data table below the chart.

From the results of Fig. 2 one can observe that the estimations made by INSPIRE's diagnostic module and the teacher coincide in 17 out of the 20 student cases. On the other hand, only in the case of 9 out of the 20 students the teacher's estimations are the same as estimations based on the percentage of correct answers. Even though the sample is rather small to reach a safe conclusion, the results indicate that INSPIRE can indeed perform diagnosis in a way that gives results similar to the way that a teacher evaluated students.

6 Conclusions

In this paper, the problem of student diagnosis was investigated, as it appears in the context of the adaptive educational hypermedia system INSPIRE, and several difficulties that arise when trying to perform student diagnosis, were pointed out.

A method making use of ideas from the fields of fuzzy logic and multicriteria decision-making has been proposed in order to deal with uncertainty and to incorporate in the system a more complete and accurate description of the expert's knowledge and flexibility in student's assessment. This way the assessment procedure takes into account the individual teacher's personal style of assessing as well as the current knowledge level of the student and accordingly adapts the relative importance of the selected criteria for assessing student's knowledge.

Experimental results have been encouraging, even performed on a limited test group and show that the student diagnosis performed by the proposed method is close to the teacher-expert estimations. Further investigation of the effect of the different parameters and structural features of the proposed diagnostic process, through a sensitivity analysis (VanLehn & Niu, 2001), is necessary in order to determine their influence in the accuracy of the assessment and adjust them accordingly.

References

1. Brusilovsky, P.: *Methods and Techniques of Adaptive Hypermedia. User Modeling and User-Adapted Interaction*, Vol.6. Kluwer Academic Publ., Netherlands (1996) 87-129
2. Carroll, J.M., Rosson, M.B. (1990). Human-computer interaction scenarios as a design representation. In *Proc. of IEEE HICSS-23, 23rd Hawaii Intl. Conf. System Sciences Vol. II*, 555-561.
3. Papanikolaou, K., Grigoriadou, M., Kornilakis, H., Magoulas, G.: INSPIRE: An Intelligent System for Personalized Instruction in a Remote Environment. In: Reich, S., Tzagarakis, M.M., De Bra, P. M.E. (eds.): "Hypermedia: Openness, Structural Awareness and Adaptivity". *Lecture Notes in Computer Science Vol. 2266*. Springer-Verlag, Berlin (2001)
4. Jameson, A.: Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues. *User Modeling and User-Adapted Interaction* **5:3/4** (1996) 193-251
5. Lin, C.T., Lee, C.S.G.: *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*. Prentice Hall P T R Upper Saddle River, New Jersey (1996)
6. Magoulas, G.D., Papanikolaou, K.A., Grigoriadou, M.: Neuro-fuzzy Synergism for Planning the Content in a Web-based Course. *Informatica* **25:1** (2001) 39-48
7. Merrill, M.D.: *Component Display Theory*. In: C.M.Reigeluth (ed.), *Instructional design theories and models: An overview of their current status*. Lawrence Erlbaum Associates Hillsdale NJ (1983)
8. Panagiotou M., Grigoriadou M.: An Application of Fuzzy Logic to Student Modeling. In: *Proceedings of the IFIP World conference on Computer in Education. (WCCE95)*, Birmigham (1995)
9. Reece, I., Walker, S.: *Teaching, Training and Learning. A Practical Guide*. Third Edition. Business Education Publishers Limited, Sunderland (1997)
10. Saaty T.: *The Analytic Hierarchy Process*, McGraw-Hill, New York (1980)
11. VanLehn, K.: Student Modeling. In: Polson, M. C., Richardson, J.J. (eds.): *Foundations of Intelligent Tutoring Systems*. Lawrence Erlbaum Associates Hillsdale, New Jersey (1988) 55-78
12. VanLehn, K., Niu, Z.: Bayesian student modeling, user interfaces and feedback: A sensitivity analysis. *International Journal of Artificial Intelligence in Education* **12** (2001)
13. Zadeh, L. A.: Fuzzy Sets. *Information and Control* **8:3** (1965) 338-353