Johanyák, Z.C.: Survey on Three Fuzzy Inference Based Student Evaluation Methods, Proceedings of the 10th International Symposium of Hungarian Researchers – CINTI 2009, November 12-14, Budapest, pp. 185-192. http://johanyak.hu

Survey on Three Fuzzy Inference Based Student Evaluation Methods

Zsolt Csaba Johanyák

Institute of Information Technologies, Kecskemét College, H-6000 Kecskemét, Izsáki út 10, johanyak.csaba@gamf.kefo.hu

Abstract: The evaluation of students' academic performance in cases when a total automated scoring is not possible can result in quite significant differences between the marks given by different evaluators or at different occasions. The problem partly can be traced back to the uncertainty in the opinion of the evaluator that cannot be expressed properly in the traditional one-value-based scoring. Fuzzy set theory based evaluation methods can reduce the mentioned differences. In this paper after defining a criterion set for the evaluation and comparison we do a survey on three fuzzy inference based student scoring methods.

Keywords: fuzzy student evaluation, rules based system, fuzzy inference, fuzzy rule interpolation

1 Introduction

In case of the evaluation of students' answerscripts containing narrative responses quite often there is vagueness in the opinion of the evaluator that hardly can be fitted in the frames of the traditional evaluation techniques where a response is rated by a single crisp value. Therefore this area could be a very good application field for fuzzy theory based evaluation methods.

Recently several fuzzy methods have been published in order to deal with this problem. They can be classified in two main groups: (1) methods applying fuzzy inference, and (2) methods applying "only" fuzzy arithmetic. The advantage of the first approach is that the rules are easily readable and understandable. Their drawback is however that they usually require a tedious preparation work done by human expert graders. Besides, such a system is usually task/subject specific, i.e. minor modifications in the aspects can lead to a demand on a completely redefinition of the rule base (rigidity of the system). Another problem arises from the fact that rule based systems can only operate with a low number of fuzzy sets owing to the exponentially growing number of necessary rules in multidimensional cases if a full coverage of the input space should be ensured.

The advantage of the second approach is its simplicity and easy adaptability. Besides, it can operate with a higher resolution of the input space. However, as its disadvantage it should be mentioned the lack of the humanly easy-to-interpret rules.

2 Criteria for comparison of fuzzy evaluation methods

In this section, we introduce a set of criteria for fuzzy methods aiming the evaluation of the students' academic performance. We consider these requirements as properties that help the reader to compare the overviewed methods. The criteria are the followings.

1. The method should not increase the time needed for the assessment compared to the traditional evaluation techniques.
2. The method should help the graders to express the vagueness in their opinion.
3. The method should be transparent and easy to understand for both parties involved in the assessment process, i.e. the students and the graders.
4. The method should ensure a fair grading, i.e. it should be beneficial for all students.
5. The method should allow the teacher to express the final result in form of a total score or percentage as well as in form of grades using a mapping between them.
6. The method should be easy implementable in software development terms.
7. The method should be compatible with the traditional scoring system, i.e. when the grader provides crisp scores for each response the total score and the final grade should be identical with the one calculated by the traditional way.

3 Fuzzy Inference Based Student Evaluation Methods

3.1 Evaluation Based On Fuzzy Classification

Nolan published in [8] the development and successful application of a fuzzy rule based evaluation method aiming the rating of writing samples of fourth grade students. Previously in course of the evaluation the teachers used a predefined comprehensive scoring guide that defined which skills have to be measured by the evaluator and which ones have to be determined from them.

The rule base was created from this scoring guide involving the participation of a group of expert evaluators. In order to reduce the complexity of the rule base they defined input partitions with a quite low resolution each containing only three fuzzy sets.

In course of the evaluation the rater measures skills like character recognition, text understanding, understanding elements of the plots, and understanding ideas. The system infers the evaluation of skills like reading comprehension. For example a rule of the system is

IF understanding=*high* AND character-recognition=*strong* AND elements-of-plot=*all* AND generates-ideas=*expand* THEN reading-comprehension=*high*.

The main advantage of the method compared to the traditional evaluation form was that it reduced the time necessary for the learning of the scoring technique and the difference between the scores given by different evaluators decreased significantly. The drawback of the method is that it does not support the fuzzy input; the evaluators can express their opinion only in form of crisp values, which will be fuzzyfied later by the method. Based on the description given in the literature we can summarize that the method fulfils the criteria 1, 3, 4, and 6. Furthermore it surely does not fulfil criteria 2 and 5.

3.2 Bai-and-Chen's method

In order to reduce the subjectivism in student evaluation Bai and Chen (further on we will refer to it as BC method) suggested a quite complex solution in [1]. However, their method addresses only a part-task of the evaluation, namely the ranking of the students that obtained the same total score.

The BC method is applied as a follow-up of a conventional scoring technique. First, in case of each student ($S_j, 1 \le j \le n$) each question ($Q_i, 1 \le i \le m$) is evaluated independently by an accuracy rate a_{ij}, where $a_{ij} \in [0,1]$. Then, the evaluator calculates a total score for the student by

$$TS_j = \sum_{i=1}^{m} a_{ij} \cdot g_i \,, \tag{1}$$

where $\mathrm{g_i}$ is the maximum achievable score assigned to the question Q_i ($\sum_{i=1}^{m} g_i = 100$).

In order to rank the students having the same total score Bai and Chen propose an adjustment of their scores. The adjustment is based on introduction of new aspects in the evaluation, i.e. the importance and the complexity of the questions, which are based on fuzzy sets determined by the evaluator or by domain experts. The measurement part of the evaluation is also extended by including the time necessary for answering the individual questions divided by the maximum time allowed to solve the question (answer-time rate, $t_{ij} \in [0,1]$).

Although it is used only in cases when two or more students achieve the same total score, the answer-time rate has to be measured for each student during the exam because it can not be obtained posteriorly.

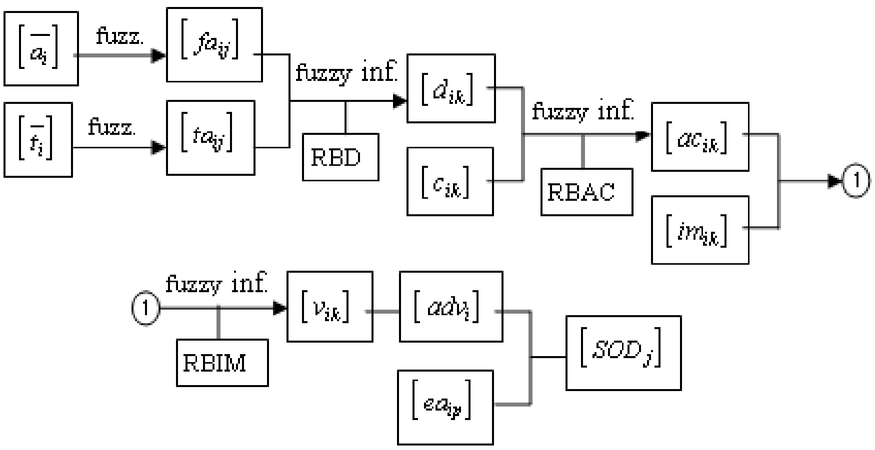

Figure 1

Block diagram of the BC method

The modified scores are determined in six steps applying a three-level fuzzy reasoning process whose block diagram is presented in figure 1. After calculating the average of the accuracy rates ($\overline{a_i}$) and the average of the answer-time rates

($\overline{t_i}$) for each question these are fuzzyfied by calculating their membership values in the corresponding predefined partitions resulting in the fuzzy grade matrices $\left[fa_{ik} \right]$ and $\left[ft_{ik} \right]$.

In the second step of the method one determines the fuzzy difficulty ($\left[d_{ik} \right]$) of each question using a special kind of fuzzy reasoning applying a predefined rule base (*RBD*) and a weighted average of the previously calculated membership values. The third step of the method concentrates on the calculation of the answer-cost of each question (a_{ik}) from the difficulty and the complexity values. The complexity of each question (c_{ik}) is expressed as membership values in the five sets of the predefined complexity partition. The $\left[c_{ik} \right]$matrix is defined by domain experts. This step uses the same fuzzy inference model as the previous one applying a predefined rule base (*RBAC*).

The fourth step of the method calculates the adjustment values (v_{ik}) of each question from the answer-cost and the importance values. The importance of each question (im_{ik}) is expressed as five membership values in the five sets of the predefined importance partition. The $\left[im_{ik} \right]$ matrix is defined by domain experts. This step uses the same fuzzy inference model as the previous one applying a predefined rule base (*RBIM*). Next, one calculates the final adjustment value (adv_i) for each question as a weighted average of the individual adjustment values (v_{ik}) corresponding to the question.

In step 5 a new grade matrix ($\left[ea_{ip} \right]$) is constructed that contains only that k columns of the original accuracy rate matrix, which correspond to the students having the same total score.

The modified score values of each student ($SOD_j, 1 \le j \le n$) are calculated in the last step by

$$SOD_j = \sum_{p=1}^{k} \left[\left[\sum_{\substack{i=1 \\ i \neq j}}^{m} \left(ea_{pj} - ea_{pi} \right) \right] \cdot g_p \cdot \left(0.5 + adv_p \right) \right]. \qquad (2)$$

The main advantages of the method are that it does not increase the time needed for the evaluation and it allows the evaluators to make a ranking among students achieving the same score in the traditional scoring system. However, one has to pay a too high price for this result. In course of the exam preparation two matrices have to be defined by domain experts, one describing the complexity $\left[c_{ik} \right]$ and one describing the importance $\left[im_{ik} \right]$ of each question. It introduces redundancy in the evaluation process because these aspects presumably already have been taken into consideration in course of the definition of the vector $\left[\,g_i\right]$.

Thus it is hardly avoidable the occurrence of cases when the achievable score of a question is not in accordance with its complexity and importance evaluation. Besides, the level of subjectivity is also increased by the fact that seven weights have to be determined by domain experts as well and there is no formalized way to determine their optimal values. Another drawback of the method is that it does not allow the evaluator to give a fuzzy set as evaluation.

The real novel aspect of the evaluation is the answer-time rate. However, it is not clear how the base time for each question is defined. Besides, it seems not too efficient to measure the answer time for each student for each question and then to use it in case of students having the same total score unless it can be done by software automatically. Thus the BC method is not applicable in case of non computer-based exams. We can summarize that it fulfils criteria 1, 4, 5, and 6.

3.3 SEFRI

The method SEFRI (Student Evaluation based on Fuzzy Rule Interpolation) [5] offers a solution using a rule base containing only the most relevant rules. In course of the rating the evaluator takes into consideration three aspects, namely the accuracy of the response, the time necessary for answering the questions, and the correct use of the technical terms. In course of the preparation the 100 achievable marks are divided between the questions. They are the weights associated to the questions.

In case of the second aspect one works with the total time necessary for answering all of the questions, which is determined automatically and reported to the allowed total response time. The resulting relative time is fuzzyfied (*TR*) using singleton type fuzzyfication.

The characteristics “the accuracy of the response” (*AC*), and “the correct use of the technical terms” (*CU*) are measured by the evaluator with separate fuzzy marks (fuzzy numbers) for each question. The scoring scale is in both cases the unit interval. After assigning the two fuzzy marks for each question one calculates an average *AC* and *CU* value ($\overline{AC}$ and $\overline{CU}$) for the student as a weighted average of the individual values.

Next one determines from the three fuzzy values ($\overline{AC}$, *TR*, and $\overline{CU}$) the general evaluation of the student using fuzzy inference. In order to reduce the complexity of the rule base a fuzzy rule interpolation based reasoning method called LESFRI [4] is used. Thus the underlying rule base requires only 64 rules in contrast with the 125 rules of the dense rule base owing to the fact that each input dimension contains five fuzzy sets.

The fuzzy inference results the general fuzzy evaluation of the student (*GFE*) that is defuzzyfied using Center Of Area method in order to get the total score (*TS*). Finally the grade of the student is determined using the standardized mapping of the university. For example a possible mapping is presented in Table 1.

Similar to the previous techniques this method can only applied in practice when a software support is present. Its advantage is that it contains only one-level inference with a relatively transparent rule base. The drawback of the method is that owing to the sparse character of the rule base it applies a bit complex inference technique that could require more software development work. We can summarize that the method satisfies 1, 2, 3, 4, 5, and 6.

Table 2

Mapping between scores and grades [5]

Score intervals	**Grades**
0 - 50	Unsatisfactory
51 - 60	Satisfactory
61 - 75	Average
76 - 85	Good
86 - 100	Excellent

3 Conclusions

Fuzzy student evaluation methods can be a very useful tool supporting the evaluator in handling the uncertainty that is often present in the opinion of the rater in cases when the evaluation process is not fully defined, i.e. when it cannot be fully automated. Fuzzy inference based solutions offer a transparency owing to the humanly interpretable character of the rule base.

However, their disadvantage is their rigidity and the implicit weighting. A small change in the aspects or in the weighting could require a completelly redefinition of the underlying rule base. Besides, owing to the implicit weighting the importance of the different aspects is not clear visible.

We can summarize that none of the overviewed methods fulfils all the previosly defined criteria. The lack of the compatibility with the traditional methods proved to be a common drawback of them, which probably could be solved using automatic fuzzy rule base identification methods [2][9][10]. The application of other fuzzy inference techniques like the methods presented in e.g. [6] and [3]. could also contribute to the development of evaluation techniques that better fit the applied criteria. Despite of the fuzzy character of the methods only SEFRI allows the fuzzy expression of the evaluator's opinion. As a positive evaluation one can state that all the three methods satisfy criteria 1, 4, and 6.

Acknowledgement

This research was supported by Kecskemét College GAMF Faculty grant no: 1KU31, and the National Scientific Research Fund Grant OTKA K77809.

References

[1] Bai, S.M., Chen, S. M.: Evaluating students' learning achievement using fuzzy membership functions and fuzzy rules, Expert Systems with Applications, 34 (2008), pp. 399-410.

[2] Botzheim, J., Hámori, B., & Kóczy, L.T.: Extracting trapezoidal membership functions of a fuzzy rule system by bacterial algorithm, 7th Fuzzy Days, Dortmund 2001, Springer-Verlag, pp. 218-227. [2]

[3] Hládek, D., Vaščák, J., & Sinčák, P.: Hierarchical fuzzy inference system for robotic pursuit evasion task, in Proc. of the 6th International Symposium on Applied Machine Intelligence and Informatics (SAMI 2008), January 21-22, Herľany, Slovakia, 2008, pp. 273-277.

[4] Johanyák, Z. C., Kovács, S.: Fuzzy Rule Interpolation by the Least Squares Method, 7th International Symposium of Hungarian Researchers on Computational Intelligence (HUCI 2006), November 24-25, 2006 Budapest, pp. 495-506.

[5] Johanyák, Z. C.: Student Evaluation Based on Fuzzy Rule Interpolation, International Journal of Artificial Intelligence, 2009, (to be published)

[6] Kovács, S.: Extending the Fuzzy Rule Interpolation "FIVE" by Fuzzy Observation, Advances in Soft Computing, Computational Intelligence, Theory and Applications, Bernd Reusch (Ed.), Springer Germany, 2006, pp. 485-497.

[7] Mamdani, E. H., Assilian, S. (1975): An experiment in linguistic synthesis with a fuzzy logic controller, International Journal of Man Machine Studies, Vol. 7, 1975, pp. 1-13.

[8] Nolan, J. R.: An expert fuzzy classification system for supporting the grading of student writing samples, Expert Systems With Applications, 15 (1998), pp. 59-68.

[9] Precup, R.E., Preitl S., Tar, J. K. , Tomescu, M. L., Takács, M., Korondi, P. Baranyi, P.: Fuzzy control system performance enhancement by Iterative Learning Control. IEEE Transactions on Industrial Electronics, vol. 55, no. 9, 2008, pp. 3461-3475.

[10] Škrjanc, I., Blažič, S., and Agamennoni, O.E.: Interval fuzzy model identification using l_∞-norm, IEEE Transactions on Fuzzy Systems, vol. 13, no. 5, 2005, pp. 561-568.