

Survey on Five 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 In case of non-automated examinations the evaluation of students' academic achievements involves in several cases the consideration of impressions and other subjective elements that can lead to differences between the scores given by different evaluators. The inherent vagueness makes this area a natural application field for fuzzy set theory based methods aiming the reduction of the mentioned differences. After introducing a criterion set for the comparison the paper surveys five relevant fuzzy student evaluation methods that apply fuzzy inference for the determination of the students' final score.

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

1 Introduction

The evaluation of students' answerscripts containing narrative responses or assignments that cannot be rated fully automatically is from nature vague, which can lead to quite different scores given by different evaluators. This problem usually is solved by defining scoring guides that become more and more complex after the developers face new and new cases that seemed to be unimaginable previously. The more specific the guides are the more tedious they become, which leads to inconsistency in their application and increases the time need of the scoring. Owing to the increased complexity and hard-to-learn character of the comprehensive scoring guides evaluators often use ad hoc inference methods that lack a formal mechanism. Beside the demand on consistency of the evaluation the easy-to-explain/confirm character is also important not only for the teachers but also for other interested parties, like students, parents, etc.

A completely new approach appeared in the late 90s in field of evaluation methods by emerging the fuzzy set theory based evaluation techniques, which make possible a good trade-off between the demand on quick evaluation and high consistence of the results. Biswas [2] proposed a particular (FEM) and a generalized

(GFEM) method that were based on the vector representation of fuzzy membership functions and a special aggregation of the grades assigned to each question of the student's answerscripts. Chen and Lee [4] suggested a simple (CL) and a generalized (CLG) method that produced improvements by applying a finer resolution of the scoring interval and by including the possibility of weighting the four evaluation criteria. Wang and Chen [18] extended the CL/CLG method pair by introducing the evaluator's optimism as a new aspect, and by using type-2 fuzzy numbers for the definition of the satisfaction. Johanyák suggested a fuzzy arithmetic based simple solution (FUSBE) in [7] for the aggregation of the fuzzy scores. Nolan [13] introduced a fuzzy classification model for supporting the grading of student writing samples in order to speed up and made more consistent the evaluation. Bai and Chen [1] developed a method for the ranking of students that obtained the same total score during the traditional evaluation. They used a three-level fuzzy reasoning process. Saleh and Kim [16] enhanced the BC method by excluding some subjective elements and applying Mamdani [12] type inference. Rasmani and Shen [15] introduced a data driven fuzzy rule identification method. Johanyák suggested a low complexity fuzzy rule interpolation based method (SEFRI) in [9].

The fuzzy student evaluation techniques can be classified in two main groups depending on their algorithm: (1) methods applying fuzzy inference (e.g. [1][9][13][15][16]), and (2) methods applying "only" fuzzy arithmetic (e.g. [2][4][7][18]). The advantage of the first approach is that the rules are close to the traditional human thinking, they 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. This feature makes the system rigid. Another problem arises from the fact that in general the 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. Furthermore the methods based on it can operate with a higher resolution of the input space. However, as its disadvantage one should mention the lack of the humanly easy-to-interpret rules. The rest of this paper is organized as follows. Section 2 introduces the criterion set considered as relevant for fuzzy student evaluation methods. Section 3 gives a short survey on fuzzy inference based evaluation methods. The conclusions are drawn in section 4.

2 Criteria for comparison of fuzzy evaluation methods

In this section, we introduce a criterion set [10] 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 [13] 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 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. 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 fulfill 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 ($S_j, 1 \leq j \leq n$) each question ($Q_i, 1 \leq i \leq m$) 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

$$TS_j = \sum_{i=1}^m a_{ij} \cdot g_i, \quad (3.1)$$

where 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 cannot be obtained posterior.

The modified scores are determined in six steps applying a three-level fuzzy reasoning process whose block diagram is presented in figure 3.1. After calculating the average of the accuracy rates (\bar{a}_i) and the average of the answer-time rates (\bar{t}_i) for each question these are fuzzified by calculating their membership values in the corresponding predefined partitions resulting in the fuzzy grade matrices $[fa_{ij}]$ and $[ft_{ik}]$.

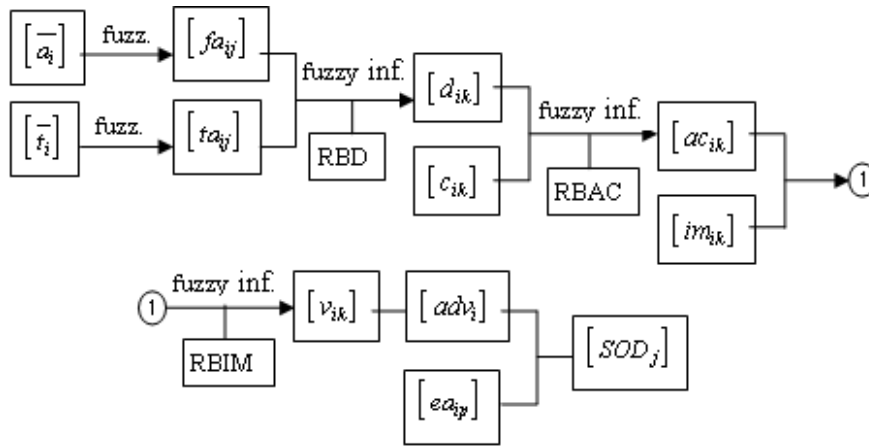


Fig. 3.1. Block diagram of the BC method

In the second step of the method one determines the fuzzy difficulty ($[d_{ik}]$) 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 $[c_{ik}]$ 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 $[im_{ik}]$ 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 ($[ea_{ip}]$) is constructed that contains only those 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 , $1 \leq j \leq n$) are calculated in the last step by

$$SOD_j = \sum_{p=1}^k \left[\sum_{\substack{i=1 \\ i \neq j}}^m (ea_{pj} - ea_{pi}) \right] \cdot g_p \cdot (0.5 + adv_p). \quad (3.2)$$

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 $[c_{ik}]$ and one describing the importance $[im_{ik}]$ 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 $[g_i]$.

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 the seven weights have to be determined by domain experts 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 express the evaluation using fuzzy sets.

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 Saleh-and-Kim's method

In order to alleviate some shortcoming of the BC method Saleh and Kim [16] suggested the so called *Three node fuzzy evaluation system* (TNFES) that applies Mamdani type fuzzy inference and COG defuzzification. Similar to the BC method TNFES works with five inputs, namely the original grade vector ($[g_i]$), the accuracy grade matrix ($[a_i]$), the time rate matrix ($[t_i]$), the complexity matrix ($[c_{ik}]$), the importance matrix ($[im_{ik}]$), as well as with three rule bases, one for the difficulty (*RBD*), one for the effort (*RBE*), and one for the adjustment (*RBA*). The accuracy rate and answer time rate matrices are results of the examination. The complexity and importance matrices as well as the rule bases are defined by do-

main experts. The output of the system is a new grade vector, which contains the adjusted score values.

TNFES defines three fuzzy nodes (difficulty, effort and adjustment) that attain a three level fuzzy inference schema as follows:

- Difficulty node: $D = I([a_i], [t_i], RBD)$,
 - Effort node: $E = I(D, [c_{ik}], RBE)$,
 - Adjustment node: $W = I(E, [im_{ik}], RBA)$,
- where I represents the Mamdani type fuzzy inference.

Each of the nodes contains a fuzzy logic controller with two scalable inputs and one output. The scalable inputs make possible the weighting of the different aspects, however, the authors do not use this possibility, they consider each input of equal influence. Each node consists of three steps (fuzzyfication, inference, defuzzyfication), which modularity can be also considered as a drawback owing to the redundancy introduced by the consecutive defuzzyfications and fuzzyfications. The result of the third node ($W=[w_i]$) is used for the calculation of the adjusted grade vector $[ga_i]$ by

$$ga_i = g_i \cdot (1 + w_i), \quad (3.3)$$

followed by a scaling operation

$$ga_i = g_i \cdot \frac{ga_i \cdot \sum_{j=1}^m g_j}{\sum_{j=1}^m ga_j}, \quad (3.4)$$

where m is the number of the questions. The final total score is determined by

$$TS = [a_i]^T \cdot [ga_i]. \quad (3.5)$$

Owing to the similarity between TNFES and the BC approaches, the advantages and the drawbacks of the method are also similar to the features of BC. We can summarize that it fulfils criteria 1, 4, 5, and 6.

3.4 Student Evaluation based on Fuzzy Rule Interpolation

The method *Student evaluation based on fuzzy rule interpolation* (SEFRI) [9] offers a solution using a rule base containing only the most relevant rules. The me-

thod 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 3.1.

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 3.1. Mapping between scores and grades [9]

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

3.5 Rasmani-and-Shen's method

Rasmani and Shen proposed in [15] a special fuzzy inference technique and the use of a data driven fuzzy rule identification method that also allowed the addition of expert knowledge. Their main aim was to obtain user comprehensible knowledge from historical data making also possible the justification of any evaluation. The suggested inference technique is the so called weighted fuzzy subsethood based reasoning, which was developed for multiple input single output (MISO) fuzzy systems that apply rules of form

$$\begin{aligned}
 & \text{IF } A_1 \text{ is } [w(E_i, A_{11}) \cdot A_{11} \text{ OR } w(E_i, A_{12}) \cdot A_{12} \text{ OR } \dots \text{ OR } w(E_i, A_{1j}) \cdot A_{1j} \text{ OR } \dots \\
 & \text{OR } w(E_i, A_{1n_1}) \cdot A_{1n_1}] \text{ AND} \\
 & A_2 \text{ is } [w(E_i, A_{21}) \cdot A_{21} \text{ OR } w(E_i, A_{22}) \cdot A_{22} \text{ OR } \dots \text{ OR } w(E_i, A_{2j}) \cdot A_{2j} \text{ OR } \dots \\
 & \text{OR } w(E_i, A_{2n_2}) \cdot A_{2n_2}] \text{ AND } \dots \text{ AND} \\
 & A_k \text{ is } [w(E_i, A_{k1}) \cdot A_{k1} \text{ OR } w(E_i, A_{k2}) \cdot A_{k2} \text{ OR } \dots \text{ OR } w(E_i, A_{kj}) \cdot A_{kj} \text{ OR } \dots \\
 & \text{OR } w(E_i, A_{kn_k}) \cdot A_{kn_k}] \text{ AND } \dots \text{ AND} \\
 & A_m \text{ is } [w(E_i, A_{m1}) \cdot A_{m1} \text{ OR } w(E_i, A_{m2}) \cdot A_{m2} \text{ OR } \dots \text{ OR } w(E_i, A_{mj}) \cdot A_{mj} \text{ OR } \dots \\
 & \text{OR } w(E_i, A_{mm_m}) \cdot A_{mm_m}] \text{ THEN } B \text{ is } E_i
 \end{aligned} \tag{3.6}$$

where m is the number of antecedent dimensions, A_k $k \in [1, m]$ are the antecedent linguistic variables, n_k is the number of linguistic terms in the k th antecedent dimension, B is the consequent linguistic variable, E_i $i \in [1, N]$ is the i th consequent linguistic term, N is the number of consequent linguistic terms, and $w(E_i, A_{kj})$ is the relative weight of the antecedent linguistic term A_{kj} . The weight expresses the influence of the set A_{kj} towards the conclusion drawn. One determines the weight as a result of the normalization of the fuzzy subsethood value of the set

$$w(E_i, A_{kj}) = \frac{S(E_i, A_{kj})}{\max_{l=1..n_k} S(E_i, A_{kl})}. \tag{3.7}$$

The fuzzy subsethood value S represents in this case the degree to which the fuzzy set A_{kj} is the subset of a the fuzzy set E_i . It is calculated as

$$S(E_i, A_{kj}) = \frac{\sum_{x \in U} \nabla(\mu_{E_i}(x), \mu_{A_{kj}}(x))}{\sum_{x \in U} \mu_{E_i}(x)}, \tag{3.8}$$

where U is the universe of discourse, μ is the membership function, and ∇ is an arbitrary t-norm.

The rule base contains only one rule for each consequent linguistic term. The first step of the fuzzy inference is the calculation of the overall weight of each rule by applying the arbitrary disjunction and conjunction operators [5] to the antecedent side. Next, one selects the rule having the highest weight, whose consequent will represent the final score of the student.

One identifies the rule base in the following steps:

1. Create the input and output partitions.
2. Divide the training dataset into subgroups depending on the output linguistic terms.
3. Calculate fuzzy subsethood values for each subgroup.
4. Calculate weights for each linguistic term.
5. Create rules of form (3.6).
6. Test the rule base using a test dataset.

The main advantage of the method proposed by Rasmani and Shen is that it requires a rule base with a low number of rules, which number is equal with the number of output linguistic terms. Besides, it allows the evaluation of a question/test to be made by fuzzy numbers. However, it is not clear how the antecedent and consequent are determined and what is the meaning of the fuzzy subsethood values in case of the evaluation of the students' academic performance. We can summarize that the method satisfies 1, 2, 4, 5, and 6.

4 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 completely 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 previously 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 [3][14][17]. The application of other fuzzy inference techniques like the methods presented in [6] and [11] could also contribute to the development of evaluation techniques that better fit the applied criteria. Despite of the fuzzy character of the methods only the last two me-

thods (SEFRI and the method proposed by Rasmani and Shen) allow the fuzzy expression of the evaluator's opinion. As a positive evaluation one can state that all the methods satisfy criteria 1, 4, and 6.

Acknowledgments 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, pp. 399-410 (2008)
- [2] Biswas, R.: An application of fuzzy sets in students' evaluation. *Fuzzy Sets and System*, 74(2), pp. 187–194 (1995)
- [3] Botzheim, J., Hámori, B., & Kóczy, L.T.: Extracting trapezoidal membership functions of a fuzzy rule system by bacterial algorithm, In: *Proceedings of the 7th Fuzzy Days, Dortmund 2001*, Springer-Verlag, pp. 218-227 (2001)
- [4] Chen, S. M., and Lee, C. H.: New methods for students' evaluating using fuzzy sets, *Fuzzy Sets and Systems*, 104(2), pp. 209–218 (1999)
- [5] Fodor, J.: *Aggregation Functions in Fuzzy Systems, Aspects of Soft Computing, Intelligent Robotics and Control, Studies in Computational Intelligence*, Ed: Fodor, J., Kacprzyk, J., Springer-Verlag, pp. 25-50 (2009)
- [6] Hládek, D., Vaščák, J., & Sinčák, P.: Hierarchical fuzzy inference system for robotic pursuit evasion task, In *Proceedings of the 6th International Symposium on Applied Machine Intelligence and Informatics (SAMII 2008)*, January 21-22, Herľany, Slovakia, pp. 273-277 (2008)
- [7] Johanyák, Z.C.: Fuzzy set theory based student evaluation, In *Proceedings of the IJCCI 2009 - International Joint Conference on Computational Intelligence, ICFC 2009 - International Conference on Fuzzy Computation*, 5-7 October, Funchal-Madeira, Portugal, pp. 53-58 (2008)
- [8] Johanyák, Z. C., Kovács, S.: Fuzzy Rule Interpolation by the Least Squares Method, In: *Proceedings of the 7th International Symposium of Hungarian Researchers on Computational Intelligence (HUCI 2006)*, November 24-25, 2006 Budapest, pp. 495-506.
- [9] Johanyák, Z. C.: Student Evaluation Based on Fuzzy Rule Interpolation, *International Journal of Artificial Intelligence*, (to be published) (2009)
- [10] Johanyák, Z.C.: Survey on Three Fuzzy Inference Based Student Evaluation Methods, In: *Proceedings of the 10th International Symposium of Hungarian Researchers – CINTI 2009*, November 12-14, Budapest, pp. 185-192 (2009)
- [11] 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, pp. 485-497 (2006)
- [12] Mamdani, E. H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Man Machine Studies*, Vol. 7, 1975, pp. 1-13 (1975)
- [13] Nolan, J. R.: An expert fuzzy classification system for supporting the grading of student writing samples, *Expert Systems With Applications*, 15, pp. 59-68 (1998)
- [14] 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, pp. 3461-3475 (2008)
- [15] Rasmani, K. A., Shen, Q.: Data-driven fuzzy rule generation and its application for student academic performance evaluation, *Appl. Intell.*, 25, pp. 305-319 (2006)

- [16] Saleh, I., Kim, S.: A fuzzy system for evaluating students' learning achievement, *Expert Systems with Applications*, 36, pp. 6236-6243 (2009)
- [17] Škrjanc, I., Blažič, S., and Agamennoni, O.E.: Interval fuzzy model identification using 1-norm, *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 5, pp. 561-568 (2005)
- [18] Wang, H.Y., and Chen, S.M.: New methods for Evaluating the Answerscripts of Students Using Fuzzy Sets, *Advances in Applied Artificial Intelligence, Lecture Notes in Computer Science*, Vol. 4031, pp. 442-451 (2006)