Rules Optimization Based Fuzzy Model for Predicting Distance Education Students' Grades

Osman Yildiz, Abdullah Bal, Sevinc Gulsecen, and Fulya D. Kentli

Abstract—**Distance education has been rapidly becoming widespread all around the world. This rapid growth brings along both success and failure. One of the causes of failure in distance education is the lack of student observation, which is a part of traditional education. We can overcome this deficit through an analysis of logs kept in the learning management system. These logs make it possible to estimate academic performance of a student attending to distance education. The instructor having such information can take precautions in order to prevent failure. In this study, a mathematical model has been formed by using Fuzzy Logic. In order to enhance the success of model, fuzzy logic rules have been optimized by using genetic algorithm. Three different variableshave been introduced to the proposed model. These are recency, stating the number of days that pass from the date of registering course contents until the date of student's admission to the system; frequency, stating the frequency of logging on to the system and monetary, stating the period of time spent online on the system. End of term scores of students have been estimated on the basis of the first 6-week data and then the results have been verified by the real scores.**

Index Terms—**Academic performance, fuzzy logic, genetic algorithm, rule optimization.**

I. INTRODUCTION

Recently, distance education system has become a popular and important concept in most countries due to their advantages. Students are enabled to log onto the learning system at any time from their homes, offices or anywhere and to continue their education. Despite all these advantages, the system brings along a disadvantage; lack of monitoring students' academic performance. Traditional education enables the instructor to directly monitor the student in class environment and thus to take necessary precautions as a result of such observations. However, unlike traditional education, the instructors cannot observe students during distanceeducation. Lack of observation, considered to a disadvantage for distanceeducation, can be transformed into an advantage by analyzing the logs kept in the learning management system (LMS) because LMS keeps logs of all student-related activities.

This study used three types of data; recency, frequency and monetary. Recency expresses the number of days that pass

from the date of entering the content until the date of logging onto the system; frequency expresses the student's frequency of logging onto the system and monetary expresses the total time spent on the system by the student. The fuzzy logic model has three inputs and one output was generatedon the basis of the data on the inputs. Besides, 27 rules specific to this model were formed. Then, these rules were optimized by using genetic algorithm in order to enhance the success of this model.

The rest of the study was created as follows. The second part presents the study-related findings in the academic literature. Then, it briefly explains Fuzzy Logic and Genetic Algorithm. The fourth part presents the classic fuzzy model and genetic fuzzy model used to estimate the academic performance. Finally, the fifth part lists the conclusions and implications for further studies.

II. RELATED WORKS

Several studies focus on the field of analyzing student performance in distance education. [1] have conducted a supervised machine learning algorithm in which the training set was comprised of students' key demographic characteristics and their mark on a few written assignments. [2] have employed genetic algorithms and decision trees to estimate academic performance of distance education students. [3] have used neural network, decision tree and linear regression to estimate students' academic performance. In this work, they have employed demographic profile and students' first semester cumulative grade point averages (CGPA) to predict their final CGPA. [4] have studied a model by means of neural networks procedure in SAS/Enterprise mining. They have categorized students into three groups as "low-risk", "medium-risk" and "high-risk" depending on their probability of failure. [5] have performed a student achievement prediction method applied to 10-week introductory level e-learning. [6] have incorporated multiple instance genetic algorithms to predict whether students will fail or pass for a certain course. This prediction has been based on students' activities such as quizzes, assignments, and forums.

In the literature, prediction studies involving Fuzzy Logic method are generally carried out in association with Artificial Neural Network (ANN). [7] have built an evaluation of student's performance and learning efficiency based on ANFIS, too.ANFIS stands for Adaptive Neuro–Fuzzy Inference System, which is used to enhance certain characteristics such as flexibility, speed, and adaptability. In that study, neural network and fuzzy have been used for predicting student's performance based on four criteria, namely scores earned, time spent, number of attempts and

help needed. [8] have presented a study that introduces a systematic approach to design of a fuzzy inference system based on a class of neural networks to assess the students' academic performance. This study is a continuation of the previous study [9].

III. Fuzzy Logic and Genetic Algorithm

A. Fuzzy Logic

Fuzzy Logic was developed by Dr. LotfiZadeh in 1960s. Fuzzy logic includes the degrees of belonging to a set. The syllogistic logic defines belonging to a set with 1 and not belonging to a set with 0. In syllogistic logic, people are either tall or short. In Fuzzy logic, there are degrees of being tall. In that case, a person might be neither tall nor short but can be 0.75 degree tall and 0.25 degree short.

Fuzzy sets and fuzzy logic are known as efficient tools used to overcome uncertainties related to vagueness, ignorance and imprecision. Fuzzy Logic involves three main stages; fuzzification, rule evaluation and defuzzification. Fuzzification is the initial step that converts new inputs into degrees of match with linguistic values. Rule evaluation is where knowledge provided by experts are formed, which is then called fuzzy rules. Defuzzification transposes the fuzzy outputs to new values [7].

B. Genetic Algorithm

The Genetic Algorithm (GA), based on principles of genetics and natural selection, is an optimization and search method [10].Genetic algorithm starts working randomly or an existing, initial population. Each element of the population is called chromosome. Population is tested with fitness function and the chromosomes with the best results will continue to the process. Fitness function has a decision making role for decision whether a finding is good or bad. Generally, values of the chromosomes will be processed by being converted into binary system. After that stage, two chromosomes having best results are randomly intercrossed to create new chromosomes. After crossover, the chromozomes are subjected to mutation. Mutation prevents the algorithm to be trapped in a local minimum. A simple mutation can consist in inverting value of each gene with a small probability. Mutation of a bit involves flipping a bit, changing 0 to 0 and vice-versa [11]. Next, the chromosomes are retested and the chromosomes having the best results are determined again. This cycle continues until a determined iteration or a desired outcome is achieved.

IV. Methods and Results

In this study, the informatics course of the Yildiz Technical University (YTU), Matlab course is delivered in web-based education during 2010/2011 spring semester for Chemistry students.

Moodle (Modular object oriented developmental learning environment), a free learning LMS is utilized for this course. Moodle offers many activities that can be combined with learning materials. It logs every activity report for each student that indicates the activities such as operation on course materials, quizzes, and messages for different times and days. This fact enables the authors to collect data only from the tutors involved in this learning process.

The Moodle records for Matlab course have supplied the data for the Fuzzy-RFM model. Matlab course is composed of 12 weeks and leads to an elective course. The total number of registered students is 62, 46 of whom are selected and the model is formed with RFM values being entered for the first six weeks. As the desired output, grade scores of 46 students are selected and introduced to model.

We have first used Classical Fuzzy Model with expert opinion to evaluate students' academic performance. Then, we have improved the model by employing Genetic Optimization Algorithm to obtain higher prediction accuracy.

In this model, Fuzzy Memberships are organized for three input data sets; Recency, Frequency and Monetary. Academic Performance (AP) is used as output value of the model. Input values are determined as low (R1), medium (R2) and high (R3) for Recency, low (F1), medium (F2) and high (F3) for Frequency, low (M1), medium (M2) and high (M3) for Monetary. The output values are determined to be low (AP1), medium (AP2) and high (AP3). Fig. 1 shows Classic Fuzzy Model.
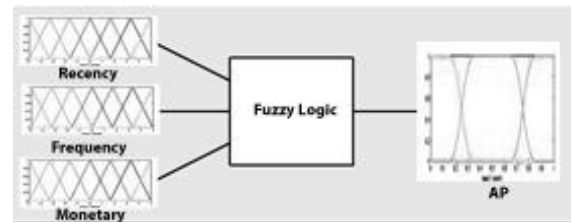

Fig. 1. Fuzzy model.

Input values R, F and M are shown in Fig. 2- Fig. 4 respectively.
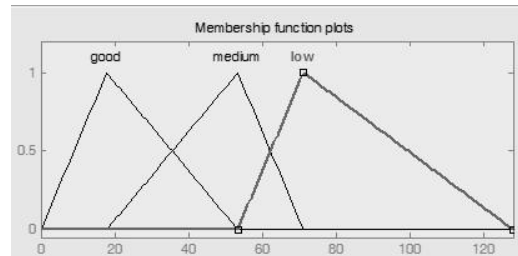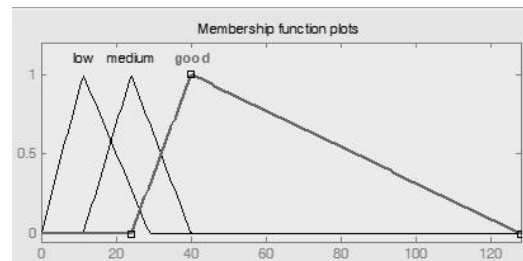

Fig. 2. Recency membership function plots.
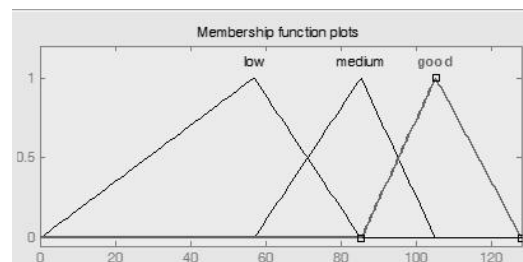

Fig. 3. Frequency membership function plots.

Fig. 4. Monetary membership function plots.

Fuzzy model has 2 significant steps following the determination of input and output variables. These are determining membership function intervals and establishing fuzzy rules. These steps can be created depending on data or by one consulting experts. This study created models individually by using the data available and by consulting experts. The success rate of the model created with data available was 65 % whereas the success rate of the model created in cooperation with experts was 75 %. Rules of the model created in the second way are as follows;

1. If (Recency is good) and (Frequency is low) and (Monetary is low) then (AP is low)
2. If (Recency is good) and (Frequency is low) and (Monetary is medium) then (AP is medium)
3. If (Recency is good) and (Frequency is low) and (Monetary is good) then (AP is medium)
….
24. If (Recency is low) and (Frequency is medium) and (Monetary is good) then (AP is medium)
25. If (Recency is low) and (Frequency is good) and (Monetary is low) then (AP is medium)
26. If (Recency is low) and (Frequency is good) and (Monetary is medium) then (AP is medium)
27. If (Recency is low) and (Frequency is good) and (Monetary is good) then (AP is good)

The ratio of accuracy between the alteration rules according to the expert and prediction of students' academic performance has risen to 74%.

Rules Optimization Based Fuzzy Model

For Fuzzy logic model (see Fig. 5), the rules were created in cooperation with experts and this is pointed out as a defect of fuzzy logic. The rules were optimized by using genetic algorithm in order to overcome this deficit. Matlab environment was used for optimization. Initially, the genetic algorithm started processing with a population randomly created with 15 chromosomes. Each chromosome has 27 rules. These rules are illustrated as vectors of 1 × 4. For example:

Rule1

If (Recency is low) and (Frequency is medium) and (Monetary is good) then (AP is medium)

Rule1=[3 2 3 2]

One set of rule population matrix size is 27x4 and GA has generated 15 sets of initial population.

The algorithm stopped after 100.000 iterations because the fitness function values remained the same. The set of rules with the best results is as follows.

1. If (Recency is good) and (Frequency is low) and (Monetary is low) then (AP is medium)
2. If (Recency is low) and (Frequency is medium) and (Monetary is low) then (AP is low)
3. If (Recency is good) and (Frequency is medium) and (Monetary is good) then (AP is low)
4. If (Recency is low) and (Frequency is low) and (Monetary is medium) then (AP is medium)
5. If (Recency is low) and (Frequency is low) and (Monetary is medium) then (AP is good)
6. If (Recency is good) and (Frequency is low) and (Monetary is medium) then (AP is good)
7. If (Recency is low) and (Frequency is low) and (Monetary is medium) then (AP is low)
8. If (Recency is good) and (Frequency is medium) and (Monetary is good) then (AP is good)
9. If (Recency is good) and (Frequency is low) and (Monetary is low) then (AP is low)

10. If (Recency is low) and (Frequency is medium) and (Monetary is good) then (AP is good)
11. If (Recency is medium) and (Frequency is good) and (Monetary is good) then (AP is medium)
12. If (Recency is low) and (Frequency is low) and (Monetary is low) then (AP is good)
13. If (Recency is low) and (Frequency is medium) and (Monetary is low) then (AP is medium)
14. If (Recency is low) and (Frequency is low) and (Monetary is low) then (AP is medium)
15. If (Recency is medium) and (Frequency is low) and (Monetary is medium) then (AP is medium)
16. If (Recency is low) and (Frequency is medium) and (Monetary is low) then (AP is medium)
17. If (Recency is medium) and (Frequency is low) and (Monetary is good) then (AP is medium)
18. If (Recency is low) and (Frequency is low) and (Monetary is medium) then (AP is good)
19. If (Recency is low) and (Frequency is good) and (Monetary is low) then (AP is medium)
20. If (Recency is good) and (Frequency is low) and (Monetary is good) then (AP is good)
21. If (Recency is good) and (Frequency is good) and (Monetary is good) then (AP is medium)
22. If (Recency is low) and (Frequency is low) and (Monetary is medium) then (AP is low)
23. If (Recency is good) and (Frequency is low) and (Monetary is good) then (AP is low)
24. If (Recency is low) and (Frequency is medium) and (Monetary is low) then (AP is medium)
25. If (Recency is medium) and (Frequency is low) and (Monetary is medium) then (AP is good)
26. If (Recency is medium) and (Frequency is good) and (Monetary is medium) then (AP is medium)
27. If (Recency is low) and (Frequency is medium) and (Monetary is good) then (AP is medium)
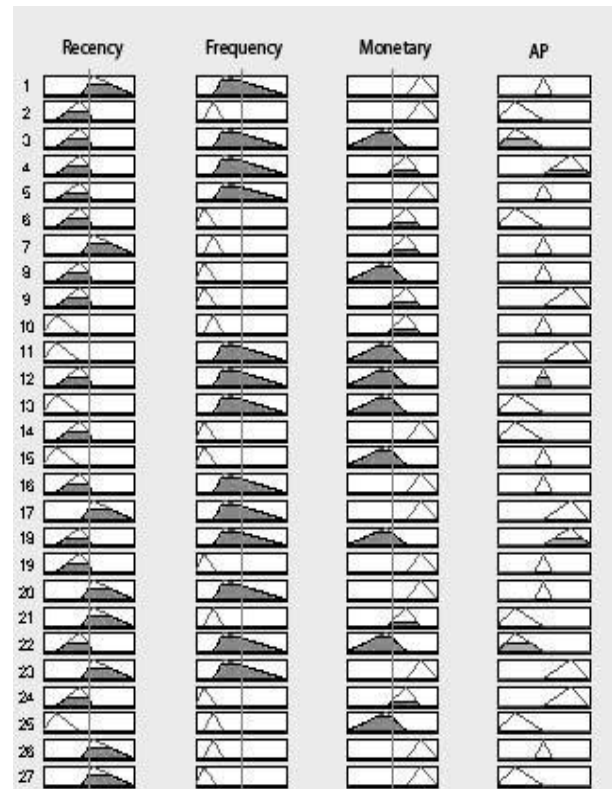

Fig. 5. Fuzzy rules after optimization.

As a result of the optimization process, the accuracy of estimating passing grades of distanceeducation students increased up to 84.90 %.

V. CONCLUSION

The proposed rules optimization based Fuzzy model has

been used to estimate academic performance of distanceeducation system students. Initially, the rules of Fuzzy System have been determined according to the expert opinion. Accuracy of the results of this modelis 74.7 %. Then the fuzzy rules have been optimized by genetic algorithm. The new model, optimizing the fuzzy rules, has produced new results having higher accuracy ratio, 85 %, compared to the previous model.Further studies could focus on optimizing fuzzy rules as well as simultaneously optimizing fuzzy membership functions with intervals.

REFERENCES

[1] S. B. Kotsiantis and P. E. Pintelas, "Predicting Students' Marks in Hellenic Open University," in *Proc. Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, Taiwan: IEEE Computer Society, 2005, pp. 664-668.

[2] D. Kalles and C. Pierrakeas, "Analyzing Student Performance in Distance Learning with Genetic Algorithms and Decision Trees," *Applied Artificial Intelligence*, pp. 655-674, 2006.

[3] Z. Ibrahim and D. Rusli, "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree And Linear Regressıon," *21st Annual SAS Malaysia Forum*, Kuala Lumpur, pp. 1-6, 2007.

[4] J. P. Vandamme, N. Meskens, and J. F. Superby, "Predicting Academic Performance by Data Mining Methods," *Education Economics*, pp. 405-419, December 2007.

[5] I. Lykourentzou, I. Giannoukos, G. Mpardis, V. Nikolopoulos, and V. Loumos, "Early and dynamic student achievement prediction in e-learning courses using neural networks," *JASIST*, pp. 372-380, 2009.

[6] A. Zafra and S. Ventura, "Predicting Student Grades in Learning Management Systems with Multiple Instance Genetic Programming," *Educational Data Mining 2009*, pp. 307-314, 2009.

[7] N. Yusof, N. A. Zin, N. M. Yassin, and P. Samsuri, "Evaluation of Student's Performance and Learning Efficiency based on ANFIS," in *Proc. 2009 International Conference of Soft Computing and Pattern Recognition*, IEEE Computer Society, 2009, pp. 460-465.

[8] O. Taylan and B. Karagozoglu, "An adaptive neuro-fuzzy model for prediction of student's academic performance," *Computers & Industrial Engineering*, pp. 732-741, 2009.

[9] O. Yildiz, A. Bal, S. Gulsecen, and F. D. Kentli, "A Genetic-Fuzzy Based Mathematical Model To Evaluate The Distance Education Students' Academic Performance," presented at the INT-E International Conference on New Horizons in Education ,Prague, June 05-07, 2012.

[10] R. L. Haupt and S. Haupt, *Practical Genetic Algorithms,* New Jersey: John Wiley & Sons, 2004.

[11] S. N. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms.* New York: Springer, 2008.

Osman Yildiz received his BS degree in mathematics from Marmara University in 1998, his MS degree in econometry from Marmara University in 2001 and 2005. He is still PhD student in informatics in Istanbul University in 1997 and 2002. He is a lecturer at Yildiz Technical University. His research interests include distance education, fuzzy logic, genetic algorithm and clustering.

Abdullah Bal received his BS degree in electronics and telecommunication engineering from Istanbul Technical University in 1993, his MS and PhD degrees in electrical engineering from Yildiz Technical University in 1997 and 2002. He is a faculty member at Yildiz Technical University. His research interests include digital/optical signal and image processing, pattern recognition, artificial neural networks, machine learning, Wavelet theory, and data/decision fusion.

Sevinç Gülseçen received her BS degree from İstanbul University, Faculty of Science, Department of Astronomy and Space Sciences in 1984 and MS and PhD degrees from Istanbul University, Faculty of Business Administration, Departmentof Quantitative Methods in 1989 and 1993 respectively. Currently she is professor and Chair of Informatics Department and Director of Computer Research and ApplicationCenter at IstanbulUniversity. Her research interests includeICT, System Analysis and Design, Constructivist Learning, e-Learning and Knowledge Management.

Fulya Damla Kentli completed her PhD at the Faculty of Education Science at University of Hamburg in 2008. She is a faculty member at Yildiz Technical University. Her research interests include curriculum development, research methods in education, vocational choice.