Fuzzy Logic Model for Predicting the Number of Online Courses Needed from Number of Students Enrolled in Higher Education

Rosa Leonor Ulloa Cazarez[1], Cuauhtémoc López-Martín[2]
[1, 2] Information Systems Department, CUCEA, Universidad de Guadalajara, Jalisco, Mexico
[1] rosi_ulloa@cucea.udg.mx, [2] cuauhtemoc@cucea.udg.mx

Abstract –

Context: In Higher Education where online courses are offered, one need is to predict the number of courses to be open. At date, some types of models have been used for this goal, such as models based upon machine learning, statistical and softcomputing approaches.
Goal: To propose a softcomputing model for predicting the number of online courses (NOC) needed from the number of students enrolled in Higher Education.
Hypothesis: Prediction accuracy of a fuzzy model is better or equal than a statistical regression model.
Results: Prediction accuracy of a fuzzy model was slightly better than that of a statistical regression model
Conclusion: Fuzzy logic could be applied for predicting the NOC needed from the number of students enrolled in Higher Education.

Keywords: Higher Education, Online courses prediction, fuzzy logic, statistical regression.

1 Introduction

Online education or e-learning, are terms used for the instruction facilitated and delivered online, through different technologies and typically have no face to face meetings [1].

Several institutes have incorporated into their curriculum online courses, in order to respond different purposes: improve quality of instruction, improvement of the use of spaces and attend the demand for enrollment [5]. Some institutes have offered full online programs and there are others which have born as online educative institutions. As a result of this variety, the definition of the online higher education system is a difficult job and is more accurate to the purposes of this study, called it as the offer of online education.

Because the nature of the online education, institutions have acquired benefits as the opportunity to provide more students with their educational services and the reduction of expenditures, and adult learners have been able to combine higher education and training with their work and home responsibilities [5].

These characteristics have allowed the growth of the online education offerings [3] whom have attracted an increasing number of participants and as a result, some projections of online education demand growth indicates that this effect continue for the next ten years [1] [7].

We believe that the faster growing of the enrollments in online education offerings, needs more attention in order to improve the planning processes through a proper analysis of the trends [1].

In this study we propose a fuzzy logic model for predicting the number of online courses (NOC) needed from the number of students enrolled in Higher Education. We have found studies using other dependent and independent variables, which have been used for generating and validating models based on (1) machine learning, such as neural networks, support vector machines, and genetic algorithms, (2) statistical models, such as simple and multiple regressions, and softcomputing, such as fuzzy logic.

The accuracy of our proposed model is compared with that of a statistical regression model. Data for generating and validating the models were obtained from a study realized by Kardan et al. [3] who predicted course selection of students in the context of two online courses. Data from the year 2005 to 2011 were used for generating the two models, and data of 2012 were used for validating the models.

Hypothesis of this research is the following:

> Prediction accuracy of a fuzzy model is better or equal than that of a statistical regression model.

The rest of this study is structured as follows: in section 1.1 a brief introduction to fuzzy logic is done, in Section 1.2 the studies related to ours are analyzed, whereas in Section 1.3, the criterion for evaluating the model is described. In Section 2, the data and the process followed for generating and validating the models are detailed. Sections 3 and 4, the models are generated and validated, respectively. Finaly, in section 5 the conclusions and future work are mentioned.

1.1 Fuzzy logic

The prediction techniques related to on-line courses have as characteristic that their variables use to be described using categorical data (nominal or ordinal scale) such as *small*, *medium*, *average*, or *high* named linguistic values. A more comprehensive approach to deal with linguistic values is by using fuzzy set theory [13].

A fuzzy model has the following two main properties [9]: (1) It operates in at the level of linguistic terms (fuzzy sets), and (2) it represents and process uncertainty.

A fuzzy logic model facilitates the representation and manipulation of uncertain, incomplete, imprecise or noisy data. Specifically, fuzzy logic offers a particularly convenient way to generate a keen mapping between input and output spaces thanks to natural expression of fuzzy rules [13]. In accordance with on-line courses in Higher Education issue, two considerations justify the decision of implementing a Fuzzy model: first, it is impossible to develop a precise mathematical model of the domain; second, measures only produce predictions of the real complexity. Thus, according to the previous assertions, formulating a set of natural rules describing interactions between the number of students and the number of on-line courses in Higher Education estimation could reveal their correlation. In this study a rule induction system replacing the crisp facts with fuzzy inputs, an inference engine uses a base of rules to map inputs to a fuzzy output which can either be translated back to a crisp value, is constructed.

A software tool was used to generate the fuzzy logic system with type: mamdani, *and* method: *min*, *or* method: *max*; implication: *min*, aggregation: *max*, and defuzzyfication: *centroid*.

1.2 Related work

Our literature review was focused in studies related with the following two main searches: (1) online higher education; and (2) prediction, estimation and forecasting models, (once we identified that these three words have indistinctly been used).

In the papers found, we targeted the following questions by study: (1) what was the purpose of the study? (2) What models/techniques were used? (3) What variables were involved? (4) What was the result? and (4) How the models/techniques were generated and validated?

Lykourentzou, I. GIannoukos, V. Nikolopoulos, G. Mpardis and V. Loumos [4], had the main purpose of predicting students dropout through the combination of three machine learning techniques. The three were a Feed-Forward neural network, a Support Vector Machine and a Probabilistic Ensembled Simplified Fuzzy named ARTMAP. The dependent variable was student dropout and the independent variables were of two types, (1) demographic (gender, residency, working experience in the field of the course, educational level, fluency of language) and (2) time variant (student’s progress, level of engagement and participation). According to the results of the study, the combination of the three machine learning techniques, leads to more accurate and pront identifying of the students dropout.

N. Nistor and K. Neubauer [8], predicted dropouts from student’s participation through Discriminant Analysis statistical technique. The dependent variable was students dropout and the independent variable was participation. In this study there were not intention of evaluating the discriminant analysis technique used, but a propose of strategie to measure and predict dropout takes place.

O. Yildiz, A. Bal, S. Gulsecen and F. Damla Kentli [11], evaluated academic performance in Distance Education using a Genetic-fuzzy model to predict academic performance from recency, frecuency and monetary. The prediction accuracy of students’ academic performance using the Genetic-Fuzzy model got an acceptable value.

S. Huang and N. Fang [2], predicted student performance in an engineering dynamics course through the comparation of four mathematical models: Multiple linear regression, two neural network models (Multilayer Perception neural network, and Radial Basis Function network), and a Support Vector Machine. The predictor variables were student’s problem solving skills, level of statisticals knowledge, student’s mathematical skills and students understanding of physical concepts. The analysis of results revealed that the type of mathematical model had only a slight effect on the average prediction accuracy and on the percentage of accurate predictions; on the other hand, the combination of predictor variables had only a slight effect on the Average Prediction Accuracy (APA) but a profound impact on the Percentage of Accurate Predictions (PAP). Support vector machine models had the highest PAP; and adding more predictor variables did not help improve the average prediction accuracy of any of the models.

A. A. Kardan, H. Sadeghi, S. S. Ghidary and M. R. Fani Sani [3], predicted students course selection through a Multi-layer Perceptron neural network from nine predictor variables: course characteristics, instructor’s characteristics, student’s workload, course grade, course type, course time, number of time conflicts, final examination time and student demands. In this study, the neural network proved to be effective and outperformed other methods predicting course selection considering students demands, but without considering the student demands, the neural network was not as accurate as the model with this consideration.

H.-W. Vivian Tang and M.-S. Yin [10], compared two grey prediction models and exponential smoothing for accuracy in prediction of the education expenditure from school enrollment. Forecasting efficiency of one of the grey prediction models used, was superior to exponential smoothing and the other grey prediction model used.

Jung Jae Yup [12], used statistical techniques derived from self-determination theory, expectancy-value theory and research on occupational indecision to predict enrollment in higher education from amotivation and indecision. The results of the study proved that both models had a good fit.

Based on the advantages described in section "1.1 fuzzy logic" as well as the previous related work analysis in which we did not find any study where a fuzzy model has been applied for predicting the number of on-line courses in higher education, we proposed a fuzzy logic model.

1.3 Accuracy criterion

A common criterion used to assess prediction models is the Magnitude of Relative Error (MRE) [6]. The MRE is defined as follows:

$$MRE_i = \frac{\left| Actual\ Courses_i \ \text{-}\ Predicted\ Courses_i \right|}{Actual\ Courses_i}$$

The MRE value is calculated for each observation *i* whose number of courses is predicted. The aggregation of MRE over multiple observations (N) can be achieved through the Mean MRE (MMRE) as follows:

$$MMRE = (1/N)\sum_{i=1}^{N} MRE_i$$

The accuracy of a prediction model is inversely proportional to the MMRE.

2 Experimental design

2.1 Data description

Table 2 shows the data of courses from 2005 to 2011 years obtained from [3] where CNE and ITME are the names of the courses (CNE: Computers Network Engineering, and ITME: Information Technology and Management Engineering).

2.2 Process for generating and validating the models

The steps for generating and validating the models were the following:

1. Selection of a sample involving data from 2005 to 2011 years.
2. Scatter plot analysis. The dependent variable versus independent variable.
3. Calculation of the coefficient of correlation.
4. Calculation of the coefficient of determination.
5. Linear regression equation generation.
6. ANOVA for the linear regression equation.
7. Fuzzy rules determination based on a correlation analysis.
8. Membership function selection.
9. Fuzzy logic model generation based on adjusting membership function parameters.
10. Validating the linear regression equation and fuzzy logic models using data of the 2012 year.

Table 2. Data for generating the models

Year	Season	Program	Number of students	Number of courses
2005	Spring	CNE	194	23
2005	Fall	CNE	180	22
2005	Spring	ITME	146	19
2005	Fall	ITME	161	21
2006	Spring	CNE	205	24
2006	Fall	CNE	188	23
2006	Spring	ITME	156	20
2006	Fall	ITME	177	22
2007	Spring	CNE	172	21
2007	Fall	CNE	186	22
2007	Spring	ITME	158	20
2007	Fall	ITME	184	23
2008	Spring	CNE	208	24
2008	Fall	CNE	197	23
2008	Spring	ITME	152	20
2008	Fall	ITME	160	21
2009	Spring	CNE	224	25
2009	Fall	CNE	196	23
2009	Spring	ITME	169	21
2009	Fall	ITME	175	22
2010	Spring	CNE	180	22
2010	Fall	CNE	215	24
2010	Spring	ITME	172	21
2010	Fall	ITME	191	23
2011	Spring	CNE	231	25
2011	Fall	CNE	194	23
2011	Spring	ITME	183	22
2011	Fall	ITME	179	22

3 Generation of models

3.1 Data analysis and linear regression model

Figure 1 shows a scatter plot correlating the number of students to number of courses. The relationship between these two variables is the following: The higher the value of students, the higher the number of courses.

The correlation value is r = 0.97, that is, there is a strong relationship between the two variables. The linear regresion model generated is the following:

Number of Courses = 9.3751 + 0.0698416*Number of students

The coefficient of determination is r^2 = 0.94, it means that the model as fitted explains 94% of the variability in courses.

The p-value of the analysis of variance (ANOVA) of the linear regression equation (Table 2) shows that there is a statistically significant relationship between the number of courses and the number of students at the 99.0% confidence level.

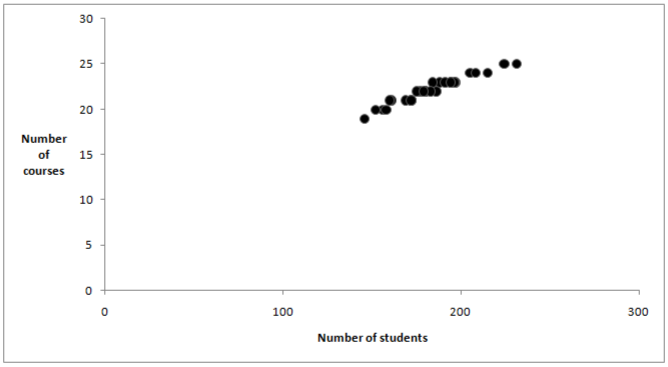

Figure 1. Scatter plot of Students vs. Courses

Table 2. ANOVA for the linear regression equation

Source	Sum of Squares	Degrees of freedom	Mean Square	F-Ratio	P-Value
Model	58.9738	1	58.9738	489.35	0.0000
Residual	3.1333	26	0.1205		
Total	62.1071	27			

3.2 Fuzzy model

The fuzzy model can be created from the expert knowledge that in a verbal form is translated into a set of if–then rules. A certain model structure is created, and parameters of this structure, such as membership functions and weights of rules, can be tuned using input and output data.

The following fuzzy rules were formulated based on the correlation (*r*) showed in Figure 1:

1) If (*Number of students* is *Small*) then (*Number of courses*) is *Small*
2) If (Number of students is *Big*) then (*Number of courses*) is *Big*

Implementing a fuzzy system requires that the different categories of the different inputs be represented by fuzzy sets, which in turn is represented by membership functions (MF). The MF type considered to this experiment is triangular (because of this type has demonstrated acceptable results when it has been applied to prediction [4]).

The input (*Number of students*) and output (*Number of courses*) was composed with two membership functions: *small* and *big*.

Parameters of membership functions for the input and for the output were iteratively adjusted until obtaining the smallest MMRE possible.

A triangular MF is a three-point (parameters) function, defined by minimum (a), maximum (c) and modal (b) values, that is, MF(a,b,c) where a ≤ b ≤c [9]. Their scalar parameters (a, b, c) are defined as follows:

$$MF(x) = 0 \text{ if } x < a \qquad MF(x) = 1 \text{ if } x = b \qquad MF(x) = 0 \text{ if } x > c$$

Figures 2 and 3 show the membership functions, whereas Table 3 shows the parameters for the input and output by triangular membership function.

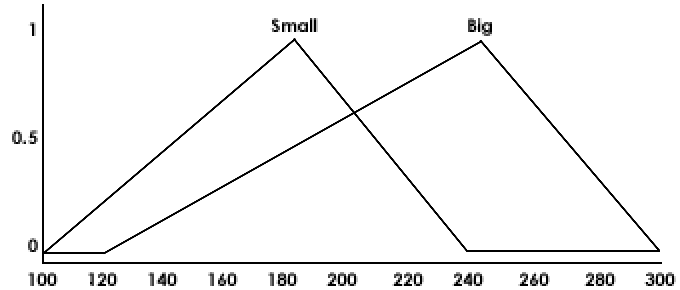

Figure 2. Membership functions for number of students (input)

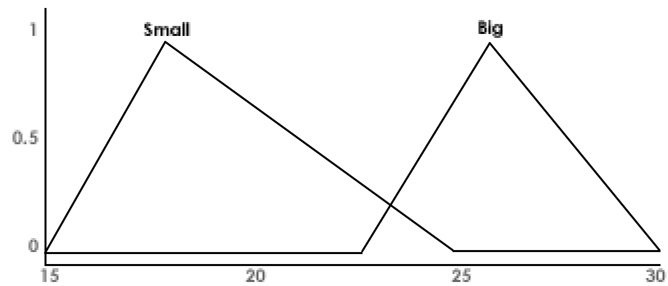

Figure 3. Membership functions for number of courses (output)

Table 3. Parameters of fuzzy model membership functions

Type of variable	**Variable**	**MF**	**Parameters**		
			a	**c**	**b**
Input	*Number of students*	Small	100	182	238
		Big	121	243	300
Output	*Number of courses*	Small	15	18	25
		Big	23	26	30

3.3 Model adequacy checking

The linear regression equation and the fuzzy model were applied to original data set (Table 1). The accuracy obtained (MMRE) by model is showed in Table 4.

4 Validation of models

Once that the linear regression equation and the fuzzy model were generated and their adequacy was checked, the two models were applied to a new data set obtained from [3] involving data of the 2012 year. The accuracy obtained (MMRE) is showed in Table 5. It can be showed that the fuzzy logic model had a slightly better accuracy (MMRE= 0.28).than the regression model (MMRE= 0.30).

Table 4. Prediction accuracy by model (LRE: Linear regression Equation; FLM: Fuzzy Logic Model)

Number of students	Number of courses	LRE	MRE	FLM	MRE
194	23	22.92	0.004	22.01	0.043
180	22	21.94	0.003	21.68	0.015
146	19	19.57	0.030	21.08	0.110
161	21	20.61	0.018	21.37	0.018
205	24	23.68	0.013	22.42	0.066
188	23	22.50	0.022	21.84	0.050
156	20	20.26	0.013	21.29	0.064
177	22	21.73	0.012	21.63	0.017
172	21	21.38	0.018	21.55	0.026
186	22	22.36	0.016	21.79	0.009
158	20	20.40	0.020	21.32	0.066
184	23	22.22	0.034	21.75	0.054
208	24	23.89	0.004	22.57	0.060
197	23	23.13	0.005	22.10	0.039
152	20	19.98	0.001	21.22	0.061
160	21	20.54	0.022	21.36	0.017
224	25	25.01	0.000	23.76	0.050
196	23	23.06	0.002	22.07	0.040
169	21	21.17	0.008	21.50	0.024
175	22	21.59	0.019	21.60	0.018
180	22	21.94	0.003	21.68	0.015
215	24	24.38	0.016	22.99	0.042
172	21	21.38	0.018	21.55	0.026
191	23	22.71	0.013	21.92	0.047
231	25	25.50	0.020	24.70	0.012
194	23	22.92	0.004	22.01	0.043
183	22	22.15	0.007	21.73	0.012
179	22	21.87	0.006	21.66	0.015
		MMRE	0.013		0.038

Table 5. Prediction accuracy by model (LRE: Linear regression Equation; FLM: Fuzzy Logic Model)

Number of students	Number of courses	LRE	MRE	FLM	MRE
235	25	25.78	0.031	25.46	0.018
223	24	24.94	0.039	23.65	0.014
166	21	20.96	0.002	21.45	0.022
180	23	21.94	0.046	21.68	0.057
		MMRE	0.030		0.028

5 Conclusions

In this study, data of the number of students enrolled in on-line Higher Education and number of courses were used for generating and validating independent data samples. The models for predicting the number of courses were a linear regression and a fuzzy model. These models were generated from a data set composed with 28 records that were obtained from 2005 to 2011 years. These two models were validated when they were used for predicting the number of courses of four records corresponding to 2012 year. The hypothesis accepted was the following:

> Prediction accuracy of a fuzzy model was better than a statistical regression model.

This result suggests that a fuzzy model using as independent variable (input) the number of students, can be used for predicting the number of on-line courses needed.

Future research involves the use of additional independent and dependent variables, obtained from other datasets, in models based on fuzzy logic and neural networks.

6 References

[1] Allen & Seaman, “Class Differences. Online Education in the United States, 2010”. United States of America: Babson Survey Research Group. 2010. Available at: http://sloanconsortium.org/publications/survey/class_differences

[2] Huang, S. & Fang, N. “Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models.” Computers & Education. 61 (133-145) (2013).

[3] Kardan, A., Sadeghi, H., Ghidary S. & Fani Sani, M. “Prediction of student course selection in online higher education institutes using neural network”. Computers & Education. Thompson Reuters. 65 (1-11) (2013)

[4] Lopez-Martin, C. Yañez-Marquez, C. & Gutierrez-Tornes, A. “Predictive Accuracy Comparison of Fuzzy Models for Software Development Effort of Small Programs”. Journal of Systems and Software, Vol. 81, No. 6, Elsevier, 2008

[5] Lykourentzou, I., GIannoukos, I., Nikolopoulos, V., Mpardis, G. & V. Loumos. “Dropout prediction in e-learning courses through the combination of machine learning techniques”. Computers & Education. Elsevier. 53. (950-965) (2009).

[6] Myrtveit, I., Stensrud, E. “Validity and reliability of evaluation procedures in comparative studies of effort prediction models”. Empirical Software Engineering. Springer. 17 (1-2) (2012) 23-33. DOI: 10.1007/s10664-011-9183-7

[7] National Center for Education Statistics, National Center for Education Statistics, Institute of Education Sciences, January 2013. Available at: http://nces.ed.gov/programs/projections/projections2021/. [Accessed 14 March 2013]

[8] Nistor, N. & Neubauer, K. “From participation to dropout: Quantitative participation patterns in online university courses”. Computers & Education.

Thompson Reuters. 55 (663-672) (2010).

[9] Pedrycz W., Gomide F. “An Introduction to Fuzzy Sets”; The MIT Press, 1998.

[10] Vivian Tang, H. & Yin, M. “Forecasting performance of grey prediction for education expenditure and school enrollment”. Economics of Education Review. Elsevier. 31 (452-462) (2012).

[11] Yildiz, O., Bal, A., Gulsecen, S. & Damla Kentli, F. “A Genetic-Fuzzy based mathematical model to evaluate the Distance Education students' academic performance” Procedia. Social and Behavioral Sciences. Elsevier. 55 (409-418) (2012).

[12] Yung, J. “Amotivation and indecision in the decision-making processes associate with University entry. Research in Higher Education. Elsevier. 54 (115-136) (2012).

[13] Zadeh. Lotfi A. “From Computing with Numbers to Computing with Words – From Manipulation of Measurements to Manipulation of Perceptions”. IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications, vol. 45, no. 1, pp. 105-119, 1999