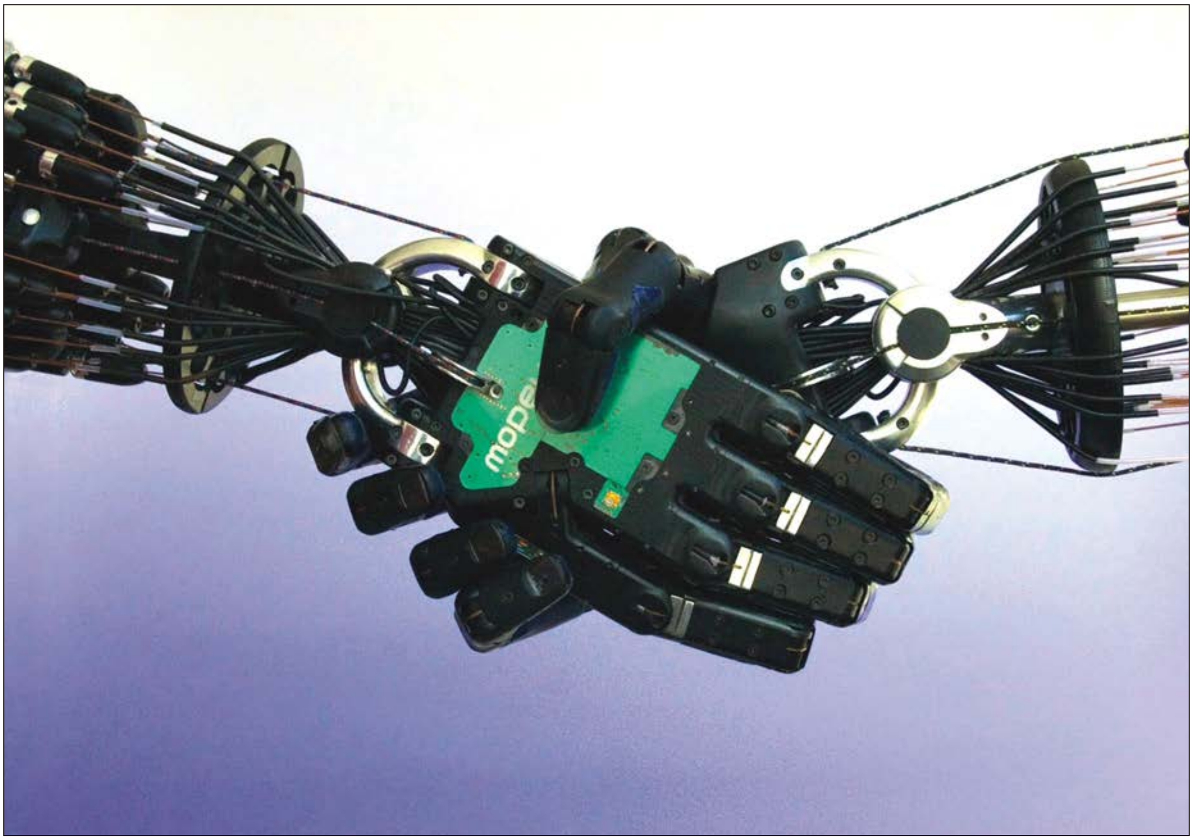

Journal of Artificial Intelligence

ISSN 1994-5450

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Artificial Intelligence 4 (1): 110-118, 2011
ISSN 1994-5450 / DOI: 10.3923/jai.2011.110.118

Enrollment Forecasting based on Modified Weight Fuzzy Time Series

Z. Ismail and R. Efendi
Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia

Corresponding Author: Zuhaimy Ismail, Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia Tel: +60197133940

ABSTRACT

Many different methods and models have been proposed by researchers using fuzzy time series for many different applications. The main issue in forecasting is in improving forecast accuracy. This paper presents the development of weight fuzzy time series based on a collection of variation of the chronological number in the Fuzzy Logical Group (FLG). The aim here is to develop an appropriate weight on fuzzy time series for forecasting of trend series data. A data set of university enrollment for Alabama University and Universiti Teknologi Malaysia (UTM) are used for forecasting. Results from this study shows that the proposed approach gave a lot of improvement. The forecasting fitness function used are the Means Square Error (MSE) and average error.

Key words: Fuzzy time series, forecast accuracy, enrollment modeling, forecasting, fuzzy sets, Linguistic value

INTRODUCTION

In the last decade, fuzzy time series has been widely used for forecasting data of dynamic and non-linear in nature. Many studies on forecasting using fuzzy logic time series have been discussed such as enrollment (Song and Chissom, 1993a, b; Chen, 1996; Chen and Hsu, 2004; Kuo *et al.*, 2009) the stock index (Huarng, 2001; Yu, 2005; Cheng *et al.*, 2006; Huarng *et al.*, 2007; Jilani and Burney, 2008; Yu and Huarng, 2008; Chu *et al.*, 2009), temperature (Chen, 2000) and financial forecasting (Lee *et al.*, 2006). Forecast accuracy is one of the main issues discussed (Ismail *et al.*, 2009a). Many approaches such as regression models, time series models and artificial intelligences were used for enrollment forecasting (Weiler, 1980; Gardner, 1981; Shaw, 1984; Ismail *et al.*, 2009b). Most of them were conducted using real data or numerical data. Enrollment of students in a university is very important as many decisions can be elaborated from them. However, obtaining accurate forecast of student enrolment is not an easy task, as many factors determine the impact of the enrollment numbers.

Many different methods and models have been proposed by researchers using fuzzy time series for enrollments forecasting. Song and Chissom (1993a, 1994) initiated a study on time-invariant and time-variant models for forecasting with fuzzy time series using enrollment data of Alabama University in 1993. Yu proposed the used of weight fuzzy time series models for Taiwan Stock Index (TAIEX) forecasting in 2004. It is assigned by the recurrent Fuzzy Logical Relationships (FLRs) in Fuzzy Logical Group (FLG). In establishing fuzzy relationship and forecasting are important step to consider the weight. Cheng *et al.* (2006). Also presented the trend-weight fuzzy time series model for TAIEX forecasting in 2005. On the other hand, the enrollment forecasting has been initiated by other researchers such as Sullivan and Woodall (1994), Chatman (1986), Pope and Evans (1985), Warrack and Russel (1983) and Hoenack and Weiler (1979). Their method do not consider fuzzy time series concept.

In this study, we consider a modified weight for fuzzy time series. It shows a significant reduction of the MSE (mean square error) and average error of forecasting when compared with Yu, (2005) and Cheng *et al.* (2006).

MATERIALS AND METHODS

The basic theory on fuzzy time series: Fuzzy time series is an application of fuzzy logic to time series analysis. In Zadeh (1975), introduced the fuzzy theory which the modern concept of uncertainty to deal with linguistics terms. Currently, fuzzy theory is frequently studied and applied in expert system, approximate reasoning, controls, pattern recognition, database and information retrieval systems, etc. (Huarng *et al.*, 2007). The definitions and classical fuzzy time series method have been presented as follows.

Definition 1: Let Y(t) (t = ..., 0, 1, 2, ...) is a subset of real number. Let Y(t) be the universe of discourse defined by the fuzzy set $\mu_i(t)$. If F(t) consists of $\mu_i(t)$ (i = 1, 2, ...), F(t) is defined as a fuzzy time series on Y(t) (t = ..., 0, 1, 2, ...) (Song and Chissom, 1993a, b).

Definition 2: Suppose F(t) is caused by F(t-1) denoted by F(t-1) → F (t), then this relationship can be represented by:

$$F(t)=F(t-1)^{\circ}R(t,t-1)$$

where, R(t, t-1) is a fuzzy relationship between F(t) and F(t-1) and is called the first-order model of F(t) (Song and Chissom, 1993a, b).

Definition 3: Let F(t-1) = A_i and F(t) = A_j. The relationship between two consecutive observations, F(t) and F(t-1), referred to as a Fuzzy Logical Relationship (FLR), can be denoted by $A_i \rightarrow A_j$, where A_i is called the left-hand side (LHS) and A_j is the Right-Hand Side (RHS) of the FLR (Song and Chissom, 1993a; Huarng *et al.*, 2007).

Definition 4: All fuzzy logical relationships in the training dataset can be further grouped together into different fuzzy logical relationship groups according to the same left-hand sides of the fuzzy logical relationship. For example, there are two fuzzy logical relationships with the same left-hand side (A_i): $A_i \rightarrow A_{j1}$ and $A_i \rightarrow A_{j2}$. These two fuzzy logical relationships can be grouped into a fuzzy logical relationship group as bellow (Song and Chissom, 1993a).

$$A_i \rightarrow A_{j1}, A_{j2}, \ldots \quad 1 \leq i, j \leq n$$

MODIFIED WEIGHT ON FUZZY TIME SERIES

The fundamental reason for weight: As described in the previous sections, this study proposed the modified weight for fuzzy time series forecasting. There are three main reasons why it is necessary to develop these modified weight for forecasting. Firstly, it is to resolve recurrent fuzzy

Table 1: The Fuzzy Logical Relationship (FLRs) and recurrence

FLR_s	Previous studies	Yu's
$A_1 \rightarrow A_1$		
$A_1 \rightarrow A_2$	$A_1 \rightarrow A_1, A_2$	$A_1 \rightarrow A_1, A_2, A_1,$
$A_2 \rightarrow A_1$		A_1
$A_1 \rightarrow A_1$	$A_2 \rightarrow A_1$	$A_2 \rightarrow A_1$
$A_1 \rightarrow A_1$		

Table 2: The chronological order on Fuzzy Logical Group (FLG)

FLR_s	Previous studies
$A_1 \rightarrow A_1$	
$A_1 \rightarrow A_2$	
$A_1 \rightarrow A_3$	$A_1 \rightarrow A_1, A_2, A_3, A_4, A_5$
$A_1 \rightarrow A_4$	
$A_1 \rightarrow A_5$	

relationships; secondly, it is to resolve the chronological order of Fuzzy Logical Relationships (FLRs), where both of them are used in different interpretation from Yu (2005) for weight determining and finally it is to propose a modified weight based on a collection of variation of chronological number in the Fuzzy Logical Group (FLG).

Recurrence of Fuzzy Logical Relationships (FLRs): Our reviews on previous studies in related areas show that the descriptions of the recurrent fuzzy relationships are not clearly given. The repeated FLRs were simply ignored when fuzzy relationships were established. The following examples as given by Yu (2005) can be used to explain this. Let there are FLR_s in chronological order as in Table 1.

Referring to Table 1, there are four out of five FLR that has the same LHS, A_1. The occurrences of the same FLR given in column 2 are regarded as if there were only one occurrence. In other words, the recent identical FLR are simply ignored in (1). It is questionable if the recurrence is ignored. The occurrence of a particular FLR represents the number of its appearances in the past. For instance, in column 3, $A_1 \rightarrow A_1$ appears three times and $A_2 \rightarrow A_1$ only once. The recurrence can be used to indicate how the FLR may appear in the future. Hence, to cover all of the FLR, an approach to representing the fuzzy relationship is suggested below:

$$A_1 \rightarrow A_1, A_2, A_1, A_1 \tag{1}$$

The various recurrences of FLR are viewpoint to assign of the modified weight.

Chronological order on Fuzzy Logical Group (FLG): This is the second type of FLR where there exist a chronological order between A_i and A_j. For example, it will be described as in Table 2. From column 2, it can be denoted that FLR_s between A_1 with others are in a chronological order. Each A_i and has a different linguistic value so that we can determine the various weight for fuzzy time series before forecasting. The collection of variation of the chronological number in the Fuzzy Logical Group (FLG) can be used as modified weight.

Table 3: The computational of weight on fuzzy time series

FLR_s	Chronological number	Weight W_n
$A_i \to A_j$	$j = c_1$	$w_1 = \frac{c_1}{(c_1+c_2+c_3+c_4)}$
$A_i \to A_k$	$k = c_2$	$w_2 = \frac{c_2}{(c_1+c_2+c_3+c_4)}$
$A_i \to A_l$	$l = c_3$	$w_3 = \frac{c_3}{(c_1+c_2+c_3+c_4)}$
$A_i \to A_m$	$m = c_4$	$w_4 = \frac{c_4}{(c_1+c_2+c_3+c_4)}$

Modified weight on fuzzy time series: In this study is proposed the weight fuzzy time series based on the various recurrences of FLR. Its computational is assigned as follows:

Suppose $A_i \to A_j, A_k, \ldots, A_p$ is a FLG and the weights are specified as follows: For $j = c_1$, $k = c_2$, ... , $p = c_n$, the computational can be determined as below:

$$W(t) = [w_1\, w_2 \ldots w_n] = \left[\frac{j}{(j+k+\ldots+p)}\ \frac{k}{(j+k+\ldots+p}\ \frac{p}{j+k+\ldots+p}\right]$$

$$= \left[\frac{c_1}{(c_1+c_2+\ldots+c_n)}\ \frac{c_2}{(c_1+c_2+\ldots+c_n)}\ \frac{c_n}{(c_1+c_2+\ldots+c_n)}\right] \qquad (2)$$

$$= \left[\frac{c_1}{\sum_{h=1}^{n} c_h}\ \frac{c_2}{\sum_{h=1}^{n} c_h} \cdots \frac{c_n}{\sum_{h=1}^{n} c_h}\right]$$

where, n is number of FLR in FLG.

Here, the modified weight can be elicited by using recurrences and the chronological number of FLRs in FLG as follows:

Let is assumed that there are possibility that i, j,... , p a re equal or can be at least one i, j,... , p not equal. Let be a recurrence and can be assumed t hat $j = k = m \neq l$ where $A_i \to A_j$ appears two time and $A_i \to A_l$ only once in FLG. Then, the computational for weights can be done as in Table 3.

$$w_1 + w_2 + w_3 + w_4 = \frac{c_1}{\sum_{h=1}^{4} c_h} + \frac{c_1}{\sum_{h=1}^{4} c_h} + \frac{c_1}{\sum_{h=1}^{4} c_h} + \frac{c_1}{\sum_{h=1}^{4} c_h} = \frac{3c+c_3}{3c+c_3} = 1$$

Further, from column 3 in Table 3 the weight can be proven to have a necessary condition as follows because of c_1, c_2 and c_4 are equal so that they can be denoted as c, or it can be written as below:

$$\sum_{h=1}^{4} w_h = 1 \qquad (3)$$

In addition, the computational of the weight based on chronological order can be assigned as follow:

Let $A_i \rightarrow Aj$, A_k, A_l, A_m be a chronological order and can be assumed that $j \neq k \neq l \neq m$ where between A_i with A_j, A_k, A_l, and A_m only once in FLG. Then, the computational for weights can be done as like in Table 3. From Table 3 column 3 the weight can be proven that is the necessary condition as follows:

$$w_1 + w_2 + w_3 + w_4 = \frac{c_1}{\sum_{h=1}^{4} c_h} + \frac{c_2}{\sum_{h=1}^{4} c_h} + \frac{c_3}{\sum_{h=1}^{4} c_h} + \frac{c_4}{\sum_{h=1}^{4} c_h} = \frac{\sum_{h=1}^{4} c_h}{\sum_{h=1}^{4} c_h} = 1$$

or it can be written as below:

$$\sum_{h=1}^{4} w_h = 1 \qquad (4)$$

From the Eq. 8 and 9 it can be denoted that the weight as below:

$$\sum_{h=1}^{4} w_h = 1 \qquad (5)$$

Here, the first modification can be seen from chronological number usage. This is the main differences between our approach and that of Yu where the weights were assumed to be $w_1 = 1$, $w_2 = 2$, ..., $w_m = m$ (m is natural number). These weight values increased gradually. In addition, if weights were multiplied by the midpoint intervals then the forecast values will also increase. On the other hand, Cheng *et al.* (2006). proposed the trend-weight fuzzy time series where they considered the recurrent classification of fuzzy relationships into three different types of trends and assign a proper weight to individual fuzzy relationships. The second modification can be described in reversal of transpose matrix elements on forecasting rule.

Proposed fuzzy time series forecasting method: In this proposed method of fuzzy time series, Yu (2005) used the weights and midpoint intervals as the forecasting method. The weights are determined by using a collection of variation of the chronological number in the FLG and the detail is given by the following example. The Forecasting method as follow:

Let $A_i \rightarrow A_j$, A_k,..., A_p is a FLG and the corresponding weights for are w_1, w_2, ..., w_n. The defuzzified of the midpoints of are m_j, m_k, ..., m_p. It can be denoted in the product of the defuzzified matrix and the transpose of the weight matrix:

$$F(t) = M(t) \times W(t)^T = [m_j\, m_k \ldots m_p] \times [w_1 w_2 \ldots w_n]^T$$

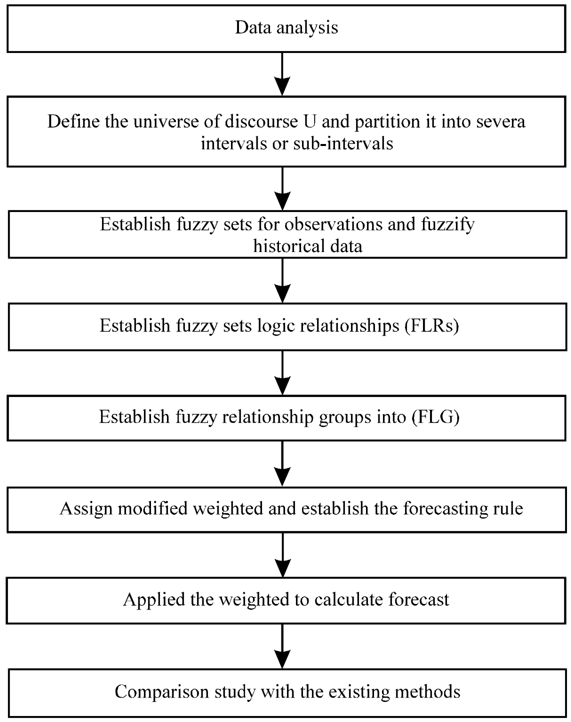

Fig. 1: The forecasting steps

$$= \left[\mathbf{m}_j \ \mathbf{m}_k \ \dots \ \mathbf{m}_p \right] \times \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_n \end{bmatrix} \tag{6}$$

where the number elements in matrix M(t) and W(t) are equal.

From Eq. 6 the modification can be done with reversal of weight elements in transpose matrix as follows:

$$= \left[m_j \ m_k \ \dots \ m_p \right] \times \begin{bmatrix} w_n \\ \vdots \\ w_2 \\ w_1 \end{bmatrix} \tag{7}$$

where, M(t) is a 1 x n matrix and $W(t)^T$ is a n x 1 matrix, respectively. By using Eq.(7), the increasing of forecasting error can be reduced. The forecasting steps are described in Fig. 1.

RESULTS AND DISCUSSION

The performance of the proposed method will be compared with Yu's method and Cheng's method by sing the enrollment of Alabama University from 1972 to 1992. It can be illustrated in Table 4.

Table 4: The enrollment forecasting of Alabama University using different methods

Year	Enrollments	Proposed method	Cheng's method	Yu's method
1971	13055	-	-	-
1972	13563	13863	13680.5	14250
1973	13867	13933	13731.3	14250
1992	18876	18876	19033.7	19500
MSE	16248.7	192084.3	259357.5	
Average error of forecasting	0.50%	2.08%	14.50%	

Table 5: The enrollment forecasting of Universiti Teknologi Malaysia (UTM) based on the proposed method

Year	Postgraduate	Proposed method	Undergraduate	Proposed method
1990	175	-	4158	-
1991	205	260.0	4615	4615.0
1992	305	310.0	5227	5441.3
1993	439	527.2	5383	5331.3
2007	3958	3958.0	16417	16320.9
2008	4721	4721.0	15010	15010.0
MSE		12546.8		101.7
Average error forecasting		0.074%		0.769%

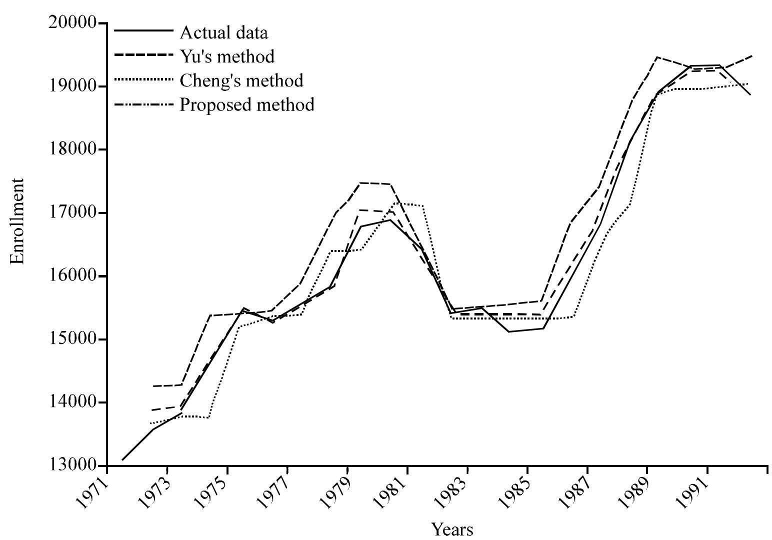

Fig. 2: Comparing proposed method with other methods

Table 4, the MSE and average error of forecasting are 16248.7 and 0.5% of the proposed method respectively. On the other hand, the MSE and average error of forecasting of Cheng's method are 192084.3 and 2.08% consecutively. In addition, Yu's method given the MSE 259357.5 and average error of forecasting is 14.47%. Thus, it is that obvious that the proposed method has a smaller MSE and lower average error than the existing methods. The trends in forecast by above mentioned methods are being illustrated in Fig. 2.

Following the steps then performance of UTM enrollment forecasting can be tested and trained by using the proposed method. The forecasting results can be seen in Table 5.

Table 5 indicates that MSE and average error of forecasting for postgraduate and undergraduate enrollment are 12546.8, 0.074% and 101.7, 0.769% respectively.

CONCLUSION

In the proposed method, it has been presented the weight approach for analyzing fuzzy time series forecasting. The modified weight are assigned by using the chronological number of FLRs in FLG and a modification is also done in reversal of weight elements on transpose matrix for forecasting rule. The proposed method showed smaller values of MSE lower percentage of average error of forecasting.

The error of forecasting most likely will be smaller as the length of intervals reduces that based on fuzzy logical relationships and the midpoint intervals. In fuzzy time series, many various methods have been designed in different form due to non standard rule to be followed. Further study may explore an appropriate weight that can be extended to reach a higher forecasting accuracy for each component of time series data and investigate new methods which may be adapted with multi-variate fuzzy time series and out-sample forecasts.

ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education (MOHE) under Fundamental Science Fund Grant Vote No. 78313 and the Department of Mathematics, Faculty of Science, University Technology Malaysia.

REFERENCES

Chatman, S.P., 1986. Short-term forecasts of the number and scholastic ability of enrolling freshmen by academic divisions. Res. Higher Educ., 25: 68-81.

Chen, S.M., 1996. Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst., 81: 311-319.

Chen, S.M. and C.C. Hsu, 2004. A new method to forecast enrollments using fuzzy time series. Int. J. Applied Sci. Eng., 3: 234-244.

Chen, S.M., 2000. Temperature prediction using fuzzy time series. IEEE Trans. Syst. Man Cybern B Cybern, 30: 263-275.

Cheng, C.H., T.L. Chen and C.H. Chiang, 2006. Trend-weight fuzzy time series model for TAIEX forecasting. Neural Inf. Process., 4234: 469-477.

Chu, H.H., T.L. Chen, C.H. Cheng and C.C. Huang, 2009. Fuzzy dual-factor time series for stock index forecasting. Expert Syst. Appl. Int. J., 36: 165-171.

Gardner, D.E., 1981. Weight factor selection in double exponential smoothing enrollment forecasts. Res. Higher Educ., 14: 49-56.

Hoenack, S.A. and W.C. Weiler, 1979. The demand for higher education and institutional enrollment forecasting. Econ. Inquiry, 17: 89-113.

Huarng, K., 2001. Heuristic models of fuzzy time series for forecasting. Fuzzy Sets Syst., 123: 369-386.

Huarng, K.H., T.H. Yu and Y.W. Hsu, 2007. A multivariate heuristic model for fuzzy time series forecasting. IEEE Trans. Syst. Man Cybern. B Cybern., 37: 836-846.

Ismail, Z., A. Yahaya and A. Shabri, 2009a. Forecasting gold prices using multiple linear regression method. Am. J. Applied Sci., 6: 1509-1514.

Ismail, Z., A. Yahya and K.A. Mahpol, 2009b. Forecasting peak load electricity demand using statistics and rule based approach. Am. J. Applied Sci., 6: 1618-1625.

Jilani, T.A. and S.M.A. Burney, 2008. A refined fuzzy time series model for stock market forecasting. Physica A, 387: 2857-2862.

Kuo, I.H., S.J. Horng, T.W. Kao, T.L. Lin, C.L. Lee and Y. Pan, 2009. An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. Expert Syst. Appl., 36: 6108-6117.

Lee, C.H.L., A. Liu and W.S. Chen, 2006. Pattern discovery of fuzzy time series for financial prediction. IEEE Trans. Knowledge Data Eng., 18: 613-625.

Pope, J.A. and J.P. Evans, 1985. A forecasting system for college admissions. Coll. Univ., 60: 113-131.

Shaw, R.C., 1984. Enrollment forecasting: What methods work the best. NASSP Bull., 68: 52-58.

Song, Q. and B.S. Chissom, 1993a. Forecasting enrollments with fuzzy time series. Fuzzy Sets Syst., 54: 1-9.

Song, Q. and B.S. Chissom, 1993b. Fuzzy time series and its models. Fuzzy Sets Syst., 54: 269-277.

Song, Q. and B.S. Chissom, 1994. Forecasting enrollments with fuzzy time series. Fuzzy Sets Syst., 62: 1-8.

Sullivan, J. and W.H. Woodall, 1994. A comparison of fuzzy forecasting and Markov modeling. Fuzzy Sets Syst., 64: 279-293.

Warrack, B.J. and C.N. Russel, 1983. Forecasting demand for postsecondary education in Manitoba: The motivational index and the demand index as an enrollment forecasting tool. Res. Higher Educ., 19: 335-349.

Weiler, W.C., 1980. A model for short-term institutional enrollment forecasting. J. Higher Educ., 51: 314-327.

Yu, H.K., 2005. Weighted fuzzy time series models for TAIEX forecasting. Phys. A Statistical Mechanics Appl., 349: 609-624.

Yu, T.H.K. and K.H. Huarng, 2008. A bivariate fuzzy time series model to forecast the TAIEX. Expert Syst. Appl., 34: 2945-2952.

Zadeh, L.A., 1975. The concept of a linguistic variable and its application to approximate reasoning-I. Inform. Sci., 8: 199-249.