Multidisciplinary Informatics Journal
Vol. 2, No. 2, December 2019, Pg. 117 – 125

FUZZY TIME SERIES FORECASTING MODEL BASED ON SECOND ORDER FUZZY LOGICAL RELATIONSHIP AND SIMILARITY MEASURE

Nik Muhammad Farhan Hakim Nik Badrul Alam and Nazirah Ramli

Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA (UiTM) Pahang,
26400 Bandar Jengka
Author Correspondence e-mail: farhanhakim@uitm.edu.my

Received: 6 September 2019 / Accepted: 5 November 2019 / Published online: 15 December 2019

ABSTRACT

Various fuzzy time series (FTS) forecasting methods have been proposed to cater for data in linguistic values. In this paper, an improved FTS forecasting method based on second order fuzzy logical relationship is proposed and it is used to forecast the enrollment of students in the University of Alabama. The performance of the forecasted results is compared to the actual data by using seven different similarity measures. The hybrid similarity measure based on geometric distance, centre of gravity, area, perimeter and height gives the best performance.

Keywords: Fuzzy Time Series; Similarity Measure; Second Order; Trapezoidal Fuzzy Number.

1. INTRODUCTION

Fuzzy time series (FTS) is very powerful in forecasting historical data with linguistic values since it was first introduced by [1] and [2]. The forecasting methods have been changed and analysed by [3-5]. In their studies, the fuzzy set was used to define the linguistic values and the methods used in the first order forecasting method. [6] used high-order fuzzy time series to forecast the student enrollments in the University of Alabama. However, [1-6] used the fuzzy set in defining the linguistic values which are unable to produce the forecasted range under different degrees of confidence. [7] first used trapezoidal fuzzy number (TrFN) to forecast the student enrollments in the University of Alabama, then the methods of forecasting were changed and developed by [8-11]. To analyse the performance of the forecasting results, the TrFNs were defuzzified into crisp values to calculate the mean absolute percent error (MAPE), mean square error (MSE) and root mean square error (RMSE). During the defuzzification process, some information on the data was lost.

An alternative to evaluate the performance of the forecasting model is by using similarity measure concept, but it is only applied in a limited number of studies. In this paper, we propose an improved FTS forecasting method using second order FTS and TrFNs, and the performance is evaluated using multiple types of similarity measures. This paper is organised as follows; section 2 consists of some definitions of fuzzy time series and trapezoidal fuzzy number, section 3 presents the proposed forecasting model, section 4 illustrates the proposed method using a numerical example, section 5 consists of the discussion of the results, and section 6 concludes this paper.

2. PRELIMINARIES

In this section, some basic definitions are listed to support the proposed forecasting model.

Definition 1 [1]: *Let* $X(t)$ $(t=...,0,1,2,...)$ *be a subset of* $\mathbb{R}$ *and* $X(t)$ *is the universe of discourse defined by fuzzy set* $U_i(t)$ $(t=...,0,1,2,...)$, *then* $F(t)$ *is called fuzzy time series on* $X(t)$ $(t=...,0,1,2,...)$.

Definition 2 [1]: *If there exists a fuzzy relationship* $R(t-1,t)$ *such that* $F(t)=F(t-1)*R(t-1,t)$ *where * represents the fuzzy operator, then* $F(t)$ *is said to be caused by* $F(t-1)$. *The relationship can be noted as* $F(t-1)\rightarrow F(t)$.

Definition 3 [7]: *A trapezoidal fuzzy number* $\tilde{N}$, *denoted by* $\tilde{N}=(n_1,n_2,n_3,n_4)$ *is defined as:*

$$\mu_{\tilde{N}}(x)=\begin{cases} 0 & , \quad x<n_1 \\ \dfrac{x-n_1}{n_2-n_1} & , \quad n_1\le x\le n_2 \\ 1 & , \quad n_2\le x\le n_3 \\ \dfrac{n_4-x}{n_4-n_3} & , \quad n_3\le x\le n_4 \\ 0 & , \quad x>n_4 \end{cases} \tag{1}$$

3. SECOND ORDER FTS FORECASTING MODEL

In this section, we propose the FTS forecasting model. After the historical data is collected, the following steps is done.

Step 1. Define the universe of discourse P as $[T_{\min}-T_1, T_{\max}+T_2]$ where $T_{\min}$ and $T_{\max}$ are the minimum historical data and the maximum historical data respectively, while T_1 and T_2 are two proper positive numbers.

Step 2. Partition universe of discourse P into seven intervals of equal length. These intervals are labelled as P_i where $i = 1,2,3,...,7$.

Step 3. Establish the trapezoidal fuzzy numbers to represent the linguistic values of the intervals of P. Suppose $P_1 = [n_1, n_2]$, $P_2 = [n_2, n_3]$, …, $P_6 = [n_6, n_7]$ and $P_7 = [n_7, n_8]$, according to [7], the trapezoidal fuzzy numbers are defined as:

$$\begin{aligned} \tilde{N}_1 &= (n_0, n_1, n_2, n_3) \\ \tilde{N}_2 &= (n_1, n_2, n_3, n_4) \\ \vdots & \qquad \vdots \\ \tilde{N}_6 &= (n_5, n_6, n_7, n_8) \\ \tilde{N}_7 &= (n_6, n_7, n_8, n_9) \end{aligned}$$

Step 4. Fuzzify the historical data. According to [7], if the value of the historical data is located in the range of P_i where $i = 1,2,3,...,7$, then it belongs to the fuzzy number $\tilde{N}_i$ where $i = 1,2,3,...,7$.

Step 5. Develop the second order fuzzy logical relationship. [6] denoted "if the data value of time $t-1$ and t are $\tilde{N}_i$ and $\tilde{N}_j$ respectively, then that of time $t+1$ is $\tilde{N}_k$" as $\tilde{N}_i, \tilde{N}_j \rightarrow \tilde{N}_k$. However, the repeated relationships are not accounted [2]. Next, generate the second order fuzzy logical relationship group as follows:

Table 1. Fuzzy Logical Relationship Group

Group	**Fuzzy logical relationship**
Group 1	$\tilde{N}_a, \tilde{N}_b \rightarrow \tilde{N}_c, \tilde{N}_a, \tilde{N}_b \rightarrow \tilde{N}_d, ...$
Group 2	$\tilde{N}_b, \tilde{N}_c \rightarrow \tilde{N}_d, \tilde{N}_b, \tilde{N}_c \rightarrow \tilde{N}_e, ...$
$\vdots$	$\vdots$
Group m	$\tilde{N}_e, \tilde{N}_f \rightarrow \tilde{N}_c, \tilde{N}_e, \tilde{N}_f \rightarrow \tilde{N}_d, ...$

Step 6. Calculate the forecasted value, $\tilde{F}_t$ of time t by using the rules proposed by [6] as follows:

(i) If the fuzzy logical relationship group of $\tilde{N}_i, \tilde{N}_j$ is empty, that is $\tilde{N}_i, \tilde{N}_j \rightarrow \phi$, then $\tilde{F}_t = \dfrac{\tilde{N}_i + \tilde{N}_j}{2}$.

(ii) If the fuzzy logical relationship group of $\tilde{N}_i, \tilde{N}_j$ is one-to-one, that is $\tilde{N}_i, \tilde{N}_j \rightarrow \tilde{N}_k$, then $\tilde{F}_t = \tilde{N}_k$.

(iii) If the fuzzy logical relationship group of $\tilde{N}_i, \tilde{N}_j$ is one-to-many, that is $\tilde{N}_i, \tilde{N}_j \rightarrow \tilde{N}_{k1}, \tilde{N}_i, \tilde{N}_j \rightarrow \tilde{N}_{k2}, \ldots, \tilde{N}_i, \tilde{N}_j \rightarrow \tilde{N}_{kp}$, then $\tilde{F}_t = \dfrac{\tilde{N}_{k1} + \tilde{N}_{k2} + \ldots + \tilde{N}_{kp}}{p}$.

4. FORECASTING ENROLLMENT

The proposed second order FTS forecasting approach is illustrated using the data of students' enrollment in the University of Alabama, which is adopted from [2] and shown in Figure 1.

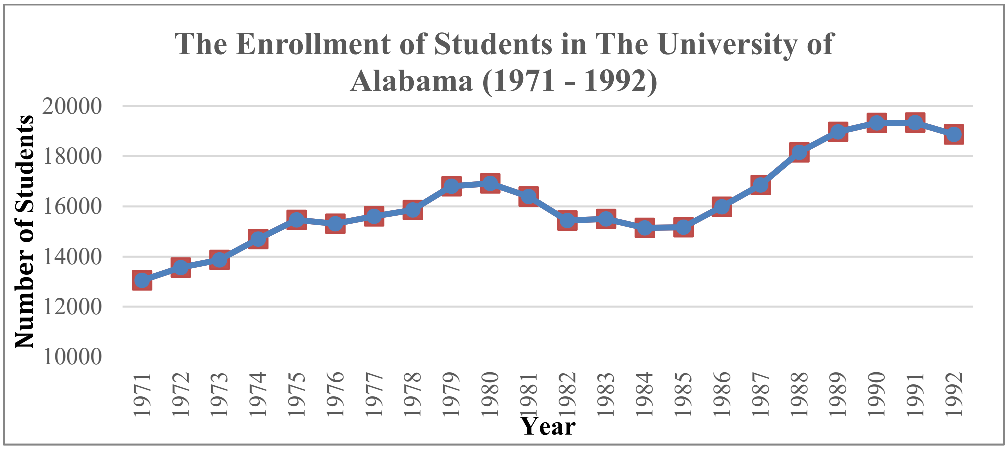

Fig.1. Enrollments in the University of Alabama

Step 1. From Figure 1, $T_{\min} = 13055$ and $T_{\max} = 19337$. $T_1 = 55$ and $T_2 = 663$ are chosen such that $P = [13000, 20000]$.

Step 2. Divide P into seven partitions, namely $P_1 = [13000, 14000]$, $P_2 = [14000, 15000]$, $P_3 = [15000, 16000]$, $P_4 = [16000, 17000]$, $P_5 = [17000, 18000]$, $P_6 = [18000, 19000]$ and $P_7 = [19000, 20000]$.

Step 3. Establish the trapezoidal fuzzy numbers as follows:

Table 2. Trapezoidal Fuzzy Numbers for the Enrollment in the University of Alabama

Intervals	Trapezoidal Fuzzy Numbers
$P_1 = [13000, 14000]$	$\tilde{N}_1 = (12000, 13000, 14000, 15000)$
$P_2 = [14000, 15000]$	$\tilde{N}_2 = (13000, 14000, 15000, 16000)$
$P_3 = [15000, 16000]$	$\tilde{N}_3 = (14000, 15000, 16000, 17000)$
$P_4 = [16000, 17000]$	$\tilde{N}_4 = (15000, 16000, 17000, 18000)$
$P_5 = [17000, 18000]$	$\tilde{N}_5 = (16000, 17000, 18000, 19000)$

$P_6 = [18000, 19000]$	$\tilde{N}_6 = (17000, 18000, 19000, 20000)$
$P_7 = [19000, 20000]$	$\tilde{N}_7 = (18000, 19000, 20000, 21000)$

Step 4. Fuzzify data of historical enrollments at the University of Alabama. The fuzzified data is shown in Table 3 below.

Table 3. The Fuzzified Data of Enrollments at the University of Alabama

Year	TrFNs	Year	TrFNs
1971	(12000,13000,14000,15000)	1982	(14000,15000,16000,17000)
1972	(12000,13000,14000,15000)	1983	(14000,15000,16000,17000)
1973	(12000,13000,14000,15000)	1984	(14000,15000,16000,17000)
1974	(13000,14000,15000,16000)	1985	(14000,15000,16000,17000)
1975	(14000,15000,16000,17000)	1986	(14000,15000,16000,17000)
1976	(14000,15000,16000,17000)	1987	(15000,16000,17000,18000)
1977	(14000,15000,16000,17000)	1988	(17000,18000,19000,20000)
1978	(14000,15000,16000,17000)	1989	(15000,16000,17000,18000)
1979	(15000,16000,17000,18000)	1990	(18000,19000,20000,21000)
1980	(15000,16000,17000,18000)	1991	(18000,19000,20000,21000)
1981	(15000,16000,17000,18000)	1992	(17000,18000,19000,20000)

Step 5. From Table 3, develop the second order fuzzy logical relationships and classify them into some groups to obtain the following result:

Table 4. Fuzzy Logical Relationship Groups of Enrollment at the University of Alabama

Group	Fuzzy logical relationships	Group	Fuzzy logical relationships
Group 1	$\tilde{N}_1, \tilde{N}_1 \to \tilde{N}_1$, $\tilde{N}_1, \tilde{N}_1 \to \tilde{N}_2$	Group 7	$\tilde{N}_4, \tilde{N}_3 \to \tilde{N}_3$
Group 2	$\tilde{N}_1, \tilde{N}_2 \to \tilde{N}_3$	Group 8	$\tilde{N}_4, \tilde{N}_6 \to \tilde{N}_6$
Group 3	$\tilde{N}_2, \tilde{N}_3 \to \tilde{N}_3$	Group 9	$\tilde{N}_6, \tilde{N}_6 \to \tilde{N}_7$
Group 4	$\tilde{N}_3, \tilde{N}_3 \to \tilde{N}_3$, $\tilde{N}_3, \tilde{N}_3 \to \tilde{N}_4$	Group 10	$\tilde{N}_6, \tilde{N}_7 \to \tilde{N}_7$
Group 5	$\tilde{N}_3, \tilde{N}_4 \to \tilde{N}_4$, $\tilde{N}_3, \tilde{N}_4 \to \tilde{N}_6$	Group 11	$\tilde{N}_7, \tilde{N}_7 \to \tilde{N}_6$
Group 6	$\tilde{N}_4, \tilde{N}_4 \to \tilde{N}_4$, $\tilde{N}_4, \tilde{N}_4 \to \tilde{N}_3$	Group 12	$\tilde{N}_7, \tilde{N}_6 \to \#$

Step 6. Calculate the forecasted enrollments and the results are listed in Table 5.

Table 5. The Forecasted Enrollments for Years 1978 – 1993

Year	Actual Enrollment	Trapezoidal Fuzzy Numbers	Forecasted Enrollments
1978	15861	(14000,15000,16000,17000)	(14500,15500,16500,17500)
1979	16807	(15000,16000,17000,18000)	(14500,15500,16500,17500)
1980	16919	(15000,16000,17000,18000)	(16000,17000,18000,19000)
1981	16388	(15000,16000,17000,18000)	(14500,15500,16500,17500)
1982	15433	(14000,15000,16000,17000)	(14500,15500,16500,17500)
1983	15497	(14000,15000,16000,17000)	(14000,15000,16000,17000)
1984	15145	(14000,15000,16000,17000)	(14500,15500,16500,17500)
1985	15163	(14000,15000,16000,17000)	(14500,15500,16500,17500)
1986	15984	(14000,15000,16000,17000)	(14500,15500,16500,17500)
1987	16859	(15000,16000,17000,18000)	(14500,15500,16500,17500)
1988	18150	(17000,18000,19000,20000)	(16000,17000,18000,19000)
1989	18970	(17000,18000,19000,20000)	(17000,18000,19000,20000)
1990	19328	(18000,19000,20000,21000)	(18000,19000,20000,21000)
1991	19337	(18000,19000,20000,21000)	(18000,19000,20000,21000)
1992	18876	(17000,18000,19000,20000)	(17000,18000,19000,20000)
1993			(17500,18500,19500,20500)

5. DISCUSSION

Based on the forecasted results obtained in the previous section, we calculate the similarity measure to see how close they are as compared to the actual fuzzy numbers. Several similarity measures [12-18] are used and the results are compared and analysed to acquire the best performance.

Table 6. Similarity Measures for the Enrollments in the University of Alabama

Year	[12]	[13]	[14]	[15]	[16]	[17]	[18]
1973	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1974	0.980100	0.990000	0.990528	0.990000	0.980199	0.999911	0.969798
1975	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1976	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1977	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985001
1978	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985001
1979	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985057
1980	0.980100	0.990000	0.990528	0.990000	0.980199	0.999911	0.970470
1981	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985057
1982	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985001
1983	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

1984	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985001
1985	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985001
1986	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985001
1987	0.990025	0.995000	0.995264	0.995000	0.990050	0.999978	0.985057
1988	0.980100	0.990000	0.990528	0.990000	0.980199	0.999911	0.970681
1989	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1990	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1991	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1992	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
Average	0.992526	0.996250	0.996448	0.996250	0.992552	0.999977	0.988806

The similarity measure for years 1973, 1975, 1976, 1983, and 1989 until 1992 is 1.00, which shows that the calculated forecasted data is exactly the same as the historical. Among all the similarity measures, [17] shows the highest similarity measure for each year making its average outperforms among all. The performance of the similarity measures can be ranked as in Table 7.

Table 7. Ranking of Performance of Similarity Measures

Ranking	Similarity Measures	Performance (%)
		Enrollments in the University of Alabama
1	[17]	99.9977
2	[14]	99.6448
3	[15]	99.6250

Among all similarity measures used, the similarity measure proposed by [17] has the highest performance based on their average. This means that the similarity measure based on geometric distance, centre of gravity (COG), area, perimeter and height is the best to be used to compare the forecasted and actual trapezoidal fuzzy numbers in further studies. Next, the performance of the second order fuzzy logical relationship obtained from Table 7 is compared to the performance of the first order fuzzy logical relationship from [19] as shown in Table 8.

Table 8. Comparison of Performance of Similarity Measures between First Order and Second Order Fuzzy Logical Relationship

Similarity Measures	[17]		[14]		[15]	
	First Order [19]	Second Order	First Oder [19]	Second Order	First Order [19]	Second Order
Performance (%)	99.7868	99.9977	99.7643	99.6448	99.5858	99.6250

The similarity measure from [17] and [15] produce higher similarity values for the second order fuzzy logical relationship compared to the first order. While, the second order fuzzy logical relationship similarity from [14] is slightly lower than that of the first order. Furthermore, [17] outperforms others for both first and second order fuzzy logical relationship.

6. CONCLUSION

This paper has presented the forecasting method based on the second order FTS. The performance of the forecasting results were then evaluated using the similarity measures proposed by [12-18]. Among all these similarity measures, [17] shows the best performance. When the performance is compared to the first order fuzzy logical relationship of the same data from [19], [17] still shows the best performance. However, the second order does not necessarily produce higher similarity measure compared to the first order fuzzy logical relationship. The similarity measure used can be directly obtained from the fuzzy forecasted value without going through the defuzzification process.

7. REFERENCES

[1] Q. Song and B. S. Chissom (1993). Fuzzy time series and its model. *Fuzzy Sets and Systems, 54*, 269-277.

[2] Q. Song and B. S. Chissom (1993b). Forecasting enrollments with fuzzy time series – Part I. *Fuzzy Sets and Systems, 54*, 1-9.

[3] S. M. Chen (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems, 81*, 311 - 319.

[4] K. Huarng (2001). Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets and Systems, 123*, 369-386.

[5] S. M. Chen and C. C. Hsu (2004). A New Method to Forecast Enrollments Using Fuzzy Times Series. *International Journal of Applied Science and Engineering, 2*(3), 234-244.

[6] S. M. Chen (2002). Forecasting enrollments based on high-order fuzzy time series. *Cybernetics and Systems, 33*(1), 1-16.

[7] H. T. Liu (2007). An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optimization and Decision Making, 6*(1), 63-80.

[8] H. T. Liu (2009). An integrated fuzzy time series forecasting system. *Expert Systems with Applications, 36*(6), 10045-10053.

[9] S. Rajaram and V. Vamitha (2012). A Modified Approach on Fuzzy Time Series Forecasting. *Annals of Pure and Applied Mathematics, 2*(1), 96-106.

[10] D. K. Yadav, S. K. Chaturvedi and R. B. Misra (2012). Forecasting Time-Between-Failures of Software using Fuzzy Time Series Approach. *Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS),* IEEE, 1-8.

[11] A. Syigit, C. Ulu and M. Guzelkaya (2014). A new fuzzy time series model using triangular and trapeazoidal membership functions. *International Work-conference on Time Series (ITISE2014)*, 634-644.

[12] S. Chen and S. Chen (2003). Fuzzy Risk Analysis Based on Similarity Measures of Generalized Fuzzy Numbers. *IEEE Transactions on Fuzzy Systems, 11*(1), 45-56.

[13] S. H. Wei and S. M. Chen (2009). A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Expert Systems with Applications, 36*(1), 589-598.

[14] Z. Xu, S. Shang, W. Qian and W. Shu (2010). A method for fuzzy risk analysis based on the new similarity of trapezoidal fuzzy numbers. *Expert Systems with Applications, 37*(3), 1920-1927.

[15] K. Patra and S. K. Mondal (2015). Fuzzy risk analysis using area and height based similarity measure on generalized trapezoidal fuzzy numbers and its application. *Applied Soft Computing, 28*, 276-284.

[16] J. Li and W. Zeng (2017). Similarity measures of generalized trapezoidal fuzzy numbers for fault diagnosis. *Soft Computing, 23*(6), 1999-2014.

[17] H. A. Khorshidi and S. Nikfalazar (2017). An improved similarity measure for generalized fuzzy numbers and its application to fuzzy risk analysis. *Applied Soft Computing, 52*, 478 - 486.

[18] R. Chutia and M. K. Gogoi (2018). Fuzzy risk analysis in poultry farming using a new similarity measure on generalized fuzzy numbers. *Computer & Industrial Engineering, 115*, 543-558.

[19] N. M. F. Alam and N. Ramli (2019). Fuzzy Time Series Forecasting Model based on Various Types of Similarity Measure Approach. *Gading Journal of Science & Technology, 2*(2), 17-25.