Journal of Computer Science & Computational Mathematics, Volume 8, Issue 4, December 2018
DOI: 10.20967/jcscm.2018.04.010

Fuzzy Time Series Forecasting Model based on Centre of Gravity Similarity Measure

Nazirah Ramli[1*], Siti Musleha Ab Mutalib[2], Daud Mohamad[2]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Pahang, 26400 Bandar Jengka, Pahang, Malaysia
[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[*]Corresponding author email: nazirahr@pahang.uitm.edu.my

Abstract: This paper proposes a new method for measuring fuzzy forecasting accuracy (FFA) based on centre of gravity (COG) similarity measure approach. Fuzzy time series (FTS) data represented in trapezoidal fuzzy numbers (TrFNs) form, average based length partitioning method, and first order fuzzy logical relation are used in developing the FTS forecasting model. The COG similarity measure is calculated between the fuzzified historical data and fuzzy forecasted values. The distance of COG similarity measure represents the error of the forecasting model which is the uniqueness of the FFA method. The proposed forecasting model is applied in a numerical example of unemployment rate with the forecasting error of 0.0241 obtained. The new FFA can be directly obtained from the fuzzy forecasted values without going through the defuzzification process as compared to other fuzzy forecasting models. The historical data and forecasted values remained in the TrFNs form and, thus, this proposed forecasting model preserved the information that has been kept during the forecasting procedure from being lost. The proposed model can be applied in other time series data such as forecasts on finance, tourism and weather.

Keywords: centre of gravity similarity measure, defuzzification, fuzzy time series forecasting model, trapezoidal fuzzy numbers, unemployment rate.

1. Introduction

Song and Chissom [1] have proposed the fuzzy time series (FTS) forecasting model to overcome the drawback in the classical time series methods. They [1] used discrete fuzzy set in representing the time series data and produced the forecasted values in terms of discrete fuzzy sets. Many improvements and modifications have been made [1] such as alterations on the interval length [2, 3], types of fuzzy logical relation (FLR) [4] and types of defuzzification [5].

Instead of using discrete fuzzy set, trapezoidal fuzzy numbers (TrFNs) [6, 7] were used to represent the linguistic terms of the time series data and produced the forecasted values in TrFNs form. The forecasted values in some studies [1-7] were defuzzified to crisp values and the forecasting accuracy such as mean square error (MSE) and root mean square error (RMSE) were calculated. During the defuzzification process, some information that has been kept had dissipated from the data.

This paper proposes a new fuzzy forecasting accuracy (FFA) approach using centre of gravity (COG) similarity measure concept. This paper is organized as follows: in Section 2, the basic definition on FTS and fuzzy similarity measure are introduced; Section 3 presents the proposed FTS forecasting model based on COG similarity measure; Section 4 illustrates the proposed method using the data of unemployment rate in Malaysia. Discussions and conclusion are presented in Sections 5 and 6 respectively.

2. Preliminaries

In this section, FTS and fuzzy similarity measure are briefly reviewed.

Definition 1. Let $X(t)$ ($t = \ldots, 0,1,2,\ldots$) be subset of $\Re$ and $X(t)$ the universe of discourse defined by fuzzy set $u_i(t)$ ($i = 1, 2, \ldots$). If $F(t)$ consists of $u_i(t)$ (i=1, 2, …), then $F(t)$ is called fuzzy time series on $X(t)$ ($t = \ldots, 0,1,2,\ldots$) [1].

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

Definition 3. Let $F(t–1) = A_i$ and $F(t) = A_j$. The fuzzy logical relationship (FLR) between $F(t)$ and $F(t–1)$ can be denoted by $A_i\rightarrow A_j$, whereby A_i and A_j are the left-hand side and right-hand side of the FLR respectively[6].

Definition 4. FLR can be further grouped together into FLR group based on the same left-hand side of FLR. For n FLR with left-hand side A_i, $A_i\rightarrow A_{j1}$, $A_i\rightarrow A_{j2}$, …, $A_i\rightarrow A_{jn}$, then the FLR can be grouped as $A_i\rightarrow A_{j1}, A_{j2}, \ldots, A_{jn}$[8].

Definition 5. Let $A = (a_1, a_2, a_3, a_4; h_A)$ and $B = (b_1, b_2, b_3, b_4; h_B)$ be two generalized trapezoidal fuzzy numbers. The degree of similarity measure between two fuzzy numbers denoted by $S(A,B)$ is defined as[9]:

$$S(A,B)=1-\frac{1}{8}\sum_{i=1}^{4}\left|a_i-b_i\right|-\frac{d(A,B)}{2} \quad \text{whereby}$$

$$d(A,B)=\frac{\sqrt{\left(x_A^*-x_B^*\right)^2+\left(y_A^*-y_B^*\right)^2}}{\sqrt{1.25}},$$

$$y_A^*=\begin{cases}\dfrac{h_A\left(\dfrac{a_3-a_2}{a_4-a_1}+2\right)}{6} & \text{if } a_4\neq a_1\\[2ex] \dfrac{h_A}{2} & \text{if } a_4=a_1\end{cases},$$

$$x_A^*=\begin{cases}\dfrac{y_A\left(a_3-a_2\right)+\left(a_4-a_1\right)\left(h_A-y_A^*\right)}{2h_A} & \text{if } h_A\neq 0\\[2ex] \dfrac{a_4+a_1}{2} & \text{if } h_A=0\end{cases},$$

$$y_B^* = \begin{cases} \dfrac{h_B\left(\dfrac{b_3-b_2}{b_4-b_1}+2\right)}{6} & \text{if } b_4 \neq b_1 \\ \dfrac{h_B}{2} & \text{if } b_4 = b_1 \end{cases},$$

$$x_B^* = \begin{cases} \dfrac{y_B(b_3-b_2)+(b_4-b_1)(h_B-y_B^*)}{2h_B} & \text{if } h_B \neq 0 \\ \dfrac{b_4+b_1}{2} & \text{if } h_B = 0 \end{cases}.$$

3. Proposed Model

In this section, the proposed FTS forecasting model based on COG similarity measure is presented. The proposed FTS forecasting model is given as follows:

Step 1: Collect the historical data D_t. Determine the minimum data D_{min} and the maximum data D_{max}.

Step 2: Determine the universe of discourse UD which is defined as $UD = [D_{min} - k_1, D_{max} + k_2]$ whereby k_1 and k_2 are two appropriate positive real numbers.

Step 3: Determine the appropriate length l and partition the universe of discourse, UD as follows:

i. Calculate the absolute difference between the D_{t+1} and D_t whereby t = 1, 2, .. $n - 1$ and n is the number of observation. Calculate the average of the difference between two consecutive data.
ii. Take half of the average in (i) as the length L.
iii. Determine the range and base for length L and round L to l using the base mapping table in Table 1.
iv. Calculate the number of interval using $m = \dfrac{D_{max} + D_2 - D_{min} + D_1}{l}$.
v. Partition UD based on the interval obtained in (iv) as $u_1=[d_1,d_2]$, $u_2=[d_2,d_3]$ … , $u_{m-1}=[d_{m-1}, d_m]$, $u_m=[d_m, d_{m+1}]$.

Step 4: Develop the TrFNs based on the intervals obtained in Step 3(v) to represent the linguistic term as follows:

$A_1 = (d_0, d_1, d_2, d_3)$, $A_2 = (d_1, d_2, d_3, d_4)$, …,
$A_{m-1} = (d_{m-2}, d_{m-1}, d_m, d_{m+1})$,
$A_m = (d_{m-1}, d_m, d_{m+1}, d_{m+2})$.

Step 5: Fuzzify the historical data D_t. If the value of historical data is located in the range of u_m, then it belongs to TrFNs A_m.

Step 6: Establish the FLR between the fuzzified values from Definition 3 and develop the FLR group from Definition 4.

Step 7: Calculate the fuzzy forecasted value F_t based on the heuristic rules from [10].

Step 8: Normalize the fuzzified historical data and fuzzy forecasted value.

Step 9: Calculate the COG similarity measure [9] between the normalized fuzzy data in Step 8.

Step 10: Calculate the average degree of similarity as $avg\,S(A,B) = \dfrac{\sum_{i=1}^{n} S(A,B)}{n}$. Then, determine the error based on the distance as $D_{A,B}$ = 1- $avg\,S(A,B)$.

Table 1. Base mapping table [2]

Range	Base
0.1 – 1.0	0.1
1.0– 10.0	1
11.0 – 100.0	10
101.0 – 1000.0	100

4. Numerical Example

The proposed FTS forecasting model is illustrated using the data of unemployment rate in Malaysia from the year 1982 to 2013 (shown in Figure 1).

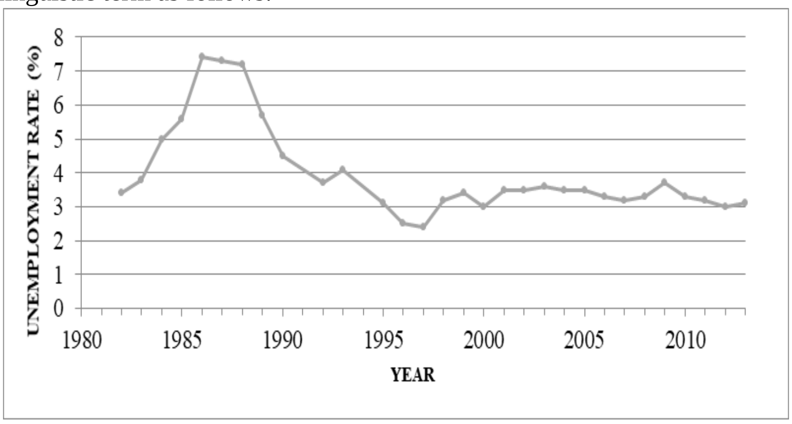

Figure 1. Time series data of unemployment rate in Malaysia from 1982 to 2013

Step 1: The minimum (D_{min}) and maximum (D_{max}) rates of unemployment are 2.4% and 7.4% respectively.

Step 2: By choosing two appropriate numbers as k_1 = 0.4 and k_2 = 0.6, we obtain UD as UD = [2.0, 8.0].

Step 3: By using the average based length, 30 intervals are obtained with length interval 0.2. The intervals obtained are as follows:

u_1=[2.0, 2.2], u_2=[2.2, 2.4], u_3=[2.4,2.6], …, u_{28}=[7.4,7.6],

u_{29}=[7.6,7.8], u_{30}=[7.8,8.0].

Step 4: The linguistic terms of unemployment rate in TrFNs form are given as follows:

A_1=(1.8, 2.0, 2.2, 2.4), A_2=(2.0, 2.2, 2.4, 2.6), … ,

A_{29}=(7.2, 7.4, 7.6, 7.8), A_{30}=(7.4, 7.6, 7.8, 8.0).

Step 5: The fuzzified unemployment rates in Malaysia for the year 2006 to 2010 are shown in Table 2.

Table 2. Fuzzified unemployment rates in TrFNs form

Year	Unemployment rate	TrFNs
2006	3.3	A_7
2007	3.2	A_6
2008	3.3	A_7
2009	3.7	A_9
2010	3.3	A_7

Step 6: The FLR group is presented in Table 3.

Table 3. FLR groups of the unemployment

Group	FLR	Group	FLR
1	$A_2 \to A_6$	9	$A_{13} \to A_9$, $A_{13} \to A_{11}$
2	$A_3 \to A_2$	10	$A_{15} \to A_{18}$
3	$A_5 \to A_6$, $A_5 \to A_8$	11	$A_{18} \to A_{27}$
4	$A_6 \to A_3$,$A_6 \to A_5$, $A_6 \to A_7$	12	$A_{19} \to A_{13}$
5	$A_7 \to A_5$,$A_7 \to A_6$, $A_7 \to A_9$	13	$A_{26} \to A_{19}$
6	$A_8 \to A_6$,$A_8 \to A_7$, $A_8 \to A_8$	14	$A_{27} \to A_{26}$, $A_{27} \to A_{27}$
7	$A_9 \to A_7$,$A_9 \to A_{11}$, $A_9 \to A_{15}$	15	$A_6 \to \phi$
8	$A_{11} \to A_8$, $A_{11} \to A_9$		

Step 7: Based on the heuristic rules in a study [10], the forecasted value of F_t is calculated. The forecasted values of the unemployment rates from the year 2001 until 2010 are shown in Table 4.

Table 4. Fuzzy forecasted unemployment rates for year 2001 until 2010

Year	Fuzzy historical data	Fuzzy forecasted
2001	(3.2, 3.4, 3.6, 3.8)	(3.0, 3.2, 3.4, 3.6)
2002	(3.2, 3.4, 3.6, 3.8)	(3.0, 3.2, 3.4, 3.6)
2003	(3.2, 3.4, 3.6, 3.8)	(3.0, 3.2, 3.4, 3.6)
2004	(3.2, 3.4, 3.6, 3.8)	(3.0, 3.2, 3.4, 3.6)
2005	(3.2, 3.4, 3.6, 3.8)	(3.0, 3.2, 3.4, 3.6)
2006	(3.0, 3.2, 3.4, 3.6)	(3.0, 3.2, 3.4, 3.6)
2007	(2.8, 3.0, 3.2, 3.4)	(2.93, 3.13, 3.33, 3.53)
2008	(3.0, 3.2, 3.4, 3.6)	(2.6, 2.8, 3.0, 3.2)
2009	(3.4, 3.6, 3.8, 4.0)	(2.93, 3.13, 3.33, 3.53)
2010	(3.0, 3.2, 3.4, 3.6)	(3.8, 4.0, 4.2, 4.4)

Step 8: To normalize the fuzzified historical data A_t and fuzzy forecasted value F_t, A_t and F_t are divided by Cheng *et al.* [10].

Step 9: The COG similarity measure [9] is calculated between the normalized A_t and normalized F_t and shown in Table 5.

Step 10: The average degree of similarity is calculated and produces the *avg* $S(A,B)$ = 0.9759 and error based on distance as $D_{A,B}$ = 0.0241.

Table 5. The COG similarity measure (Year 2001- 2010)

Year	Normalized fuzzy historical data	Normalized fuzzy forecasted	COG similarity measure
2001	(0.32, 0.34, 0.36, 0.38)	(0.30, 0.32, 0.34, 0.36)	0.9756
2002	(0.32, 0.34, 0.36, 0.38)	(0.30, 0.32, 0.34, 0.36)	0.9756
2003	(0.32, 0.34, 0.36, 0.38)	(0.30, 0.32, 0.34, 0.36)	0.9756
2004	(0.32, 0.34, 0.36, 0.38)	(0.30, 0.32, 0.34, 0.36)	0.9756
2005	(0.32, 0.34, 0.36, 0.38)	(0.30, 0.32, 0.34, 0.36)	0.9756
2006	(0.30, 0.32, 0.34, 0.36)	(0.30,0.32, 0.34, 0.36)	1
2007	(0.28, 0.30, 0.32, 0.34)	(0.293, 0.313, 0.333, 0.353)	0.9841
2008	(0.30, 0.32, 0.34, 0.36)	(0.26, 0.28, 0.30, 0.32)	0.9512
2009	(0.34, 0.36, 0.38, 0.40)	(0.293, 0.313, 0.333, 0.353)	0.9426
2010	(0.30, 0.32, 0.34, 0.36)	(0.38, 0.40, 0.42, 0.44)	0.9024

5. Discussion and Conclusion

In this paper, the similarity measure between the fuzzy historical data and fuzzy forecasted values were calculated. The distance of the average similarity measure represents the error between the fuzzy historical data and fuzzy forecasted values. The lower the distance, the lower the error and the better the accuracy of the model will be. Table 6 shows a comparison of FFA with methods by Ramli *et al.* [11]. The RMSE of the proposed model is also calculated and the result shows that it is better than the two methods by Ramli *et al*. [11]. The fuzzy forecasted values in Ramli *et al.* [11] were defuzzified to crisp values and, thus, the COG similarity measure cannot be obtained.

Table 6. Comparison of FFA

Method	RMSE	COG similarity measure
[11] with first order FLR	0.4470	*
[11] with second order FLR	0.3603	*
Proposed model	0.2995	0.0241

* cannot be obtained

The proposed FTS forecasting model used the COG similarity measure in determining the forecasting accuracy. No defuzzification process is involved and, thus, the proposed model preserves the information that has been kept during the forecasting procedure from being lost. The forecasting accuracy based on COG similarity measure can be directly obtained from the forecasted values without going through the defuzzification process which is the benefit of the proposed forecasting model.

Acknowledgement

This research is supported by the Ministry of Education Malaysia (MOE) and Universiti Teknologi MARA (UiTM) under the Academic & Research Assimilation 0092/2016.

References

[1] Q. Song, B. S. Chissom, “Forecasting enrollments with fuzzy time series-Part I”, *Fuzzy Sets and Systems*, vol. 54, no. 1, pp. 1-9, 1993.

[2] K. Huarng, “Effective length of intervals to improve forecasting in fuzzy time series”, *Fuzzy Sets and Systems*, vol. 123, no. 3, pp. 387-394, 2001.

[3] R. C. Tsaur, T. C. Kuo, “The adaptive fuzzy time series model with an application to Taiwan’s tourism”, *Expert Systems with Applications*, vol. 38, no. 8, pp. 9164-9171, 2011.

[4] S. M. Chen, “Forecasting enrollments based on high-order fuzzy time series”, *Cybernetics and Systems*, vol. 33, no. 1, pp. 1-16, 2002.

[5] N. Ramli, A. O. M. Tap, “T-norm of Yager class of subsethood defuzzification: Improving enrolment forecast in fuzzy time series”, *Gading Business and Management Journal*, vol.10, no. 2, pp. 1-11, 2006.

[6] H. T. Liu, “An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers”, *Fuzzy Optimization and Decision Making*, vol. 6, no. 1, pp. 63-80, 2007.

[7] H. T. Liu, “An integrated fuzzy time series forecasting system”, *Expert Systems with Applications*, vol. 36, no. 6, pp. 10045-10053, 2009.

[8] F. Radmehr, N. S. Gharneh, “Forecasting based on higher order fuzzy time series and simulated annealing technique”, *South African Journal of Industrial Engineering*, vol. 23, no. 2, pp. 176-190, 2012.

[9] Z. Xu, S. Shang, W. Qian, W. Shu, “A method for fuzzy risk analysis based on the new similarity of trapezoidal fuzzy numbers”, *Expert Systems with Applications*, vol. 31, no. 3, pp. 1920-1927, 2010.

[10] C. Cheng, J. Wang, C. Li, “Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix”, *Expert Systems with Applications*, vol. 34, no. 4, pp. 2568-2575, 2008.

[11] N. Ramli, S. M. A. Mutalib, D. Mohamad, “Fuzzy time series forecasting model with natural partitioning length approach for predicting the unemployment rate under different degree of confidence”, *AIP Conference Proceedings*, vol. 1870, no. 1, p. 040026 , 2017.