International Journal of Engineering & Technology, 7 (4.30) (2018) 281-284

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Fuzzy Time Series Forecasting Model based on Frequency Density and Similarity Measure Approach

Nazirah Ramli[1]*, Siti Musleha Ab Mutalib[2], Daud Mohamad[3]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Pahang, 26400, Bandar Jengka, Pahang, Malaysia
[2,3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, 40450, Shah Alam, Selangor, Malaysia
**Corresponding author E-mail: nazirahr@pahang.uitm.edu.my*

Abstract

This paper proposes an enhanced fuzzy time series (FTS) prediction model that can keep some information under a various level of confidence throughout the forecasting procedure. The forecasting accuracy is developed based on the similarity between the fuzzified historical data and the fuzzy forecast values. No defuzzification process involves in the proposed method. The frequency density method is used to partition the interval, and the area and height type of similarity measure is utilized to get the forecasting accuracy. The proposed model is applied in a numerical example of the unemployment rate in Malaysia. The results show that on average 96.9% of the forecast values are similar to the historical data. The forecasting error based on the distance of the similarity measure is 0.031. The forecasting accuracy can be obtained directly from the forecast values of trapezoidal fuzzy numbers form without experiencing the defuzzification procedure.

Keywords: *Area and Height Similarity Measure; Forecasting Accuracy; Frequency Density; Fuzzy Time Series; Unemployment Rate.*

1. Introduction

To overcome the drawback in the classical time series method, [1] proposed the fuzzy time series (FTS) prediction model. The discrete fuzzy set was used to represent the time series data, and the forecast value in terms of discrete fuzzy set was produced. A large number of studies have been carried out to improve the procedure of FTS in [1] such as by [2-4]. [2] proposed a model to improve the length of the interval by utilizing a new method and [3] proposed the FTS forecasting model which can deal with seasonal time series data. In another study, [4] proposed a higher order-forecasting model based on automatic grouping strategy and generalized FLR.

[5-6] used trapezoidal fuzzy numbers (TrFNs) to denote the linguistic term of the data, and produced the forecast values of TrFNs form. [1-6] defuzzified the forecast values to crisp values, and the forecasting accuracy such as mean absolute percentage error (MAPE), mean square error (MSE), and root mean square error (RMSE) was calculated. The defuzzification procedure produces the forecast values of single point form, and thus some information under a various level of confidence that kept throughout the forecasting procedure has been dissipated from the data. Figure 1 shows the process for obtaining the forecasting accuracy for [1-6].

This paper proposes an improved fuzzy forecasting model based on frequency density [7], and area and height similarity measure [8]. The frequency density partitioning method redefines the intervals based on the frequency of data onto each interval. This partitioning method reflects the distribution of data, and discards the interval of no distribution of data. The similarity measure concept, which portrays the level of likeness between two comparing objects, is generally utilized as a part of numerous applications, for example, pattern recognition, decision-making, and machine learning. In this study, the similarity measure is applied to compare the similarity between the forecast values and historical values. The forecasting accuracy of this FTS model is based totally on the degree of similarity between the forecast values and historical values.

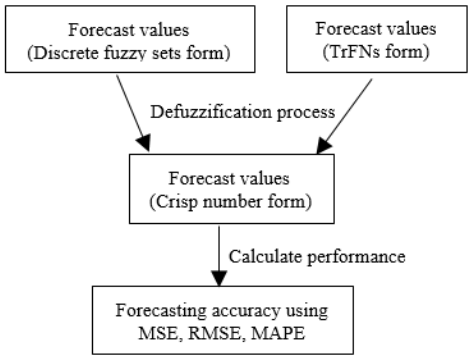

Fig. 1: The process for obtaining the forecasting accuracy from the previous methods [1-6]

This paper is organized as follows: The basic definition of FTS, TrFNs, and fuzzy similarity measure are presented in section two. Section three presents the proposed FTS model based on frequency density and similarity measure; section four illustrates the proposed technique by using the data of unemployment rate in Malaysia. The conclusion is presented in section five.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2. Preliminaries

This section briefly reviewed some fundamental concepts on FTS [1, 5, 9], TrFNs [10] and the similarity measure of area and height type [8].

Definition 1:

Let Y(t) (t = …, 0,1,2,…) be a subset of $\Re$ and Y(t) be the universe of discourse defined by fuzzy set $\mu_i(t)(i = 1,2,…)$, then H(t) is named as FTS on Y(t) (t = …, 0,1,2,…), [9].

Definition 2:

Let H(t) is an FTS. H(t) is produced from H(t−1) if there exists a fuzzy relationship Q(t−1, t) such that H(t) = H(t−1) ⊗ Q(t−1, t) whereby ⊗ denotes as a fuzzy operator. The relationship can be denoted as H(t−1)→ H(t), [9].

Definition 3:

Suppose that $H(t-1) = C_i$ and $H(t) = C_j$. The fuzzy logical relationship (FLR) can be defined as $C_i \rightarrow C_j$ where C_i is the left-hand side and C_j is the right-hand side of FLR. If the FLR on the left-hand side has the same fuzzy set, then the FLR can be further classified into the same FLR group, [5].

Definition 4:

A trapezoidal fuzzy number (TrFN) denoted as P = (p,q,r,s) is defined by the membership function as follows, [10]:

$$\mu_P(x) = \begin{cases} 0, & x < p \\ \dfrac{x-p}{q-p}, & p \le x \le q \\ 1, & q \le x \le r \\ \dfrac{s-x}{s-r}, & r \le x \le s \\ 0, & x > s \end{cases}$$

Definition 5:

Let $M = (m_1,m_2,m_3,m_4;h_M)$ and $N = (n_1,n_2,n_3,n_4;h_N)$ be two generalized TrFNs. The degree of similarity between M and N is denoted by S(M, N) and defined as [8],

$$S(M,N) = \left(1 - \frac{1}{4}\sum_{i=1}^{4}|m_i - n_i|\right) \times \left(1 - \frac{1}{2}\left\{|ar(M) - ar(N)| + |h_M - h_N|\right\}\right)$$

whereby ar(M) is the area of TrFN M defines as $ar(M) = \dfrac{(m_4 + m_3 - m_2 - m_1) \times h_M}{2}$.

3. Proposed Fuzzy Time Series Forecasting Model

This section presents the proposed FTS forecasting model that consists of 10 steps. Steps 1 to 9 is the development of the model for producing the forecast values of TrFNs form. Step 10 is the process to calculate the forecasting accuracy (as shown in Figure 2).

Step 1: Collect the historical data D_t and determine the minimum and maximum data denoted by D_{min} and D_{max} respectively.

Step 2: The universe of discourse is defined as

$V = [D_{min} - a_1, D_{max} + a_2]$

whereby a_1 and a_2 are two appropriate positive real numbers.

Step 3: By using the randomly chosen length method, divide the universe of discourse V into m equal length intervals $v_1, v_2, v_3, \ldots, v_m$.

Step 4: Count the frequency of the historical data included within the interval. Classify and sub-partition the interval based on their frequency density. The interval with the highest frequency is classified as Class 1 and is divided into four sub-partition. Table 1 shows the detailed classification and sub-partition of the interval.

Table 1: Partition of Sub-interval

Level of Frequency Numbers of Interval	Class	Number of Sub-interval
Highest	1	4
Second highest	2	3
Third highest	3	2
Fourth highest and above	4	1

Step 5: Based on Table 1, list all the new sub-intervals $w_1, w_2, w_3, \ldots, w_k$.

Step 6: Based on the new sub-intervals obtained in **Step 5**, establish the new TrFNs as follows:

$C_1 = (d_0,d_1,d_2,d_3)$, $C_2 = (d_1,d_2,d_3,d_4)$, …, $C_{k-1} = (d_{k-2},d_{k-1}, d_k, d_{k+1})$, $C_k = (d_{k-1}, d_k, d_{k+1}, d_{k+2})$.

Step 7: Transform the historical data D_t to TrFNs form. If the value of the historical data is located in the range of w_k, then it belongs to TrFN C_k.

Step 8: Establish the FLR and FLR group based on Definition 3.

Step 9: Based on the heuristic rules from [11], calculate the fuzzy forecasted value H_t in the form of TrFNs. Normalize the H_t and D_t.

Step 10: Calculate the similarity of H_t and D_t by using area and height similarity measure approach from Definition 5 [8]. Then, calculate the forecasting error which is defined as

$D = 1 - avgS(M, N)$ whereby $avgS(M, N)$ is the average degree of similarity between TrFNs M and N.

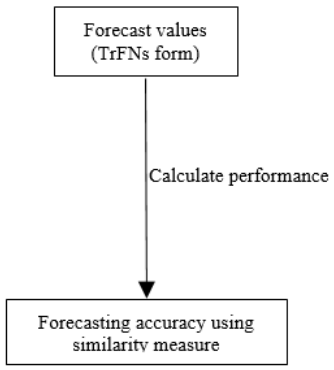

Fig. 2: The process for obtaining the forecasting accuracy for the proposed method

4. Numerical Example

The proposed FTS forecasting model is illustrated using the data of the unemployment rate in Malaysia from the year 1982 to 2013 [12] (shown in Figure 3).

Step 1: Based on the unemployment rate data from [12], D_{min} = 2.4% and D_{max} = 7.4%.

Step 2: By choosing two appropriate numbers as a_1 = 0.4 and a_2 = 0.6, the universe of discourse is defined as V = [2.0, 8.0].

Step 3: By choosing at random the interval length as 0.75, the universe of discourse V is divided into eight equal length as follows: v_1=[2,2.75], v_2=[2.75,3.5], v_3=[3.5,4.25], v_4=[4.25,5], v_5=[5,5.75], v_6=[5.75,6.5], v_7=[6.5,7.25], v_8=[7.25,8].

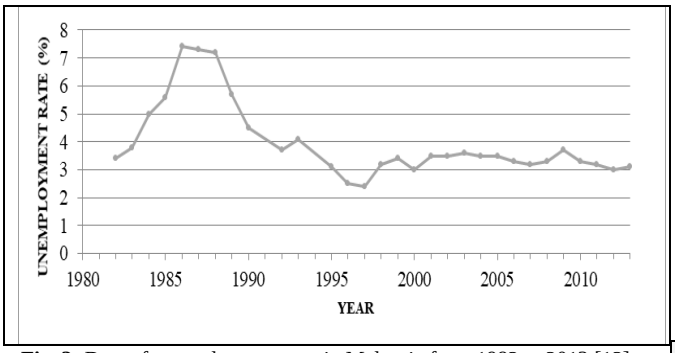

Fig. 3: Data of unemployment rate in Malaysia from 1982 to 2013 [12]

Step 4: The frequency, classification, and sub-interval are shown in Table 2.

Table 2: Classification and sub-interval of unemployment rate data

Interval	Frequency	Class	Number of Sub-interval
v_1=[2,2.75]	2	3	2
v_2=[2.75,3.5]	16	1	4
v_3=[3.5,4.25]	7	2	3
v_4=[4.25,5]	2	3	2
v_5=[5,5.75]	2	3	2
v_6=[5.75,6.5]	0	4	1
v_7=[6.5,7.25]	1	4	1
v_8=[7.25,8]	2	3	2

Step 5: Based on the number of sub-interval in Table 2, there are 17 new sub-intervals obtained which are w_1=[2,2.375], w_2=[2.375,2.75], w_3=[2.75, 2.938], w_4=[2.938,3.125], w_5=[3.125,3.313], w_6= [3.313,3.5], w_7=[3.5,3.75], w_8=[3.75,4], w_9=[4,4.25], w_{10}=[4.25,4.625], w_{11}=[4.625,5], w_{12}=[5,5.375], w_{13}=[5.375,5.75], w_{14}=[5.75,6.5], w_{15}=[6.5,7.25], w_{16}=[7.25,7.625], w_{17}=[7.625,8].

Step 6: The linguistic term of unemployment rate in TrFNs form are given as follows:

C_1=(1.625, 2, 2.375, 2.75), C_2=(2, 2.375, 2.75, 2.938), … , C_{16}=(6.5, 7.25, 7.625, 8), C_{17}=(7.25, 7.625, 8, 8.375).

Step 7: Table 3 shows the fuzzified unemployment rate in Malaysia for the year 2008 to 2013.

Step 8: Table 4 presents the FLR group of the unemployment rate.

Table 3: Fuzzified unemployment rate in TrFNs form

Year		Unemployment rate	TrFNs
2008		3.3	A_5
2009		3.7	A_7
2010		3.3	A_5
2011		3.2	A_5
2012		3	A_4
2013		3.1	A_4

Table 4: FLR group of unemployment rate

Group	FLR	Group	FLR
	$C_2 \rightarrow C_2$, $C_2 \rightarrow C_5$	6	$C_8 \rightarrow C_{11}$
2	$C_4 \rightarrow C_2$, $C_4 \rightarrow C_2$,	7	$C_9 \rightarrow C_7$
	$C_4 \rightarrow C_2$	8	$C_{10} \rightarrow C_7$, $C_{10} \rightarrow C_9$
3	$C_5 \rightarrow C_4$, $C_5 \rightarrow C_5$,	9	$C_{11} \rightarrow C_{13}$
	$C_5 \rightarrow C_6$, $C_5 \rightarrow C_7$	10	$C_{13} \rightarrow C_{10}$, $C_{13} \rightarrow C_{16}$
4	$C_6 \rightarrow C_4$, $C_6 \rightarrow C_5$,	11	$C_{15} \rightarrow C_{13}$
	$C_6 \rightarrow C_6$, $C_6 \rightarrow C_5$,	12	$C_{16} \rightarrow C_{15}$, $C_{16} \rightarrow C_{16}$
	$C_6 \rightarrow C_8$	13	$C_4 \rightarrow \phi$
5	$C_7 \rightarrow C_4$, $C_7 \rightarrow C_5$,		
	$C_7 \rightarrow C_6$, $C_7 \rightarrow C_9$		

Step 9: Based on the heuristic rules from [11], the fuzzy forecast value H_t is calculated. The values of H_t for the year 2008 to 2013 are shown in Table 5. To normalize the historical data C_t and forecast H_t, C_t and H_t are divided by 10.

Table 5: Fuzzy forecasted unemployment rate for year 2008 until 2013

Year	Fuzzy historical data	Fuzzy forecasted
2008	(2.938, 3.125, 3.313, 3.5)	(3.031, 3.219, 3.422, 3.641)
2009	(3.313, 3.5, 3.75, 4)	(3.031, 3.219, 3.422, 3.641)
2010	(2.938, 3.125, 3.313, 3.5)	(3.141, 3.344, 3.547, 3.797)
2011	(2.938, 3.125, 3.313, 3.5)	(3.031, 3.219, 3.422, 3.641)
2012	(2.75,2.938,3.125, 3.313)	(3.031, 3.219, 3.422, 3.641)
2013	(2.75,2.938,3.125, 3.313)	(2.625, 2.875, 3.125, 3.333)

Step 10: Based on the normalized C_t and normalized H_t, the area and height similarity measure [8] is calculated as shown in Table 6.

Table 6: The area and height similarity measure (Year 2008- 2013)

Year	Similarity	Year	Similarity	Year	Similarity
1983	0.953	1994	1	2004	0.995
1984	1	1995	0.955	2005	0.997
1985	1	1996	0.944	2006	0.976
1986	0.851	1997	0.958	2007	0.988
1987	0.963	1998	0.958	2008	0.988
1988	0.963	1999	0.991	2009	0.966
1989	1	2000	0.957	2010	0.974
1990	0.851	2001	0.953	2011	0.988
1991	1	2002	0.997	2012	0.969
1992	1	2003	0.978	2013	0.990
1993	0.924				

Table 6 shows the degree of similarity for the years 1984, 1985, 1989, 1991, 1992 and 1994 are equal to one. It demonstrates that 19.4% of the forecast unemployment rate is precisely similar with the actual value. 100% of the forecast values have more than 85% degree of similarity and on average the forecast values have 96.9% similarity with the actual value. The forecasting error based on the distance of the similarity measure is 0.031 compared to 7.62% for the MAPE value. According to [13], MAPE is the most helpful measurement to investigate the accuracy of forecasts between various elements as it measures relative performance. Based on the scale of forecasting accuracy from [13], the MAPE value indicates that the forecasting model has a good forecast. However, in order to obtain the MAPE values, the forecast values of TrFNs need to be transformed to crisp numbers via defuzzification process.

5. Conclusion

In this paper, the forecasting accuracy is developed based on the area and height type of similarity measure concept. The forecasting accuracy can be obtained directly from the forecast values of TrFNs form without experiencing the defuzzification procedure as compared to most of the previous studies of FTS such as [1-6] (as shown in Figures 1 and 2). The proposed FTS model preserves the forecast values of TrFNs form and thus able to keep some information under various level of confidence from being lost.

Acknowledgement

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

References

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

[2] Yu HK (2005), A refined fuzzy time series model for forecasting. *Physica A*: *Statistical Mechanics and its Application* 346(3-4), 657-681.

[3] Liu HT & Wei ML (2010), An improved fuzzy forecasting method for seasonal time series. *Expert Systems with Applications* 39(9), 6310-6318.

[4] Qiu W, Zhang P & Wang Y (2015), Fuzzy time series forecasting model based on automatic clustering techniques and generalized fuzzy logical relationship. *Mathematical Problems in Engineering*, 1−8.
[5] Liu HT (2007), An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optimization and Decision Making* 6(1), 63-80.
[6] Liu HT (2009), An integrated fuzzy time series forecasting system. *Expert Systems with Applications* 36 (6), 10045-10053.
[7] Hsu CC & Chen SM (2002), A new method for forecasting enrollments based on fuzzy time series. *Proceedings of the Seventh Conference on Artificial Intelligence and Applications*, 17-22.
[8] Patra K & Modal SK (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.
[9] Song Q & Chissom BS (1994), Forecasting enrollments with fuzzy time series – Part II. *Fuzzy Sets and Systems* 62, 1-8.
[10] Wang X (1997), An investigation into relations between some transitivity related concept. *Fuzzy sets and Systems* 89(2), 257-262.
[11] Cheng C, Wang J & Li C (2008), Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix. *Expert Systems with Applications* 34(4), 2568-2575.
[12] Department of Statistic Malaysia. *Time series data of unemployment*. https://www.dosm.gov.my. Accessed January 13, 2014.
[13] Lewis CD, *Industrial and business forecasting methods*, Butterworths, London, (1982).