COMPARISON OF INTERVAL LENGTHS FOR THE INTUITIONISTIC FUZZY TIME SERIES FORECASTING MODEL

Nik Muhammad Farhan Hakim Nik Badrul Alam[1*], Asyura Abd Nassir[1], Ainun Hafizah Mohd[1], Nazirah Ramli[1]

[1]*Faculty of Computer and Mathematical Sciences*
Universiti Teknologi MARA Pahang, 26400 Bandar Jengka, Pahang

[*]*Corresponding author: farhanhakim@uitm.edu.my*

Abstract

A fuzzy time series forecasting model can cater for the time series data described by linguistic terms. The use of fuzzy sets in forecasting time series data is evidently better at predicting data compared to the classical time series model. The fuzzy set concept was extended to the intuitionistic fuzzy set, in which its performance in forecasting time series data is extensively better than the classical fuzzy set. In the intuitionistic fuzzy time series forecasting model, the universe of discourse is defined and divided into several intervals before the data are fuzzified. The objective of this study is to compare the forecasting performance using different interval lengths. The historical data of student enrollments at the University of Alabama were adopted, in which 7, 14, and 21 intervals were used to perform the forecasting process. The results have shown that the model with 21 sub-intervals outperformed the other models. In the future, it is recommended that researchers determine an effective interval length at the early stage of forecasting to obtain the best performance result for time series forecasting.

Keyword: Interval length; intuitionistic fuzzy sets; fuzzy time series; forecasting model

Introduction

Time series data are observations in a time-ordered sequence (Wei, 2013). Many statistical models are developed in analysing and forecasting time series data. The emergence of the fuzzy set theory (Zadeh, 1965) has enabled the forecasting of time series data in linguistic terms.

Fuzzy time series was proposed by Song and Chissom (1993a) using the proposed theory in forecasting student enrollments at the University of Alabama (Song & Chissom, 1993b). In their model, the universe of discourse was divided into seven intervals of the same length. By adopting the same data, Chen (1996) simplified the model by considering the midpoints of intervals in the defuzzification process instead of using the max-min composition operation. Many fuzzy time series forecasting models have since been developed and improved.

Liu (2007), for instance, had used trapezoidal fuzzy numbers to forecast time series data. The use of fuzzy numbers is better than fuzzy sets due to the fact that the single-point values of fuzzy sets mimic the traditional time series method (Liu, 2007). Many other improvements were done from time to time to improve the interval partition and defuzzification methods.

The fuzzy set theory was generalised into an intuitionistic fuzzy set (IFS) by Atanassov (1986), which comprises the membership and non-membership functions that represent the degree of belongingness and non-belongingness to the fuzzy sets, respectively. The IFS has a better capability of handling uncertainty and vagueness in real applications (Joshi & Kumar, 2012).

The knowledge of IFS is applied in the fuzzy time series forecasting model for better forecasting performance. Therefore, many models have been developed by numerous researchers based on the IFS, such as Joshi and Kumar (2012), Kumar and Gangwar (2015), Abhishekh et al. (2018), and Abhishekh et al. (2020). The IFS-based models have also shown better performance compared to the classical fuzzy set (Alam et al., 2021).

The existing models used different interval lengths in the early stage of the forecasting process. For instance, Joshi and Kumar (2012) and Kumar and Gangwar (2015) used a randomly chosen number of intervals of equal length. In different contexts, Abhishekh et al. (2018) applied the statistical distribution method to divide the universe of discourse, while Abhishkeh et al. (2020) used the average-based length method to determine the number of intervals.

In analysing the various types of interval lengths used in developing the forecasting model, the different interval lengths are assumed to affect the forecasting performance. Therefore, this study aims to find the effective length for the intuitionistic fuzzy time series forecasting model by implementing the model using the historical data of student enrollments at the University of Alabama. The forecasting performance using different interval lengths is then compared.

This paper is organised as follows: Section 1 presents the introduction; Section 2 reviews some related mathematical preliminaries; Section 3 presents the methodology; Section 4 illustrates numerical examples; Section 5 presents and discusses results; Section 6 concludes the paper.

Preliminaries

In this section, some mathematical preliminaries related to this study are reviewed. A fuzzy number encompasses a subset of real numbers, in which the definition of a triangular fuzzy number (TFN) is given below.

Definition 2.1: Let A be a TFN. Then, A is characterised by the membership function $\mu_A : x \to [0,1]$ defined by

$$\mu_A(x) = \begin{cases} \dfrac{x-a}{b-a} & , \quad a \le x < b \\ 1 & , \quad x = b \\ \dfrac{c-x}{c-b} & , \quad b < x \le c \\ 0 & , \quad \text{otherwise} \end{cases} \tag{1}$$

The IFS generalises the classical fuzzy set by comprising a non-membership function instead of the membership function alone. The IFS is defined by Atanassov (1986) as follows:

Definition 2.2: An IFS, I on the universe of discourse, U can be written in the form of

$$I = \left\{ \left(x, \mu_I(x), \nu_I(x) \right) \mid x \in U \right\} \tag{2}$$

where $\mu_I(x)$ and $\nu_I(x)$ represent the membership and non-membership functions, respectively.

To compare between two intuitionistic fuzzy numbers (IFN), a score function is used. The score function of an IFN was defined by Xu and Yager (2006) as follows:

Definition 2.3: The score of an IFN, $I = \langle \mu_I, \nu_I \rangle$ is given by

$$S(I) = \mu_I - \nu_I \tag{3}$$

where $S(I) \in [-1,1]$. The higher the value of $S(I)$, the greater the IFN.

The intuitionistic fuzzy time series defined by Bisht et al. (2018) is shown in Definition 2.4.

Definition 2.4: Assume $Y(t), (t = ..., 0, 1, 2, ...)$ is a sequence of data over time and a subset of real numbers. If $I_j(t), (j = 1, 2, ...)$ are IFS defined in $Y(t)$, then the collection of $I_j(t)$ is called intuitionistic fuzzy time series. $I_j(t-1) \rightarrow I_j(t)$ indicates the relationship that $I_j(t)$ is caused by $I_j(t-1)$.

Jurio et al. (2010) proposed a method for converting the classical fuzzy set into an IFS:

Definition 2.5: Let $A_F \in FS(U)$, where $FS(U)$ is a collection of all fuzzy sets in *U*. Let $\alpha, \beta : U \rightarrow [0,1]$, then $f(x,\alpha,\beta) = (f_\mu(x,\alpha,\beta), f_\nu(x,\alpha,\beta))$ where

$$f_\mu(x,\alpha,\beta) = x(1-\alpha\beta) \quad \text{and} \quad f_\nu(x,\alpha,\beta) = 1 - x(1-\alpha\beta) - \alpha\beta. \tag{4}$$

Methodology

This section presents the methodology used in this study. The following entails the seven steps performed in the intuitionistic fuzzy time series forecasting model.

Step 1: The universe of discourse is defined, $U = [D_{\min} - k_1, D_{\max} + k_2]$, where $D_{\min}$ and $D_{\max}$ are the minimum and maximum historical data, respectively, and $k_1, k_2 \in \square$.

Step 2: The universe of discourse is divided into several intervals u_k, for $k = 1, 2, 3 \ldots$.

Step 3: The historical data are fuzzified using the triangular fuzzy number in Equation (1).

Step 4: The fuzzy sets are then converted into intuitionistic fuzzy sets (IFS) using Atanassov's conversion of IFS as shown in Equation (4).

Step 5: The score functions are calculated using Equation (3) and the IFS are defined for the data.

Step 6: The intuitionistic fuzzy logical relationships (IFLR) are defined and grouped.

Step 7: The IFS are then defuzzified. The forecasted data are finally computed using the following rules:

(i) If the IFLR is $I_{j1} \rightarrow I_{j2}$, then the forecasted output is the midpoint of the interval u_{j2}.

(ii) If the IFLR is $I_{j1} \rightarrow I_{j2}, I_{j3}, ..., I_{jn}$, then the forecasted output is the average of the midpoints of intervals $u_{j2}, u_{j3}, ..., u_{jn}$.

(iii) If the IFLR is $I_{j1} \rightarrow \phi$, then the forecasted output is the midpoint of the interval

u_{j1}.

The comparison of interval lengths began by dividing the universe of discourse into three different lengths at Step 2. To illustrate this comparison, numerical examples of student enrollments at the University of Alabama were used.

Numerical Examples
Student enrollments at the University of Alabama from 1971 until 1992 are shown in **Figure 1**.

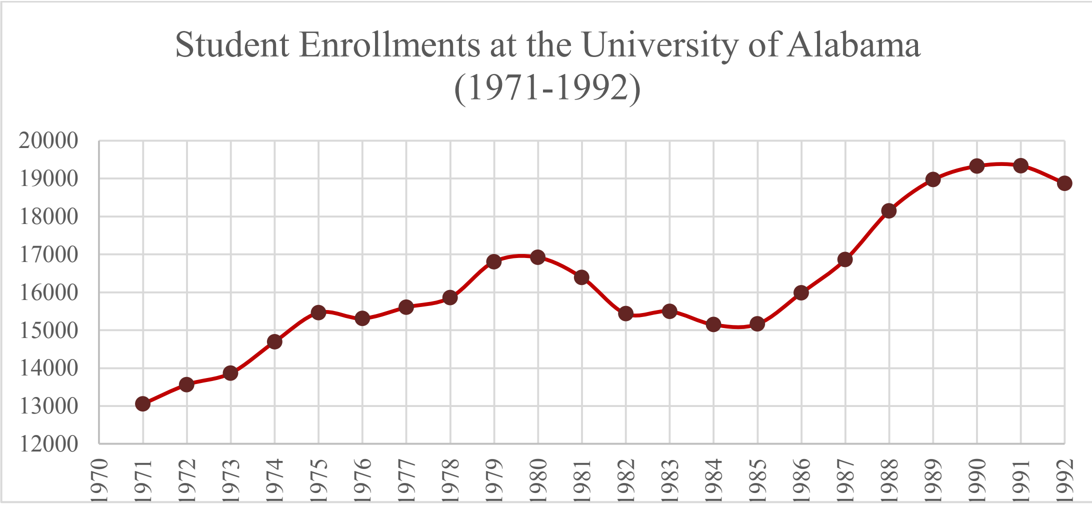

Figure 1 Student enrollments at the University of Alabama (1971-1992)

The minimum and maximum data denote 13055 and 19337, respectively. Hence, $k_1 = 55$ and $k_2 = 663$ were chosen such that the universe of discourse is defined as $U = [13000, 20000]$. Next, U was divided into 7, 14, and 21 intervals, as shown in **Table 1**.

Table 1 Intervals using different lengths

Length	Intervals
Length 1	$u_1 = [13000,14000]$, $u_2 = [14000,15000]$, $u_3 = [15000,16000]$, $u_4 = [16000,17000]$, $u_5 = [17000,18000]$, $u_6 = [18000,19000]$, $u_7 = [19000,20000]$
Length 2	$u_1 = [13000,13500]$, $u_2 = [13500,14000]$, $u_3 = [14000,14500]$, $u_4 = [14500,15000]$, $u_5 = [15000,15500]$, $u_6 = [15500,16000]$, $u_7 = [16000,16500]$, $u_8 = [16500,17000]$, $u_9 = [17000,17500]$, $u_{10} = [17500,18000]$, $u_{11} = [18000,18500]$, $u_{12} = [18500,19000]$, $u_{13} = [19000,19500]$, $u_{14} = [19500,20000]$
Length 3	$u_1 = [13000,13333]$, $u_2 = [13333,13667]$, $u_3 = [13667,14000]$, $u_4 = [14000,14333]$, $u_5 = [14333,14667]$, $u_6 = [14667,15000]$,

$u_7 = [15000,15333]$, $u_8 = [15333,15667]$, $u_9 = [15667,16000]$,
$u_{10} = [16000,16333]$, $u_{11} = [16333,16667]$, $u_{12} = [16667,17000]$,
$u_{13} = [17000,17333]$, $u_{14} = [17333,17667]$, $u_{15} = [17667,18000]$,
$u_{16} = [18000,18333]$, $u_{17} = [18333,18667]$, $u_{17} = [18333,18667]$,
$u_{18} = [18667,19000]$, $u_{19} = [19000,19334]$, $u_{20} = [19334,19667]$,
$u_{21} = [19667,20000]$

The historical data were then fuzzified using triangular fuzzy numbers that are defined based on the intervals in **Table 1**. Next, the fuzzy sets were converted into IFS using Equation (4). Hence, the score functions were calculated to determine the IFS of the data. After the IFS of each datum was defined, intuitionistic fuzzy logical relationships (IFLR) were formed. **Table 2** presents the IFLR groups obtained using different interval lengths.

Table 2 IFLR obtained using different interval lengths

	Length 1	**Length 2**	**Length 3**
IFLR	$I_1 \to I_1, I_2$	$I_1 \to I_1, I_2$	$I_1 \to I_2$
	$I_2 \to I_2, I_3$	$I_2 \to I_3$	$I_2 \to I_3$
	$I_3 \to I_2, I_3, I_4$	$I_3 \to I_5$	$I_3 \to I_5$
	$I_4 \to I_3, I_4, I_5$	$I_4 \to I_4, I_6$	$I_5 \to I_7$
	$I_5 \to I_6$	$I_5 \to I_4, I_5, I_6$	$I_6 \to I_6, I_9$
	$I_6 \to I_7$	$I_6 \to I_8$	$I_7 \to I_6, I_7, I_8$
	$I_6 \to \phi$	$I_7 \to I_5$	$I_8 \to I_9$
		$I_8 \to I_8, I_{10}$	$I_9 \to I_{11}, I_{12}$
		$I_{10} \to I_{12}$	$I_{10} \to I_7$
		$I_{12} \to I_{13}$	$I_{11} \to I_{12}$
		$I_{13} \to I_{12}, I_{13}$	$I_{12} \to I_{10}, I_{15}$
		$I_{12} \to \phi$	$I_{15} \to I_{18}$
			$I_{18} \to I_{19}$
			$I_{19} \to I_{18}, I_{19}$
			$I_{18} \to \phi$

Results and Discussion

The forecasted enrollments using different interval lengths are illustrated in **Figure 2**. The interval lengths 1, 2, and 3 denote the lengths with 7, 14, and 21 intervals, respectively.

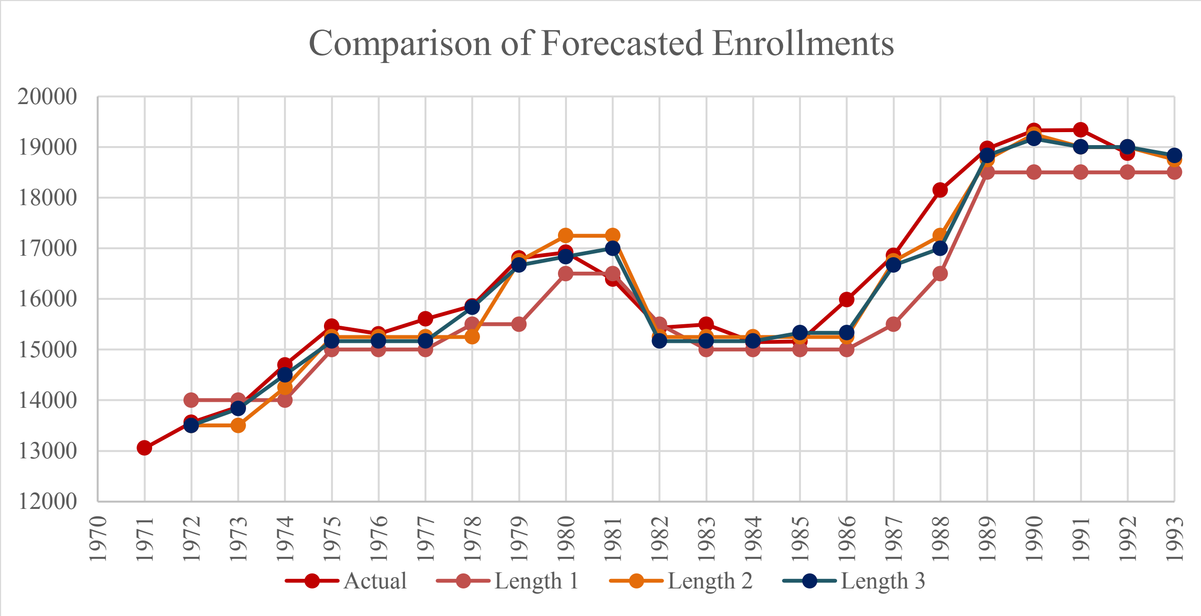

Figure 2 Comparison of forecasted enrollments using different interval lengths

The performance of the intuitionistic fuzzy time series forecasting model using different interval lengths was then analysed by computing the mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) to measure the errors between the actual and forecasted values. The error values are presented in **Table 3**.

Table 3 Comparison of forecasting performance using different interval lengths

	Length 1	Length 2	Length 3
MSE	522706.5	161730.3	138515.9
RMSE	722.9845	402.1571	372.1772
MAE	581.6667	308.8095	265.3413

The forecasting model with 21 intervals (Length 3) outperformed the other models. The number of intervals has affected the number of the IFLR groups, which control the crucial part in the defuzzification process. This fact is supported by the result obtained by Ramli et al. (2020), which compared the forecasting performance of unemployment rate using average-based length, frequency density-based, and randomly chosen length methods. In their finding, the average-based length method with the highest number of intervals among all outperformed the other methods. Lasaraiya et al. (2021) also obtained a similar result when comparing the forecasting performance of tuberculosis cases using the average-based length method (42 intervals), half of the average-based length (21 intervals), and natural partition (20 intervals). In their findings, the average-based length method outperformed the other methods. However, Huarng (2001) states that the effective interval length should not be too large or small.

Conclusion

This paper has investigated the effect of changing the number of intervals on the performance of the intuitionistic fuzzy time series forecasting model. Three interval lengths were used in which the universe of discourse was divided into 7, 14, and 21 intervals. Using MSE, RMSE and MAE, the forecasting performance using different interval lengths was compared. The results showed that the application of 21 intervals has resulted in the lowest error values among all. However, it should be noted that when the interval length is reduced, more fuzzy sets and intuitionistic fuzzy sets will be defined, which makes the forecasting procedure longer. Hence, an effective interval length should be determined at the early stage of forecasting so that time series forecasting can be done optimally with the best performance results.

Acknowledgement
This research is supported financially by Geran Lestari Khas UiTM Cawangan Pahang 600-TNCPI 5/3/DDN (06) (009/2020).

Conflict of interest
All authors declare that there is no conflict of interest.

References
Abhishekh, Gautam, S. S., & Singh, S. R. (2018). A Score function-based method of forecasting using intuitionistic fuzzy time series. *New Mathematics and Natural Computation*, *14*(01), 91-111. https://doi.org/10.1142/S1793005718500072

Abhishekh, Gautam, S. S., & Singh, S. R. (2020). A new method of time series forecasting using intuitionistic fuzzy set based on average-length. *Journal of Industrial and Production Engineering*, *37*(4), 175-185. https://doi.org/10.1080/21681015.2020.1768163

Alam, N. M. F. H. N. B., Ramli, N., & Mohd, A. H. (2021, July). Intuitionistic fuzzy set-based time series forecasting model via delegeration of hesitancy degree to the major grade de-i-fuzzification and arithmetic rules based on centroid defuzzification. In *Journal of Physics: Conference Series* (Vol. 1988, No. 1, p. 012014). IOP Publishing. https://doi.org/10.1088/1742-6596/1988/1/012014

Atanassov, K. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, *20*(1), 87-96. https://doi.org/10.1016/S0165-0114(86)80034-3

Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy sets and systems*, *81*(3), 311-319. https://doi.org/10.1016/0165-0114(95)00220-0

Huarng, K. (2001). Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy sets and systems*, *123*(3), 387-394. https://doi.org/10.1016/S0165-0114(00)00057-9

Joshi, B. P., & Kumar, S. (2012). Intuitionistic fuzzy sets based method for fuzzy time series forecasting. *Cybernetics and Systems*, 43(1), 34-47. https://doi.org/10.1080/01969722.2012.637014

Jurio, A., Paternain, D., Bustince, H., Guerra, C., & Beliakov, G. (2010, July). A construction method of Atanassov's intuitionistic fuzzy sets for image processing. In *2010 5th IEEE International Conference Intelligent Systems* (pp. 337-342). IEEE. https://doi.org/10.1109/IS.2010.5548390

Kumar, S., & Gangwar, S. S. (2015). A fuzzy time series forecasting method induced by intuitionistic fuzzy sets. *International Journal of Modeling, Simulation, and Scientific Computing*, 6(04), 1550041. https://doi.org/10.1142/S1793962315500415

Lasaraiya, S., Zenian, S., & Hasim, R. M. (2021, August). An effective length of intervals for forecasting in fuzzy time series. In *the 9th International Conference on Research and Education in Mathematics (ICREM9)*. Universiti Putra Malaysia, Malaysia.

Liu, H. T. (2007). An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optimization and Decision Making*, *6*(1), 63-80. https://doi.org/10.1007/s10700-006-0025-9

Song, Q., & Chissom, B. S. (1993a). Fuzzy time series and its models. *Fuzzy sets and systems*, *54*(3), 269-277. https://doi.org/10.1016/0165-0114(93)90372-O

Song, Q., & Chissom, B. S. (1993b). Forecasting enrollments with fuzzy time series—Part I. *Fuzzy sets and systems*, *54*(1), 1-9. https://doi.org/10.1016/0165-0114(93)90355-L

Wei, W. W. (2013). Time series analysis. In Little T. D. (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis*, New York: Oxford University Press, pp. 458-485.

Xu, Z., & Yager, R. R. (2006). Some geometric aggregation operators based on intuitionistic fuzzy sets. *International Journal of General Systems*, *35*(4), 417-433. https://doi.org/10.1080/03081070600574353

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X