# A Fuzzy Time Series Forecasting Model Based on Yearly Difference of the Student Enrollment Number

Hongxu Wang
College of tourism management, QiongZhou University,
Sanya, China
whx16233@126.com

Hui Wang
College of science and Engineering, QiongZhou University,
Sanya, China
hwan072@163.com

Jianchun Guo
College of tourism management,
QiongZhou University,
Sanya, China
gjc92@163.com

Hao Feng
College of Science and Engineering
QiongZhou University
Sanya, China
feng_h007 @126.com

***Abstract*—A number of forecasting models have been proposed based on fuzzy time series in the past 20 years, and forecasting accuracy rate continues to increase. This paper establishes a fuzzy time series forecasting model based on yearly difference of the student enrollment number. The method uses the yearly difference of the student enrollment number as domain to establish a fuzzy number, inverse fuzzy number and prediction formula. The forecasting process is illustrated by applying history student enrollment number of the University of Alabama, and forecasting accuracy rate is higher than that of the existing method.**

***Keywords— difference of enrollment number; domain; inverse fuzzy number; fuzzy time series; forecasting model***

## I. INTRODUCTION

Fuzzy sets theory founded by Zadeh[1] is widely used and fruitful. The combination of fuzzy techniques and time series forecasting technique opens up new application field. Song and Chissom first proposed fuzzy time series forecasting model[2]. Many scholars have joined this field [5-20], and have put forward different models. The study scope has involved in student registration number prediction [2-15], the temperature prediction [16], the stock index prediction [17], the exchange rate forecast [18], data mining [19], automobile accident prediction [20] and et al., but the prediction accuracy is still not high. This paper uses the idea of data mining, integrated historical data yearly difference technique of Hwang and Chen and Lee [8], inverse fuzzy numbers technique of Jilani and Burney and Ardil [6], and forecasting model of Saxena and Sharma and Easo [5], to propose a new time series forecasting model fuzzy. Section II describes the preliminaries; Section III gives the basic steps of a new fuzzy time series forecasting model; Section IV presents in detail the prediction process of new fuzzy time series forecasting model by applying the number of new students registered predicted example of the University of Alabama from 1971 to 1992; Section V compares the new model with existing models; Section VI is the conclusion.

## II. PRELIMINARIES

Here are the basic concepts, which are used in many literature. For example

Definition 1 [7, 8, 9] Yearly difference of historical data is defined as

$$E_i - E_{i-1} \qquad (1)$$

Where $E_i$ and $E_{i-1}$ are the historical data in $i$ and $i-1$ year.

Definition 2 [7] AFER (Average Forecasting Error Rate) calculation formula is

$$AFER = \frac{1}{n}\sum_{i=1}^{n}\frac{|E_i - F_i|}{E_i} \times 100\% ,$$

Where $E_i$ and $F_i$ are the historical data and forecasting data in $i$ year, respectively.

Definition 3 [7] MSE (Mean Square Error) calculation formula is

$$MSE = \frac{1}{n}\sum_{n=1}^{n}(E_i - F_i)^2 ,$$

Where $E_i$ and $F_i$ are the historical data and forecasting data in $i$ year, respectively.

## III. NEW FUZZY TIME SERIES FORECASTING MODEL

The new model is based on historical yearly difference of Hwang and Chen and Lee [8], inverse fuzzy numbers of Jilani and Burney and Ardil [6], and improved forecasting model of Saxena and Sharma and Easo [5]. First, no need to create a domain interval; Second, no need to find the midpoint of each subinterval; Third, do not use yearly change percentage in historical data,

while use the yearly difference in historical data [7]; fourth, greater adjustment for prediction formula of the inverse fuzzy numbers. A new model "a fuzzy time series forecasting model based on the yearly difference of new students registered number " is proposed. Its application steps are following.

Introduce the application steps of the new model and compare with existing models based on the prediction question of new students enrollment number of the University of Alabama from 1971 to 1992, and exhibit application steps of "a fuzzy time series forecasting model based on new student enrollment number difference" in detail.

***Step 1 List historical data tables***

The enrollment numbers of new students in University of Alabama from 1971 to 1992 are shown in Table 1 [2].

***Step 2 Calculate yearly differences of historical data***

Apply formula (1) to calculate new student enrollment number difference in the University of Alabama from 1971 to 1992 [7], for example $d_{1972} = E_{1972} - E_{1971} = 13563 - 13055 = 508$.

So calculate yearly difference for each year from 1972 to 1992 $d_i (i = 1972, 1973, \mathrm{L}\ , 1992)$, fill in Table 1. This work has been done by Hwang and Chen and Lee [8] and Sah and Degtiarev and MIEEE [9].

TABLE I THE FORECASTS

| Year | Enrollment $E_i$ | Element $d_i$ | Forecast $F_i$ | $E_i - F_i$ | $(E_i - F_i)^2$ | $\lvert E_i - F_i \rvert / E_i$ |
|---|---|---|---|---|---|---|
| 1971 | 13055 | | | | | |
| 1972 | 13563 | 508 | 13563 | 0 | 0 | 0.000000 |
| 1973 | 13867 | 304 | 13867 | 0 | 0 | 0.000000 |
| 1974 | 14696 | 829 | 14695 | 1 | 1 | 0.000068 |
| 1975 | 15460 | 764 | 15460 | 0 | 0 | 0.000000 |
| 1976 | 15311 | －149 | 15311 | 0 | 0 | 0.000000 |
| 1977 | 15603 | 292 | 15603 | 0 | 0 | 0.000000 |
| 1978 | 15861 | 258 | 15861 | 0 | 0 | 0.000000 |
| 1979 | 16807 | 946 | 16796 | 1 | 1 | 0.000059 |
| 1980 | 16919 | 112 | 16919 | 0 | 0 | 0.000000 |
| 1981 | 16388 | －531 | 16388 | 0 | 0 | 0.000000 |
| 1982 | 15433 | －955 | 15432 | 1 | 1 | 0.000065 |
| 1983 | 15497 | 64 | 15497 | 0 | 0 | 0.000000 |
| 1984 | 15145 | －352 | 15144 | 1 | 1 | 0.000066 |
| 1985 | 15163 | 18 | 15163 | 0 | 0 | 0.000000 |
| 1986 | 15984 | 821 | 15982 | 2 | 4 | 0.000125 |
| 1987 | 16859 | 875 | 15859 | 0 | 0 | 0.000000 |
| 1988 | 18150 | 1291 | 18150 | 0 | 0 | 0.000000 |
| 1989 | 18970 | 820 | 18970 | 0 | 0 | 0.000000 |
| 1990 | 19328 | 358 | 19327 | 1 | 1 | 0.000052 |
| 1991 | 19337 | 9 | 19337 | 0 | 0 | 0.000000 |
| 1992 | 18876 | －146 | 18874 | 2 | 4 | 0.000106 |
| AFER | | | | | | 0.0026% |
| MSE | | | | | 1 | |

***Step 3 Establish a discrete domain D***

Establish a discrete domain D based on yearly difference of historical data as element, as follows:

$$\begin{aligned} D = \{ & d_{1972} = 508, d_{1973} = 304, d_{1974} = 829, \\ & d_{1975} = 764, d_{1976} = -149, d_{1977} = 292, \\ & d_{1978} = 258, d_{1979} = 946, d_{1980} = 112, \\ & d_{1981} = -531, d_{1982} = -955, d_{1983} = 64, \\ & d_{1984} = -352, d_{1985} = 18, d_{1986} = 821, \\ & d_{1987} = 875, d_{1988} = 1291, d_{1989} = 820, \\ & d_{1990} = 358, d_{1991} = 9, d_{1992} = -461\}. \end{aligned}$$

This domain has four characteristics: the year is order, the yearly differential is element, the element is a real number and a discrete domain.

***Step 4 Establish inverse fuzzy number of consecutive years***

Establish inverse fuzzy numbers of consecutive years based on domain D, as follows:

$$\left.\begin{aligned} v_{1972} &= \frac{1 + 0.0001}{\frac{1}{d_{1972}} + \frac{0.0001}{d_{1973}}}, \\ v_{\alpha} &= \frac{0.0001 + 1 + 0.0001}{\frac{0.0001}{d_{\alpha-1}} + \frac{1}{d_{\alpha}} + \frac{0.0001}{d_{\alpha+1}}}, 1973 \le \alpha \le 1991, \\ v_{1992} &= \frac{0.0001 + 1}{\frac{0.0001}{d_{1991}} + \frac{1}{d_{1992}}}. \end{aligned}\right\} \quad (2)$$

***Step 5 Establish a new forecasting formula to forecast***

The forecasting formula of this model is:

$$F_{\alpha} = E_{\alpha-1} + v_{\alpha}, (\alpha = 1972, 1973, \mathrm{L}\ , 1992). \quad (3)$$

Apply real number discrete domain and inverse fuzzy numbers formula (2) and forecasting formula (3) to forecast the enrollment number of new students in the

University of Alabama from 1972 to 1992, for example

$$v_{1972} = \frac{1+0.0001}{\frac{1}{d_{1972}}+\frac{0.0001}{d_{1973}}} = \frac{1.0001}{\frac{1}{508}+\frac{0.0001}{304}} =$$

$$= \frac{1.0001}{0.001968503937+0.000000328947}$$

$$= \frac{1.0001}{0.001968832884} = 508$$

$$F_{1972} = E_{1971} + v_{1972} = 13055 + 508 = 13563$$

Fill in table 1 the forecasting results of new students enrollment number in the University of Alabama from 1972 to 1992, the forecasting error, the forecasting variance, the forecasting error rate, et al. Then calculate the mean forecasting error rate AFER and mean square error MSE and fill in Table 1.

## IV. Comparison with Existing Models

Forecast the results and compare different forecasting results by using existing different forecasting model of fuzzy time series with this model to examine the merits of this model, namely the application of the models in references [2], [4], [10], [12], [7], [6], [13], [14], [5], [11], [15] and the model proposed in this paper to forecast new students enrollment number in the University of Alabama. The results obtained are shown in Table 2-1 and Table 2-2, (where most of the content is taken from the Table 6 in reference [5], a small part taken from the Table VI in the reference [6]). The forecasting models respectively proposed by Saxena and Sharma and Easo [5], and Stevenson and Porter [7] are good after evaluating the merits of forecasting models by applying forecast data and AFER (average prediction error rate) and MSE (mean square error). The results of fuzzy time series forecasting model presented in this paper are better than existing models.

TABLE 2 COMPARISON OF DIFFERENT FORECASTING MODELS[5],[6]

| Year | Enrollments | Chen[12] | Stevenson, Porter[7] | Saxena,Sharma, Easo[5] | Hwang,Chen, Lee[11] | Huarng[15] | Proposed Model |
|---|---|---|---|---|---|---|---|
| 1971 | 13055 | - | - | - | - | | |
| 1972 | 13563 | - | 13410 | 13486 | - | 14000 | 13563 |
| 1973 | 13867 | - | 13932 | 13896 | - | 14000 | 13867 |
| 1974 | 14696 | 14500 | 14664 | 14698 | - | 14000 | 14695 |
| 1975 | 15460 | 15500 | 15423 | 15454 | - | 15500 | 15460 |
| 1976 | 15311 | 15500 | 15847 | 15595 | 16260 | 15500 | 15311 |
| 1977 | 15603 | 15500 | 15580 | 15600 | 15511 | 16000 | 15603 |
| 1978 | 15861 | 15500 | 15877 | 15844 | 16003 | 16000 | 15861 |
| 1979 | 16807 | 16500 | 16773 | 16811 | 16261 | 16000 | 16796 |
| 1980 | 16919 | 16500 | 16897 | 16916 | 17407 | 17500 | 16919 |
| 1981 | 16388 | 16500 | 16341 | 16425 | 17119 | 16000 | 16388 |
| 1982 | 15433 | 15500 | 15671 | 15657 | 16188 | 16000 | 15432 |
| 1983 | 15497 | 15500 | 15507 | 15480 | 14833 | 16000 | 15497 |
| 1984 | 15145 | 15500 | 15200 | 15214 | 15497 | 15500 | 15144 |
| 1985 | 15163 | 15500 | 15218 | 15184 | 14745 | 16000 | 15163 |
| 1986 | 15984 | 15500 | 16035 | 15995 | 15163 | 16000 | 15982 |
| 1987 | 16859 | 16500 | 16903 | 16861 | 16384 | 16000 | 15859 |
| 1988 | 18150 | 18500 | 17953 | 17965 | 17659 | 17500 | 18150 |
| 1989 | 18970 | 18500 | 18879 | 18964 | 19150 | 19000 | 18970 |
| 1990 | 19328 | 19500 | 19303 | 19329 | 19770 | 19000 | 19327 |
| 1991 | 19337 | 19500 | 19432 | 19378 | 19928 | 19500 | 19337 |
| 1992 | 18876 | 18500 | 18966 | 18984 | 15837 | 19000 | 18874 |
| AFER | 3.22% | 1.52% | 0.57% | 0.34% | 2.44% | 1.53% | 0.0026% |
| MSE | 423027 | 86696 | 21575 | 9169 | 226611 | 86694 | 1 |

## V. Conclusion

In this paper, we establish a new model based on data mining theory by integrating yearly difference of historical data in Hwang and Chen and Lee [8], inverse fuzzy numbers in Jilani and Burney and Ardil [6], and forecasting models in Saxena and Sharma and Easo [5], et al. This new model provides minimal AFER and MSE, but here only study training samples and lack the study of test samples. We expect to develop fuzzy time series forecasting methods with not only a high prediction accuracy of training samples and that of the test samples.

## References


[1] Zadeh L A. Fuzzy sets. Information and Control, Vol.8, pp.338-353，1965

[2] Q Song, B S Chissom. Forecasting enrollments with fuzzy time series--- part 1. Fuzzy Set and Systems, Vol. 54, pp: 1-9, 1993

[3] Q Song, B S Chissom. Fuzzy time series and its models. Fuzzy Set and Systems, 1993, Vol. 54, pp: 269-277, 1993

[4] Q Song, B S Chissom. Forecasting enrollments with fuzzy time series--- part 2. Fuzzy Set and Systems, Vol. 62, pp: 1-8, 1994

[5] Preetika Saxena, Kalyani Sharma, Santhosh Easo. Forecasting enrollments based on fuzzy time series with higher forecast accuracy rate. Int. J. Computer Technology& Applications, Vol.3, No.3, pp: 957-961, 2012

[6] T A Jilani, S M A Burney, C Ardil. Fuzzy metric approach for fuzzy time series forecasting based on frequency density based

partitioning. Proceedings of World Academy of Science, Engineering and Technology, Vol. 34, pp: 1-6, 2007

[7] Meredith Stevenson and John Porter. Fuzzy time series forecasting using percentage change as the universe of discourse. Proceedings of World Academy of Science, Engineering and Technology, Vol. 55, pp: 154-157, 2009

[8] Jeng-Ren Hwang, Shyi-Ming Chen, Chia-Hoang Lee. Handling forecasting problems using fuzzy time series. Fuzzy Sets and Systems, Vol. 100, pp: 217-228, 1998

[9] Melike Sah and Konstantin Y Degtiarev, MIEEE. Forecasting enrollment model based on first-order fuzzy time series. Procedings of World Academy of Science, Engineering and Technology Volume 1 January 2005 ISSN 1307-6884, pp.375-378, 2005

[10] S M Chen. Forecasting enrollments based on fuzzy time series. Fuzzy Sets and Systems, Vol. 81, pp: 311-319, 1996

[11] Jeng-Ren Hwang, Shyi-Ming Chen, Chia-Hoang Lee. Handling forecasting problems using fuzzy time series. Fuzzy Sets and Systems, Vol. 100, pp: 217-228, 1998

[12] S M Chen. Forecasting enrollments based on high-order fuzzy time series. Cybernetics and Systems: An International Journal, Vol. 33, pp: 1-16, 2002

[13] T A Jilani, S M A Burney. M-factor high order fuzzy time series forecasting for road accident data [M]. IEEE-IFSA 2007, World Congress, Cancun, Mexico, June 18-21, Forthcoming in Book series Advances in Soft Computing, Springer-Verlag, 2007

[14] T A Jilani, S M A Burney, C Ardil. Multivariate high order fuzzy time series forecasting for car road accidents. International Journal of Computational Intelligence, Vol. 4, No. 1, pp: 15-20, 2007

[15] K Huang. Heuristic models of time series for forecasting. Fuzzy Sets and Systems, Vol. 123, pp. 369-386, 2001

[16] S M Chen, J R Hwang. Temperature prediction using fuzzy time series. IEEE Transaction on Systems, Man, and Cybernetics-Part B: Cybernetics Vol. 30, pp: 263-275, 2000

[17] Huarng K H, Yu H K. A type 2 fuzzy time series model for stock index forecasting. Physical A: Statistical Mechanics and Applications, Vol. 353, No.1-4, pp: 445-462, 2005

[18] Leu Y H, Lee C P, Jou Y Z. A distance-based fuzzy time series model for exchange rates forecasting. Expert Systems with Applications, Vol. 36, No. 4, pp: 8107-8114, 2009

[19] Fu T-C. A review on time series data mining. Engineering Applications of Artificial Intelligence, Vol. 24, pp: 164-181，2011

[20] Jilani T A, Burncy S M A, Ardil C. Multivariate high order fuzzy time series forecasting for car road accidents, International Journal of Computational Intelligence, Vol. 4, No. 1 pp: 15-20, 2007