Fuzzy Time Series Forecasting Model of Inverse Fuzzy Number Based on Percentage Year by Year of Continuous Point

Haifeng Wang[1, a], Kun Zhang[1, b], Zhuang Li[1, c] and Hongxu Wang[1, d]

[1]College of Computer Engineering, Hainan Tropical Ocean University, Sanya, Hainan, 572022, China

[a]wyfxxz@163.com, [b]zk0588@163.com, [c]hitlz@163.com, [c]whx16233@163.com

Keywords: Percentage; Continuous point; Inverse fuzzy number; Fuzzy time-series forecasting

Abstract. Fuzzy time-series forecasting model of inverse fuzzy number based on percentage year to year of continuous point is proposed. We improved the forecasting model of Saxena. The new model puts the percentage year to year of historical data as for the domain of discourse, uses percentage year to year of continuous point to define fuzzy number, and then defines the corresponding inverse fuzzy number. At last we again provide the predictor formula and study the prediction problem of freshman registration number at the University of Alabama, in order to demonstrate the application of a new prediction model. The results show that AFER and MSE of the model are very small compared with the existing models.

Introduction

Zadeh published pioneering thesis fuzzy sets in 1965 which made the research of uncertainty for added a powerful fuzzy set theory. As the time-series prediction model has been applied in many fields by many scholars. For example in literature, data mining is studied. In literature, Stock Market are studied. Because there are many indetermination about prediction problem, the fuzzy set theory is combined with time series that become the new direction for time series prediction model research. Many scholars have done a helpful job. Such as in literature Song first discussed to propose the first-order time-invariant fuzzy time series model and time-variant fuzzy time series model. And the model is applied to the enrollments prediction problem in University of Alabama at Tuscaloosa. Literature respectively put forward improved fuzzy time series forecasting model. Especially the improved model that was proposed by Saxena is applied to the enrollments prediction problem in University of Alabama at Tuscaloosa. The model makes AFER (Average Forecasting Error Rate) and MSE (Mean Square Error) to greatly reduce. This thesis has improved base on the improved model proposed in the literature. Fuzzy time-series forecasting model of inverse fuzzy number based on percentage year by year of continuous point is gained. The model is applied to the enrollments prediction problem in University of Alabama at Tuscaloosa. Compared with other algorithms, the proposed algorithm parameters such as AFER and MSE have been greatly decreased.

Basic Concept

This section describes some basic concepts that are used in this thesis by literature.

Definition 1 Let $Z=\{Z_1,Z_2,\cdots,Z_n\}$ be nonempty set, W be named fuzzy subset in Z. A map $\mu_W : Z \to [0,1]$ is given from domain of discourse Z to closed interval [0, 1]. μ_W is named membership function of fuzzy subset W. $\mu_W(Z_i),(Z_i \in Z, i=1,2,\cdots,n)$ is named membership degree of fuzzy set W for the element Z_i, as follow: $W=\mu_W(Z_1)/Z_1+\mu_W(Z_2)/Z_2+\cdots\mu_W(Z_n)/Z_n$

Definition 2 When the element $Z_i(1\le i\le n)$ be real number, fuzzy subset sometimes is named fuzzy number in nonempty set $Z=\{Z_1,Z_2,\cdots,Z_n\}$. For example when $Z_i(1\le i\le n)$ be real number, Triangular fuzzy number in nonempty set $Z=\{Z_1,Z_2,\cdots,Z_n\}$ can be defined as follows:

$W_1 = \frac{1}{Z_1} + \frac{0.5}{Z_2} + \frac{0}{Z_3} + \cdots + \frac{0}{Z_n}$, $W_2 = \frac{0.5}{Z_1} + \frac{1}{Z_2} + \frac{0.5}{Z_3} + \frac{0}{Z_4} + \cdots + \frac{0}{Z_n}$, $\cdots$

$W_\beta = \frac{0}{Z_1} + \cdots + \frac{0.5}{Z_{\beta-1}} + \frac{1}{Z_\beta} + \frac{0.5}{Z_{\beta+1}} + \cdots + \frac{0}{Z_n}, 2 \le \beta \le n-1$ $\cdots$ $W_n = \frac{0}{Z_1} + \cdots + \frac{0}{Z_{n-2}} + \frac{0.5}{Z_{n-1}} + \frac{1}{Z_n}$.

Definition 3 The definition of YYCR (Year to year changes rate) percentage for truthful data is as follow:

$$Z^{x+i} = \frac{(A_{x+i} - A_{x+i-1})}{A_{x+i-1}} \times 100\%, i \in \{1,2,\cdots,n\} \tag{1}$$

Among A_i, A_{i-1} is respectively truthful data in $x+i, x+i-1$ year.

Definition 4 The difference between truthful data A_i and predicted data F_i is named *FE* (Forecasting Error), as follow:

$$FE = A_i - F_i \tag{2}$$

Definition 5 Computational formula of MSE is defined as follow:

$$MSE = \sum_{i=1}^{n} (A_i - F_i)^2 \Big/ n \tag{3}$$

Definition 6 Computational formula of FER (Forecasting Error Rate) is defined as follow:

$$FER = |A_i - F_i| \Big/ A_i \tag{4}$$

Definition 7 Computational formula of AFER is defined as follow:

$$AFER = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{|A_i - F_i|}{A_i} \right) \tag{5}$$

Suggested New Model

New model improves the prediction model in. Don't need to set up domain of discourse range firstly. Don't need to request the mid-point of each interval secondly. At last the content of prediction formula for the inverse fuzzy number also is improved bigger. Application steps of the new model are as follows:

Need to study the history of the data list $A_{x+i}, (i = x+1, x+2, \cdots, x+n)$;

Using Eq. (1) to calculate the percentage of YYCR for the historical data, $Z^{x+i}, (i = 1, 2, \cdots, n)$;

With the percentage of YYCR for the historical data as the element, theory of discrete domain $Z = \{Z^{x+1}, Z^{x+2}, \cdots, Z^{x+n}\}$ is established;

Establish the inverse fuzzy number of the continuous point in domain of discourse Z as following:

$$k_\beta = \begin{cases} \dfrac{1+0.003}{1/Z_1 + 0.003/Z_2}, & \beta = 1 \\ \dfrac{0.003+1+0.003}{0.003/Z_{\beta-1} + 1/Z_\beta + 0.003/Z_{\beta+1}}, & 2 \le \beta \le n-1 \\ \dfrac{0.003+1}{0.003/Z_{n-1} + 1/Z_n}, & \beta = n \end{cases} \tag{6}$$

Set up the new predictor formula (7) as following:

$$F_\beta = A_{\beta-1} \times (1+k_\beta\%) \quad (7)$$

Among $A_{\beta\text{-}1}$ is historical data in $\beta-1$ year. F_β is prediction data in β year. The formula shows that applying historical data in $\beta-1$year to predict forecast data in β year.

Appling prediction formula for historical data to make predictions.

Application Examples of Prediction Research about the Enrollment of the University of Alabama

A case study of what enrollment prediction problem at the University of Alabama, the improved model will be introduced in this paper. Taking the data of enrollment prediction problem first in literature by Song at the University of Alabama as the research object, the data were listed Table 1.

Table 1 History of enrollment and the percentage of change year by year at the University of Alabama

Table 1

Year	Enrollment	Year to year	The percentage of change	Element	Forecasts (F_i)	$A_i - F_i$	$(A_i - F_i)^2$	$FER = \frac{\lvert A_i - F_i \rvert}{A_i}$
1971	13055							
1972	13563	1971-72	3.89%	$Z_{14}^{1972} = 3.89$	13562	1	1	0.00737%
1973	13867	1972-73	2.24%	$Z_{13}^{1973} = 2.24$	13867	0	0	0
1974	14696	1973-74	5.98%	$Z_{20}^{1974} = 5.98$	14697	-1	1	0.00680%
1975	15460	1974-75	5.20%	$Z_{16}^{1975} = 5.20$	15460	0	0	0
1976	15311	1975-76	-0.96%	$Z_{5}^{1976} = -0.96$	15300	11	121	0.07184%
1977	15603	1976-77	1.91%	$Z_{12}^{1977} = 1.91$	15603	0	0	0
1978	15861	1977-78	1.65%	$Z_{10}^{1978} = 1.65$	15860	1	1	0.00630%
1979	16807	1978-79	5.96%	$Z_{19}^{1979} = 5.96$	16807	0	0	0
1980	16919	1979-80	0.67%	$Z_{9}^{1980} = 0.67$	16919	0	0	0
1981	16388	1980-81	-3.14%	$Z_{2}^{1981} = -3.14$	16388	0	0	0
1982	15433	1981-82	-5.83%	$Z_{1}^{1982} = -5.83$	15435	-2	4	0.01296%
1983	15497	1982-83	0.41%	$Z_{8}^{1983} = 0.41$	15497	0	0	0
1984	15145	1983-84	-2.27%	$Z_{4}^{1984} = -2.27$	15146	-1	1	0.00660%
1985	15163	1984-85	0.12%	$Z_{7}^{1985} = 0.12$	15163	0	0	0
1986	15984	1985-86	5.41%	$Z_{17}^{1986} = 5.41$	15984	0	0	0
1987	16859	1986-87	5.47%	$Z_{18}^{1987} = 5.47$	16859	0	0	0
1988	18150	1987-88	7.66%	$Z_{21}^{1988} = 7.66$	18149	1	1	0.00551%
1989	18970	1988-89	4.52%	$Z_{15}^{1989} = 4.52$	18970	0	0	0
1990	19328	1989-90	1.89%	$Z_{11}^{1990} = 1.89$	19328	0	0	0
1991	19337	1990-91	0.05%	$Z_{6}^{1991} = 0.05$	19337	0	0	0
1992	18876	1991-92	-2.38%	$Z_{3}^{1992} = -2.38$	18876	0	0	0
AFER								0.00559%
MSE							130	

Using Eq. (1) to Calculate the Percentage of YYCR for the Historical Data. In literature, Saxena has calculated the percentage of YYCR for enrollment at the University of Alabama in 1971-1992.In this case, let $x = 1971$, using Eq. (1) to calculate.

$$Z^{1972} = \left(A_{1972} - A_{1971}\right)/\,A_{1971} = \left(13563 - 13055\right)/13055 = 3.89\%.$$

With the Percentage of YYCR for the Historical Data as the Element, Theory of Discrete Domain is Established. The elements of domain are the percentage of the historical data YYCR by size. Theory of discrete domain that is formed by the percentage of enrollment YYCR at the University of Alabama in 1971-1992 as follows:

$$\begin{aligned} Z = \{ & Z_1^{1982} = -5.83, Z_2^{1981} = -3.14, Z_3^{1992} = -2.38, Z_4^{1984} = -2.27, Z_5^{1976} = -0.96, Z_6^{1991} = 0.05, Z_7^{1985} = 0.12, \\ & Z_8^{1983} = 0.41, Z_9^{1980} = 0.67, Z_{10}^{1978} = 1.65, Z_{11}^{1990} = 1.89, Z_{12}^{1977} = 1.91, Z_{13}^{1973} = 2.24, Z_{14}^{1972} = 3.89, \\ & Z_{15}^{1989} = 4.52, Z_{16}^{1975} = 5.20, Z_{17}^{1986} = 5.41, Z_{18}^{1987} = 5.47, Z_{19}^{1979} = 5.96, Z_{20}^{1974} = 5.98, Z_{21}^{1988} = 7.66\} \end{aligned} \quad (8)$$

Among subscript of the elements indicate their serial number. Superscripts of the elements indicate the year of the elements. For example $Z_2^{1981} = -3.14$, it shows that the serial number of the element is 2 and the year of the element is 1981. In order to calculate conveniently, Eq. (8) is rewritten into domain of discourse that superscript in chronological order and the subscript is still according to the serial number of the elements. Domain of discourse is as follows:

$$\begin{aligned} Z = \{ & Z_{14}^{1972} = 3.89, Z_{13}^{1973} = 2.24, Z_{20}^{1974} = 5.98, Z_{16}^{1975} = 5.20, Z_5^{1976} = -0.96, Z_{12}^{1977} = 1.91, Z_{10}^{1978} = 1.65, \\ & Z_{19}^{1979} = 5.96, Z_9^{1980} = 0.67, Z_2^{1981} = -3.14, Z_1^{1982} = -5.83, Z_8^{1983} = 0.41, Z_4^{1984} = -2.27, Z_7^{1985} = 0.12, \\ & Z_{17}^{1986} = 5.41, Z_{18}^{1987} = 5.47, Z_{21}^{1988} = 7.66, Z_{15}^{1989} = 4.52, Z_{11}^{1990} = 1.89, Z_6^{1991} = 0.05, Z_3^{1992} = -2.38\} \end{aligned} \quad (9)$$

Application of Prediction Formula for Historical Data to Make Predictions. Applying the inverse fuzzy number Eq. (6) which fuzzy number of normal distribution and the prediction Eq. (7), enrollment is simulated prediction in 1971-1992 at the University of Alabama, among n=21. For example:

$$k_{1972} = \frac{0.003 + 1 + 0.003}{\dfrac{0.003}{Z_{13}^{1973}} + \dfrac{1}{Z_{14}^{1972}} + \dfrac{0.003}{Z_{15}^{1989}}} = \frac{1.006}{\dfrac{0.003}{2.24} + \dfrac{1}{3.89} + \dfrac{0.003}{4.52}} = 3.883090$$

$$F_{1972} = F_{1971} \times (1 + k_{1972}\%) = 13055 \times (1 + 3.883090\%) = 13562$$

$$A_{1972} - F_{1972} = 13563 - 13562 = 1, \left|A_{1972} - F_{1972}\right| / A_{1972} = 1 \div 13563 = 0.000074$$

The computed data are filled in Table 1 too.

The Comparison of Different Forecasting Models

Applying the different models that were proposed in this paper and literature, we come into studying enrollment prediction at the University of Alabama. The given results are listed in Table 2. If use AFER and MSE of these two indexes to evaluate the merits of the forecasting model, the forecasting model that proposed by Saxena in literature and presented in this paper are all better.

Conclusion

For more than 20 years, people unremitting efforts, the original prediction model is continuously improved, and new ideas emerge continuously. For the application at the University of Alabama in enrollment prediction problem, AFER and MSE of the proposed model in this paper are less than the existing methods. Combining calendar year enrollment plan, these studies will make people having a more detailed understanding for history change law about enrollment at the University of Alabama. To study and analysis the reason of new student erosion and formulate the new recruitment plan that it is good. Developing fuzzy time series prediction method of new higher prediction accuracy, especially developing fuzzy time series forecasting model to predict the unknown year data and expanding application field of fuzzy time series forecasting model that should be a research direction of fuzzy time series forecasting method.

Table 2 Comparison of Different Forecasting Models

Year	Enrollments	Song	Porter	Jilani	Chen	Hwang	Jilani	Jilani	Huarng	Saxena	Proposed
1971	13055	-	-	13579	-	-	-	14464	-	-	-
1972	13563	-	13410	13798	-	-	-	14464	14000	13486	13562
1973	13867	-	13932	13798	-	-	-	14464	14000	13896	13867
1974	14696	-	14664	14452	14500	-	14730	14710	14000	14698	14697
1975	15460	14700	15423	15373	15500	-	15615	15606	15500	15454	15460
1976	15311	14800	15847	15373	15500	16260	15614	15606	15500	15595	15300
1977	15603	15400	15580	15623	15500	15511	15611	15606	16000	15600	15603
1978	15861	15500	15877	15883	15500	16003	15611	15606	16000	15844	15860
1979	16807	15500	16773	17079	16500	16261	16484	16470	16000	16811	16807
1980	16919	16800	16897	17079	16500	17407	16476	16470	17500	16916	16919
1981	16388	16200	16341	16497	16500	17119	16469	16470	16000	16425	16388
1982	15433	16400	15671	15373	15500	16188	15609	15606	16000	15657	15435
1983	15497	16800	15507	15373	15500	14833	15614	15606	15500	15480	15497
1984	15145	16400	15200	15024	15500	15497	15612	15606	16000	15214	15146
1985	15163	15500	15218	15024	15500	14745	15609	15606	16000	15184	15163
1986	15984	15500	16035	15883	15500	15163	15606	15606	16000	15995	15984
1987	16859	15500	16903	17079	16500	16384	16477	16470	16000	16861	16859
1988	18150	16800	17953	17991	18500	17659	18482	18473	17500	17965	18149
1989	18970	19300	18879	18802	18500	19150	18481	18473	19000	18964	18970
1990	19328	17800	19303	18994	19500	19770	19158	19155	19000	19329	19328
1991	19337	19300	19432	18994	19500	19928	19155	19155	19500	19378	19337
1992	18876	19600	18966	18916	18500	15837	18475	18473	19000	18984	18876
AFER	3.22%	3.11%	0.57%	1.02%	1.52%	2.44%	1.40%	2.39%	1.53%	0.34%	0.00559%
MSE	423027	407507	21575	41426	86696	226611	82269	227194	86694	9169	130

Acknowledgements

This work was financially supported by the key project of scientific research of Hainan Province(No.Hnky2015ZD-14), the scientific and technological cooperative project for college and region of Sanya (2015YD16, 2014YD30), the key Laboratory of Sanya Project(No.L1410).

References

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

[2] Chen C-H, Hong T-P: Tseng V S. fuzzy data mining for time-series data. Applied Soft Computing, vol.12 (2012), pp. 536-542.

[3] Chu H-H, Chen T-L, Cheng C-H and Huang C-C: Fuzzy dual-factor time-service for stock index forecasting. Expert Systems with Application, vol.36 (2009), pp. 165-171.

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

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

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

[7] J R Hwang, S M Chen, C H Lee: Handling forecasting problems using fuzzy time series. Fuzzy Sets and Systems, 1998, vol.100, pp. 217-228.

[8] Stevension Porter: Fuzzy time series forecasting using percentage change as the universe of discourse. Proceedings of World Academy of Science. Engineering and Technology. vol.55(2009), pp.154-157

[9] Tahseen Ahmed Jilani, Syed Muhammad Aqil Burney, and Cemal Ardil: Fuzzy metric approach for fuzzy time series forecasting based on frequency density based partitioning. World Academy of Science, Engineering and Technology, Vol. 34(2007), pp. 1-6.

[10] Preetika Saxena, Kalyani Sharma and Santhosh Easo: Foreca enrollment based on fuzzy time series with higher forecast accuracy rate. International Journal of Computer Technology and Applications, vol. 3. No. 3(2012).pp.957-961.

[11] T A Jilani, S M A Burney: M-factor high order fuzzy time series forecasting for road accident data. Analysis and design of intelligent systems using soft computing techniques. Springer Berlin Heidelberg, (2007).p.246-254.

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

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