ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A NEW METHOD FOR POPULATION FORECASTING BASED ON FUZZY TIME SERIES WITH HIGHER FORECAST ACCURACY RATE

Preetika Saxena*, Satyam Shrivastava, Avinash S. Bundela

* Computer Science and Engineering, Acropolis Institute of Technology and Research, Indore (MP), India
Computer Science and Engineering, Acropolis Institute of Technology and Research, Indore (MP), India
Computer Science and Engineering, Acropolis Institute of Technology and Research, Indore (MP), India

ABSTRACT

The Time-Series models have been used to make predictions in whether forecasting, academic enrollments, etc. Song & Chissom introduced the concept of fuzzy time series in 1993. Over the past 19 years, many fuzzy time series methods have been proposed for population forecasting. But the forecasting accuracy rate of the existing methods is not good enough. These methods have either used actual population or difference of population as the universe of discourse. And either used frequency density based partitioning or nature- ratio based partitioning. In this paper, we proposed a method based on fuzzy time series, which gives the higher forecasting accuracy rate than the existing methods. The proposed method used the percentage change as the universe of discourse and mean based partitioning. To illustrate the forecasting process, the historical population of Azerbaijan is used.

KEYWORDS: fuzzy set, fuzzy time series, time variant model, first order model, forecast error.

INTRODUCTION

Forecasting plays an important role in our day- to- day life. If there is an uncertainty about the future, decision makers need to forecast. Forecasting is the process of predicting future outcomes, by which decision makers analyze the related data and graphs to decide and take the best decisions for the future. The time series forecasting problem has attracted more attention for at least two reasons. First, most business, economic, and financial data are time series. Second, the technology for making and evaluating time series forecast is well developed. During last few decades, various approaches have been developed for time series forecasting, but the classical time series methods can not deal with the forecasting problems in which the values of time series are linguistic terms represented by fuzzy sets [5]. To overcome this drawback, Song & Chissom presented the theory of fuzzy time series in 1993[13]. Fuzzy time series model deal with the both linguistic & numerical values.

Fuzzy forecasting methods have been used to model enrollment data for the University of Alabama ([1] - [9], [11] - [13], [18]-[19]), car fatalities ([14] - [15]), food grain production [16], and population forecasting [17]. Hwang, Chen & Lee [2] and Sah & Degtiarev[6] used the differences of the enrollments and Stevenson & Porter [8] used the percentage change of year- to- year enrollments of the University of Alabama as the universe of discourse. For more than one decade, different fuzzy time series models have also been applied to solve various domain problems, such as financial forecasting, temperature forecasting, etc. Fuzzy time series models also used for forecasting ICT products.
The objective of the present work is to build, implement and test the fuzzy time series method for population forecasting, and comparison of the result with existing forecasting methods.

In section 2, we briefly review the basic concepts and definitions of fuzzy time series. In section 3, we proposed a new method by using percentage change as the universe of discourse and mean based partitioning. The proposed method has been implemented on the historical population of Azerbaijan. In section 4, we compared the forecasting result of proposed method with the forecasting results of the existing methods. Finally the concluding remarks are discussed in section 5.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

CONCEPT OF FUZZY TIME SERIES

Definition 1: *Fuzzy Set*

Fuzzy sets are sets whose elements have degree of membership.

Let U be the universe of discourse, U= {u1, u2,…, un}, and let A be a fuzzy set in the universe of discourse U defined as follows:

$A = fA(u1) / u1 + fA(u2) / u2 + \ldots + fA(un) / un$,

Where fA is the membership function of A, fA: U → [0, 1], $fA(ui)$ indicates the grade of membership of ui in the fuzzy set A, $fA(ui) \in [0, 1]$, and $1 \le i \le n$, [5].

Definition 2: *Time Series*

A time series is a sequence of data points, measured typically at successive time spaced at uniform time intervals.

Definition 3: *Fuzzy Time Series*

Chronological sequences of imprecise data are considered as time series with fuzzy data. A time series with fuzzy data is referred to as fuzzy time series. [7]. Let X(t) (t = …, 0, 1, 2, …) be the universe of discourse and be a subset of R, and let fuzzy set $fi(t)$ (i = 1, 2, …) be defined in X(t). Let F(t) be a collection of $fi(t)$ (i=1, 2, …). Then, F(t) is called a fuzzy time series of X(t) (t = …, 0, 1, 2, …)[5].

Definition 4: *Time variant and Time-invariant fuzzy time series*

Let F(t) be a fuzzy time series and let R(t, t - 1) be a first-order model of F(t). If R(t, t - 1) = R(t - 1, t - 2) for any time t, then F(t) is called a time-invariant fuzzy time series.

If R(t, t - 1) is dependent on time t, that is, R(t, t-1) may be different from R(t - 1, t - 2) for any t , then F(t) is called a time-variant fuzzy time series, [5].

Definition 5: *First order model*

If F(t) is caused by F(t - 1), denoted by F(t - 1) → F(t), then this relationship can be represented by F(t) = F(t - 1) ∘ R(t, t - 1), where the symbol " ∘ " denotes the Max-Min composition operator; R(t, t – 1) is a fuzzy relation between F(t) and F(t - 1) and is called the first-order model of F(t), [5].

Definition 6*: Forecast Error*

A forecast error is the difference between the actual or real and the predicted or forecast value of a time series.

Error = Actual value - forecasted value

PROPOSED METHOD

In this section, we proposed a new method by using percentage change as the universe of discourse and mean based partitioning. The aim of this study is to propose a method that is aimed to attain better forecasting accuracy by using fuzzy time series. It should be emphasized that for forecast it uses only historical data in the numerical form (actual population) without any additional pieces of knowledge. The historical populations of Azerbaijan are shown in Table I [17].

Finally, step- by- step forecasting process looks as follows:

Step 1: Define the universe of discourse (universal set U). Partition U into equally length intervals.

Step 2: Define fuzzy sets X_i, and fuzzify actual historical data.

Step 3: Forecast and defuzzify the forecasted output.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

TABLE.1
THE HISTORICAL POPULATION OF AZERBAIJAN [17]

Year	Population
1988	6928.0
1989	7021.2
1990	7131.9
1991	7218.5
1992	7324.1
1993	7440.0
1994	7549.6
1995	7643.5
1996	7726.2
1997	7799.8
1998	7879.7
1999	7953.4
2000	8016.2
2001	8081.0

Step 1: Define the universe of discourse U and partition it into intervals of equal length. The percentage change of enrollments from year to year is given in Table II and ranges from 0.79% to 1.58%. For example, assume that the universe of discourse U= [0.7, 1.7] is partitioned into five equal intervals.

TABLE.2
THE YEAR- TO- YEAR PERCENTAGE CHANGE OF POPULATION

Year to Year	Change
1988-89	1.35%
1989-90	1.58%
1990-91	1.21%
1991-92	1.46%
1992-93	1.58%
1993-94	1.47%
1994-95	1.24%
1995-96	1.08%
1996-97	0.95%
1997-98	1.02%
1998-99	0.95%
1999-2000	0.79%
2000-2001	0.81%

Step 2: Get a mean of the original data. Get the means of frequency of each interval shown in Table III. Compare the means of original and frequency of each interval and then split the five intervals into number of sub-intervals respectively.

TABLE.3
MEANS OF ORIGINAL DATA AND FREQUENCY OF INTERVALS DATA

Interval	Number of Data
[0.7, 0.9]	2
[0.9, 1.1]	4
[1.1 1.3]	2
[1.3, 1.5]	3
[1.5, 1.7]	2

Step 3: Define each fuzzy set Xi based on the re-divided intervals and fuzzify the historical data shown in Table I, where fuzzy set Xi, denotes a linguistic value of the year to year percentage change represented by a fuzzy set.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

TABLE.4
FUZZY INTERVALS USING MEAN BASED PARTITIONING

Linguistic	Interval
X_1	[0.7, 0.8]
X_2	[0.8, 0.9]
X_3	[0.9, 0.95]
X_4	[0.95, 1]
X_5	[1, 1.05]
X_6	[1.05, 1.1]
X_7	[1.1, 1.2]
X_8	[1.2, 1.3]
X_9	[1.3, 1.367]
X_{10}	[1.367, 1.434]
X_{11}	[1.434, 1.5]
X_{12}	[1.5, 1.6]
X_{13}	[1.6, 1.7]

Step 4: Defuzzify the fuzzy data shown in Table V, using the forecasting formula [7].

$$t_j = \begin{cases} \dfrac{1+0.5}{\dfrac{1}{a_1}+\dfrac{0.5}{a_2}} & , if\ j=1, \\[2ex] \dfrac{0.5+1+0.5}{\dfrac{0.5}{a_{j-1}}+\dfrac{1}{a_j}+\dfrac{0.5}{a_{j+1}}} & , if\ 2 \le j \le n-2, \\[2ex] \dfrac{0.5+1}{\dfrac{0.5}{a_{n-1}}+\dfrac{1}{a_n}} & , if\ j=n. \end{cases}$$

Where a_{j-1} , a_j, a_{j+1} are the midpoints of the fuzzy intervals X_{j-1}, X_j, X_{j+1} respectively. t_j yields the predicted year to year percentage change of population. Use the predicted percentage on the previous year's population to determine the forecasted population. The forecasted population is given in Table V.

A COMPARISON OF DIFFERENT FORECASTING METHODS

As in [8], we use the average forecasting error rate (AFER) and mean square error (MSE) to compare the forecasting results of different forecasting methods shown in Table VI:

$AFER = (|Ai - Fi\ | \ / \ Ai) \ / \ n \times 100\%$

$$MSE = \left(\sum_{i=1 \text{ to } n} (Ai - Fi)^2 \right) / \text{n}$$

Where A_i denotes the actual population and F_i denotes the forecasting population of year i, respectively. In A. M. Abbasov & M. H. Mamedova's fuzzy time series method AFER comes to 0.13%, where as for proposed fuzzy time series method has 0.0149%. Figure 1 shows the comparison of different forecasting methods.

TABLE.5
FORECASTING RESULT OF THE PROPOSED METHOD

Year	Population (A_i)	Percentage	Fuzzy set	Predicted Percentage	Forecast (F_i)	A_i-F_i	$(A_i-F_i)^2$	$\|A_i-F_i\| / A_i$
1988	6928.0	-	-	-	-	-	-	-
1989	7021.2	1.35%	X_9	1.33%	7020.1	1.1	1.21	0.000156
1990	7131.9	1.58%	X_{12}	1.55%	7130.0	1.9	3.61	0.000266
1991	7218.5	1.21%	X_8	1.24%	7220.3	-1.8	3.24	0.000249
1992	7324.1	1.46%	X_{11}	1.47%	7324.6	-0.5	0.25	0.000068
1993	7440.0	1.58%	X_{12}	1.55%	7437.6	2.4	5.76	0.000322
1994	7549.6	1.47%	X_{11}	1.47%	7549.6	0	0	0

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

1995	7643.5	1.24%	X_8	1.24%	7643.5	0	0	0
1996	7726.2	1.08%	X_6	1.08%	7726.2	0	0	0
1997	7799.8	0.95%	X_3	0.92%	7797.3	2.5	6.25	0.000320
1998	7879.7	1.02%	X_5	1.02%	7879.7	0	0	0
1999	7953.4	0.95%	X_3	0.92%	7952.2	1.2	1.44	0.000150
2000	8016.2	0.79%	X_1	0.78%	8015.4	0.8	0.64	0.000099
2001	8081.0	0.81%	X_2	0.84%	8083.5	-2.5	6.25	0.000309
							MSE=2.2038	AFER= 0.0149%

TABLE.6
COMPARISON OF DIFFERENT FORECASTING METHODS

Year	Population	A. M. Abbasov & M. H. Mamedova [17]	Proposed Method
1988	6928.0	6926.7	-
1989	7021.2	7028.0	7020.1
1990	7131.9	7114.5	7130.0
1991	7218.5	7224.9	7220.3
1992	7324.1	7308.5	7324.6
1993	7440.0	7425.1	7437.6
1994	7549.6	7544.3	7549.6
1995	7643.5	7647.9	7643.5
1996	7726.2	7736.5	7726.2
1997	7799.8	7812.0	7797.3
1998	7879.7	7884.0	7879.7
1999	7953.4	7962.6	7952.2
2000	8016.2	8034.4	8015.4
2001	8081.0	8093.4	8083.5
AFER	-	0.13%	0.0149%
MSE	-	-	2.2038

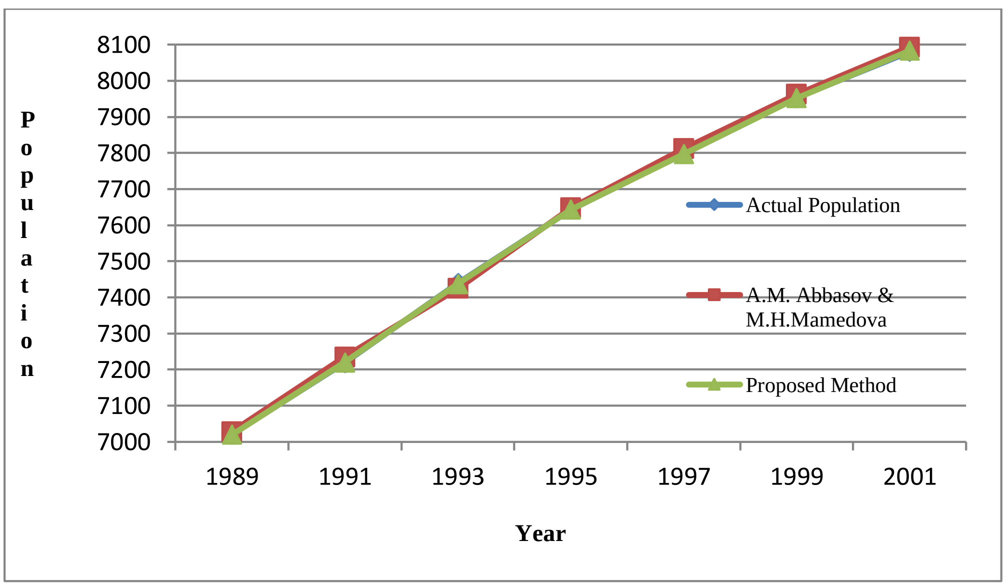

Fig1. Comparison of different forecasting methods

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

CONCLUSION

In this paper, we proposed a new method by using percentage change as the universe of discourse and mean based partitioning. From Table VI, one sees that the proposed method provide the smallest AFER and MSE. For future work, we will develop new method for forecasting data based on different intervals to get a higher forecasting accuracy.

REFERENCES

[1] S. M. Chen, Forecasting enrollments based on fuzzy time series, Fuzzy Sets and Systems, vol. 81, pp. 311-319, 1996.
[2] J. R. Hwang, S. M. Chen, C. H. Lee, Handling forecasting problems using fuzzy time series, Fuzzy Sets and Systems, vol. 100, pp. 217-228, 1998.
[3] K. Huarng, Effective length of intervals to improve forecasting in fuzzy time series, Fuzzy Sets and systems, vol. 12, pp. 387-394, 2001.
[4] S. M. Chen, Forecasting enrollments based on high-order fuzzy time series, Cybernetics and Systems: An International Journal, vol. 33, pp. 1-16, 2002.
[5] S. M. Chen and Hsu, A new method to forecasting enrollments using fuzzy time series, International Journal of Applied Science and Engineering, vol. 2, no. 3, pp. 234-244, 2004.
[6] S. Melike, K.Y. Degtiarev, Forecasting enrollments model based on first- order fuzzy time series, Proceedings of World Academy of Science, Engineering and Technology, vol. 1, pp. 132-135, 2005.
[7] 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. 333-338, 2007.
[8] Stevenson, 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.
[9] Z. Ismail, R. Efendi, Enrollment forecasting based on modified weight fuzzy time series, Journal of Artificial Intelligence, vol. 4(1), pp. 110-118, 2011.
[10] Arutehelvan G., Sivatsa S. K, Jaganathan R., Inaccuracy minimization by partitioning fuzzy data sets- validation of an analytical methodology, (IJCSIS) International Journal of Computer Science and Information Security, vol. 8, no. 1, 2010.
[11] Q. Song, B. S. Chissom, Fuzzy time series and its models, Fuzzy Sets and systems, vol. 54, pp. 269-277, 1993.
[12] Q. Song, B. S. Chissom, Forecasting enrollments with fuzzy time series: Part I, Fuzzy Sets and systems, vol. 54, pp. 1-9, 1993.
[13] Q. Song, B. S. Chissom, Forecasting enrollments with fuzzy time series: Part II, Fuzzy Sets and systems, vol. 62: pp. 1-8, 1994.
[14] T. A. Jilani, S. M. A. Burney, M-factor high order fuzzy time series forecasting for road accident data, In IEEE-IFSA 2007, World Congress, Cancun, Mexico, June 18-21, Forthcoming in Book series Advances in Soft Computing, Springer-Verlag, 2007.
[15] 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.
[16] R. K. Shrivastava, Fuzzy- Neural Techniques for short term forecast of food grain production, International Journal of Information Technology & Knowledge Management, vol. 4, no. 2, pp. 385-390, 2011.
[17] A. M. Abbasov, M. H. Mamedova, Application of fuzzy time series to population forecasting, CORP, pp. 545-552, 2003.
[18] Liang Zhao, A new method to predict enrollments based on fuzzy time series, proceeding of the 8th World Congress on Intelligent Control and Automation (IEEE), pp. 3201-3206, 2010.
[19] W. K. Wong, E. Bai, A. W. Ching Chu, Adaptive time-variant models for fuzzy time series forecasting, IEEE Transactions on Systems, Man, And Cybernatics- Part B: Cybernatics, vol. 40, no. 6, pp. 1531- 1542, 2010.