BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **18**, No 1
Sofia • 2018 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2018-0001

A Unique Computational Method for Constructing Intervals in Fuzzy Time Series Forecasting

Shilpa Jain[1], *Prakash C. Mathpal*[2], *Dinesh Bisht*[3], *Phool Singh*[1]

[1]*School of Engineering and Technology, NorthCap University, 122017 Gurugram, India*
[2]*Department of Mathematics, Govt. P.G. College, Berinag, 262531 Pithoragarh, India*
[3]*Department of Mathematics, Jaypee Institute of Information Technology, 201304 Noida, India*
E-mails: *shilpajain1606@gmail.com, mathpal.p@gmail.com, drbisht.math@gmail.com, phool.singh24@gmail.com*

Abstract**:** *This research article suggests a computational method for constructing fuzzy sets in absence of expert knowledge. This method uses concepts of central tendencies mean and variance. This study gives a solution to the critical issue in designing of fuzzy systems, number of fuzzy sets. Proposed computational method helps in finding intervals and thereby fuzzy sets for fuzzy time series forecasting. Proposed computational method is implemented on the authentic data for the enrolments of University of Alabama, which is considered as benchmark problem in the field of fuzzy time series. The forecasted values are compared with the results of other methods to state its supremacy. Projected computational method along with Gaussian membership function gave promising results over other methods for fuzzy time series for the above said benchmark data.*

Keywords**:** *Fuzzy logic, central tendencies, membership function, prediction, fuzzy time series.*

1. Introduction

Initially fuzzy logic found its application in engineering systems, commercial products and gradually non-engineering applications like medicine, social science etc. were added to its wide range of domains. The notion of fuzzy set and linguistic variable was instigated by Prof. L. A. Zadeh [1, 2]. Fuzzy set is considered a more general concept of the crisp classical set. In a fuzzy set, an element can be a part of the set with some membership value (i.e., degree of belongingness). Each fuzzy set represents a linguistic label. For example, 1 to 15th day of a month has been found to have hot temperature can be represented as

$$H = 0.02/1 + 0.1/2 + 0.5/3 + 0.4/4 + 0.55/5 + 0.7/6 + 0.875/7 + 1.0/8 + 0.875/9 + \\ + 0.7/10 + 0.4/11 + 0.4/12 + 0.3/13 + 0.6/14 + 0.4/15.$$

Hot is a linguistic label with prior knowledge of intervals (i.e. each day is considered one interval) but in absence of domain knowledge division into intervals and linguistic label is a tedious task. Fuzzy logic found its applications in different fields successfully [3-5].A lot of literature is available on fuzzy partitioning. Some popular methods on shared partitioning are given in [3], where choice of number of fuzzy sets depends on user. Further some authors proposed deriving K_i (Number of fuzzy sets) from data [6]. Researchers are also using fuzzy clustering for partitioning [7-10]. Fuzzy time series proposed in 1993 by Song and Chissom [11, 12] include various approaches and procedures developed thereafter for forecasting fuzzy time series. It has been used for forecasting in various data, some of them are treated as benchmark data sets[11-17]. Moreover, hybrid methods are used nowadays which involve evolutionary computational techniques like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA), etc., along with fuzzy for their improved partitioning and prediction accuracy [3, 13, 18-20]. In 2006 Chen and Chung [13] forecasted enrolments of students using genetic algorithms and fuzzy time series. The suggested method adjusts the length of every interval in the universe of discourse for predicting the enrolments of University of Alabama. In 2012 Mahnam and Ghomi [21] proposed a PSO algorithm for forecasting based on time variant fuzzy time series in which length of each interval in the universe of discourse and degree of membership value are determined simultaneously. Various researches have been done in the field of fuzziness [22-26]. Partition of universe of discourse is still a matter of guess and needed expert knowledge. This article deals with the method to partition and making them gaussian fuzzy sets.

2. Need of computational method

Whenever we start implementing fuzzy logic on any practical problem two most trivial questions that come to our mind are number of fuzzy sets and membership functions to use. Both are answered in this research article. In this research article, we have suggested a method based on central tendencies to state the intervals for construction of fuzzy systems, instead of going for computationally expensive techniques. Our second objective in this research article is to help the designer in choosing and applying membership functions.

3. Advantages of proposed method

Number of Fuzzy sets is a debatable issue in Fuzzy literature. An easy solution to it is:

1. Instead of Computationally expensive techniques, this method is based on central tendencies.
2. Expert of domain knowledge is not required, naïve user can fuzzify data.
3. Implementation is easy and fast.
4. Once the users have identified number of fuzzy sets, another big question which membership function should be used is also answered in this research article.

4. Methodology

In this segment we propose the complete procedure of the suggested approach of fuzzifying intervals of fuzzy sets and identifying suitable membership functions.

Step 1. Specify the Universe of discourse X, based on the range of available data.

Step 2. Divide the Universe of discourse into intervals as described below-

Step 2.1. Calculate mean (μ) and standard deviation (σ) of the data of each input and output variable.

Step 2.2. Calculate n_{right} and n_{left} to find number of points to the right and left of mean (μ) which is treated as center of intervals (accuracy up to two decimal places):

(1) $$n_{\text{right}} = \frac{2(P_{\text{right}} - \mu)}{\sigma},$$

(2) $$n_{\text{left}} = \frac{2(\mu - P_{\text{right}})}{\sigma}.$$

Generalized fuzzy intervals are presented in Fig. 1.

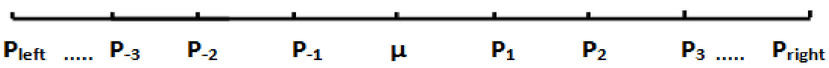

Fig. 1. Generalized fuzzy intervals

Step 2.3. Let the mean μ be treated at center of intervals. Points to the right of the mean are calculated as

(3) $$P_n = \mu + n\frac{\sigma}{2}, \quad n = 1, 2, ..., n_{\text{right}}.$$

Similarly points to the left are calculated as

(4) $$P_{-n} = \mu - n\frac{\sigma}{2}, \quad n = 1, 2, ..., n_{\text{left}}.$$

Step 2.4. Equation (5) is used to calculate total number of fuzzy sets

(5) $$n_{\text{tot}} = n_{\text{right}} + n_{\text{left}} - 1.$$

Step 3. Divide the Universe of discourse X at the intervals

$u_{-n} = [P_{-n-2}, P_{\text{left}}]$, ..., $u_{-4}=[P_{-4}, P_{-2}]$, $u_{-3}=[P_{-3}, P_{-1}]$, $u_{-2}=[P_{-2}, \mu]$, $u_{-1}=[P_{-1}, P_1]$, $u_0=[\mu, P_2]$, $u_1=[P_1, P_3]$, $u_2=[P_2, P_4]$, $u_3=[P_3, P_5]$, ..., $u_n= [P_{n-2}, P_{\text{right}}]$.

Let the fuzzy sets A_1, A_2, ..., A_n on universe of discourse, having linguistic values be assigned to these intervals:

For instance, A_1 = EL (Extremely Low) is assigned to u_{-n} and A_n = EH (Extremely High) to u_n and in between other linguistic values can be assigned to other fuzzy intervals, respectively.

Step 4. Construction of membership function

Gaussian membership function: Fuzzy sets are constructed using intervals. To apply Gauss membership for each fuzzy set we require two parameters, i.e., mean and standard deviation as follows:

(i) All values of data lying between upper and lower boundary of the interval are clipped to calculate the mean (μ) of fuzzy sets.

(ii) Standard deviation is calculated as $\sigma = \frac{\text{U.L.} - \mu}{3}$. (Assuming data are normally distributed with accuracy of α level), where μ and σ are mean and standard deviation; U.L. is upper limit of fuzzy sets as depicted in Fig. 2.

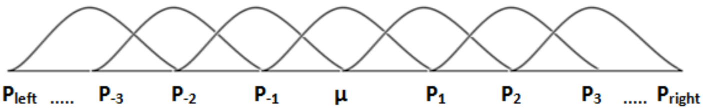

Fig. 2. Gaussian membership function

Step 5. Data are used to construct the rule base. Rule base is constructed from training data using Wang and Mendel [27] method.

Step 6. Evaluation is done using Mamdani Fuzzy Inference System which uses Gaussian membership functions, the min intersection operator and correlation product as inference procedure. Defuzzification technique used is centroid method.

5. Simulation: Forecasting for enrollments of Alabama's university

Here, in this research article we suggest a method to determine fuzzy intervals and thereby fuzzy sets. Examples from totally different fields, from hydrology which is a continuous data and from fuzzy time series, University of Alabama enrolment problem which is discrete data are used to test the results. An instance to illustrate this concept, we have used data of Alabama University, which was used initially by Song and Chissom [11].

Step 1. Specify the Universe of discourse X, based on the range of available data (i.e., 13,0000 – 20,000).

Step 2. Divide the Universe of discourse into intervals as described below Fig. 3.

Step 2.1. Calculate mean μ and standard deviation σ of the data of input are 16,194.18 and 1,816.49, respectively.

Step 2.2. Calculate n_{right} and n_{left} to find number of points to the right and left of mean μ, which is treated as center of intervals (Fig. 3),

(6) $$n_{\text{right}} = \frac{2(P_{\text{right}} - \mu)}{\sigma} = 4,$$

(7) $$n_{\text{left}} = \frac{2(\mu - P_{\text{right}})}{\sigma} = 4.$$

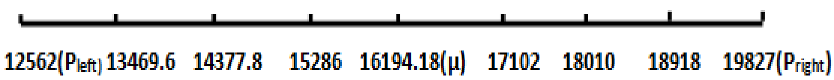

Fig. 3. Fuzzy intervals

Step 2.3. Let mean μ be treated at center of intervals. Points to the right of the mean are calculated using Equation (3):

P_1=17,102, P_2=18,010, P_3=18,918... P_{right}=19827.

Similarly points to the left are calculated using Equation (4):

P_{-1}=15,286, P_{-2}=14,377.8, P_{-3}=13,469.6, ..., P_{left}=12,562.

Step 2.4. To calculate total number of fuzzy sets, Equation (5) is used.

Step 3. Divide the Universe of discourse X at the intervals

x_1=[12,562, 14,377.8], x_2=[13,469.6, 15,286], x_3=[14,377.8, 16,194.18], x_4=[15,286, 17,102], x_5=[16,194.18, 18,010], x_6=[17,102, 18,918], x_7=[18,010, 19,827].

Let the Fuzzy sets Q_1, Q_2, ..., Q_7 on universe of discourse, having linguistic values as given below:

Q_1 = EL (Extremely Low),
Q_2 = VL (Very Low),
Q_3 = L (Low),
Q_4 = M (Medium),
Q_5 = H (High),
Q_6 = VH (Very High),
Q_7 = EH (Extremely High).

Step 4. Using above stated method for predicting intervals for forecasting enrolments of University of Alabama. Mean and standard deviation of data given in Table 1 are 16,194.8 and 1,816.49 respectively. Fuzzy sets are constructed using intervals with the help of Fig. 4 in following steps.

Step 4.1. To construct fuzzy set 1, all data values from data lying between 12,562 and 14,377.8 are 13,055, 13,563, and 13,867 so the mean is calculated as 13,495.

Step 4.2. Standard deviation is 294.2667 is calculated as stated in Step 4 in methodology.

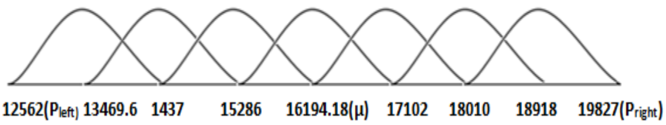

Fig. 4. Gaussian membership function

Using above calculated mean and standard deviation Gaussian membership is constructed; similarly other six fuzzy sets are constructed. One important consideration is of a fuzzy set where no data lies between these intervals for constructing the set. For this the mean could be taken as mean of upper and lower bound for fuzzy set and standard deviation is given by minimum standard deviation of all fuzzy sets for a given input.

Step 5. Rule base is constructed as follows, from training data using of Wang and Mendel [27] method:

$$Q_1 \to Q_1,\ Q_1 \to Q_2,$$
$$Q_2 \to Q_2,\ Q_2 \to Q_3,$$
$$Q_3 \to Q_3,\ Q_3 \to Q_4,$$
$$Q_4 \to Q_3,\ Q_4 \to Q_4,\ Q_4 \to Q_5,$$
$$Q_5 \to Q_5,\ Q_5 \to Q_4,\ Q_5 \to Q_6,$$
$$Q_6 \to Q_7,$$
$$Q_7 \to Q_7.$$

Step 6. Centroid method is used to convert hence find fuzzified output into crisp value.

6. Results and comparison

Table 1 shows MSE using Gaussian membership function is 258,200. The values of R and R^2 for Gaussian membership function are 0.985541 and 0.971 respectively. In order to validate our findings, we compared it to eminent researchers work [11, 12, 14, 15, 23] and results stated in Table 2.

Table 1. Results for data of Alabama University

Function	MSE	R	R^2
Gaussian membership function	258,200	0.985541	0.971

Table 2. Enrollments of Alabama University for different methods

Year	Observed	SC_time variant [12]	SC_time invariant [11]	Chen [28] method	Huarng [14] heuristic	Lee and Chou [15]	Proposed
1971	13,055						14,011
1972	13,563						14,091
1973	13,867						14,402
1974	14,696		14,000	14,000	14,000	14,568	15,145
1975	15,460	14,700	15,500	15,500	15,500	15,654	15,918
1976	15,311	14,800	16,000	16,000	15,500	15,654	15,768
1977	15,603	15,400	16,000	16,000	16,000	15,654	16,052
1978	15,861	15,500	16,000	16,000	16,000	15,654	16,194
1979	16,807	15,500	16,000	16,000	16,000	16,197	16,827
1980	16,919	16,800	16,813	16,833	17,500	17,283	16,932
1981	16,388	16,200	16,813	16,833	16,000	17,283	16,374
1982	15,433	16,400	16,789	16,833	16,000	16,197	15,892
1983	15,497	16,800	16,000	16,000	16,000	15,654	15,952
1984	15,145	16,400	16,000	16,000	15,500	15,654	15,594
1985	15,163	15,500	16,000	16,000	16,000	15,654	15,612
1986	15,984	15,500	16,000	16,000	16,000	15,654	16,194
1987	16,859	15,500	16,000	16,000	16,000	16,197	16,877
1988	18,150	16,800	16,813	16,833	17,500	17,283	18,918
1989	18,970	19,300	19,000	19,000	19,000	18,369	18,918
1990	19,328	17,800	19,000	19,000	19,000	19,454	18,918
1991	19,337	19,300	19,000	19,000	19,500	19,454	18,918
1992	18,876	19,600		19,000	19,000		18,918

Table 3 demonstrates MSE obtained by various methods whereas Fig. 5 gives a comparative plot.

Table 3. Comparison of MSE for forecasting enrolments for different methods

Methods	MSE
SC_time invariant [11]	458,437.5
SC_time variant [12]	775,686.8
Chen [28] method	439,420.8
Huarng [14] heuristic	239,483.1
Lee and Chou [15]	240,047
Proposed computational method	157,190

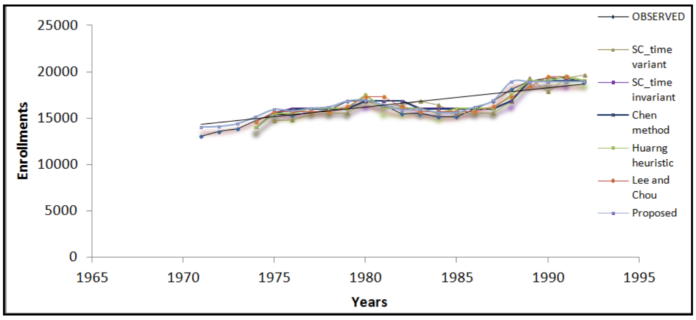

Fig. 5. Comparative graph

7. Summary and conclusion

In this research article, we have proposed a computational method for stating intervals and membership functions for fuzzy sets. In absence of expert knowledge based on available data, one can make fuzzy intervals thereafter fuzzy sets. To illustrate the method, it is implemented on University of Alabama student enrollment data. Key points summarized in our research article are:

1. Instead of Computationally expensive techniques, this method is based on central tendencies.
2. Using methodology stated in this research article one can find intervals and thereby fuzzy sets.
3. Expert of domain knowledge is not required, naïve user can fuzzily data.
4. Implementation is easy and fast.
5. Gaussian membership functions have been applied along with our computational method to give best of results.
6. The proposed method can be applied to various fields of science and management.

References

1. Z a d e h, L. A. Fuzzy Sets. – Information and Control, Vol. **8**, 1965, No 3, pp. 338-353.
2. Z a d e h, L. A. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning. I. – Information Sciences, Vol. **8**, 1975, No 3, pp. 199-249.
3. J a i n, S., D. B i s h t. Genetic Algorithms Based Fuzzy Time Series Prediction for Water Table Elevation Fluctuation. – Aloy Journal of Soft Computing and Applications, Vol. **3**, 2015, pp. 14-23.
4. B i s h t, D., S. J a i n, M. M. R a j u. Prediction of Water Table Elevation Fluctuation Through Fuzzy Logic and Artificial Neural Networks. – International Journal of Advanced Science and Technology, Vol. **51**, 2013, pp. 107-120.
5. B i s h t, D. C. S., M. R a j u, M. J o s h i. Simulation of Water Table Elevation Fluctuation Using Fuzzy-Logic and ANFIS. – Computer Modelling and New Technologies, Vol. **13**, 2009, No 2, pp. 16-23.
6. B o r t o l e t, P. Modelisation et Commande Multivariables Floues: Application a la Commande D’un Moteur Thermique. PhD Diss., Toulouse, INSA, 1998.
7. D u n n, J. C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. – Journal of Cybernetics, Vol. **3**, 1973, No 3, pp. 32-57.

8. K r i s h n a p u r a m, R., J. M. K e l l e r. A Possibilistic Approach to Clustering. – IEEE Transactions on Fuzzy Systems, Vol. **1**, 1993, No 2, pp. 98-110.
9. K r i s h n a p u r a m, R., J. K i m. A Note on the Gustafson-Kessel and Adaptive Fuzzy Clustering Algorithms. – IEEE Transactions on Fuzzy Systems, Vol. **7**, 1999, No 4, pp. 453-461.
10. Y a g e r, R. R., D. P. F i l e v. Generation of Fuzzy Rules by Mountain Clustering. – Journal of Intelligent & Fuzzy Systems, Vol. **2**, 1994, No 3, pp. 209-219.
11. S o n g, Q., B. S. C h i s s o m. Forecasting Enrollments with Fuzzy Time Series. Part I. – Fuzzy Sets and Systems, Vol. **54**, 1993, No 1, pp. 1-9.
12. S o n g, Q., B. S. C h i s s o m. Forecasting Enrollments with Fuzzy Time Series. Part II. – Fuzzy Sets and Systems, Vol. **62**, 1994, No 1, pp. 1-8.
13. C h e n, S. M., N. Y. C h u n g. Forecasting Enrollments of Students by Using Fuzzy Time Series and Genetic Algorithms. – International Journal of Information and Management Sciences, Vol. **17**, 2006, No 3, pp. 1-17.
14. H u a r n g, K. Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series. – Fuzzy Sets and Systems, Vol. **123**, 2001, No 3, pp. 387-394.
15. L e e, H. S., M. T. C h o u. Fuzzy Forecasting Based on Fuzzy Time Series. – International Journal of Computer Mathematics, Vol. **81**, 2014, No 7, pp. 781-789.
16. Y e, F., L. Z h a n g, D. Z h a n g, H. F u j i t a, Z. G o n g. A Novel Forecasting Method Based on Multi-Order Fuzzy Time Series and Technical Analysis. – Information Sciences, Vol. **367**, 2006, No C, pp. 41-57.
17. H w a n g, J. R., S. M. C h e n, C. H. L e e. Handling Forecasting Problems Using Fuzzy Time Series. – Fuzzy Sets and Systems, Vol. **100**, 1998, No 1-3, pp. 217-228.
18. C r u z-V e g a, I., C. A. R. G a r c í a, P. G. G i l, J. M. R. C o r t é s, J. de J. R. M a g d a l e n o. Genetic Algorithms Based on a Granular Surrogate Model and Fuzzy Aptitude Functions. – In: 2016 IEEE Congress on Evolutionary Computation (CEC'16), 2016, pp. 2122-2128.
19. E ğ r i o ğ l u, E. A New Time-Invariant Fuzzy Time Series Forecasting Method Based on Genetic Algorithm. – Advances in Fuzzy Systems, Vol. **2012**, 2012, p. 2.
20. H s u, L. Y., S. J. H o r n g, T. W. K a o, Y. H. C h e n, R. S. R u n, R. J. C h e n, J. L. L a i, I. H. K u o. Temperature Prediction and TAIFEX Forecasting Based on Fuzzy Relationships and MTPSO Techniques. – Expert Systems with Applications, Vol. **37**, 2010, No 4, pp. 2756-2770.
21. M a h n a m, M., S. G h o m i. A Particle Swarm Optimization Algorithm for Forecasting Based on Time Variant Fuzzy Time Series. – International Journal of Industrial Engineering & Production Research, Vol. **23**, 2012, No 4, pp. 269-276.
22. A n g e l o v a, V. Investigations in the Ares of Soft Computing. Targeted State of the Art Report. – Cybernetics and Information Technologies, Vol. **9**, 2009, No 1, pp. 18-24.
23. I l i e v a, G. A Fuzzy Approach for Bidding Strategy Selection. – Cybernetics and Information Technologies, Vol. **12**, 2012, No 1, pp. 61-69.
24. I l i e v a, G. TOPSIS Modification with Interval Type-2 Fuzzy Numbers. – Cybernetics and Information Technologies, Vol. **16**, 2016, No 2, pp. 60-68.
25. I l i e v a, G. Group Decision Analysis with Interval Type-2 Fuzzy Numbers. – Cybernetics and Information Technologies, Vol. **17**, 2017, No 1, pp. 31-44.
26. K h e r a b a d i, H. A., S. E. M o o d, M. M. J a v i d i. Mutation: A New Operator in Gravitational Search Algorithm Using Fuzzy Controller. – Cybernetics and Information Technologies, Vol. **17**, 2017, No 1, pp. 72-86.
27. W a n g, L. X., J. M. M e n d e l. Generating Fuzzy Rules by Learning from Examples. – IEEE Transactions on Systems, Man, and Cybernetics, Vol. **22**, 1992, No 6, pp. 1414-1427.
28. C h e n, S. M. Forecasting Enrollments Based on Fuzzy Time Series. – Fuzzy Sets and Systems, Vol. **81**, 1996, No 3, pp. 311-319.

Received 17.10.2017; *Accepted 20.11.2017*