Fuzzy Metric Approach for Fuzzy Time Series Forecasting based on Frequency Density Based Partitioning

Tahseen Ahmed Jilani, Syed Muhammad Aqil Burney, and Cemal Ardil

***Abstract*—**In the last 15 years, a number of methods have been proposed for forecasting based on fuzzy time series. Most of the fuzzy time series methods are presented for forecasting of enrollments at the University of Alabama. However, the forecasting accuracy rates of the existing methods are not good enough. In this paper, we compared our proposed new method of fuzzy time series forecasting with existing methods. Our method is based on frequency density based partitioning of the historical enrollment data. The proposed method belongs to the kth order and time-variant methods. The proposed method can get the best forecasting accuracy rate for forecasting enrollments than the existing methods.

***Keywords*—**Fuzzy logical groups, fuzzified enrollments, fuzzy sets, fuzzy time series.

I. Introduction

It is obvious that forecasting activities play an important role in our daily life. During last few decades, various approaches have been developed for time series forecasting. Among them ARMA models and Box-Jenkins model building approaches are highly famous. But the classical time series methods can not deal with forecasting problems in which the values of time series are linguistic terms represented by fuzzy sets [11], [23]. Therefore, Song and Chissom [18] presented the theory of fuzzy time series to overcome this drawback of the classical time series methods. Based on the theory of fuzzy time series, Song et al. presented some forecasting methods [16], [18], [19], [20] to forecast the enrollments of the University of Alabama. In [1] Chen and Hsu and in [2], Chen presented a method to forecast the enrollments of the University of Alabama based on fuzzy time series. It has the advantage of reducing the calculation, time and simplifying the calculation process. In [8], Hwang, Chen and Lee used the differences of the enrollments to present a method to forecast the enrollments of the University of Alabama based on fuzzy time series. In [5] and [6], Huang used simplified calculations with the addition of heuristic rules to forecast the enrollments using [2]. In [4], Chen presented a forecasting method based on high-order fuzzy time series for forecasting the enrollments of the University of Alabama. In [3], Chen and Hwang presented a method based on fuzzy time series to forecast the daily temperature. In [15], Melike and Konstsntin presented a new first order time series model for forecasting enrollments of the University of Alabama. In [14], Li and Kozma presented a dynamic neural network method for time series prediction using the KIII model. In [21], Su and Li presented a method for fusing global and local information in predicting time series based on neural networks. In [22], Sullivan and Woodall reviewed the first-order time-variant fuzzy time series model and the first-order time-invariant fuzzy time series model presented by Song and Chissom [18], where their models are compared with each other and with a time-variant Markov model using linguistic labels with probability distributions. In [13], Lee, Wang and Chen presented two factor high order fuzzy time series for forecasting daily temperature in Taipei and TAIFEX. In [9], Jilani and Burney and in [10], Jilani, Burney and Ardil presented new fuzzy metrics for high order multivariate fuzzy time series forecasting for car road accident casualties in Belgium.

In this paper, we present a comparison of our proposed method and existing fuzzy time series forecasting methods to forecast the enrollments of the University of Alabama. Our proposed method belongs to the class of k-step first-order univariate time-variant method. The proposed method gives the best forecasting accuracy rate for forecasting enrollments when compared with existing methods. The rest of this paper is organized as follows. In Section 2, we briefly review some basic concepts of fuzzy time series. In Section 3, we present our method of fuzzy forecasting based on frequency density based partitioning of the enrollment data. In Section 4, we compared the forecasting results of the proposed method with the existing methods. The conclusions are discussed in Section 5.

II. Some Basic Concepts of Fuzzy Time Series

There are number of definitions for fuzzy time series.

Manuscript received June 15, 2007. This work was supported in part by the Higher Education Commission of Pakistan.

T. A. Jilani is with the Department of Computer Science, University of Karachi, Karachi-75270, Pakistan (phone: +92-21-9363131-37; fax: 92-21-9261340; e-mail: tahseenjilani@ieee.org).

S. M. A. Burney is Distinguished Professor in the Department of Computer Science, University of Karachi, Pakistan (e-mail: burney@computer.org).

C. Ardil is with National Academy of Aviation, AZ1045, Baku, Azerbaijan, Bina, 25th km, NAA (e-mail: cemalardil@gmail.com).

Definition 1: Imprecise data at equally spaced discrete time points are modeled as fuzzy variables. The set of this discrete fuzzy data forms a fuzzy time series.

Definition 2: 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.

Definition 3: Let $Y(t),(t=...,0,1,2,...)$ be the universe of discourse and $Y(t)\subseteq R$. Assume that $f_i(t),\ i=1,2,...$ is defined in the universe of discourse $Y(t)$ and $F(t)$ is a collection of $f(t_i),(i=...,0,1,2,...)$, then $F(t)$ is called a fuzzy time series of $Y(t),\ i=1,2,....$ Using fuzzy relation, we define $F(t)=F(t-1)\circ R(t,t-1)$ where $R(t,t-1)$ is a fuzzy relation and "$\circ$" is the max–min composition operator, then $F(t)$ is caused by $F(t-1)$ where $F(t)$ and $F(t-1)$ are fuzzy sets.

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. In [19], Song et al. proposed the time-variant fuzzy time-series model and forecasted the enrollments of the University of Alabama based on the model.

III. A New Method for Forecasting Enrollments Using Fuzzy Time Series

In this section, we present our method to forecast the enrollments of the University of Alabama based on fuzzy time series based on [9] and [10]. The historical enrollments of the University of Alabama are shown in Table I, [19].

Firstly, based on [1], we defined the partition the universe of discourse into equal length intervals. Then based on frequency density portioning, we redefine the intervals. After this, define some membership function for each interval of the historical enrollment data to obtain fuzzy enrollments to form a fuzzy time series. Then, it establishes fuzzy logical relationships (FLRs) based on the fuzzified enrollments in Table IV. Finally, it uses our proposed method. The proposed method bases on Hsu and Chen approach, [7] of partitioning universe of discourse are as follows:

Step 1: Define the universe of discourse U and partition it into several even and equal length intervals $u_1,u_2,...,u_n$. For example, assume that the universe of discourse U= [13000, 20000] is partitioned into seven even and equal length intervals.

TABLE I
THE HISTORICAL ENROLLMENTS OF THE UNIVERSITY OF ALABAMA [19]

YEAR	ENROLLMENTS	YEAR	ENROLLMENTS
1972	13055	1982	15433
1972	13563	1983	15497
1973	13847	1984	15145
1974	14696	1985	15163
1975	15460	1986	15984
1976	15311	1987	16859
1977	15603	1988	18150
1978	15861	1989	18970
1979	16807	1990	19328
1980	16919	1991	19337
1981	16388	1992	18876

Step 2: Get a weighted aggregation [24] of the fuzzy distribution of the historical enrollments in each interval. Sort the intervals based on the number of historical enrollment data in each interval from the highest to the lowest as given in [1]. Find the interval having the largest number of historical enrollment data and divide it into four sub-intervals of equal length. Find the interval having the second largest number of historical enrollment data and divide it into three sub-intervals of equal length. Find the interval having the third largest number of historical enrollment data and divide it into two sub-intervals of equal length. Find the interval with the fourth largest number of historical enrollment data and let the length of this interval remain unchanged. If there are no data distributed in an interval then discard this interval. For example, the distributions of the historical enrollment data in different intervals are summarized as shown in Table II, [7].

After executing this step, the universe of discourse [13000, 20000] is re-divided into the following intervals [7], see Table III.

Step 3: Define each fuzzy set A_i based on the re-divided intervals and fuzzify the historical enrollments shown in Table I, where fuzzy set A_i denotes a linguistic value of the enrollments represented by a fuzzy set. We have used triangular membership function to define the fuzzy sets A_i [10]. The reason for fuzzifying the historical enrollments into fuzzified enrollments is to translate crisp values into fuzzy sets to get a fuzzy time series.

Step 4: Establish fuzzy logical relationships based on the fuzzified enrollments where the fuzzy logical relationship "$A_p,A_q,A_r \to A_s$" denotes that "if the fuzzified enrollments of year p, q and r are A_p,A_q and A_r respectively, then the fuzzified enrollments of year (r) is A_r".

TABLE II
THE FREQUENCY DENSITY BASED DISTRIBUTION OF THE HISTORICAL ENROLLMENT DATA [7]

Intervals	Number of historical enrollment data
[13000, 14000]	3
[14000, 15000]	1
[15000, 16000]	9
[16000, 17000]	4
[17000, 18000]	0
[18000, 19000]	3
[19000, 20000]	2

TABLE III
FUZZY INTERVALS USING FREQUENCY DENSITY BASED PARTITIONING

Linguistic	Intervals
u1	[13000, 13500]
u2	[13500, 14000]
u3	[14000, 15000]
u4	[15000, 15250]
u5	[15250,15500]
u6	[15500,15750]
u7	[15750,16000]
u8	[16000,16333]
u9	[16333,16667]
u10	[16667,17000]
u11	[18000,18500]
u12	[18500,19000]
u13	[19000,20000]

Intervals

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

where a_{j-1}, a_j, a_{j+1} are the mid points of the fuzzy intervals A_{j-1}, A_j, A_{j+1} respectively. Based on the fuzzify historical enrollments obtained in Step 3, we can get the fuzzy logical relationship group (FLGR) as shown in Table IV.

Divide each interval derived in Step 2 into subintervals of equal length with respect to the corresponding frequency density of the interval. We have assumed thirteen partitions of the universe of discourse of the main factor fuzzy time series. Assuming that $0 \ne A_i, \forall A_i, i = 1,2,...,13$. The proposed method satisfies the following axioms:

TABLE IV
THIRD-ORDER FUZZY LOGICAL RELATIONSHIP GROUPS [9]

Group 1:	$X_2, X_2, X_3 \rightarrow X_5$	Group 10:	$X_9, X_5, X_5 \rightarrow X_4$
Group 2:	$X_2, X_3, X_5 \rightarrow X_5$	Group 11:	$X_5, X_5, X_4 \rightarrow X_4$
Group 3:	$X_3, X_5, X_5 \rightarrow X_6$	Group 12:	$X_5, X_4, X_4 \rightarrow X_7$
Group 4:	$X_5, X_5, X_6 \rightarrow X_7$	Group 13:	$X_4, X_4, X_7 \rightarrow X_{10}$
Group 5:	$X_5, X_6, X_7 \rightarrow X_{10}$	Group 14:	$X_4, X_7, X_{10} \rightarrow X_{11}$
Group 6:	$X_6, X_7, X_{10} \rightarrow X_{10}$	Group 15:	$X_7, X_{10}, X_{11} \rightarrow X_{12}$
Group 7:	$X_7, X_{10}, X_{10} \rightarrow X_9$	Group 16:	$X_{10}, X_{11}, X_{12} \rightarrow X_{13}$
Group 8:	$X_{10}, X_{10}, X_9 \rightarrow X_5$	Group 17:	$X_{11}, X_{12}, X_{13} \rightarrow X_{13}$
Group 9:	$X_{10}, X_9, X_5 \rightarrow X_5$	Group 18:	$X_{12}, X_{13}, X_{13} \rightarrow X_{12}$

Axiom 1: $t(\underline{0}) = 0\ and\ t(\underline{1}) = 1$ (Boundary Condition)

Axiom 2: $t^{\alpha_i}(a_1, a_2, ..., a_n) \le t^{\alpha_i}(b_1, b_2, ..., b_n)$ provided $a_i \le b_i, i = 1,2,...,n$ (Monotonicity)

Axiom 3: $t^{\alpha}(a_1, a_2, ..., a_n)$ is continuous

Axiom 4: $(A_k)_{\min} \le t^{\alpha}(a) \le (A_k)_{\max}$; $k = 1,2,...,8$ (Symmetry)

Axiom 5: $t^{\alpha}(a_1, a_2, ..., a_n) = a, \forall a \in [0,1]$ (Idempotency)

In the next section, we have given comparison of different fuzzy time series forecasting methods.

IV. A COMPARISON OF DIFFERENT FORECASTING METHODS

In the following, Table V summarizes the forecasting results of the proposed method from 1972 to 1992, where the universe of discourse is divided into thirteen intervals based on frequency density based partitioning. In the following, we use the average forecasting error rate (AFER) and mean square error (MSE) to compare the forecasting results of different forecasting methods, where A_i denotes the actual enrollment and F_i denotes the forecasting enrollment of year i, respectively.

$$AFER = \frac{|A_i - F_i| / A_i}{n} \times 100\%$$

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

In Table VI, we compare the forecasting results of the proposed method with that of the existing methods. From Table III, we can see that when the number of intervals in the universe of discourse is thirteen and the intervals are sub-partitioned based on frequency density, the proposed method shows smallest values of the MSE and AFER of the forecasting results as compared to other methods of fuzzy time series forecasting. That is, the proposed method can get a higher forecasting accuracy rate for forecasting enrollments than the existing methods.

TABLE V
ACTUAL ENROLLMENTS AND FORECASTED ENROLLMENTS OF THE UNIVERSITY OF ALABAMA BASED ON FREQUENCY DENSITY BASED PARTITIONING

Year	Enrollments (A_i)	Fuzzy Rule	FLRG	Forecast (F_i)	$A_i - F_i$	$(A_i - F_i)^2$	$\frac{\lvert A_i - F_i \rvert}{A_i}$
1971	13055	A_1	A_1, A_2	13579	-524	274778	0.040153
1972	13563	A_2	A_1, A_2, A_3	13798	235	55344	0.017345
1973	13847	A_2	A_1, A_2, A_3	13798	-49	2376	0.003520
1974	14696	A_3	A_2, A_3, A_4	14452	-244	59427	0.016588
1975	15460	A_5	A_4, A_5, A_6	15373	-87	7575	0.005630
1976	15311	A_5	A_4, A_5, A_6	15373	62	3840	0.004047
1977	15603	A_6	A_5, A_6, A_7	15623	20	400	0.001282
1978	15861	A_7	A_6, A_7, A_8	15883	22	487	0.001391
1979	16807	A_{10}	A_9, A_{10}, A_{11}	17079	272	73765	0.016160
1980	16919	A_{10}	A_9, A_{10}, A_{11}	17079	160	25471	0.009433
1981	16388	A_9	A_8, A_9, A_{10}	16497	109	11800	0.006629
1982	15433	A_5	A_4, A_5, A_6	15373	-60	3604	0.003890
1983	15497	A_5	A_4, A_5, A_6	15373	-124	15384	0.008004
1984	15145	A_4	A_3, A_4, A_5	15024	-121	14599	0.007978
1985	15163	A_4	A_3, A_4, A_5	15024	-139	19272	0.009155
1986	15984	A_7	A_6, A_7, A_8	15883	-101	10188	0.006315
1987	16859	A_{10}	A_9, A_{10}, A_{11}	17079	220	48223	0.013025
1988	18150	A_{11}	A_{10}, A_{11}, A_{12}	17991	-159	25136	0.008735
1989	18970	A_{12}	A_{11}, A_{12}, A_{13}	18802	-168	28221	0.008856
1990	19328	A_{13}	A_{12}, A_{13}	18994	-334	111886	0.017306
1991	19337	A_{13}	A_{12}, A_{13}	18994	-343	117988	0.017764
1992	18876	A_{12}	A_{11}, A_{12}, A_{13}	18916	40	1600	0.002119
						MSE=41425.56	**AFER=1.0242%**

TABLE VI
A COMPARISON OF THE FORECASTING RESULTS OF DIFFERENT FORECASTING METHODS

Year	Enrollments	Song Chissom Method [18]	Song Chissom Method [19]	Chen's [2]	Hwang, Chen & Lee's [8]	Huarng's [5]	Chen's [4]	Jilani and Burney [9]	Jilani, Burney and Ardil [10]	Proposed Method
1971	13055	--	--	--	--	--	--	--	14464	13579
1972	13563	14000	--	14000	--	14000	--	--	14464	13798
1973	13867	14000	--	14000	--	14000	--	--	14464	13798
1974	14696	14000	--	14000	--	14000	14500	14730	14710	14452
1975	15460	15500	14700	15500	--	15500	15500	15615	15606	15373
1976	15311	16000	14800	16000	16260	15500	15500	15614	15606	15373
1977	15603	16000	15400	16000	15511	16000	15500	15611	15606	15623
1978	15861	16000	15500	16000	16003	16000	15500	15611	15606	15883
1979	16807	16000	15500	16000	16261	16000	16500	16484	16470	17079
1980	16919	16813	16800	16833	17407	17500	16500	16476	16470	17079
1981	16388	16813	16200	16833	17119	16000	16500	16469	16470	16497
1982	15433	16789	16400	16833	16188	16000	15500	15609	15606	15373
1983	15497	16000	16800	16000	14833	16000	15500	15614	15606	15373
1984	15145	16000	16400	16000	15497	15500	15500	15612	15606	15024
1985	15163	16000	15500	16000	14745	16000	15500	15609	15606	15024
1986	15984	16000	15500	16000	15163	16000	15500	15606	15606	15883
1987	16859	16000	15500	16000	16384	16000	16500	16477	16470	17079
1988	18150	16813	16800	16833	17659	17500	18500	18482	18473	17991
1989	18970	19000	19300	19000	19150	19000	18500	18481	18473	18802
1990	19328	19000	17800	19000	19770	19000	19500	19158	19155	18994
1991	19337	19000	19300	19000	19928	19500	19500	19155	19155	18994
1992	18876	--	19600	19000	19537	19000	18500	18475	18473	18916
MSE	423027	775687	407507	321418	226611	86694	86694	82269	227194	41426
AFER	3.2238%	4.3800%	3.1100%	3.1169%	2.4452%	1.5294%	1.5294%	1.4064 %	2.3865%	1.0242%

V. CONCLUSION

In this paper, we have presented frequency density based partitioning of the historical enrollment data of the University of Alabama and applied improved fuzzy metric for forecasting. The proposed method belongs to the first order and time-variant methods. From Table VI, we can see that the AFER and MSE of the forecasting results of the proposed method are the smallest than that of the existing methods. In the future, we will extend the proposed method to deal with other forecasting problems based on fuzzy time series. We also will develop new methods for forecasting enrollments based on fuzzy parametric and semi-parametric approaches to get a higher forecasting accuracy.

REFERENCES

[1] Chen S. M. and Hsu C.-C. 2004. A new method to forecasting enrollments using fuzzy time series, International Journal of Applied Science and Engineering, 2, 3: 234-244.

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

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

[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] K. Huarng, “Heuristic models of fuzzy time series for forecasting”, *Fuzzy Sets and Systems*, Vol. 123, pp. 369-386, 2002.

[6] K. Huarng, “Effective lengths of intervals to improve forecasting in fuzzy time series”, *Fuzzy Sets and Systems*, Vol. 12, pp. 387-394, 2001.

[7] C. C. Hsu, S. M. Chen, “A new method for forecasting enrollments based on fuzzy time series”, Proceedings of the Seventh Conference on Artificial Intelligence and Applications, Taichung, Taiwan, Republic of China, pp. 17-22.

[8] 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.

[9] 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.

[10] 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.

[11] G. J. Klir, T. A. Folger, *Fuzzy Sets, Uncertainty, and Information*, Prentice-Hall, New Jersey, U.S.A, 1988.

[12] G. J. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, New Jersey, U.S.A, 2005.

[13] L. W. Lee, L. W. Wang, S. M. Chen, “Handling forecasting problems based on two-factors high-order time series”, *IEEE Transactions on Fuzzy Systems*, Vol. 14, No. 3, pp.468-477, 2006.

[14] H. Li, R. Kozma, “A dynamic neural network method for time series prediction using the KIII model”, Proceedings of the 2003 International Joint Conference on Neural Networks, 1: 347-352, 2003.

[15] S. Melike, Y. D. Konstsntin, “Forecasting enrollment model based on first-order fuzzy time series”, in proc. International Conference on Computational Intelligence, Istanbul, Turkey, 2004.

[16] Q. Song, “A note on fuzzy time series model selection with sample autocorrelation functions”, *Cybernetics and Systems: An International Journal*, Vol. 34, pp. 93-107, 2003.

[17] Q. Song, B. S. Chissom, “Fuzzy time series and its models”, *Fuzzy Sets and Systems*, Vol. 54, pp. 269-277, 1993.

[18] Q. Song, B. S. Chissom, “Forecasting enrollments with fuzzy time series Part I”, *Fuzzy Sets and Systems*, 54: 1-9.

[19] Q. Song, B. S. Chissom, “Forecasting enrollments with fuzzy time series: Part II”, *Fuzzy Sets and Systems*, Vol. 62: pp. 1-8, 1994.

[20] Q. Song, R. P. Leland, “Adaptive learning defuzzification techniques and applications”, *Fuzzy Sets and Systems*, Vol. 81, pp. 321-329, 1996.
[21] S. F. Su, S. H. Li,“ Neural network based fusion of global and local information in predicting time series”, Proceedings of the 2003 IEEE International Joint Conference on Systems, Man and Cybernetics, No. 5: pp. 4445-4450, 2003.
[22] J. Sullivan, W. H. Woodall, “A comparison of fuzzy forecasting and Markov modeling”, *Fuzzy Sets and Systems*, Vol. 64, pp.279-293, 1994.
[23] L. X. Wang, J. M. Mendel, “Generating fuzzy rules by learning from examples”, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, pp.1414-1427, 1992.
[24] H.-J. Zimmermann, *Fuzzy Set Theory and Its Applications*, Kluwer Publishers, Boston, USA. 2001.