Journal of Computer Science 6 (7): 735-739, 2010
ISSN 1549-3636

Constructing Fuzzy Time Series Model Based on Fuzzy Clustering for a Forecasting

[1]Ashraf K. Abd Elaal, [2]Hesham A. Hefny and [3]Ashraf H. Abd Elwahab
[1]Department of Computer and Information Sciences, The High Institute of Computer Science, Al-Kawser City at Sohag, Egypt
[2]Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Egypt
[3]Department of Computer Sciences, Electronics Research Institute, National Center for Research, Egypt

Abstract: Problem statement: In this study researchers introduced a fuzzy time series model depending on fuzzy clustering to solve the problem in which the membership values are assumed as Song and Chissom model and to increase the performance of fuzzy time series model. **Approach:** Proposed model employed seven main procedures in time-invariant fuzzy time-series and time-variant fuzzy time series models. In the first step: clustering data, in the second step: determine membership values for each cluster, the third step: define the universe of discourse, in the fourth step: partition universal of discourse into equal intervals, in the fifth step: fuzzify the historical data, in the sixth step: build fuzzy logic relationships and the last step: calculate forecasted outputs to increase the performance of the proposed fuzzy time series model. **Results:** From the evaluations, the proposed model can further improve the forecasting results than the other model. **Conclusion:** The proposed model is a good model for forecasting values. Selecting membership functions based on fuzzy clustering offers an alternative approach to let the data determine the nature of the membership functions. Our results showed that this approach can lead to satisfactory performance for fuzzy time series.

Key words: Fuzzy time series, fuzzy clustering, fuzzy logical relationship, forecasting, enrollments

INTRODUCTION

Song and Chissom (1993a) presented the concept of fuzzy time series based on the historical enrollments of the University of Alabama. They presented the time-invariant fuzzy time series model and the time-variant fuzzy time series model based on the fuzzy set theory for forecasting the enrollments of the University of Alabama.

The fuzzy forecasting methods can forecast the data with linguistic values. Fuzzy time series do not need to turn a non-stationary series into a stationary series and do not require more historical data along with some assumptions like normality postulates. Although fuzzy forecasting methods are suitable for incomplete data situations, their performance is not always satisfactory (Kirchgassner and Wolters, 2007; Palit and Popovic, 2005).

The proposed fuzzy time series model is introduced to handle forecasting problems and improving forecasting accuracy. Each value (observation) is represented by a fuzzy set. The transition between consecutive values is taken into account in order to model the time series data.

The forecast accuracy is compared by using Normalized Root Mean Square Error (NRMSE). The Normalized Root Mean Square Error (NRMSE), in statistic is the square root of the sum of the squared deviations between actual and predicted values divided by the sum of the square of actual values:

$$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\text{actual}_i - \text{predict}_i)^2}{\sum_{i=1}^{N}(\text{actual}_i)^2}} \quad (1)$$

Fuzzy Clustering (FCMI): Fuzzy C Mean iterative assume the existence of pattern space X = {x1, x2,…, xm) and c fuzzy clusters, whose centers have initial values y10, y20,…,yc0. Every iteration the membership

Corresponding Author: Ashraf K. Abd Elaal, Department of Computer and Information Sciences, The High Institute of Computer Science, Al-Kawser City at Sohag, Egypt

function values updated and the cluster centers also. The process terminates when the difference between two consecutive clusters centers do not exceed a given tolerance (Friedman and Kandel, 1999).

Step 1: Determine the number of all iteration N, At iteration k = 0, initialize $y_i = y_{i0}$, $1\leq i\leq c$

Step 2: For $1\leq i\leq c$ and $1\leq j\leq m$:

$$d_{ij}^{(k)} = \left\| x_j - y_i^{(k)} \right\| \quad (2)$$

Step 3: For $1\leq i\leq c$ and $1\leq j\leq m$:

$$X_{ij}^{(k)} = \left[\sum_{l=1}^{c} \left(\frac{d_{ij}^{(k)}}{d_{lj}^{(k)}} \right)^{2/(\beta-1)} \right]^{-1} \quad (3)$$

Step 4: If $d_{lj}^{(k)} = 0$ for $l=l_0$, set $X_{l_0 j}^{(k)} = 1$ and $X_{ij}^{(k)} = 0$ for all $i \neq l_0$, for $1\leq i \leq c$ update the cluster centers by:

$$y_i^{(k+1)} = \frac{\sum_{j=1}^{m} X_{ij}^{(k)} x_j}{\sum_{j=1}^{m} X_{ij}^{(k)}} \quad (4)$$

Step 5: If:

$$\left[\sum_{i=1}^{c} \left\| y_i^{(k+1)} - y_i^{(k)} \right\|^2 \right]^{1/2} < \varepsilon \quad (5)$$

Set $y_i = y_i^{(k+1)}$ for $1\leq i\leq c$; $X_{ij} = X_{ij}^{(k)}$ for $1\leq i\leq c$ and $1\leq j\leq m$ and stop.

Step 6: If k = N then stop else k = k + 1 and go to Step 2.

Fuzzy time series: Song and Chissom (1993b) presented the concept of fuzzy time series based on the historical enrollments of the University of Alabama. Fuzzy time series used to handle forecasting problems. They presented the time-invariant fuzzy time series model and the time-variant fuzzy time series model based on the fuzzy set theory for forecasting the enrollments of the University of Alabama. The definitions and processes of the fuzzy time-series presented by Song and Chissom (1993a) are described as follows (Liu, 2007; Huarng, 2001).

Definition 1 (FTS): Assume Y (t) (t = .. 0, 1, 2,...) is a subset of a real numbers. Let Y (t) is the universe of discourse defined by the fuzzy set fi (t). If F(t) is a collection of f1(t), f2(t)... then F(t) is defined as a fuzzy time-series on Y (t) (t = ..., 0, 1, 2,...).

Definition 2 (FTSRs): If there exists a fuzzy logical relationship R(t-1, t), such that F(t) = F(t-1) × R(t-1, t), where "×" represents an operation, then F(t) is said to be induced by F(t-1). The logical relationship between F(t) and F(t-1) is:

$$F(t-1) \rightarrow F(t)$$

Definition 3 (FLR): Suppose F(t-1) = Ai and F(t) = Aj. The relationship between two consecutive observations, F(t) and F(t-1), referred to as a fuzzy logical relationship, can be denoted by:

$$A_i \rightarrow A_j$$

Where:

A_i = Called the Left-Hand Side (LHS)

A_j = The Right-Hand Side (RHS) of the FLR

Definition 4: (FLRG) all fuzzy logical relationships in the training dataset can be grouped together into different fuzzy logical relationship groups according to the same Left-Hand Sides of the fuzzy logical relationship. For example, there are two fuzzy logical relationships with the same Left-Hand Side (A_i): $A_i \rightarrow A_{j1}$ and $A_i \rightarrow A_{j2}$. These two fuzzy logical relationships can be grouped into a fuzzy logical relationship group:

$$A_i \rightarrow A_{j1}\ A_{j2}$$

Definition 5 (IFTS and VFTS): Assume that F(t) is a fuzzy time-series and F(t) is caused by F(t-1) only and F(t) = F (t-1) × R (t-1, t). For any t, if R (t-1, t) is independent of t, then F(t) is named a time-invariant fuzzy time-series, otherwise a time-variant fuzzy time-series.

Song and Chissom (1993a) employed five main procedures in time-invariant fuzzy time-series and time-variant fuzzy time series models as follows:

Define the universe of discourse U: Define the universe of discourse for the observations. According to the issue domain, the universe of discourse for observations is defined as:

$$U = [D_{min} - D_1, D_{max} + D_2] \quad (6)$$

Where:

D_{min} = The minimum value

D_{max} = The maximum value
D_1 and D_2 = The positive real numbers to divide the U into n equal length intervals

Partition universal of discourse U into equal intervals: After the length of the intervals, is determined, the U can be partitioned into equal-length intervals $u_1, u_2,...,u_n$.

Define the linguistic terms: Each linguistic observation, A_k can be defined by the intervals $u_1, u_2,...,u_n$, as follows:

$$A_k = \begin{cases} \frac{1}{u_1}+\frac{0.5}{u_2} & k=1 \\ \frac{0.5}{u_{k-1}}+\frac{1}{u_k}+\frac{0.5}{u_{k+1}} & 2 \le k \le n-1 \\ \frac{0.5}{u_{n-1}}+\frac{1}{u_n} & k=n \end{cases} \quad (7)$$

Fuzzify the historical data: Each historical data can be fuzzified into a fuzzy set.

Build fuzzy logic relationships: Build fuzzy logic relationships. Two consecutive fuzzy sets $A_i(t-1)$and $A_j(t)$ can be established into a single FLR as $A_i \rightarrow A_j$.

MATERIALS AND METHODS

Proposed model: we introduce a proposed fuzzy time series model depend on fuzzy clustering. Most of authors in fuzzy time series field took the same path according to processes of the fuzzy time-series, which are presented by Song and Chissom (1993a), but we introduce this novel model to solve the problem in which the membership values are assumed as Song and Chissom model and this membership values have an important role in the forecasting values. Proposed model employed seven main procedures in time-invariant fuzzy time-series and time-variant fuzzy time series models as follows:

Step 1: Cluster data into c clusters: Apply fuzzy clustering on a time series Y(t) with n observation to cluster this time series into c ($2 \le c \le n$) clusters. FCMI is used because it is the most popular one and well known in fuzzy clustering field:

$$\begin{aligned} &\text{At iteration } k = 0, \text{initialize } y_i = y_{i0},\ 1 \le i \le c \\ &y_{i0} = \text{Dmin} / (10 * i) \end{aligned} \quad (8)$$

Where:
D_{min} = The minimum value
C = The number of clusters

Step 2: Determine membership values for each cluster: In this step, membership values is determining after doing fuzzy cluster. The proposed model, use this membership values according to the forecasting values.

Step 3: Define the universe of discourse U: In this step, the proposed model defines the universe of discourse as Song and Chissom (1993b) were defined it as Eq. 6.

Step 4: Partition universal of discourse U into equal intervals: According to this step, the proposed model, partition the universe of discourse into c intervals.

Step 5: Fuzzify the historical data: In this step, proposed model fuzzufy historical data, where the proposed model determine the best fuzzy cluster to each actual data.

Step 6: Build fuzzy logic relationships: Proposed model in this step build fuzzy logic relationship as Definition 3. if $F(t-1) = A_i$ and $F(t) = A_j$ then the relationship between two consecutive observations:

$$A_i \rightarrow A_j \quad (9)$$

Step 7: Calculate forecasted outputs: Proposed model forecasting values based on fuzzy logic relationship, if $A_i \rightarrow A_j$ then the forecasting value is the midpoint of A_j.

RESULT AND DISCUSSION

Empirical study: Previous studies on fuzzy time series often used the enrollments data at the University of Alabama as the forecasting target in many forecasting studies.

Based on the enrollments of the University of Alabama from 1971-1992, we can get the universe of discourse U = [13055, 16919], partition U into 7 equal intervals, $D_1 = 31$ and $D_2 = 55$. Hence, the intervals are:

$u_1; u_2; u_3; u_4; u_5; u_6; u_7$

Where:
u_1 = [13024, 13588]
u_2 = [13588, 14153]
u_3 = [14153, 14717]
u_4 = [14717, 15281]
u_5 = [15281, 15845]
u_6 = [15845, 16410]
u_7 = [16410, 16974]

Table 1: Data of enrollments of the University of Alabama

Years	Actual	Fuzzy set
1971	13055	A_1
1972	13563	A_1
1973	13867	A_2
1974	14696	A_3
1975	15460	A_5
1976	15311	A_5
1977	15603	A_5
1978	15861	A_6
1979	16807	A_7
1980	16919	A_7
1981	16388	A_6
1982	13055	A_1
1983	13563	A_1
1984	13867	A_2
1985	14696	A_3
1986	15460	A_5
1987	15311	A_5
1988	15603	A_5
1989	15861	A_6
1990	16807	A_7
1991	16919	A_7
1992	16388	A_6

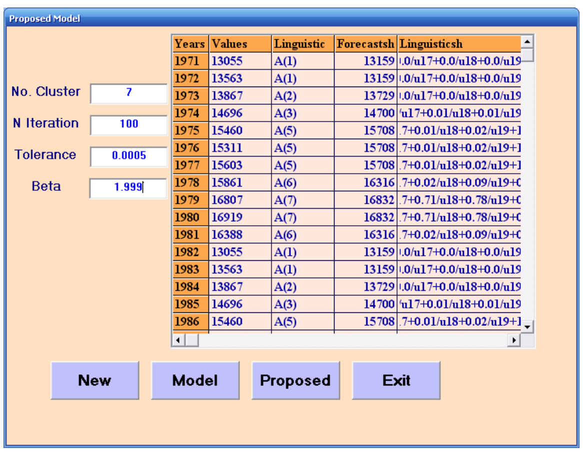

Fig. 1: Forecasting enrollments of the University of Alabama by proposed model

Table 1 lists the enrollment of the University of Alabama from 1971-1992, linguistic value for each observation and fuzzy set as Eq. 7.

The forecasted value for each year is a fuzzy set, A_i, which is defined by:

$$A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}$$

$$A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}$$

...

$$A_6 = \frac{0}{u_2} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0.5}{u_5} + \frac{1}{u_6} + \frac{0.5}{u_7}$$

$$A_7 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0.5}{u_6} + \frac{1}{u_7}$$

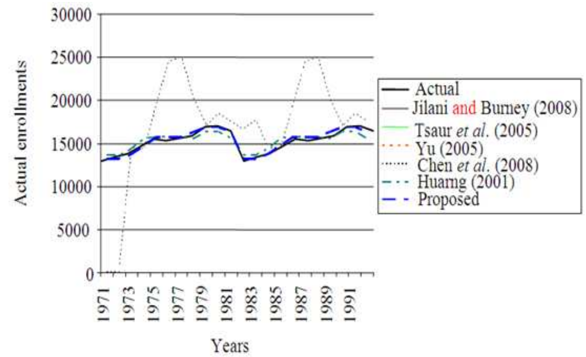

Fig. 2: Comparisons of the forecasting results of different models

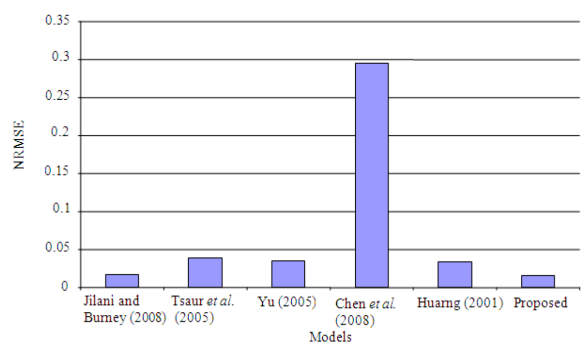

Fig. 3: Comparisons NRMSE of different models

The forecasting value for year 1971 is 13159 while the actual value was 13055. Figure 1 shows linguistic terms and forecasting values deduced by proposed model.

This study uses the same data to evaluate the proposed model and compare with other methods. The forecasting values for (Jilani and Burney, 2008; Tsaur *et al*., 2005; Yu, 2005; Chen *et al*., 2008; Cheng *et al*., 2008) and our proposed model are given for comparison purpose in Table 2. By using line-chart the comparison of the forecasting results between these models in Fig. 2.

The NRMSE in (Jilani and Burney, 2008) is 0.0171, in (Tsaur *et al*., 2005) is 0.0393, in (Yu, 2005) is 0.0349, in (Chen *et al*., 2008) is 0.2955, in (Huarng, 2001) is 0.0332 and in proposed model is 0.0158. From the evaluations, the proposed model can further improve the forecasting results than the other model. Figure 3 shows the comparisons by NRMSE of different models.

Table 2: Comparisons of the forecasting results of different models

Year	Actual	Jilani and Burney (2008)	Tsaur *et al*. (2005)	Yu (2005)	Chen *et al*. (2008)	Huarng (2001)	Proposed
1971	13055	13489	13588	13645	0	13588	13159
1972	13563	13489	13588	13645	0	13588	13159
1973	13867	13859	14435	14435	14474	14435	13729
1974	14696	14424	15563	15563	14333	15563	14700
1975	15460	15553	15845	15805	19361	15751	15708
1976	15311	15553	15845	15805	24438	15751	15708
1977	15603	15553	15845	15805	25018	15751	15708
1978	15861	16118	14999	15563	20100	15563	16316
1979	16807	16499	16410	16353	17008	16410	16832
1980	16919	16499	16410	16353	18410	16410	16832
1981	16388	16118	14999	15563	17533	15563	16316
1982	13055	13489	13588	13645	16670	13588	13159
1983	13563	13489	13588	13645	17704	13588	13159
1984	13867	13859	14435	14435	14474	14435	13729
1985	14696	14424	15563	15563	14333	15563	14700
1986	15460	15553	15845	15805	19361	15751	15708
1987	15311	15553	15845	15805	24438	15751	15708
1988	15603	15553	15845	15805	25018	15751	15708
1989	15861	16118	14999	15563	20100	15563	16316
1990	16807	16499	16410	16353	17008	16410	16832
1991	16919	16499	16410	16353	18410	16410	16832
1992	16388	16118	16410	15563	17533	15563	16316
NRMSE		0.0171	0.0393	0.0349	0.2955	0.0332	0.0158

CONCLUSION

A novel fuzzy time series method based on fuzzy clustering has been proposed. The method of FCMI is integrated in the processes of fuzzy time series to partition datasets. Experimental results on enrollments at the University of Alabama and the comparison with other models: (Jilani and Burney, 2008; Tsaur *et al*., 2005; Yu, 2005; Chen *et al*., 2008; Huarng, 2001) show from results that the proposed model is a good model for forecasting values.

Selecting membership functions based on fuzzy clustering offers an alternative approach to let the data determine the nature of the membership functions.

Our results show that this approach can lead to satisfactory performance for fuzzy time series.

REFERENCES

Chen, T.L., C.H. Cheng and H.J. Teoh, 2008. High-order fuzzy time-series based on multi-period adaptation model for forecasting stock markets. Physica A., 387: 876-888. DOI: 10.1016/j.physa.2007.10.004

Cheng, C.H., J.W. Wang and G.W. Cheng, 2008. Multi-attribute fuzzy time series method based on fuzzy clustering. Exp. Syst. Appli., 34: 1235-1242. DOI: 10.1016/j.eswa.2006.12.013

Friedman, M. and A. Kandel, 1999. Introduction to Pattern Recognition Statistical, Structural, Neural and Fuzzy Logic Approaches. Imperial College Press, London, ISBN: 9810233124, pp: 329.

Huarng, K., 2001. Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets Syst., 123: 387-394.

Jilani, T.A. and S. Burney, 2008. Multivariate stochastic fuzzy forecasting models. Exp. Syst. Appli., 35: 691-700. DOI: 10.1016/j.eswa.2007.07.014

Kirchgassner, G. and J. Wolters, 2007. Introduction to Modern Time Series Analysis. Springer-Verlag, Berlin, Germany, ISBN: 978-3-540-73290-7, pp: 153.

Liu, H.T., 2007. An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. Fuzzy Optim. Dec. Mak., 6: 63-80. DOI: 10.1007/s10700-006-0025-9

Palit, A.K. and D. Popovic, 2005. Computational Intelligence in Time Series Forecasting Theory and Engineering Applications. Springer-Verlag, London, UK., ISBN: 1852339489, pp: 18.

Song, Q. and B.S. Chissom, 1993a. Forecasting enrollments with fuzzy time series. I. Fuzzy Sets Syst., 54: 1-9. DOI: 10.1016/0165-0114(93)90355-L

Song, Q. and B.S. Chissom, 1993b. New models for forecasting enrollments: fuzzy time series and neural network approaches. ERIC, pp: 27. http://www.eric.ed.gov

Tsaur, R.C., J.C. Yang and H.F. Wang, 2005. Fuzzy relation analysis in fuzzy time series model. Comput. Math. Appli., 49: 539-548. DOI: 10.1016/j.camwa.2004.07.014

Yu, H.K., 2005. Weighted fuzzy time series models for TAIEX forecasting. Physica A., 349: 609-624. DOI: 10.1016/j.physa.2004.11.006