Load Forecasting based on Fuzzy Time Series

*AO Pei

Engineering Technology Research Center for Computing Intelligence & Data Mine Henan Province

College of Computer and Information Engineering

Henan Normal University

Xinxiang Henan, China

e-mail: 34204339@qq.com

Key words: Fuzzy time series; K-means algorithm; Fuzzidication; Fuzzy relation;Load forecasting

Abstract. Load forecasting is a traditional research issue in the field of electric power system. In this paper, an improved fuzzy time series approach is used to forecast load. Firstly, a method of unequal-sized intervals partitioning based on K-means algorithm is proposed. Secondly, improved fuzzification method is proposed to overcome the defect of traditional fuzzification method. Finally, the model is used to forecast load and the relation between the number of clustering and the prediction accuracy and the relation between the order and the prediction accuracy are studied. The validity of model is verified by prediction results.

Introduction

The classical time series is a forecasting method based on the theory of mathematical statistics to predict the future trend of the future by using historical data. Time series analysis tends to rely on a large amount of historical data. However, it produces uncertain and incomplete data because of the interaction between variable data, collection error and time delay. If the traditional time series analysis method is used to predict, prediction accuracy of model can be influenced. Fuzzy time series approach based on fuzzy set theory [1] firstly is proposed by Song and Chissom [2-4]. The aim of the studies on the fuzzy time series in the literature is to contribute to three steps of the approach. One of these is the fuzzification step [5-9]. Another step is defining fuzzy relations [10,11]. And the last step is the defuzzification [12,13]. Since then, a large number of researches have focused on the domain partition, fuzzification and defining fuzzy relations in order to improve the prediction accuracy and reduce computational complexity. In the initial stage of fuzzy time series, a lot of researches are limited to the method of equal-sized intervals partitioning. Afterwards, with the development of the research, some scholars realized that the intervals have a great influence on the prediction results and proposed a method of unequal-sized intervals partitioning. In this paper, a method of unequal-sized intervals partitioning based on K-means algorithm is proposed and the fuzzy time series model is built based on improved fuzzification method.

Model for the Optimal Fuzzy Measure on the Attribute Set

A. The domain partition based on K-means algorithm

According to the K-means algorithm process, different clustering centers $k_1, k_2, \ldots, k_{c_g}$ can be got. The midpoint of two adjacent cluster centers is taken as the boundary point of the domain partition. The boundary points are defined as $d_1, d_2, \ldots, d_{c_g}$. The domain U are divided into c_g intervals $u_1, u_2, \ldots, u_{c_g}$, where $u_1 = [x_{\min} + s, d_1]$, $u_2 = [d_1, d_2], \ldots,$ $u_{c_g} = [d_{c_g - 1}, x_{\max} + s]$, $s = \min(k_{j+1} - k_j), 1 \le j \le c_g - 1$.

B. Definition of fuzzy sets and fuzzification of data

The domain U are divided into c_g intervals $U = \{u_1, u_2, \ldots, u_{c_g}\}$, then $A_k = (u_1(x_k), u_2(x_k), \ldots, u_{c_g}(x_k))$ is fuzzy set on U, where x_k is a data point in the sample data, $u_1(x_k), u_2(x_k), \ldots, u_{c_g}(x_k)$ is the membership degree of x_k.

When the difference between the sample data is not too large, using triangular fuzzy membership degree function to fuzzy data will make some sample data fall in the same interval and lead to fuzzy different sample data into the same fuzzy set. The data of sensitivity can not to be made full use. Prediction accuracy of model is reduced. In order to solve this problem, a method of fuzzification is used as follows:

Let $D_1, D_2, \ldots, D_{c_g+1}$ is the boundary points by using the method of unequal-sized intervals partitioning on U, and $D_1 = x_{\min} + s$, $D_2 = d_1$, …, $D_{c_g+1} = x_{\max} + s$. When sample data x_k belongs to u_i, $1 \le i \le c_g$. Let

$$\begin{cases} u_i(x_k) = 1 & x_k \text{ belongs to } u_i, 1 \le i \le c_g. \\ u_j(x_k) = \dfrac{D_{\min}}{\left|D_{j+1} - x_k\right| + \left|x_k - D_j\right|} & \\ x_k \text{ does not belong to } u_i, 1 \le j \le c_g,\ j \ne i, 1 \le k \le n & \end{cases} \tag{1}$$

where $D_{\min}$ is the minimum of $D_{m+1} - D_m$ when $m = 1, 2, \ldots, c_g$. All of sample data is fuzzed and corresponding fuzzy set is obtained as follows:

$$\begin{cases} A_1 = \dfrac{u_1(x_1)}{u_1} + \dfrac{u_2(x_1)}{u_2} + \cdots + \dfrac{u_{c_g}(x_1)}{u_{c_g}} \\ A_2 = \dfrac{u_1(x_2)}{u_1} + \dfrac{u_2(x_2)}{u_2} + \cdots + \dfrac{u_{c_g}(x_2)}{u_{c_g}} \\ \qquad \cdots \\ A_n = \dfrac{u_1(x_n)}{u_1} + \dfrac{u_2(x_n)}{u_2} + \cdots + \dfrac{u_{c_g}(x_n)}{u_{c_g}} \end{cases} \tag{2}$$

By using this method, x_k meet that the membership degree of the intervals in the two sides of u_i present decreasing trend reduced when U is divided by using the method of unequal-sized intervals partitioning.

C. Defining fuzzy relation

After fuzzification, each sample data is fuzzed into a fuzzy value. In fuzzy time series analysis, historical data have important influence on prediction. Data are closer to the predicted point data and the correlation to the forecasting values is stronger, the effect on the forecasting values is greater. If a fuzzy set relationship between two corresponding data is only considered when the fuzzy logical relation is built, while ignoring the influence of non-adjacent data on fuzzy set relation by of fuzzy sets, the fuzzy relation may lose a lot of information and inaccurate results of prediction are lead to.

Fuzzy relation of k order model can be built. Let $A_i^{t-k}, A_j^{t-k+1}, \ldots, A_l^{t-1}$ is the corresponding fuzzy sets of $F(t-k), F(t-k+1), \ldots, F(t-1)$. If there is a derivation relationship $F(t-k), F(t-k+1), \ldots, F(t-1) \to F(t)$, $A_i^{t-k}, A_j^{t-k+1}, \ldots, A_l^{t-1} \to A_m^t (i, j, l, m = 1, 2, \ldots, c_g)$. The standard vector of fuzzy time series is defined as $C(t) = f(t-1) = \lfloor C_1, C_2, \ldots, C_{c_g} \rfloor$, the operation matrix about $F(t)$ is shown as follows:

$$O^k(t) = \begin{bmatrix} f(t-1) \\ f(t-2) \\ \cdots \\ f(t-k) \end{bmatrix} = \begin{bmatrix} O_{11} & \cdots & O_{1c_g} \\ O_{21} & \ldots & O_{2c_g} \\ \cdots & \cdots & \cdots \\ O_{(k-1)1} & \cdots & O_{(k-1)c_g} \end{bmatrix} \tag{3}$$

where $f(t-1)$ is the fuzzification change $F(t)$ of between $t-1$ and $t-2$, c_g is the number of intervals on the domain. Then fuzzy relation $R(t)$ is shown as follows:

$$R(t)=C(t)*O^k(t)=\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1c_g} \\ R_{21} & R_{22} & \ldots & R_{2c_g} \\ \cdots & \cdots & \cdots & \cdots \\ R_{(k-1)1} & R_{(k-1)2} & \cdots & R_{(k-1)c_g} \end{bmatrix} \tag{4}$$

where $C_j, O_{ij} \in [0,1], i \in [1, k-1], j \in [1, c_g], R_{ij} = C_j * O_{ij}$. Form the fuzzy relation $R(t)$, the membership degree of prediction in t time point is shown as follows:

$$\begin{aligned} f(t) &= \left(\max\left(R_{11}, \ldots, R_{(k-1)1}\right), \ldots, \max\left(R_{1c_g}, \ldots, R_{(k-1)c_g}\right)\right) \\ &= \left(f_{t1}, \ldots, f_{tc_g}\right) \end{aligned} \tag{5}$$

D. Forecasting and defuzzification

The method of gravity center method is used to defuzzification of the prediction results.

$$F_t = \frac{\sum_{t=1}^{c_g} k_t f_{ti}}{\sum_{t=1}^{c_g} f_{ti}} \tag{6}$$

where k_t is the clustering center, f_{ti} is the membership degree of prediction, F_t is value of prediction, c_g is the number of the clustering center.

Empirical analysis

In this paper, we study the load data of Electric Power Company National in May 23, 2007 and data are shown in Table 1. Using the method of fuzzy time series proposed in this paper, specific steps are as follows:

TABLE 1 THE LOAD DATA OF ELECTRIC POWER COMPANY NATIONAL IN MAY 23, 2007

Series number	1	2	3	4	5	6	7	8	9	10	11	12
Load value	1176	1129	1095	1098	1093	1080	1195	1327	1509	1567	1614	1640
Series number	13	14	15	16	17	18	19	20	21	22	23	24
Load value	1610	1600	1591	1547	1528	1482	1418	1637	1633	1515	1417	1293

Firstly, In order to address the relation between the number of clustering and the prediction accuracy, forecasting is done when the number of clustering is made to change from 2 to 12. K-means clustering algorithm is used to partitioning on the domain. The domain partition is shown in Table 2.

Secondly, Fuzzification is done according to Eq. (1). Four order fuzzy relation is built according to Eq. (3) and (4). Forecasting is done according to Eq. (5) and (6). Prediction results are shown in Fig. 1.

Then, according to $MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|(x_t - \hat{x}_t)/x_t\right|$, mean absolute percent error (MAPE) is calculated to evaluate the prediction accuracy and shown in Table 3. From Table 3, the MAPE gets the minimum when the number of clustering is 8. Namely, the prediction accuracy is highest when the number of clustering is 8.

In order to address the relation between the order and the prediction accuracy, the forecasting with 8 clustering partitioning on domain is done when the order changes from 2 to 7. Prediction results are shown in Fig. 2. MAPE is calculated and shown in Table 4. From Table 4, the MAPE gets the minimum when the order is 4. Namely, the prediction accuracy is highest when the order is 4.

TABLE 2 PARTITION ON THE DOMAIN

The number of clustering	Partition on the domain
2	[691,1359],[1359,1943]
3	[869,1244],[1244,1469],[1469,1786]
4	[954,1217],[1217,1394],[1394,1541],[1541,1731]
5	[1026,1244],[1244,1436],[1436,1538],[1538,1595],[1595,1676]
6	[994,1142],[1142,1248],[1248,1364],[1364,1471],[1471,1571],[1571,1704]
7	[1043,1110],[1110,1157],[1157,1248],[1248,1364],[1364,1471],[1471,1571],[1571,1655]
8	[1075,1142],[1142,1275],[1275,1436],[1436,1533],[1533,1580],[1580,1618],[1618,1636],[1636,1644]
9	[1047,,1110],[1110,1157],[1157,1248],[1248,1364],[1364,1463],[1463,1533],[1533,1580],[1580,1620],[1620,1670]
10	[1051,1110],[1110,1157],[1157,1248],[1248,1364],[1364,1460],[1460,1520],[1520,1552],[1552,1585],[1585,1620],[1620,1666]
11	[1075,1110],[1110,1153],[1153,1186],[1186,1253],[1253,1364],[1364,1463],[1463,1533],[1533,1580],[1580,1618],[1618,1636],[1636,1644]
12	[10651088],[1088,1112],[1112,1157],[1157,1239],[1239,1310],[1310,1372],[1372,1460],[1460,1520],[1520,1552],[1552,1585],[1585,1620],[1620,1652]

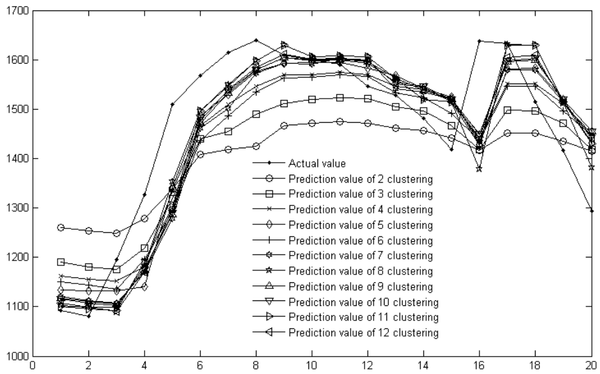

FIG. 1 PREDICTION RESULTS

TABLE 3 MAPE OF PARTITION

The number of clustering	2	3	4	5	6	7	8	9	10	11	12
MAPE	0.0812	0.0636	0.0577	0.0549	0.0580	0.0567	0.0490	0.0552	0.0541	0.0510	0.0532

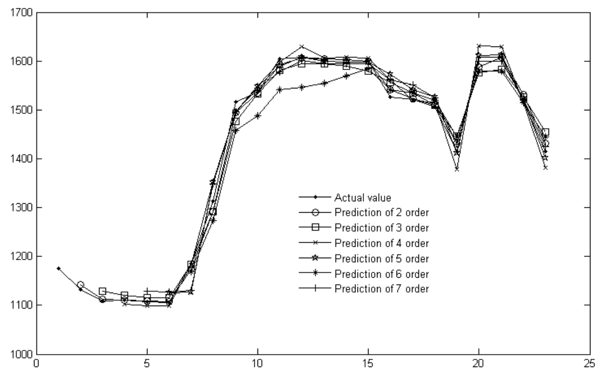

FIG. 2 PREDICTION RESULTS

TABLE 4 MAPE OF PARTITION

Order	2	3	4	5	6	7
MAPE	0.0502	0.0524	0.0490	0.0551	0.0548	0.0637

Conclusion

In this paper, an improved fuzzy time series approach is used to forecast load. Firstly, a method of unequal-sized intervals partitioning based on K-means algorithm is proposed. Secondly, improved fuzzification method is proposed to overcome the defect of traditional fuzzification method. Finally, the model is used to forecast load and the relation between the number of clustering and the prediction accuracy and the relation between the order and the prediction accuracy are studied. From MAPE, it is shown that the prediction accuracy is the highest by using 4 order fuzzy time series forecasting model with 8 clustering partition on domain.

Acknowledgment

This work was supported by basic and advanced technology research project of science and technology department in Henan province (No.13A413506) and basic research project of key project of scientific and technological research education department in Henan province (No.142300410163).

References

[1] Zadeh LA. Fuzzy sets [J]. Information & Control, 1965, 8: 338-352

[2] Song Q., Chissom B.S.. Fuzzy time series and its models[J]. Fuzzy Sets and Systems, 1993a, 54:269-277

[3] Song Q., Chissom B.S.. Forecasting enrollments with fuzzy time series-Part I [J]. Fuzzy Sets and Systems, 1993b, 54:1-10

[4] Song Q., Chissom B.S.. Forecasting enrollments with fuzzy time series-Part II [J]. Fuzzy Sets and Systems, 1994, 62(1):1-8

[5] Huanrng K. Heuristic models of fuzzy time series for forecasting [J]. Fuzzy Sets and Systems, 2001, 123(3):369-386

[6] Huanrng K, Yu T.H.-K. Ratio-based lengths of intervals to improve fuzzy time series forecasting [J]. IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, 2006b, 36:328-340

[7] Chen S.M., Chung N.Y.. Forecasting enrollment using high order fuzzy time series and genetic algorithms [J]. International Journal of Intelligent Systems, 2006, 21:485-501

[8] Lee L.W., Wang L.H., Chen S.M.. Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms [J]. Expert Systems with Application, 2007, 33:539-550

[9] Li S.-T, Cheng Y.-C, Lin S.-Y. A FCM-based deterministic forecasting model for fuzzy time series [J]. Computer and Mathematics with Application, 2008, 56:3052-3063

[10] Egrioglu E., Aladag C.H., Yolcu U., et al. A new approach based on artificial neural networks for high order multivariate fuzzy time series [J]. Expert Systems with Application, 2009, 36:10589-10594

[11] Sullivan J., Woodall W.H.. A comparison of fuzzy forecasting and Markov modeling [J]. Fuzzy Sets and Systems, 1994, 64(3):279-293

[12] Jilani T., Burney S.M.A.. A refined fuzzy time series model for stock market forecasting [J]. Physica A, 2008, 387:2857-2862

[13] Cheng C.-H, Chen T.-L, Teoh H.J., et al. Fuzzy time-series based on adaptive expectation model for TAIEX forecasting [J]. Expert Systems with Application, 2008, 34:1126-1132