Fuzzy Logic and T-Test for Load Forecasting

***Phan Thi Thanh Binh*[1]*, Dinh Xuan Thu*[1]*, Vo Viet Cuong*[2,*]**

[1]*HCMC University of Technology, No. 268 Ly Thuong Kiet Street, District 10, HCMC, Vietnam*

[2]*HCMC University of Technology and Education, No. 1 Vo Van Ngan Street, HCMC, Vietnam*

Received: October 03, 2017; Accepted: November 26, 2018

Abstract

The forecasting models based on regression function have the analytic form with proving that there is some rule expressing the correlation between forecasting value and other related fators. In reality, forecasted load is not always in linear form of factors, such as: temperature, population, GDP or historical load data. This paper applied fuzzy rules to approximate the relationship between loads and other factors using the subtractive clustering. The implementation is carried out for one substation in Ho Chi Minh city. Results show that the proposed approach gives better accuracycy of forecasting, and the effort of finding crisp function for forecasting is not helping to have better results.

Keywords: subtractive clustering, fuzzy rule, correlation, T-test, load forecasting

1. Introduction

By tradition, the forecasting models in regression function have an analytic form, such as $Y = f(x_1, x_2, ..., x_n)$ or $\log Y = f(\log x_1, \log x_2, ..., \log x_n)$. These models are linear and are used only when the linear correlation is significant (expressed by the correlation coefficient) [1]. Relationship between load and correlation factors GDP and economic, social factors such as electricity consumption per person, energy consupmtion per unit production, electricity price to be effected by time (cheaper technology, more electrification …). All of theses make relationship between load and correlation factors is not the analytic form. So in reality, the crisp form of $Y = f(x_1, x_2, ..., x_n)$ are not easy or sometimes not necessary to be found.

Recently, the AI techniques such as Neural network, Wavelet, and Fuzzy logic [2-4], [6], [7] are widely used in forecasting. The advantages of these techniques are focused on approximation of $Y = f(x_1, x_2, ..., x_n)$ without concerns about proving the existence of analytic function of forecasting. Many works as [2][3][4] concentrated on the regression with others factors such as temperatures and on the FCM algorithm (Fuzzy C mean) for finding fuzzy rules. The quality of FCM depends strongly on the choice of initial clusters centers.

Yager and Filev proposed a simple and effective algorithm, called the mountain method, for estimating the number and initial location of cluster centers. Their method is based on gridding the data space and computing a potential value for each grid point. Although this method is simple and effective, the computation grows exponentially with the dimension of the problem. Chiu [5] proposed an extension of Yager and Filev's mountain method, called subtractive clustering, in which each data point, rather than the grid point, is considered as a potential cluster center. Using this method, the number of effective "grid points" to be evaluated is simply equal to the number of data points, independent of the dimension of the problem.

This paper focused on using Fuzzy rules to approximate the relationship between loads and external factors. Theres rules are found based on the method proposed by Chiu in 1994 [5]. The correlation between load at one moment and itself in the past will be mentioned. The correlation estimation is based on the T-test. Combination of fuzzy rules deliver approximate modle of relationship between load and correlation factors.

2. Test for correlation estimation of electricity consumption, temperature

The T-test is based on the correlation r. This expresses the correlation of variable X (electricity consumption) and Y (temperature, electricity consumption of previous days) with the test for hypothesis H_0:

$H_0 : \rho = 0$ (no correlation between X and Y)

$H_1 : \rho \neq 0$ (is correlation between X and Y)

Test value:

*Corresponding author: Tel.: (+84) 986.523.475
Email: cuongvv@hcmute.edu.vn

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \quad (1)$$

Test rule: for meaning level α, H_0 will be denied if:

$$\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} < -t_{n-2,\frac{\alpha}{2}} \quad \text{or} \quad \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} > t_{n-2,\frac{\alpha}{2}} \quad (2)$$

With:

$$r = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}}$$

The set of daily electricity consumption may be treated as one time series. From the T-test result, the correlation between daily electricity consumption At, itself in the past and temperature will be determined. Suppose there are the correlation between t day, one day, two days, seven days before, the temperature, then the input-output matrix has the following forms:

$$\begin{bmatrix} T_8^o & A_1 & A_6 & A_7 & A_8 \\ T_9^o & A_2 & A_7 & A_8 & A_9 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_n^o & A_{n-7} & A_{n-2} & A_{n-1} & A_n \end{bmatrix}$$

input *output*

Or:

$$\left\{ A_{t-7}, A_{t-2}, A_{t-7}, T^0 \quad A_t \right\}$$

input y *output z*

The forecasting function will be:

$$A_t = f\left(A_{t-h}, T^0\right) \quad (3)$$

Where h is the backship day and T^0 is the temperature at the day of t.

3. Determining fuzzy rules

Supposing that electric load and correlation factors is vector x which include 2 parts of input which consist of correlation factors, and output is electric load. Those vectors to be clasified that deliver certain groups. By that way, (3) can be approximated by some rules. The number of rules is the number of cluster centers. The subtractive clustering in [5] is developed. Consider a collection of n data points $\{x_1, x_2, \ldots x_n\}$ in an M dimensional space. Using the subtractive clustering proposed by Chiu, the set of centers $\left\{x_i^*\right\}$ will be determined without loss of generality, we assume that the data points have been normalized in each dimension so that they are bounded by a unit hypercube. If each data point is considered as a possible cluster center, then the potential of data point x_i will be:

$$P_i = \sum_{k=1}^{n} e^{-\alpha\|x_k - x_i\|^2} \quad (4)$$

With: $$\alpha = \frac{4}{r_a^2} \quad (5)$$

||.|| denotes the Euclidean distance, and r_a is a positive constant. A data point with many neighboring data points will have a high potential value. The constant r_a is effectively the radius defining a neighborhood. The data point with the highest potential is selected as the first cluster center. Let x_1* be the location of the first cluster center and P_1* be its potential value. The potential of each data point x_i is revised by the formula:

$$P_i \Leftarrow P_i - P_1^* e^{-\beta\|x_i - x_1^*\|^2} \quad (6a)$$

With:

$$\beta = \frac{4}{r_b^2} \quad (6b)$$

where r_b is the effective radius and be equal to 1.25 r_a . The data points near the first cluster center will have greatly reduced potential, and therefore will be unlikely to be selected as the next cluster center. The data point with the highest remaining potential is selected as the second cluster center. The process is then continued further until the remaining potential of all data points falls below some fraction of the potential of the first cluster center P_1*.

The algorithm of subtractive clustering is illustrated in Fig.1.

Each center $\left\{x_i^*\right\}$ consists of input $\left\{y^*\right\}$ and output $\left\{z^*\right\}$ and will be regarded as one fuzzy rule. For each input vector y, its degree to satisfying the i-fuzzy rule is:

$$\mu_i = e^{-\alpha\|y - y_i^*\|^2} \quad (7)$$

The output will be:

$$z = \frac{\sum_{i=1}^{c}\mu_i z_i^*}{\sum_{i=1}^{c}\mu_i} \quad (8)$$

where c is number of centers.

Yager and Filev [5] suggested that Z_{ij} in (8) will be the linear function of the inputs as following:

$$\overset{*}{z}_{ij} = G_i y + h_i \qquad (9)$$

Here G_i is the matrix of constants with *(N-1)x1* dimension; h is the column vector of constants with *(N-1)* elements where *(N-1)* is the dimension of input.

Now denoting:

$$\rho_i = \frac{\mu_i}{\sum_{j=1}^{c} \mu_j} \qquad (10)$$

Then (8) is rewritten as:

$$z = \sum_{i=1}^{c} \rho_i \overset{*}{z}_i = \sum_{i=1}^{c} \rho_i \left(G_i y + h_i \right) \qquad (11)$$

With a set of n inputs $\{y_1, y_2, \ldots y_n\}$, the set of outputs will be:

$$\begin{bmatrix} z_1^T \\ \vdots \\ z_n^T \end{bmatrix} = \begin{bmatrix} \rho_{1,1} y_1^T & \rho_{1,1} & \cdots & \rho_{c,1} y_1^T & \rho_{c,1} \\ & & \vdots & & \\ \rho_{1,n} y_n^T & \rho_{1,n} & \cdots & \rho_{c,n} y_n^T & \rho_{c,n} \end{bmatrix} \begin{bmatrix} G_1^T \\ h_1^T \\ \vdots \\ G_c^T \\ h_c^T \end{bmatrix} \qquad (12)$$

where T is the tranpose symbol.

The estimation of G and h in (12) can be realised by mean least square method. After evaluating G and h, for given y at moment $t+1$, we can calculate the output z_t+1 as the one step ahead forecasting using (11).

4. Case study

4.1. Forecasting the daily electricity consumption of Go Vap substation in the year of 2012

The historic data are the daily temperature and daily electricity consumption from 02/01/2012 to 07/24/2012. 165 data will be used for identification and training, 15 data are used for testing (validation). The T test shows that daily electricity consumption is depended on the daily mean temperature, the consumption of one day, two days, and seven days before. The results for 15 days are presented in the Table. 1. The MAPE for 15 days forecasting is 2.11%. Meanwhile, if we focused only on the correlation between load and the temperature, the results are given in Table 2 and the MAPE is of 2.59%. Crisp modle is also to be test in the paper, the best trying fuction is:

$$y = 35.648271x$$

with MAPE of the last 10 days is 2.655% (see table 5).

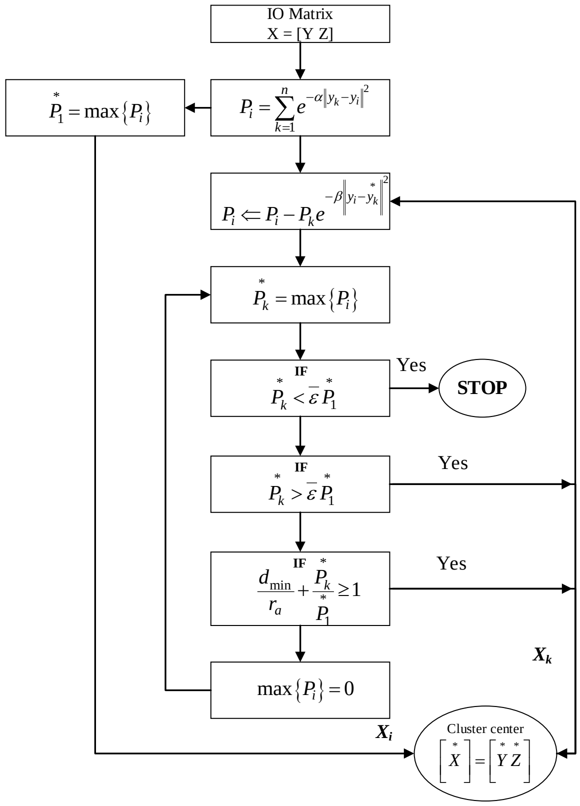

Fig. 1. The cluster centers identification.

4.2. Forecasting the peak hours electricity consumption of Go Vap substation

The series of electricity consumption in peak hours are examined. As in the above section, the influence of daily temperature, peak consumption of one day, two days, and seven days before will be included in (3). The results for 15 days are presented in Table. 3 and the MAPE of 15 days is 2.34%. Meanwhile, if we focused only on the correlation between load and the temperature, the results are given in Table 4 and the MAPE is of 2.86%. While, after trying different regression forms, the best crisp fuction is:

$$y = -525.132 - 0.542x^2 + 40.9131x$$

with MAPE of the last 10 days is 2.954% (see table 5).

Table 1. Forecasting results with correlation of temperature and consumption of previous days

Day	7/10	7/11	7/12	7/13	7/14
Forecasting (MWh)	1404.388	1382.24	1372.014	1348.185	1370.993
Real value (MWh)	1394.1	1325.1	1365.7	1346.1	1402.9
Error	0.00738	0.043122	0.004623	0.001549	0.022743
Day	7/15	7/16	7/17	7/18	7/19
Forecasting (MWh)	1403.494	1433.228	1451.123	1375.076	1400.899
Real value (MWh)	1355.6	1536.5	1468.9	1361.2	1406
Error	0.035331	0.067212	0.012102	0.010194	0.003628
Day	7/20	7/21	7/22	7/23	7/24
Forecasting (MWh)	1378.054	1404.325	1411.243	1438.145	1436.133
Real value (MWh)	1395.1	1423	1333.6	1470.6	1431.4
Error	0.012219	0.013123	0.058221	0.022069	0.003307

Table 2. Forecasting results with correlation of temperature only

Day	7/10	7/11	7/12	7/13	7/14
Forecasting (MWh)	1399.185	1367.276	1364.205	1308.274	1370.04
Real value (MWh)	1394.1	1325.1	1365.7	1346.1	1402.9
Error	0.003647	0.031828	0.001095	0.028101	0.023423
Day	7/15	7/16	7/17	7/18	7/19
Forecasting (MWh)	1436.401	1478.026	1404.522	1349.309	1431.445
Real value (MWh)	1355.6	1536.5	1468.9	1361.2	1406
Error	0.059606	0.038057	0.043827	0.008736	0.018097
Day	7/20	7/21	7/22	7/23	7/24
Forecasting (MWh)	1375.777	1404.975	1415.641	1446.838	1391.524
Real value (MWh)	1395.1	1423	1333.6	1470.6	1431.4
Error	0.013851	0.012667	0.061518	0.016158	0.027858

Table 3. The peak hours consumption forecasting with correlation of temperature and of the peak consumption of previous days

Day	7/10	7/11	7/12	7/13	7/14
Forecasting (MWh)	213.0536	210.3872	208.4978	202.89	206.6889
Real value (MWh)	210.6	205.8	201.2	205.2	208.7
Error	0.01165	0.022289	0.036271	0.011257	0.009636
Day	7/15	7/16	7/17	7/18	7/19
Forecasting (MWh)	213.5647	220.7682	219.6089	204.9288	213.3576
Real value (MWh)	213.7	239	218.5	211.7	213
Error	0.01165	0.022289	0.036271	0.011257	0.009636
Day	7/20	7/21	7/22	7/23	7/24
Forecasting (MWh)	208.8112	213.4879	213.4774	220.9498	217.6144
Real value (MWh)	216.1	208.9	203.5	227.7	219.7
Error	0.033729	0.021962	0.049029	0.029645	0.009493

Table 4. The peak hours consumption forecasting with correlation of temperature ony

Day	7/10	7/11	7/12	7/13	7/14
Forecasting (MWh)	211.7097	206.679	206.2511	197.3768	207.1493
Real value (MWh)	210.6	205.8	201.2	205.2	208.7
Error	0.005269	0.004271	0.025105	0.038125	0.00743
Day	7/15	7/16	7/17	7/18	7/19
Forecasting (MWh)	217.56	224.1803	212.6203	203.8448	216.8835
Real value (MWh)	213.7	239	218.5	211.7	213
Error	0.018063	0.062007	0.02691	0.037105	0.018233
Day	7/20	7/21	7/22	7/23	7/24
Forecasting (MWh)	208.1032	212.7412	214.381	219.3055	210.6149
Real value (MWh)	216.1	208.9	203.5	227.7	219.7
Error	0.037005	0.018388	0.053469	0.036867	0.041352

Table 5. Forecasting with crisp function

Day	7/15	7/16	7/17	7/18	7/19	7/15	7/16	7/17	7/18	7/19
Forecasting daily consumption (MWh)	1444.2	1478.3	1427.1	1354.7	1435.6	1371.7	1384.5	1371.7	1427.1	1333.5
Real daily consumption (MWh)	1355.6	1536.5	1468.9	1361.2	1406	1395.1	1423	1333.6	1470.6	1431.4
Error	0.065	0.0378	0.0284	0.0047	0.0211	0.0167	0.027	0.0286	0.0295	0.0683
Forecasting peak hours consumption (MWh)	218.0	224.0	213.2	203.7	217.3	208.3	213.2	214.8	219.6	211.0
Real peak hours consumption (MWh)	213.7	239	218.5	211.7	213	216.1	208.9	203.5	227.7	219.7
Error	0.0206	0.063	0.024	0.0378	0.0202	0.0357	0.0206	0.0559	0.0354	0.0394

5. Conclusion

The T-test is necessary for finding the correlation between load at one moment and at the previous moments. These correlations are expressed by fuzzy rules based on the subtractive methods. Examining for one substation shows that the proposed approachbased Fuzzy Logic with T-test has the good results. The proposed forecasting model do not need to know form of the regresion function, and to determinate level of the correlations of variables or parameters. Forecasting results are (1) more arccurate with correlation of temperature and previous days data rather than with temperature only, and (2) the effort of finding crisp function for forecasting is not help to have better results.

References

[1] D.N. Dinh, Power System, Science and Technics Publishing House. (1986) Hanoi, Vietnam.

[2] B.K. Chauhan, M. Hanmandlu, Load forecasting using wavelet fuzzy neural network. International Journal of Knowledge-Based and Intelligent Engineering Systems. 14 (2010) 57-71.

[3] Xiaoxi Li, Electricl Load Forecasting Based on Fuzzy Wavelet Neural Networks. Conference on Future Biomedical Information Engineering. (2008) 122-125.

[4] Yuancheng Li; Bo Li; Tingjian Fang, Short-term load forecast based on fuzzy wavelet support vector machine, Intelligent Control and Automation, WCICA Fifth World Congress, 6 (2004) 5194-5198.

[5] S. Chiu, Fuzzy Model Identification Based on Cluster Estimation, Journal of Intelligent and Fuzzy Systems, 2 (1994) 267-278.

[6] P.T.T.Binh, N.T.Hung, P.Q.Dung, Lee-Hong Hee, Load Forecasting Based on Wavelet Transform and Fuzzy Logic, POWERCON 2012, Aukland, (2012).

[7] Juneho Park. Short-term Electric Load Forecasting Based on Wavelet Transform and GMDH. Journal of Electrical Engineering & Technology, Vol.10, 3 (2015) 832-837.