Similarity Clustering and Combination Load Forecasting Techniques Considering the Meteorological Factors

YI-XIONG JIN[1]; JUAN SU[2]
1. Department of Electrical Engineering; 2. Department of Information Center of Library
Shanghai University of Electric Power
Yang-pu District, Shanghai, 200090
CHINA

Abstract: - A similar historical load data search technique, which took the sum of meteorological load and secular trend load as clustering center, was put forward. This method can improve the similarity of the loads between forecast day and sample days, and so as to improve the reliability and precision of load forecast. Manifold load forecast methods assembled by optimal weight were also applied. Technical application manifest these techniques can represent the load character of different areas, types and weather sensitivities, so as to have robust adaptability, and then have higher precision even to those small, big amplitude of vibration and weather-sensitive load.

Key-Words: - Load forecast, Meteorological factors, Linear regression, Time series, Gray model, Neural network, Combination forecast

1. Introduction

Short-term load forecasting is important to the economical and safe operation of power system. A very important characteristic of short-term load forecasting is that it's affected by the weather changes. A lot of researches have carried on load forecasting considering the meteorological factor now [1-4]. In order to reduce the neuron of input layer of neural network, meteorological factor is combined by human adaptive factor in paper [5], and so a lot of meteorological factors are united in to one factor. It can accelerate the calculate speed, but the human adaptive factor is an experience data, and that paper did not provide the calculation method of it. In addition, human adaptive factor is unable to provide such relations as flooded field drainage load, small hydro-power plant generation with regard to meteorological factor. Considering these methods taking the meteorological factor into account synthetically, two questions exist. First, they considered meteorological factors independently, while not considered their interrelations, such as the relation between temperature & humidity, temperature & wind-force, and rainfall & temperature. Second, when the weather state changes bigger, the forecasting results are out of the confidential interval, so it will reduce the precision of forecasting greatly. In order to improve the precision of load forecasting, load history is divided into secular trend load and meteorological load which is influenced by meteorological factors, and meteorological load history is assembled by linear regression model to calculate the meteorological load of forecasting day, then it was added with the secular trend load as the cluster centre, and then predicting samples can be found by the cluster centre. In addition, the prediction information and precision offered by different prediction methods are different. In order to improve the precision, this paper combined linear regression model, time series, gray model and neural network by optimal weight dynamically.

2. Load decomposing

There are lots of factors influencing the load. How to draw the meteorological load out of total load is very important. Generally speaking, load can be divided into secular trend load and meteorological load:

$$\boldsymbol{L} = \boldsymbol{L}' + \boldsymbol{L}'' \qquad (1)$$

$\boldsymbol{L}'$ is secular trend load no affected by meteorological factor; $\boldsymbol{L}''$ is meteorological load affected by meteorological factor.

Secular trend load $\boldsymbol{L}'$ can be represented by the following linear function:

$$l'_t = at + b \qquad (2)$$

l'_t is secular trend load at moment *t*; *t* is time (the serial number of the load samples); *a* and *b* are the coefficient. There estimate coefficient *a* and *b* by least square method, using total load value l_t to replace l'_t, because the value of l'_t can't be known in advance:

$$Q(a,b) = \sum_{i=1}^{n} (l_{ti} - at_i - b)^2 \qquad (3)$$

n is number of sample.

Extreme value principle is $\dfrac{\partial Q}{\partial a} = \dfrac{\partial Q}{\partial b} = 0$, i.e.

$$\left.\begin{aligned} \frac{\partial Q}{\partial a} &= -2\sum_{i=1}^{n}(l_{ti} - at_i - b)t_i = 0 \\ \frac{\partial Q}{\partial b} &= -2\sum_{i=1}^{n}(l_{ti} - at_i - b) = 0 \end{aligned}\right\} \qquad (4)$$

a and *b* can be solved out:

$$\left.\begin{aligned} a &= \frac{n\sum_{i=1}^{n} t_i l_{ti} - \sum_{i=1}^{n} t_i \sum_{i=1}^{n} l_{ti}}{n\sum_{i=1}^{n} t_i^2 - (\sum_{i=1}^{n} t_i)^2} \\ b &= \frac{1}{n}\sum_{i=1}^{n} l_{ti} - \frac{a}{n}\sum_{i=1}^{n} t_i \end{aligned}\right\} \qquad (5)$$

Then we can solve out (*a*1, *b*1) ~ (*a*96, *b*96) for 96 points load every day. Meteorological load can be deduced by $\boldsymbol{L}'' = \boldsymbol{L} - \boldsymbol{L}'$.

3. Linear regression model of meteorological factor and meteorological load

Because the history data of weather usually can only offer the real-time data including temperature, wind-force, humidity and rainfall of the 24-hour integral point for one day, while the weather forecast only provide the highest, minimum temperature, average wind-force, the highest, minimum humidity and the weather state, etc., it need to make the real-time weather history data keep in line with the form of the weather forecast data. The treatment process is as follows:

$\overline{T}_t = \max(T_t^1, T_t^2, \cdots T_t^{24})$, which is the maximum temperatures of sample day *t*;

$\underline{T}_t = \min(T_t^1, T_t^2, \cdots T_t^{24})$, which is the minimal temperatures of sample day *t*;

$\overleftrightarrow{W}_t = \dfrac{1}{24}\sum_{i=1}^{24} W_t^i$, which is the average wind-force of sample day *t*;

$\overline{H}_t = \max(H_t^1, H_t^2, \cdots H_t^{24})$, which is the maximum humidity of sample day *t*;

$\underline{H}_t = \max(H_t^1, H_t^2, \cdots H_t^{24})$, which is the minimal humidity of sample day *t*;

Then the whether factors can be grouped into:

$(\boldsymbol{L}''_t, \overline{T}_t, \underline{T}_t, \overleftrightarrow{W}_t, \overline{H}_t, \underline{H}_t, P_t) \quad (t = 1,2,\cdots,n)$,

in which P_t is the rainfall of sample day *t*.

Linear regression equation can be set up as following:

$$\begin{aligned} l''_t = x_0 + x_1\overline{T}_t + x_2\underline{T}_t + x_3\overleftrightarrow{W}_t \\ + x_4\overline{H}_t + x_5\underline{H}_t + x_6 P_t \end{aligned} \qquad (6)$$

The 96 point of forecasting meteorological load of forecasting day can be calculated by linear regression equation (6) using least square method.

4. Samples selection based on similar clustering considering the meteorological factors

Samples selection is an important factor influencing the precision of load forecasting. Samples selection based on similar clustering considering the meteorological factors is as follows:

a) Classify the history load data by date type, for instance such types as normal day, Saturday, Sunday, festivals or holidays, etc.
b) Load centre of forecasting day can be calculated as follows:

$$\boldsymbol{L}(t) = \boldsymbol{L}'(t) + \boldsymbol{L}''(t) \qquad (7)$$

$\boldsymbol{L}(t)$: Load centre of forecasting day;

$\boldsymbol{L}'(t)$: Secular trend load of forecasting day, which is calculated by linear regression equation (2);

$\boldsymbol{L}''(t)$: Meteorological load of forecasting day, which is calculated by linear regression equation (6);

Then, the samples data can be got by the load centre. That is selecting several top samples whose distances to the load centre is minimal.

This samples selection method can improve the similarity of the loads between forecasting day and sample days, and so as to improve the reliability and precision of load forecast. It can represent the load character of different areas, types and weather sensitivities, so as to have robust adaptability.

5. Combination load forecasting techniques

Different load forecasting method can offer different information and precision. If simply choose the method whose error is minimal as the optimum method, some important information may be lost, especially for those area with many kinds of load characteristic. So, a better method is to combine several forecasting method by dynamic optimal weight. Studied methods in this paper include multiple linear regressions [6, 7], time series [8-10], gray model [11-13] and neural network [14, 15]. These methods can be looked up in relevant documents.

If there have forecasting results $\widehat{Y}_{ji}$ (j=1, 2, …,m; i=1, 2, …,n) calculated by m kind of load forecasting methods, and Y_i(i =1, 2, …,n) is real load data for day i. $\varepsilon_{ji} = \widehat{Y}_{ji} - Y_i$ (i=1, 2, …, n) are the load error for day i of method j; following combination method can be use to calculate the final forecasting results:

$$\widehat{Y}(i) = \sum_{j=1}^{m} \omega_j \widehat{Y}_{ji} \quad (i=1,\ 2,\ \ldots,n) \qquad (8)$$

ω_j is the optimal weight for method j, and subject to the following restrictions:

$$\left.\begin{aligned} &\sum_{j=1}^{m} \omega_j = 1 \\ &\omega_j \ge 0,\ j = 1,\ 2,\ \cdots,\ m \end{aligned}\right\} \qquad (9)$$

Then the load error of the combined method is $\varepsilon_i = \widehat{Y}(i) - Y_i$, i.e.:

$$\begin{aligned} \varepsilon_i &= \sum_{j=1}^{m} \omega_j \hat{Y}_{ji} - Y_i = \sum_{j=1}^{m} \omega_j \hat{Y}_{ji} - Y_i \sum_{j=1}^{m} \omega_j \\ &= \sum_{j=1}^{m} \omega_j \varepsilon_{ji} \end{aligned} \qquad (10)$$

Least square method can be used to calculate the optimal weight ω_j, and then following result can be anticipated:

$$\boldsymbol{W} = (\boldsymbol{H}^{-1}\boldsymbol{e}) \big/ (\boldsymbol{e}^T \boldsymbol{H}^{-1} \boldsymbol{e}) \qquad (11)$$

In which:

$\boldsymbol{W} = (\omega_1, \omega_2, \cdots, \omega_m)^T$, $\boldsymbol{e} = (1,1,\cdots,1)^T$

$$\boldsymbol{H} = \begin{bmatrix} h_{11} & \cdots & h_{1m} \\ \cdots & \cdots & \cdots \\ h_{m1} & \cdots & h_{mm} \end{bmatrix}$$

$$h_{jk} = h_{kj} = \sum_{i=1}^{n} \varepsilon_{ji} \varepsilon_{ki} \quad (j, k = 1,\ 2,\ \ldots,\ m)$$

6. Simulation

The similarity clustering and combination load forecasting techniques has already used in Liu-an Utilities Electric Co. of An-hui Province. The city is under the foot of Da-bie Mountain, whose average load is about 250,000-450,000kW. Its load is smaller, has big fluctuating, and affected by small hydro-power plant in mountain area, whose total installation capacity is about 100,000kW. Therefore, it is quite obviously to be influenced by meteorological factors. It is relatively difficult for load forecasting. The meteorological data is sent from meteorological observatory by FTP.

Figure 1 shows the load decomposing of its load on August 10, 2006 of Liu-an.

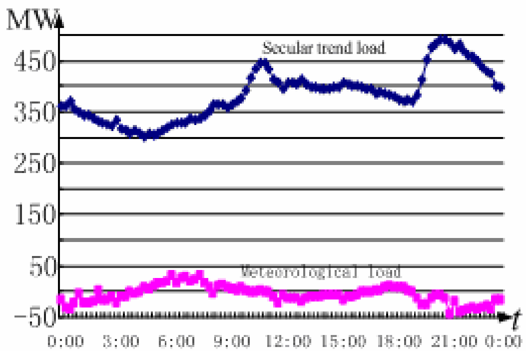

Figure 1 decomposing of the Load curve

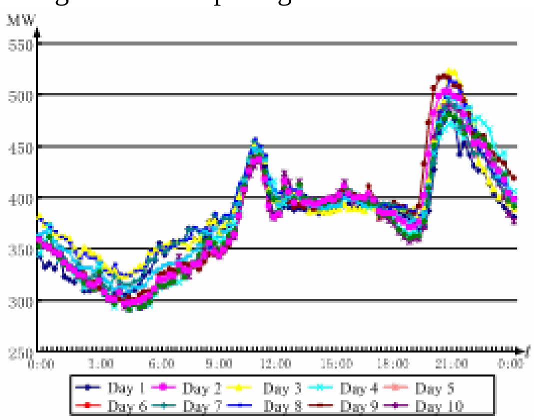

Figure 2 sample data curves based on similarity clustering

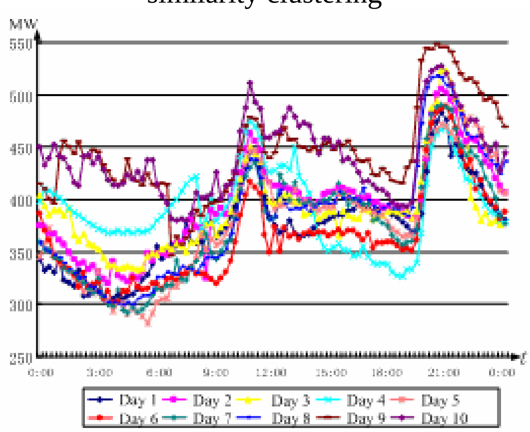

Figure 3 sample data curves got by ordinary method

Figure 2 shows the load samples got by similarity clustering, while Figure 3 shows the load samples got by ordinary method, which simply chose the previous ten days' load data. Comparing Figure 2 and Figure 3, it can easily find out that the samples got by similarity clustering method putted forward in this paper is quite alike, which can increase the credibility and precision of forecasting result.

The dynamic optimization weights of four kinds of methods calculated by program are shown in Figure 4. Figure 5 shows the forecasting results and its real-time load values.

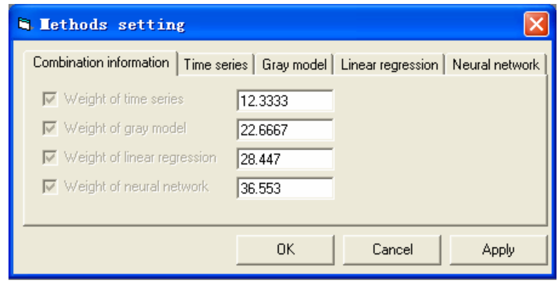

Figure 4 Optimization weight

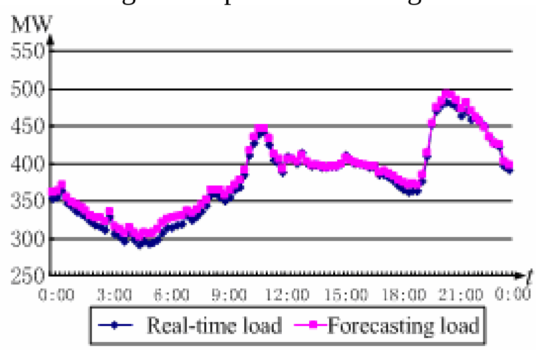

Figure 5 Forecast results

Statistics result from July to August of 2006 of Liu-an are shown in table 1:

Table 1 the load forecasting statistics result

The biggest error (%)	7.13	Minimal error (%)	0.02
Mean error (%)	3.47	Square error (%)	2.31

As can be seen from table 1, the similar clustering and combination load forecasting technology has higher precision, and the square error is relatively little, which manifests the result all have higher precision.

7. Conclusion

This paper analyzed the load forecasting methods considering the meteorological factors in existence, and their flaws, then put forward a similar historical load data clustering technique, which divide the load into two parts, constitute their linear regression models respectively, and select several top history data whose distances to the load cluster centre are minimal as samples. This method can improve the similarity of the loads between forecast day and sample days, and so as to improve the reliability and precision of load forecast. It can represent the load character of different areas, types and weather sensitivities, so as to have robust adaptability. Dynamic optimal weight combination load forecasting method is also applied to improve the precision of the forecasting results. The program based on these methods has already succeeded in applying to the Liu-an Utilities Electric Co. in An-hui Province.

References:

[1] Taylor JamesW, Roberto Buizza. Neural network load forecasting with weather ensemble predictions [J]. *IEEE Transaction on Power Systems*, 2001, 17(3) : 626-632.

[2] Zhu Sheng, Jiang Chuanwen, Hou Zhijian. Application of a weather component based elman neural network to short-term load forecasting [J]. *Proceedings of the CSU-EPSA*, 2005, 17(1) : 23-26.

[3] Yang Zhengling, Tian Yong, Lin Kongyuan. Application of multi-step regression considering climate factors in method synthesizing double period s and chaotic components to short term load forecasting [J]. *Power System Technology*, 2004, 28(12) : 20-24.

[4] Li Jinghua, Li Ran, Niu Dongxiao. Application of mining default rules based on rough set in power system short-term load forecasting [J]. *Power System Technology,* 2006, 30(5): 18-23.

[5] Qin Haichao, Wang Wei, Zhou Hui. Short-term electric load forecast by inducting human body amenity [J]. *Journal of Electric Power*, 2006, 21(2): 143-145.

[6] Kyung-Bin Song, Young-Sik Baek, Dug Hun Hong. Short-term load forecasting for the holidays using fuzzy linear regression method [J]. *IEEE Transaction on Power Systems*, 2005, 20(1) : 96-101.

[7] He Jing, Wei Gang, Xiong Lingling. Fuzzy improvement of linear regression analysis for load forecasting [J]. *East China Electric Power*, 2003, 11: 21-23.

[8] Nima Amjady. Short-term hourly load forecasting using time-series modeling with peak load estimation capability [J]. *IEEE Transaction on Power Systems*, 2001, 16(3) : 498-505.

[9] Lei Shaolan, Sun Caixin, Zhou Quan et al. Method of multivariate time series of short-term load forecasting [J]. *Transactions of China Electriotechnical Society*, 2005, 20(4): 62-67.

[10] Zhang Lin, Liu Xianshan, Yin Hejun. Application of support vector machines based on time sequence in power system load forecasting [J]. *Power System Technology*, 2004, 28(19): 38-41.

[11] Yun Lu, Xin Lin. Short-term Load Forecasting Based on Grey Model and Correction System [C]. *2005 IEEE/PES Transmission and Distribution Conference & Exhibition*: Asia and Pacific Dalian, China: 1-5.

[12] Cao Guojian, Huang Chun, Long Hui et at. Load forecasting based on improved GM(1,1) Model [J]. *Power System Technology*, 2004, 28(13) : 50-53.

[13] Yu Jianming, Yan Fei, Yang Wenyu et at. Gray variable weight combination model for middle and load term load forecasting [J]. *Power System Technology*, 2005, 29(17): 26-29.

[14] Zhou Dianmin, Guan Xiaohong,Sun Jie, Huang Yong. A short-term load forecasting system based on BP artificial neural network [J]. *Power System Technology*, 2002, 26(2): 10-13, 18.

[15] Ma Wenxiao, Bai Xiaomin, Mu Lianshun. Short term load forecasting using artificial neural network and fuzzy inference [J]. *Power System Technology*, 2003, 27(5): 29-32.

Supported by 2007 scientific research project of Shanghai Educational Commission (07ZZ145); The special fund for outstanding young teacher's scientific research of Higher Education of Shanghai; Shanghai leading academic discipline project (P1301); and keystone project of Shanghai Science and Technology Commission(061612040).