Short-term Load Forecasting Using Improved Similar Days Method

Qingqing Mu
College of hydroelectricity and digitalization
Eng. HUST
WuHan ,China
mqq-510@163.com
496673668@qq.com

Yonggang Wu
College of hydroelectricity and digitalization
Eng. HUST
WuHan ,China
ygangwu@163.com

Xiaoqiang Pan, Liangyi Huang, Xian Li
Hainan power grid corporation
Dispatching centre. Hainan power grid corporation
HaiKou,China

Abstract—**Short-term load forecasting is the basis for the safe operation of power systems. The accuracy of forecasting will have a direct impact on the load distribution of the entire power grid. There are many factors affecting the load, while the method based on similar historical days' data can fully consider these factors. It forecasts load by selecting similar historical days' data and then obtaining a weighted average from them. However, in previous studies, the weights of similar days selected are not obvious, which cannot reflect the importance of the most similar days, and results in a big forecasting error. In this paper, the weight of the most similar days is increased so as to embody the influence of the most similar days on the forecasting load,and then weighted average of the selected similar days is used to predict the load of 96 periods. At the same time, it makes an analysis on how to select similar days and situations without similar days. Moreover, it forecasts the load of a certain week of June in Hainan, and the forecasting results are more desirable than previous methods.**

Keywords- Short-term load forecasting, Similar Days Method, Setting of similarity weight, Selecting similar days

I. INTRODUCTION

Short-term load forecasting is the basis for safe and economic operation of power systems, which plays an important role of power planning of regions or areas under the control of the power grid. Currently, there are many ways used for short-term load forecast [1-2], such as time series method, artificial neural networks and so on. Time series method is based on the law of load development to carry out the forecasting. It needs less data and the calculation is simple, but it cannot fully take into account the load impact factors. Artificial neural network is to forecast through training the relationship between the load and each factor, but it requires a large number of samples for training in order to obtain load variation law. It is insensitive to the new law embodied by a small number of emerging samples [5].

With the modern socio-economic development, factors affecting the load are becoming increasingly diversified, in particular, the impact of meteorological factors on the load is becoming greater [3-4]. At the same time, people found that the load of two days with similar impact factors such as meteorological conditions, week type (workday or day off) is relatively close to each other too when doing short-term load forecasting [6]. Similar days method is set up according to this principle. It first selects similar historical days with similar characteristics, and then uses weighting and extrapolation to forecast the daily load.

However, in the previous study of similar days method [7], as for the simple weighted average method, the weight of the most similar days is set to be almost the same as the one of others when normalizing the similarity. As the load of two days with similar impact factors is close to each other, it cannot reflect the importance of the most similar days and results in a large error of load forecasting. In this paper, when normalizing the similarity, the similarity of each similar day is given a processing of nth power so as to increase the weight of the most similar days and less similar days, accordingly reducing the load forecast error. Moreover, it makes an analysis on how to select h similar days, how to determine h, and on situations with no similar days.

II. PRINCIPLE OF SIMILAR DAYS METHOD

A. Basic principle

Many factors influencing the daily load of power system, such as weather conditions, temperature, day type, and so on. An index-mapping database is designed for each factor to obtain the mapping value of each factor [7], and then, the similarity of day characteristics is introduced to evaluate the similarity between the historical day and the forecasting day, and finally h similar days are selected to forecast the load.

B. Establishment of mapping database

For Hainan Province, there are the following characteristics: date type, weather type, week type, maximum temperature, minimum temperature, date difference. Because some of the characteristics do not possess value, and cannot be compared, the characteristics of each factor needs to be give a non-dimensional treatment. Under normal circumstances, these characteristics must be mapped to the interval [0, 1] so that they can be compared. But in order to embody the "leading" role of a certain factor, it can be mapped to the interval[0,a], where a>1. The mapping database is built as shown in Table 1.

TABLE I. INDEX-MAPPING DATABASE

Name of characteristics	**Description of characteristics**	**Mapping value**	**Name of characteristics**	**Description of characteristics**	**Mapping value**
Day type	Normal day	0	Maximum temperature	37°C	4
Day type	New Year's Day	2	Maximum temperature	36°C	3.6
Day type	The Spring Festival	2.5	Maximum temperature	35°C	3.4
Day type	…	…	Maximum temperature	…	…
Weather type	sunny	0.1	Minimum temperature	5°C	3.9
Weather type	cloudy	0.1	Minimum temperature	6°C	3.6
Weather type	ovdercast	0.2	Minimum temperature	7°C	3.4
Weather type	…	…	Minimum temperature	…	…
Week type	Monday	0.1	Date difference	0 day	0
Week type	Tuesday	0.2	Date difference	1 day	0.2
Week type	…	…	Date difference	2 days	0.4
Week type	Sunday	2	Date difference	…	…

The following non-dimensional treatment is carried out to the above characteristics. There are two date types, one is ordinary day, and the other is holiday. The map value of ordinary days is 0. The map value of holidays can be greater than 1 so as to reflect the dominant role of holidays.

For weather impact factors, according to the actual weather situation in Hainan, the map value falls between the interval of [0, 1]. Rainstorms and heavy rainstorms have significant influence on the load, their values are set to be 1.2 and 1.5.

For week type, according to the actual working situations in Hainan Province, the load of Monday to Saturday is similar, and the mapping value is set to be 0.1,0.2, ... 0.6, respectively. Sunday is an off day, which has great impact on the load, therefore, its mapping value is set to be 2.

For temperature factor, taking the highest temperature as an example, 29 degree and 36 degree is assumed to be the threshold. When the temperature is below 29 degree, the linear mapping is adopted, with little change in the value. When the temperature is between 29 degree and 36 degree, the linear mapping is also adopted, but with a relatively large change in the value. When the temperature is above 36 degree, non-linear mapping is used. Every time the temperature is augmented by one degree, the change becomes much greater. It is similar to the handling of situations with minimum temperature.

For date difference, the "near is small ,far is great " principle is adopted, and a linear mapping is used. The value of the forecasting day is set to be 0, the day before the forecasting day is set to be 0.2, the day followed by is 0.4,0.6 ....respectively.

C. Similarity of different days

After the index-mapping database has being built and the mapping value of the characteristic of each factor has being obtained, the similarity can be introduced to describe the similarity between two days. The quantitative indicator of each day is set as follows: $W_i=(w_{i1},w_{i2},\dots,w_{im})$, which is the mapping value of m characteristics of the ith day. The similarity[8] of two days is computed as follows:

$$r_{ij} = \frac{\sum_{k=1}^{m} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{m} w_{ik}^2 \sum_{k=1}^{m} w_{jk}^2}} \tag{1}$$

where r_{ij} represents the similarity of the daily characteristics. It is the cosine of the angle between two vectors in the m-dimensional space. It reflects the distance of the characteristics of the two days in the m-dimensional space. Larger r_{ij} indicates that the factor of day i and day j is similar in terms of integrated meaning. On the contrary, smaller r_{ij} is corresponding to lower similarity between the two days.

D. Prediction algorithm

After the above index-mapping database has being built and non-dimensional treatment has being adopted for each factor influencing the load, select the historical load and meteorological factors of n days before the day to be forecasted. In this paper, the daily load is 96 periods, denoted as $L_i=(l_{i1},l_{i2},\dots,l_{iT})$, which represents the load value of each period of the ith day (where T = 96). The daily characteristics is obtained through the index-mapping database based on the factors of forecasting day and historical days, and then the similarity of n days is computed through formula (1). Select h days with high similarity.

1) Normalizing the similarity of the h days:

$$r_{i0} = \frac{r_{i0}}{\sum_{i=1}^{h} r_{i0}} \tag{2}$$

Where r_{i0} is the similarity of the *i*th day.

2) The load of forecasting day is the weighted average of h days' load:

$$l_{t0} = \sum_{k=0}^{h} r_{k0} l_{kt} \qquad t=1,\ 2,\ \dots,\ T \tag{3}$$

where l_{t0} is the forecasting value of the tth period of the forecasting day, l_{kt} is the historical load of the tth period of the kth day in h days. T=96 means that there are altogether 96 periods. This way the load of each period is forecasted.

III. IMPROVEMENT OF SIMILAR DAYS METHOD

A. Setting of similarity weight

In the previous similarity normalization formula (2) in the second part, there is no n-th power, only simple normalization is used, which results in some problems, that's, there is no obvious distinction between the similarity of the most similar days and less similar days, and the load forecasted cannot reflect the characteristics of the most similar days, accordingly, there is a big error of the load forecasted. Therefore, the following formula (4) is used to increase the weight of the most similar days.

$$r_{i0}' = \frac{r_{i0}'^{n}}{\sum_{i=1}^{h} r_{i0}'^{n}} \tag{4}$$

Where the power coefficient n needs to be determined later according to the actual load in the area to be forecasted. In the above formula, nth power is used, which can increase the weight of the most similar days and less similar days. Therefore, the forecasting load contains more information about the most similar days, which results in a lower error of forecasting load. In this paper, the power coefficient n is set to be 110. We found that when n is less than 110, the error decreases as n increases, and as n reaches 110, the error reaches the lowest point; while when n is greater than 110, the error remains unchanged, thus 110 is the most appropriate value for n.

B. Selecting h similar days

When selecting h similar days to predict load, as there are many historical days selected and the similarity of each day is not the same at all, some historical days with low similarity with the forecasting day can be abandoned, and only days with high similarity are taken so that a more accurate load can be forecasted.

The selection of h days can be determined as follows: first sort various similarities in descending order, and then set up a similarity threshold; those greater than this threshold can be retained, while those less than this threshold are abandoned. In this paper, the threshold is set to be 0.6. Those less than 0.6 are almost dissimilar at all and cannot be selected. A part of historical similar days are selected through the threshold, and then only the most similar h days are kept. h is set according to the actual situations. If h is set to be very large, the forecasting load will be affected by the daily load of those dissimilar days. If h is set to be very small, the daily load of the historical days is too single, which is not good for load forecasting. Thus it is inappropriate to assign h a very big value or a very small value.Here 7 days are being selected, and the total historical days are 29 days. However, if the historical days greater than the threshold are less than 7 days, then only those days（h<7） greater than the threshold can be selected.

C. Treatment of situations with no similar days

When there are no historical similar days greater than the threshold, i.e. the similarity of the historical days all is less than 0.6, we can think that there is no historical days similar to the day to be forecasted. Such situation generally happens when there is more than 20 continuous days in the high temperature, but suddenly rain follows. Under this situation, the similarity obtained is very low, almost unlike at all. If so, n days before this n historical days should be used again to forecast the load, that's, the n days before the forecasting day are abandoned, while the n days before the abandoned n days are kept to predict the load.

For holiday forecasting, as holidays have been listed in the index-mapping database, and a high value is assigned, the "dominant" factor can be automatically recognized when computing the similarity. This way, the nearest holidays and the selected historical days can be grouped together for the selection of similar days, and then take use of formula (3) and (4) to forecast the load.

IV. CASE STUDY

Select the load and weather data of a week in June 2008 in Hainan to forecast the load. The historical days selected is 29 days. Then use the above methods to forecast the shot-term load of 96 periods. With respect to the error index used in Hainan, 5 error index are used in this paper, namely, maximum error, minimum error, electrical quantity error, root mean square error, average relative error. The parameters are shown as table 2. The comparison between the forecasted load and the actual load of June 18 is made, as figure 1 shows.

A. Maximum error:

$$E_{max} = \left| \frac{Y_{max} - Y'_{max}}{Y_{max}} \right| * 100 \tag{5}$$

where Y_{max} is the actual maximum load value, and Y'_{max} is the forecasted load value of the actual point with maximum load value.

B. Minimum error:

$$E_{min} = \left| \frac{Y_{min} - Y'_{min}}{Y_{min}} \right| * 100 \tag{6}$$

where Y_{min} is the actual minimum load value, and Y'_{min} is the forecasted load value of the actual with minimum load value.

C. Electrical quantity Error:

$$E_{zon} = \left| \frac{Y_{zong} - Y'_{zong}}{Y_{zong}} \right| * 100 \tag{7}$$

where Y_{zong} is the actual total load value of the day, and Y'_{zong} is the forecasted total load value for this day.

D. Root mean square error:

$$E_{MSE} = \sqrt{\frac{1}{T} \sum_{i=1}^{T} \left(\frac{Y_i - Y'_i}{Y_i} \right)^2} * 100 \tag{8}$$

where Y_i is the actual load, Y'_i is the forecasted load, and T = 96.

E. Average relative error:

$$E_{MAPE}=\frac{1}{T}\sum_{i=1}^{T}\left|\frac{Y_i-Y_i^{'}}{Y_i}\right|*100 \tag{9}$$

As can be seen from the figure, the improved method can well forecast the load trends.

TABLE II. ERROR INDICATORS OF FORECASTED LOAD（%）

error	Maximum error		Minimum error		Electrical quantity error		Root mean squre error		Average relative error	
date	***Before improvement***	***After improvement***	***Before improvement***	***After improvement***	***Before improvement***	***After improvement***	***Before improvement***	***After improvement***	***Before improvement***	***After improvement***
2008-06-15	3.23	0.00	6.93	0.00	3.98	0.47	4.51	0.84	4.13	0.51
2008-06-16	2.26	0.00	5.45	0.50	4.71	1.04	4.95	1.46	4.75	1.20
2008-06-17	1.29	0.00	1.01	2.02	3.13	1.23	3.33	2.00	3.06	1.67
2008-06-18	0.32	0.00	0.51	2.02	1.70	0.39	2.01	1.07	1.73	0.84
2008-06-19	0.32	0.00	1.52	0.51	0.56	0.07	1.23	0.35	1.05	0.24
2008-06-20	0.96	1.27	0.50	1.00	1.23	0.91	1.46	0.99	1.24	0.91
2008-06-21	1.27	1.27	0.00	1.00	1.31	0.91	1.47	0.99	1.29	0.91
平均值	1.38	0.36	2.27	1.01	2.37	0.72	2.71	1.1	2.46	0.90

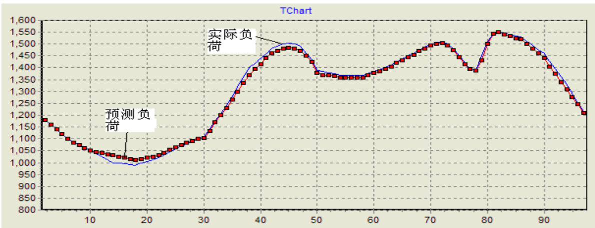

Figure 1. comparison between the actual load and the forecasted load of June 18

From the above table, the following analysis can be made:

- From the error of each day, we can see that the improved method decreases the load error greatly. Each kind of error is in agreement with the requirement of 2%, and the average error is within the range of 1.5%.
- Among the five kinds of error, electrical quantity error, root mean square error, and average relative error is more inconsistent with each other. When electrical quantity error is small, root mean square error and average relative error is also small. Anyway, the error after improvement is much less than the one before improvement. However, the maximum error and the minimum error is inconsistent. When the error before improvement is large(or small), it is small after improvement.

V. CONCLUSIONS

In this paper, through increasing the weight of the most similar days, the forecasting error decreases greatly. At the same time, we made a discussion on how to select similar days and situation without similar days. Under both situations, the results are satisfactory. As the information affecting load is not comprehensive, if the amount of information is increased, the forecasting load will be more accurate. At the same time, some adjustment on certain characteristics must be made in time according to weather variance and the change of some dominant factors. Similar days method can also combined with other methods like gray theory for load forecasting, and the result would be better.

REFERENCES

[1] Pockard N H. Geometry form a Time Series . Phy., Rev, Lett., 1980, 712:45.

[2] Lee K Y, Cha Y T , Park J H. Short-term Load Forecasting Using an Artificial Neural Network. IEEE Trans on Power Systems , 1992 , 7 (1) :534-537.

[3] KANG Chongqing , ZHOU Anshi , WANG Peng , et al . Impact analysis of hourly weather factors in short-term load forecasting and its processing strategy. Power System Technology , 2006 ,30 (7) : 5-10.

[4] XIE Hong , CHEN Zhiye , NIU Dongxiao. Principle component analysis of data processing for short-term load forecasting. Power System Technology , 2000 , 24 (1) : 43-46.

[5] TOMONOBU S , HITOSHI T , KATSUMI U , et al . One-hour-ahead load forecasting using neural network. IEEE Trans on Power Systems, 2002 , 17 (1) : 113-118.

[6] JU Ping , JIANG Wei, ZHAO Xiayang. Ninety-six points short-term load forecasting — theory and applications.Automation of Electric Power Systems , 2001 , 25 (22) : 32-36.

[7] KANG Chongqing , CHENG Xu , XIA Qing , et al . A new unified approach to short-term load forecasting considering correlated factors. Automation of Electric Power Systems ,1999 , 23 (18) : 32-35.

[8] MO Weiren , ZHANG Boming , SUN Hongbin , et al . Method to select similar days for short-term load forecasting. Journal of Tsinghua University : Natural Sciences , 2004 , 44 (1) : 106-109.