The Delicate Analysis of Short–Term Load Forecasting

Changwei Song\textsuperscript{1} and Yuan Zheng\textsuperscript{2}

\textsuperscript{1} Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University, Jinan, China
\textsuperscript{2} State Grid Jiangxi Electric Power Company Jiujiang Electric Power Supply Company, China

Corresponding author’s e-mail address: 535224087@qq.com

Abstract. This paper proposes a new method for short-term load forecasting based on the similar day method, correlation coefficient and Fast Fourier Transform (FFT) to achieve the precision analysis of load variation from three aspects (typical day, correlation coefficient, spectral analysis) and three dimensions (time dimension, industry dimensions, the main factors influencing the load characteristic such as national policies, regional economic, holidays, electricity and so on). First, the branch algorithm one-class-SVM is adopted to select the typical day. Second, correlation coefficient method is used to obtain the direction and strength of the linear relationship between two random variables, which can reflect the influence caused by the customer macro policy and the scale of production to the electricity price. Third, Fourier transform residual error correction model is proposed to reflect the nature of load extracting from the residual error. Finally, simulation result indicates the validity and engineering practicability of the proposed method.

1. Introduction

As an important part of Energy Management System (EMS), load forecasting has been a hot spot of research in power systems [1-3]. With the acceleration development of electric power industry market-oriented reform in recent years, the importance of load forecasting, especially short-term load forecasting is becoming more and more obvious. There are many methods of short-term load forecasting [4-7], the commonly used prediction methods are time series method, similar day method, neural network method and so on. Similar day method [5] has been proved to be an effective approach to study the rule of load change. In this paper, the selection of typical day is obtained by the branch algorithm one-class-SVM. It is an expansion and extension of the traditional SVM, using hyper-sphere instead of hyper-plane to split the data is the essence of one-class-SVM.

Also, due to the complex nonlinear characteristics when different factors affect the load at the same time, using curve fitting or experience method to revise the load difference caused by the difference of factors cannot achieve good effect and stability. This paper also proposes Fourier transform residual error correction model. Fourier transform is an infinite superposition of a series of different frequency sine waves which can extract frequency components. Taking advantage of the powerful noise reduction ability of Fourier transform, putting the residual energy as a time series with limited energy, the information reflecting the nature of load can be extracted from the residual error. So, it is feasible to improve the residual error according to Fourier transform in theory.
With the rise of new energy and the maturity of distributed generation technology, the inherent attributes of the network load have been broken, the ambiguous boundary of source load aggravate the difficulty of load forecasting. Now user data can only be analyzed after the fact. We can only grasp the whole rule of load change roughly from the macro level. We can neither analyse a certain customer in a certain industry nor find the reason of the change of load characteristic.

Based on this background, by means of analyzing electric power market step by step, the typical customers, customer base and load characteristic can be stripped. By mean of researching the load analysis system from all aspects of multi-time and multi-level, taking advantage of the strong and flexible condition screening function and custom function, the change rules of the load of the various study object including the typical load curve modelling and the key factors filtering can be better analyzed by professionals.

This paper is organized as follows: In Section II, this paper proposes the theory and method. In Section III, it gives simulation results. Finally, Section IV concludes the paper.

2. Theory and Method

2.1. The selection of typical day model
In this paper, the branching algorithm of one class support vector machines (one-class SVM) is used for the selection of the typical day, this method is an expansion and extension of the traditional SVM, introducing the statistical learning theory into the unsupervised learning theory. The principle of one-class SVM to selecting typical day will be briefly introduced in the following text.

The essence of one-class SVM is the substitution of hyper-planes by hyper-sphere to split the data. The initial problem of the objective function is:

\[ \min_{R,c} R^2 + \frac{1}{v} \sum_{i=1}^{l} \xi_i \]

s.t. \[ \|x_i - c\|^2 \leq R^2 + \xi_i \]
\[ \xi_i \geq 0, i = 1, 2, \ldots, l \] (1)

By setting parameter \(0 \leq v \leq 1\), we can strike a compromise between the spherical radius and its capacity for the number of training samples. When \(v\) is small, putting the data inside the ball as far as possible. When \(v\) is large, compressing the size of the ball as far as possible.

Solving this problem by Lagrange function:

\[ L = R^2 + \frac{1}{v} \sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i \left( R^2 + \xi_i - \|x_i - c\|^2 \right) - \sum_{i=1}^{l} \beta_i \xi_i \] (2)

So:

\[ \frac{\partial L}{\partial R} = 2R - \sum_{i=1}^{l} \alpha_i \cdot 2R = 0 \] (3)

\[ \frac{\partial L}{\partial c} = \sum_{i=1}^{l} \alpha_i \cdot 2(x_i - c) = 0 \] (4)

\[ \frac{\partial L}{\partial \xi_i} = \frac{1}{v} - \alpha_i - \beta_i = 0 \] (5)

get the dual problem:
\[
\min_{\alpha} \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) - \sum_i \alpha_i (x_i \cdot x_i)
\]

\[s.t. \quad 0 \leq \alpha_i \leq \frac{1}{\sqrt{l}}, i = 1, 2, \cdots, l \]

\[
\sum_{i=1}^l \alpha_i = 1
\]

The optimized solution \( \alpha \) can be obtained by solving this dual problem through the QP optimization method.

Taking \( \alpha \) into the type (4), the value of circle \( c \) can be obtained, and this is the typical day.

As we can see, the main function of one-class SVM is to make a compromise between the empirical risk and confidence risk by clustering sample and adjusting parameter \( v \). The selection of typical day itself is also a kind of clustering. If the measurement data of every day as a sample, then the sample is a point in the 24d. Using a hyper-sphere which can cover all the sample points as far as possible, the centre of the super ball is the centre of the measured data, and that is the place where the data of the typical day is.

2.2. The correlation coefficient

Correlation coefficient indicates the strength and direction of the linear relationship between two random variables. The absolute value of the correlation coefficient is not more than 1. When the linear relationship between the two variables enhances, the correlation coefficient is tending to -1 or 1. As two variables both increase, the correlation coefficient is greater than zero. When one variable increases while the other variable decreases, the correlation coefficient is less than 0. If two variables are independent, the correlation coefficient is equal to 0. However, the converse is not true. The calculation formula of the correlation coefficient is as shown below:

\[
r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}
\]  

(7)

By each user's typical daily load curve, we can analyse the correlation of power rules between users with users and customers with the overall industry. First, calculating the correlation coefficient of \( n \) users typical daily load curve respectively, the result can be stored in the \( n \times n \) symmetric matrix with diagonal elements of 1. Off-diagonal elements represent the correlation coefficient of two users typical daily load curve, if many correlation coefficient curves converge nearby the place where the correlation coefficient of user A is 1, supposing that these curves correspond to user collection \( S \), it means that the load curve of the user A is similar to the load curves of the users in \( S \) set in shape.

2.3. Spectrum analysis

Fast Fourier transform (FFT) is a fast and efficient algorithm of DFT. The main reason why the finite length sequences occupies a very important position in digital technology is that its frequency spectrum can be discrete. The DFT of Finite length sequences itself can fully express sequence spectrum so it can directly analyze the spectrum of signals. FFT can transform a signal from time domain to frequency domain, the frequency spectrum of a signal can also be extracted by this way. Sometimes it is hard to see the signal's characteristic in the time domain, but it is easy in the time domain.

The computation to meet a DFT arithmetic of a finite length sequence \( x(n) \) is as shown below:
$$X(k) = DFT[x(n)] = \sum_{n=0}^{N-1} x(n)w_N^{nk} \quad k = 0,1,\ldots,N-1 \quad (8)$$

By this method to calculate DFT, for each K value of $X(k)$, it needs $4N$ times multiplication and $(4N-2)$ times together. For N K values, it totally needs $4N*N$ multiplication and $N*(4N-2)$ real number together. The improved DFT algorithm which uses the periodic and symmetry in the DFT, making whole DFT calculation into a series of iterative computation, can greatly reduce the computational complexity and cost, and this is the basic idea of FFT.

Here we take the time series of user load as discrete sampling points of the Fourier transform and then set the discrete Fourier transform. After the 24 point data of users has been transformed by DFT, we can grasp the frequency characteristic of power fluctuations.

3. example analysis

Large industrial power is an important part of the total electricity consumption, to make a better analysis of large industrial power law, 28 large industrial users in Weifang area have been selected to analyse the load characteristics.

3.1. The typical daily load curve analysis

After getting the typical day through the selection model of the typical day, the relevant features of the daily load curve can be analysed. From January to August 2014, the typical daily load curve of 28 large industrial users and industrial overall typical daily load curve are respectively shown in figure 1 and figure 2.

By the typical daily load curve of large industrial users, it can be seen that the local peak occurs at 3 am, 11 am and 17 pm. The intraday volatility of load is frequent, but the amplitude is small.

According to calculating the correlation coefficient between the typical daily load curves of 28 users respectively, the correlation characteristic of each user load curve can be analyzed. The figure 3 shows the correlation coefficient of load curves between users. There are many correlation coefficient curves converging nearby correlation coefficient 1 of users of number 19, 21, 22, 23. Remarkably, the typical daily load curve of users of number 21, 22, 23, 28 converge near -1 with others, showing a strong negative correlation. It illustrates that some users’ electricity law have a strong complement to them.

Figure 4 and figure 5 give typical daily load curve of user 28 and user 19, the correlation coefficient of them is 0.95.
3.2. Customer base with typical customers

By analysing the correlation coefficient between users, we can classify the users according to their electricity law. As user 21 (Litas auto parts Co., Ltd) for example, there are two users whose correlation coefficient is greater than 0.95, 15 and 16. The two users are JinTong pipe Co., Ltd and Jintai casting Co., Ltd. It can be seen that the higher load characteristics correlation coefficient with Litas auto parts Co., Ltd are mostly foundry related enterprises, and the main products of Litas auto parts Co., Ltd are casting parts for cars. We can suppose that the load curve of user 21 can represent the typical daily load curve of Casting production enterprise, user 21 is the typical user of customers base of casting production enterprises. The typical daily load curve of user 21 is as shown in figure 6 below.

In figure 6, the electricity law of casting production enterprises customer base can be seen. This kind of customers' daily maximum load occurs at 1 am, minimum load at 13 pm. The all-day electricity law is more in the day and less in the night. This is mainly due to its high energy consumption and abundant production capacity. Under the influence of time-sharing electricity price policy, through the reasonable arrangement of the production plan, the running cost of the enterprise can be effectively reduced.

Figure 7 shows the correlation coefficient of typical daily load curve between the whole large industry and each user. Table 1 lists the top 5 users in descending order by their correlation coefficients.
Table 1. The overall phase meter.

| Sort | Correlation coefficient | User number |
|------|-------------------------|-------------|
| 1    | 0.728074                | 3           |
| 2    | 0.694527                | 15          |
| 3    | 0.617865                | 23          |
| 4    | 0.583684                | 21          |
| 5    | 0.575963                | 8           |

3.3. Load characteristic of time domain analysis

As a result of the large span of the data, it is difficult to reflect the changes of electricity law between mouths. So the typical daily load curve of large industrial is established according to the mouth.

From figure 8, it is easy to see that from January to May, the shape of typical daily load curve of five months is very similar and the load fluctuation is small. The most similar correlation coefficient between them are February and March, and the correlation coefficient is 0.92. The rest is all in 0.9 which means there was some difference of typical daily load curve between different months. The daily maximum load all appear at 1 am. The local valley of load appear at 8 am, 12 am and 11 pm, and the power all decrease at 13 pm. The load levels of June, July and August are significantly higher than that of the other five months, and the intraday volatility of the load is more obvious. The reason may be that in summer, the more use of cooling equipment improves the overall load level.

3.4. Load characteristic of spectrum analysis

Now we take a large industrial user in Weifang area for example, to make spectrum analysis about the load characteristic. The spectrum of load curve and the corresponding power curve of user 15 are respectively shown in figure 9 and figure 10:

First, we analyze the situation of load fluctuation in a day (considering the part frequency greater than or equal to 30). As the spectrum figure 9, we can see that the frequency 30, 60, 90, of user 15 have large amplitude, and the amplitudes in frequency 120, 150, 180 are also significant. This frequency corresponds to month frequency which means the times the periodic change occurs in a month. Converted to day characteristics, it changes once, twice, three times in a day, the appropriate change period is 24h, 12h, 8h. The variation rule can also be roughly seen in the power curve of user 15.

To grasp the long-term load variation rules of users better, we can analyze the part which is less than 30 in the spectrum figure. In this part, the corresponding change cycle is greater than one day. The local spectrum figure which is less than 30 and the mid-long term load curve are respectively shown in figure 11 and figure 12.
By figure 11 we can see, there are no significant peaks in the part of which frequency is less than 30. This illustrates that there is no existence of a long time scale (more than a day) as the cycle fluctuations. Corresponding to the mid-long term load curve about user 15, we can also find that mid-long term load fluctuation is not obvious.

4. Conclusion
In this article, the measurement information of power is flexibly screened from different industries, times and regions. According to this way, the change law of load can be analysed on a micro level, the electricity information of users can be researched more accurately. The electric power sector can take advantage of the internal connection between users, industries and the whole society to find out the main factor which leads to the anomaly fast and accurately.

References
[1] Yan D, Zhao J G. Practicable method for load forecasting of distribution network. Automation of electric power system. 2001, 25(22):45-48.
[2] Ju P, Jiang W, Zhao X Y, Wang J K, et al. Ninety-six points short-term load forecasting-theory applications. Automation of electric power system. 2001, 25(22):32-36.
[3] Zhang X X, Zhou Q, Ren H J, et al. Input parameters in short-term load forecasting model based on incremental reduction algorithm. Automation of electric power system. 2005, 29(13):40-44.
[4] Park D C, El-Aharkawi M A, Marks R J. Load Forecasting Using an Artificial Neural Network. IEEE Trans on Power Systems, 1991, 6(6): 442-449.
[5] Cheng Song A New Approach to Load Forecasting Based on Similar Day Proceeding of Jiangsu Electrical Engineering Association 1999,18(4): 28-32.
[6] 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.
[7] Jaehak Y, Byung-Bok Lee DaeHeon Park. Real-time cooling load forecasting using a hierarchical multi-class SVDD. Multimedia Tools and Applications. July 2014, 71(1) : 293-307.