Support Vector Machine based on clustering algorithm for interruptible load forecasting

Support Vector Machine based on clustering algorithm for interruptible load forecasting

Xiang Yu[1,a], Guangfeng Bu[1],Bingyue Peng[1] ,Chen Zhang[1],Xiaolan Yang[1],Jun Wu[1], Wenqing Ruan[1],Yu Yu[1],Liangcai Tang[1]and Ziqing Zou[2]

1. State Grid Yangzhou Power Supply Company, China
2. School of Electrical Engineering, Southeast University, China

a. Corresponding author: Xiang Yu@18795969086@163.com

Abstract. Accurately forecast interruptible load can help to alleviate the power supply tension during peak load and make scheduling more flexible. Support vector machine (SVM) method which has been widely used in load forecasting usually selects a period of date close to the forecast day without considering the information characteristics of itself. An interruptible load forecasting method based on clustering algorithm is proposed in this paper. This method puts forward a new idea to select the sample of prediction model which takes full account of the weather and date information of the forecast day and solve the problem that the traditional SVM method cannot properly reflect it. In this paper, the principles of clustering algorithm and support vector machine are introduced firstly. Then K-means clustering algorithm is used to classify the historical data, and the support vector machine forecasting model is constructed by using the categories of the forecast day membership. Finally, the prediction is carried out by combining with the actual data. The results show that the prediction accuracy of this method is more than 95%, and it has higher precision.

1 Introduction

As an important component of demand side management (DSM) in power system, interruptible load management (ISM) makes use of the flexibility of users to alleviate the power supply tension during peak load. So as to avoid or reduce the expensive rotating reserve and the generation capacity investment needed to meet the growth of electricity demand, it is conducive to the safe and economic operation of the power system, weaken the impact of market power in the electricity market, and suppress price spikes. Therefore, interruptible load forecasting is an important part of power system energy management. Interruptible load forecasting not only provides a guarantee for the safe and economic operation of power system, but also is the basis of power market scheduling and power supply planning [1-2].

Support vector machine (SVM) algorithm has the characteristics of high fitting accuracy, strong generalization ability and global optimization, and has been applied to load forecasting by some scholars [3-4]. But this kind of traditional load forecasting based on support vector machine usually chooses a period of date close to the forecasting date, and the closer the forecasting date is, the more similar the characteristics of the load forecasting date are. However, even if it is only one or two days apart, the actual load can be very different, for example, the load on the day before the holiday and during the holiday is markedly different, and the adjacent days may also be greatly affected by the weather. Direct prediction can lead to a certain deviation.

Therefore, this paper proposes a support vector machine forecasting based on load clustering, which takes both meteorological factors and time factors into account. Firstly, the meteorological data and date information of historical loads are sorted out, then the historical data are clustered by K- means clustering algorithm, and the clustering results are classified according to the similarity of historical loads. The relevant information of the forecast day is obtained by querying the information and multiple historical loads are found in which are the same as the information of the forecast day. Then the support vector machine prediction model is constructed with the historical date of this category. The method is applied to the interruptible load forecasting of an enterprise in Yangzhou, and the higher forecasting accuracy is obtained, which proves the accuracy and superiority of the support vector machine forecasting method based on clustering algorithm.

2 Principle of support vector machine based on clustering

2.1 Clustering algorithm

Clustering is an analysis method which divides a group of data sets into several categories according to the similarity between the set elements [5-6]. Through clustering analysis, the attributes and characteristics of each category can be studied for further processing. Cluster analysis requires that the distance between elements in the same category is small, while the distance between different categories is very large.

2.1.1 Normalization processing

First of all, the clustering should be normalized. Normalization is a simplified way of calculating, that is, a dimensional expression, transformed into a dimensionless expression, become a scalar. Many operations can be handled in this way, not only to ensure the convenience of the operation, but also to highlight the essence of the meaning of the physical quantity [7].
Normalization definition: normalization is to process the data that needs to be processed (by some algorithm) to limit to a certain range you need. First of all, the normalization is for the convenience of later data processing, and the second is to accelerate the convergence of the preservative program when it runs. The specific function of normalization is to induce and unify the statistical distribution of samples. The normalization between 0-1 is a statistical probability distribution, and the normalization is a statistical coordinate distribution on a certain interval.
User load data obtained by power system load measurement devices will vary greatly in numerical range, and these differences will have a great impact on the results of classification. Therefore, the sample data should be normalized before classification in order to eliminate the impact of these differences. The load curves referred to in this paper are all representative daily load curves of users after normalization, so the classification of power system users is transformed into the classification of user load curves.

2.1.2 K-means clustering algorithm

According to the different principles of classification, clustering analysis can be divided into partition clustering method, hierarchical clustering method, density clustering method, grid and model-based method and so on. In this paper, we choose the k-means clustering algorithm in partition method. K-means is an iterative clustering algorithm, in which the objects in the cluster are constantly moved until an ideal cluster is obtained, and each cluster is represented by the average value of the objects in the cluster. Using k-means algorithm to get the cluster, the similarity of objects in the cluster is very high, and the difference between objects in different clusters is also very high.
The main steps of the algorithm are [8]:

(1) Randomly selecting k objects from n data objects as the initial cluster center;

(2) The average value of each cluster is calculated, and the average value is used to represent the corresponding cluster;

(3) According to the distance between each object and the center of each cluster, the cluster is assigned to the nearest cluster;

(4) The second step is to recalculate the average value of each cluster.

This process repeats until it is satisfied that a criterion function is no longer significantly changed or that the clustered objects are no longer subject to change. The criterion function of the general K-means algorithm is the squared error criterion, which is defined as:

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} \left| P - mi \right|^2 \quad (1)$$

Where E is the sum of the mean variance of all the objects in the dataset and the corresponding cluster center, p is the given data object, and m_i is the mean value of the c_i clusters (p and m are both multidimensional). The k-means algorithm is relatively scalable and efficient for large databases, and generally ends in a local optimal solution.

2.2 Principle of support vector machine (SVM)

The support vector machine (SVM) algorithm is based on statistical theory. SVM method is a successful implementation of statistical learning theory [9-10]. It is based on the VC theory of statistical learning theory and the principle of structural risk minimization. According to the limited sample information, the best tradeoff is found between the complexity of the model (the learning accuracy of a specific training sample) and the learning ability (the ability to identify arbitrary samples without errors), in order to obtain better generalization ability. It has the following characteristics:

(1) The SVM principle is implemented, which can minimize the upper bound of generalization error rather than the training error, so it has better generalization performance.

(2) Compared with neural network method, SVM has less free parameters. There are only three free parameters in the SVM algorithm, but there are a lot of free parameters in the neural network, which need to be selected subjectively by experience.

(3) The neural network may not converge to the global optimal solution, so it is easy to fall into the local optimal solution. In SVM algorithm, training SVM is equivalent to solving a quadratic convex programming problem with linear constraints, so its solution is unique, global and optimal.

The load forecasting algorithm based on support vector machine is as follows:

For the sample set $\{x_i, y_i\}_{i=1}^{l}$, , $x_i \in R^n$ as the input variable and $y_i \in R$ as the corresponding output value, a nonlinear mapping $\varphi(x)$ is introduced to map the n -dimensional input and 1-dimensional output sample sets D from the original sample space to the high-dimensional feature space, and the regression function is in the form of the following[11-12]:

$$f(x) = \omega \cdot \varphi(x) + b \quad (1)$$

In the formula: $\omega \in R^n$ is the weight and $\varphi(x)$ is the x mapped eigenvector, b is the threshold.

According to the principle of structural minimization, SVM can determine the regression function by minimizing the target, that is:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{m} \xi_i^2 \quad (2)$$

$$s.t. \quad y_i[\omega \cdot \varphi(x_i) + b] = 1 - \xi_i$$
$$\xi_i \geq 0, i = 1, 2, ..., m \quad (3)$$

Where $C > 0$ is the penalty function， ξ_i is the relaxation factor. For the constraint condition, the corresponding multiplier α_i is introduced, and the kernel method is introduced, the form of feature

mapping is considered, then the final decision function is given as:

$$\varphi(x)=\sum_{i=1}^{l}(\alpha_i^*-\alpha_i)K(x_i,x)+b \tag{4}$$

The commonly used kernel functions and their parameters are as follows [13-14]:

(1) Polynomial kernel function

$$K(x_i,x_j)=(x^T{}_i x_j)^d \tag{5}$$

(2) Linear kernel function

$$K(x_i,x_j)=(x^T{}_i x_j) \tag{6}$$

(3) Gauss kernel function

$$K(x_i,x_j)=\exp(-\frac{\left\|x_i-x_j\right\|^2}{2\sigma^2}) \tag{7}$$

Where $\sigma>0$ is the width of the kernel function.

(4) Laplace kernel function

$$K(x_i,x_j)=\exp(-\frac{\left\|x_i-x_j\right\|}{\sigma}) \tag{8}$$

Where $\sigma>0$ is the width of the kernel function.

(5) Sigmoid kernel function

$$K(x_i,x_j)=\tanh[\beta(x_i^T x_j)+\theta] \tag{9}$$

As for complex prediction problems, because of the different sources of data, their composition may also be heterogeneous data sets. This leads to the complexity of data distribution. If we only choose a single kernel function SVM model, we may not be able to achieve satisfactory prediction performance. For the common kernel functions mentioned above, they can be divided into global kernel function and local kernel function according to their influence on the regression prediction problem. The typical representative of global kernel function is polynomial kernel function, whose outstanding feature is its strong generalization performance. It can extract the whole situation of the sample data very well, but the disadvantage is that the learning ability is relatively weak. Gauss kernel function, as one of the typical representatives of local kernel function, has good learning ability for samples in a certain distance. The disadvantage is that its generalization ability is relatively poor. Therefore. Considering the merits and demerits of the two kinds of kernel functions, the two kinds of kernel functions are combined according to certain weights to form the final kernel function, which can improve the comprehensive performance of the model.

According to the above analysis, if the weight coefficient of Gauss kernel function K_G is α and the weight coefficient of polynomial kernel K_P is $1\text{-}\alpha$, then the final mixed kernel function after combination is as follows:

$$K=\alpha K_G+（1\text{-}\alpha）K_P \tag{10}$$

3 Prediction model

3.1. Load clustering based on k-means clustering algorithm

In this paper, the K-means algorithm is used to cluster the characteristic variables corresponding to 96 points daily load data of interruptible load, and then the load pattern clustering is realized.

Firstly, the influencing factors of interruptible load are analysed [15]. The interruptible load fluctuates greatly because of the weather and date. The selected weather factors are daily maximum temperature, daily average temperature, and the definition of the specific weather conditions, defining rainy days and non-rainy days, rainy days for the category 1, non-rainy days for the category 0. The date factor is reflected as whether it is a holiday or not, the load curve of the ordinary working day and the legal holiday will have great difference, so the definition of the statutory holiday is class 2, the ordinary working day is category.

3.2 Establishment of support vector machine model

From the above analysis, we can see that there are five model parameters to be chosen in this paper, namely regularization parameter C, Gauss kernel width σ 、The weight coefficient α of the Mixed kernel function K、Polynomial coefficient d and ε. Among them, C and σ play a very important role in support vector machine (SVM) algorithm.

C determines the training error and generalization ability. Too small a value will cause underlearning to the training data, and too much will easily cause an overlearning phenomenon to the training data, resulting in deterioration of generalization performance. The suitable value of C should be between 10 and 100. σ reflects the distribution characteristics of the training sample data. If the value of σ is too small, the training set will be overlearned, and the training set will be underlearned if the value of σ is too small. The suitable value of σ should be between 1 and 10.The performance of support vector machine algorithm is insensitive to ε, and the change of ε has little effect on the standard mean square error of training set and test set, so it is not affected by the variation of ε. In general, the number of support vectors decreases with the increase of ε. However, the large value of ε can also reduce the approximation accuracy of the data points, so the value of ε can not be too large. In order to give full play to the advantages of the two kinds of kernels, the mixed kernels take half of each function, that is, α is 0.5. The range of the polynomial degree d is $[0,4]$.Therefore, this paper selects the parameters of support vector machine C=50, σ=2， ε=0.001.

Therefore, the forecast flow is as follows:

1） Input historical load and historical weather date data;
2） K-means clustering method for load clustering;
3） Get 4 kinds of loads;
4） Query the information of the days to be forecasted and find out the similar load of the days to be forecasted.
5） Load forecasting of SVM with similar date load input;
6） Output prediction results.

The specific flow chart is shown in the following figure:

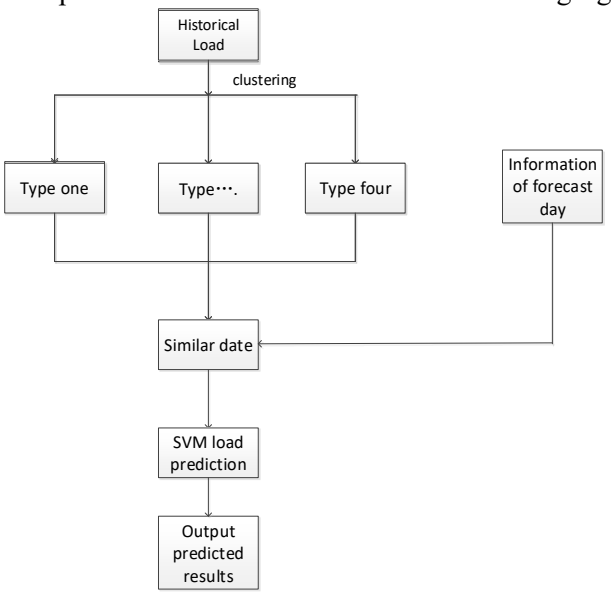

Figure 1. Prediction flow chart

4 Analysis of example

In this paper, the interruptible load of an enterprise in Yangzhou is forecasted by the above methods, and the forecasting results are as follows:

4.1 The analysis of clustering results

Firstly, the daily load data of interruptible load in December 2017 are clustered. Because of the similar temperature in December, the load is divided into four categories according to whether it rains or not, that is, sunny day holiday, rainy day holiday. Sunny working day and rainy working day, get the result below.

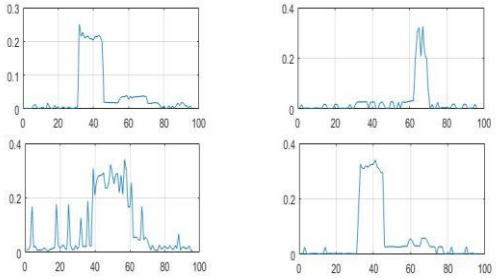

Figure 2. Cluster center curve

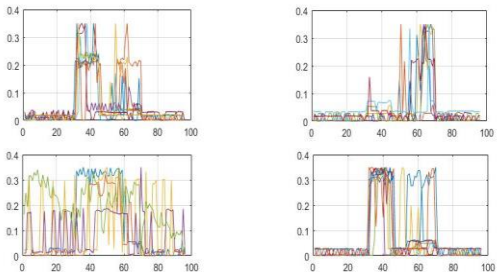

Figure 3. Cluster results

After clustering the load curves with k-means method, all the 31 samples to be clustered are divided into their own categories, and none of them is left out. And basically, the curve in each class has a high similarity with the original cluster center.

4.2 The analysis of prediction results

To forecast the daily load curve of interruptible load on December 31, the basic information of 31 December is searched, the corresponding category is found, and the forecast is carried out with the input data. The forecast results are as follows:

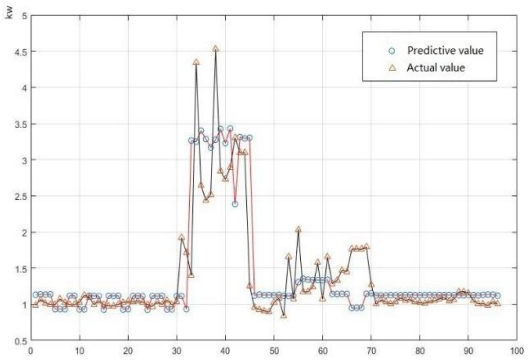

Figure 4. Comparison of forecasted and actual loads as at 31 December

Table 1. Prediction results of 31 December

Times	Actual load (MW)	Forecast load (MW)	Relative error
0:00	1.1422	1.0818	0.053
2:00	1.1189	1.0754	0.039
4:00	0.9267	0.9432	0.018
6:00	1.1131	1.0945	0.016
8;00	3.4834	4.2538	0.249
10:00	3.2275	2.89630	0.103
12:00	1.1237	1.0559	0.060
14:00	1.3532	1.3967	0.032
16:00	1.1429	1.1841	0.036
18:00	1.1237	1.1423	0.016
20:00	1.1263	1.1202	0.005
22:00	1.1346	1.1235	0.009
24:00	1.1189	1.1156	0.003

As can be seen from the results in figure 4 and table 1, the predicted value and the actual value almost coincide in a curve, except that there are differences at a few points. The big difference between the forecast load and the actual load at 8 o'clock may be due to the adjustment of the enterprise's production plan at the early peak. In general, the prediction accuracy of this method is high, up to 95%.

5. Conclusion

There are many factors affecting interruptible load forecasting, and the relationship between them is complex, and the correlation among them is uncertain. Support vector machine (SVM) algorithm can be widely used to establish a nonlinear mathematical relationship between various complex factors

and loads. However, the traditional support vector machine (SVM) method usually chooses a period of date close to the forecast date as a training sample, which can not fully reflect the information characteristics of the forecast day, and the prediction error is large.

Clustering analysis can solve the problem of sample selection well by clustering the similar samples into one group according to the similarity of each sample. Therefore, this paper proposes an interruptible load forecasting method based on clustering algorithm of support vector machine, that is, after clustering, the load similar to the information characteristics of the forecast day to be forecasted is found as a forecasting sample, and then the support vector machine method is used to forecast the load. This method provides a new idea for the selection of forecasting model samples and overcomes the problem that the selection of training date based on traditional support vector machine (SVM) method can not reflect the characteristics of the information of the forecast day. From the forecasting results, it can be seen that the method is accurate and reliable, and it is feasible to be used in interruptible load forecasting.

References

[1] T. Zhang, J.H. Song，X.L. Cheng, "*A Summary of Interruptible Load Study*". NEP **6**，46(2007)

[2] C.Q. Kang, Q. Xia, B.M. Zhang, "*Summary of Power system load forecasting Research and discussion on its Development Direction*". AUTOMAT ELECTRON POWER SYS **28**, 1(2008).

[3] D.F. Zhao, M. WANG, J.S Zhang, "*Short-term load forecasting based on support Vector Machine*"， PRON CHIIN SOC ELECTRICAL ENG **4**, 26(2002).

[4] Y.C. Li，T.J. Fang，E.K. Yu, "Research on support Vector Machine for Short-term load forecasting",（PRON CHIIN SOC ELECTRICAL ENG, 2003）.

[5] J. Wang, S.T. Wang, Z.H. Deng, "*Survey on challenges in clustering analysis research*", CONTROL DECIS,**27**,321(2012).

[6] Q. Wang, C. Wang, Z.Y. Feng, "*Review of K-means clustering algorithm*", EDE **20**， 21（2012）.

[7] L.J. Cheng, C. Jing, Y.X Wu," Research *on normalization method of mathematical expression*". J. Zhejiang Univ-TECHNO. 40, 229(2012).

[8] T.B. Zhu, J. Fu, Y.F. Yang, "*Cluster analysis of load characteristic based on electricity consumption information acquisition system*", EM & I **53**, 70 (2016).

[9] V.Vapnik ， S. Golowich，A. Smola." *Support vector method for function approximation*", (MIT ， 1997).

[10] A.J. Smola, "*Regression estimation with support vector learning ma-chines",* (TS,1996).

[11] S.K. Shevade，S.S. Keerthi，C. Bhattacharyy," *Improvements to SMO algorithm for SVM regression*", IEEE Trans **5**, 1188(2000).

[12] L. Zhang, X. S Liu, H.J. Yin, "*Improvements to SMO algorithm for SVM regression*" . IEEE Trans **5**, 566(2006).

[13] G.H. Liu, X. Yang, "*Application of support vector machine based on time sequence in power system load forecasting*" SA **11**, 19(2017).

[14] L. Li , Y.L. Lu, M.M Zhou , "*A New Fuzzy Support Vector Machine Based on Kernel Method*", POWER SYS TECHNO 28, 38(2004).

[15] X.D. Niu, Z.H. Gu, M. Xing, "*Study on Forecasting Approach to Short-term Load of SVM Based on Data Mining* ", PRO CSEE **26**, 6(2006).