Research Article

Load Prediction Based on Hybrid Model of VMD-mRMR-BPNN-LSSVM

Gang Zhang,1 Hongchi Liu,2 Pingli Li,3 Meng Li,3 Qiang He,3 Hailiang Chao,4 Jiangbin Zhang,2 and Jinwang Hou2

1State Key Laboratory of Eco-hydraulics in Northwest Arid Region, Xi’an University of Technology, Xi’an 710048, China
2The Institute of Water Resources and Hydro-Electric Engineering, Xi’an University of Technology, Xi’an 710048, China
3The Shaanxi Gas Storage Transportation and Comprehensive Utilization Engineering Research Center, Xi’an 710016, China
4The Shaanxi Gas Group New Energy Development Co., Ltd., Xi’an 710016, China

Correspondence should be addressed to Gang Zhang; zhanggang3463003@163.com

Received 30 April 2019; Revised 27 July 2019; Accepted 14 August 2019; Published 20 January 2020

Academic Editor: Michele Scarpiniti

Copyright © 2020 Gang Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Power system load forecasting is an important part of power system scheduling. Since the power system load is easily affected by environmental factors such as weather and time, it has high volatility and multi-frequency. In order to improve the prediction accuracy, this paper proposes a load forecasting method based on variational mode decomposition (VMD) and feature correlation analysis. Firstly, the original load sequence is decomposed using VMD to obtain a series of intrinsic mode function (IMF), it is referred to below as a modal component, and they are divided into high frequency, intermediate frequency, and low frequency signals according to their fluctuation characteristics. Then, the feature information related to the power system load change is collected, and the correlation between each IMF and each feature information is analyzed using the maximum relevance minimum redundancy (mRMR) based on the mutual information to obtain the best feature set of each IMF. Finally, each component is input into the prediction model together with its feature set, in which back propagation neural network (BPNN) is used to predict high-frequency components, least square-support vector machine (LSSVM) is used to predict intermediate and low frequency components, and BPNN is also used to integrate the prediction results to obtain the final load prediction value, and compare the prediction results of method in this paper with that of the prediction models such as autoregressive moving average model (ARMA), LS-SVM, BPNN, empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and VMD. This paper carries out an example analysis based on the data of Xi’an Power Grid Corporation, and the results show that the prediction accuracy of method in this paper is higher.

1. Introduction

Power load forecasting is the basis for power system to arrange power generation plan, economically dispatch the grid, and determine spare capacity, and a high forecasting accuracy plays an important role in the safe and reliable operation of the power grid [1]. With the continuous expansion of the power grid and the rapid development of science and technology, the grid system can obtain a large amount of data, thus providing data support for the research load forecasting system.

In recent years, with the rapid development of computer technology, artificial intelligence algorithms have been introduced into the field of load forecasting. At present, the main power load forecasting techniques are mainly divided into two categories: parametric model based methods and nonparametric model based methods [2]. Among them, the parametric model based methods mainly include time series model [3], regression analysis model [4], trend extrapolation model [5], and other methods. However, the nonlinearity, uncertainty and time-varying nature of electrical load cannot be completely expressed by mathematical formulas, therefore, such models are complicated and difficult to implement. While the nonparametric model method overcomes these weaknesses, and it can adapt to electrical load forecasting with nonlinearity, uncertainty and time-varying nature [6]. The methods based on nonparametric models mainly include wavelet analysis method [7], grey model method [9], support vector machine
(SVM) [10] and artificial neural network method [11–13], etc. These methods could consider the nonlinearity of the electrical load, and the law of the electrical load could be found through fitting and approximating of the original data. Therefore, compared with the traditional method, such a model is superior and the prediction result is more accurate.

Among them, the wavelet analysis methods applying in the electrical load prediction, different wavelets are used to identify and classify loads of different nature and predict them separately, finally, the predicted sequences are integrated as the final load prediction results [14]. In literature [7, 8], the wavelet transform is used to project each sequence component onto different scales, the data processing on different sub-load sequences is performed, and the models matching them are used for prediction, finally, through the wavelet reconstruction, we obtain the load prediction results. However, in the wavelet transform, once the wavelet base is selected, it cannot be replaced during the whole decomposition and reconstruction process. Therefore, it is possible that the wavelet base is only optimal globally. In addition, what criteria are used to select the wavelet base in theoretically and practical applications are still a difficult point [15]. The grey model is a research method suitable for systems with small sample data and poor information. It can obtain useful information from the grey system and build a model based on this useful information to predict unknown data [16]. Literature [9] combined with the new initial conditions and rolling mechanism, designed a new optimized grey prediction model, which could accurately predict industrial electricity consumption and total electricity consumption. However, for the grey model, the greater the degree of data dispersion, the worse the prediction accuracy. And its differential equation exponential solution is more suitable for the load index with exponential growth trend. The SVM model uses a kernel function to replace the nonlinear mapping, with few controlled parameters, the algorithm is simple and easy to implement [17]. It can establish a decision function based on a small number of support vectors, effectively avoiding dimensionality disasters. In literature [10], the improved bird swarm algorithm (IBSA) is used to optimize the parameters of the LS-SVM, and then conduct the prediction using the optimized model. However, SVM is limited to small cluster samples. When observing too many samples, it is less efficient, and relatively difficult to find a suitable kernel function. Artificial neural networks have powerful nonlinear processing capabilities, parallel computing capabilities, self-learning and self-adaptive capabilities, and better memory ability and fault tolerance. When using neural network for electrical load forecasting, there is no need to special processing the data or model, the neural network can achieve better fitting effect by its self-learning ability and self-adaptive ability. Therefore, neural networks have a wide range of applications in the field of power load forecasting. For example, literature [11] proposed a short term electrical load forecasting model based on the hybrid GA-PSO-BPNN algorithm. In literature [12], considering the influencing factors such as temperature, relative humidity, date type, and historical load, a new short-term load prediction method based on kernel function extreme learning machine model is proposed. Literature [13] uses wavelet neural networks to predict electrical load. However, the artificial neural network has the disadvantage of a local minuscule and determining difficulty the number of hidden layer units.

In addition, in order to make full use of the multi-frequency and nonlinearity of power system load, some scholars use the first decomposition and then integration prediction method to predict the power system load. For example, the literature [18] uses empirical mode decomposition (EMD) to decompose the electrical load and then use support vector regression (SVR) to conduct the modeling prediction, in literature [19], the original load data is decomposed using the ensemble empirical mode decomposition (EEMD), and then predicted using the Elman neural network. Although EMD and EEMD can automatically decompose modal components based on data, it is convenient to study, however, the EMD and the improved EEMD that is adding the white noise suffer “endpoint effects” during the decomposition process where divergence phenomena occur at the edge of the data. As the iterations progress, the data sequence may grow severely distorted with modal aliasing and spurious components [20]. Therefore, the variational mode decomposition method that can overcome the end-effect is applied in many fields. For example, the literature [21] uses EMD and VMD to decompose the price data respectively, then uses GRNN to predict each component and uses GRNN to integrate each component. The experimental results show that the prediction accuracy obtained by using the VMD method is higher. Literature [22] uses VMD to decompose the power load data and then predict the individual sub-components using the extreme learning machine (ELM), and then obtain the final power load prediction value by integrating the individual sub-components. Although these two literatures have improved the prediction accuracy compared with the single prediction model, they do not consider the information loss that may occur in the decomposition process and the influence of other factors on the data. As we all know, the load on the power system is also affected by factors such as holidays and temperature. Therefore, some scholars have begun to consider the impact of these factors on the power load. For example, the literature [23] uses the autocorrelation function (ACF) to find the dependency relationship between one time and other times in the time series, then the model is established combining with the LS-SVM to predict the electrical load. However, this method only considers the relationship between the power load time series and does not further consider the influence of other influencing factors on the load. The literature [24] uses the mutual information method to consider the impact of influencing factors on power system load. However, mutual information will generate relatively greater redundancy, increase the model input, and reduce the run speed of the model [25].

Based on the discussions of the above methods, this paper firstly uses the VMD which is completely different from the EMD and EEMD theory to decompose the original loads and to improve the modal aliasing appearing in EMD and EEMD, after obtaining IMFs with different fluctuation characteristics, all IMFs are divided into high-frequency component, intermediate-frequency component and low-frequency component, and then different prediction methods are used to predict
each kinds of frequency components to improve the limitation of the single prediction model. In the prediction of high frequency components, this paper uses BPNN for prediction. This is because high frequency components have strong nonlinearity, and a single hidden layer BPNN can usually approximate any nonlinear function with arbitrary precision. This characteristic makes BPNN very popular in predicting complex nonlinear systems [26]. Although BPNN has the problem of easily falling into the local minimum, and the RBF neural network can improve this problem, the number of hidden layer neurons in the RBF neural network is much higher than that of BPNN when the training samples increase, which makes the complexity of the RBF neural network increased, the structure is also too large, so the amount of calculation has also increased. For the convolutional neural network (CNN), which is the representative algorithm of deep learning algorithm, although sharing the convolution kernel, that is, has no pressure on high-dimensional data processing, and there is no need to manually select features, it requires large amount of samples to adjust parameters because it has more structure layer compared to the BPNN network. Moreover, in order to overcome the shortcomings of BPNN, many scholars have proposed different methods to optimize the initial connection weight value and thresholds of traditional BPNN. For example, the literature [11] uses genetic algorithms and particle swarm optimization (PSO) to optimize BPNN parameters. Both of the convergence speed and prediction accuracy of optimized model are superior to the traditional BP neural network model [26]. For intermediate and low frequency components, LS-SVM is used for prediction. LS-SVM is a kind of extension based on SVM, which is obtained by transferring standard SVM inequality constraints into equality constraints. The basic idea is to map the input vector to the high-dimensional feature space by nonlinear transformation. When constructing the optimal decision function in this space, the principle of structural risk minimization is applied, and the kernel function of the original space is used to replace the dot product operation in the high-dimensional space [27]. The LS-SVM has a simple structure and few parameters under controlled, which is beneficial for predicting intermediate and low frequency components with moderate fluctuations. In addition, in order to improve the prediction accuracy of LS-SVM, some scholars use optimization algorithms to optimize the parameters of LS-SVM. For example, the literature [10] uses IBSA to optimize the parameters to improve the accuracy of load prediction. In the process of predicting each frequency component, in order to analyze the influence of time, holiday and other influencing factors on each component to compensate for the information that may be missing in the process of power load decomposition and eliminate the large redundancy caused by mutual information, and by this, the input quantity is reduced, this paper will use the maximum relevance minimum redundancy (mRMR) based on mutual information to select the set of characteristic factors that have the greatest influence on each component from the feature set of influential factors established in advance, and then use the set together with each component as the input of model. Finally, BP neural network is used to integrate the prediction results of each component to obtain the final power system load prediction value.

2. Materials and Methods

2.1. VMD. In order to resolve the endpoint effect caused by EMD and EEMD and make full use of the multi-frequency and volatility of power system load, this paper uses VMD to decompose the power system load data. VMD was proposed by Dragomiretskiy and Zosso in 2014 [28], the essence is a set of adaptive Wiener filter banks, which could use nonrecursive mode decomposition to estimate the modalities of different center frequencies simultaneously.

The decomposition process of VMD algorithm is the process of solving the variational problem. The algorithm can be divided into the construction and seeking solution of the variational problem. It mainly involves three important concepts: classical Wiener filtering, Hilbert transform, and frequency mixing [29].

2.1.1. The Construction of Variational Problems. Assuming that each modality is a finite bandwidth with a center frequency, the variational problem can be described as seeking to preset k modal functions \( u_k(t) \) under the condition that makes the sum of the estimated bandwidths of each modality smallest, with the constraint condition being the sum of each modality is the original input signal \( f \) and the specific construction steps are as follows:

1. Through the Hilbert transform, the analytical signal of each modal function \( u_k(t) \) is obtained, in order to obtain its unilateral frequency spectrum:
   \[
   \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t).
   \]  
2. Adding an estimated center frequency \( e^{-j\omega_k t} \) to the parsed signals of each mode, so that the spectrum of each modal can be modulated to the corresponding base frequency band:
   \[
   \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) e^{-j\omega_k t}.
   \]  
3. Calculate the square norm L2 of the above-mentioned demodulated signal gradient, and estimate the bandwidth of each modal signal. The variational problem is expressed as follows:
   \[
   \min_{\{u_k, \omega_k\}} \sum_k \parallel \partial_t \left( \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) e^{-j\omega_k t} \right) \parallel^2 \]
   s.t. \( \sum_k u_k = f \),

where \( u_k = \{u_1, u_2 \ldots u_k\} \) is the modal function set; \( \omega_k = \{\omega_1, \omega_2 \ldots \omega_k\} \) is the central frequency set; \( \partial_t \) is the partial derivative of the function for time \( t \); \( \delta(t) \) is the unit pulse function; \( j \) is the imaginary unit; * indicates the convolution.

2.1.2. Seeking Solution of Variational Problem. An augmented Lagrangian function \( \zeta \) is introduced to transform the above optimization problem into a problem of obtaining the
minimum value of $\zeta$ via alternating direction multiplier method

$$\zeta([u_k], \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^{K} \left\| \bar{d}(\frac{\delta(t) + \frac{f}{\pi t}}{\pi}) \ast u_k(t) e^{-jut} \right\|^2_2$$

$$+ \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|^2_2$$

$$+ \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle,$$

where $\alpha$ represents the bandwidth parameter; $\lambda(t)$ represents the Lagrange multiplier.

The specific solution steps are as follows [30].

1. Find the minimum value of the function $u_k$:

$$u_k^{n+1}(\omega) = \frac{\bar{f}(\omega) - \sum_{i=1}^{K} \bar{u}_i(\omega) + \lambda(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2}. \tag{5}$$

2. Calculate the minimum value of the function $\omega_k$:

$$\omega_k^{n+1} = \int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega \int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega \tag{6}$$

3. Apply the following iteration constraint conditions:

$$\sum_{k=1}^{K} \frac{\left| \bar{u}_k^{n+1} - \bar{u}_k^n \right|^2}{\left| \bar{u}_k^n \right|^2} < \epsilon. \tag{7}$$

The implementation process of the VMD method is as follows [30].

1. Parameter initialization $\{\bar{u}_k^1\}, \{\omega_k^1\}, \lambda^1, n \leftarrow 0$;
2. Select the number of iterations $n = n + 1$;
3. Let the scale $k$ starts from 1 up to the preset scale $k$ and update the function $\bar{u}_k$ for all $\omega \geq 0$:

$$\bar{u}_k^{n+1}(\omega) \leftarrow \frac{\bar{f}(\omega) - \sum_{i=1}^{K} \bar{u}_i^{n+1}(\omega) - \sum_{i=1}^{K} \bar{u}_i^n(\omega) + \lambda(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2},$$

then update the function $\tilde{\omega}_k$:

$$\tilde{\omega}_k^{n+1} \leftarrow \int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega \int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega \tag{9}$$

(4) Double promotion for all $\omega \geq 0$:

$$\tilde{\lambda}^{n+1}(\omega) \leftarrow \tilde{\lambda}^n(\omega) + \gamma \left[ \bar{f}(\omega) - \sum_{k=1}^{K} \bar{u}_k^{n+1}(\omega) \right] \tag{10}$$

where $\gamma$ represents the noise tolerance coefficient, which is set to zero to ensure effective denoising.

(5) Repeat Steps (2) to (4) until the constraint of Eq. (7) is satisfied.

2.2 mRMR. The mRMR is a method of using mutual information to measure the dependencies between two variables, considering not only the relevant information between the feature and the target variable, but also obtaining the redundant information between the features.

2.2.1. The Solution of Maximum Relevance. Based on the mRMR method, the maximum relevance criterion can be expressed as the average of the mutual information between the feature $x$ and the target variable $y$ as [31]:

$$\max D(J, y) = \frac{1}{|J|} \sum_{x \in J} I(x, y), \tag{11}$$

where: $x$ is the characteristic, indicating the influencing factors of each component; $y$ is the component; $J$ is the feature set, which is the collection of features $x$; $|J|$ is the number of features in feature set $J$; $D$ is the mean of the mutual information between each feature $x$ in feature set $J$ and the target variable $y$; $I(x, y)$ is the mutual information between the feature $x$ and the target variable $y$, and its expression is as follows [31]:

$$I(x, y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy, \tag{12}$$

where: $p(x)$, $p(y)$, and $p(x, y)$ are the edge probability density function and the joint probability density function of the random variable $x, y$, respectively. The greater the correlation between the variable $x$ and the variable $y$, the larger the value $I(x, y)$ of the mutual information; when the two variables are independent of each other, the mutual information value is zero, meaning that there is no interdependence between the two variables.

2.2.2. The Solution Process of Minimum Redundancy. Since the features selected by the maximum relevance criterion may have some redundancy, while the input of the redundant features not only increases the number of input features, but also reduces the accuracy of the prediction model. Therefore, in the feature selection process, the redundancy degree between features needs to be calculated. The overlapping information between any two feature variables, that is, redundant information.
The minimum redundancy requires that the dependency relationship between each feature be minimized, it can be expressed by the following Equation [31]:

$$\min R(J) = \frac{1}{|J|^2} \sum_{x_i \in J, x_j \in J} I(x_i, x_j).$$  \hspace{1cm} (13)

Then mRMR can be expressed by Equations (11) and (13) as:

$$\max \psi(D, R), \quad \psi = D - R.$$  \hspace{1cm} (14)

In general, the features can be expressed using the incremental search method [29]. Assuming that the n – 1 features that have been selected from the feature set $F_m$ together constitute the feature set $I_{n-1}$, the selection of the n-th feature formula from the set $\{F_m - I_{n-1}\}$ according to the incremental search method, the formula is as follows:

$$\max_{x_i \in F_m - I_{n-1}} \left( I(x_i, y) - \frac{1}{n-1} \sum_{x_j \in I_{n-1}} I(x_j, x_i) \right).$$  \hspace{1cm} (15)

### 2.3. Various Frequency Component Prediction Model

The BPNN has the ability to process different information in parallel, excellent nonlinear fitting ability, self-learning ability, and very effective ability to deal with complex problems. Therefore, BPNN is used to predict high frequency components. The least squares support vector machine (LS-SVM) model has a simple structure, fast learning speed, good generalization, and good regression prediction performance. Therefore, the LS-SVM is used to predict the intermediate and low frequency components. The method of BPNN and LS-SVM are very mature and have been widely used in power system load forecasting [11, 23, 25], and will not be described here.

### 2.4. Prediction Model

This paper uses the prediction model of "decomposition-feature analysis-multi-frequency prediction-integration". The model first uses VMD to decompose the original power system load data to obtain the modal components of different fluctuation characteristics, and divides these components into high frequency components, intermediate frequency components and low frequency components according to their fluctuation characteristics. Then, through mRMR, the factor set that have the greatest influence on each component is selected, and the set and the corresponding modal component are simultaneously used as model inputs. Finally, BPNN is used to predict the high frequency components, LS-SVM is used to predict the intermediate and low frequency components, and BPNN is also used to integrate the predicted values of each frequency component to obtain the final power system load prediction value. The model flow chart is shown in Figure 1.

In order to evaluate the performance of the model, three evaluation indicators: mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) were used. The formula is as follows:

$$\text{MAE} = \frac{1}{S} \sum_{s=1}^{S} |Y_s - F_s|.$$  \hspace{1cm} (16)

$$\text{RMSE} = \sqrt{\frac{1}{S} \sum_{s=1}^{S} (Y_s - F_s)^2},$$  \hspace{1cm} (17)

$$\text{MAPE} = \frac{1}{S} \sum_{s=1}^{S} \left| \frac{Y_s - F_s}{Y_s} \right|.$$  \hspace{1cm} (18)

### 3. Results and Discussion

#### 3.1. Decompose the Original Load Data

In this paper, EMD, EEEMD, and VMD will be used to decompose the data in this paper, then compare and analyze them. The original data is shown in Figure 2, and the exploded view of each model is shown in Figure 3.

In this paper, there are 96 data points per day. It can be seen from Figure 2 that the local power load value began to decline after entering February, and then in mid-February, it increased again and the fluctuation tended to be stable. In addition, by observing the data, it can be found that the power system load data is with a cycle of 96, that is, a similar fluctuation occurs every 96 points, and it can be seen that the daily load change situation are very similar. For a more intuitive view of the data, Table 1 describes the basic information of the load data.

It can be clearly seen from Table 1 that the maximum load value of this group of load value is 5355.7 MW, and the minimum value is 1788.3 MW, and the two differ by 3567.4 MW, it can be seen that the volatility of the load is very strong. In addition, the average value of the load value of the group is 3305.3 MW, and the standard deviation is 794.6 MW, it can be seen that the measures of dispersion of the data is large, which further explains the strong volatility and nonlinearity of the load data.

As can be seen from Figure 3, the EMD decomposes the load data into 10 modal components and 1 remainder. The fluctuations of the first four components (IMF1–IMF4) are relatively strong, and starting from the component IMF1, the fluctuation begins to moderate, and the fluctuations of the IMF8–IMF10 and the remainder are all moderate. The EEEMD decomposes the load data into 12 modal components and 1 remainder. The fluctuations of IMF1–IMF5 are relatively strong, and the fluctuation of components from IMF6–IMF8 begins to ease, and the fluctuations of IMF9–IMF12 and the remainder are all moderate. Since EMD and EEEMD have the same decomposition merit condition, their remainders are monotonic functions.

The VMD decomposes the load data into 10 modal components. Since there are different decomposition termination conditions comparing with EMD and EEEMD,
Complexity

The original load data → VMD →
- High frequency component
- Intermediate frequency component
- Low frequency component
- Influencing factor feature

Analysis of correlation features of mRMR based on mutual information → Various frequency component + Input feature set →
- BPNN (high frequency)
- LS-SVM (intermediate frequency)
- LS-SVM (low frequency)

Integrate using BPNN → Comparison and analysis

**Figure 1:** The model flow chart.

![Power system load data](image)

**Figure 2:** Power system load data.
Figure 3: The exploded view of (a) EMD, (b) EEMD, and (c) VMD.
Figure 4: Continued.
It can be seen from Figure 4 that EMD and EEMD inevitably arise modal aliasing in the decomposition process, that is, one modal component is decomposed into multiple decomposition results. For example, the frequency band of IMF4 in Figure 4(a) overlaps with that of IMF3 and IMF5; the frequency band of IMF3 and IMF4 in Figure 4(b) overlaps. At the meanwhile, the above-mentioned overlap phenomenon does not occur in the Figure 4(c), that is, the VMD decomposition method.

It can be seen that the VMD improves the modal aliasing well during the decomposition process. Based on Figures 3 and 4, we can conclude that EMD, EEMD, and VMD can effectively decompose the power system load with multi-frequency and volatility and obtain the components of different fluctuation characteristics. In addition, although EEMD is theoretically improved on the basis of EMD, the empirical decomposition mode makes it inevitable that illusive components and modal aliasing occur. However, VMD improves this phenomenon well, enabling the power system load data to be more fully decomposed and to obtain more physically meaningful components, to make a good foundation of
conducting load prediction with fully utilizing the volatility and multi-frequency of the power system load data.

3.2. Feature Information Selection. The magnitude of the power system load value is affected by factors such as season, time, and holidays. Therefore, this paper will establish a set of influencing factor features \( F \) and select the input feature set \( I_{IN} \) suitable for each component that is after VMD decomposition. The influencing factor set built and the representation method are shown in Table 2.

In Table 2, \( F_i \) represents time. In this paper, there are 96 data points in one day, that is, four data points in one hour. Each day from 0.00 to 23.45 indicates the change of time; \( F_M \) represents the month, where 1 represents January and 2 represents 2 Month and so on; \( F_W \) stands for the week, where Monday is 1, Tuesday is 2, and so on; \( F_H \) indicates whether it is a holiday (1 means that it is holiday); \( F_S \) indicates season, spring, summer, autumn and winter respectively are represented by 1, 2, 3, 4; \( L \) represents the load value before moment \( t \), for example, \( L_{t-15} \) represents the value before 15 min of moment \( t \), and so on.

After establishing the set of influencing factor features, the incremental search method is used to find the feature satisfying Eq. (15) to form candidate feature set \( J \), and then calculate the mRMR value of each feature in \( J \), and then sort the mRMR values in descending order and input mRMR one by one into prediction mode to calculate the MAPE error, and the calculation formula is as shown in the Eq. (18). Finally, the number of features with the smallest error is taken as the final input feature set \( I_{IN} \).

This article uses IMF6 as an example. The mRMR value of the candidate feature set \( J \) obtained after the incremental search is as shown in Figure 5. The sorting names of each feature in Figure 5, that is, the order of the IMF6 candidate feature sets are as shown in Table 3.

Combined with Table 3 and Figure 5, it can be seen that the characteristics of mRMR values exceeding 1 is \( L_{t-150} \), \( L_{t-120} \), \( L_{t-120} \), \( L_{t-120} \), \( L_{t-105} \), \( L_{t-105} \), \( L_{t-240} \), \( L_{t-225} \) a total of 15. By observing these characteristics, it can be found that except that the third is the month, all the other are load history values, and these the historical load value is concentrated on the load data from 2 hours to 4 hours before moment \( t \), which shows that IMF6 is greatly affected by historical load data 2 hours to 4 hours before.

As described above, the mRMR values of the candidate feature set \( J \) are sorted in descending order and input into the prediction model one by one, and then the MAPE is used to calculate the error, and the set composed of the number of features with the smallest error is taken as the input feature set \( I_{IN} \). The relationship between IMF6 input dimension and error is shown in Figure 6.

It can be seen from Figure 6 that as the number of input features increases, the prediction error of IMF6 is constantly fluctuating, and a small error occurs when the number of inputs is 11, 19, and 24, but in comparison, when the input number of features is 11, the error is the smallest, so the number of features of the input feature \( I_{IN} \) of the IMF 6 is 11. Its input features are shown in Table 4.

As can be seen by observing Table 4, the input feature of IMF6 are included in the 15 components of the previously analyzed mRMR value exceeding one. The features of mRMR values who is in descending order are input one by one into the prediction model to calculate the error and obtain the number of features with minimum error, which greatly reduces the matrix dimension and increases the running speed.

The input feature set \( I_{IN} \) of other components is obtained by using the method that obtains the IMF6 component input feature set. The relationship between the input dimension of each component and the error is shown in Figure 7.

It can be seen from Figure 7 that the relationship between the number of input features and the error can be roughly divided into the following categories: the error firstly decrease and then increase, for example, Figures 7(a) and 7(d), the error first decreases, and after the input features quantity reaches a certain value, it starts to increase again; the error is always increasing, for example, Figures 7(b) and 7(c), the error reaches a minimum at the beginning or after a small fluctuation, and then increases continuously with no trend of decreasing or the trend is not obvious; the error is always decreasing, for example, Figures 7(e) and 7(g), their final input characteristic quantity is determined according to the minimum error; the change trend of error is fluctuating, that is, as the number of features increases, the error curve fluctuates continuously, as shown in Figures 7(f) and 7(h)–7(j), in the process of error fluctuation, select the number of input features with the smallest error. The number of features and their names selected through Figure 7 are shown in Table 5.

For convenience of representation in Table 5, the feature name order is not expressed in descending order of mRMR values like the IMF6. Table 5 shows that except for IMF8–IMF10, other components have historical load
Complexity

(a) IMF1

(b) IMF2

(c) IMF3

(d) IMF4

(e) IMF5

(f) IMF6

Figure 7: Continued.
In IMF7, there are 18, 17, and 11 historical load features respectively, while IMF8–IMF10 are mainly affected by time, month, week, and season.

In addition, the correlation coefficient method and mutual information (MI) are used to select the influence feature of each component. The selection results are shown in Table 6. Comparing with Table 5 and it can be seen from Table 6, for the correlation coefficient method, the number of features of the IMF1, IMF4, and IMF5 is close to that of the mRMR, and the number of features of other components is greater than that of the mRMR. Similarly, the number of features selected by MI, except for that of IMF4 and IMF5, is greater than mRMR. Therefore, these two methods increase the matrix dimension and reduce the speed of the operation. In addition, comparing with Table 5, it can be found that except for IMF2, IMF4, and IMF5, other components will be affected by features other than historical load values, but after correlation coefficient selection and MI selection, the effects of other features are not obvious. This is due to the correlation coefficient method and mutual information only considering the correlation between variables but neglecting the characteristics and account for a large proportion of input characteristics. It can be seen that historical load values have a greater impact on IMF1–IMF7, especially IMF1, IMF5, and IMF7, there are 18, 17, and 11 historical load features respectively, while IMF8–IMF10 are mainly affected by time, month, week, and season.

Comparing with Table 5 and it can be seen from Table 6, for the correlation coefficient method, the number of features of the IMF1, IMF4, and IMF5 is close to that of the mRMR, and the number of features of other components is greater than that of the mRMR. Similarly, the number of features selected by MI, except for that of IMF4 and IMF5, is greater than mRMR. Therefore, these two methods increase the matrix dimension and reduce the speed of the operation. In addition, comparing with Table 5, it can be found that except for IMF2, IMF4, and IMF5, other components will be affected by features other than historical load values, but after correlation coefficient selection and MI selection, the effects of other features are not obvious. This is due to the correlation coefficient method and mutual information only considering the correlation between variables but neglecting the characteristics and account for a large proportion of input characteristics. It can be seen that historical load values have a greater impact on IMF1–IMF7, especially IMF1, IMF5, and IMF7, there are 18, 17, and 11 historical load features respectively, while IMF8–IMF10 are mainly affected by time, month, week, and season.

Comparing with Table 5 and it can be seen from Table 6, for the correlation coefficient method, the number of features of the IMF1, IMF4, and IMF5 is close to that of the mRMR, and the number of features of other components is greater than that of the mRMR. Similarly, the number of features selected by MI, except for that of IMF4 and IMF5, is greater than mRMR. Therefore, these two methods increase the matrix dimension and reduce the speed of the operation. In addition, comparing with Table 5, it can be found that except for IMF2, IMF4, and IMF5, other components will be affected by features other than historical load values, but after correlation coefficient selection and MI selection, the effects of other features are not obvious. This is due to the correlation coefficient method and mutual information only considering the correlation between variables but neglecting the characteristics and account for a large proportion of input characteristics. It can be seen that historical load values have a greater impact on IMF1–IMF7, especially IMF1, IMF5, and IMF7, there are 18, 17, and 11 historical load features respectively, while IMF8–IMF10 are mainly affected by time, month, week, and season.

Comparing with Table 5 and it can be seen from Table 6, for the correlation coefficient method, the number of features of the IMF1, IMF4, and IMF5 is close to that of the mRMR, and the number of features of other components is greater than that of the mRMR. Similarly, the number of features selected by MI, except for that of IMF4 and IMF5, is greater than mRMR. Therefore, these two methods increase the matrix dimension and reduce the speed of the operation. In addition, comparing with Table 5, it can be found that except for IMF2, IMF4, and IMF5, other components will be affected by features other than historical load values, but after correlation coefficient selection and MI selection, the effects of other features are not obvious. This is due to the correlation coefficient method and mutual information only considering the correlation between variables but neglecting the characteristics and account for a large proportion of input characteristics. It can be seen that historical load values have a greater impact on IMF1–IMF7, especially IMF1, IMF5, and IMF7, there are 18, 17, and 11 historical load features respectively, while IMF8–IMF10 are mainly affected by time, month, week, and season.

Comparing with Table 5 and it can be seen from Table 6, for the correlation coefficient method, the number of features of the IMF1, IMF4, and IMF5 is close to that of the mRMR, and the number of features of other components is greater than that of the mRMR. Similarly, the number of features selected by MI, except for that of IMF4 and IMF5, is greater than mRMR. Therefore, these two methods increase the matrix dimension and reduce the speed of the operation. In addition, comparing with Table 5, it can be found that except for IMF2, IMF4, and IMF5, other components will be affected by features other than historical load values, but after correlation coefficient selection and MI selection, the effects of other features are not obvious. This is due to the correlation coefficient method and mutual information only considering the correlation between variables but neglecting the characteristics and account for a large proportion of input characteristics. It can be seen that historical load values have a greater impact on IMF1–IMF7, especially IMF1, IMF5, and IMF7, there are 18, 17, and 11 historical load features respectively, while IMF8–IMF10 are mainly affected by time, month, week, and season.
redundancy between the two variables, while mRMR not only considers the correlation between the two variables but also considers their redundancy, so the selected influencing features are more diversifying than the other two methods. In summary, using mRMR for feature selection reduces the matrix dimension and improves the speed of operation and better reflects the influence of various feature factors on each component.

3.3. Various Frequency Component Prediction. According to the fluctuation characteristics of each frequency component in Figure 3(c), the first five components with small fluctuation periods (IMF1–IMF5) are classified as high frequency components, the IMF10 with longest period is classified as low frequency data, and the remaining components (IMF6–IMF9) are as intermediate frequency data. Regarding the division of frequency data, and the remaining components (IMF6–IMF9) are classified as high frequency components. The first five components with small fluctuation characteristics of each frequency component is better than that of the LS/uni2010SVM, while for the high frequency component, the prediction performance of BPNN increases again. In summary, in this paper, among a total of 11,520 load data points from January 2, 2016 to April 30, 2016, 10,752 load data points from January 2, 2016 to April 22, 2016 will be used as intra-sample data, and then use the selected model to conduct the prediction on a total of 768 load data points from April 23, 2016 to April 30, 2016.

Besides, before adding the influencing factors, BPNN and LS-SVM are used to directly predict the various frequency components decomposed by VMD in 4.1, and use MAPE to count the prediction error. The results are shown in Table 8.

It can be seen from Table 8 that for high frequency components IMF1–IMF5, the prediction accuracy of BPNN is significantly higher than that of LS-SVM. For the intermediate frequency and low frequency components, at IMF6, the prediction accuracy of them are very close, at other IMF, for the prediction accuracy of intermediate and low frequency components, LS-SVM is significantly higher than BPNN, especially after the IMF8, it is becoming more and more obvious. In summary, in comparison, for the high frequency component, the prediction performance of BPNN is better than that of the LS-SVM, while for the high frequency and intermediate frequency components, the prediction performance of the LS-SVM is better than that of the BPNN.
Figure 8: High frequency component prediction figure. Prediction results of (a) IMF1, (b) IMF2, (c) IMF3, (d) IMF4, and (e) IMF5.
3.3.1. The Prediction of High Frequency Component. According to the above, the high frequency component is predicted by BPNN, and the predicted result is shown in Figure 8. In the modeling process, the number of iterations is set to 1000, the learning speed rate is 0.1, and the expected error is 0.0004. The structure of the neural network is 3-input and 1-output. The three inputs in this paper refer to the three original points of a certain component and the features corresponding to the three original points are used as inputs. In addition, in this paper, we use the method of literature [32], which combines the "trial and error method" with the "ten-fold cross-validation method" to determine the number of hidden layer nodes of the BP neural network. That is, in the vicinity of the number of hidden layer nodes calculated by formula (19), the BP neural network is constructed respectively, and then the mean square error (MSE) of each group of data is obtained using the "ten-fold cross-validation method", finally take the number of nodes, that has the smallest average value of the MSE, as the number of hidden layer nodes of the BP neural network. Where Equation (19) is as follows [32]:

\[ n = \sqrt{0.43m + 0.12l^2 + 2.54m + 0.77l + 0.35 + 0.51}, \quad (19) \]

where \( n \) is the number of hidden layer nodes, \( m \) is the number of input layer nodes, \( l \) is the number of output layer nodes. According to the above method, the number of hidden layer nodes of each high frequency component is as shown in Table 9.

After completing the above discussion, the feature set of various frequency components, selected in 4.2, is taken into consideration in the prediction process.

3.3.2. Prediction of Other Components. In this paper, LS-SVM is used to predict the intermediate frequency and low frequency component. The radial basis function (RBF) is used in the modeling process and the regularization parameters and kernel parameters are optimized using the particle swarm optimization algorithm (PSO), and the parameters of LS-SVM model of each intermediate and low frequency component are shown in Table 10.

The intermediate and low frequency component are predicted according to the LS-SVM model of Table 10. The prediction results are shown in Figure 9.

As can be seen from Figures 9 and 8, the fluctuation of intermediate frequency component (IMF6-IMF9) and the low frequency component (IMF10) are much more moderate than the high frequency component fluctuation. Therefore, as its fluctuation gradually eases, the prediction results are getting better and better. It can be seen that using the LS-SVM with fast learning speed and good generalization to predict the intermediate and low frequency components that the fluctuation is moderate have smaller error.

3.4. Combination of Each Frequency Component Result. In this paper, BPNN is used to integrate high, intermediate, low frequency components, and compare the integrated results with that of using LS-SVM, in the process of integration, with the real value of each component in training data set as input, the real load value of training data set as the output to train it, and finally use the predicted value of each component in test data set as input to obtain the final power system load prediction value. The combination result is shown in Figure 10.

The upper one of Figure 10 is the power load prediction result, and the lower one is the error between the actual value and the predicted value, where the green point is the distribution of the predicted value. As can be seen by observing Figure 10, most of the predicted points can be distributed around the actual value except that few points deviate from the actual values. However, in Figure 10, predicted value distribution point in (a) is significantly more concentrated than (b), especially the head and the end, which indicates that the integration effect of BPNN is better than LS-SVM.

### Table 9: The neural network structure of each frequency component.

| Component | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 |
|-----------|------|------|------|------|------|
| Structure | 60-15-1 | 9-6-1 | 21-9-1 | 30-10-1 | 54-14-1 |

### Table 10: The parameters of LS-SVM model of each intermediate and low frequency component.

| Parameters | IMF6 | IMF7 | IMF8 | IMF9 | IMF10 |
|------------|------|------|------|------|-------|
| C          | 98.5037 | 46.6631 | 74.2140 | 32.2762 | 94.1287 |
| \( \sigma^2 \) | 0.2142 | 0.1292 | 0.1301 | 0.1134 | 0.1252 |

### Table 11: Various model prediction results.

| Model | MAE | RMSE | MAPE (%) |
|-------|-----|------|----------|
| VMD + mRMR + BPNN + LS-SVM + BPNN | 8.1794 | 9.0265 | 0.33 |
| VMD + mRMR + BPNN + LS-SVM+LS-SVM | 9.0726 | 11.2312 | 0.41 |
| VMD + correlation coefficient + BPNN + LS-SVM+BPNN | 9.1572 | 11.3363 | 0.44 |
| VMD + MI + BPNN + LS-SVM + BPNN | 8.9248 | 10.1486 | 0.39 |
| ARMA | 28.9115 | 37.0648 | 1.03 |
| BPNN | 28.8477 | 36.4898 | 1.03 |
| LS-SVM | 30.4543 | 38.5108 | 1.09 |
| EEMD + BPNN + LS-SVM + BPNN | 19.1876 | 24.6827 | 0.73 |
| EMD + BPNN + LS-SVM + BPNN | 22.6169 | 26.6001 | 0.91 |
| VMD + BPNN + LS-SVM + BPNN | 9.4434 | 11.6906 | 0.46 |
Figure 9: Intermediate frequency and low frequency component prediction figure. Prediction results of (a) IMF6, (b) IMF7, (c) IMF8, (d) IMF9, and (e) IMF10.
F/i.sc/g.sc/u.sc/r.sc 10:/uni00A0/T_he /figure of load prediction. Integrated results of (a) BPNN, and (b) LS-SVM.

F/i.sc/g.sc/u.sc 11:/uni00A0/Continued.
Complexity

3.5. Comparison of Multiple Prediction Model Results. In order to test the prediction effect of the proposed model of this paper and further analysis of the prediction results of Figure 10, we compared the proposed model against the single model (ARMA, BPNN, LS-SVM) and decomposition models (EMD, EEMD, VMD) to validate its effectiveness. Among them, the EMD, EEMD and VMD decomposition models all use the decomposition method to decompose the original load data, then divide the sub-component into high, intermediate and low frequency components, and use BPNN to predict high-frequency components and use LS-SVM predicts intermediate and low-frequency components, finally, BPNN is used for integration. The results are shown in Figure 11.

It can be clearly seen from Figure 11 that the prediction points of the single prediction model deviate farther comparing with multi-frequency combination prediction models (EMD, EEMD, and VMD), while in the multi-frequency combined prediction model, it can be seen that the improved EEMD based on EMD is better than EMD, and the prediction results of VMD is the best, the principle of VMD is completely different from the other two. This shows that the prediction...
accuracy of the multi-frequency combined prediction model is greater than that of the single prediction model, while the VMD multi-frequency combination prediction model that resolves the problem of modal aliasing and illusive components has the highest prediction accuracy.

The results of the above-mentioned evaluation indexes which we used to ensure a comprehensive comparison are shown in Table 11.

Through the comparison of model results of Figure 10 and Figure 11 (the comparison as shown in Table 11) can clearly see that the error of the single prediction model is larger than that of the multi-frequency combination prediction model. For the prediction accuracy of different multi-frequency combination prediction models, the VMD model is better than the EEMD model and the EMD model. In addition, after mRMR is combined with VMD, the prediction accuracy of mRMR-VMD is higher than the VMD model, and for models considering influencing factors, the effect of using BPNN integration is better than that of using LS-SVM. For the VMD multi-frequency combined prediction model using different feature selection methods, the prediction accuracy of the VMD multi-frequency combined prediction model using mRMR is better than the VMD multi-frequency combined prediction model using the correlation coefficient method and MI. In summary, the multi-frequency combined prediction model based on mRMR feature selection and VMD proposed in this paper is superior to other models.

4. Conclusion

This paper proposes a power system load forecasting model based on VMD and feature selection. By comparing the decomposition results of three decomposition methods of EMD, EEMD, and VMD, it can be seen that VMD can resolve the problem of modal aliasing and illusive components well. Through the example analysis, it can be seen that the prediction accuracy is improved after resolving the problem of modal aliasing and illusive components.

In this paper, after using VMD to decompose the original loads, all IMFs are divided into high frequency component, intermediate frequency component and low frequency component according to the fluctuation characteristics of each frequency component. In addition, when using mRMR to pick features related to each component, the candidate feature set is first established by the incremental search method, then the mRMR values of each feature are calculated and arranged in descending order. Finally, the arranged features are input into the prediction model one by one to calculate MAPE and choose the minimum MAPE as the number of input features of the component. In addition, this paper also compares the three feature selection methods containing mRMR, correlation coefficient method and MI, the results show that mRMR can reduce the matrix dimension and better select the influencing feature of each component.

After completing the above work, each component is input into the model together with the corresponding feature matrix for prediction. In the prediction process, BPNN is used to predict the high frequency components and LS-SVM predicts the intermediate frequency and low frequency components to improve the limitation of single prediction model. Finally, the BPNN with strong nonlinear mapping ability, high self-learning, and self-adaptive ability is used to combine the predicted values of high, intermediate, low frequency component, that is, with each component sample data as input and with the actual load value as output to train the model, and then the prediction values of each components are used as inputs to predict the final power system load prediction value. Finally, we use the MAE, RMSE, and MAPE to compare the prediction results of prediction model in this paper with that of single prediction model (BPNN, ARMA, LS-SVM) and multi-frequency combined prediction model of (EMD, VMD, EEMD) and the models that VMD multi-frequency combined prediction model respectively combined with correlation coefficient and MI. By comparison, it is found that the prediction accuracy of multi-frequency combined prediction model is higher than that of single prediction model, and the prediction accuracy of model in this paper is higher than that of other multi-frequency combination prediction model.

In the prediction process, it is found that when the mRMR values in descending sort are input into the prediction model to calculate the MAPE value, there is a case where the MAPE fluctuates as the number of input features increases. For example, in Figure 7 (i), it can be seen that when the number of features is 2 and 3, the MAPE values are very close, while in a sense, there is not much difference in the input dimension when the number of input features is 2 and 3. Then how to deal with this situation better in the feature selection process and whether there is a more reasonable feature selection method to analyze the relationship between features and power load will become the future research object of this paper.

Data Availability

The power load data used to support the findings of this study have not been made available because policy limitation of China’s Power Grid.

Conflicts of Interest

The author declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant No. 2016yfc0401409), and the National Natural Science Foundation of China (Grant Nos. 51679186, 51679188, 51979221, 51709222), and the Research Fund of the State Key Laboratory of Eco-Hydraulics in Northwest Arid Region, Xi’an University of Technology, (Grant No. 2019JJCXTD-5), and the Key Research and Development Plan of Shaanxi Province, (Grant No. 2018-ZDCXL-GY-10-04), and the Natural Science Basic Research Program of Shaanxi (Program No. 2019JLZ-15).
References

[1] T. Zhang and J. Gu, “High proportion renewable energy and markov short-term load forecasting method,” Power Grid Technology, vol. 42, no. 4, pp. 1071–1078, 2018.

[2] J. S. Lin and J. Yao, “Combination forecasting based on BP neural networks and its application in load of power system,” Journal University of Shanghai for Science and Technology, vol. 27, no. 5, pp. 79–83, 2005.

[3] Z. C. Chen, “Power system load forecasting using time series,” Automation of Electric Power Systems, vol. 4, pp. 10–18, 1982.

[4] T. Zhang and J. L. Zhu, “The application of wavelet transform and regression analysis used in short-term load power forecasting,” Journal of Harbin University of Science and Technology, vol. 13, no. 1, pp. 74–76, 2008.

[5] S. Li, L. Liu, and Y. Liu, “Application of trends extrapolation on the power load forecast,” Journal of Shenyang Institute of Engineering (Natural Science), vol. 1, no. 2, pp. 64–65, 2005.

[6] L. Dai, “Research and implementation of electric load forecasting method based on neural network,” University of Electronic Science and Technology of China, 2012.

[7] N. L. Tai, Z. J. Hou, T. Li, and C. W. Jiang, “New principle based on wavelet transform for power system short-term load forecasting,” Proceedings of the CSEE, vol. 23, no. 1, pp. 46–51, 2003.

[8] P. Fryzlewicz, S. Van Bellegem, and R. von Sachs, “Forecasting nonstationary time series by wavelet process modelling,” Annals of the Institute of Statistical Mathematics, vol. 55, no. 4, pp. 737–764, 2003.

[9] S. Ding, K. W. Hipel, and Y.-G. Dang, “Forecasting China’s electricity consumption using a new grey prediction model,” Energy, vol. 149, pp. 314–328, 2018.

[10] L.-L. Li, Z.-F. Liu, M.-L. Tseng, and A. S. F. Chiu, “Enhancing the lithium-ion battery life predictability using a hybrid method,” Applied Soft Computing, vol. 74, pp. 110–121, 2019.

[11] Y. Hu, J. Li, M. Hong et al., “Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm—a case study of papermaking process,” Energy, vol. 170, pp. 1215–1227, 2019.

[12] N. Zhang and T. J. Liu, “Kernel function ELM method for short-term load forecasting considering influencing factors,” Engineering Journal of Wuhan University, vol. 51, no. 8, pp. 703–714, 2018.

[13] D. H. Zhang, S. F. Jinag, Y. Q. Bi, and G. B. Zhou, “Study of power system load forecast based on wavelet neural networks,” Electrical Power Automation Equipment, vol. 23, no. 8, pp. 29–32, 2003.

[14] N. Zhang, “Short-term load forecasting model based on EEMD and extreme learning machine,” Journal of Guizhou University (Natural Sciences), vol. 34, no. 6, pp. 39–48, 2017.

[15] Y. Q. Zheng, Research on short-term power load forecasting based on empirical mode decomposition and artificial neural network, Chongqing University, China, 2009.

[16] S. F. Liu, T. B. Guo, and Y. G. Dang, Grey System Theory and Its Application, Science Press, Beijing, 2000.

[17] R. G. Brereton and G. R. Lloyd, “Support vector machines for classification and regression,” The Analyst, vol. 135, no. 2, pp. 230–276, 2010.
