A Hybrid Method Based on Wavelet Analysis for Short-term Load Forecasting

Chenhao Niu, Ling Ji

1. First Author
School of Economics and Management, Beijing Jiaotong University, bjtnch@163.com

2. Corresponding Author
School of Economics and Management, North China Electric Power University, hdjiling@126.com

Abstract

Power load is a typical time series with the characters of nonlinearity and volatility. Accuracy load forecasting can help ensure the security of power grid operation. In this paper, a new hybrid method based on wavelet analysis for short-time load forecasting application is proposed. The key idea is that the original time series can be decomposed into various components with different frequencies through wavelet analysis. Then the low frequency part and high frequency components are predicted by RBF and Markov respectively. The final forecasting result is obtained by wavelet reconstruction. Though the experimental study, the hybrid method was found to perform better with higher accuracy compared to RBF and Markov models. The experimental results show the potential of this combined models based on wavelet analysis in the application of load forecasting.

Keywords: Time Series, Load Forecasting, Wavelet Analysis, RBF, Markov

1. Introduction

Nonlinearity and volatility is the main characteristics of time series, and time series prediction is to infer the future value by constructing certain model and rules. Power load belongs to the typical time series, the change of load has obvious randomness and complexity [1]. Two different trends reflect the special rule of power load. One is the increasing trend, and the other is the cycle fluctuation with day, week, month and year. Additional, climate factors like temperature, rainfall, wind and other factors would cause the random fluctuation. Due to these characteristics, there are lots of difficulties and challenges for efficient short-time load forecasting. In the last decades, numerous studies have been investigated on the nonlinear forecasting model such as neural network [2], support vector model [3-5], PSO [6] and RBF [7]. However, in practice, any method is limited with its own advantages and disadvantages and can’t be the perfect one. Like neural network technology, an excellent tool to describe the nonlinear relationship, is easy to fall into local minimum, besides its learning speed is slow.

Since power load series is consist of stable sequence and random sequence, a single prediction model for two parts will increase mutual interference error. Hence, various theories and methods are studied to dig out the relationship between load and all kinds of influence factors. An improved singular spectral analysis method is adopted in paper [8] for short term load forecasting, and the simulation results show that the proposed method has a good ability in characterizing and prediction of the desired load time series. Wavelet analysis has good localization characteristics in both time domain and frequency domain, and can extract information from signal. Therefore, it shows the great superiority on the treatment of the non-stationary time series. In recent years, hybrid approaches with wavelet analysis have been applied for forecasting successfully [6-9]. Paper [10] use a two-dimensional wavelet based state dependent parameter (SDP) modelling approach for daily peak power demand prediction. In paper [11], the wavelet analysis (WA) combined with the fuzzy support vector kernel regression method was proposed for load forecasting. Thus, wavelet analysis will be adopted for load series decomposition, then according to the different features of the high frequency and the low frequency, different algorithm will be employed to forecast respectively. The specific process is shown in figure 1 below.
Figure 1. The key idea of this paper

The remainder of this paper is organized as follows. In Section 2 we introduce the necessary theory background. The hybrid model is described in Section 3. In Section 4, we conduct the application of the new hybrid model is demonstrated, followed by the comparison experiments and analysis. Finally, our concluding remarks are presented in Section 5.

2. Methodology

2.1. Wavelet analysis

Preliminary studies have indicated that wavelet analysis appears to be a more effective tool than the Fourier Transform in analyzing non-stationary time series. Though wavelet analysis, the complex time series can be decomposed into simple signals with different frequencies. The signal with single frequency is more stable. According to the multi-resolution analysis in the basic principles of wavelet analysis, the input signal is decomposed into a rough approximation with low frequency and the details of the high frequency. The former shows the long time trend of the signal, while the latter reflect the short-term volatility. For the decomposition of the original time series signal $S$, after the first decomposition, $S$ is broken into a high-frequency details $W_1$ and a signal approximate $S_1$, and $S = S_1 + W_1$. Thus, after $J$ times decompositions like this, signal $S$ final is broken into $S = W_1 + W_2 + ... + W_J + S_J$, among them, $S_J$ for the rough approximation reflecting the whole trend of signal $S$, $W_1, W_2, ..., W_J$ show the fluctuation details with different frequencies.

Consider a function $\psi(t)$ as a quadratically integrable function, $\psi(t) \in L^2(\mathbb{R})$. If its Fourier transform satisfies the admissible condition $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\hat{\psi}(w)|^2 dw < \infty$, $\psi(t)$ can be called as mother wave. A series of sub-wavelets can be developed though dilating and translating the $\psi(t)$. It can be described as:

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right), \quad a \neq 0, b \in \mathbb{R}$$  \hspace{1cm} (1)

Where, $a$ is the wavelet scale, $b$ is the translation parameter, and $t$ is the time.

Wavelet transform is to approximate the signal by wavelet function, and it can be divided into continuous wavelet transform and discrete wavelet transform. For signal $f(t) \in L^2(\mathbb{R})$, the continuous wavelet transform is described as below:

$$W_\psi f(a,b) = |a|^{1/2} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$  \hspace{1cm} (2)

Where, $a, b$ and $t$ are all continuous variables.

Since in load forecasting research, discrete time series are most studies, here, we introduce the discrete wavelet transform and multi-resolution analysis. In discrete wavelet analysis, both $a$ and $b$ are restricted to discrete values such that $a = a_m^m$, $m \in \mathbb{Z}$; $b = nb_m^m$, $n \in \mathbb{Z}$, and obtain the corresponding discrete set $\psi_{a_m^m, b_m^m, n}$. Though multi-resolution analysis, function $f$ can be described as a series of approximate function limits. Each approximate is the smooth version of function $f$ and it is more and more subtle. These are obtained on different scales. Suppose $\psi_{a_m^m, b_m^m, n}$ is the orthogonal wavelet base of $L^2(\mathbb{R})$, for arbitrary $f(x)$ can be spread as:
\[ f(x) = \sum_{m,n} <f, \varphi_{m,n}> \varphi_{m,n}(x) \]  

Mallat is a fast algorithm for wavelet transform proposed on the multi-resolution analysis. Mallat algorithm for discrete wavelet decomposition can be expressed as:

\[
\begin{align*}
C_{k,n} &= \sum_l a_{n-2l} C_{k,l} \\
d_{k,n} &= \sum_l b_{n-2l} C_{k,l}
\end{align*}
\]

The decomposed components can be assembled back to the original signal without losing information. This process is called reconstruction. And the Mallat algorithm for wavelet reconstruction is:

\[ C_{k-1,n} = \sum_i \left( p_{n-i} C_{i,2l} + q_{n-i} d_{i,2l} \right) \]

Where, \( k \) means the decomposition level.

How to select appropriate decomposition level \( k \) is important in wavelet analysis. With the increasing of \( k \), the frequency band can be divided more carefully with more stable, which is helpful to fit. However, the whole forecasting error is greater, and the accuracy decease. In practice, the number of decomposition level is determined according to the principle of minimizing the overall prediction error.

2.2. RBF model for low frequency component

The stable sequence is nonlinear, and in the field of nonlinear modeling, neural network like BP (Back Propagation) and RBFN (Radial Basis Function Network) have been applied successful. The structure of RBFN includes input layer, hidden layer and output layer. The transform from input layer to hidden layer is nonlinear and that from hidden layer to output layer is linear, which is quite different from BP network. The activation function of hidden layer is radial basis function [16-18]. The framework of standard RBFN is depicted in Figure 2.

[Figure 2. The typical structure of RBFN]

The input of the network is the stable load sequence, and the output is the load need to be predicted. Caussian function is usually selected as activation function of the hidden layer for the network. And there is no need to define the number of units in hidden layer. The structure of the hidden layer is adjusted automatically by the training data. The model can be explained by:

\[ y = \sum_{i=1}^{N} w_i g \left( \| x - c_i \| / \sigma \right) + b \]
Where, $x \in \mathbb{R}^n$ is the input of the neural network, $w_i$ denote the weights of output layer, $g(*)$ is the radial basis function, $c_i$ means the center of the radial basis function, $\sigma_i$ is the width of the network, $b$ is the threshold of the output layer, $n_e$ is the number of the units in hidden layer, and $\|x_i\|$ represents the distance between the input $x$ and the center $c_i$.

At currently, the optimization algorithm based on gradient is the commonest used for training RBF network, namely, for training samples $D_m = \{x_i, y_i\}, i=1,2,...,M$, we search for $\Theta = (c_i, \sigma_i, n_e, w_i)$ to minimize the error, which can be expressed as:

$$\min F_{\text{RBF}} = \frac{1}{M} \sum_{i=1}^{M} |Y_i - f_n(x_i, \Theta)|^2$$  \hspace{1cm} (7)

$$f_n(x_i, \Theta) = \sum_{i=1}^{n_e} w_i g \left( \|x_i - c_i\| / \sigma_i \right)$$  \hspace{1cm} (8)

Where, $Y_i$ is the object value, and Gaussian function for the radial basis function, the value of the threshold in output layer $b$ is 0, and the output of unit $i$ is:

$$R_i(x) = \exp \left( -\|x - C_i\|^2 / 2\sigma_i^2 \right) i=1,2,...,m$$  \hspace{1cm} (9)

Where, $X$ is the input vector with $n = n_u + n_v$ dimensions, $C_i$ is the center vector. And the mapping relationship of nonlinear function for RBF network with one output is as following:

$$y = f(X) = WR = \sum_{i=1}^{n_e} w_i R_i(x)$$  \hspace{1cm} (10)

Where, $W$ is the weight vector from hidden layer to output layer, and $R$ is the output vector. The center vector $C_i$, parameter $\sigma_i$ and the network weight $W$ need to be determined in the application of RBF.

In order to facilitate calculation, the original data must be normalized, and the normalized formula is as follows:

$$x'_i = 2x_i \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} - 1$$  \hspace{1cm} (11)

Though the treatment, the sample data is mapped into interval [-1, +1]. The output value must be rollback to primary space for forecasting.

2.3. High frequency components with Markov

In this work, Markov is employed to predict the coefficient series $b^j_l$ with high frequency. According to the fluctuation range of the original series, the coefficient series are divided into $m$ state spaces, $[B_1, T_1), [B_2, T_2), \ldots, [B_m, T_m]$, where, $B_l$ and $T_l$ are the lower limit and upper limit of state space $l (l=1,2,\ldots,m)$, respectively. Use frequency approximate method to establish the state transition matrix of the coefficient sequence $Q$, $Q = (q_{lh})_{m \times m}$, and $q_{lh}$ is the transition probability from state $l$ to state $h$. Suppose from the $N$ times observing of the original series, the frequency of sample falling into
state $l$ is $n_l$, $N = \sum_{l=1}^{N} n_l$. When it change from current state $l$ to the next state $h$ with the frequency of $n_{lh}$, hence, $q_{lh} = n_{lh}/n_l$. Therefore, if the last state is $l$, and the forecasting value at this state is [21-24]:

$$b^l = \sum_{h=1}^{W} q_{lh} \cdot \frac{B_h + T_h}{2}$$ (12)

According to Eq.(12), we can calculate the forecasting value $b^l$ for each high frequency coefficient sequence.

3. Hybrid forecasting model

Depending on the theory analysis above, the steps of the proposed hybrid forecasting model are described as follows:

Step1: Use wavelet method to decompose original time series into a number of different sub-series (depending on level of decomposition).

Step2: After the treatment, the low frequency is forecasted by RBFN, while the high frequency part is forecasted by Markov.

Step3: Though reconstructing the forecasting value from sub-series, the final forecasting value can obtained.

Step4: Error measure indexes are used to evaluate the forecasting precision and effect.

To verify the predictive ability of the proposed scheme, we perform numerical simulations. The data used in the simulation are history load of a city in North of China ranging from July 1, 2010 to July 19, 2010. And the 96 point load on July 20 is the forecasting sample. According to the characters of load series, it is suitable for this paper to adopt ‘bior4.4’ as the wavelet function and set the level $N$ at 3. Fig. 3 displays the original load series and decomposed results.

![Wavelet decomposition result of electric load](image)

Figure 3. Wavelet decomposition result of electric load

The performance of the proposed model was compared to conventional RBF models and Markov for hourly load forecasting. In Markov forecasting, the load data are divided into 16 states and the corresponding state transition probability matrix is established.

In order to compare the proposed new algorithm with others, three forecasting error measures are employed for model evaluation and model comparison, the mean absolute percentage error (MAPE) and Predict accuracy (PPD).
A Hybrid Method Based on Wavelet Analysis for Short-term Load Forecasting
Chenhao Niu, Ling Ji

\[ MAPE = \frac{1}{N} \sum_{y} \left| \frac{y - \hat{y}}{y} \right| \]  

(13)

\[ PPD = \left( 1 - \frac{1}{N} \sum_{y} \left| \frac{y - \hat{y}}{y} \right| \right) \times 100\% \]  

(14)

4. Results and comparative analysis

Figure 4-6 show the comparisons between the real records of hourly load and the predicted ones by different models.

**Figure 4.** Forecasting results with the proposed hybrid model

**Figure 5.** Forecasting results with RBF

**Figure 6.** Forecasting results with Markov
From the fitting curves illustrated in Figure 4-6, it is obvious that the forecasting result from new hybrid model with better performance than the other two single. With the proposed hybrid model, the mean absolute error and standard deviation is 1.42% and 0.737%. The PPD is equal to 98.41%. Although the forecasting result of other two algorithms are acceptable, the mean absolute percentage error of RBF and Markov is 2.08% and 2.51% respectively. Besides, the forecasting value is much more fluctuate with the standard deviation equal to 0.946% and 0.971%. Therefore, the PPD with RBF is 97.12%, a little lower than the proposed method, while, that of Markov is the lowest, only 94.72%. According to this analysis, the experiment has verified that the hybrid model by forecasting the subseries with different frequencies is helpful to improve the accuracy.

5. Conclusions

For the special volatility and periodicity characteristics of the short-time load, a new hybrid model based on wavelet analysis has been developed. From the prediction model combining RBF and Markov for non-stationary time series, it can be found that the whole prediction is efficient not only the variation trend, but also describes details on high frequency. Compared with the standard RBF and Markov, the proposed model obviously over performs with higher accuracy.

It should be noted that in the Markov forecasting process, the key is how to determine the state transition probability matrix which impact the prediction accuracy. In general, once it is determined, this state transition probability matrix is always used to forecast. Over time, the original data may no longer to adapt to the new situation, namely it can’t reflect the real system state transition law. If the new information is complemented in time, it can further improve the accuracy of prediction.

6. Acknowledgement

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant Nos. 12QX23.

7. References