A PSO-SVM-based 24 Hours Power Load Forecasting Model

Xiaoxu Yu & Haisen Ji
State Grid Liaoyang Electric Power Supply Company, Liaoyang, Liaoning, China

ABSTRACT: In order to improve the drawbacks of over-fitting and easily get stuck into local extremes of BACK propagation Neural Network, a new method of combination of wavelet transform and PSO-SVM (Particle Swarm Optimization- Support Vector Machine) power load forecasting model is proposed. By employing wavelet transform, the authors decompose the time sequences of power load into high-frequency and low-frequency parts, namely the low-frequency part forecast with this model and the high-frequency part forecast with weighted average method. With PSO, which is a heuristic bionic optimization algorithm, the authors figure out the preferable parameters of SVM, and the model proposed in this paper is tested to be more accurately to forecast the 24h power load than BP model.

Keywords: wavelet transform; particle swarm optimization; support vector machine

1 INTRODUCTION

Power load forecasting is a predictive mathematical model based on the characteristics of the power load, integration of economics and social and meteorological data to predict and prepare for future power load changes.

According to the time this model forecasted, the power load forecasting can be categorized as long-term, medium-term, short-term and super short-term power load forecasting. According to the power supply department, the power load forecasting can be categorized as industrial, commercial, agricultural and residential power load forecasting. According to the purpose of power using, the power load forecasting can be categorized as power, illumination and electric heating power load forecasting.

In all the above classifications, the short-term power load forecasting, especially the 24 hours power load forecasting, is a major work of power grid facilities. Accurate 24 hours power forecasting can help the power grid staff develop a balance power supply and demand plan and estimate the production, transmission, distribution and consumption situation in advance [8].

The present power load forecasting method is mainly known as the time series method, the regression analysis method, the exponential smoothing method, the artificial neural network method, the support vector machine (SVM) method and the chaos forecasting method and so on [3]. With an earlier start in research, company and researchers from abroad take widely use of neural network method combined with other model to predict the power load. Chinese researchers also develop many kinds of predicting methods despite the relatively short time putting into this field [5].

This paper proposed a 24 hours power load fore-casting model using support vector machine (SVM) optimized by wavelet transform and particle swarm algorithm, avoiding the shortcoming of artificial neural network method. A simulation test of the optimized SVM model is also carried out in this paper.

2 SUPPORT VECTOR MACHINE (SVM)

2.1 Algorithm of SVM

The basic idea of support vector machine (SVM) algorithm is to construct an optimal hyperplane from the sample space or feature space to make the margin between different types of data points as far as possible. As shown in Figure 1, squares and circles represent two types of data points. The squares and circles on the boundary of each type is the type specific SVM, which forms hyperplane H_1 and H_2 as the hyperplane of different classification data's hyperplane. The margin between the two hyperplane is the two data classifications' max margin, which contains the optimum classification hyperplane.
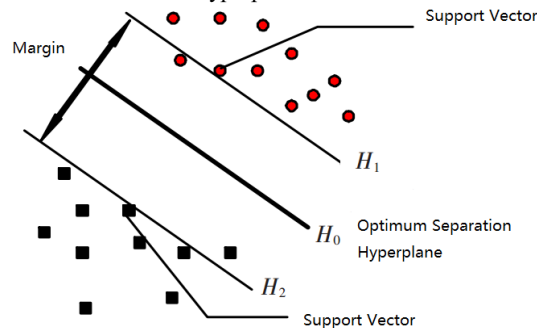

Figure 1. Theory of SVM

If the sample set $S = \{(x_i, y_i) \in R^n \times \{\pm 1\}, i = 1, \cdots, m\}$, x is the input sample, and y is the output sample. One type of data is represented as $x \in A, y = 1$, and the other is represented as $x \in B, y = -1$. The Algorithm's goal is to use training data to estimate a classfier, which is the function $f : R^n \to \{\pm 1\}$. If the training data is linear and integrable, then the linear boundary of this training data is $(w, x) \in R^n \times R$ $w^T x + b = 0$, $w^T x + b \geq 1(x \in A)$ and $w^T x + b \leq -1(x \in B)$, which can be shown as $y_i[(w^T \bullet x_i) + b - 1] \geq 0$; the decision function is $f_{w,b}(x) = sign(w^T x + b)$, w is the weighed vector, and b is deviation. Finally, this question can be transformed into solving the quadratic linear optimization equation:

$$\begin{cases} \min \dfrac{\|w\|^2}{2}, \\ s.t. y_i[(w^T \bullet x_i) + b - 1] \geq 0, i = 1, \cdots, m. \end{cases}$$
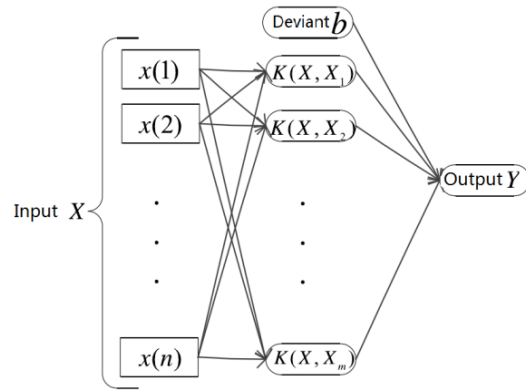

Figure 2. Structure of SVM

2.2 *Kernel function of SVM*

The system structure of SVM can be shown in Figure 2.

In Figure 2, K is the kernel function. With proper kernel function, a non-linear problem can be transformed into a high-dimension linear problem, which can use the method in chapter 2.1 to get the optimum classification plane.

Two types of kernel function have been frequently used [9]:

Polynomial function:

$$K(x_i, x_j) = [(x_i \bullet x_j) + 1]^q, q \in N^*$$

Gauss Radial Basis Function:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$$

The polynomial function has been widely used. However, when dealing with high dimensional problem and compared with the Gauss radial basis function, the calculating progress can be tedious. Therefore, this paper uses Gauss radial basis function [10].

3 INPUT DATA OPTIMIZATION WITH WAVELET TRANSFORM

3.1 *Wavelet transform theory*

Power load time series can be divided into high frequency and low frequency series. The prediction of low frequency series data can be done with SVM; the prediction of high frequency series data can be done with strong randomness; the weighted average method will be applied [6]. The combination and reconstruction of high and low frequency predict data will reduce the overall randomness of power load prediction. Therefore, the power load time series data must be firstly degenerated with wavelet transform.

3.2 *Wavelet function*

Communication signals have a strong non-linear characteristic. For one mother wavelet $\phi(t)$ and in order to analyze its characteristic at one fixed time, we need to scale, translate and integrate the mother wavelet.

Assume that $\phi(t)$ is a square integral function. If the Fourier transform of $\phi(t)$ meets the follow condition:

$$C_\psi = \int_R \frac{\left|\phi(w)\right|^2}{|w|} dw < \infty$$

Then, $\phi(t)$ is called as the basic wavelet or the mother wavelet function. With scaling and translating, the wavelet function is equal to:

$$\phi_{a,b}(t) = a^{-\frac{1}{2}} \phi(\frac{t-b}{a}) a, b \in R; a \neq 0$$

$\phi_{a,b}(t)$ is wavelet function, a and b is the scaling and translation factor. When a is increased, the wavelet function's waveform is widened but the general form is remained; when a is decreased, the wavelet function's waveform is narrowed but the general form is remained. So we say the frequency increases. The reciprocal of the scaling factor 1/a corresponds to frequency to a certain context, and the translation factor usually corresponds to the time factor.

3.3 *Mallat algorithm of wavelet transform*

Assuming that the scale function $\phi(t)$ corresponds to wavelet function $\psi(t)$, then:

$$\begin{cases} \phi(t) = \sum_n h(n)\phi(2t-n) \\ \psi(t) = \sum_n g(n)\phi(2t-n) \end{cases}$$

Orthogonal projection $f(t)$ to space V_j, W_j gets approximation signal $H_j f$ and detailed signal $D_j f$ under the resolution of 2^j:

$$H_j f = \left\langle f(t), \phi_{j,k}(t) \right\rangle = 2^{-\frac{j}{2}} \int_R f(t)\phi(2^{-j}t-k)dt$$

$$D_j f = \left\langle f(t), \phi_{j,k}(t) \right\rangle = 2^{-\frac{j}{2}} \int_R f(t)\psi(2^{-j}t-k)dt$$

Then $H_j f$ can be decomposed to be the sum of approximation signal $H_{j+1}f$ under resolution 2^{j+1}, and the detailed signal $D_{j+1}f$ between resolutions 2^{j+1} and 2^j is:

$$\begin{cases} H_{j+1}f = \sum_k h(k-2n)H_j f \\ D_{j+1}f = \sum_k g(k-2n)D_j f \end{cases}$$

So under different resolutions, original signals can be decomposed into approximation signals and detailed signals. If $f(t) = H_0 f$ represents the original signals, the number of iterations is $J(1 \le j \le J)$, and $H_0 f$ can be decomposed as:

$$f(t) = H_0 f = H_J f + \sum_{j=1}^{J} D_j f$$

Meanwhile, $H_{j+1}f$ and $D_{j+1}f$ can reconstruct signal $H_j f$ as:

$$H_j f = \sum_k h(n-2k)H_{j+1}f + \sum_k g(n-2k)D_{j+1}f$$

The Mallat algorithm of signal decomposition process can be shown in Figure 3.

The reverse process of decomposing is a reconstruction process, which is shown in Figure 4.

Based on the method described above, the input signals of support vector machine (SVM) can be decomposed into high and low frequency signals, among which the low frequency signals are always to be useful signals and the high frequency signals to be the noise [11]. To remove the high frequency signals of the input data, it's a method to reduce the noise in input data, whose characteristic will still be reflected in the remaining low frequency signals.
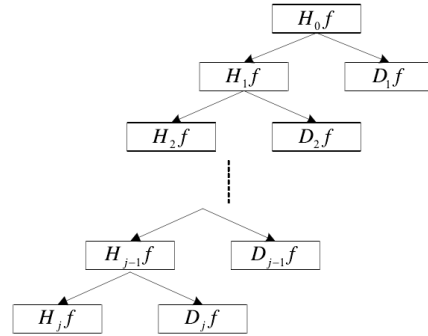

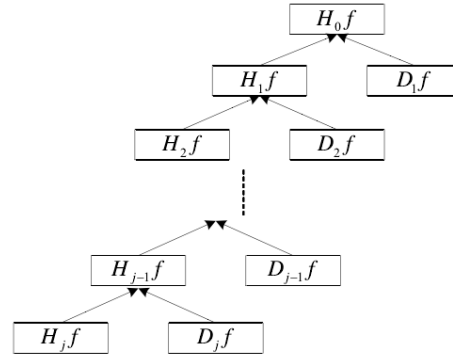Figure 3. Signal decomposing process through Mallat algorithm

Figure 4. Signal reconstruction process through Mallat algorithm.

4 OPTIMIZATION OF SVM PARAMETER SELECTION WITH PSO

4.1 *Main parameters of SVM*

Error penalty parameter c and kernel function variance g are the two main parameters affecting the SVM [12].

Currently, the most commonly-used method of optimizing SVM parameter selection is to generate groups of value c and g in a certain range, then predict the forecast accuracy with K-CV method of these groups' value c and g, which results the highest forecast accuracy will be selected [4]. In order to prevent the happening of over-learning, in general, a group among the highest forecast accuracy groups will be selected with the minimal value of variance c. If one corresponds to the multiple values of g, then we select the one that comes first in searching.

The results of training the SVM with generating values of c and g in the range of [0,100] are shown in Table 1:

Table 1. The forecast accuracy of randomly generating parameter SVM

No.	Value of c	Value of g	Forecast Accuracy
1	43.6473	92.3564	38.3618%
2	35.1584	84.3479	81.4181%
3	5.5184	33.8137	43.7197%
4	82.4631	79.5135	61.5478%
5	8.1615	18.3518	48.6812

Through the table above, it can be seen that the random generating value of c and g can't guarantee and idealize the final prediction accuracy. The selection of proper value through experience wouldn't be a reliable way, so the particle swarm optimization (PSO) algorithm is needed for selection of parameters of SVM.

4.2 Particle swarm optimization (PSO) algorithm

Particle swarm optimization algorithm will first determine the particle's initial state, including initial position and velocity. N is the size of initial particles, and the position and velocity of number i particle in a D dimension space can be represented as follows:

$$X_i = [x_{i1}, x_{i2}, \cdots, x_{iD}]$$
$$V_i = [v_{i1}, v_{i2}, \cdots, v_{iD}]$$

Though evaluation of the fitness of the particle, the particle's best position P_{best} and the group's best position are as follows:

$$P_i = [p_{i1}, p_{i2}, \cdots, p_{iD}]$$
$$P_g = [p_{g1}, p_{g2} \cdots, p_{gD}]$$

To find the fitness position of particle I, we can use the equation as follows:

$$P_i(k+1) = \begin{cases} P_i(k) & if \ f(X_i(k+1)) \geq f(P_i(k)) \\ X_i(k+1) & if \ f(X_i(k+1)) < f(P_i(k)) \end{cases}$$

The best position of all the particles in the swarm is as follows:

$$P_8 \in \{[P_1(k), P_2(k), \cdots P_D(k)] \big| f(P_8) = \min(f(P_1(k), P_2(k), \cdots, P_D)))\}$$

Through continuous iteration, the update of the velocity and position of particles can be tracked by P_{best} and g_{best}:

$$V_{ij}(k+1) = \omega V_{ij}(k) + c_1 r_1 [P_{ij}(k) - X_{ij}(k)] + c_2 r_2 [P_{8j}(k) - X_{ij}(k)]$$
$$X_{ij}(k+1) = X_{ij} + V_{ij}(k+1), i = 1,2,\cdots,N, j = 1,2,\cdots,D$$

In the above two formulas, the inertia weight factor is ω, and the normal acceleration are c_1 and c_2,

whose value are often between 0 and 2. Meanwhile, with proper velocity interval $[-v_{max}, v_{max}]$ and position range $[x_{min}, x_{max}]$, the particles movement will be effectively restricted.

4.3 Process of particle swarm optimization algorithm

With the analysis above, the process of particle swarm optimization (PSO) can be shown in Figure 5:
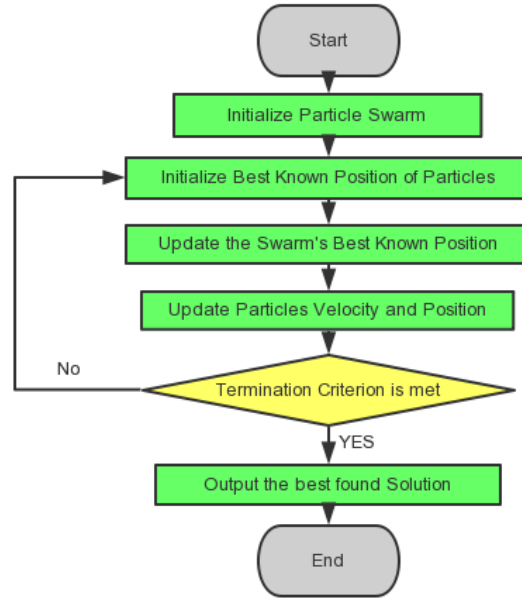

Figure 5. Process of PSO algorithm

(1) Initialize all the particles in the swarm randomly.
(2) Once getting the particle's optimal position, we use P_{best} to lock the particle's location, search for g_{best} to save all the optimal location information and adapt to the information of target identification [7].
(3) Compare each particle's optimal position to the previous optimal location information in order to update the information to the particle's newest location.
(4) Compare the global optimal position information to the entire particle's optimal location information. If the later one is smaller, then we update the variance g_{best}'s value.
(5) If the termination condition is met, that is, has a good enough optimal value or has reached the maximum number of iterations, thus output g_{best} and the optimal position information and terminate the process.

5 TWENTY-FOUR HOURS POWER LOAD FORECASTING MODEL

5.1 Preparing data

The first step of power load forecast is to have a clear

object, and collect historical data, and then apply the model to learn the pattern. A large number of historical power load data are essential to get an accurate predicting power load [13]. Therefore, to increase the power load predicting accuracy, just gathering the historical power load data is not enough, the data must be processed in certain degree.

An abnormal data criterion must be set to screen the abnormal data, such as a standard county. When deviance exceeds the standard interval, the data from the county can be regarded as illegal data.

5.2 Wavelet transform

Wavelet toolbox in Matlab has the wavelet transform function. According to the time series, it can decompose the data into 4 layers, including the high frequency data and the low frequency data.
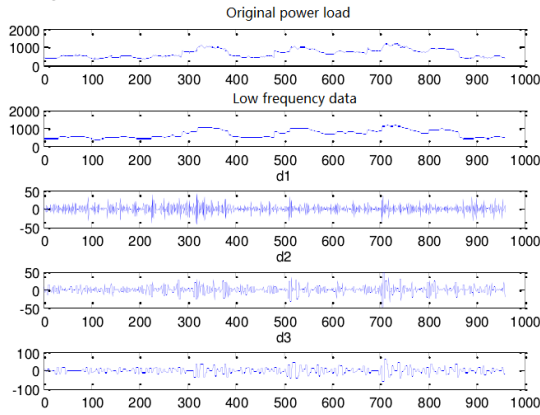
Figure 6 shows the wavelet transform results:

Figure 6. Results of wavelet transform

5.3 Kernel function and parameter selection

Due to the high dimension input vector and large amount of training samples, the Gauss Radial Basis Function has been used as follows:

$$K(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2), \gamma > 0$$

Meanwhile, optimize the parameter selection with particle swarm optimization algorithm to get the final results: $c=110$, $g=1.6$.

6 SIMULATION AND VERIFICATION

In order to verify the method in this paper, a prediction has been carried out with power load data from 1st June, 2014 to 31th July, 2014 at a municipal electricity company in a southwest coastal city. These three days, 21th, 22th and 23th of July in 2014, have been predicted. First, we use db3 wavelet transform to decompose the power load data into 4 layers wavelet, then predict the

low frequency data with particle swarm optimization algorithm with SVM, and predict the high frequency data with weighed average method. The results are shown in Figures 7, 8 and 9 corresponding to the power load forecast data on 21th, 22th and 23th respectively. In Figures 7, 8 and 9, BP for BP neural network model predicting data, PSO-SVM is the method in this paper.
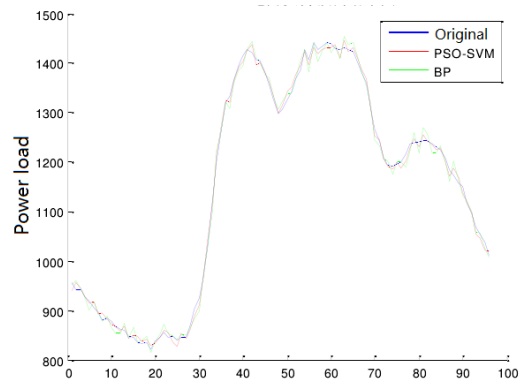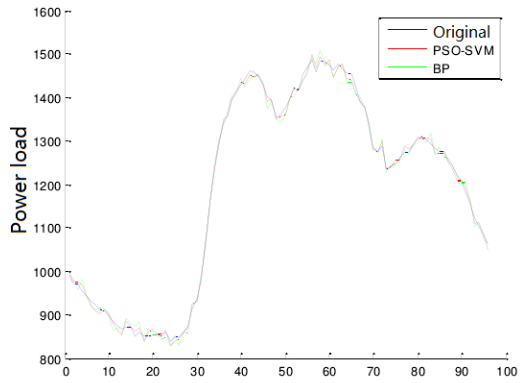

Figure 7. Forecasting curve on 21th June, 2014

Figure 8. Forecasting curve on 22th June, 2014

Table 2. Forecasting accuracy on 21th July, 2015

	BP	PSO-SVM
MAPE	3.2824%	2.5811%

Table 3. Power load forecasting accuracy on 22th July, 2015

	BP	PSO-SVM
MAPE	4.1289%	2.6835%

Table 4. Power load forecasting accuracy on 23th July, 2015

	BP	PSO-SVM
MAPE	4.1269%	2.6008%

BP network and PSO-SVM 24h power load forecasting error and accuracy rate are listed in Tables 2, 3 and 4. In conclusion, the PSO-SVM method proposed in this paper on the prediction precision is better than BP network. It fully shows that the method used for 24h power load forecasting has a strong practical value.
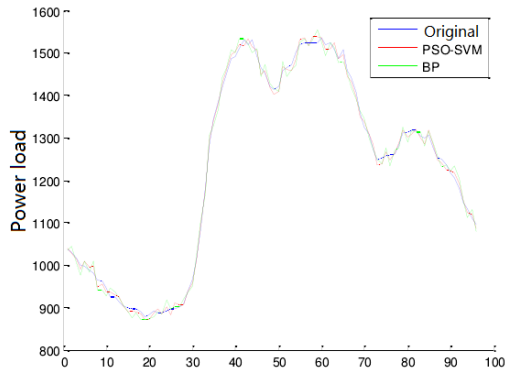

Figure 9. Forecasting curve on 23th June, 2014

7 CONCLUSION

A method of using wavelet transform to reduce the noise in input data and optimizing the parameter selection with particle swarm optimization algorithm to predict the 24 hours power load is proposed in this paper. This power prediction model is used to reduce the input data randomness and deal well with small, nonlinear sample and sample with extreme minimum point. By testing this model with 2 months' actual data from a municipal power supply department to predict 24 hours power load, this model's results are better than which is obtained from the BP neural network model. Thus the model proposed in this paper has a great prospect in power system scheduling and distribution.

REFERENCES

[1] Zheng Jin. 2012. Short-term power load forecasting method research [D]. Zhengzhou University.

[2] axlk. 2012. Research on short-term power load forecasting. China's new technology and new products, (14): 131-132.

[3] xu. 2012. *Load forecasting new method research*. Nanchang university,

[4] Wang Shuange. 2012. Small sample event prediction based on support vector machine (SVM). Huazhong University of science and technology.

[5] philosophy, ShuQin. 2015. Short-term power load forecasting based on decomposition of ESPRIT algorithm. *Power system protection and control*, (7): 90-93.

[6] FanYiNa. 2012. Short-term power load forecasting based on wavelet transform. *Journal of qinghai university: natural science edition*, (1): 25 to 29.

[7] Wang Yijun & Li Dianwen, etc. 2015. Based on improved PSO - SVM's short-term power load forecasting. *Electric measurement and instrument*, 52 (3): 22-24.

[8] Yang Fangyuan. 2015. Distribution network planning in the power load forecasting method of application. *Journal of engineering construction and design*, (2): 50-54.

[9] Zhao Yuan zhang, & xia fe. 2011. Nonparametric regression method in the short-term power load forecasting application. *High voltage technology*, 37 (2): 429-435.

[10] gui-feng wu, Wang Xuan, & Chen Donglei. 2013. Short-term power load forecasting based on the optimization of the neural network. *Computer simulation*, 30 (11): 95-99.

[11] Yi Tao, feng-yun Chen, & Lin Yangyu, etc. 2014. Based on wavelet transform analysis and monthly load hybrid prediction. *Computer simulation*, 31 (6): 137-141.

[12] Liang Danya, hong-wei li, & Wang Ke, etc. 2015. The particle swarm optimization (GPSR - BB compression perception reconstruction algorithm. *Journal of air force engineering university: natural science edition*, (1): 81-84.

[13] liu, wen-jie fu. 2011. Short-term power load forecasting model evaluation. *Journal of China electric industry: technology edition*, (12): 52-55.