A Method for Power Load Forecasting Base on SVM and Wavelet Neural Network

Tongna Liu

***Abstract*—This paper put forward a new method of the SVM and wavelet neural network model for short-term load forecasting. The neural call function is basis of nonlinear wavelets. We overcome the shortcoming of single train set of SVM. It can be seen from the example this method can improve effectively the forecast accuracy and speed. The forecast model was tested and the result showed that it was an effective way to forecast short-term electric load.**

***Index Terms*—SVM, Wavelet Neural Network, Electric Load Forecasting**

I. INTRODUCTION

A short-term electric load forecasting means the load forecasting that time units are the hour, day or month. Because its tendency has a strong randomness, the determination of mathematical models is difficult. The improvement of forecasting accuracy is difficult.

Traditional load forecasting methods, including time series, regressive analysis and so on, can't meet the need of predicted accuracy demand in practical community. On the contrary, modern artificial intelligence methods, represented by artificial neural network (ANN), support vector machines (SVM) etc, have a certain ability of self-leaning, self-adapting and powerful continuous function approximation. As their outstanding performance in the field of non-linear application, modern methods are more feasible than traditional ones.

Comparing with the way of ANN's machine learning, that of SVM is structural risk minimization (SRM) inductive principle instead of empirical risk minimization (ERM) inductive principle. So, outfitting phenomenon hardly appears in SVM's learning process of finite amount of training data. Consequently, SVM's generalization performance is better than ANN's. In addition, due to its training process equals to solve a quadratic programming, its optimal solution must not be local extreme value.

II. PRINCIPLE OF SVM

Suppose an set of data $\{(x_1, y_1), \Lambda, (x_N, y_N)\}$, where, x_i is the i_{th} input vector and y_i is the corresponding desired output. For $i = 1,2,\Lambda, N$, where N is size of the samples. The estimating function takes the form as follows:

$$f(x) = (w \cdot \varphi(x)) + b \tag{1}$$

Where, w is the weight vector, b the bias and $\varphi(x)$ the high dimension feature space which is non-linearly mapped from the input space and $(\cdot)$ represents the inner product.

This leads to the optimization problem for standard SVM.

$$\min\ R_{str} = \frac{1}{2}\|w\|^2 + \gamma R_{emp} \tag{2}$$

Where, γ is a positive real constant which determines penalty to estimation errors, and $R_{emp}(w,b) = \frac{1}{N}\sum_{i=1}^{N}|y_i - f(x_i)|_\varepsilon$ is the estimation errors by experimental risk and measured by loss function. Usually the ε- insensitive loss function is adopted for its excellent sparsity.

$$|y - f(x)|_\varepsilon = \begin{cases} 0, & |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon, & elsewhere \end{cases} \tag{3}$$

For LS-SVM, the 2-norm of estimation error is adopted as the loss function in the objective function and equality constrains instead of inequality constrains. So the optimization problem is described as follows:

$$\begin{cases} \min\limits_{w,b,\xi_i} & \frac{1}{2}w^T w + \frac{1}{2}\gamma\sum_{i=1}^{N}\xi_i^2 \\ s.t. & y_i = w^T\varphi(x_i) + b + \xi_i \quad i = 1,2,\Lambda, N \end{cases} \tag{4}$$

Where, ξ_i is a slack variable.

After the introduction of Lagrange multipliers α_i, constructing the Lagrange function as following:

$$L = \frac{1}{2}w^T w + \frac{1}{2}\gamma\sum_{i=1}^{N}\xi_i^2 - \sum_{i=1}^{N}\alpha_i\{w^T\varphi(x_i) + b + \xi_i - y_i\} \tag{5}$$

According to KKT Conditions, the equation as follows can be obtained:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \\ \frac{\partial L}{\partial b} = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \\ \frac{\partial L}{\partial \alpha_i} = 0 \end{cases} \Rightarrow \begin{cases} w = \sum_{i=1}^{N}\alpha_i\varphi(x_i) \\ \sum_{i=1}^{N}\alpha_i = 0 \\ \alpha_i = \gamma\xi_i \\ w^T\varphi(x_i) + b + \xi_i - y_i = 0 \end{cases} \tag{6}$$

After eliminating w and γ, we can obtain:

$$\begin{bmatrix} 0 & \Theta^T \\ \Theta & \Omega + \frac{1}{\gamma} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (7)$$

where, $\Theta = [1, \Lambda ,1]_{1\times N}$, I is an unit matrix, Ω is a square matrix, and the element of Ω is expressed as follow: $\Omega_{ij} = \varphi(x_i)^T \varphi(x_j)$. $\alpha = [\alpha_1, \Lambda , \alpha_N]$, $y = [y_1, \Lambda , y_N]$.

Solving the equation (11), the values of α and b are gotten. According to Mercer's Condition, there exists a kernel function that the value of which equals to the inner product of two vectors xi and xj in the feature space $\varphi(x_i)$ and $\varphi(x_j)$, that is $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$. Then, LS-SVM model for regression is expressed as follows:

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \quad (8)$$

II. The Wavelet Concept and the Wavelet Transform

For setting up the load forecast model of wavelet neural network, first we introduce some basic concepts of wavelet and wavelet transform.

We call the square integral function $\varphi(t) \in L^2(R)$ that it is asked to satisfy the admissible condition:

$$\int_{-\infty}^{+\infty} [\hat{\varphi}(\omega)]^2 [\omega]^{-1} d\omega < +\infty \quad (9)$$

basic wavelet or mother wavelet. $\hat{\varphi}(\omega)$ is the Fourier Transform of the φ(t). Let

$$\varphi_{ab}(t) = \frac{1}{\sqrt{|a|}} \varphi(\frac{t-b}{a}) \quad (10)$$

Among them, a, b are real number, and a≠0, φ_{ab} is called the wavelet basis that it is generated by mother wavelet and it depends on the parameter a,b.

Let $f(t) \in L^2(R)$, The f(t) is tendency function that it shows the variance law of the short-term electric load.

Let the wavelet transform of the f(t)

$$W_f(a,b) = (f, \varphi_{ab}) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \varphi(\frac{t-b}{a}) dt \quad (11)$$

Due to the specificity of the load, this transform is discussed only in the real numbers. From(3), the variance of the parameter b in the wavelet basis has the use of translation. The parameter a not only changes the frequency spectrum structure but also changes the length of its window. Thus, a, b are called respectively the expansion and contraction factor and the translation factor of φ_{ab} (t). The similarity of the Fourier analysis, the wavelet analysis resolves the signal function into the wavelet normal orthogonal basis. It constructs a series to approximate the signal function. This linear combination has an optimum recognition capacity.

The mother wavelet is asked to satisfy condition (9), Thus we get

$$\int |\varphi(t)|^2 dt < \infty \quad (12)$$

The condition (1) determines the locality behavior of the wavelet. It equals 0 without a limited interval or approximates 0 fast. The formula (4) determines that the wavelet has the limited energy and is an oscillation (the positive number part equals the negative number part). The wavelet's name is produced from here.

The mother wavelets with the better locality property and smooth property have the spline wavelet and Merlot wavelet usually. Its system of the expansion and contraction and translation composes the normal orthogonal basis of $L^2(R)$. The wavelet series generated can approximate f(t) optimally.

The similarity of the Fourier transform, the wavelet transform has also a inversion formula.

$$f(t) = \frac{1}{C_\varphi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_f(a,b) \varphi_{ab}(t) \frac{da db}{a^2} \quad (13)$$

Among them,

$$C_\varphi = \int_{-\infty}^{+\infty} |\varphi(\omega)|^2 |\omega|^{-1} d\omega \quad (14)$$

III. The Load Forecasting Model of the Wavelet Neural Network

In the wavelet neural network, we replaces the nonlinear sigmoid function with the nonlinear wavelet basis. The fitting of a load historical sequence is completed with the linear superposition of the nonlinear wavelet basis. The limited terms of a wavelet series can approximate the load historical curve. The load curve can be fitted with the wavelet basis $\varphi_{ab}(t)$

$$\hat{y}(t) = \sum_{k=1}^{L} \omega_k \varphi(\frac{t-b_k}{a_k}) \quad (15)$$

In (15), $\hat{y}$ (t) shows the forecast values sequence of load curve y(t). ω_k are weight numbers. b_k are translation factors. a_k are expansion and contraction factors. L is wavelet basis number. In Figure(1), we give the structure of the wavelet neural network. (see Figure 1)The network is a single implicit layer. It only contains a importation node and a exportation node. We need to determine the network parameters ω_k a_k b_k and L. Let the sequence y(t) and sequence $\hat{y}(t)$ are fitted optimally. The parameter ω_k a_k and b_k can be optimized on the basis of the minimum square error energy function EL.

$$E_L = \frac{1}{2} \sum_{t=1}^{L} [y(t) - \hat{y}(t)]^2 \quad (16)$$

We determine L with the method of stepwise test, Thus the network structure is determined also. to determined every L, we .compute optimal parameters ω_k a_k and b_k by(16) as follows.

First, we use Morlet mother wavelet in(15)

$$\varphi(t)=\cos(1.75t)\exp(-\frac{t^2}{2}) \qquad (17)$$

Let $T=\frac{t-b_k}{a_k}$, then the gradient of (16) is showed respectively as follows.

$$g(\omega_k)=\frac{\partial E}{\partial \omega_k}=-\sum_{t=1}^{L}[y(t)-\hat{y}(t)]$$
$$\cos(1.75T)\exp(-\frac{T^2}{2})$$
$$g(b_k)=\frac{\partial E}{\partial b_k}=-\sum_{t=1}^{L}[y(t)-\hat{y}(t)]\omega_k$$
$$[1.75\sin(1.75T)\exp(-\frac{T^2}{2})$$
$$+\cos(1.75T)\exp(-\frac{T^2}{2}T)]/a_k$$
$$g(a_k)=\frac{\partial E}{\partial a_k}=-\sum_{t=1}^{L}[y(t)-\hat{y}(t)]\omega_k$$
$$[1.75\sin(1.75T)\exp(-\frac{T^2}{2})T$$
$$+\cos(1.75T)\exp(-\frac{T^2}{2})T]/a_k=Tg(b_k)$$

Network parameters ω_k b_k and a_k are optimized with the conjugate gradient method. Let vector $\omega=(\omega_1,\omega_2,\cdots\omega_k\cdots\omega_L)$, $g(\omega)=(g(\omega_1),g(\omega_2)\cdots g(\omega_k)\cdots g(\omega_L))$. $S(\omega)^i$ shows the ith cyclic search direction of the function of w. Then

$$S(\omega)^i=\begin{cases}-g(\omega)^i & i=1\\ -g(\omega)^i+\dfrac{g(\omega)^{iL}g(\omega)^i}{g(\omega)^{(i-1)L}g(\omega)^{i-1}}\cdot s(\omega)^{i-1} & i\neq 1\end{cases}$$

The weight vector is regulated as follows.

$$\omega^{i+1}=\omega^i+\alpha_w^i S(\omega)^i \qquad (18)$$

IV. The Applied Study of Short-term Load Forecasting In China xx city Power

Applying the new method put forward in the paper, we study the short-term load forecasting in xx city Power Network. The selected forecasting data is the history data and weather factors in JUNE 2016.The short-term electric load of 24 o'clock of JUNE, 17, 2008 is to be forecast. To compare the two models, forecast methods are elected by SVM and wavelet neural network model (SWANN) and artificial neural network model (ANN) respectively. Through the imitation computation, we know that the accuracy and speed of the SWANM are raised obviously. (see Figure 1,2,3)

Table 1 The new method forecasting analysis of the load in China xx city power system

Time	Actual load	ANN Forecasting	ANN Relative error%	SWANN Forecasting	SWANN Relative error%
1	2209.912	2162.392	-2.15%	2215.225	0.24%
2	2114.247	2102.179	-0.57%	2121.111	0.32%
3	2088.159	2062.752	-1.22%	2094.544	0.31%
4	2148.301	2143.994	-0.20%	2143.399	-0.23%
5	2207.209	2206.375	-0.04%	2200.454	-0.31%
6	2224.643	2215.408	-0.42%	2225.837	0.05%
7	2270.604	2292.645	0.97%	2275.001	0.19%
8	2233.972	2242.781	0.39%	2221.111	-0.58%
9	2314.449	2303.499	-0.47%	2309.175	-0.23%
10	2523.906	2530.878	0.28%	2513.067	-0.43%
11	2544.783	2565.338	0.81%	2563.816	0.75%
12	2474.724	2508.995	1.38%	2472.217	-0.10%
13	2426.435	2454.469	1.16%	2407.710	-0.77%
14	2542.619	2502.070	-1.59%	2517.922	-0.97%
15	2501.070	2546.494	1.82%	2518.017	0.68%
16	2565.568	2560.285	-0.21%	2558.050	-0.29%
17	2726.226	2720.560	-0.21%	2744.481	0.67%
18	2762.281	2773.929	0.42%	2740.895	-0.77%
19	2800.654	2781.051	-0.70%	2801.701	0.04%
20	2629.729	2596.536	-1.26%	2618.138	-0.44%
21	2569.603	2587.328	0.69%	2563.304	-0.25%
22	2575.274	2618.458	1.68%	2569.316	-0.23%
23	2490.948	2474.521	-0.66%	2475.831	-0.61%
24	2332.599	2297.764	-1.49%	2317.893	-0.63%

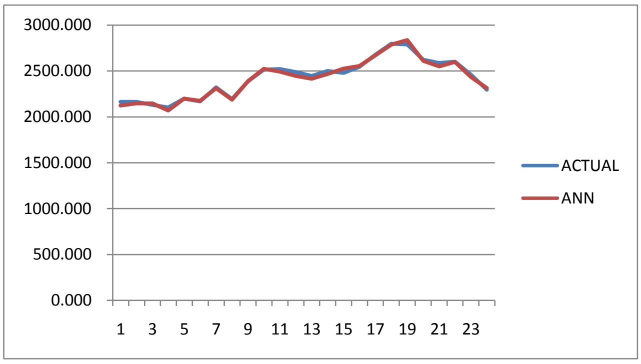

Fig1. The ANN analysis of the load in China xx city power system

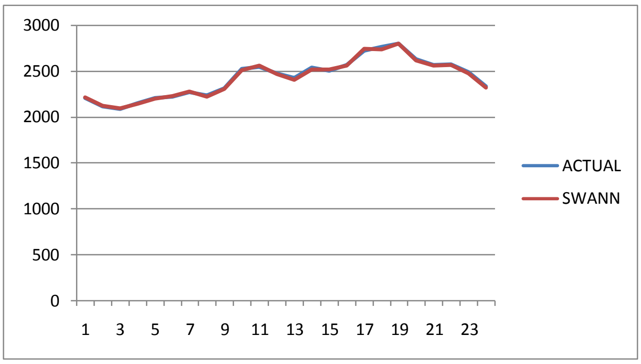

Fig2. The SWANN analysis of the load in China xx city power system

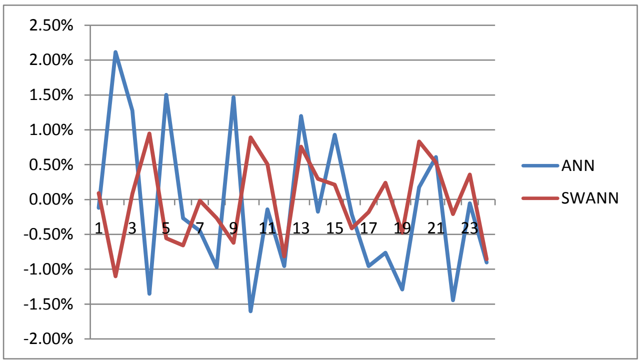

Fig 3. Relative error analysis of the load in China xx city power system

V. Conclusions

In this paper, we propose the SVM and wavelet neural network forecast model of short-term electric load. It overcomes the intrinsic defects of a artificial neural network that its learning speed is slow, its network structure is difficult to determine rationally and it produces local minimum points. Its nervous cell function is basis of nonlinear wavelets. We get the global optimum fitting effect. Then the accuracy is improved. The network structure is determined rationally with the stepwise test method, Because the network is a single implicit layer structure, Its speed is improved obviously It can be used to forecast short-term electric load.Through the imitation computation, we prove that the accuracy and speed are improved obviously. This is a new and effective method of short-term electric load forecasting.

Acknowledgment

This research was supported by the Fundamental Research Funds for the Central Universities 2018MS092

References

[1] Tai Neng-ling, Hou Zhi-jian. New short-term load forecasting principle with the wavelet transform fuzzy neural network for the power systems[J]. Proceedings of CSEE, Vol.24, No.1, 2014, pp. 24-29.

[2] Freund Y, Schapire RE, A decision-theoretic generalization of on-line learning and an application to boosting, In Proceedings of the 2nd European Conference on Computational Learning Theory, Barcelona, 2013, pp.23-37.

[3] Hendrik B, David P, Multi-Instance tree learning, In Proceedings of the 22nd international conference on Machine learning, 2015, pp.57-64.

[4] Lee J, Chai J, Paul S. A. Reitsma, Jessica K. Hodgins and Nancy S. Pollard, Interactive Control of Avatars Animated with Human Motion Data, In proceedings: SIGGRAPH, 2012, pp.491-500.

[5] Liu F, Liang RH, Motion Path Synthesis for Intelligent Avatar, Proceedings from the 4th International Working Conference on Intelligent Virtual Agents, LNAI 2792, 2013, pp. 141-149.

[6] Liu F., Zhuang YT, Wu F, Pan YH, 3d Motion Retrieval with Motion Index Tree, Computer Vision and Image Understanding, Vol.92(2-3), 2013, pp.265-284

Tongna Liu in Department of Electronic and Communication Engineering North China Electric Power University, Baoding China. The main research new energy revenue management.