Short-term Load Forecasting Based on FCM and Complex Gaussian Wavelet SVM

Yongkang Zheng Weirong Chen Chaohua Dai Shengyong Ye

School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, P. R. China

Abstract

Complex Gaussian wavelet support vector machine (CGW-SVM) is constructed with complex Gaussian wavelet kernel function for short-term load forecasting (STLF). Based on the chaotic characteristics of short-term load time series, the series is reconstructed with phase space reconstruction theory (PSRT). Then the vector of phase space is used as the input of CGW-SVM. Considering the periodical feature of power loads, the fuzzy c-means (FCM) clustering algorithm is introduced to reduce sample data. The experiments show that, by applying the proposed method, the accuracy of load forecasting results is improved and the forecasting process is speeded up.

Keywords: complex Gaussian wavelet support vector machine, fuzzy c-means clustering; short-term load forecasting, phase space reconstruction, complex Gaussian wavelet kernel, support vector machine

1. Introduction

Short-term load forecasting (STLF) is a crucial issue for power systems. Load forecasting helps the system operator to schedule reserve allocation efficiently and is also useful to power system security. Since the power system is a complicated nonlinear system which exhibits chaotic behavior, precise forecasting of short-term load is still a difficult task [1].

Many techniques for STLF have been tested with different degrees of success. Traditional techniques include time-series models, regression models, Kalman filtering models, autoregressive (AR) model, and so on [2]. These models and techniques are basically linear methods and have limited ability to capture nonlinearities in the short-term load series. Artificial neural networks (ANN) have also been proposed for STLF [3]. The ANN extracts the implicit non-linear relationship among input variables by learning from training data without making complex dependency assumptions among input variables. But learning algorithm of ANN lacks quantitative analysis and perfect results since it adopts empirical risk minimization (ERM) principle according to statistical learning theory (SLT), which tries only to minimize experience risk.

Chaotic time series analysis has been studied to investigate complicated nonlinear power systems in recent years [4]. It's based on phase space reconstruction theory (PSRT), which is supported by embedded theory. How to construct load forecasting model with PSRT is a key issue.

Support vector machine (SVM) proposed by Vapnik in 1995 has been applied to load forecasting [5]. It is a small-sample theory firmly grounded on the framework of SLT [6]. SVM is based on the structural risk minimization (SRM) principle to minimize the generalization error rather than the empirical error. According to SVM theory, regression problems can be converted into linear ones, and finally deduced to mathematical problems of quadratics programming. For STLF, RBF function is often selected as the kernel function and the load and weather data is used as the inputs of SVM.

In this paper, we prove that complex Gaussian wavelet can satisfy the kernel function condition, and try to build a kind of complex Gaussian wavelet kernel SVM (CGW-SVM). The short-term load series is reconstructed with PSRT, and the vector of phase space is used as the inputs of CGW-SVM. The fuzzy c-means (FCM) clustering algorithm is introduced to reduce sample data to increase convergence speed. The experimental results show that the proposed method can be believed as one of the most promising methods and has high application value in STLF.

2. CGW-SVM

A common d-dimensional (d-D) wavelet function can be written as the product of one-dimensional (1-D) wavelet function [7]: $\psi_d(x)=\prod_{i=1}^{d}\psi(x_i)$, then the translation-invariant wavelet kernel that satisfies the translation-invariant kernel theorem is:

$$k(x,x')=\prod_{i=1}^{d}\psi\left(\frac{x_i-x_i'}{a}\right) \tag{1}$$

The translation-invariant kernel is an admissive SVM kernel if it satisfies the following theorem [8]:

Theorem: A translation-invariant kernel $k(x, x') = k(x - x')$ is an admissible SVM kernel if and only if the Fourier transform

$$F[k(\omega)] = (2\pi)^{-d/2} \int_{R^d} \exp(-j\omega x) k(x) dx \quad (2)$$

is non-negative.

2.1. Complex Gaussian wavelet kernel

Complex Gaussian wavelet is constructed with the n^{th}-derivative of complex Gaussian function $\psi(t) = C_n (\mathrm{e}^{-\mathrm{j}x} \mathrm{e}^{-x^2})$. C_n is such that the 2-norm of the n^{th}-derivative of $\psi(t)$ is equal to 1. Let $C_n = 1$ in this paper and it won't decrease the performance of SVM. We select the imaginary part of the 1^{st}-derivative, and get complex Gaussian wavelet as follows:

$$f(x) = (-\cos x + 2x \sin x) \exp(-x^2) \quad (3)$$

The SVM kernel of this mother wavelet is

$$k(x, x') = \prod_{i=1}^{d} (-\cos \frac{x_i - x_i'}{a} + 2 \frac{x_i - x_i'}{a} \sin \frac{x_i - x_i'}{a}) \times \exp\left(-\frac{\|x_i - x_i'\|^2}{a^2} \right) \quad (4)$$

It is an admissible SVM kernel function when d is an even number which is proved as follows.

Proof: For all x,

$$\begin{aligned} k(x) &= \prod_{i=1}^{d} \psi(\frac{x_i}{a}) \\ &= \prod_{i=1}^{d} (-\cos \frac{x_i}{a} + 2 \frac{x_i}{a} \sin \frac{x_i}{a}) \exp\left(-\frac{\|x_i\|^2}{a^2} \right) \end{aligned} \quad (5)$$

Substituting (5) into (2), we can calculate the integral term

$$\begin{aligned} &\int_{R^d} \exp(-j(\omega x)) k(x) dx \\ &= \int_{R^d} \exp(-j(\omega x)) \\ &\quad \times \prod_{i=1}^{d} (-\cos \frac{x_i}{a} + 2 \frac{x_i}{a} \sin \frac{x_i}{a}) \exp\left(-\frac{\|x_i\|^2}{a^2} \right) dx \\ &= \prod_{i=1}^{d} \int_{-\infty}^{+\infty} \exp(-j(\omega x_i)) \\ &\quad \times (-\cos \frac{x_i}{a} + 2 \frac{x_i}{a} \sin \frac{x_i}{a}) \exp\left(-\frac{\|x_i\|^2}{a^2} \right) dx \\ &= \prod_{i=1}^{d} [-a^2 \sqrt{\pi} \omega \sinh(wa/2) \\ &\quad \times \exp(-1/4 - \omega^2 a^2/4)] \\ &= (-1)^d \prod_{i=1}^{d} [a^2 \sqrt{\pi} \omega \sinh(wa/2) \\ &\quad \times \exp(-1/4 - \omega^2 a^2/4)] \end{aligned} \quad (6)$$

Substituting (6) into (2), we can obtain the Fourier transform

$$\begin{aligned} &F[k(\omega)] \\ &= (2\pi)^{-d/2} \times (-1)^d \times \\ &\prod_{i=1}^{d} a^2 \sqrt{\pi} \omega \sinh(wa/2) \exp(-1/4 - \omega^2 a^2/4) \end{aligned}$$

Obviously, when d is an even number, $F[k(\omega)] \geq 0$.

2.2. SVM

$x \in R^d$ is the input vector of the SVM, and $y \in R$ is the output. The non-linear function $\Phi(x)$ maps the sample of input space to output space. Generally, the optimization problem for ε-insensitive SVM is given as follow quadratic programming problem:

$$\begin{aligned} &\min_{w,b,\xi} \ \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \\ &s.t. \ \ | y_i - \langle w \cdot \Phi(x) \rangle - b | \leq \varepsilon + \xi_i, \\ &\qquad \xi_i \geq 0, \xi_i^* \geq 0, i = 1, 2, \cdots n \end{aligned} \quad (7)$$

where ξ_i is a slack variable and $C > 0$ is a constant which determines penalties. We solve the optimization problem and get the estimation function as follows:

$$f(x) = \sum_{x \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (8)$$

where $\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i \leq C, 0 \leq \alpha_i^* \leq C$, and $K(x_i, x)$ is SVM kernel. The typical examples of kernel function are as follows:

Linear: $K(x, x_i) = \langle x \cdot x_i \rangle$.

Sigmoid: $K(x, x_i) = \tanh(v \langle x \cdot x_i \rangle + c)$.

Polynomial: $K(x, x_i) = (\langle x \cdot x_i \rangle + 1)^d$.

Radial basis function (RBF):

$$K(x, x_i) = \exp\left(-((x - x_i)/\sigma)^2\right).$$

The common kernel function is RBF kernel. In this paper, we choose complex Gaussian wavelet function as the kernel, and construct a kind of CGW-SVM.

3. FCM

For STLF with SVM, too small training data will cause increase of test error, and too many training data will cause the increase of training time. Considering the periodical feature of power loads, for a great deal of historical load data, FCM clustering algorithm is proposed to reduce training data. Thus the forecasting process can be speeded up and the accuracy of forecasting can be improved.

c clusters of $\{x_i\}_{k=1}^{N}$ with the standard FCM objective function is given by [9]

$$J_p(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} d_{ik}^{2} \tag{9}$$

where $d_{ik} = \|x_k - v_i\|$, and v_i are the prototype or center of the i^{th} cluster. The array $[u_{ik}] = U$ represents a partition matrix, and the parameter m determines the amount of fuzziness of the resulting classification.

FCM Algorithm:

1) Fix c, m and U^0. Then at step $l = 0$.

2) Calculate

$$v_i^{(l)} = \sum_{k=1}^{n}(u_{ik}^{(l)})^m x_k \Bigg/ \sum_{k=1}^{n}(u_{ik}^{(l)})^m .$$

3) Calculate

$$u_{ik}^{(l+1)} = 1\Bigg/\sum_{j=1}^{c}(d_{ik}/d_{jk})^{2/m-1} \ \forall i, \forall k .$$

4) For $\varepsilon > 0$, if

$$\left\|U^{(l+1)} - U^{(l)}\right\| \le \varepsilon ,$$

then stop. Otherwise, set $l = l + 1$, and return to 2).

4. Simulation example

In this paper, we use the proposed method compared with the commonly used RBF-SVM to show the forecasting performances. The experiment adopts load data of New South Wales, Australia from June 23, to July 22, 2006 [10]. We predict the load of July 23.

4.1. Data pretreatment

The load series $\{x(i), i = 1,2,..., N\}$ is normalized first. Supposing the maximum of $x(i)$ is $x_{\max}$ and the minimum is $x_{\min}$, then we get $\hat{x}(i)$ after normalization:

$$\hat{x}(i) = \frac{x(i) - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

Considering the chaotic characters of short-term load series, we reconstruct it with PSRT which based on Takens' embedding theory [11]. For the normalized load series, we can get the phase space point

$$X(k) = [\hat{x}(k), \hat{x}(k+\tau), ..., \hat{x}(k+(d-1)\tau)] \tag{11}$$

where $k = 1,2,..., M$, M is the point number in reconstructed phase space, $M = N - (d-1)\tau$, d is the embedding dimension, and τ is the time delay. With the load series, we get $d = 6$ and $\tau = 3$ using the method proposed by [11]. Since d is an even number, the CGW-SVM can be used for the STLF.

4.2. Parameter selection

For CGW-SVM, there are the parameters a, C, ε, and for RBF-SVM, there are σ, C, ε. The common method for parameter selection is based on experience. In recent years, the genetic algorithm has been proposed for the purpose. In this paper, we use the cloud theory-based genetic algorithm proposed by [12]. It is based on both the idea of genetic algorithm and the properties of randomness and stable tendency of a normal cloud model. We can get the optimized parameters with this algorithm.

4.3. Forecasting results

With FCM, the sample data of phase space are classified into 400 clusters as the training data of CGW-SVM. The Actual load and forecasting load with CGW-SVM and RBF-SVM are shown in figure 1. Figure 2 presents the error results corresponding to figure 1. Table 1 shows the numerical comparisons of CGW-SVM and RBF-SVM, including the mean absolute percentage error E_{mape}, the maximum relative error $E_{\max}$, the training time T, and the number of support vector N_{sv}. The simulation programs run on a 2.93 GHz Pentium Ⅳ processor under Windows XP and Matlab 7.4.0 compiler. We use LOQO optimizer for the training of the SVM which consists on solving quadratic programming problem.

From the figures and the table, it can be seen that the two SVM methods are efficient, but the proposed method is a better promising one. It has the lower forecasting error of E_{mape} and $E_{\max}$. Furthermore, with FCM clustering, the elapsed time is only 592 seconds, and the number of support vector is only 229.

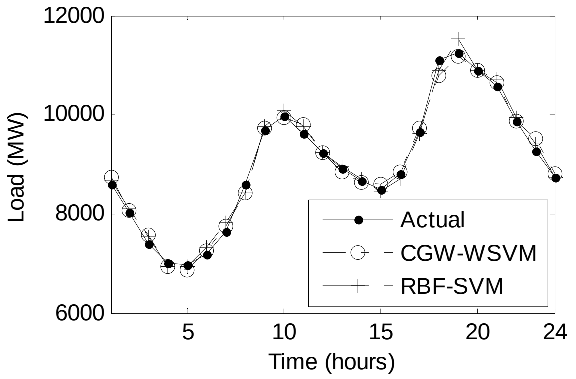

Fig.1 Actual load and load forecasting results

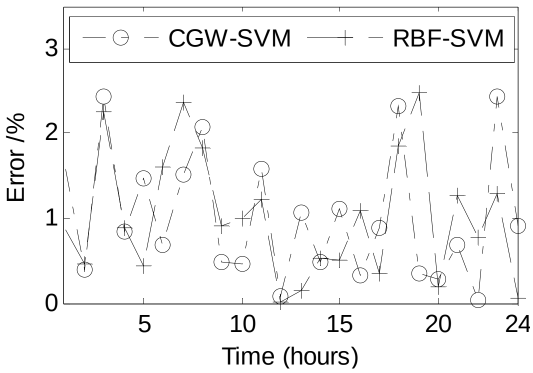

Fig.2 Error results corresponding to Fig.1

Models	E_{mape} /%	E_{max} /%	T /s	N_{sv}
CGW-SVM	1.06	2.44	592	229
RBF-SVM	1.18	2.49	17315	847

Tab.1 Numerical comparisons of CGW-SVM and RBF-SVM

5. Conclusions

As a novel machine learning method, SVM is powerful for the STLF. In this paper, a new kernel function of SVM, complex Gaussian kernel, was proposed, and it was proved that the function satisfies the translation-invariant kernel condition. Considering its chaotic characteristics, the short-term load series was reconstructed based on PSRT. The FCM clustering method was also adopted to reduce sample data. The experiment results show that the proposed method can improve the forecasting accuracy and speed up the forecasting processing.

The proposed model is a single-step method. Our forthcoming research is to propose adaptive multi-step forecasting method.

References

[1] S. P. Mchanos, A. C. Tsakoums, P. Fessas, *et al*, Short-term load forecasting using a chaotic time series, *International Symposium on Signals, Circuits and Systems*, pp. 437-440, 2003.

[2] I. Moghram, S. Rahman, Analysis and evaluation of five short-term load forecasting techniques, *IEEE Trans. Power Syst*, pp. 1484-1491, 1989.

[3] AK. Topalli, Erkmen I, A hybrid learning for neural networks applied to short term load forecasting, *Neurocomputing*, 51: 495-500, 2003.

[4] C.W. Jiang, L. Tao, Forecasting method study on chaotic load series with high embedded dimension, *Energy Conversion and Management*, 46: 667-676, 2005.

[5] M.G. Zhang, short-term load forecasting based on support vector machines regression, *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, pp. 4310-4314, Guangzhou, 2005.

[6] V. N Vapnik, Statistical learning theory, New York: Springer-Verlag, 2000.

[7] C.J.C Burges, Geometry and invariance in kernel based methods, *Advance in Kernel Methods-support Vector Learning*. pp. 89-116, 1999.

[8] L. Zhang, W. Zhou, L. Jiao, Wavelet support vector machine, *IEEE Transactions on systems, man, and cybernetics-Part B: cybernetics*, 34: 34-39, 2004.

[9] J. C. Bezdek and S. K. Pal, Fuzzy Models for Pattern Recognition, *NJ: IEEE Press*, Piscataway, 1991.

[10] http://www.nemmco.com.au.

[11] F. Takens, Detecting Strange Attractors in Turbulence, *Lecture Note in Mathematics*, 898: 366-381, 1981.

[12] C. Dai, Y. Zhu, W. Chen, Cloud Theory-Based Genetic Algorithm, *Journal of Southwest Jiaotong University*, 41: 729-732, 2006.