Available online at www.sciencedirect.com

SciVerse ScienceDirect

Energy Procedia 12 (2011) 387 – 393

Energy Procedia

ICSGCE 2011: 27-30 September 2011, Chengdu, China

Short-Term Load Forecasting Based on Gaussian Wavelet SVM

Yongkang Zheng[a*], Li Zhu[b], Xing Zou[b]

[a] *The Sichuan Electric Power Research Institute, Chengdu, Sichuan, China*
[b] *Chengdu Electric Power Bureau, Sichuan, China*

Abstract

Gaussian wavelet support vector machine (SVM) is constructed for short-term load forecasting (STLF). It is proved that the p^{th}-derivative Gaussian wavelet is an admissible translation-invariant kernel function of SVM when p is an even number. The Gaussian wavelet SVM is constructed with wavelet kernel function, and improved stochastic focusing search (SFS) algorithm is used to optimize the parameters of SVM and its kernel function. The experiments of STLF are conducted using the proposed SVM, the conventional Gaussian SVM and Morlet wavelet SVM respectively. The comparison shows that the proposed method is efficient and superior, and has some application value in STLF.

Selection and/or peer-review under responsibility of University of Electronic Science and Technology of China (UESTC).

Keywords: Gaussian wavelet support vector machine; short-term load forecasting; Gaussian wavelet kernel; stochastic focusing search.

1. Introduction

STLF 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, precise forecasting of STLF is still a difficult task [1].

Many techniques for STLF have been tested with different degrees of success. Traditional techniques include regression models, time-series 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

* Corresponding author. Tel.: +86-28-87082140.
E-mail address: zyk555@163.com.

1876-6102 © 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of University of Electronic Science and Technology of China (UESTC). Open access under CC BY-NC-ND license.
doi:10.1016/j.egypro.2011.10.052

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.

SVM is a small-sample theory firmly grounded on the framework of SLT [4]. 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.

In this paper, Gaussian wavelet can satisfy the kernel function condition is proved, and a kind of Gaussian wavelet kernel SVM is build. The short-term load series is reconstructed with phase space reconstruction theory (PSRT) [5], and the vector of phase space is used as the inputs of Gaussian wavelet SVM. Aimed at solving the problems of choosing parameters of Gaussian wavelet SVM and its kernel function, an improved SFS optimization algorithm is proposed [6]. 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. Gaussian wavelet 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 1: A translation-invariant kernel $k(x, x') = k(x - x')$ is an admissible SVM kernel if and only if the Fourier transform

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

Is non-negative.

2.1. Gaussian wavelet kernel

Gaussian wavelet is constructed with the p^{th}-derivative of Gaussian function:

$$f(\boldsymbol{x}) = (-1)^{p/2} \left[\exp(-\boldsymbol{x}^2)\right]^{(p)} \tag{3}$$

The SVM kernel of this Gaussian wavelet is

$$k(\boldsymbol{x}, \boldsymbol{x}') = \prod_{i=1}^{d} (-1)^{p/2} \left[\exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_i'\|^2}{a^2}\right)\right]^{(p)} \tag{4}$$

Theorem 2: Formula (4) It is an admissible SVM kernel function when p is an even number.

It is proved as follows:

Proof: For all x ,

$$k(\boldsymbol{x})=\prod_{i=1}^{d}\psi(\frac{\boldsymbol{x}_i}{a})=\prod_{i=1}^{d}(-1)^{p/2}\left[\exp\left(-\frac{\|\boldsymbol{x}_i\|^2}{a^2}\right)\right]^{(p)} \tag{5}$$

Substituting (5) into (1), we can calculate the integral term

$$\int_{R^d}\exp(-\mathrm{j}(\omega x))k(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

$$=\int_{R^d}\exp(-\mathrm{j}(\omega\boldsymbol{x}))\prod_{i=1}^{d}(-1)^{p/2}\left[\exp\left(-\frac{\|\boldsymbol{x}_i\|^2}{a^2}\right)\right]^{(p)}\mathrm{d}\boldsymbol{x}=\prod_{i=1}^{d}\int_{-\infty}^{+\infty}\exp(-\mathrm{j}(\omega\boldsymbol{x}_i))(-1)^{p/2}\left[\exp\left(-\frac{\|\boldsymbol{x}_i\|^2}{a^2}\right)\right]^{(p)}\mathrm{d}\boldsymbol{x}$$

$$=\prod_{i=1}^{d}(-1)^{p/2}(\mathrm{j}\omega)^p\times\int_{-\infty}^{+\infty}\exp(-\mathrm{j}(\omega\boldsymbol{x}_i))\exp\left(-\frac{\|\boldsymbol{x}_i\|^2}{a^2}\right)\mathrm{d}\boldsymbol{x}=\prod_{i=1}^{d}(-1)^{p/2}(\mathrm{j}\omega)^p\times|a|\sqrt{\pi}\exp(-\frac{a^2\omega^2}{4}) \tag{6}$$

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

$$F[k(\omega)]$$
$$=(2\pi)^{-d/2}\prod_{i=1}^{d}(-1)^{p/2}(\mathrm{j}\omega)^p\times|a|\sqrt{\pi}\exp(-\frac{a^2\omega^2}{4}) \tag{7}$$

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

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:

$$\begin{aligned}&\min_{w,b,\xi}\ \frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}(\xi_i+\xi_i^*)\\ &s.t.\ \ |y_i-<w\cdot\Phi(x)>-b|\le\varepsilon+\xi_i,\\ &\qquad \xi_i\ge 0,\xi_i^*\ge 0,i=1,2,\cdots n\end{aligned} \tag{8}$$

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:

$$f(x)=\sum_{x\in SV}(\alpha_i-\alpha_i^*)K(x_i,x)+b \tag{9}$$

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

Linear: $K(x,x_i)=<x\cdot x_i>$.

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

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

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

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

3. Improved SFS

The SVM and its kernel function have several parameters. For solving the problems of choosing these parameters, we use SFS optimization algorithm and proposes an improved method.

The SFS models the human randomized searching behaviors based on their memory, experience, uncertainty reasoning and communication with each other. In searching, people will focus on the best position so far and step forward a neighborhood of it stochastically. The SFS's performance was studied by using a challenging set of typically complex functions with comparison of differential evolution (DE) and three modified particle swarm optimization (PSO) algorithms, and the simulation results showed that SFS is competitive to solve most parts of the benchmark problems [6].

In SFS, each particle is treated as a point in the *M*-dimensional problem space. Let *s* denote the population size. Each individual particle *i* ($1\le i\le s$) has the following attributes: the current position $\vec{x}_i(t)=[x_{i1},x_{i2},\cdots,x_{ij},\cdots,x_{iM}]$ (*j* stands for the *j*th dimension of variable $\vec{x}_i$, and *M* is the dimension size) for each iteration *t* in the search space, and the global best position $\vec{g}_{best}$ so far. The rate of the position change (velocity) for particle *i* is represented as $\vec{v}_i(t)=[v_{i1},v_{i2},\cdots,v_{ij},\cdots,v_{iM}]$. The optimization problems to be solved assumed as minimization problems, the particles are manipulated according to the equations as [6]. From this equations, we can see that all particles are searching around a $\vec{g}_{best}$. Though it adopted neighbor space R_t to avoid a local optimum, the choosing of parameter *w* for R_t is difficulty. In this paper, we improved the SFS algorithm to solve this problem. The improved SFS is as the following equations:

$$\vec{v}_i(t)=\begin{cases} Rand()\times(R_{ti}-\vec{x}_i(t-1)) \\ \qquad \text{if } fun(\vec{x}_i(t-1)) \ \ge \ fun(\vec{x}_i(t-2)) \\ \vec{v}_i(t-1) \\ \qquad \text{if } fun(\vec{x}_i(t-1)) \ < \ fun(\vec{x}_i(t-2)) \end{cases} \tag{10}$$

$$\vec{x}_i(t)=\vec{v}_i(t)+\vec{x}_i(t-1) \tag{11}$$

$$\vec{x}_i(t)=\begin{cases} \vec{x}_i(t) & \text{if } fun(\vec{x}_i(t)) \ < fun(\vec{x}_i(t-1)) \\ \vec{x}_i(t-1) & \text{if } fun(\vec{x}_i(t)) \ \ge fun(\vec{x}_i(t-1)) \end{cases} \tag{12}$$

where $fun(\vec{x}_i(t))$ is the objective function value of $\vec{x}_i(t)$, R_{ti} is a random selected point (position) in the neighbor space R_t of $\vec{p}_{best}(i)$. $\vec{p}_{best}(i)$ is defined as a better particle than $\vec{x}_i(t)$ when sort by fitness values from best to worst. R_t is defined as:

$$\left[\vec{p}_{best}(i)-\frac{w(\vec{p}_{best}(i)-\vec{x}_{\min})}{(\vec{x}_{\max}-\vec{x}_{\min})^{1-w}},\vec{p}_{best}(i)+\frac{w(\vec{x}_{\max}-\vec{p}_{best}(i))}{(\vec{x}_{\max}-\vec{x}_{\min})^{1-w}}\right] \tag{13}$$

where $\vec{x}_{\max}$ and $\vec{x}_{\min}$ are the search space borders. When *w* is linearly decreased from 1 to 0, R_t is deflated from the entire search space to point $\vec{p}_{best}(i)$. The larger R_t is, the wider the search scope is. That means the particle swarm has the potential to avoid a local optimum. On the contrary, the smaller R_t is, the narrower the search scope is. This characteristic helps the swarm to improving the global searching ability and avoiding the astringency of a local extremum.

w is defined as:

$$w = (\frac{G-t}{G})^{\delta} \tag{14}$$

where G is the maximum generation, δ is a positive number. It is indicated that w is decreased from 1 to 0 with the increasing of iteration *t*.

The pseudocode of the improved SFS is presented in Fig. 1.

```
begin
t←0;
generating s positions uniformly and randomly in search space;
evaluating each particle;
repeat
   t←t+1;
   evaluating subpopulations and sort by fitness values;
   find p_best(i) for every subpopulations;
   updating each particle's position using (10) and (11);
   evaluating each particle;
   evaluating x_i(t) use (12);
until the stop condition is satisfied
end.
```

Fig. 1. The pseudo code of the main algorithm

4. Simulation example

In this paper, we use the proposed method compared with the conventional Gaussian SVM and the Morlet wavelet SVM to show the forecasting performances. For the proposed method, we choose the 2^{nd} -derivative Gaussian wavelet SVM. The experiment adopts load data of New South Wales, Australia from June 23, to July 22, 2006 [9]. 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:

$$\hat{x}(i) = \frac{x(i) - x_{\min}}{x_{\max} - x_{\min}} \tag{15}$$

Considering the chaotic characters of short-term load series, we reconstruct it with PSRT which based on Takens' embedding theory [10]. For the normalized load series, we can get the phase space points:

$$X(k) = [\hat{x}(k), \hat{x}(k+\tau), ..., \hat{x}(k+(d-1)\tau)] \tag{16}$$

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 $d = 14$ and $\tau = 7$ using the method proposed by [11]. Then we can get the input vector of SVM as (17) and output as (18).

$$X(k) = [\hat{x}(k), \hat{x}(k+7), ..., \hat{x}(k+(14-1)\times 7)] \tag{17}$$

$$y(k) = \hat{x}(k+(14-1)\times 7+1) \tag{18}$$

4.2. Parameter selection

For the 2^{nd} -derivative Gaussian wavelet SVM, conventional Gaussian SVM and Morlet wavelet SVM, there are the parameters: a, C, ε . The common method for parameter selection is based on experience. In this paper, we use the improved SFS algorithm.

For a, C, ε , we set searching range as: $a \in [0,100]$, $C \in [10^2, 10^{10}]$, As a result of optimizing, for the 2^{nd} -derivative Gaussian wavelet SVM: $\varepsilon \in [0, 0.1]$. $a = 2.4$, $C = 4.9\times 10^8$, $\varepsilon = 2.1\times 10^{-4}$; for Gaussian SVM: $a = 23.7$, $C = 6.8\times 10^9$, $\varepsilon = 2.7\times 10^{-4}$; for Morlet wavelet SVM: $a = 26.3$, $C = 1.4\times 10^9$, $\varepsilon = 1.2\times 10^{-3}$.

4.3. Forecasting results

Table 1 shows the numerical comparisons of the 2^{nd} -derivative Gaussian wavelet SVM, Gaussian SVM and Morlet wavelet SVM, including the mean absolute percentage error E_{mape} and the maximum relative error $E_{\max}$. The Actual load and forecasting load with the 2^{nd} -derivative Gaussian wavelet SVM are shown in Fig. 2.

From the figures and the table, it can be seen that the three SVM methods are efficient, but the proposed 2^{nd} -derivative Gaussian wavelet SVM is a better promising one. It has the lowest forecasting error of E_{mape} and $E_{\max}$. The E_{mape} is only 0.95% and the $E_{\max}$ is only 2.42%. But for Morlet wavelet SVM, the E_{mape} is 0.98% and the $E_{\max}$ is 2.5%; for Gaussian SVM, the E_{mape} is 1.03% and the $E_{\max}$ is 2.93%.

Table 1. The load forecasting results

Result	Gaussian wavelet SVM	Morlet wavelet SVM	Gaussian SVM
Emape /%	0.95	0.98	1.03
Emax /%	2.42	2.5	2.93

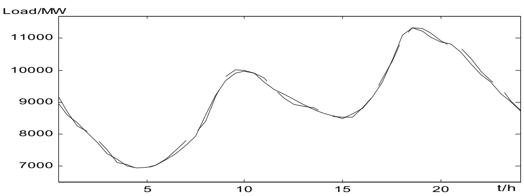

Fig. 2. Actual load (solid line) and load forecasting results with the 2^{nd} -derivative Gaussian wavelet SVM (dotted line)

5. Conclusion

As a novel machine learning method, SVM is powerful for the STLF. In this paper, a new kernel function of SVM, Gaussian wavelet 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. For choosing parameters of Gaussian wavelet SVM and its kernel function, an improved SFS algorithms was proposed. The experiment results show that the proposed method can improve the forecasting accuracy and speed up the forecasting processing.

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

References

[1] Changhao Xia, Jian Wang, and Karen McMenemy, "Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks," International Journal of Electrical Power & Energy Systems, Volume 32, Issue 7, September 2010, Pages 743-750.

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

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

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

[5] Minglun Cai, Feng Cai, Aiguo Shi, "Chaotic time series prediction based on local-region multi-steps forecasting model," Advances in Neural Networks - ISNN 2004: International Symposium on Neural Networks, Part II, Dalian, China, 2004, pp. 418-423.

[6] ZHENG YongKang, CHEN WeiRong, "Stochastic focusing search: A novel optimization algorithm for real-parameter optimization," Systems Engineering and Electronics, 2009, vol.20, no.4.

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

[8] Zhang Li, Zhou Wei-da, Jiao Li-cheng, "Wavelet support vector machine," IEEE Transactions on systems, man, and cybernetics-Part B: cybernetics, Vol. 34, pp. 34-39, 2004.

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

[10] F. Takens, "Detecting Strange Attractors in Turbulence, " Lecture Note in Mathematics, Vol. 898, pp. 366-381, 1981.

[11] Minglun Cai, Feng Cai, Aiguo Shi, "Chaotic time series prediction based on local-region multi-steps forecasting model," Advances in Neural Networks - ISNN 2004: International Symposium on Neural Networks, Part II, Dalian, China, 2004.8: 418-42