Liying WANG[1], Zheng LIU[2,3]

1 Hebei University of Engineering
2 School of Computer Science and Technology, Shandong University of Finance and Economics
3 Shandong Provincial Key Laboratory of Digital Media Technology

RVM with wavelet kernel combined with PSO for short-term load forecasting in electric power systems

Abstract*. This paper presents a new hybrid method for the short-term load forecasting in electric power systems based on particle swarm optimization (PSO) and relevance vector machine (RVM). In this method, we firstly develop a type of kernel as the kernel function of the RVM model, and then its parameter is optimized by the PSO, finally the established RVM forecast mode is applied to short-term load forecasting in electric power systems in a city. The simulation results show the parameter of the wavelet kernel is well optimized using the PSO, and the acquired RVM model is more sparse and can obtain higher forecast accuracy compared with the RVM model with Gaussian kernel, so the proposed method is effective for forecasting the short-term load in electric power systems.*

Streszczenie. *W artykule zaprezentowano nową hybrydową metode krótkoterminowego prognozowania obciążeń sieci energetycznej bazująca na algorytmie mrówkowym PSO i narzędzia RVM (relevance vector machine). W pierwszym etapie wyznaczane jest falkowe jądro (kernel) jako RVM co znacznie poprawia skuteczność algorytmu PSO. (**Hybrydowe połączenie funkcji PSO i RVM jako narzędzie do krótkoterminowego prognozowania obciążenia sieci energetycznej**)*

Keywords: Load forecasting, Wavelet kernel, PSO, RVM
Słowa kluczowe: prognozowanie obciążeń, jądro falki, algorytm mrówkowy PSO**.**

Introduction

Short-term load forecasting has become increasingly important with the rIse of the competitive energy market and become one of major area in electrical engineering in recent years, because it can positively contribute to electric network's reliable and economic development which can be applied to power system for unit commitment, optimal dispatch, maintenance, dynamic configure of power systems and so on [1]. Through analysizing the historical data, short-term load forecasting in electrical power systems studies on the internal relationship in electrical power load and the law of its own development and change, and then estimates and predicts the electrical power demand in the future, which is based on its change in load, meteorological information and the economy law. So the load forecasting is very significant to the operational planning of electric power systems, which is effective to make full use of electrical energy and relieve the contradictions between supply and demand, if the load forecast is inaccurate,. the contradictions are becoming increasingly acute between supply and demand. [2]. Because the load has complex and nonlinear relationships with some uncertain factors such as weather information, temperature conditions, day sort and so on, so how to find a effective method to forecast load is burning task. As a nonlinear model, the Support Vector Machine (SVM) proposes some significant shortcomings, it is wasteful both of data and computation, and its output is a point estimate rather than a conditional distribution, moreover, owing to the requirement of Mercer's conditions, the SVM do not allow for the free use of an arbitrary kernel function [3, 4, 5].

RVM [6] is a probabilistic sparse kernel model identical in functional form to SVM, where a Bayesian approach to learning is adopted, introducing a prior over the weights governed by a set of hyperparameters, the RVM suffers from none fo the above disadvantages and examples demonstrate that for comparable generalisation performance, it can make us improve its kernel function more flexibly to establish a better RVM model with fewer training samples. In this paper, a new thought which can improve the accuracy of load forecasting is presented, the wavelet kernel is used as the kernel function of the RVM model, whose parameter is optimized by the PSO. Then the proposed method is applied to short-term load forecasting in electric power systems in a city, the simulation results illustrate the superiority of the proposed method over its counterpart.

Particle swarm optimization

PSO is a method for performing numerical optimization without explicit knowledge of the gradient of the problem to be optimized [7]. It is simplified and first intended for simulating social behaviour. PSO is basically developed through simulation of bird flocking in two-dimension space. This modification can be represented by the concept of velocity. Velocity of each agent can be modified by the following equation:

$$v_i^{k+1} = wv_i^k + c_1 rand \times (pbest_i - x_i^k) + c_2 rand \times (gbest - x_i^k) \quad (1)$$

where, v_i^k : velocity of agent i at iteration k; w : inertia weight; c_j : weight factor; *rand* : random number between 0 and 1; x_i^k: current position of agent i at iteration k; $pbest_i$: *pbest* of agent *i*.

Using the above equation, a certain velocity, which gradually gets close to *pbest* and *gbest* can be calculated. The current position (searching point in the solution space) can be modified by the following equation:

$$x_i^{k+1} = x_i^k + v_i^{k+1} . \quad (2)$$

For the algorithm, the inertia weight is very important in balancing the exploration and exploitation, to keep its availability and flexibility, the linear decreasing scheme is used, where an initially large weight (ω_{max}=0.9) is linearly decreased to the small one (ω_{min}=0.4), the linear decreasing inertia weight is defined as follows:

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{Iteration_{\max}} \times Iteration \quad (3)$$

where $Iter_{max}$ is the maximum iteration number and the Iteration is the current iteration number.

Relevance vector machine

Tipping [6,8] proposed the Relevance Vector Machine in 2000. For a regression problem, given a training dataset $\{x_n, t_n\}_{n=1}^N$,

(4) $t_n = y(x_n, \mathbf{w}) + \varepsilon_n$ $\mathbf{t} = \mathbf{y} + \boldsymbol{\varepsilon}$

Where the errors $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)$ are modeled probabilistically as independent zero-mean Gaussian, with variance σ^2, so $p(\varepsilon) = \prod_{n=1}^{N} N(\varepsilon_n | 0, \sigma^2)$, $\mathbf{w} = (w_1 ....., w_M)$ is the parameter vector and $y(x_n, \mathbf{w})$ can be expressed as a linearly weighted sum of some basis functions $\phi(\mathbf{x})$:

(5) $$y(\mathbf{x}, \mathbf{w}) = \sum_{m=1}^{M} w_m \phi_m(\mathbf{x}) + w_0 \quad \mathbf{y} = \boldsymbol{\Phi}\mathbf{w}$$

The likelihood of the complete dataset can be written as

(6) $$p(\mathbf{t} | \mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \| \mathbf{t} - \Phi \mathbf{w} \|^2\right\}$$

RVM adopts a Bayesian perspective, and constrains the parameter by defining an explicit prior probability distribution over them, the posterior distribution over the weights is thus given by:

(7) $$p(\mathbf{w} | \boldsymbol{\alpha}) = (2\pi)^{-M/2} \prod_{m=1}^{M} \alpha_m^{1/2} \exp(-\alpha_m w_m^2 / 2)$$

Given $\boldsymbol{\alpha}$, the osteriori parameter distribution conditioned on the data is given by combining the likelihood and prior within Bayes's rules:

(8) $$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}) / p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)$$

is a Gaussian distribution $N(\mu, \Sigma)$ with

(9) $$\mu = \sigma^{-2} \Sigma \Phi^T \mathbf{t} \quad \Sigma = (A + \sigma^{-2} \Phi^T | \Phi)^{-1}$$

Sparse Bayesian learning can be formulated as a type–II maximum likelihood procedure; that is, a most probable point estimate α_{MP} may be found throughout the maximization of the marginal likelihood with respect to the hyperparameters a_i

(10) $$L(\boldsymbol{\alpha}) = -\frac{1}{2}\left[N \log 2\pi + \log |\mathbf{C}| + \mathbf{t}^T \mathbf{C}^{-1} \mathbf{t}\right],$$

The predictive distribution for a new datum $\mathbf{x}_*$ is defined as follows:

(11) $$p(t_* | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) = \int p(t_* | \mathbf{w}, \sigma_{MP}^2) p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) d\mathbf{w}$$

which is easily computed due to the fact that both integrated terms are Gaussian, resulting in a Gaussian too.

(12) $$p(t_* | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) = N(t_* | y_*, \sigma_*^2),$$

with $\sigma_*^2 = \sigma_{MP}^2 + \phi(\mathbf{x}_*)^T \Sigma \phi(\mathbf{x}_*)$

RVM with Wavelet Kernel Combined With PSO

The choice of kernel is importment to the performance of the RVM, Gaussian kernel function outperforms others when the lack of a prior knowledge in the learning process, which is as follows:

(13) $$K(\boldsymbol{x}, \boldsymbol{y}) = \exp(\frac{(\boldsymbol{x} - \boldsymbol{y})^2}{2\sigma^2})$$

where σ is a width factor which shapes the width of the Gaussian kernel function, so we often use Gaussian kernel function as the kernel of RVM, but Gaussian function is a local kernel function, the map characteristic of the Gaussian function for Y_i =0.3 is shown in Fig. 1 according to equation (13), it can be seen that there exists larger kernel function value only near the test point Y_i =0.3, and the farther the point is from the test point, the smaller its kernel function value becomes, finally it approaches zero rapidly. So, the Gaussian kernel function only has an effect on samples near the neighborhood of the test point instead of that far from the test point [9].

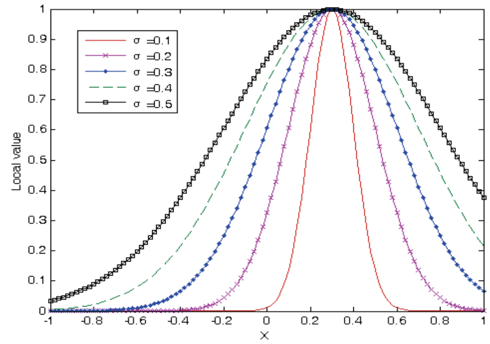

Fig.1. The map characteristic of Gaussian kernel function for Y_i =0.3

With the development of wavelet theory, various types of wavelet functions emerge as an alternative to approximate any function, in general, let $\varphi(x)$ be a mother wavelet, a and b are the dilation and translation respectively, $a,b \in R$, $x,y \in R^d$, d is the number of dimensions of the input space, the wavelet kernel function is defined as:

(14) $$K(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{d} \varphi(\frac{x_i - b_i}{a}) \varphi(\frac{y_i - b_i}{a})$$

Especially, the translation-invariant wavelet kernel function is as follows:

(15) $$K(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{d} \varphi(\frac{x_i - y_i}{a})$$

Compared with other kernel functions, wavelet kernel function has better approximation ability on complex input, in this paper, we use Morlet wavelet kernel as the kernel function of RVM, which is as follows:

(16) $$K(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{d} (\cos(1.75 \times (\frac{x_i - y_i}{a})) \times \exp(-\frac{\| x_i - y_i \|^2}{2a^2}))$$

The map characteristic of Morlet wavelet kernel for Y_i =0.3 is shown in Fig. 2 according to equation (16), it is clear that it has both the local characteristic and the global characteristic.

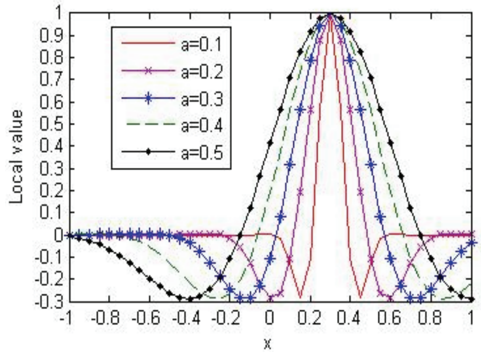

Fig. 2. The map characteristic of Morlet wavelet kernel function for Y_i =0.3

The determination of the kernel parameters need to meets a specific criterion. The following mean relative error is used as the performance criterion; meanwhile, it is also used as the fitness evaluation function of the PSO which is given by

(17) $$J_{fitness} = \frac{1}{N} \sum_{i=1}^{N} \frac{| y(i) - y^*(i) |}{y(i)}$$

RVM has excellent characteristics of nonlinear approximation and has been widely used in many fields, which has many advantages many other similar algorithms have not. To greatly improve the performance of the RVM, it is important for us to define the kernel parameter, in this section, we describe the proposed hybrid PSO-RVM method for load forecasting, PSO algorithm can gain the optimal parameter of RVM that can lead to best prediction accuracy in present circumstances, and it uses the fittest particles to contribute to the next generation of candidate particles, the flowchart of the RVM model based on PSO is shown in Fig. 3 and the procedure is summarized as follows:

Step 1: Input historical load data.

Step 2: Analyize the influencing factors and select input samples.

Step 3: Normalize the imput samples and range form 0 to 1, then form the training samples..

Step 4: Establish the RVM model and initialize the kernel parameter.

Step 5: Initialize parameters *m, w,* c_1*,* c_2*, θ,* where:*m*: number of population; *w*: inertia weight; c_i: weight factor; *θ*: parameter of identification (coefficient of nonlinear rectification equation); the velocity and position of each particle are initialized randomly.

Step 6: Each particle's velocity is updated according to (1) and each particle's position is updated according to (2).

Step 7: Each particle's fitness is evaluated. The Mean square Error (MSE) is used as the fitness function to guide the particle population for searching for the optimum solution, we use all the training samples to calculate the error for each particle, so as to generate the training error of particles with training samples.

Step 8: The personal best position *pbest* and and the global best position *gbest* are updated.

Step 9: If the maximum of the iteration is not achieved or the optimum solution is acquired, then return to Step 6.

Step10: Establish the best RVM forecasting model using the optimum solution, and implement the load forecasting.

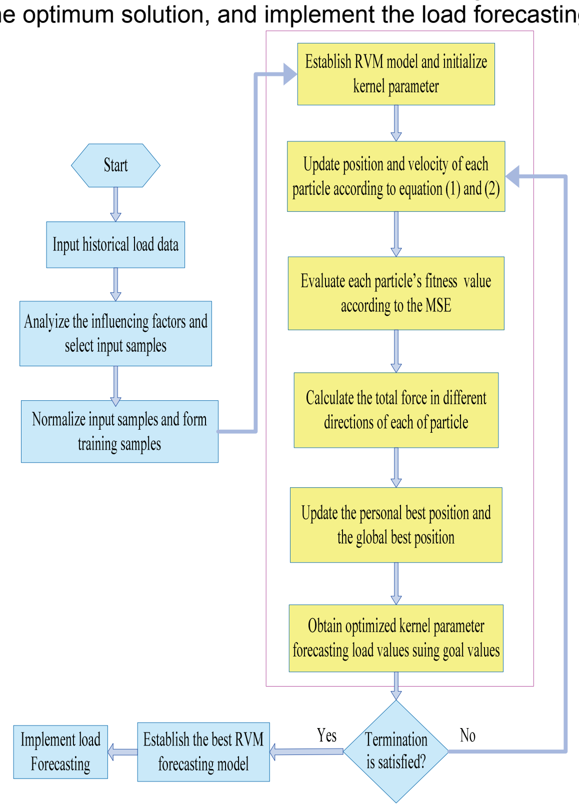

Fig.3. The flowchart of the RVM based on PSO

Experiment results and analysis

The results of load forecasting of power systems are influenced by such factors as weather information, day type, day temperature and related events, these factors should be taken into account which are included in one eigenvector at the day's value [10]. In order to verify the affectivity of the proposed method, a set of historical data on the load forecasting from a city in Hebei province, China, is used. For the comparison, the proposed method in this paper and its counterpart are also used to forecast the short-term load. In the dataset, we choose load values every one hour, the power load data from 0:00 at 08/01/2008 to 0:00 at 08/04/2008 is used as training samples (96 samples), and the power load data from 0:00 at 08/04/2008 to 0:00 at 08/05/2008 is used as test samples (24 samples). The involved parameters of PSO algorithm we used are follows: the population sizes of the particles is set at 50, the maximum iteration is 100, the acceleration factors c_1 and c_2 are both 2.0, a linear decaying inertia weight *w* starting at 0.9 and ending at 0.2 is used.

For the comparison, the RVM models with the Gaussian and the Morlet wavelet kernel based on the PSO are all used to this experiment, for the RVM model with the Morlet wavelet kernnel, after the model is optimized by the PSO, when the iteration is over the $J_{fitness}$ achieves 0.00962 and 17 relevance vectors (RVs) are obtained when the optimal kernel parameters are *a*=2.3148; for the RVM model with Gaussian kernel, when the iteration is over the $J_{fitness}$ achieves 0.01251 and and 25 relevance vectors are obtained when the optimal kernel parameter is σ= 38.215, finally the optimal kernel parameters are put into the RVM model to establish the load forecasting models.

To exhibit the influence of each training sample to the final solution, the weight of each train sample is used as the important degree which contributes to the final model according to the maximum likelihood II method [8], Table 1 summarizes the performance comparison for the RVM model corresponding to the two kernels, it can see that the influence of each training sample for the two RVM model, for the RVM model with the Morlet wavelet kernel, the number of relevance vector is 17, for the RVM model with the Gaussian kernel, the number of relevance vector is 25, it shows the former is more sparse then the latter, which requiring less than a quarter of training samples can establish a good RVM model, and the training samples used as establishing the RVM model with the Morlet wavelet kernel, are used to the RVM model with the Gaussian kernel too, it can be seen that not only the numbers of RV for the two RVM models, but only the MSE of the test samples are listed, there are a great difference between the two model, to establish a good RVM model, the RVM model with the i Morlet wavelet kernel requires only 18% of the training samples, is much less than 26% of the training samples required by the RVM model with the Gaussian kernel implying the great sparsity of the Morlet wavelet kernel, and then the MSE of training and test samples for the RVM model with the Morlet wavelet kernel are smaller than those for the RVM model with the Gaussian kernel, this is because the dataset could contain attributes of very different natures, and throught the Morlet wavelet kernel, we can obtain more appropriate parameters that give the right weights to the right characteristic, moreover, the Morlet wavelet kernel exhibits that it has more flexibility to reduce the number of relevance vector.

Table 1 Performance comparison between two RVM model

Kernel method	Number of RV	Percent of RV	MSE of training set
Gaussian kernel	25	26%	1.251%
Morlet wavelet kernel	17	18%	0.962%

We use the established RVM models to forecast the power load data from 0:00 at 08/04/2008 to 0:00 at 08/05/2008 which consists of 24 test samples, the comparison results of load forecasting on test samples between the two methods are obtained, it can be noticed that the relative error of most of test samples for the RVM withMorlet wavelet kernel method is less than that of the RVM with Gaussian kernel method, the average relative error for the former is only 1.05%, and the average relative error for the latter is 1.52%, so the proposed method in this paper improves the accuracy of the short-term load forecasting than it counterpart.

To make a clearer comparison of forecasting performance between the two methods, their forecasting curves are depicted in Fig.3, it can be seen that the acquired load values using the proposed method is mostly closer to the actual load values, whereas the forecasting load values using the RVM with Gaussian kernel is apart from the actual load. The comparison of forecasting relative error curve between the two methods is shown in Fig. 4, the relative error of forecasting results obtained by the proposed method has small range ability with a maximum error of 2.43%, most of errors are less than 2%; for the RVM with Gaussian kernel, the relative error of forecasting results obtained has relative big range ability with a maximum error of 3.01%, and many errors are more than 2%. It can be seen from the above analysis that the proposed method performs better forecasting effectiveness than its counterpart.

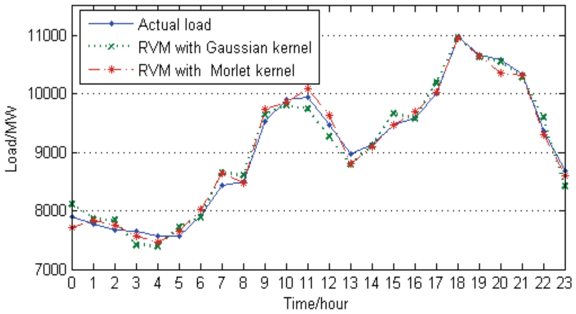

Fig. 3. Comparison of forecasting curve between the two methods

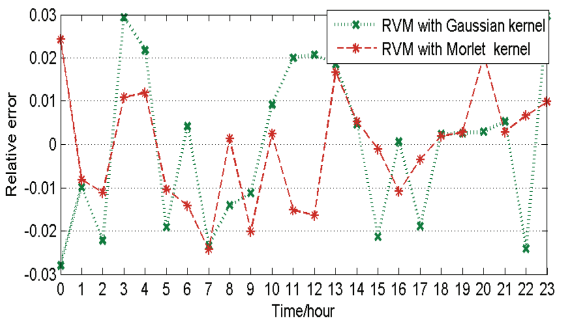

Fig. 4 Comparison of forecasting relative error between the two methods

Conclusions

The proper load forecasting model can save a huge amount of operational cost as well as do efficient generation and distribution planning. In this paper, a new thought which can improve the accuracy of load forecasting is presented, it believes the key which can improve the accuracy is to how to establish a RVM model and then optimize its kernel parameter, based on this, a hybrid method combined PSO with RVM model with Morlet wavelet kernel is presented, which is applied to the short-term load forecasting in electric power systems. The simulation results show the parameter of the Morlet wavelet kernel is well optimized using the PSO, and the acquired RVM model can precisely forecast the load and the less training samples can establish the accurate RVM model than its counterpart, implying its sparsity and efficiency for load forecast in electric power systems.

Acknowledgment

This work is supported by the Science and Technology Research Project of University of Hebei Province No. 2011138, and the Natural Science Foundation of Hebei Province of China No. E2010001026.

REFERENCES

[1] Md. Khairul Hasan†, Mohammad Asif A Khan, Suman Ahmmed, Ahmed Yousuf Saber. An Efficient Hybrid Model to Load Forecasting. International Journal of Computer Science and Network Security, (2010)10:8, 61-68

[2] Yongli Wang, Dongxiao Niu, Xiaoyong Ma. Optimizing of SVM with Hybrid PSO and Genetic Algorithm in Power Load Forecasting. Journal of Networks, VOL. 5, NO. 10, (2010)5:10,1192-1200

[3] Belousov AI, Versakov SA, Von Frese J. A flexible classification approach with optimal generalization performance: support vector machines. Chemon. Intell. Lab. Syst. (2002)64, 15–25

[4] Belousov AI, Versakov SA, Von Frese J. Application aspects of supports vector machines. J. Chemom., (2002)16, 482–489

[5] Abdulaziz Alshareef. Next 24-Hours Load Forecasting for the Western Area of Saudi Arabia Using Artificial Neural Network and Particle Swarm Optimization. Journal of Engineering and Computer Sciences, (2010)3:2, 97-117

[6] Tipping ME. The relevance vector machine. Advances in Information Processing System, (2000)2, 652-658

[7] Kennedy, J., Eberhart, R.C. Particle swarm optimization. Proc. IEEE Int'l. Conf. on Neural Networks, (1995), 1942-1948

[8] Noslen Hernández, Isneri Talavera, Angel Dago, Rolando J. Biscay, Marcia M. Castro Ferreira, Diana Porro. Relevance vector machines for multivariate calibration purposes. Chemometrics, (2008)22, 686–694

[9] YANG Liu1, ZHANG Lei, ZHANG Shaoxun, LIU Jianwei. Comparison Research of Single Kernel and Multi-kernel Relevance Vector Machine, Computer Engineering, (2010)36:12,195-197

[10] Pan Duan, Kaigui Xie, Tingting Guo,Xiaogang Huang. Short-Term Load Forecasting for Electric Power Systems Using the PSO-SVR and FCM Clustering Techniques. Energies, (2011) 4, 173-184

***Authors**: Dr Liying Wang, College of Hydropower, Hebei University of Engineering, Handan 056021, China, E-mail: 2000wangly@163.com, Dr Zheng Liu, School of Computer Science and Technology, Shandong University of Finance and Economics, Shandong Provincial Key Laboratory of Digital Media* Technology, Ji'nan 250014, China, E-mail: *Liuzheng48@gmail.com.*