http://dx.doi.org/10.14257/ijmue.2015.10.6.31

A Hybrid Model for Short-Term Load Forecasting Based on Non-Parametric Error Correction

Kun Yang[a,] and Shuang Liu[a, *]

College of Quality and Technology Supervision, Hebei University, Baoding, 071002, China
liushuang99@hotmail.com

Abstract

In this paper, we presented the performance of forecasting model and error correction will affect the accuracy of short-term load forecasting. Least squares support vector machines (LS-SVM) based on improved particle swarm optimization is selected as load forecasting model. Forecasting accuracy and generalization performance of LS-SVM depend on selection of its parameters greatly. Adaptive particle swarm optimization (APSO) based on fitness function was put forward to optimize the kernel parameter σ and regularization parameter γ of LS-SVM. Based on the optimized forecasting model, non-parametric error correction model is also presented by iterative method. The error forecasted by non-parametric model was used to update the forecasted load so as to improve the forecasting accuracy. Load data selected from some area in South China as training and forecasting data is used to analyze. Case study illustrates that the proposed forecasting model (NP-APSO-SVM) has more generalized performance and better forecasting accuracy compared with the method of standard SVM.

Keywords: *Least square support vector machines, improved particle swarm optimization, non-parametric model, error correction*

1. Introduction

Load forecasting is the premise and foundation of planning decisions and economic operation in power system. Short-term load forecasting is susceptible for a wide variety of facts, such as climate conditions and previous load demand data. Intelligent algorithm has been applied to short-term load forecasting widely. SVM is a promising method for regression proposed by Vapnik based on statistical learning theory and structural risk minimization [1], which is closely related to the selection of parameter σ and C. As a good solution to problems of the small sample, nonlinearity, high dimension and local optimization, SVM has been successfully extended to the pattern recognition, information fusion, time series forecasting and other fields [2-3].

In practice, different parameters would have a great effect on generalization ability and forecast accuracy of SVM. There is not a fixed method of choosing the optimal parameters in SVM, and Cross-validation is the most widely used method. However, this kind of method generally has considerable blindness and randomness, which is lack of appropriate theoretical basis and fully depends on their experience, so it is difficult to find the optimal points. These shortcomings limit the application of support vector machines model to some extent. To further improve the performance of SVM, many optimization methods have been used for parameter selection [4-5]. HUO Ming, LUO Dian-sheng *et al*. [6] set an objective function of the combination optimization problem and used improved mutative scale chaos optimization algorithm to search global optimal value. Thereby the optimal parameters combination was obtained by chaos optimization method. Ant colony algorithm can also be used to parameters selection. Qi Liang [7] applied ant colony algorithm to search the value of optimal objective function with the fine performance of robustness and distributed computing. The simulations results show that ACA was an effecting method for selecting parameters of SVM, and could obtain the performance for the function approximation. Moreover, Niu Dong xiao Liu Da *et al*. [8] proposed another optimization method that using genetic algorithm to select parameters in the SVM models automatically.

Some other parameters selection methods have also been presented in many fields. ZHAO Ying, LIU Hong-xing [9] proposed A new criterion which was called the sum of the squared derivatives. Compared with the already well-known hyper parameters optimization method such as k-fold cross-validation method or Radius/Margin Bound method, the class-separating hyper-surface designed based on this criterion could ‘leg and leg’ the whole original input space for all the samples, thus it supports the structural risk minimization principle better. With the leave one out (LOO) prediction error on the entire training sample being the object of optimization, TAO Shao-hui, CHEN De-zhao [10] put forward the gradient-based optimal algorithm to find the optimal parameters of LSSVM based on the fast (LOO) method selecting hyper parameters of LSSVM. XIE Hong, WEI Jiang-ping *et al*. [11] presented a new approach of selecting learning parameters based on analysis of model structure, which was an optimizing algorithm of Gauss kernel parameter based on local optimal searching direction.

By the above analysis, different parameters should be selected to reduce the modeling error of SVM and testing error for different SVM models. At present, a variety of improved particle swarm optimization algorithms are used as methods of training the parameters of SVM [12-14]. A new kind of adaptive PSO is put forward to optimize the parameters of SVM in this paper. Adaptive inertia weight was adopted in the improved algorithm and the particles studies not only from itself and the best one but also from the mean value of some other particles.

Meanwhile, analysis of forecasting errors will have influence on the improvement of forecasting accuracy to some extent. The forecasting model only considered the main factors that affected load changes, and many secondary factors were ignored. Actually, the error is an objective reflection of the impact of these secondary factors. So the error will be a certain trend under long-term impact of these minor factors. In order to improve the prediction accuracy of the original forecasting model, forecasting error correction method is added to the current forecasting model, which considered the ignoring secondary factors in the process of load forecasting. MA Xiao-guang, MENG Wei [15] put forward a fuzzy linear regression—residual error amendment model based on real number output and fitting residual error to improve the accuracy of load forecasting. The residual error amendment model is derived from fuzzy linear regression model, so it can find the most suitable linear function to make the line difference sum in ideal linear regression minimum. Zhou Ming, YAN zheng *et al*. [16] also presented forecasting error models by iterative method and used the predicted errors to update the forecasted prices so as to gradually improve the forecasting accuracy. The results showed that the presented approach improved the accuracy of forecasting significantly. In other forecasting fields, error correction model is also applied widely. Ren Hongli [17] presented a new prediction method of predictor-based error correction (PREC) in order to effectively use statistical experiences in dynamical prediction. Analyses showed that the PREC can reasonably utilize the significant correlations between predictors and model prediction errors and correct prediction errors by establishing statistical prediction model. An-Sing Chen, Mark T. Leung [18] proposed the two-stage models that general regression neural network was used to correct the errors of the estimates. Trading simulation experiments suggested that the proposed hybrid approach produced better exchange rate forecasts. James M.W. Wong, Albert P.C. Chan *et al.* [19] developed a vector error correction (VEC) model for manpower demand forecast.

Based on the above analysis, a non-parameter error correction model is presented to correct the predicted results of load forecasting by the forecasting error to improve prediction accuracy in this paper. In reality, the relationship between variables is unknown, so the model setting error usually exits in the traditional linear and nonlinear models, which cannot meet the needs of practical application. On the contrary, the non-parameter regression model assumes that the relationship between variables is unknown, so non-parametric regression model is a more realistic model to estimate regression function.

The remainder of this paper is organized as follows: The LS-SVM is introduced in Section 2, including the formulation of LS-SVM. Improved PSO algorithm is presented in Section 3. Non-parametric regression model and simulation are described in Section 4. The proposed combined forecasting model will be compared with other individual models in Section 5. Conclusions are discussed in Section 4.

2. Methodology

2.1 Least Squares Support Vector Machine（LSSVM）

The essential idea of SVM regression is to use a kernel function to map the initial input data into a high-dimensional space(Hilbert space) so the two classes of data become, as far as possible, linearly separable. However, Standard SVM regression has some drawbacks, mainly associated with its formulation and efficiency. LS-SVM are a modification of the standard SVM formulation introduced to overcome these disadvantages, and the resulting optimization problem therefore has half the number of parameters and the model is optimized by solving a linear system of equations instead of a quadratic problem [20].

Given a set of function samples $\{(x_1, y_1),...,(x_N, y_N)\}\in R^d\times R$, where N is the number of samples and d is the dimensionality of the input data, the classic unbiased LS-SVM model for regression tries to approximate a zero-mean function $y=f(x)$ that relates the inputs X to the output Y by solving the following optimization problem:

$$\min_{\omega,e} : J(\omega,e)=\frac{1}{2}\omega^{\mathrm{T}}\omega+\gamma\sum_{i=1}^{N}e_i^2 \tag{1}$$

Constraints are as follows:

$$y_i=\omega^{\mathrm{T}}\varphi(x_i)+b+e_i \tag{2}$$

Where ω denotes adjustable weighted vector, b denotes the bias, and $\varphi(x)$ denotes non-linear mapping from an input space to a high-dimensional space. J is the function to be optimized which depends on the weight vector $\omega\in R^d$, γ is a regularization factor, and e_i is the error committed when approximating the ith sample. The problem can be solved using Lagrange multipliers, and Lagrange function is in the following:

$$L(\omega,b,e;\alpha)=J(\omega,e)-\sum_{i=1}^{n}\alpha_i\{\omega^T\varphi(x_i)+b+e_i-y_i\} \tag{3}$$

Where Lagrange multiplier $\alpha_k\in R$. Parameters can be obtained by solving the following linear system according to partial derivative of the Lagrange function:

$$\begin{bmatrix} 0 & I^{\mathrm{T}} \\ I & \boldsymbol{k}+\frac{1}{\gamma}\boldsymbol{E} \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix}=\begin{bmatrix} 0 \\ y \end{bmatrix} \tag{4}$$

Where k is the scalar product between a pair of input points, $I=[1\ldots1]$, $\alpha=[\alpha_1\ldots\alpha_n]$, and E is a unit matrix. The scalar product operation in the input space is substituted by a scalar product in a feature space given by a kernel function $k(x, x_i)$, so the function of LS-SVM model is given by:

$$y(x)=\sum_{i=1}^{n}\alpha_i k(x,x_i)+b \qquad (5)$$

Generally, the specific expression of mapping and high dimensional feature space does not need to be known in regression model, as long as the form of kernel function can be given. Polynomial functions, radial basis function, sigmoid function, etc. are commonly used [21]. In this paper, radial basis function is adopted in the following form:

$$k\left(x_i,x\right)=\exp[-\|x-x_i\|^2/(2\sigma^2)] \qquad (6)$$

Where σ is the kernel parameter.

A correct procedure for the setting and optimization the parameters of a LS-SVM model for regression is of critical importance in enabling us to obtain good performances and avoid excessive computation times. Load forecasting is an especially difficult case of regression, in that it is very sensitive to the values of the parameters in order to get models with good accuracy and generalization abilities.

The regularization factor γ in Eq. (4) is in charge of the trade-off between the smoothness of the model and its accuracy. The bigger the regularization factor the more importance is given to the error of the model in the minimization process. An excessively large value of γ suggests that the model over-fits the data, thus losing its generalization capabilities. Therefore, it is most important to give a proper value to the regularization factor. The σ value is related to the distance between training points and the smoothness of the interpolation of the model [22]. As a general rule, the higher the σ is, the smoother the interpolation between two consecutive points is.

2.2 Parameter Selection using Improved Particle Swarm Optimization

Recently, PSO developments and applications have been widely explored in engineering and science mainly due to its distinct favorable characteristics. Refs. [23] is an excellent reference that analyzed and studied the PSO promising convergence characteristics With regard to its mathematical development. The authors successfully established some mathematical foundation to explain the behavior of a simplified PSO model in its search for an optimal solution. Just like in the case of other evolutionary algorithms, PSO has many key features that attracted many researchers to employ it in different applications [24-25].

In PSO, a number of particles form a ‘‘swarm” that evolve or fly throughout the feasible hyperspace to search for fruitful regions in which optimal solution may exist. Each particle has two vectors associated with it, the position (X_i) and velocity (V_i) vectors. In *N*-dimensional search space, $X_i=(x_{i1}, x_{i2}, \ldots, x_{iN})$ and $V_i=(v_{i1}, v_{i2}, \ldots, v_{iN})$ are the two vectors associated with each particle *i*. During their search, members of the swarm interact with each other’s in a certain way to optimize their search experience. There are different variants of particle swarm paradigms but the most commonly used one is the *gbest* model where the whole population is considered as a single neighborhood throughout the flying experience [26]. In each iteration, particle with the best solution shares its position coordinates (*gbest*) information with the rest of the swarm. Each particle has its own best position $P_i=(p_{i1}, p_{i2}, \ldots, p_{iD})$ based on the personal best objective value obtained and the global best particle, which is denoted by $P_g=(p_{g1}, p_{g2}, \ldots, p_{gD})$.

In the adjustable parameters, inertia weight is the most important parameter. The larger weight will help improve the global search capability, while the smaller weight will increase the weight capacity of the local search algorithm. In this paper, adaptive inertia weight ω (t) based on fitness function was presented to update the velocity of each particle. It is beneficial to global search at the beginning of iteration and local search at the end of iteration. This improvement not only makes the adaptive PSO converge faster, but also not easily fall into the local optima.

Each particle updates its coordinates based on its own best search experience (p_{best} and g_{best}) according to the following equations:

$$\begin{aligned} v_i(t+1) &= \omega \times v_i(t) + c_1 \times r_1 \times [p_i(t) - x_i(t)] \\ &+ c_2 \times r_2 \times [p_g(t) - x_i(t)] \end{aligned} \tag{7}$$

$$v_i(t+1) = \begin{cases} v_{\max} & v_i(t+1) \geq v_{\max} \\ -v_{\max} & v_i(t+1) < -v_{\max} \end{cases} \tag{8}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{9}$$

where c_1 and c_2 are two positive acceleration constants, they keep balance between the particle's individual and social behavior when they are set equal; r_1 and r_2 are two randomly generated numbers with a range of [0-1] added in the model to introduce stochastic nature in particle's movement; and ω is the inertia weight and it keeps a balance between exploration and exploitation.

Inertia weight ω is updated in accordance with the value of the objective function of each particle in the end of each iteration, and the adaptive adjustment of the specific formula is as follows:

$$\begin{aligned} v_i(t+1) &= \omega(t) \times v_i(t) + c_1 \times r_1 \times [p_i(t) - x_i(t)] \\ &+ c_2 \times r_2 \times [p_g(t) - x_i(t)] \end{aligned} \tag{10}$$

$$\omega(t) = \begin{cases} \lambda \cdot \omega(t-1) + \theta \dfrac{f(p_g(t)) - f(x_i(t))}{f(p_g(t)) - f(x_{\min}(t))} & \text{if } \omega(t) \geq \omega_{\min} \\ \omega_{\min} & \text{otherwise} \end{cases} \tag{11}$$

Where λ and θ are the constraint factors in the range of (0-1), ω_{min} is the minimum inertia weight. $f(p_g(t))$ is the value of fitness function corresponding to the best particle in the tth iteration, while $f(x_{min}(t))$ is the value of fitness function corresponding to the worst particle. $f(x_i(t))$ is the value of fitness function corresponding to x_i in the tth iteration.

The definition of fitness function is as follows:

$$f = -\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{12}$$

In Eq. (12), y_i is the desired output value and $\hat{y}_i$ is the training output value, N is the number of samples.

The steps of APSO optimization are as follows:

Step1. Initialize the parameters of PSO. Set the acceleration factors c_1, c_2, λ, v_{max}, T_{max}; the current evolution generation is set to t. In this paper, D=2.

Step2. Let the current position of each particle be p_i, and p_g be the current position of the best particle in the whole population

Step3. Calculate the population $x(t)$. The smaller particle's value of fitness function is, the better the position of particle is.

Step4. If the current fitness value of some particle is better than p_i, substitute the current position for p_i; if the best fitness value of all particles is better than p_g, substitute the best position of the best fitness value for p_g.

Step5. Update $x(t)$, $v(t)$ and $\omega(t)$ by Eq. (7) and Eq. (8).

Step6. Determine whether it gets to the global convergence. If reached, continue to step 7; otherwise, go to step 3;

Step7. Output global optimal value, that is optimal parameter vector(C, σ)

2.3 Non-Parametric Error Correction

Non-parametric regression model is characterized by any form of regression function, without any constraints, and few constraints for the distribution of explanatory variables, so the model has greater flexibility. Non-parametric regression model has a better fit than the classical assumption model and higher accuracy in the past analysis of regression. Based on the load error data series, non-parametric regression model is put forward as error model, and local linear estimation is used to estimate the model.

There is no boundary effect to local linear estimation, which means that the convergence of boundary points and internal points is as small as the bias, and the deviation is independent with density function of explanatory variables. Local linear estimation is to minimize the following formula:

$$\sum_{i=1}^{n}\left\{Y_i - m(x) - m'(x)(X_i - x)\right\}^2 K_{h_n}(X_i - x) \tag{13}$$

Where $K_{hn}(u)= h_n^{-1}\mathrm{K}(h_n^{-1}u)$, h_n is the bandwidth, and $k(\cdot)$ is the probability density function in[-1,1]. If $k(\cdot)$ is the probability density function, local linear estimation of $m(x)$ is the least squares estimation of partial model(X_i, Y_i), of which X_i is in[x-h_n, x+h_n]. Y_i is as follows:

$$Y_i = m(x) + m'(x)(X_i - x) + e_i \tag{14}$$

When the X_i is closer to x, the weight of Y_i is greater. Local linear estimate of $m(x)$ can be expressed as matrix:

$$m(x, h_n) = e_1^T (X_x^T W_x X_x)^{-1} X_x^T W Y \tag{15}$$

Where $e_1=(1,0)^T$, $X_{x,i}=(1,(X_i-x))^T$, $\mathrm{W}_x=\mathrm{diag}\{K_{hn}(X_1-x),..., K_{hn}(X_n-x)\}$, $Y=[Y_1,..., Y_n]^T$

A. Bandwidth selection

When $h_n\to 0$, $m_n(X_i) \to K(0)Y_i/K(0)= Y_i$ and $m_n(X_i)\to 0$, ($x\neq X_i$, i=1,...,n). If bandwidth is too small, the function values of the points are zero in addition to data points. So random error noise has not been ruled out and there is no meaningful estimate. When $h_n\to\infty$, $K((X_i-x)/h_n)\to \mathrm{K}(0)$ and $m_n(X_i)\to n^{-1}\sum Y_i$, ($i$=1,...,n). However, too large bandwidth value will obtain too smooth curves, which is close to a straight line. Therefore, estimation at this time is of no significance.

Seen from the above, the bandwidth is an important parameter to control estimation accuracy. Cross-validation method is used to select bandwidth in this paper. Kernel estimation is as follows:

$$\hat{m}_{n,-i}(X_i) = \sum_{j\neq i} W_{nj}(X_i) Y_j \tag{16}$$

Finally, we can get the minimum bandwidth h_n by comparing the squared fitting error:

$$CV(h_n) = n^{-1}\sum_{i=1}^{n}(Y_i - \hat{m}_{n,-i}(X_i))^2 W(X_i) \tag{17}$$

B. Simulation

In order to test the effect of local linear estimate, the following fitting model is used to simulate:

$$Y_i = \sin(2\exp(X_i + 1)) + u_i \tag{18}$$

Where $X_i=i/100$, $u_i\sim N(0,1)$, i=1,...,n, and h_n=0.08. Gaussian kernel is selected as kernel function and the experiments are simulated on Mtalab 7.1 software platform.

The simulation results are shown in Figure 1 and Figure 2. Figure 1 is the function graph of m(x). Scatter plot for (X_i, Y_i) and the local linear estimation curve are in Figure 2. The local linear estimation shows good performance compared with Figure 1.

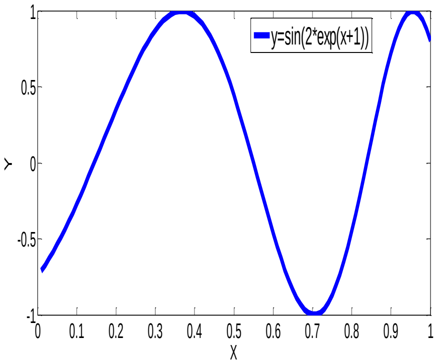

Figure 1. *y*=sin(2*exp(*x*+1))

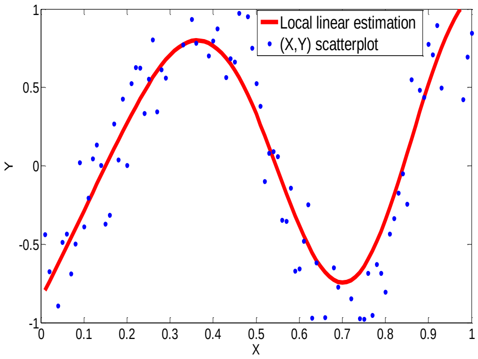

Figure 2. Local Linear Estimated Fitting Curve and (*X, Y*) Scatterplot

Adjustment process of non-parametric error correction model is as follows:

(1) Use the model of load forecasting mentioned above to forecast the historical load $\hat{y}_i$

(2) Calibrate the forecasted value with forecasted error $E_i(1)$:

$$\hat{y}_{1i} = \hat{y}_i + E_i(1) \tag{19}$$

(3) Calculate the residual $e_i(1)=y_i-\hat{y}_{1i}$. If $E_i(1)$ meets the required accuracy, go to step (6); otherwise, go to step (4).

(4) Establish another new model of error forecasting according to $e_i(1)$ and obtain new forecasted error $E_i(2)$; then calibrate the forecasted value:

$$\hat{y}_{2i} = \hat{y}_i + E_i(1) + E_i(2) \tag{20}$$

(5) Calculate the residual $e_i(2)=y_i-\hat{y}_{2i}$. If $e_i(2)$ meets the required accuracy, go to step (6), otherwise, go to step (4).

(6) Models of error forecasting can be obtained by iterative method, which will be used to calibrate load value.

After completion of the modeling process above, integrated forecasting models can be formed: one load forecasting model (APSO-SVM) and several non-parametric error correction models. Therefore, the combined forecasting model established by training samples is applied to the test samples:

$$\hat{y}_i' = \hat{y}_i + \sum_k E_i(k) \quad i = 1, 2, \cdots, n \quad k = 1, 2, \cdots \tag{21}$$

In Eq. (21), $\hat{y}'_i$ is the final revised load forecasting value, $\hat{y}_i$ is the load value forecasted by APSO-SVM, $E_i(k)$is the kth grey error forecasting model.

3. The Case Study

24 points load data from some area in South China is selected as training and testing samples. Appropriate input variables are selected, mainly including the previous one-week load data, previous three-day load data, previous one-day load data, previous three-hour load data, previous two-hour load data, previous one-hour load data, day type, humidity, average temperature, maximum temperature and minimum temperature. Load data from 1/7/2009 to 30/9/2009 is selected as training samples and used to establish forecasting model, and three months of which is training samples and load data from 1/10/2009 to 31/10/2009 is selected as testing samples.

Moreover, inner product of feature vectors needs to be calculated, and large feature values may cause numerical difficulties, such as the linear kernel and polynomial kernel. To avoid difficulty of numerical calculation, we will transform samples to the range of [0-1]:

$$x' = (x - x_{\min})/(x_{\max} - x_{\min}) \tag{22}$$

Where x is the original load data；x_{min} is the minimum value；x_{max} is the minimum value；x' is the standardized data.

A. *Load forecasting of APSO-SVM*

Parameters setting of the proposed model are as follows: m=20, c_1=c_2=2, λ=0.8, θ=0.5, v_{max}=100, T_{max}=100.

The relative error (RE), absolute percentage error (APE), and mean absolute percentage error (MAPE) between the actual and the forecasted values are used to compare the forecasting models and defined as follows:

$$RE = (y_i - \hat{y}_i)/y_i \tag{23}$$

$$APE = |y_i - \hat{y}_i|/y_i \tag{24}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{y}_i|/y_i) \tag{25}$$

In addition, standard SVM forecasting model is also used for prediction as compared model, and forecasting results and fitting curves are shown in Figure.3 and Figure.4.

Figure 3 illustrates that actual load curve and fitting curve of training sample at 14:00. From Figure 3, the forecasting model of APSO-SVM fits the original load data well and has a higher accuracy. Meanwhile, standard SVM is also used to establish forecasting model. Figure 4 shows forecasting curves of testing samples using both models compared to actual load curve. The results of the two models are shown in Figure 5 and Table 1 MPAE is 3.11% with SVM model and 2.71% with APSO-SVM which has a higher accuracy than SVM. Compared with SVM model, the advantage of APSO-SVM is obvious. It makes more accurate load forecasting and less deviation than SVM. It can also be seen that APSO-SVM model has better stability and generalization performance.

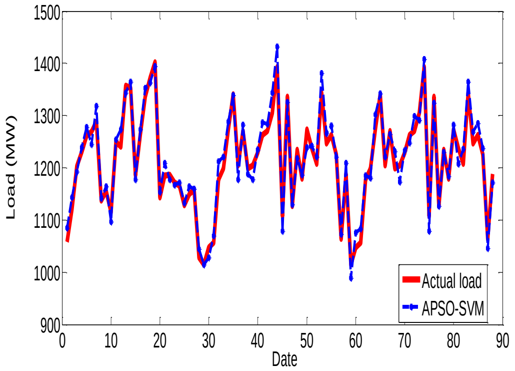

Figure 3. Fitting Curve of APSO-SVM

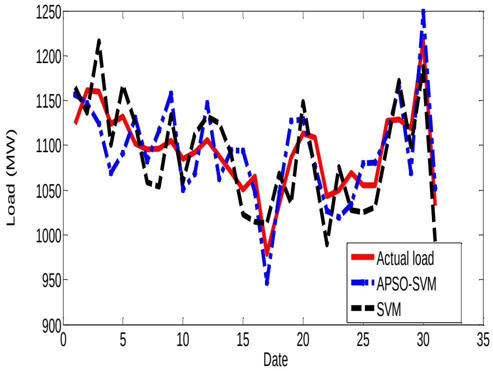

Figure 4. Fitting Curve of Testing Samples

Table 1. Absolute Percentage Error of Different Forecasting models(%)

Date	SVM	APSO-SVM	NP-APSO-SVM	Date	SVM	APSO-SVM	NP-APSO-SVM
1	3.6	2.88	2.6	17	3.48	3.3	3.48
2	2.15	1.26	2.15	18	3.34	1.18	2.34
3	4.85	3.08	1.85	19	4.72	3.77	0.72
4	2.11	4.88	0.11	20	3.15	1.41	3.15
5	3.13	3.6	2.13	21	3.54	2.85	2.54
6	2.1	2.64	2.1	22	5.18	1.53	1.18
7	3.35	1.18	0.35	23	2.51	2.87	2.51
8	3.85	1.86	3.85	24	3.92	3.34	3.92
9	2.57	4.67	2.57	25	2.89	2.33	2.89
10	2.44	3.2	0.44	26	2.38	2.37	1.38
11	1.92	2.15	1.92	27	2.41	1.34	2.41
12	2.34	3.68	2.34	28	3.83	3.2	1.83
13	3.37	2.37	3.37	29	1.82	4.47	1.82
14	1.91	2.31	1.91	30	2.41	2.59	2.41
15	2.59	4.19	2.59	31	3.88	1.89	2.88
16	4.75	1.67	1.75	MAPE	3.11	2.71	2.18

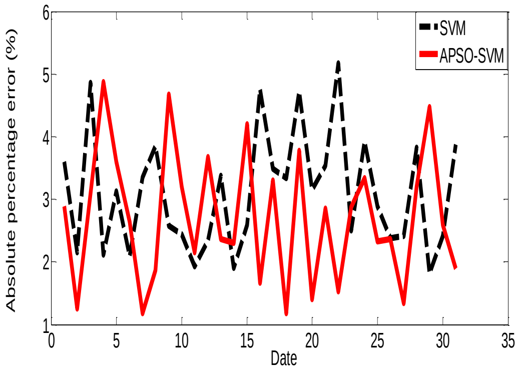

Figure 5. APE Curves of Testing Samples

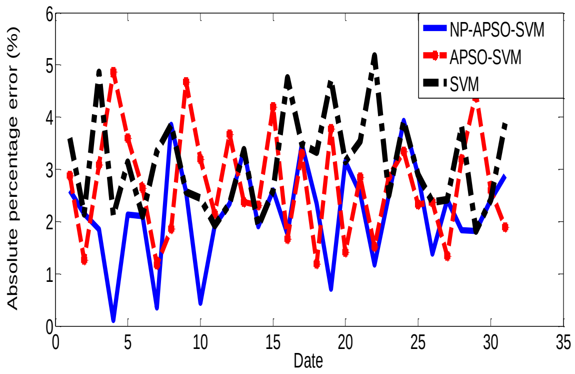

Figure 6. APE Curves of Testing Samples After the First Error Correction

B. Non-parametric error correction

In order to improve the prediction accuracy of PSO-SVM forecasting model, non-parametric error correction method is added to the current forecasting model, which considered the ignoring secondary factors in the process of load forecasting. After the first error correction, absolute percentage error curves of forecasting results are shown in Figure 6 and Table 1 MAPE is 2.18% with NP-APSO-SVM model which is 0.93% and 0.53% lower than SVM and APSO-SVM. According to the results, prediction accuracy has reached the standard and there is no need for a second error correction. Apparently, model of NP-APSO-SVM can narrow the scope of the deviation and improve the accuracy of load forecasting.

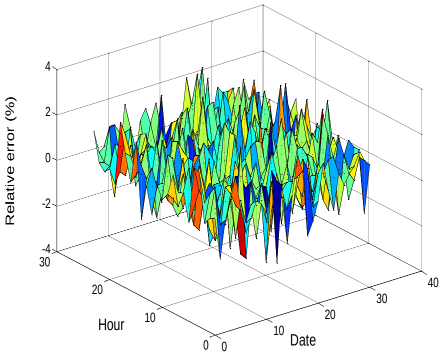

Figure 7. Forecasting Performance of NP-APSO-SVM Model

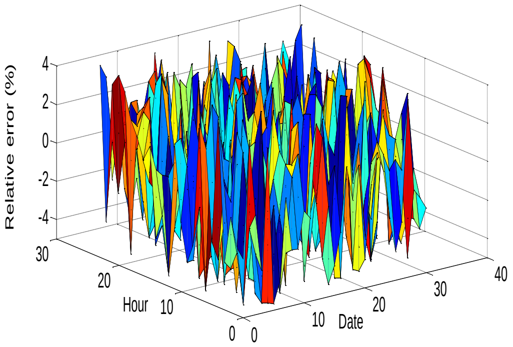

Figure 8. Forecasting Performance of APSO-SVM Model

Figure 7 and Figure 8 describe the surface curve of the load forecasting deviation performance of NP-APSO–SVM model and APSO–SVM model in next month. It depicts the relative forecasting error of every day's 24 time points of the following 31 days which ranges from 01/10/2009 to 31/10/2009. From the two figures, the relative error is in the scope [-3%, 3%] with NP-APSO-SVM and [-5%, 5%] with APSO-SVM. It can be seen that the load forecasting curve of NP-APSO-SVM makes smaller fluctuation than APSO-SVM. The NP-APSO-SVM model makes not only accurate load forecasting, but also the load forecasting curve to be stable in a relatively long time.

4. Summary and Conclusion

In this paper, we have presented a hybrid forecasting model based on the non-parametric error correction. We have also provided the non-parametric error correction model to improve forecasting accuracy.

(1) Combined model of short-term load forecasting is established based on adaptive inertial weight particle swarm optimization and least squares support vector machines. APSO is used to optimize the kernel parameter σ and regularization parameter C, which was proved to have better accuracy. Using improved PSO algorithm to optimize parameters of least squares support vector machine, velocity of convergence is not only faster, but optimization will not easily fall into local optima.

(2) Based on the optimized regression model, non-parametric error correction is also presented by iterative method. Experiment results show that the proposed forecasting model can obtain more generalized performance and better forecasting accuracy.

(3) The research in this paper is useful for the regional grid scheduling and power management, also can be applied to other fields widely. The results provide a guide to the initialization and orientation of the search for the parameter values for this kind of model, which is one of the most used in practice for time series modelling and prediction.

Acknowledgements

This paper is supported by Hebei Province Social Science Foundation the Grant No.HB12GL073, and Hebei Province Educational department Science Foundation the Grant No.GH121003, and National Natural Science Funds of China (Grant No. 11104058), and Hebei province natural science foundation of China (Grant No. A2011201155).

References

[1] Vapnik, V. "The nature of statistical learning," New York: Springer **(1995)**.

[2] Sanchez A. D., "Advanced support vector machines and kernel methods," Neurocomputing, vol. 55 no. 1, **(2003)**, pp. 5-20.

[3] Fei S. W., Liu C. L. and Miao Y. B., "Support vector machine with genetic algorithm for forecasting of key-gas ratios in oil-immersed transformer," Expert Systems with Applications, vol. 36 no.3, **(2009)**, pp. 6326–6331.

[4] Wu C. H., Tzeng G. H. and Lin R. H., "A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression," Expert Systems with Applications, vol. 36 no.3, **(2009)**, pp. 4725–4735.

[5] Widodo A. and Yang B., "Wavelet support vector machine for induction machine fault diagnosis based on transient current signal," Expert Systems with Applications, vol. 35 no. 1–2, **(2008)**30 pp. 7–316.

[6] Huo M., Luo D. S. and He J., "Chaos Optimization Method of SVM Parameters Selection for Short-term Load Forecasting," Proceedings of the CSU-EPSA, vol. 21 no. 5, **(2009)**, pp. 124-128.

[7] Qi L., "Parameters selection of support vector machine based on ant colony algorithm," System Simulation Technology, vol. 4 no. 1, **(2008)**, pp. 14-18.

[8] Niu D. X., Liu D., Chen G. J. and Feng, "Support vector machine models optimized by genetic algorithm for hourly load rolling forecasting," Transactions of China electrotechnical society, vol. 22 no. 6, **(2007)**, pp. 148-153.

[9] Zhao Y., Liu H. X., and Gao D. T., "Sum of squared derivatives criterion and SVM hyperparameters optimization," Journal of System Simulation, vol. 20 no. 12, **(2008)**, pp. 3150-3156.

[10] Tao S. H., Chen D. Z., and Hu W., "M. Gradient algorithm for selecting hyper parameters of LSSVM in process modeling," Journal of Chemical Industry and Engineering, vol. 58 no. 6, **(2007)**, pp. 1514-1518.

[11] Xie H., Wei J. P., and Liu H., "Parameter selection and optimization method of SVM model for short-term load forecasting," Proceedings of the CSEE, vol. 26 no. 22, **(2006)**, pp. 43-47.

[12] Kennedy J., Eberhart R. C. and Shi Y., "Swarm Intelligence," San Francisco: Morgan Kaufman Publishers, **(2006)**.

[13] Shi Y. and Eberhart R. C., "A modified particle swarms optimizer," Proceedings of IEEE Conference on Evolutionary Computation, **(1998)**, pp. 69-73.

[14] Naka S., Genji T., Yura T. and Fukuyama Y., "A hybrid particle swarm optimization for distribution state estimation," IEEE Trans. on Power Systems, vol. 18 no. 1, **(2003)**, pp. 60-68.

[15] Ma X. G. and Meng W., "Application of residual error amendment method in power load forecasting," Power System Technology, vol. 24 no. 4, **(2001)**, pp. 21-25.

[16] Zhou M., Yan Z., Ni Y. X. and Li G. Y., "A novel ARIMA approach on electricity price forecasting with the improvement of predicted error," Proceedings of CSEE, vol. 24 no. 12, **(2004)**, pp. 63-68.

[17] Ren H. L., "Predictor-based error correction method in short-term climate prediction," Progress in Natural Science, vol. 18 no. 1, **(2008)**, pp. 129-135.

[18] Chen A. S. and M. T. Leung, "Regression neural network for error correction in foreign exchange forecasting and trading," Computers & Operations Research, vol. 31 no. 7, **(2004)**, pp. 1049-1068.

[19] Wong J., A. Chan and Y. H. Chiang, "Forecasting construction manpower demand: A vector error correction model," Building and Environment, vol. 42 no. 8, **(2007)**, pp. 3030-3041.

[20] Rubio G., "A heuristic method for parameter selection in LS-SVM: Application to time series prediction," International Journal of Forecasting, **(2010)**.

[21] Avci E., "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm-support vector machines: HGASVM," Expert Systems with Applications, vol. 36 no. 2, **(2009)**, pp. 1391–1402.

[22] Rojas I., Pomares H., Gonza′lez J., Bernier J., Ros E. and Pelayo F., "Analysis of the functional block involved in the design of radial basis function networks," Neural Processing Letters, vol. 12, **(2000)**, pp. 1–17.

[23] Wu Q., "A hybrid-forecasting model based on Gaussian support vector machine and chaotic particle swarm optimization," Expert Systems with Applications, vol. 37, **(2010)**, pp. 2388-2394.

[24] Tang L. B., Tang L. X. and Sheng H., "Forecasting volatility based on wavelet support vector machine," Expert Systems with Applications, vol. 36 no. 2, **(2009)**, pp. 2901–2909.

[25] Lin S. W., Ying K. C., Chen S. C. and Lee Z. J., "Particle swarm optimization for parameter determination and feature selection of support vector machines," Expert Systems with Applications, vol. 35 no. 4, **(2008)**, pp. 1817–1824.

[26] Liu L. X., Zhuang Y. Q. and Liu X. Y., "Tax forecasting theory and model based on SVM optimized by PSO," Expert Systems with Applications, vol. 38, **(2011)**, pp. 116-120**.**