Short-term Load Forecasting with LS-SVM Based on Improved Ant Colony Algorithm Optimization

Yuansheng HUANG, Jiajia DENG†

School of Economics and Administration, North China Electric Power University, Beijing, China

Abstract

Research of short-term load forecasting has important practical application value in the field of power network dispatching. The regression models of least squares support vector machines (LS-SVM) have been applied to load forecasting field widely, and the regression accuracy and generalization performance of the LS-SVM models depend on a proper selection of its parameters. In this paper, a new regression model of LS-SVM was put forward based on adaptive ant colony optimization (AACO). AACO was used to optimize the kernel parameter $\sigma$ and regularization parameter $C$ of LS-SVM. Results of case analysis and simulation show that the proposed forecasting model can obtain more generalized performance and better forecasting accuracy compared with the method of single SVM and BP neural networks.

Keywords: Load Forecasting; Least Square Support Vector Machines; Ant Colony Optimization

1. Introduction

Short-term load forecasting is an important factor for making plans for energy transactions and dispatching scheduling of generating capacity. In order to improve safety and economy of grid operation, improve the quality of power supply and realize automation of power system, higher accuracy is necessary in short-term load forecasting for the modern power system.

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. Recently, SVM is a promising method for pattern classification and regression proposed by Vapnik, based on statistical learning theory and structural risk minimization [1]. Reference [2,3]proposed the method based on the least square support vector machine for both mid-long term and short-term load forecasting; Reference [4,5] put forward a SVM forecasting system based on data mining preprocess, which eliminated redundant information and improved the processing speed. Experiment and simulation have shown that kernel parameter $\sigma$ and regularization parameter $C$ played a significant part in the performance of SVM. Methods of gradient descend [6], genetic algorithm [7] and particle swarm optimization have been used to select parameters of SVM in many fields.

This paper proposed a method of parameters selecting based on ant colony optimization. Ant colony algorithm is a new type of simulated evolutionary algorithm, which has the features of positive feedback,
high accuracy, robustness, etc [8-12]. The results of case experiment show that the presented approach improves the accuracy of forecasting significantly.

2. SVM Regression Model

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.

Let the samples load data, \( \{x_i, y_i\} \) \( (i = 1, 2, 3, \ldots, n) \), SVM regression based on \( \varepsilon \)-insensitive loss function is adopted, so precision of the algorithm can be controlled in advanced to ensure the sparsity of dual variables.

The estimating function of using SVM regression:

\[
f(x) = \omega \varphi(x) + b
\]

In Eq. 1, \( \varphi(x) \) denotes non-linear mapping from an input space to a high-dimensional space; \( \omega \) denotes adjustable weighted vector; \( b \) denotes the bias.

The purpose of solving the problems of regression estimation is to obtain \( \omega \) and \( b \) by minimizing a regularization function with risk:

\[
R_{SVM}(\omega) = c \left[ \frac{1}{n} \sum_{i=1}^{n} L^\varepsilon (y_i, \omega \varphi(x_i) + b) + \frac{1}{2} \| \omega \|^2 \right]
\]

In Eq. 2, \( c \left[ \frac{1}{n} \sum_{i=1}^{n} L^\varepsilon (y_i, \omega \varphi(x_i) + b) \right] \) denotes empirical risk; the constant \( C \) determines the tradeoff between the empirical risk and regularization, \( \frac{1}{2} \| \omega \|^2 \) denotes the part of regularization.

The slack variables \( \xi \) and \( \hat{\xi} \) are introduced into the optimization problem above. The optimization problem solved by SVM can be formulated as:

\[
\min_{\omega} R_{SVM}(\omega) = \frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{n} (\xi_i + \hat{\xi}_i)
\]

\( \omega \) can be obtained by the method of derivation calculus. Using duality theory, it can be converted that the variables \( \alpha_i, \hat{\alpha}_i \) are obtained by solving the following dual problem.

\( \omega \) and Gaussian kernel function \( k(x, x') \) are introduced to the estimating function:

\[
f(x) = \sum_{i=1}^{n} (\alpha_i - \hat{\alpha}_i) k(x_i, x) + b
\]

In Eq. 4, \( k(x_i, x) = \exp[-\|x - x_i\|^2 / (2\sigma^2)] \).
From the estimating function mentioned above, the way of mapping the initial input data into a high-dimensional space depends on kernel function, and the regularization parameter $C$ is used for balancing between training error and model complexity.

3. Parameter Selection Based on Ant Colony Optimization

Appropriate parameters selection is crucial to the learning performance and generalization ability of SVM. Ant colony algorithm is used to select parameters $\sigma$ and $C$. This simulated evolutionary algorithm was proposed by M. Dorigo and, which has the features of Positive feedback, high accuracy and robustness. $C$ and $\sigma$ are reflected by the node value in the ant colony system, and hormones are left in each node traversed by the ants. The concentration of pheromone is not updated according to the length of path in the ant colony system when optimizing parameter, but according to the objective function value. The objective function implied all information of the nodes walked by the ants and the current accuracy of the model.

Let parameters $C$ and $\sigma$ be variables to be optimized; the effective bits and accuracy of parameters $C$ and $\sigma$ are determined by the experts before the first run. Five effective bits are allocated to each parameter and the first two bits before decimal point are percentile and decile. To facilitate the use of ant colony algorithm, these five bits of each parameter are expressed in the plane $XOY$, which is shown in Fig.1. $L1$–$L5$ and $L6$–$L10$ are assigned to five bits of parameters $C$ and $\sigma$ respectively. In the plane $XOY$, let $\knot(x_i, y_{ij})$ be a node, $x_i$ be the abscissa of segment $Li$ ($i = 1$–$15$), and $y_{ij}$ be the ordinate of $Li$ in the node $j$ ($j = 0$–$9$).

Let a set of ants starting from origin $O$ of the coordinate, and they complete a cycle when crawling to the line $L10$ at any point. For example, from the Fig.1 we can see that the optimized parameter values expressed by the most optimal path are: $C = 62.639, \sigma = 57.954$

![Fig.1 Parameters $C$ and $\sigma$ Optimized by Ant Colony Algorithm](image)

Let $\tau(x_i, y_{ij}, t)$ be the pheromone on the node $\knot(x_i, y_{ij})$ at $t$ time, and pheromone on each node is equal at the initial moment, that is $\tau(x_i, y_{ij}, 0) = \lambda$ ($i = 1$–$10, j = 0$–$9$, $\lambda$ is a constant), $\Delta r(x_i, y_{ij}, 0) = 0$.

$P_k(x_i, y_{ij}, t)$ indicates the probability of crawling from $\knot(x_{i+1}, y_{i+1})$ to $\knot(x_i, y_{ij})$, then its formula is:
In Eq. 5, \( \eta(x_i, y_{ij}, t) \) is the expectation from the knot \((x_{i-1}, y_{i-1}, j)\) to knot \((x_i, y_i, j)\), and its value can be determined according to the heuristic algorithm and the situation of the problem.

Assuming the initial time \( t = 0 \), and all the ants are in origin \( O \) of the coordinate, then after 10 time units, all the ants are crawling to the destination from starting point, then the amount of pheromone at each path point can be adjusted according to the following formula:

\[
\tau(x_i, y_i, t + 10) = \rho \tau(x_i, y_i, t) + \Delta \tau(x_i, y_i, t)
\]

\[
\Delta \tau(x_i, y_i, t) = \sum_{k=1}^{m} \Delta \tau_k(x_i, y_i)
\]

\[
\Delta \tau(x_i, y_i, t) = \begin{cases} 
Q & \text{if } F_k < \min \tau(t) \\
0 & \text{otherwise}
\end{cases}
\]

In Eq. 6, \( \rho \) is the pheromone coefficient, in Eq. 7, \( m \) is the amount of the ants; in Eq. 8, \( Q \) is a constant; \( F_k \) denotes the value of the objective function of the \( k \)th ant in this cycle, which can be calculated in the following formula:

\[
F = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

In this paper, adaptive pheromone coefficient \( \rho(t) \) was presented to update the amount of pheromone at each path point. 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 ASO converge faster, but also not easily fall into the local optima. The form of \( \rho(t) \) is expressed as follows:

\[
\tau(x_i, y_i, t + 10) = \theta \rho(t) \tau(x_i, y_i, t) + \Delta \tau(x_i, y_i, t)
\]

\[
\rho(t) = \begin{cases} 
\theta \cdot \rho(t-1) & \text{if } \theta \cdot \rho(t-1) < \rho_{\min} \\
\rho_{\min} & \text{otherwise}
\end{cases}
\]

In Eq. 11, \( \rho_{\min} \) is the minimum pheromone coefficient, \( \theta \) is the constraint factor.

Steps of optimizing parameters are as follows:

(1) Let the number of ants be \( m \) and the maximum number of cycles be \( N_{\max} \); time counter \( t = 0 \); number
of cycles \( N = 0; \tau(x_{i_{1}}, y_{i_{1}}) = 0 ; \Delta \tau(x_{i_{1}}, y_{i_{1}}) = 0 ; i=1 \). All ants are placed at the starting point \( O \).

(2) Calculate probability of these ants transferring to each node on \( L_i \) according to Eq. 5, then use method of roulette wheel to select a node on each ant \( k (k = 1 \sim m) \) according to calculated probability.

(3) \( i = i + 1, \) if \( i \leq 10 \), then jump to step (2); otherwise, go to step (4).

(4) Calculate the parameters \( C \) and \( \sigma \) according to the path crawled by the ants, then note the best parameter \( C^* \) and \( \sigma^* \) based on the objective function value.

(5) \( t = t + 1, N = N + 1; \) update the pheromone on each node by Eq. 6, Eq. 7 and Eq. 8.

(6) If \( N < N_{\text{max}} \) and the whole ant colony haven’t converged to the same path, all of the ants are back to the starting point; and then jump to step (3). If \( N < N_{\text{max}} \) and the whole ant colony have converged to the same path, the algorithm ends. Then output the best parameter \( C^* \) and \( \sigma^* \) based on the optimized path.

4. Simulation

Function \( \text{sinc} \) is used to verify the effectiveness of the proposed optimizing algorithm, which is commonly used in SVM regression. All tests are completed using MTALAB 7.0 software platform. One-dimensional function \( \text{sinc} \) is as follows:

\[
y = f(x) = \sin c(x) + \nu, \quad x \in [-2, 2]
\]

In Eq. (12), \( \nu \) is Gaussian white noise with its mean of 0 and variance of 0.1.

Range of input variables obtained in the 100 SVM data constitute the training sample of SVM is obtained by 200 data within range of input variables, using AACO algorithm to optimize regularization parameter \( C \) and parameter \( \sigma \) of RBF kernel function. Parameters of AACO are assigned as follows: \( Q = 100, \rho = 0.7, \lambda = 0.1, \theta = 0.8, \alpha = 1, \beta = 5, m = 30, N_{\text{max}} = 100 \). The results of simulation are shown in Fig. 2 and Fig. 3.
Figures above show the results of training SVM using the parameters obtained by AACO, and the effect of fitness was very good and the error value can meet the requirement, so the proposed optimizing method can be applied to load forecasting in the following analysis.

5. Case Analysis

In this paper, load data at 14:00 in some area from July 2009 to August 2009 is selected as training and test samples. Parameters setting of the proposed model are as follows: \( Q=100, \ \rho=0.7, \ \lambda=0.1, \ \theta=0.8, \ \alpha=1, \ \beta=5, \ m=30, \ N_{\max}=100 \). Parameters can be optimized by the ant colony algorithm of matlab program: \( C^* = 86.12, \ \sigma^* = 37.378 \). In addition, single SVM and BP neutral networks are applied to forecasting, and the results are analyzed compared with AACO-SVM in Table 1.

Mean absolute percentage error is taken as the criterion to judge the prediction effect:

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|
\]  

(13)

Table 1 Comparision of the Load Forcasting Result of Workday at 14:00

<table>
<thead>
<tr>
<th>Date</th>
<th>AACO-SVM</th>
<th>SVM</th>
<th>BP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE%</td>
<td>MAPE%</td>
<td>MAPE%</td>
</tr>
<tr>
<td>2009-08-01</td>
<td>2.77</td>
<td>2.65</td>
<td>3.15</td>
</tr>
<tr>
<td>2009-08-02</td>
<td>2.14</td>
<td>2.92</td>
<td>2.13</td>
</tr>
<tr>
<td>2009-08-03</td>
<td>2.43</td>
<td>2.43</td>
<td>3.47</td>
</tr>
<tr>
<td>2009-08-04</td>
<td>1.33</td>
<td>2.84</td>
<td>2.44</td>
</tr>
<tr>
<td>2009-08-05</td>
<td>1.84</td>
<td>2.71</td>
<td>3.08</td>
</tr>
<tr>
<td>2009-08-06</td>
<td>3.61</td>
<td>1.89</td>
<td>3.69</td>
</tr>
<tr>
<td>2009-08-07</td>
<td>1.97</td>
<td>3.15</td>
<td>1.78</td>
</tr>
<tr>
<td>2009-08-08</td>
<td>2.44</td>
<td>3.94</td>
<td>3.97</td>
</tr>
<tr>
<td>2009-08-09</td>
<td>2.67</td>
<td>3.59</td>
<td>3.24</td>
</tr>
<tr>
<td>2009-08-10</td>
<td>1.02</td>
<td>2.33</td>
<td>2.68</td>
</tr>
<tr>
<td>2009-08-11</td>
<td>1.73</td>
<td>3.77</td>
<td>3.84</td>
</tr>
<tr>
<td>2009-08-12</td>
<td>2.77</td>
<td>2.91</td>
<td>4.57</td>
</tr>
<tr>
<td>Average value</td>
<td>2.17</td>
<td>2.70</td>
<td>3.17</td>
</tr>
</tbody>
</table>
In order to show the advancement of the model in this paper, single SVM and BP neutral networks are also built for comparison. Forecasting results are shown in Table 1 and Fig.4, AACO-SVM has a better performance on forecasting accuracy compared with single SVM and BP neutral networks.

6. Conclusion

A new regression model of LS-SVM was put forward based on adaptive ant colony optimization (AACO) and grey error calibration. AACO was used to optimize the kernel parameter $\sigma$ and regularization parameter $C$, which was proved to have better accuracy. Simulation and case analysis results show that the proposed forecasting model can obtain more generalized performance and better forecasting accuracy compared with the method of single SVM and BP neural networks. The presented approach is suitable for extending to mid-term load forecasting, electricity price forecasting and other forecasting fields.

References


