Compare Relevance Vector Machine with Improved Support Vector Machine in Short-term Power Load Forecasting

Guo Xiaopeng and Wang Yi
School of Economics and Management, North China Electric Power University, Beijing, 102206, China

Abstract: This study explores two regression forecasting models of Relevance Vector Machine (RVM) and Support Vector Machine (SVM) to improve short-term load forecasting accuracy in small sample data sets. In the regression model of SVM, the study adopts Particle Swarm Optimization (PSO) algorithm to automatically select the penalty factor C and the parameter of kernel functions, which can overcome the drawback that some specified parameters are often determined by experience. The regression model of RVM based on sparse Bayesian theory has many advantages, such as good generalization ability, no penalty factors setting. In order to verify the forecasting performance of two methods, the study uses power load data from a certain area of Fujian province. The results obtained present that they all have good forecasting accuracy. By contrast, the regression model of RVM has better performance, more stability and efficiency.

Keywords: Support vector machine, relevance vector machine, short-term power load forecasting, particle swarm optimization

INTRODUCTION

Electric power load forecasting can effectively improve safety and reliability of electricity systems, hence, power system planning, operation dispatch and power market transaction often depend on it. Load forecasting can be divided into three categories in terms of planning horizon's duration: short-term forecasts, medium forecasts and long-term forecasts. Among them, short-term load forecasting (STLF) plays an important role in traditional monopolistic power systems, whose principal objective is to provide the load prediction for basic generation scheduling, for timely dispatcher information and for assessing the security of system operation. Therefore, STLF is the main composition module of energy management system and an important part of electricity market technology support system.

In consideration of STLF's importance, researchers employed different forecasting methods for achieving predictive accuracy. Regression analysis and time series models are frequently adopted in many fields. However, their predictive accuracy is low when the sample is small. The grey model and its extended models are suited to small data sets with limited information, but if the discrete degree of the raw data sequence is high, the forecasting accuracy is usually low (Li and Wang, 2007). As for system dynamics-based models, their parameters setting is complex and often depends on experience (Tan et al., 2011). Recently, Artificial Neural Networks (ANNs), with mature nonlinear mapping ability and data processing characteristics, emerged as an attractive tool to improve forecasting accuracy (Cui et al., 2009). However, it often encounter local extremum and over-fitting problems. SVM, developed by Vapnik (1995) and his colleagues, has solid theoretical foundation rooted in statistical learning theory and has been successfully applied to regression problems (Song et al., 2012). SVM implements the principle of structural risk minimization by constructing an optimal separating hyper-plane in the hidden feature space, hence, the performance of its predictive model heavily depends on several hyper-parameters selection (Salcedo-Sanz et al, 2011). Unfortunately, the choice of hyper-parameters of traditional SVR model is often determined by experience, which may generate larger prediction error. For this reason, researchers attempted to use evolutionary algorithms to optimize hyper-parameters for SVR in order to improve its generalization ability. These approaches includes Genetic Algorithm (GA) (Wu et al., 2009), Particle Swarm Optimization (PSO) algorithm (Salcedo-Sanz et al., 2011), Bee colony algorithm, etc. (Hong, 2011). By contrast, GA needs conduct a series of complicate operations, such as encoding, selection, intersection and variation, by
contrast, PSO algorithm has the advantages of high efficiency, fast convergence, better capabilities, etc. Therefore, an improved SVM forecasting model combined with PSO for parameter selection is adopted for STLF.

Relevance Vector Machine (RVM) is a popular learning machine motivated by the statistical learning theory, which has been employed for classification and regression problems. Tipping first demonstrated RVM model, a probabilistic nonlinear model with a prior distribution on the weights that enforces sparse solutions. Compared with SVM, RVM has many advantages, such as good sparseness, high generalization ability and no penalty factor setting, etc. (Tipping, 2001; Tipping and Faul, 2003). This work is also focused on this methodology, because RVM has shown good performance in previous applications (Shi et al., 2012).

Based on the above analysis, this study uses load data captured from a certain area in Fujian province to verify predictive performance of two models for short-term load forecasting and do comparative analysis. The rest of the study is organized as follows. Section 2 introduces RVM model and improved SVR model based on PSO. Section 3 shows forecasting steps by the proposed methods and detailed algorithm process. Section 4 employs real load data to verify and Section 5 draws conclusion.

MATERIALS AND METHODS

Support vector regression: Support Vector Regression (SVR) maps original data non-linearly into a higher dimensional feature space to yield and solve a linear regression problem and has been found especially useful in time series prediction (Muller et al., 1997). Here a brief description of SVR is introduced. Consider a training data set \( D = \{(x_i, y_i) | i = 1, 2, ..., n\} \), where \( x_i \) is the input variable and \( y_i \) is the output variable. According to statistical learning theory, the regression function between input variables and output variables can be obtained by minimizing the following objective function.

\[
\min \left\{ \frac{1}{2} \| \beta \|^2 + C \sum \xi_i + \xi_i' \right\} \\
\text{s.t.} \quad y_i - \phi(x_i)b - \xi_i \leq \varepsilon + \xi_i', \\
\phi(x_i)b - y_i \leq \varepsilon + \xi_i', \\
\xi_i, \xi_i' \geq 0
\]  

(1)

where, \( C \) is a positive constant as regularization parameter, i.e., a penalty factor, \( \varepsilon \) is the insensitive lost function and \( \xi_i \) and \( \xi_i' \) are slack variables. \( \xi_i' \) is denoted as follows:

\[
\xi_i' = \begin{cases} 
0, & \|f(x_i) - y_i\| \leq \varepsilon \\
\|f(x_i) - y_i\| - \varepsilon, & \|f(x_i) - y_i\| > \varepsilon
\end{cases}
\]  

(2)

The optimization formulation can be transformed into a dual problem by using Lagrange multipliers and the solution is expressed as:

\[
f(x) = \sum_{i=1}^{n} \alpha_i \phi(x_i) + b
\]  

(3)

where, \( \alpha_i, \alpha_i' \) (with \( 0 \leq \alpha_i, \alpha_i' \leq C \)) are Lagrange multipliers. The penalty factor \( C \) is the trade-off between the flatness \( f \) and the amount up to which deviation larger than HH is tolerated. \( K(x_i, x) \) is the kernel function of SVM, which satisfies the Mercer's conditions and performs the non-linear mapping. Those sample points that appear with non-zero coefficients in Eq. 3 are the so called SVs. Several kernel functions have been proposed in research literatures, such as Radial Basis Function (RBF), polynomial function, sigmoid function, etc. In this study, the focus is put on RBF, which is defined as follows:

\[
K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)
\]  

(4)

Parameter selection by PSO: Penalty factor \( C \), insensitive lost function and parameter \( \sigma \) of Kernel function decide the performance of SVR model. Among them, \( \sigma \) shows features of training data sets and determines complexity of the solutions and then affects generalization ability of learning machines; \( C \) constrains penalty to larger fitting deviation because large value may cause over-fitting while small value may cause under-fitting. Therefore, the optimal selection to these parameters is important for improving the forecasting quality of SVM model. In this study, an optimization algorithm based on PSO is employed for parameter selection during the SVM training.

Particle swarm optimization, proposed by Kennedy and Eberhart (1995), is a evolutionary algorithm motivated by the flocking behavior of birds. It optimizes a problem by leading a randomly initialized population of particles around the problem's search-space. The characteristics of particles are denoted by their positions, velocities and fitness values. Fitness values are computed by fitness functions (i.e., the objective function for optimization computation), which can describe the merits or defects of particles. In the computation process, the position of each particle is updated via tracking individual extremum \( P_{best} \) and global extremum \( G_{best} \). When satisfactory individual extremum and global extremum are obtained, it shows that the parameter selection for SVM model is end.
Relevance vector regression: Relevance Vector Machine (RVM) obtains solutions to classification and regression tasks under probabilistic learning framework and yet remains the excellent predictive quality of the support vector machine (Tipping, 2001). In SVM model, it is necessary to estimate the penalty parameter C and the insensitivity parameter H, while the RVM does not suffer from such limitation. Here, the basic summary of relevance vector regression proposed by Tipping (2001) is given as follows.

Given a data set of input-target pairs \( \{x_i, t_i\}_{i=1}^{N} \), the corresponding relation between \( x \) and \( t \) can be written as:

\[
t = y(x; w) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + \epsilon
\]  

where, \( w \) is weight vector, \( K(X, X) \) is kernel function and \( \epsilon \) is the corresponding weight value.

In the relevance vector regression model, it should explore the forecasting error between the forecasting value and the target value from a probability perspective. Due to the assumption of the error that obeys Gauss distribution with independent zero-mean and of independence of the \( t \), the likelihood estimation of the training data set \( \{x_i, t_i\}_{i=1}^{N} \) can be written as:

\[
p(t|w, \sigma^2) = (2\pi\sigma^2)^{-rac{N}{2}} \exp\left(-\frac{(t - \Phi w)^T}{2\sigma^2}\right)
\]

In Eq. 6, \( t = (t_1, ..., t_N)^T \), \( W = [\omega_1, \omega_2, ..., \omega_N]^T \) and \( \Phi = [\Phi(x_1), \Phi(x_2), ..., \Phi(x_N)]^T \). In \( \Phi = [1, K(x_1, x_1), K(x_1, x_2), ..., K(x_1, x_N)]^T \) in order to solve over-fitting problem in Eq. 6, Tipping defined a vector \( \alpha \) of N+1-dimensional hyper-parameter and assign a zero-mean Gaussian prior distribution over \( w \):

\[
p(w|\alpha) = \prod_{i=1}^{N+1} \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_i^2}{2}\right)
\]

According to Eq. 7, the one-to-one corresponding relation between \( W \) and \( \alpha \) is built in order to control the influence of prior distribution on \( \alpha \), which can ensure the sparseness of RVM. The parameter posteriori probability distribution is denoted by Eq. 8 on the basis of Eq. 7 and Bayesian theory:

\[
p(w|\alpha, \sigma^2) = (2\pi)^{-rac{N+1}{2}} |\Sigma|^{-rac{1}{2}} \exp\left(-\frac{(w - \mu)^T \Sigma^{-1} (w - \mu)}{2}\right)
\]

where, \( \Sigma = (\sigma^2 \Phi^T \Phi + \Lambda)^{-1} \), \( \mu = \sigma^2 \Phi^T \tilde{t} \) and \( \alpha = \text{DIAG} (\alpha_0, \alpha_1, ..., \alpha_N) \). Values of \( \alpha \) and \( \sigma \) which maximize Eq. 8 cannot be obtained in closed form and Tipping used the Eq. 9 and 10 to get their most probable values \( \alpha_{\text{opt}} \) and \( \sigma_{\text{opt}} \) by iterative computation:

\[
\alpha_{\text{opt}} = \frac{1}{N} \sum_{i=1}^{N} \frac{t_i - \Phi_i \mu}{\sigma_{\text{opt}}^2}
\]

\[
\gamma_i = 1 - \alpha_i \Sigma_{ii}
\]

where, \( \mu_i \) is the i-th posteriori mean weight and \( \Sigma_{ii} \) is the i-th diagonal element of the posterior weight covariance matrix \( \Sigma \). If we put \( \alpha_{\text{opt}} \) and \( \sigma_{\text{opt}} \) into Eq. 8, the output probability distribution of input value \( x \) can be written as

\[
p(t|x; \alpha_{\text{opt}}, \sigma_{\text{opt}}^2) = M(t|x; \alpha_{\text{opt}}, \sigma_{\text{opt}}^2)
\]

and then the forecasting value of \( t \) is \( y^* = \mu_{\text{opt}} \Phi(x) \).

MODELING PROCESS

The procedure of RVR modeling is outlined as:

1. Data pretreatment,
2. Initialize parameters, such as predictive accuracy, iterative number, etc.,
3. Conduct iterative training and obtain weight values of relevance vectors,
4. Determine whether meet the condition of termination algorithm, if not, update parameters and jump to (3),
5. Output ultimate weight vectors of RVR and

The procedure of SVM-PSO modeling is outlined as:

1. Data pretreatment,
2. Initialize the number, the position and velocity of particles, acceleration factor, penalty factor C and parameter \( \sigma \) of kernel function, etc.,
3. Input sample data to compute fitness value,
4. Obtain individual extremum and global extremum of current iteration,
5. Determine whether meet the condition of termination algorithm by iterative times or predictive accuracy, if not, update parameters and jump to (3),
6. Output optimal parameters of SVR and
7. Build SVM-PSO forecasting model.

IMPLEMENTATION AND RESULTS

The forecasting process in this study is realized via Matlab programming. The encoding and processing of RVM algorithm is realized via SPARSEBAYES Matlab Toolbox 2.0. SVM algorithm is realized via LibSVM tools box. In this study, 24 h punctual load data of a month in fourth quarter from a certain area in Fujian province are used to analyze and verify the forecasting quality of two models. The raw load data is showed in Fig. 1. The first 20 days of load data is used as the training data set. After finishing training, the data of 22th will be used for forecasting and analysis. Next, it needs to normalize raw load data. Mapminmax function is often employed to pretreat raw data when it needs Matlab to realize artificial neural network and SVR algorithms:

\[
Y = \frac{(Y_{\text{max}} - Y_{\text{min}})}{2} \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} + Y_{\text{min}}
\]

with \( X_{\text{max}} \neq X_{\text{min}} \).

Fig. 1: Raw load data

Fig. 2: Comparison of predictive result from three different models and raw data

Fig. 3: Comparison of predictive error from three different models

Root-Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used as the final evaluation indicators, which are defined as:

$$\epsilon_{RMSE} = \sqrt{\frac{1}{N} \sum |\epsilon^2|}, \quad \epsilon_{MAPE} = \frac{\sum |\epsilon|}{N}$$

SVR based on GA has been adopted for STLF, hence, in order to do comparative analysis, we also the forecasting results by this algorithm here. After the completion of the above, the raw data the forecasting results of SVR based on GA, SVR based on PSO and RVR are listed in Table 1.

Based on the prediction results, the error range of RVM model is [1.2958, 1.4905%] and RMSE and MAPE is 1.412 and 1.410%, respectively, which shows that its accuracy and dispersion degree of errors are better. The error range of PSO-SVM model is [-5.6429, 2.3343%] and RMSE and MAPE is 1.275 and 1.129%, respectively, which shows that its accuracy and dispersion degree of errors can also meet the demand of practice works. Contrasting RVM model with PSO-SVM model, the former has better quality in short-term load prediction. As for GA-SVM model, its predictive accuracy is slightly inferior to those of the two algorithms in this study. The results and errors comparison are respectively listed in Fig. 2 and 3.

### Table 1: Predict results of three models

<table>
<thead>
<tr>
<th>No.</th>
<th>Actual load</th>
<th>Error of GA-SVM(%)</th>
<th>Error of PSO-SVM(%)</th>
<th>Error of RVM(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>940</td>
<td>1.4885</td>
<td>0.87276</td>
<td>1.4698</td>
</tr>
<tr>
<td>2</td>
<td>921</td>
<td>2.1837</td>
<td>1.3934</td>
<td>1.4869</td>
</tr>
<tr>
<td>3</td>
<td>950</td>
<td>1.1298</td>
<td>0.6318</td>
<td>1.4611</td>
</tr>
<tr>
<td>4</td>
<td>925</td>
<td>2.0344</td>
<td>1.2765</td>
<td>1.4832</td>
</tr>
<tr>
<td>5</td>
<td>918</td>
<td>2.2965</td>
<td>1.4859</td>
<td>1.4896</td>
</tr>
<tr>
<td>6</td>
<td>917</td>
<td>2.3434</td>
<td>1.5146</td>
<td>1.4905</td>
</tr>
<tr>
<td>7</td>
<td>1012</td>
<td>-0.92076</td>
<td>-0.48104</td>
<td>1.4111</td>
</tr>
<tr>
<td>8</td>
<td>1007</td>
<td>-2.5403</td>
<td>-1.1541</td>
<td>1.3715</td>
</tr>
<tr>
<td>9</td>
<td>1061</td>
<td>-2.3718</td>
<td>-1.0865</td>
<td>1.3756</td>
</tr>
<tr>
<td>10</td>
<td>1043</td>
<td>-1.8516</td>
<td>-0.87826</td>
<td>1.3883</td>
</tr>
<tr>
<td>11</td>
<td>1057</td>
<td>-2.2584</td>
<td>-1.041</td>
<td>1.3784</td>
</tr>
<tr>
<td>12</td>
<td>1127</td>
<td>-4.1208</td>
<td>-1.8554</td>
<td>1.3328</td>
</tr>
<tr>
<td>13</td>
<td>1004</td>
<td>-0.67041</td>
<td>-0.36603</td>
<td>1.4172</td>
</tr>
<tr>
<td>14</td>
<td>926</td>
<td>1.9973</td>
<td>1.2479</td>
<td>1.4823</td>
</tr>
<tr>
<td>15</td>
<td>951</td>
<td>1.0946</td>
<td>0.68088</td>
<td>1.4603</td>
</tr>
<tr>
<td>16</td>
<td>964</td>
<td>0.6437</td>
<td>0.32811</td>
<td>1.4493</td>
</tr>
<tr>
<td>17</td>
<td>1009</td>
<td>-0.82735</td>
<td>-0.43865</td>
<td>1.4133</td>
</tr>
<tr>
<td>18</td>
<td>1091</td>
<td>-3.1959</td>
<td>-1.424</td>
<td>1.3555</td>
</tr>
<tr>
<td>19</td>
<td>1191</td>
<td>-5.6429</td>
<td>-2.8402</td>
<td>1.2958</td>
</tr>
<tr>
<td>20</td>
<td>1142</td>
<td>-4.4974</td>
<td>-2.0554</td>
<td>1.3247</td>
</tr>
<tr>
<td>21</td>
<td>1090</td>
<td>-3.1691</td>
<td>-1.4126</td>
<td>1.3552</td>
</tr>
<tr>
<td>22</td>
<td>1086</td>
<td>-3.6017</td>
<td>-1.3673</td>
<td>1.3588</td>
</tr>
<tr>
<td>23</td>
<td>1057</td>
<td>-2.2584</td>
<td>-1.041</td>
<td>1.3784</td>
</tr>
<tr>
<td>24</td>
<td>1000</td>
<td>-0.54373</td>
<td>-0.0030604</td>
<td>1.4203</td>
</tr>
<tr>
<td>RMSE (%)</td>
<td>2.562</td>
<td>1.275</td>
<td>1.412</td>
<td></td>
</tr>
<tr>
<td>MAPE (%)</td>
<td>2.270</td>
<td>1.129</td>
<td>1.410</td>
<td></td>
</tr>
</tbody>
</table>

where, X is raw data set, X<sub>max</sub> is the maximum value and X<sub>min</sub> is the minimum value. Y represents processed data set, Y<sub>max</sub> is the maximum value (Y<sub>max</sub> = 1), Y<sub>min</sub> is the minimum value (Y<sub>min</sub> = -1). When X<sub>max</sub> = X<sub>min</sub>, Y = X.

CONCLUSION

In this study, we explores improved SVR based on PSO and RVR model for STLF and do comparative analysis for their predictive performance. First, Improved SVR based on PSO appears good performance in predictive accuracy, by contrast, the results of RVM model is more reliable and slightly better than those achieved by PSO-SVM model. Second, RVR model don't need set hyper-parameters which is super to SVR model.
in this regard. Another advantages is that its programming is simpler and has high implementation efficiency. Especially, to improved SVR model, it need to set parameter via PSO algorithm, which will make programming difficult and cause low implementation efficiency. Third, RVR model shows good performance in handling small-sample data., but its performance will decrease if data exceed a certain amount. It will need further study to solve this problem. At last, there are many factors that would affect short load patterns such as temperature and humidity. Unfortunately, raw data of these influencing factors are actually difficult to obtain. This is why we only used load data itself for prediction in this study. Sometimes, these influencing factors will straightforward affect predictive accuracy of forecasting models. Hence, on the foundation of previous study and the further considering many influencing factors, new forecasting models based on RVR and SVR should be next research objective.

ACKNOWLEDGMENTS

Project supported by National Natural Science Fund of China (No. 71071054) and the Fundamental Research Funds for the Central Universities of China (No. 11QR34).

REFERENCES


