http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (15): 3209-3213, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3209.3213

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 parameteróof 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.

Key words: 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 andover-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 useevolutionary 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

Corresponding Author: Guo Xiaopeng, School of Economics and Management, North China Electric Power University, Beijing 102206, China

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 nonlinearly 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 $D = \{(x_i, y_i) \mid i = 1, 2, \ldots, n)\}$, 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.

$$\begin{aligned} &\min\left\{\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\left(\xi_i^* + \xi_i\right)\right\} \\ &\text{s.t.}\quad y_i - \omega\cdot\phi(x) - b \le \varepsilon + \xi_i^*; \\ &\qquad \omega\cdot\phi(x) + b - y_i \le \varepsilon + \xi_i; \\ &\qquad \xi_i, \xi_i^* \ge 0 \end{aligned} \tag{1}$$

where, c is a positive constant as regularization parameter, i.e., a penalty factor, ε is the insensitive lost function and ξ_i^* and ξ_i are slack variables. ξ_i^* is denoted as follows:

$$\xi_i^* = \begin{cases} 0, & |f(x) - y_i| \le \varepsilon \\ |f(x) - y_i| - \varepsilon, & |f(x) - y_i| > \varepsilon \end{cases} \tag{2}$$

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

$$f(x) = \sum_{i=1}^{n}(a_i - a_i^*)K(X_i, X) + b \tag{3}$$

where, α_i, α_i^*(with $0\le \alpha_i, \alpha_i^* \le C$) 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:

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{4}$$

Parameter selection by PSO: Penalty factor C, insensitive lost function and parameter σ of Kernel function decide the performance of SVR model. Among them, σ 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 SVR 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 Bayesian 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 HH, 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_i and t_i can be written as:

$$t = y(x; w) = \sum_{i=1}^{N} \omega_i K(x, x_i) + \omega_0 \tag{5}$$

where, w is weight vector, K (X, X_i) is kernel function and ω 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_i, the likelihood estimation of the training data set $\{x_i, t_i\}_{i=1}^N$ can be written as:

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

In Eq. 6, $t = (t_1, \ldots, t_n)^T$, $W = [\omega_0, \omega_1, \ldots, \omega_N]^T$ and $\Phi = [\phi(x_1), \phi(x_2), \ldots, \phi(x_n)]^T$. In $\Phi = [1, K(x_i, x_1) K(x_i, x_2), \ldots, K(x_i, x_n)]^T$ In order to solve over-fitting problem in Eq. 6, Tipping defined a vector α of N+1-dimensional hyper-parameter and assign a zero-mean Gaussian prior distribution over w:

$$p(w|\alpha) = \prod_{i=0}^{N} \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_i \omega_i^2}{2}\right) \tag{7}$$

According to Eq. 7, the one-to-one corresponding relation between W and α is built in order to control the influence of priori distribution on α, 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:

$$p(w|t, \alpha, \sigma^2) = (2\pi)^{-\frac{N+1}{2}} |\Sigma|^{-\frac{1}{2}} \bullet \exp\left\{-\frac{(w-\mu)^T \Sigma^{-1} (w-\mu)}{2}\right\} \tag{8}$$

where, $\Sigma = (\sigma^{-2}\Phi^T\Phi + A)^{-1}$, $\mu = \sigma^{-2}\Sigma\Phi^T t$ and α = DIAG ($\alpha_0, \alpha_1, \ldots, \alpha_n$). Values of α and σ which maximize Eq. 8 cannot be obtained in closed form and Tipping used the Eq. 9 and 10 to get their most probable values α_{MP} and σ_{MP}^2 by iterative computation:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2}, (\sigma^2)^{new} = \frac{\|t - \Phi\mu\|^2}{N - \sum_{i=0}^{N} \gamma_i} \tag{9}$$

$$\gamma_i = 1 - \alpha_i \Sigma_{i,i} \tag{10}$$

where, μ_I is the i-th posteriori mean weight and $\Sigma_{I,j}$ is the i-th diagonal element of the posterior weight covariance matrix Σ. If we put α_{MP} and σ_{MP}^2 into Eq. 8, the output probability distribution of input value x can be written as $p(t_*|t, \alpha_{MP}, \sigma_{MP}^2) = M(t_*|y_*, \sigma^2)$ and then the forecasting value of t_* is $y^* = \mu^T \phi(x_*)$.

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 (6) Build RVM forecasting model.

The procedure of SVR-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 σ 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. SVR 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:

$$Y = (Y_{max} - Y_{min}) \times \frac{X - X_{max}}{X_{max} - X_{min}} + Y_{min}$$
$$\text{s.t. } X_{max} \neq X_{min}$$

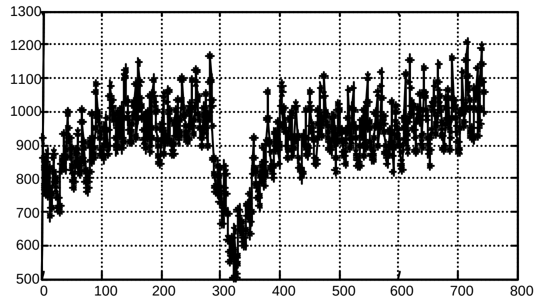

Fig. 1: Raw load data

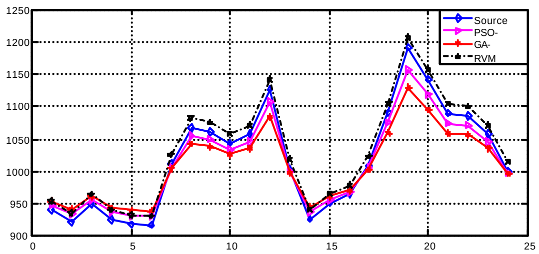

Fig. 2: Comparison of predictive result from three different models and raw data

Table 1: Predict results of three models

No.	Actual load	Error of GA-SVM(%)	Error of PSO-SVM (%)	Error of RVM (%)
1	940	1.4858	0.87276	1.4698
2	921	2.1837	1.3934	1.4869
3	950	1.1298	0.6318	1.4611
4	925	2.0344	1.2765	1.4832
5	918	2.2965	1.4839	1.4896
6	917	2.3343	1.5146	1.4905
7	1012	-0.92076	-0.48104	1.4111
8	1067	-2.5403	-1.1541	1.3715
9	1061	-2.3718	-1.0865	1.3756
10	1043	-1.8546	-0.87826	1.3883
11	1057	-2.2584	-1.041	1.3784
12	1127	-4.1268	-1.8554	1.3328
13	1004	-0.67041	-0.36603	1.4172
14	926	1.9973	1.2479	1.4823
15	951	1.0946	0.60885	1.4603
16	964	0.6437	0.32811	1.4493
17	1009	-0.82735	-0.43865	1.4133
18	1091	-3.1959	-1.424	1.3555
19	1191	-5.6429	-2.8402	1.2958
20	1142	-4.4974	-2.0554	1.3237
21	1090	-3.1691	-1.4126	1.3562
22	1086	-3.0617	-1.3673	1.3588
23	1057	-2.2584	-1.041	1.3784
24	1000	-0.54373	-0.0030604	1.4203
RMSE (%)	2.562	1.275	1.412	
MAPE (%)	2.270	1.129	1.410	

where, X is raw data set, X_{max} is the maximum value and X_{min} is the minimum value. Y represents processed data set, Y_{max} is the maximum value (Y_{max} = 1), Y_{min} is the minimum value (Y_{min}=-1). When X_{max} = X_{min}, Y = X.

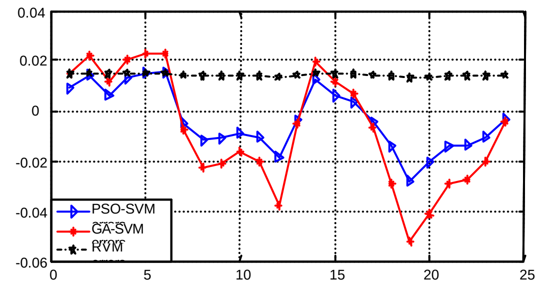

Fig. 3: Comparison of predictive error from three different models

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

$$e_{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N}},\ e_{MAPE} = \frac{\sum_{i=1}^{N} |e_i|}{N}$$

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.

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

Cui, J.F., J.X. Qi and S.D. Yang, 2009. Combined forecasting model based on BP improved byPSO and its application. J. Central South Univ. (Sci. Technol.), 40: 190-194.

Hong, W.C., 2011. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. Energy, 36: 5568-5578.

Kennedy, J. and R. Eberhart, 1995. Particle swarm optimization. Proc. IEEE Int. Conf. Neural Networks, 4: 1942-1948.

Li, C.B. and K.C. Wang, 2007. A new grey forecasting model based on BP neural network and Markov chain. J. Central South Univ. Technol., 14: 713-718.

Muller, K.R., A. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen and V. Vapnik, 1997. Predicting time series with support vector machines. Proceedings of the 7th International Conference on Artificial Neural Networks, October 8-10, 1997, Lausanne, Switzerland, pp: 999-1004.

Salcedo-Sanz, S., E.G. Ortiz-Garcia, A.M. Perez-Bellido, A. Portilla-Figueras and L. Prieto, 2011. Short term wind speed prediction based on evolutionary support vector regression algorithms. Expert Syst. Appl., 38: 4052-4057.

Shi, Y.Z., Y. Peng and H.C. Zhou, 2012. Research on mid-and long-term runoff forecast model with relevance vector machine. J. Dalian Univ. Technol., 52: 79-84.

Song, X.H., P.E. Zu, J. Yi and D. Liu, 2012. An optimally combined forecast model for long-term power demand based on improved grey and SVM model. J. Central South Univ., 43: 1803-1807.

Tan, Z.F., J.L. Zhang, L.Q. Wu, Y. Ding and Y. Song, 2011. A model integrating econometric approach with system dynamics for long-term load forecasting. Power Syst. Technol., 35: 186-190.

Tipping, M.E. and A.C. Faul, 2003. Fast marginal likelihood maximization for sparse Bayesian models. Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics, January 3-6, 2003, Florida, USA.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Machine Learn. Res., 1: 211-244.

Wu, C.H., G.H. Tzeng and R.H. Lin, 2009. A novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression. Expert Syst. Appl., 36: 4725-4735.