TELKOMNIKA, Vol.11, No.11, November 2013, pp. 6928~6934
e-ISSN: 2087-278X

Electricity Consumption Prediction based on SVR with Ant Colony Optimization

Haijiang Wang*, Shanlin Yang
School of Management of Hefei University of Technology, Anhui Hefei 230009
*Corresponding author, e-mail: tonysun800@sina.com

Abstract

Accurate forecasting of electric load has always been the most important issues in the electricity industry, particularly for developing countries. Due to the various influences, electric load forecasting reveals highly nonlinear characteristics. This paper creates a system for power load forecasting using support vector machine and ant colony optimization. The method of colony optimization is employed to process large amount of data and eliminate. The SVR model with ant colony optimization is proposed according to the characteristics of the nonlinear electricity consumption data. Then ACO-SVR model is applied to the electricity consumption prediction of Jiangsu province. The result shows better than the ANNs method and improves the accuracy of the prediction.

Keywords*: support vector regression (SVR), ant colony optimization (ACO)*

1. Introduction

Accurate electric load forecasting can provide those export oriented economies advantages through saving and efficiently distributing limited energy resources. For inaccurate electric load forecasting, it may increase operating costs. For example, over estimation of future electric load results in unnecessary spinning reserve, wastes limited energy resources, even leads to distribution inefficiency, and, furthermore, is not accepted by International energy networks owing to excess supply. In contrast, under estimation of load causes failure in providing sufficient reserve and implies high costs in the peaking unit, which discourage any economic and industrial developments. Thus, the accuracy of future electric demand forecasting have received growing attention, particularly in the areas of electricity load planning, energy expenditure/cost economy and secure operation fields, in regional and national systems.

With the higher demand for the quality of water and the acceleration of the industrialization, the need and consumption of electricity are becoming wider. In order to meet the massive requirement of industry, and to make the economy develop smoothly, it is very important to use and program electricity power. As a result, we need to predict the electric quantity, and then figure out the appointment of production plan, thus bring out the economic and social benefits.

Support Vector Machine, rendered by Vapnik in 1995 [1], can solve the phenomena of 'excess learning' with the principle of minimize the structure risk. It has great generalization capability. Applying the SVR in regression analysis, we get Support vector regression. Since the day of its birth, Experts have done lots of work to apply it into many fields and modify the model. For instance, in paper [2], the author use immune algorithm to optimize the parameters of SVR model to predict the use of electricity in Taiwan. In paper [6] the author blend the rough set theory and the reduction of property into LS-SVR, hence improving the accuracy of prediction; Paper [3] use Ant Colony Optimization to optimize trained data and speed up SVR. While paper [7] gets satisfactory results by modeling porous NiTi alloy with SVR. Also, the paper [8] gets great results by applying the LS-SVR into prediction of financial time series.

In this paper, we use the consumption of electricity and macro-economy influencing data from 2004 to 2009 in Jiangsu Province. By implementing the ant colony algorithm to optimize the parameters of SVR, we construct ACO-SVR. The results show that this model, with higher accuracy of prediction, is superior to BP-neural network both in fitting and error.

Received May 28, 2013; Revised July 29, 2013; Accepted August 9, 2013

2. Principle of SVR

Assume that:

$$f(x) = w\phi(x) + b \tag{1}$$

$\phi(x)$ is non-linear transformation, which convert the dataset x into high dimension characteristics space F, w is called weight vector, b is classification threshold. We can see $f(x)$ is a linear function of $\phi(x)$。Under the principle of minimum structure risk, the desired $f(x)$ should satisfy:

$$R(f) = \frac{1}{2}\|w\|^2 + \frac{C}{N}\sum_{i=1}^{n} L(y_i, f(x_i)) \tag{2}$$

In which, C is a positive constant. It is a compromise factor in smooth and experience error, which is also called penalty factor. $L(y_i, f(x_i))$ is loss function. Usually, we let the loss function be the ε non-sensitive loss function, which means for $i = 1, 2, \cdots, n$,

$$L(y_i, f(x_i)) = L_\varepsilon(y_i, f(x_i)) = \begin{cases} 0, & |y_i - f(x_i)| \le \varepsilon \\ |y_i - f(x_i)| - \varepsilon, & \text{ot her s} \end{cases} \tag{3}$$

Thus, the above regression solving problem becomes the optimization problem as follows:

$$\min R(w, \xi_i, \xi_i^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{4}$$

$$s.t.\begin{cases} y_i - f(x_i) \le \varepsilon + \xi_i \\ f(x_i) - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1, 2, \cdots, n \end{cases}$$

In order to solve (4), we define Lagrange function:

$$\begin{aligned} L(w,b,\xi_i,\xi_i^*,\alpha_i,\beta_i,\mu_i,\upsilon_i) &= \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ &+ \sum_{i=1}^{n}\alpha_i(y_i - w\phi(x_i) - b - \varepsilon - \xi_i) + \sum_{i=1}^{n}\beta_i(w\phi(x_i) + b - y_i - \varepsilon - \xi_i^*) \\ &- \sum_{i=1}^{n}\mu_i\xi_i - \sum_{i=1}^{n}\upsilon_i\xi_i^* \end{aligned} \tag{5}$$

For partial derivative of $L(w,b,\xi_i,\xi_i^*,\alpha_i,\beta_i,\mu_i,\upsilon_i)$ variable w, b, ξ_i, ξ_i^*, by letting the result be zero, we get:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}\alpha_i\phi(x_i) + \sum_{i=1}^{n}\beta_i\phi(x_i) = w - \sum_{i=1}^{n}(\alpha_i - \beta_i)\phi(x_i) = 0 \tag{6}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n}\beta_i - \sum_{i=1}^{n}\alpha_i = 0 \tag{7}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \tag{8}$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \beta_i - \upsilon_i = 0 \tag{9}$$

Through taking (6)~(9) to (5), while at the same time assuming that kernel function $K(x_i, x_j) = \phi(x_i) \bullet \phi(x_j)$, and changing theoptimization problem into its own antithesis problem, we have:

$$\max \sum_{i=1}^{n} y_i(\alpha_i - \beta_i) - \varepsilon \sum_{i=1}^{n}(\alpha_i + \beta_i) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \beta_i)(\alpha_j - \beta_j)K(x_i, x_j)$$
$$s.t. \begin{cases} \sum_{i=1}^{n} \beta_i - \sum_{i=1}^{n} \alpha_i = 0 \\ \alpha_i, \beta_i \in [0, C], i = 1, 2, \cdots, n \end{cases} \tag{10}$$

To solve quadratic programming (10),

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \beta_i)\phi(x_i) \bullet \phi(x) + b = \sum_{i=1}^{n}(\alpha_i - \beta_i)K(x_i, x) + b \tag{11}$$

In which,

$$w = \sum_{i=1}^{n}(\alpha_i - \beta_i)\phi(x_i) \tag{12}$$

Due to kernel function [1] of satisfying Mercer, this means corresponding to a group of dot product in high dimension space. What we have to know is the specific function that meets this condition.

Then we can get the regression function $f(x)$ even though we have no idea of specific formulation of $\phi(x)$. Here we let the kernel function be radial basis function,

$$K(x, x_i) = \exp(-|x - x_i|^2 / \sigma^2) \tag{13}$$

3. ACO for Feature Selection

The ant colony optimization algorithm (ACO) provides an alternative feature selection tool inspired by the behavior of ants in finding paths from the colony to food. Real ants exhibit strong ability to find the shortest routes from the colony to food using a way of depositing pheromone as they travel. ACO mimic this ant seeking food phenomenon to yield the shortest path (which means the system of interests has converged to a single solution).Different equally shortest paths can exist. An ACO algorithm can be generally applied to any combinatorial problem as far as it is possible to define:

Firstly, the problem can be described in a set of nodes and edges between nodes to form a graph. So the problem can be seen easily to find the main problem and overcome it.

Secondly, heuristic desirability of paths: it is a suitable heuristic measure which can find better paths from one node to every other connected node, and it can be described in a graph details.

Thirdly, construction of solutions: a feasible and complete solution of the formulated inter-cell layout problem is considered as a permutation of manufacturing cells. Each part of this solution is termed state. In the optimum process, each ant initially assigns a cell to location 1 then assigns another cell to location 2 and so on till a complete solution is obtained.

Fourthly, pheromone updating rule: firstly, it is the area pheromone updating rule, the effect is to make the desirability of edges change dynamically in order to shuffle the tour. The nodes in one ants tour will be chosen with a lower probability in building other ants tours. As a consequence, ants will favor the exploration of edges not yet visited and prevent converging to a common path.

Next, go to global updating rule, this process is performed after all ants have completed their tours. Therefore, only the globally best ant that found the best solution up to the current iteration of the algorithm is permitted to deposit pheromone.

Fifthly, probabilistic transition rule: the rule determines the probability of an ant traversing from one node in the graph to the next. The heuristic desirability of traversal and edge pheromone levels is combined to form the so-called probabilistic transition rule.

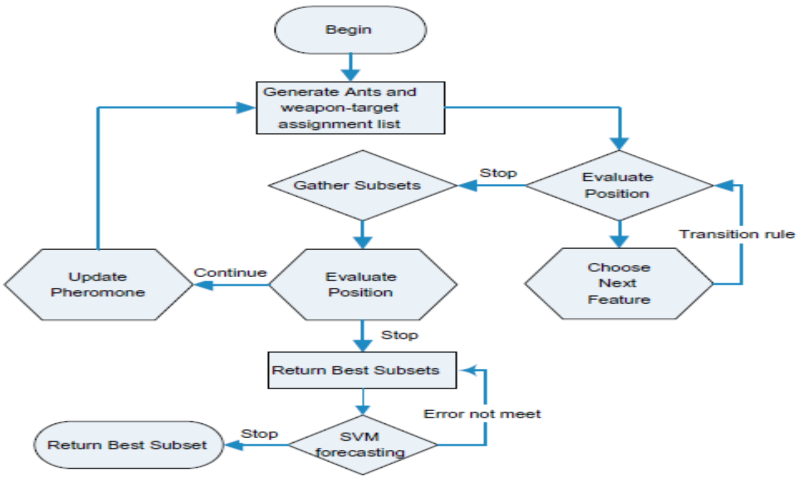

Figure 1. ACO-SVR Forecasting Process

4. Modeling and Prediction

4.1. Data Choosing and Pre-disposing

In this paper, we choose the influencing factors data and the overall electricity consumption of Jiangsu Province from January 2004 to October 2009. And we consider the data of January 2004 to July 2009 to be the trained sets, thus construct ACO-SVR model. We think the data from August to October 2009 are the test sets.

Meanwhile, in order to eliminate of dimension influence, we apply the standard 0-1, that is:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (14)$$

So the new data sets are all in $[0,1]$, and the data sets also eliminate the diversity units, which interfere the results.

As the error ε of non-sensitive loss function is too small, even though it can enhance the accuracy of trained model, it reduces generalization ability. On the contrary, if ε is too bigger, the constructed model hardly can depict the change of electricity quantity. After repeated experiment, we define ε to 0.01. It can both guarantee the fitting accuracy and the generalization ability of model.

4.2. Result of Regression Forecasting of the Model

Due to electricity consumption trends fluctuate by influencing factors. Thus we choose the influencing factors include Average Month Temperature (AMT), Social Retail Sales of Consumer Goods (SRSCC), Industry Increasing Value (IIV), Consumer Price Index (CPI) and Gross Value of Export-Import (GVEI). Hence, we select five influencing factors as input variable. Considering there has multicollinearity between the influencing factors. This paper use LS to extract the main component variable. Finally, we can get the main influencing include AMT, IIV and GVEI. As the error ε of non-sensitive loss function is too small, even though it can enhance the accuracy of trained model, and it can reduce generalization ability. On the contrary, if ε is too bigger, the constructed model hardly can depict the change of electricity consumption. After repeated experiment, we define ε to 0.01. It can both guarantee the fitting accuracy and the generalization ability of model.

4.3. Experiment Study

In this sub- section, we construct and analyze the experiment of the proposed model in this paper using matlab tools, the predictive result of regression forecasting using ACO-SVR forecasting model is showed in Figure 2.

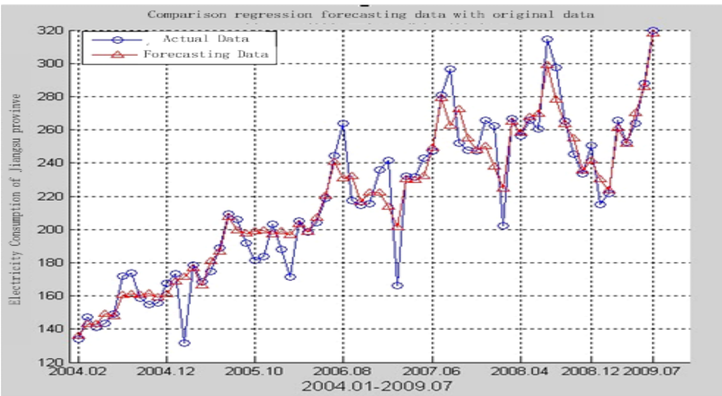

Figure 2. The Comparison of Forecasting Result

Besides, we give out the regression fitting results and Relative Errors (RE) from November 2007 to July 2009, the prediction results and relative errors from August to October 2009. It is showed in the Table 4.

Table 1. The Fitting Regression and Relative Errors

Month	Actual Value	The fitted values Value	Relative Error
2007.11	247.42	247.6876	0.08%
2007.12	265.77	251.3041	-5.82%
2008.01	262.59	237.8942	-9.41%
2008.02	202.29	224.3234	7.91%
2008.03	266.58	264.6345	-0.73%
2008.04	256.19	258.1616	0.77%
2008.05	265.68	267.5671	0.71%
2008.06	260.4	269.3890	3.42%
2008.07	314.72	299.1235	-4.97%
2008.08	297.54	277.5878	-6.61%
2008.09	265.34	263.3123	-0.73%
2008.10	245.21	254.6784	3.93%
2008.11	233.35	235.1435	0.81%
2008.12	250.74	241.1235	-3.59%
2009.01	214.96	230.4677	7.01%
2009.02	221.54	223.1236	0.85%
2009.03	265.87	260.7899	-1.84%
2009.04	252.39	251.1234	-0.36%
2009.05	263.95	270.1237	2.37%
2009.06	287.78	285.1279	-0.68%
2009.07	319.54	317.6724	-0.58%

Where the relative error is $RE = (\hat{y}_i - y_i)/y_i$, y_i is the real value, $\hat{y}_i$ is the predicted value.

From the result of Table 1, ACO-SVR leads to a satisfactory result of electricity consumption from August to October 2009, and the relative errors confines in 10%. Moreover, the relative errors of last two months are confines to 5%.

In order to assess the rationality of the model is proposed in this paper. We compare the ACO-SVR model with the PSO-SVR model and LS-SVR, the analysis result is showed as Figure 3.

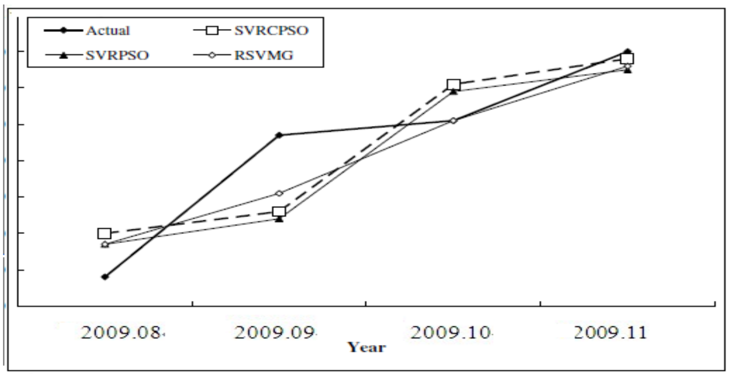

Figure 3. Relative Error Comparison with other Forecasting Method ($\varepsilon = 0.01$, $K = 3$)

Where:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \ ; \ R^2 = \frac{(\sum_{i=1}^{n}(y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}}))^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})^2},$$

y_i is the actual value, $\hat{y}_i$ is the predictive value, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i$.

MSE measures the deviation of predictive value from the actual value. The MSE is smaller shows that deviation degree is smaller, and the predictive precision of the forecasting model is more accurate.

5. Conclusion

The research of this paper is based on SVR to construct model and predict. We solve the problems of picking parameters. We use the partial swarm optimization to find out approximate optimal value and at the same time, use the cross validation to lower predicted errors and construct ACO-SVR model. By doing this, on one hand, SVR can solve the non-linear, big volatility problem. On the other hand, the parameters choosing problems can also be solved by partial swarm optimization. From the final result, we can draw a conclusion confidently that the model we build, with higher accuracy, is superior to BP-neural network

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China Grant No. 71101041 and No.71071045, National 863 Project Grant No. 2011AA05A116, Foundation of Higher School Outstanding Talents No. 2012SQRL009 and National Innovative Experiment Program No. 111035954. Thanks for the help.

References

[1] Soohyun Oh, Jin Kwak. Mutual Authentication and Key establishment mechanism using DCU certificate in Smart Grid. *Applied Mathematics & Information Sciences.* 2012; 3(1): 257-264.
[2] Vladimir N Vapnik. The Nature of Statistic Learning Theory. New York: Springer. 1995.

[3] Wei-Chiang Hong. Electric load forecasting by support vector model. *Applied Mathematical Modelling*. 2009; 33: 2444-2454.
[4] Kadir Kavaklioglu. Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. *Applied Energy*. 2011; 88: 368-375.
[5] Aldo Goia, Caterina May, Gianluca Fusai. Functional clustering and linear regression for peak load forecasting. *International Journal of forecasting*. 2010; 26: 700-711.
[6] Diyar Akay, Mehmet Atak. Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. Energy. 2007; 32: 1670-1675.
[7] Nima Amjady, Farshid Keynia. A New Neural Network Approach to Short Term Load Forecasting of Electrical Power Systems. Energies. 2011; 4: 488-503.
[8] Barker M, Rayens W. Partial least squares for discrimination. *Journal of Chemometrics*. 2003; 17: 166-173.
[9] Vladimir N Vapnik. The Nature of Statistic Learning Theory. New York, Springer. 1995.
[10] Kim K. Financial time series forecasting using support vector machines. *Neurocomputing*. 2003; 55: 07-319.