TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.4, April 2014, pp. 3068 ~ 3077
DOI: http://dx.doi.org/10.11591/telkomnika.v12i4.4834

The Predictive Method of Power Load Based on SVM

LV Feng[1], KANG Fengning*[2], SUN Hao[3]
[1]Electrical Department, Hebei Normal University, Shijiazhuang, Hebei, China Ph: 13111517578
[2]Shijiazhuang Tiedao University, Shijiazhuang, Hebei, China Ph: 15530900186
[3]Scheduling Station Shijiazhuang Power Supply Company, Shijiazhuang, Hebei, China Ph: 15833116990
Corresponding author, e-mail: lvfeng@mail.tsinghua.edu.cn[1], kangfengning@163.com*[2],
sunhaoku@163.com[3]

Abstract

Because of the problems of nonlinear and high dimension recognition existed in the power load prediction, support vector machine (SVM) is applied to the power load prediction technology based on the analysis of SVM and power load prediction theory. The predictive model of the power load is established based on SVM, which overcomes the problems of dimension disaster, over-learning, etc. The new thought of power load prediction model is more accurate, intelligent and humanized than the other method. The predictive result is more accurate by the Matlab simulation which used the historical load data from a power grid company to establish the predictive model

Keywords: *power load prediction, support vector machine (SVM), regression prediction, mesh optimization*

1. Introduction

Power load prediction is defined to predict the trend of the development in the future by the historical data of the power and the power consumption from the whole of national economy, department or areas. This prediction is a kind of scientific and reasonable inference for the power load development by the use of reliable and comprehensive methods and means under the guidance of scientific theory and the precise basis of research data. Power load prediction is the foundation of safe and economic dispatch, planning and design research in the power system. Accurate load prediction would mean the maximum use of electric power system construction fund and the acquisition of the maximum social and economic benefit under the limited investment on the premise of meet the requirements of power supply quality. The results showed that: In the UK, each 1% increase in short-term load prediction error will result in an annual increase of about 17.7 million yuan in operating costs; In Norway, every 1% increase in short-term load prediction will result in an additional 4.55~9.1 million yuan operating costs. Since the uncertainty of load prediction, the load prediction depends on not only the prediction model, but also the staff with rich experience and profound knowledge of the load forecasting. At present, the methods of power load prediction can be divided into parameter estimation method and the artificial intelligence methods, and the parameter estimation method including regression method and gray system method [1-6]; Artificial intelligence method includes evidence theory, regression trees, neural network prediction method, as well as support vector machines and other intelligent prediction method [7-11]. As a novel machine learning method, SVM (Support Vector Machine-SVM) performs many unique advantages in solving the small sample, nonlinear and high dimensional pattern recognition problems and largely overcome the "dimension disaster ", the "over-learning "and other problems [12, 13]. This paper focuses on the method for the power load prediction based on SVM.

2. Support Vector Regression

SVM was first designed and applied to the classification problem, but it can also be applied to regression [14, 15]. The biggest difference between classification and regression is: in the classification problem, but in regression problems, and the data of regression is no category. When the regression problem is handled by used of the support vector machine, a

Received September 2, 2013; Revised November 10, 2013; Accepted December 14, 2013

hyperplane based linear regression is built in low-dimensions feature space, which makes the sample point falls on both ends of the hyperplane. Within the two boundary lines, seeking the optimal regression hyperplane will be equivalent to finding maximum interval.

2.1. ε-band

Let $\varepsilon > 0$, the ε-band of a hyperplane $y = \omega \cdot \mathbf{x} + b$ means the area that the hyperplane along the y-axis successively swept up and down the translation, as shown in Figure 1.

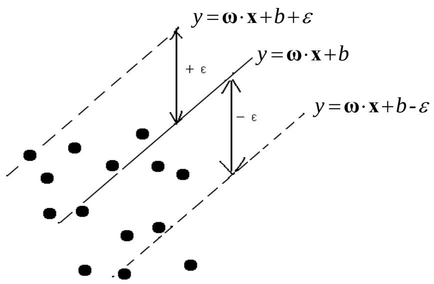


Figure 1. ε-band Schematic Diagram

The given training data set,

$$T = \{(x_1, y_1), \cdots (x_l, y_l)\} R \times R \qquad \varepsilon > 0 \tag{1}$$

If a ε-band of a hyperplane $y = \omega \cdot \mathbf{x} + b$ can contain all the training points of data set *T*, then the hperplane $y = \omega \cdot \mathbf{x} + b$ is called a hard ε-band hyperplane and satisfies:

$$-\varepsilon \le y - (\omega \cdot \mathbf{x}_i) + b \le \varepsilon, i = 1, 2, \cdots l \tag{2}$$

2.2. Linear Hard ε-band Support Vector Regression

The initial optimization problem for linear hard ε- band support vector regression:

$$\begin{cases} \min\limits_{\omega} \dfrac{\|\omega\|^2}{2} \\ s.t. (\omega \mathbf{x}_i + b) - y_i \le \varepsilon, i = 1, 2, \cdots l \\ y_i - (\omega \mathbf{x}_i + b) \le \varepsilon, i = 1, 2, \cdots l \end{cases} \tag{3}$$

The Lagrange multiplier method is introduced in order to solve the quadratic optimization problem. After constructed the Lagrange function, partial derivative of ω and b is gotten and made to 0, thus the dual problem is gotten as follows:

$$\min_{\alpha^{(*)}} \frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(\mathbf{x}_i \cdot \mathbf{x}_j) + \varepsilon \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \tag{4}$$

The constraint is following:

$$s.t. \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0,$$
$$\alpha_i^{(*)} \ge 0, i = 1, 2, \cdots, l$$

The solution to the dual problem is:

$$\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \cdots, \bar{\alpha}_l, \bar{\alpha}_l)^T \quad (5)$$

Let $\bar{\omega} = \sum_{i=1}^{l} (\bar{\alpha}^* - \bar{\alpha})\mathbf{x}_i$, according to the KKT condition:

If the positive component of $\bar{\alpha}$ is chose, $\bar{\alpha}_j \geq 0$, then:

$$\bar{b} = y_j - (\bar{\omega} \cdot \mathbf{x}_j) - \varepsilon \quad (6)$$

If the positive component of $\bar{\alpha}^*$ is chose, $\bar{\alpha}_j^* \geq 0$, then:

$$\bar{b} = y_j - (\bar{\omega} \cdot \mathbf{x}_j) - \varepsilon \quad (7)$$

Hence, $\bar{\omega}$ and $\bar{b}$ is the solution of the original optimization problem, $y = (\bar{\omega} \cdot \mathbf{x}) + \bar{b}$ is the hard ε-band hyperplane, eventually a regression approximate function of the hard ε-band linear support vector regression machine can be built as the following:

$$\sum_{i=1}^{l} (\alpha_i^* - \alpha_i)(\mathbf{x}_i \cdot \mathbf{x}) + b \quad (8)$$

2.3. Hard ε-band Support Vector Regression

By the using of kernel function technology (Radial Basis Function-RBF), the inner product $(\mathbf{x}_i \cdot \mathbf{x})$ is mapped to the Hilbert space through a kernel function $K(\mathbf{x}, \mathbf{x}_i)$, so the hard-ε-band linear support vector regression is extended to be able to solve non-linear regression with a hard ε-band support vector regression machine. The Lagrangian function is constructed by the hard ε-band support vector regression as follows:

$$\min_{\alpha^{(*)}} \frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(\mathbf{x}_i \cdot \mathbf{x}_j) + \varepsilon \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \quad (9)$$

The constraint is following:

$$s.t. \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0,$$

$$\alpha_i^{(*)} \geq 0, i = 1, 2, \cdots, l$$

Optimal solution can be obtained:

$$\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \cdots, \bar{\alpha}_l, \bar{\alpha}_l)^T \quad (10)$$

$$\bar{b} = y_j - \sum_{i=1}^{l} (\bar{\alpha}_i^* - \bar{\alpha}_i) K(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \quad (11)$$

Thus regression approximation functions are constructed:

$$\sum_{i=1}^{l} (\bar{\alpha}_i^* - \bar{\alpha}_i) K(\mathbf{x}_i, \mathbf{x}_j) + \bar{b} \quad (12)$$

2.4. Epsilon-support Vector Regression Machine

The "hard" regression limitation of the hard ε-support vector regression machine can be put aside by introducing slack variables $\xi^{(*)} = (\xi_1, \xi_1^*, \cdots, \xi_l, \xi_l^*)$ and penalty parameter C. When the division is incorrect, the slack variable $\xi^{(*)}$ is greater than 0, and when there is no error, $\xi^{(*)}$ takes 0. Penalty parameter $C > 0$ indicates the degree of punishment for the samples which exceeds the error ε. So the original optimization problem of ε-support vector regression is:

$$\begin{cases} \min\limits_{\omega \in R^n} \frac{1}{2}\|\omega\|^2 + C\frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^2) \\ s.t.((\omega \cdot \mathbf{x}_i) + b) - y_i \le \varepsilon + \xi_i, \\ i = 1, 2, \cdots, l \\ y_i - ((\omega \cdot \mathbf{x}_i) + b) \le \varepsilon + \xi_i^*, \\ i = 1, 2, \cdots, l \\ \xi_i^{(*)} \ge 0, i = 1, 2, \cdots, l \end{cases} \tag{13}$$

After constructing Lagrangian function, the optimization problem is:

$$\begin{cases} \min\limits_{\alpha^{(*)}} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(\mathbf{x}_i \cdot \mathbf{x}_j) \\ +\varepsilon\sum_{i=1}^{l}(\alpha_i^* - \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \\ s.t.\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, \\ \alpha_i^{(*)} \ge 0, i = 1, 2, \cdots, l \end{cases} \tag{14}$$

The optimal solution can be obtained through the formula (14):

$$\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \cdots, \bar{\alpha}_l, \bar{\alpha}_l)^T \tag{15}$$

The solution to ω is $\sum_{i=1}^{l}(\bar{\alpha}_i^* - \bar{\alpha}_i)\mathbf{x}_i$ .The final linear regression approximation function is:

$$\sum_{i=1}^{l}(\bar{\alpha}_i^* - \bar{\alpha}_i)K(\mathbf{x}_i, \mathbf{x}_j) + \bar{b} \tag{16}$$

$\bar{b}$ is calculated as:

If $\bar{\alpha}_j \in \left(0, \frac{C}{l}\right)$, then:

$$\bar{b} = y_j - \sum_{i=1}^{l}(\bar{\alpha}_i^* - \bar{\alpha}_i)K(x_i, x_j) + \varepsilon \cdot \tag{17}$$

If $\bar{\alpha}_k \in \left(0, \frac{C}{l}\right)$, then:

$$\bar{b} = y_k - \sum_{i=1}^{l}(\bar{\alpha}_i^* - \bar{\alpha}_i)K(x_i, x_k) + \varepsilon \tag{18}$$

3. The Simulation of Power Load Prediction by the using of SVM Model

3.1. Predicting Data and Modeling

The predictive fitting process of support vector regression model is shown in Figure 2. The figure consists of variable selection, data normalization, determination of parameters, SVM regression model, fitting prediction and performance evaluation. The simulation is some regional power grid, and the 4000 days' power load is used as the sample data sets that since January 1, 2000 to December 14, 2010. The 5-dimensional related data is taken as independent variables such as the daily electricity load, the weather of that day, the temperature and other meteorological data, the socio-economic and consumption index. The next day's load is chose as the dependent variable, so that a total of 3999 samples were used to build support vector regression model. The daily power load data was shown in Figure 3. Horizontal axis represents the number of days and the vertical axis represents the MW in the figure.

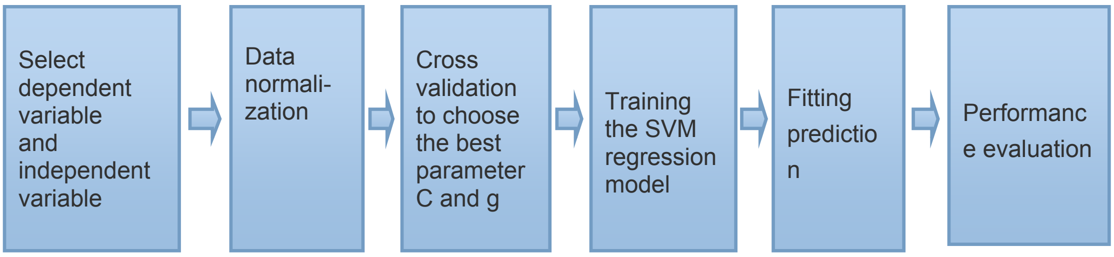


Figure 2. Regression Fitting Process

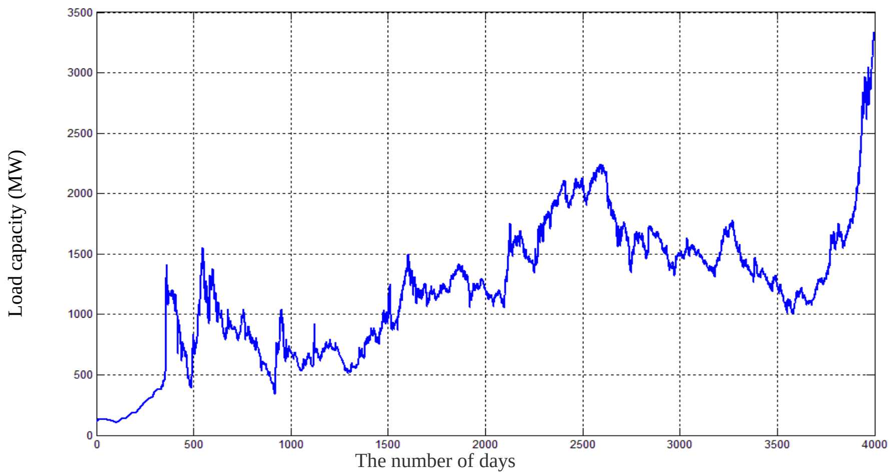


Figure 3. The Original Data of Daily Power Load

3.2. Data Preprocessing

Dependent and independent variables in the data set need to be normalized preprocessing in order to eliminate the differences between data dimension in the data set, making the curve smoother, eliminating the influence between data and improving the performance of SVM regression model. After such pretreatment, the dimensional data will become pure quantity. Usually, the normalized mapping is:

$$f : x \to y = \frac{(y_{max} - y_{min})(x - x_{min})}{x_{max} - x_{min}} + y_{min} \quad (19)$$

Among them, $[x_{min}, x_{max}]$ is the interval for the original data set, and $[y_{min}, y_{max}]$ is the normalized mapping range. According to the data characteristics of the paper, the independent variables and dependent variables are normalized to the interval $[1, 2]$, and the original normalized data set is shown in Figure 4.

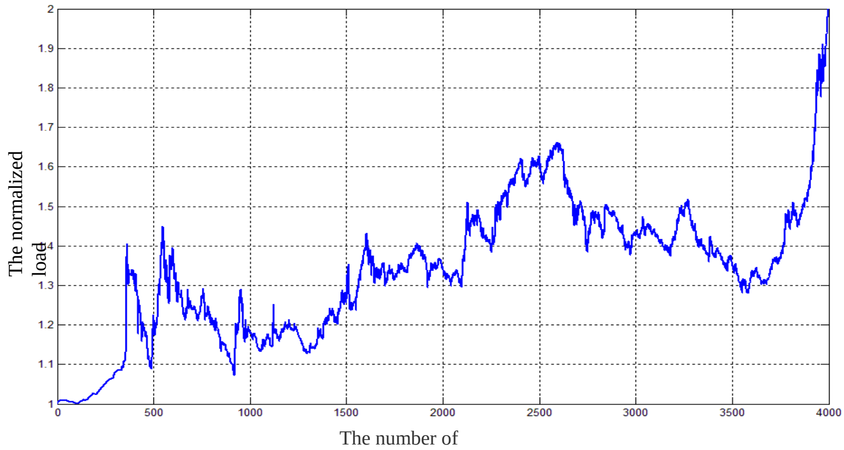


Figure 4. Normalized Original Data Set

3.3. The Establishment of Support Vector Machine Regression Model

3.3.1. Establish the ε-SVR Regression Simulation Experiment Environment

This simulation environment is built with the help of Matlab2012b and Libsvm Matlab toolkit. The Libsvm provides two kinds of Support Vector Classification such as $C-SVC$ and $\nu-SVC$, two kinds of support vector machine such as $\varepsilon-SVR$ and $\nu-SVR$, and $One-ClassSVM$ using in probability estimation. Among them, $\varepsilon-SVR$ and $C-SVC$ is the two most commonly used support vector regression and classification. Libsvm contains 3 kinds of commonly used kernel function as the following:

a. Polynomial Kernel:

$$K(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x} \cdot \mathbf{x}_i) + 1]^q \quad (20)$$

b. Radial basis function:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{|\mathbf{x} - \mathbf{x}_i|}{\gamma^2}\right\} \quad (21)$$

c. Sigmoid kernel function:

$$K(\mathbf{x}, \mathbf{x}_i) = S(\upsilon(\mathbf{x} \cdot \mathbf{x}_i) + 1) \quad (22)$$

The radial basis kernel function as the most widely used kernel function, effects perfectly when used in small sample data set modeling. Therefore, the Libsvm toolkit is used to construct $\varepsilon-SVR$ support vector regression model of RBF radial product kernel function.

3.3.2. The Problem How to Choose the Model Parameters

In the process of modeling with Libsvm, there are two parameters to determine: C and g. C is the penalty parameter involved in the optimization of the original problem solving.

That is:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l}\xi_i + C\sum_{i=1}^{l}\xi_i^* \tag{23}$$

Inside:

$$\begin{aligned} & Z_i - \omega^T\varphi(X_i) - b \le \varepsilon + \xi_i^*, \\ & \omega^T\varphi(X_i) + b - Z_i \le \varepsilon + \xi_i, \\ & \xi_i, \xi_i^* \ge 0, i = 1, \cdots, l. \end{aligned} \tag{24}$$

Among them, g is a coefficient parameter of RBF kernel function.
That is:

$$K(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2), \gamma > 0 \tag{25}$$

$$g = \gamma$$

The value of C and g can be found by stochastic method or experimental method, and the optimal value of C and g can be found by the grid search method. This paper chooses the latter. The final result by roughly optimizing is: In the sense of 3-fold cross validation, C=0.10882, g=1, MSE=0.0010881. The results are shown in Figure 5. From the rough parameter selection chart, the value of MSE is lower in the interval:

$$C \in [2^{\wedge}(-2), 2^{\wedge}2],$$

$$g \in [2^{\wedge}(-4), 2^{\wedge}4],$$

So another subdivision parameters optimization is made as well again in the interval :

$$C \in [2^{\wedge}(-2), 2^{\wedge}2]$$

$$g \in [2^{\wedge}(-4), 2^{\wedge}4]$$

The discrete steps as 0.5.

The parameter optimization range:

$$C \in \{2^{\wedge}(-2), 2^{\wedge}(-1.5), \cdots, 2^{\wedge}1.5, 2^{\wedge}2\}$$

$$g \in \{2^{\wedge}(-4), 2^{\wedge}(-3.5), \cdots, 2^{\wedge}3.5, 2^{\wedge}4\}$$

The results of segmentation optimization are as follows: in the 3-fold cross validation approach, C=0.25, g=0.35355, mean square error MSE=0.0010866, the results is increased than the rough optimization, as shown in Figure 6.

So, various parameters have been set up by using Libsvm. The value of p as the loss function is 0.01, and C and g is obtained by grid searching: C=0.25, g=0.35355. After setting up the model by using 4000 sample data set, the 4000 days' 5 dimensional data is input to the model. The output results is compared with the real results, that the regression results as shown in Figure 7, and the error is shown in Figure 8. Figure 8 (a) is the actual error, and Figure 8 (b) is the relative error.

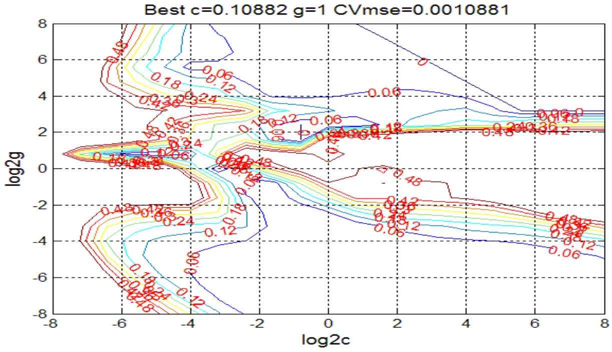


Figure 5. The Coarse Parameters Optimization Contour Map

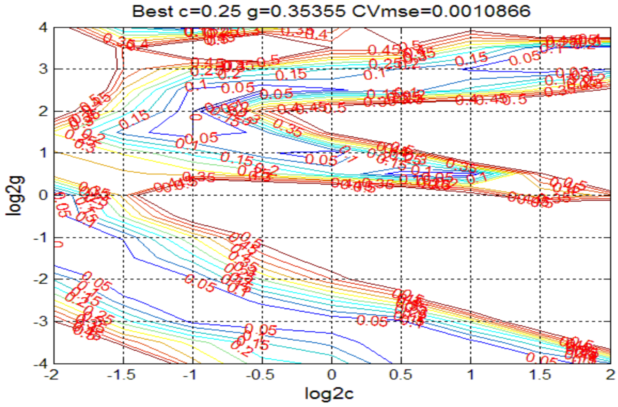


Figure 6. Fine Parameter Optimization Contour Map

The results of evaluation of the model is MSE=4.36979e-05, R2=0.998546. An ideal effect is achieved from the data set fitting results with the application of ε-SVM in the power load prediction. The regression fitting of the application of the model is successful.

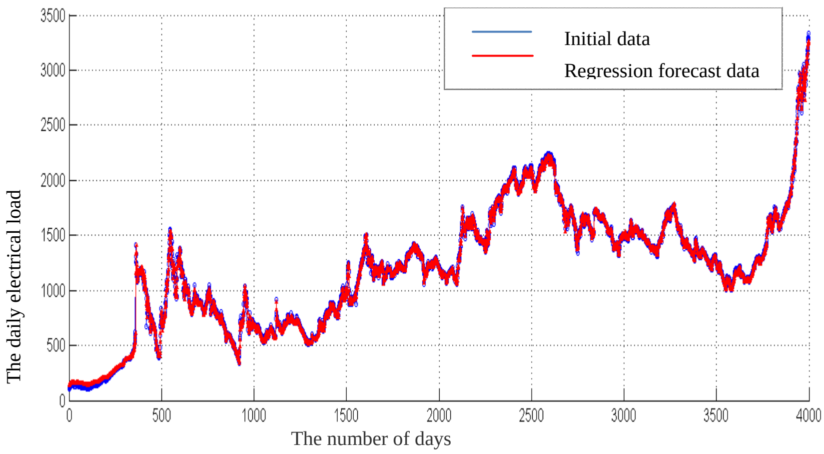
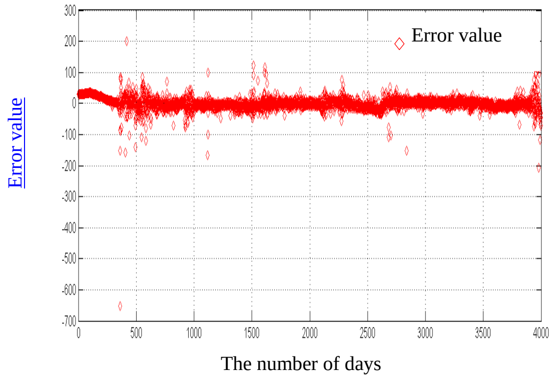
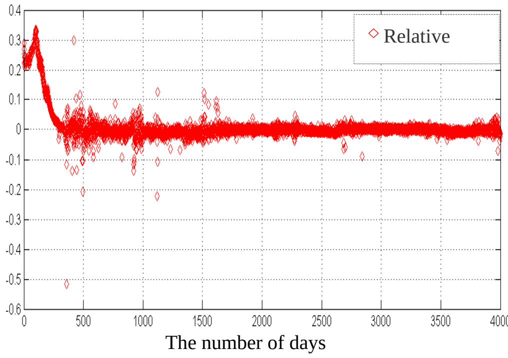


Figure 7. Regression Fitting Results Figure



(a) The actual error



(b) Relative error

Figure 8. Error Analyzing

4. Conclusion

Power load prediction scientificaly forecast the power load of some point or a certain time period in the future based on power load, socioeconomic, natural factors, a series of historical data recorded. It provides guiding significant theory basis for the power sector to explore the change rule of power load historical data and its influence on the future load, and to explore the power load and the intrinsic relationship between various influencing factors. Support vector machine (SVM) theory is applied to power load prediction technology based on the analysis of support vector machine (SVM) theory and the theory of power load forecasting. The classification problem of support vector machine (SVM) algorithm was the first returned to the application, and the hard linear binary classification algorithm based on low dimensional mathematical model was given, at last the simulation experiment results demonstrates the effectiveness of the algorithm. This method provides a new way for improving the accuracy of power load prediction system.

Acknowledgements

This paper is sponsored by the natural science foundation of China (NO: 60974063, 61175059), the nature science foundation of Hebei Province, the science and technology research projects funded by the Colleges and universities of Hebei (No: Q2012053), Hebei Normal University (L2011Q10).

References

[1] Zhu An (Wang Li). The adaptive model applied to time series of short term load forecasting. *Power system automation.* 1984; (2): 12-19.
[2] Liu Jinsheng. Railway Bureau Power Load Forecasting method research. *Railway Journal.* 1989; (1): 91-102.
[3] Zhou Qian. Intelligent engineering system and its application in power load forecasting. Doctoral Dissertation of North China Electric Power University. 2010.
[4] I Erkmen. Four methods for short-term load forecasting using the benefits of artificial intelligence. *Electrical Engineering*. 2003; 85: 29–33
[5] S Tzafestas. Computational intelligence techniques for short-term electric load forecasting. *Journal of Intelligent and Robotic Systems*. 2001; 31: 57–68
[6] Fang Rui-ming. Editors. Support Vector Machine Theory and its Application Analysis. Beijing: Electric Power Press. 2007.
[7] Pal M, Jayaraman VK, Kulkarni BD. Support Vector Machines for Classification in Remote Sensing. *International Journal of Remote Sensing*. 2005; 26(5): 1007-1011.
[8] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin. A Practical Guide to Support Vector Classification. Department of Computer Science National Taiwan University. Taiwan 2010.

[9] Wahyuningrum, Rima Tri. Efficient Kernel-based Two-dimensional Principal Component Analysis for Smile Stages Recognition. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2012; 10(1): 113-120.
[10] Li Xiao-yu, Zhang Xin-feng, Shen Lan-sun. A Selection Means on the Parameter of Radial Basis Function. *Acta Electronica Slnica*. 2005: 33(1): 2459-2463.
[11] Wang Xing-ling, Li Zhan-bin. *Identifying the Parameters of the Kernel Function in Support Vector Machines based on the Grid-Search Method*. Periodical of Ocean University of China. 2005: 53(5): 413-418.
[12] Tian Yingjie. *Support vector regression and its application.* Doctoral Dissertation of China Agricultural University. 2005.
[13] Suykens JAK, Gestel TV, Brabanter JD, Vandewalle J. Least Square Support Vector Machines. *World Scientific.* Singapore. 2002.
[14] Han, Fengqing, Li, Hongmei, etc. A new incremental support vector machine algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(6):1171-1178.
[15] Yu, Yang, Zhou, Liang. Acoustic emission signal classification based on support vector machine. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 1027-1032.