

# Short-term load forecasting based on correlation coefficient and weighted support vector regression machine

Limei LIU<sup>1, a \*</sup>

<sup>1</sup> Department of basic education, Shenyang Institute of Engineering, Liaoning, Shenyang 110136, P.R. China

<sup>a</sup> liulimei\_2006@163.com

\* corresponding author

**Keywords:** power system, support vector regression machine, short-term load forecasting, feature selection, correlation coefficient

**Abstract.** Accurate load forecasting can keep the safety of power grid operation stability and guarantee the normal production and life of society. Support vector regression machine is suitable to load forecasting. Its nature of learning method is to give a good forecasting ability according to a small amount of data information. An algorithm is combining weighted support vector regression machine with feature selection to predict electricity load. It is good significance to forecasting short-term load. The algorithm is used to have feature selection by correlation coefficient and forecast short-term load by weighted support vector regression machine on optimization of data sets. The simulation experimental results indicate that the algorithm made good predicting results and valuable attempts.

## Introduction

The current distribution system scale expands unceasingly and the operation mode is becoming more and more complex along with society's progress and the rapid development of electronic communications technology. Power load forecasting is one of the important works of the power sector. Accurate load forecasting can guarantee the normal production and effectively reduce the cost, improve the economic benefit and social benefit. The theory of load forecasting technology got great development with the advance of modern science and technology. Load forecasting results can determine the grid expansion and renovation and may decide to power grid construction and development. As a result, the level of power load forecasting work has become a measure of whether an electric power enterprise management modernization of one of the significant signs. Power load forecasting problem has become we are faced with the important and difficult task. High precision of short-term load forecasting is the basis and guarantee of the right decisions, power sector at the same time it is also important basis of power generation planning and scheduling, etc.

Prediction methods of technology are the core issue of load forecasting. Traditional load forecasting method mainly includes the time series and regression analysis. They are based on mathematical statistics. Modern forecasting methods mainly including expert system, wavelet analysis, artificial neural network and support vector machine. They are all artificial intelligence technology. Support vector machine will learn question by solving quadratic convex programming problem. SVM is on the basis of statistical learning theory a new kind of machine learning algorithm. It will obtain the global optimal solution, while neural network method can't avoid the local optimum problem [1-3]. At the same time, related to the standard SVM algorithm, such as least squares support vector machines and weighted support vector machine, etc. Support vector machine was originally used to solve the classification problem in the field of pattern recognition. After introducing the insensitive loss function, the SVM can solve the problem of linear and nonlinear regression. Therefore, it was named support vector regression machine (SVRM) when used to solve the regression problem. The SVRM algorithm effectively has become hot research topic in the field of international machine learning and domestic and lots of implementation algorithm appeared international academic circles in recent years. In electric power system, method that a large number of samples are needed cannot be used because the electricity

load samples are difficult to obtain. In comparison SVRM may tackle small sample pattern recognition problem very well. So also has been widely used in electricity load [4-6].

### Support vector regression machine

Support vector regression machine is often used for function approximation, curvilinear regression, forecasting data etc. It can take advantage of the known data on the prediction of the situation in the future and it has very high practical application value. Given a training set  $D = \{(x_i, y_i), i = 1, \mathbf{L}, l\}$ , it has data points  $\{(x_1, y_1), \mathbf{L}, (x_l, y_l)\}$  which is the input data  $x_i \in R^d$  and  $y_i$  is the expected output corresponding to them. It  $l (l \in N)$  is the number of samples and the input dimension  $n$ . When the observed value and the error between the predicted values are less than the given beforehand small positive number, the function of the sample points fitting is error-free. When the sample points between two line tapes, generally think the belt is reasonable. Regression estimate in the objective function is not supported by the area of any loss, and that is belt. Introduction of the new slack variables  $x_i, x_i^*$  was order to data processing what cannot be estimated under the accuracy  $e$  of function  $f$ . Therefore, the objective function can be the following optimization problem.

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + C \frac{1}{l} \sum_{i=1}^l (x_i + x_i^*), \\ \text{s.t. } & y_i - w \cdot f(x_i) - b \leq e + x_i, i = 1, 2, \mathbf{L}, l, \\ & -y_i + w \cdot f(x_i) + b \leq e + x_i^*, i = 1, 2, \mathbf{L}, l. \\ & x_i, x_i^* \geq 0 \end{aligned} \quad (1)$$

Among them,  $\|\cdot\|$  is the Euclidean norm. Symbol  $(\cdot)$  shows vector inner product. Which  $w$  is a  $n$  vector and safeguarding rights and  $b$  is a scalar item Numbers or skewed. And it satisfies the equation  $\|w\|^2 = w \cdot w'$  and variable  $w'$  is  $w$  transposed. While  $C > 0$  is a certainty of constant, it is a balance coefficient and the trade-off between the two. In constraints (1) the first formula says the real value is larger than predicted, but the second formula said real value smaller than predicted. The objective function of support vector machines can be expressed as a quadratic programming problem. Above optimization problem can be replaced with dual form (2).

$$\begin{aligned} \max_{a, a^*} & L(w, b, x, x^*) = -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (a_i - a_i^*)(a_j - a_j^*) f(x_i) \cdot f(x_j) + \sum_{i=1}^l (a_i - a_i^*) y_i - e \sum_{i=1}^l (a_i + a_i^*), \\ \text{s.t. } & \sum_{i=1}^l (a_i - a_i^*) = 0, \\ & 0 \leq a_i, a_i^* \leq \frac{C}{l}, i = 1, 2, \mathbf{L}, l. \end{aligned} \quad (2)$$

By solving the optimization problem (2) and can get the optimal solution for  $a^* = (a_1^*, \mathbf{L}, a_l^*)^T$ . In this way, there is regression function  $f(x) = \sum_{i=1}^l (a_i - a_i^*)(x_i \cdot x) + b$ .

### Short-term load forecasting algorithms based on a new weighted support vector regression machine

#### A new weighted support vector regression machine Algorithms

Plus or minus of slack variables should be considered in standard support vector regression machine. In order to reduce the constraint conditions and optimize the performance of support vector regression machine, it is improved and two square relaxation terms were put forward in the objective function. They are  $x_i^2, (x_i^*)^2, i = 1, 2, \mathbf{L}, l$  respectively. To give different weights  $r_1 > 0, r_2 > 0$  to slack variable is so order to meet the needs of actual. Therefore, optimization problem is following one.

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (r_1 x_i^2 + r_2 (x_i^*)^2), \\ & \text{s.t. } y_i - w \cdot f(x_i) - b \leq e + x_i^2, \quad i=1, 2, \dots, l, \\ & \quad -y_i + w \cdot f(x_i) + b \leq e + (x_i^*)^2, \quad i=1, 2, \dots, l. \end{aligned} \quad (3)$$

When it satisfies to  $r_1 = r_2 = 1$ , both weights are the same. Zero  $e$  is controlled by the parameter with the corresponding situation when the weight of the standards of the attenuation factor is least squares linear regression, also known as ridge regression. Non-zero  $e$  impact is the introduction of the extra weight of dual parameters attenuation factor. New weighted support vector regression machine is than less often than a pair of constraint conditions and it concludes ridge regression algorithm. Using the Lagrange multiplier and the dual optimization method can get the following equation.

$$L(w, b, \mathbf{x}, \mathbf{x}^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (r_1 x_i^2 + r_2 (x_i^*)^2) - a_i (y_i - (w f(x_i)) + b) - e - x_i^2 - a_i^* ((w f(x_i)) + b) - y_i - e - (x_i^*)^2, \quad (4)$$

Among them, the coefficient  $a_i$  and  $a_i^*$  are all the Lagrange multiplier. According to the KKT (Karush - Kuhn - Tucker) conditions, first solve their partial derivatives of variables to solve the problem. In the case of linear inseparable, nonlinear function is the input samples mapped to high-dimensional space could be infinite dimensional feature space  $f: R^n \rightarrow H$ . In the feature space  $H$ , the inner product condition is  $k(x_i, x_j) = f(x_i) \cdot f(x_j)$  and a lot of operation will be in the original input space rather than in the high dimensional feature space. Radial basis function (RBF) is usually used to solve the problem. Thus the nonlinear problem can be turned into a high dimensional linear problem.

$$\begin{aligned} \min_{a, a^*} L(w, b, \mathbf{x}, \mathbf{x}^*) &= \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (a_i - a_i^*) (a_j - a_j^*) k(x_i, x_j) - \sum_{i=1}^l (a_i - a_i^*) y_i + e \sum_{i=1}^l (a_i + a_i^*), \\ \text{s.t. } \sum_{i=1}^l (a_i - a_i^*) &= 0, \\ 0 \leq a_i, a_i^* &\leq \min\{r_1 C, r_2 C\}, \quad i=1, 2, \dots, l, \end{aligned} \quad (5)$$

### Feature selection

Parameter selection is important to short-term load forecasting. Short-term load forecasting is based on historical data of power load, temperature and day type as the basis. There are good methods of an effective prediction for forecasting load that can seek the historical data of the impact of load in the future and explore the short-term load and the related factors between inner relations. In the practical application of machine learning, the number of feature is often more. Among them there may be irrelevant features and there are likely to be dependent on each other between characteristics. The more characteristics Number, the longer the time it takes to training model. This easily causes the "dimension disaster" and the model will be more complex and its promotion ability will decline. Feature selection can eliminate irrelevant or more than redundant characteristics, so as to decrease the number of features to improve model accuracy and reduce the running time. On the other hand, feature may select the really relevant characteristics and simply model.

Filter measure its quality by analyzing the characteristics of internal feature subset. Use distance measure for feature selection is based on the assumption that good feature subset should make samples belonging to the same class as small as possible, belongs to the distance between the different kinds of samples as far as possible. The commonly used distance measure included to Euclidean distance, standardized Euclidean distance, markov distance, etc. Famous statistician Karl Pearson design the statistical index, the Correlation coefficient. Correlation coefficient is used to reflect the correlation between variables close degree of statistical indicators. Correlation coefficient is calculated by covariance method, as well as two variables and their average deviation as the foundation, through the two deviation multiplied to reflect the correlation degree between two variables. This paper studies linear on the number of single phase relationship. Correlation coefficient will be used to feature selection so that it gived an optimization of data sets.

### Short-term load forecasting algorithms steps

By solving the problem (5), regression function can be obtained. Here are the improved particle swarm optimization and support vector regression machine algorithm steps as follows:

Step 1 Choose a known training sets  $D = \{(x_i, y_i), i = 1, \mathbf{L}, l\}$ ,  $x_i \in R^n$ ,  $y_i \in R$  and make data normalization process.

Step 2 Feature selection depends on correlation coefficient  $r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$ . There will

be obtained a new optimization of data sets.

Step 3 Select the appropriate parameters  $e > 0, C > 0, r_1 > 0, r_2 > 0$  and the proper kernel function.

Step 4 Cross validation to obtain some parameters. If it satisfied the precision evaluation or iterations numbers, the optimal solution was output; Otherwise go to step 3.

Step 5 Structure and to solve the optimization problem (5), get the optimal solution for  $\bar{a} = (\bar{a}_1, \bar{a}_1^*, \mathbf{L}, \bar{a}_l, \bar{a}_l^*)^T$ .

Step 6 To calculate  $\bar{w}$  by following equation  $\bar{w} = \sum_{i=1}^l (\bar{a}_i^* - \bar{a}_i) f(x_i)$ , and then construct the decision function  $f(x) = \sum_{i=1}^l (\bar{a}_i^* - \bar{a}_i) k(x_i, x) + \bar{b}$ . Among them  $\bar{b}$  in the following way:

When  $\bar{a}_j^* \in (0, \min\{r_1 C, r_2 C\})$ ,  $j = 1, 2, \mathbf{L}, l$ ,  $\bar{b} = y_i - \sum_{j=1}^l (\bar{a}_j^* - \bar{a}_j) \cdot k(x_i, x_j) - e - x_i^2$ ,  $i = 1, 2, \mathbf{L}, l$ ,

And  $\bar{a}_k^* \in (0, \min\{r_1 C, r_2 C\})$ ,  $k = 1, 2, \mathbf{L}, l$ ,  $\bar{b} = y_k - \sum_{i=1}^l (\bar{a}_i^* - \bar{a}_i) k(x_i, x_k) + e + (x_i^*)^2$ ,  $i = 1, 2, \mathbf{L}, l$ .

## Simulation Experiments

### Forecasting model of short-term load

Usually the influencing one factor of short-term load forecasting is mainly meteorological conditions. It includes many aspects, such as season, wind, pressure, temperature, humidity, and clear, and many other factors. As people living standard gradually increase and social development, refrigerators and air conditioning is widely used. Those household electrical appliances make the total power load increasing. Thus the change of the meteorological factors increases the influence of load change. The characteristics of the input data include the highest temperature and minimum temperature, wind, humidity and the output data were power consumption for IPSOSVRM experiment samples. In order to eliminate the influence of the dimension, respectively the feature data and the power consumption data normalization, build the training sample, and thus established PSVM electricity consumption forecasting model.

Short-term load mainly associated with the highest temperature, lowest temperature, average temperature, wind, humidity, etc. Training samples attributes in the input data express the maximum temperature and minimum temperature, wind, humidity.  $x_i = \{(x_{i1}, x_{i2}, \mathbf{L}, x_{in})\}$ ,  $i = 1, \mathbf{L}, l$ ,  $x_i \in R^n$ . Here it equals to  $n = 5$  and that means the number of training samples attributes is five,  $x_i = \{(x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5})\}$ ,  $i = 1, \mathbf{L}, l$ ,  $x_i \in R^5$ . It  $x_{i1}$  denotes maximum temperature,  $x_{i2}$  denotes minimum temperature,  $x_{i3}$  denotes average temperature,  $x_{i4}$  denotes wind and  $x_{i5}$  denotes humidity. The output data  $y_i \in R$ ,  $i = 1, \mathbf{L}, l$  shows daily electric energy.

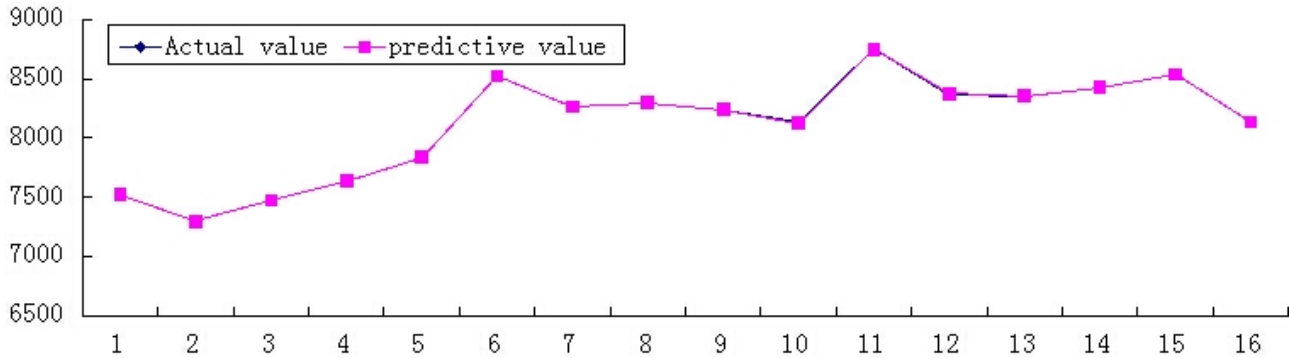
### Experiment and analysis

The experimental data were electricity data of a city in June 2014 and relevant characteristics of these data. Can be seen from the following chart, the fitting is very good between actual value and predictive value. The experiments are 65 groups of representative data what have 5 parameters. Selecting 50 samples are using to train and the others are to be used to test.

Can be seen from the following table 1, attributes of data could simplify further and there is to keep three characteristics and remove the two correlation characteristic of less. Thus a new optimization of data sets appeared and it will reduce the running time.

**Table 1: The correlation order according to correlation coefficient**

Attributes	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$	$x_{i5}$
Correlation coefficient	0.98	0.82	0.91	0.45	0.57
Correlation order	1	3	2	5	4



**Fig. 1: The fitting curve based on correlation coefficient and WSVRM algorithm**

The results are shown in Figure 1 and the fitting is very good between actual value and predictive value. The two curves almost are inseparable. Mean absolute percentage error (MAPE) is 0.006593. It reduced by about 13% of the running time. The experimental results show that based on the new algorithm is suitable for short-term load forecasting.

## Conclusions

The proposed based on correlation coefficient and weighted support vector regression machine (WSVRM) is a new method for prediction of short-term load. Among them, using correlation coefficient to select optimization of data sets combination weighted support vector regression machine training, get better power consumption performance prediction model. The results show that it has better generalization ability and application significance. Study in the future will focus on short-term load prediction with micro power grid.

## Acknowledgements

This work was financially supported by Liaoning educational scientific research general projects (L2014520) and the youth fund of Shenyang Institute of Engineering (LGQN-1401).

## References

- [1] N.Y Deng, Y.J.Tian. A New Method of Data Minging Support Vector Machines Beijing 2004.
- [2] H.B. Cheng, P.N Tan, R. Jin. IEEE Transactions on Knowledge and Data Engineering, Vol. 22(2010): 537-549.
- [3] Hong W C. Applied Mathematical Modelling, 2009, 33(5):2444-2454.
- [4] XIAOJIA WANG. System Engineering, 2012 (5): 131-136
- [5] CHEN B J, CHANG M W, LIN C J. IEEE Trans on Power Systems, 2004, 19 (4):1821-1830.
- [6] KADIR KAVAKLIOGLU. Applied Energy, 2011 (88): 368 -375.