Forecasting of Electricity Load Based on Improved Particle Swarm Optimization and Support Vector Regression Machine

Limei Liu[*]

Department of basic education, Shenyang Institute of Engineering, Liaoning, Shenyang 110136, P.R. China
[*]Corresponding author

Abstract—**Support vector regression machine is suitable for small sample decision and it is good to data forecasting capabilities. Its nature of learning method is under the condition of limited information to obtain a good ability in data mining. Accurate electricity load forecasting is an important practical value to our lives. This paper presents a new algorithm that is an improved particle swarm optimization algorithm and support vector regression machine that is proposed to predict electricity load. It is of great significance to forecasting electricity load. The algorithm can optimize training parameters of support vector regression machine by improved particle swarm optimization algorithm. The simulation experimental results indicate that the new algorithm made a meaningful exploration on forecasting electricity load.**

Keywords- power system; support vector regression machine; forecasting of electricity load; improved particle swarm optimization

I. INTRODUCTION

The electricity load forecasting is a daily routine in the operations of electric power system. Accurate electricity load forecasting plays an important rate both production planning, unit maintenance scheduling, economical and safe running, which directly influences the profit of the electric utility enterprises, and in the construction of electric power market, which need competition and economic ensure between the electric power corporations. The core technology of the forecasting of electricity load is what needs construct the mathematic model of the real load at the basis of the analysis of its specialty and design the effective algorithm to obtain accurate results. Support vector regression machine (SVRM) is integrated with the statistics learning theory. Firstly, training samples may be mapped into high dimension space. Then constructing training models are corresponded with different kernel functions. But sometimes operation speed and the accuracy of the standard support vector regression machine is not ideal. Support vector regression machine can take advantage of the known data on the prediction of the situation in the future and it has very high practical application value. For example, the process industry with large scale and complicated structure, and the characteristics of the site conditions, to predict parameters of process industry and process monitoring technology research in the field of process

control is the most challenging one of the hot topics.

The SVRM algorithm effectively has a sufficient theoretical basis and it has become hot research topic in the field of international machine learning. Support vector machine will learn question boils down to a convex quadratic programming problem. It will get the global optimal solution, but neural network method can't avoid the local optimum problem. In recent years, algorithm implementation has made breakthrough progress by domestic and international academic circles. [1-3] Expanding the scope of its application and results, practical application has a bright prospect in the field of theoretical research. SVRM is essentially a kind of nonlinear data processing tool, suitable for solution for the nonlinear forecasting problem and it is essentially a highly theoretical and practical value of one of the pattern recognition method.

In electric power system, some of the electricity load samples are difficult to obtain and the SVM can tackle small sample pattern recognition problem very well. So also has been widely used in electricity load [4-6]. At present, grid full risk is very big because the power grid construction speed is far below the power load demand growth and there is not safety. As industrial production safety problem is becoming more and more attention and more strict to the requirement of electricity load and early warning system, such as improving sensitivity, improving the USES of forecasting electricity load system, etc. Compared with the artificial neural network, SVRM is not only simple in structure and good generalization ability obviously.

II. SUPPORT VECTOR REGRESSION MACHINE

Support vector machine is suitable for processing two types of problems. Using the method of support vector machine regression problem is called support vector regression machine and it is often used for function approximation, curvilinear regression, etc. Support vector regression machine 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, it has data points $\{(x_1, y_1), \cdots, (x_l, y_l)\}$ which is the input data $x_i \in R^d$ and y_i is corresponding to the input vector x_i of the expected output. For the training set $D = \{(x_i, y_i), i = 1, \cdots, l\}$, which l is

the number of samples, as the input dimension n. The purpose is to find hyper-plane $\langle w, b \rangle$ and test data $y_1^*, \cdots, y_n^*$. Support vector regression problem is to find the appropriate real function, $f(x) = w \cdot x + b$.

In the case of linear inseparable, nonlinear function is the input samples mapped to high-dimensional space could be infinite dimensional feature space $\phi : R^n \rightarrow H$. In the feature space H, a function k can satisfy the inner product conditions $k(x_i, x_j) = \phi(x_i)\phi(x_j)$ and it is inner product of high dimensional space. In this way, 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.

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \tag{1}$$

Function based on support vector regression machine is satisfied the following structural risk minimization principle:

$$J = \frac{1}{2}\|w\|^2 + C \cdot \frac{1}{l} R_{emp}[f] \tag{2}$$

The first part of (2) shows the complexity of $f(x)$ and the second part is the average loss on the training set. While $C > 0$ is a certainty of constant, it is a balance coefficient and the trade-off between two parts. 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. Selects the insensitive loss function [8]:

$$\begin{aligned} &\min \frac{1}{2}\|w\|^2 \\ &s.t. \ y_i - w \cdot x_i - b \leq \varepsilon \\ &\qquad w \cdot x_i + b - y_i \leq \varepsilon \end{aligned} \tag{3}$$

The introduction of the new slack variables ξ_i, ξ_i^* was order to data processing what cannot be estimated under the accuracy ε of function $f(x)$. Therefore, (2) the risk function minimization can be converted into the following optimization problem.

$$\begin{aligned} &\min \frac{1}{2}\|w\|^2 + C\frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ &s.t. \ y_i - w \cdot \phi(x_i) - b \leq \varepsilon + \xi_i, i = 1, 2, \cdots, l \\ &\qquad -y_i + w \cdot \phi(x_i) + b \leq \varepsilon + \xi_i^*, i = 1, 2, \cdots, l \\ &\qquad \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{4}$$

Among them, sign $\|\cdot\|$ is the Euclidean norm. Which w is a n vector and safeguarding rights and b is a scalar item Numbers or skewed. Symbol $(\cdot)$ shows vector inner product. In constraints (4) 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. And $\|w\|^2 = w \cdot w'$ variables w' is w transposed, and $C > 0$ used to weigh the maximum boundary. About (4) optimization problem can be replaced with dual form (5).

$$\begin{aligned} &\max_{\alpha, \alpha^*} L(w, b, \vec{\xi}, \vec{\xi}^*) \\ &= -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi(x_i) \cdot \phi(x_j) \\ &\quad + \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*), \\ &s.t. \ \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, \\ &\qquad 0 \leq \alpha_i, \alpha_i^* \leq \frac{C}{l}, i = 1, 2, \cdots, l. \end{aligned} \tag{5}$$

By building the optimization problem (5) and solving, can get the optimal solution for $\alpha^* = (\alpha_1^*, \cdots, \alpha_l^*)^T$. There will also be able to get regression function $f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)(x_i \cdot x) + b$.

III. Improved Particle Swarm Optimization Algorithms Based on Support Vector Regression Machine

A. Improved Particle Swarm Optimization Algorithms

The prediction accuracy of support vector regression machines and penalty parameters, kernel function and insensitive loss of value has great relationship. As a result, appropriate optimization algorithm may optimize these parameters in order to obtain the optimal combination of these parameters. PSO is initialized with a group of random particles. Each particle flies in a d-dimensional space. The location of the i-th particle is expressed as $x_i = \{x_{i1}, x_{i2}, \cdots, x_{id}\}$. For a discrete problem expressed in a binary notation, a particle moves in a search space restricted to 0 or 1 on each dimension. In binary problem, updating a particle represents changes of a bit which should be in either state 1 or 0 and the velocity represents the probability of bit x_{id} taking the value 1 or 0. Change the location of particle velocity is shown to $v_i = \{v_{i1}, v_{i2}, \cdots, v_{id}\}$. The location of the i-th particle denotes $p_i = \{p_{i1}, p_{i2}, \cdots, p_{id}\}$ to meet the best fitness value. The best position of all particles in groups what can be

shown $p_g = \{p_{g1}, p_{g2}, \cdots, p_{gd}\}$. Each particle with the two optimal values is constantly updated in each step. In both the optimal values found, the particles can be according to following formula (6) to adjust their speed and location [8-10]. In discrete particle swarm optimization, the equations of the position and the velocity of the particles are established as:

$$\begin{cases} v_{id}(t+1) = wv_{id}(t) + c_1 rand()(p_{id} - x_{id}(t)) \\ \quad\quad + c_2 rand()(p_{gd} - x_{id}(t)), \\ S_{id}(v_{id}(t+1)) = \tanh((k+1)v_{id}(t+1)), \quad (6) \\ x_{id}(t+1) = 1, \quad if \ rand() < S(v_{id}(t+1)), \\ x_{id}(t+1) = 0, \quad else, \end{cases}$$

where $rand()$ is a random number in interval $(0,1)$. They c_1 and c_2 are two positive acceleration constants. They can be called the learning factor. The velocity v_{id} is generally conditioned by $v_{id} \in [-v_{max}, v_{max}]$ to prevent the particles from flying out of the solution area. In general, it $v_{max} = 4$ will be able to meet the condition. In order to simplify the problem, this paper proposes an improved PSO algorithm and the random probability in [7] and [8] are replaced by the trisection point in the interval. It can make computing speed faster. The improved algorithm can be stated as following.

$$\begin{cases} x_{id}(new) = x_{id}(old), \quad if \ 0 < rand() < 0.33 \\ x_{id}(new) = p_{id}, \quad if \ 0.33 \le rand() < 0.66 \quad (7) \\ x_{id}(new) = p_{gd}, \quad if \ 0.66 \le rand() < 1 \end{cases}$$

where p_{id} is the optimal value of the particle itself, p_{gd} is the global optimal value, and v_{id} is the velocity which is the random number evenly distributed interval in $[0,1]$. There should be a balance between the local and global search ability.

B. Improved Particle Swarm Optimization Algorithms Based on Support Vvector Regression Machine

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 following.

Step 1 Given a known training sets ,
$D = \{(x_i, y_i), i = 1, \cdots, l\}, x_i \in R^n, y_i \in R$.

Step 2 Select the appropriate parameters $\varepsilon > 0, C > 0$ and the proper kernel function $k(x, x')$.

Step 3 Set number of iterations, the particle number, and random initial position and velocity of a set of particles.

Step 4 Fitness function is adopted to improve the fitness evaluation. The fitness evaluation is mean absolute percentage error $F = \frac{1}{s}\sum_{i=1}^{s} \left| \frac{y_i - \overline{y}_i}{y_i} \right|$.

Step 5 If the current fitness value of particles is better than the previous best position p_{id} , replace it with the current best position p_{id} . If the current fitness value of particles is better than that of the global best position p_{gd} experienced by, will p_{gd} replaced with best the current position of the particles.

Step 6 According to the type of update the particle's speed and position.

Step 7 If it satisfied the fitness evaluation or iterations numbers, the output optimal solution; Otherwise go to step 2.

Step 8 Structure and solve the optimization problem (5) and get the optimal solution for $\overline{\alpha} = (\overline{\alpha_1}, \overline{\alpha_1^*}, \cdots, \overline{\alpha_l}, \overline{\alpha_l^*})^T$;

Step 9 To calculate $\overline{w}$ by following equation $\overline{w} = \sum_{i=1}^{l} (\overline{\alpha_i^*} - \overline{\alpha_i})\phi(x_i)$, and then construct the regression function $f(x) = \sum_{i=1}^{l} (\overline{\alpha_i^*} - \overline{\alpha_i})k(x_i, x) + \overline{b}$.

Among them $\overline{b}$ in the following way:

When $\alpha_j^* \in (0, C), j = 1, 2, \cdots, l$,

$\overline{b} = y_i - \sum_{j=1}^{l} (\overline{\alpha_j^*} - \overline{\alpha_j}) \cdot k(x_i, x_j) - \varepsilon - \xi_i^2, i = 1, 2, \cdots, l,$

And $\alpha_k^* \in (0, C), k = 1, 2, \cdots, l$,

$\overline{b} = y_k - \sum_{i=1}^{l} (\overline{\alpha_i^*} - \overline{\alpha_i})k(x_i, x_k) + \varepsilon + (\xi_i^*)^2, i = 1, 2, \cdots, l.$

IV. SIMULATION EXPERIMENTS

A. Forecasting Model of Electricity Load

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 established PSVM electricity consumption forecasting model.

Power consumption mainly associated with the highest temperature, lowest temperature, wind, humidity, etc. Training samples in the input data express the maximum temperature and minimum temperature, wind, humidity. $x_i = \{(x_{i1}, x_{i2}, \cdots, x_{in}), i = 1, \cdots, l\}, x_i \in R^n$. Here it equals to $n = 4$ and there are four attributes of each sample,

$x_i = \{(x_{i1}, x_{i2}, x_{i3}, x_{i4}), i = 1, \cdots, l\}$, $x_i \in R^4$. It x_{i1} denotes maximum temperature, x_{i2} denotes minimum temperature, x_{i3} denotes wind and x_{i4} denotes humidity. The output data shows daily electric energy. $y_i \in R$ $i = 1, \cdots, l$ .Using the following fitness function, it finished fitness evaluation $F \leq 0.00512$.

B. Experiment and Analysis

The experimental data were electricity data of a city in July 2015 and relevant characteristics of the data. Can be seen from the following chart, the fitting is very good between actual value and predictive value. The experiments are 100 groups of representative data that have 4 parameters. Selecting 85 samples are using to train and the others are to be used to test.
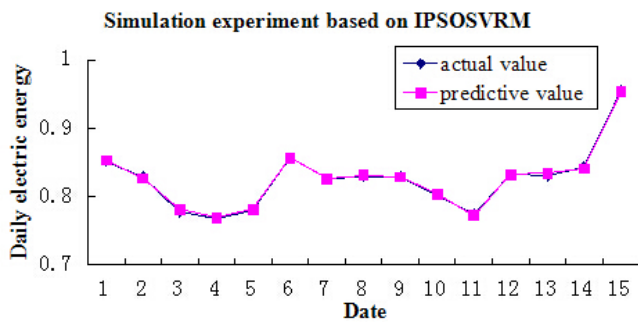

FIGURE I. THE FITTING CURVY BASED ON THE IPSOSVRM ALGORITHM.

Seen from the above chart figure 1, the fitting is very good between actual value and predictive value. The two curves almost are inseparable.

TABLE I. THE PREDICTION ERROR AND RELATIVE ERROR BASED ON THE IPSOSVRM ALGORITHM.

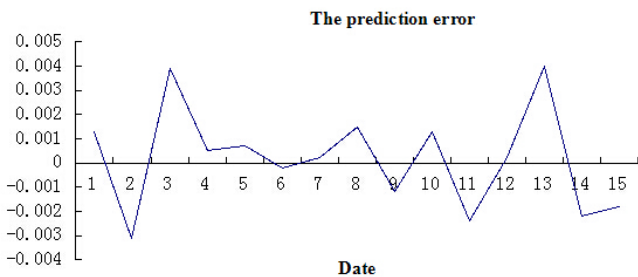| Date	1	2	3	4	5
Absolute error	0.0013	-0.0031	0.0039	0.0005	0.0007
Date	6	7	8	9	10
Absolute error	-0.0002	0.0002	0.0015	-0.0012	0.0013
Date	11	12	13	14	15
Absolute error	-0.0024	0.0002	0.004	-0.0022	-0.0018

FIGURE II. THE PREDICTION ABSOLUTE ERROR BASED ON THE IPSOSVRM ALGORITHM.

From above table what are shown, the maximum prediction absolute error is lower than 0.004 and the minimum prediction absolute error is only -0.0031. Figure 2 is shown to the changed prediction absolute error. The experimental results show that based on improved particle swarm optimization algorithm and support vector machine has the very high prediction accuracy of forecasting the electricity requirements and it is very suitable for power consumption forecast.

V. CONCLUSIONS

The proposed algorithm based on improved particle swarm optimization and support vector regression machine (IPSOSVRM) is a new method for prediction of power consumption. Among them, using the particle swarm optimization algorithm to select the optimal parameter combination support vector machine training, it gets better power consumption performance prediction model. The results show that it has better generalization ability and application significance. The next study will focus on load prediction with micro power grid.

REFERENCES

[1] N.Y Deng, Y.J.Tian, "A New Method of Data Minging-Support Vector Machines" Beijing 2004, p p.174-175.

[2] M. Varewyck, J.-P.Martens, "A Practical Approach to Model Selection for Support Vector Machines With a Gaussian Kernel", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,Vol. 41 (2011), pp. 330 – 340.

[3] Taylor J W, Buizza R., "Neural network load forecasting with weather ensemble predictions",.IEEE Trans.Power Syst.2002,17 (3) , pp.626-632.

[4] H.B. Cheng, P.N Tan, R. Jin, "Efficient Algorithm for Localized Support Vector Machine", IEEE Transactions on Knowledge and Data Engineering, Vol. 22(2010), pp. 537-549.

[5] DONGXIAO NIU, YONGLI WANG, DESHENG DASH WU,"Power load forecasting using support vector machine and ant colony optimization", Expert Systems with Applications, 2010 (37), pp.2531- 2539

[6] KADIR KAVAKLIOGLU, "Modeling and prediction of Turkey's electricity consumption using support vector regression", Applied Energy, 2011 (88) , pp.368 - 375

[7] Q. Shen, J. Jianhui., "Modified Particle Swarm Optimisation Algorithm for Variable Selection in MLR and PLS Modeling: QSAR Studies of Antagonism of Antagonism of Angiotensin II Antagonists",European Journal of Pharmaceutical Science, 2004, pp.145–1524.

[8] J. Kennedy, R. Eberhart., "A Discrete Binary Version of the Particle Swarm Optimization Algorithm",Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia, 1997, pp.4104–4108.