Hyper-sphere support vector machine based on collaborative knowledge mining and SMO optimization in power load forecasting

Resheng PAN¹, Hui LI¹, Zhidong WANG¹, Dong PENG¹, Lang ZHAO¹, Yongli WANG², Fang LI²
(1. State Grid Beijing Economic Research Institute, Beijing 102209, China; 2. School of Economics and Management North China Electric Power University, Changping District, Beijing 102206 China)

Abstract

Background: Due to the influence of power market reform policies, the conversion of power loads has become more and more complicated. The current load forecasting methods have long calculation times and inaccurate volatility load forecasting. The difficulty of forecasting is becoming greater and more accurate. It becomes very important to predict the electrical load. Under this background, this paper proposes the application methods of collaborative knowledge mining and SMO in solving prediction models based on hyperball support vector machine (CKM / SMO-SVM).

Methods: This study first analyzes the impact of historical data on samples and different parameters. The prediction of power load, sample data and various parameters have a significant impact on the prediction results. Secondly, applying weak entropy theory for collaborative knowledge mining, preprocessing sample data and historical information. Third, a short-term power load forecasting system based on the hypersphere support vector machine model is established and the problem is solved by SMO. Finally, the SVM model and BP model are selected for prediction to verify the new model.

Results: Our research proves that the rms relative error of the CKM / SMO-SVM model is only 2.32%, which is 0.67% and 1.56% lower than the SVM and BP models, respectively, and the optimization speed is faster.

Conclusions: The model proposed in this paper utilizes Hyper-sphere SVM which is suitable for Gaussian kernel function to achieve faster and more accurate load forecasting, which can provide more accurate services for energy spot transactions and energy scheduling plans.

Keywords: SMO; collaborative knowledge mining; sequence optimization; hyper-sphere SVM; power load forecasting

1. INTRODUCTION

Short-term load forecasting of power system plays an important role to reliable and economic operation, especially with the development of power market, short-term load forecasting obtains more and more attention [1]. SMO (sequential minimal optimization) put forward by Platt and applied to solve support vector machine (SVM) [2]. Zhou considered the shortcomings of conventional algorithms, simplified the SMO algorithm of solving the non-positive definite kernel, and only need to consider two Lagrange multiplier [3]. Zhao, Jiang, Jin and Liu et al. (2008) improved the disadvantages of slow training speed and unstable training results [4]. Aiming at this problem, Men and Wang introduced a parameter selection and optimization method of support vector machine (SVM) [5]. Wang, Liu and Wang combined the SMO algorithm with the differential evolution algorithm which is applied to the soft-measurement in the flue gas oxygen content, and exhibited excellent performance in multimodal function optimization problems, this method has achieved good results [6]. In [7-8] discusses how SMO can effectively reduce the sampling range.

Compared with the algorithms mentioned above, this paper combines the collaborative mining association technology with weaken entropy theory, clusters early samples and historical information and then deals with correlation pretreatment between subsets. Super sphere support vector machine (SVM) theory solved with the SMO algorithm is used finally to forecast the power load, a good prediction effect was finally achieved.

2. The collaborative knowledge mining

2.1 Data clustering mining

At present, weather is the most important influence factors for short-term daily load curve, in order to significantly improve daily load forecasting accuracy, the influence of meteorological factors must be considered. In addition, the time factors (season, whether holidays and weekend or not) also exert a tremendous influence on the power load. Among these weather factors, daily highest temperature, daily minimum temperature, daily average temperature and rainfall, humidity, cloud cover, wind speed have a great influence on the load and can be given by weather forecast. These factors were summarized as ten levels, respectively are the daily highest temperature, daily minimum temperature, daily average temperature and rainfall, humidity, cloud cover, wind speed, season, whether holidays or weekends, designated symbolically as [Z1, Z2, Z3, Z4, Z5, Z6, Z7, Z8, Z9, Z10]. Z1 to Z3, Z5 to Z7 was divided into three comparative value--[low, medium, high] and respectively assigned as [1, 2, 3]; Z4 was divided into four comparative value based on fuzzy classification as [no rain, light rain, moderate rain, heavy rain], respectively assigned as [0, 1, 2, 3]; Z8 classified as spring, summer, autumn and winter, respectively assigned as [1, 2, 3, 4]; Z9, Z10 classified as yes or no, yes was evaluated as 0, no was evaluated as 1 [Niu, Wang, Wu, 2010]. The historical daily load can be classified as follows:

$$Z = \begin{cases} Z_1 = 1, 2, 3 \\ Z_2 = 1, 2, 3 \\ Z_3 = 1, 2, 3 \\ Z_4 = 0, 1, 2, 3 \\ Z_5 = 1, 2, 3 \\ Z_6 = 1, 2, 3 \\ Z_7 = 1, 2, 3, 4 \\ Z_8 = 0, 1 \\ Z_9 = 0, 1 \end{cases}$$
The low, medium and high standards of meteorological factors can be determined by the actual situation in the region or fuzzy clustering method. Input history daily load curve into database every day, and input the influence factors of daily fuzzy classification tags at the same time. Each day have a fuzzy classification and can form the corresponding database in this way.

If we know the category of predict day is \((3, 2, 1, 1, 2, 3, 3, 1, 1, 1)\) according to the date and the weather forecast, we then extract all historical days that have category characteristics of \((3, 2, 1, 1, 2, 3, 3, 1, 1, 1)\) reversely from historical load database to form a fuzzy classification database, the library of all the days of this database have the characteristics of \((3, 2, 1, 1, 2, 3, 3, 1, 1, 1)\). The new database which is extracted of the data source that contains mining theme called mining subject database. Mining attribute of the project is no longer contained in each data records.

2.2 data correlation
Calculate the support \((s)\) and the confidence \((c)\) to universal set for each mining subject database. Support is the measure of the importance for association rules. It shows the representative compared with all things of universal set. Obviously, the larger the value of support is, the more representative it indicates. Support values are generally small on account of the abundant data in practice. Confidence is the measure of the accuracy for association rules.

We define correlation for the product of support and confidence: \(r_i = s \times c\) to solve the correlation of the mining subject database and universal set.

2.3 weaken entropy theory
It may appear a few small probability event in the new database, in order to avoid the small probability events, impact on association rule of mining, we import weaken entropy theory. Entropy expresses the measurement of uncertainty, sudden and randomness in data set. We give the probability \(p_1, p_2, \ldots, p_n\), where \(\sum p_i = 1\), the entropy is:

\[
H(p_1, p_2, \ldots, p_n) = \sum p_i \times \lg(1/p_i)
\]  

(1)

This paper uses the technology of weaken entropy, only to operate the probability of the individual elements.

It shows as follows:

\[
H(p_i) = p_i \times \lg(1/p_i)
\]  

(2)

According to the analysis for mathematics characteristics of eq. (2), if \(p_i \to 0, H(p_i) \to +\); else if \(p_i \to 1, H(p_i) \to 0\). This result shows that weaken entropy’s amplitude of variation in large probability events is limited.

2.4 correlation analysis for coordinated mining
According to the above model, the data is analyzed. And then modified association rules can be achieved. The new database marks an essentially change from previous database. We can consider that each attribute (project) of the maximum frequent item sets which form the association rules appears approximate probability in the new database, and its frequency is the highest.

3. Hyper-sphere support vector machine(SVM) regression theory
3.1 Hyper-sphere support vector machine
This paper uses the hyper-sphere OC - SVM algorithm. The idea is to find a hyper-sphere, making its radius as small as possible at the same time, the training sample contains as much as possible.

\[
\min_R R^2 + C \sum_{i=1}^{l} \zeta_i
\]  

(3)

\[
s.t. \Phi(x_i) - a^2 \leq R^2 + \zeta_i
\]  

(4)

\[
\zeta_i \geq 0
\]  

(5)

Where \(R\) is the radius of the sphere, \(a\) is the center and \(\zeta\) is slack variable, \(l\) is the number of training samples. \(C\) is the error penalty factor which controls the degree of punishment to the sub samples, achieves a compromise between the size of the ball and the number of samples contained.

When the sample points of the nonlinear separable, the samples are mapped to high dimensional feature space through nonlinear mapping \(\Phi(x_i)\), where the sample is linearly separable, and we choose kernel function \(K(x, y) = \Phi(x), \Phi(y)\) which satisfies the Mercer condition. This paper makes use of the most popular kernel function, Gauss kernel function. Among them, \(\Phi(x), \Phi(y)\) is the inner product of \(\Phi(x)\) and \(\Phi(y)\).

The Lagrange function is derived from the above formula.

\[
L = \left( R, a, \zeta_i, \alpha_i, \beta_i \right) = R^2 + C \sum_{i=1}^{l} \zeta_i - \sum_{i=1}^{l} \alpha_i \left( R^2 + \zeta_i - \Phi(x_i) - a^2 \right) - \sum_{i=1}^{l} \zeta_i \beta_i
\]  

(6)

Transforming to dual problem:

\[
W(a) = \min_a \sum_i \alpha_i \left( \sum_j \alpha_j k(x_i, x_j) - \sum_i \alpha_i x_i \right)
\]  

(7)

Which is satisfied with:

\[
\sum_{i=1}^{l} \alpha_i = 1
\]  

(8)

Solving the above QP problem, we can get Lagrange multiplier \(\alpha_i\), and determine the center and radius. The point on the sphere plays a key role to determine the sphere, which is called support vector.

Detail research has been carried out in statistical learning theory at abroad as the basis of support vector machine theory, but the support vector machine is used
as a core content of statistical learning theory, its application is far from its expected effect theory. The reason, one is the depth and the computational efficiency of the support vector machine algorithm is not enough; the two is the choice of the kernel function and the parameters of the support vector machine. Therefore, in this paper, the SMO optimization algorithm and support vector machine are combined to find out the relationship between the prediction accuracy and the parameters, which will greatly reduce the complexity of the calculation.

3.2 The influence of penalty parameter C
The function of the penalty parameter C is to adjust the ratio of the confidence range of the learning machine and the experience risk in the determined data subspace, so as to make the learning machine’s generalization ability best. The optimal C in different data subspace is different, in the determination of the data subspace, the small value of C indicates that the punishment of the empirical error is small, and the complexity of the learning machine is small and the value of experience risk is larger, and vice versa. The former is called “under learning”, while the latter is “over learning”. When C exceeds a certain value, the complexity of SVM reaches the maximum value of the data subspace, and then the experience risk and promotion ability almost no longer change. There is at least one suitable C for each data subspace, which makes the best generalization ability, as is shown in Fig. 1. As can be seen from the figure, when the C value is smaller, the error is larger, and decreases with the increase of C; When the value of C in the gradual increase in the process, in a certain stage it will be in a stable, but as the C is too large, the error increases with the increase of C.

3.3 The influence of estimation accuracy δ
The slack variables ζi control the width of the insensitive band ζ and affect the number of support vectors. If the value ζ is too small, the regression estimation accuracy is high, but the number of support vectors is increased, the complexity of SMO algorithm is increased in the paper; if the value is too large, the regression estimation precision is reduced, the number of support vectors is reduced, and the complexity of SMO algorithm is reduced. And the estimation accuracy δ is similar to the effect of relaxation variables ζ on the system. Therefore, in the standard support vector machine, the parameters δ and the C control the complexity of the model in different ways.

3.4 The influence of nuclear parameters σ
Gauss kernel parameter σ reflects the distribution or range of the training sample data, it determines the width of the local neighborhood. Larger σ means lower variance. As shown in Fig. 2. When σ → 0, the error is small and the test error is large, it is a learning phenomenon, and the algorithm is short of generalization ability; When σ → ∞, training and test errors are large, it is a phenomenon that is under study; When the value σ is less than a certain value, the correct rate of classification is decreased, the value σ of the description is not in the mathematical sense to tend to 0. Therefore, it is of great significance to improve the classification accuracy.

4. SMO algorithm
4.1 The principle of SMO algorithm
According to the theory of optimum, it can be regarded as a solution of the original equations when all meet KKT conditions of the objective function, therefore, we provide a method which is to judge whether αi is violated KKT condition or not. For ∀αi:

1) When αi = 0, \(d_i^2 \leq R^2\); When \(0 < α_i < C\), \(d_i^2 = R^2\);
When \(α_i = C\), \(d_i^2 \leq R^2\)

2) \(\sum_{i=1}^{l} α_i = 1\)

\(α_i\) is Lagrange multiplier, \(d_i^2\) is the square of the distance from \(x_i\) to the center of sphere a, \(R^2\) is the square of the sphere radius.

The above conditions don’t have to be satisfied exactly,
instead, it only has to satisfy the stop condition. Assuming \( \alpha_1 \) and \( \alpha_2 \) are the two multipliers that don’t satisfy KKT condition, and assuming other multipliers as a constant temporarily when optimizing these two multipliers, the optimized value of \( \alpha_1 \) and \( \alpha_2 \) respectively designated symbolically as \( \lambda_1 \) and \( \lambda_2 \).

The feasible region of \( \lambda_1 \) and \( \lambda_2 \) is \( 0 \leq \lambda_1 \leq C \), \( L \leq \lambda_2 \leq H \), among them, \( L = \max(0, \alpha_1 + \alpha_2 - C) \), \( H = \min(\alpha_1 + \alpha_2, C) \).

The optimized value of \( \alpha_1 \) and \( \alpha_2 \) are as follows:

\[
\lambda_1 = \alpha_1 + \alpha_2 - \lambda_2, \quad \lambda_2 = \begin{cases} \lambda_1, & L \leq \lambda_2 \leq H, \\ \lambda_2, & H \leq \lambda_2 \leq L, \end{cases}
\]

\[
\lambda_2 = \alpha_2 + \frac{d_2^2 - d_1^2}{2\eta}.
\]

\[
\eta = K(x_1, x_1) - 2K(x_1, x_2) + K(x_2, x_2)
\]

\[
d_i^2 = \Phi(x_i) - a^2 = K_{ij} - 2\sum_{j=1}^{l}\alpha_j K_{ij} + a^2 \tag{9}
\]

\( K_{ij} \) is the abbreviation of \( K(x_i, x_j) \).

In order to decompose the training set to different subsets, we call the point which satisfy the condition of \( \alpha_i = 0 \) or \( \alpha_i = C \) the boundary point and call the point which satisfy the condition of \( 0 < \alpha_i < C \) the point in bounds.

4.2 Steps of algorithm

Step 1: Initialize \( \alpha_i = 0 \) and \( R = 0 \) for a training set, calculate \( d_i^2 \) according to eq. (9);

Step 2: Search all points in bounds, look for the point called \( x_i \) that violates KKT conditions, if the program finds the point, jump to step 4;

Step 3: Search all boundary points, look for the point called \( x_i \) that violates KKT conditions, if the program can’t find the point, jump to step 7;

Step 4: Select the point \( x_i \) that can achieve maximum value of \( d_i^2 \), optimize \( \alpha_1 \) and \( \alpha_2 \), according to the above method;

Step 5: Update \( \alpha^2 \) according to the optimized value of \( \lambda_1 \) and \( \lambda_2 \), where \( \alpha^2 = \alpha_i \alpha_j K_{ij} \), then update

\[
d_i^2 \quad \text{and} \quad d_2^2, \quad \text{calculate} \quad R^2 = \frac{d_1^2 + d_2^2}{2};
\]

Step 6: Return to step 2;

Step 7: The algorithm convergence, train is over.

5. Load forecasting based on CKM/SMO - SVM optimization program

Load forecasting based on CKM/SMO - SVM optimization steps are divided into: firstly, combining the sample data and historical relevant information using collaborative technology of mining association with weaken entropy theory, getting the fuzzy classification of prior samples and historical information.

Utilizing the chosen sample data and historical relevant information as well as digging into the collaborative knowledge about the combination of weaken entropy theory, classifying the prior sample and historical information, making clustering and correlation analysis, and completing the pretreatment of the upfront input information;

Algorithm of parameter selection method is: firstly, carrying out the training of SMO algorithm according to a set of parameter values which are given on the basis of experience;
Then choose another group according to the size of the target parameter for training again, until the satisfaction of training model is obtained, and the process is suitable to be solved by utilizing bi-level optimization algorithm. The method is divided into two layers of optimization calculation; the upper optimization goal is to find the optimal set of parameters that can lead to the optimal model which forming through the SMO training;

The task of the upper constraints is obtaining training model by SMO algorithm, which is under the optimization of the parameters in the upper constraints.

In this part, the concept of the bi-level optimization and the method of function fitting are the first to be introduced, then introducing the structure and application of steps about the usage of bi-level optimization to deal with function fitting problems.

Load forecasting based on CKM/SMO - SVM optimization steps (see Fig. 3) are as follows:

Combine the sample data and historical relevant information using collaborative technology of mining association with weaken entropy theory, getting the fuzzy classification of prior samples and historical information clustering and correlation analysis, completing the preliminary data preprocessing.

Then, apply the radial basis function as the kernel function, the optimization object to punish coefficient C and kernel parameters $\sigma$ and estimation precision $\delta$, now the basic steps will be presented as follows:

1) Reading the sample data, making assignment classification of the information about weather, temperature, holiday, season, taking the advantage of collaborative mining technique to form database.

Extracting the whole data records containing mining subject from the data source then a new database called mining subject database is formed. In each of the data records does not contain mining subject attribute of the project any longer.

Then apply the collaborative knowledge mining technology to search out the correlation $R_i$ between the subject and target value, it can be divided into three steps:

(a) Utilizing the correlation analysis to calculate the database all child database sample data sequence and the complete sequence correlation $R_i$, forming the database with association rules.

(b) Applying weaken entropy theory.

(c) Calculating the correlation $R_i = \frac{r_i \times (1 - H(p_i))}{H(p_i)}$.

(3) Permuting them from big to small on the basis of relationship then get them added together, influence factors of relationship that is 85% and eliminate the residual influence factors, repeat step (1).

(4) Reading the sample data in the database, then produce a set of $\{C, \sigma, \delta\}$.

(5) According to the current $\{C, \sigma, \delta\}$ carry out the SMO training, the steps are:

(a) Initialize $\alpha_0 = 0$, $R = 0$.

(b) Seek out a point which violates the condition of KKT.

(c) Optimal Lagrange multiplier and the threshold value.

(d) Making a judgment about whether it meet the stop condition, if not, return to step (b) above.

(6) Compute the prediction precision of the test sample.

(7) Repeat step (4) to increase the corresponding number of iterations until meet the accuracy of load forecasting.

(8) Set for load forecasting based on the search for the optimal parameter $\{C, \sigma, \delta\}$.

6. Application and Analysis

6.1 Choosing Samples

The data used in this study were provided by the electric power corporation in a region. The power load data from 0:00 at 7/1/2013 to 23:00 at 6/30/2014 are as training sample and from 0:00 at 7/1/2014 to 23:00 at 7/12/2014 are as test sample.

The parameters are chosen as follows: $1 \leq C \leq 10000$, $0.0001 \leq \sigma \leq 0.0180$, $0.0001 \leq \delta$.

In order to test the accuracy of this model, we predicted the 24-hour load at 7/1/2014 first, then calculated the relative error and root-mean-square error and compared with the prediction accuracy of SVM and BP neural network. Results of each matrix are shown in Table 1 and Fig. 4.

Test sample are predicted afterwards, the relative error is calculated and are shown in Fig. 5. The main goal is to analyze the stability of this model.

Relative error and root-mean-square relative error are used as the final evaluating:

$$E_r = \frac{X_r - Y_r}{X_r} \times 100\%$$

$$RMSRE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{X_i - Y_i}{X_i} \right)^2}$$

| Time point | CKMSOM-SVM | SVM | BP |
|------------|------------|-----|----|
| 0          | 1.27       | 1.42| 2.1 |
| 1          | 2.01       | 2.51| -3.8|
| 2          | -1.41      | -3.6| 2.7 |
| 3          | 2.3        | -2.64| 4.2 |
| 4          | -1.05      | 3.85| -3.44|
| 5          | 2.53       | -1.75| 2.88|
| 6          | 2.61       | 2.39| 1.12|
| 7          | 3.07       | 2.41| -1.60|
| 8          | 2.74       | 1.16| -2.76|
| 9          | 2.33       | 0.88| 3.62|
| 10         | 3.09       | -1.66| 3.23|
| 11         | -0.57      | 1.71| 4.77|
Root-mean-square relative error is used to contrast the predict effect of different models. The train speed is compared in Table 2.

| Methods         | CKM/SMO-SVM | SVM  | BP   |
|-----------------|-------------|------|------|
| Speed (s)       | 23          | 161  | 21   |

Figure 4. Error analysis with different models

(1) From the method of optimization speed analysis, using collaborative technology of mining association conducts pretreatment to previous samples and historical data, and after the use of SMO optimized parameters, the training speed of using electric power load forecasting SVM is only 23 seconds, the prediction accuracy can reach 2.32% and the training speed of only using SVM to predict is 161 seconds, which is the CKM / SMO-SVM algorithm more than six times. However, the training speed of using BP neural network is faster, but there is no obvious advantage with the CKM / SMO-SVM algorithm comparing, but the former is up 1.56 percentage point than the latter of prediction accuracy.

(2) From Fig. 4, the comparison shows that the approach is scientific and rational. Relevant data [24] shows that if absolute value of relative error less than 3%, the model a reasonable level of accuracy. To measure the predict results that the relative error is not more than 3%, in the 24 points, 22 points’ absolute value of relative error less than 3% used CKM/SMO – SVM model, points evenly distributed up and down close to zero; For the SVM prediction model, 8 error values are above 3%, and the other 16 points in addition to 9 in the 2% range, the error of the other 7 points are between 2% and 3%; And predicted results of BP only have 14 points whose absolute value of relative error within 3%, and these 14 points close to 3% for the most part, the largest error achieve 4.77%.

(3) According to the root mean square error comparison, the root means square error of CKM/SMO-SVM algorithm is only 2.32%, 0.67 and 1.56 percentage points lower than that of SVM model and BP neural network. So no matter from the forecast error of the prediction is from the load forecast precision points compared. Prediction accuracy of the new proposed CKM/SMO-SVM prediction model of were higher than the level of prediction accuracy of the SVM model of simple and traditional BP neural network model.

(4) According to the analysis of the relative error of predicted on the 12 day of the test set, the prediction results of CKM/SMO-SVM algorithm has good stability. The relative errors were less than 4%, most of which isless than 3%. The root mean square error is 2.37%, and the prediction accuracy of the first day is very poor, which further proves the feasibility of the algorithm.
7. Conclusion
In this paper, through the detailed analysis of the use of CKM/SMO-SVM optimized support vector machine prediction model, the combination of data utilization and weaken entropy theory is used to preprocess the sample and historical information. Meanwhile, it analyzes the support to the amount of machine parameters to determine the effect to predict the results. Then, the prediction model of hyper sphere support vector machine based on Collaborative mining association technique and SMO optimization is established. Finally, the method of obtaining high accuracy prediction results is obtained. The following conclusions are obtained by comparing the proposed method with the SVM model and the BP neural network prediction method in the actual short-term power load forecasting:

(1) In load forecasting, early data without anticipation on subsequent training speed constitute greater impact and late affects prediction accuracy, because the input information is more complex, quantitative and qualitative information cross strong. So in this paper, we use the collaborative mining association technology to fuzzy clustering and correlation analysis.

(2) By support vector machine parameters to choose a value analysis of influence of different influence on predicting accuracy degree of different, I think in the use of support vector machine prediction and parameter choice to the key role and parameter selection is reasonable tend to cause calculation under learning and studying the phenomenon, which directly affect the prediction accuracy and running time.

(3) With the passage of time prediction, CKM/SMO-SVM optimization support vector machine (SVM) model not only has higher prediction precision of stability is relatively high, and the training speed has also been greatly improved and has very strong practicability.

Supplementary information
Acknowledgements
The authors would like to thank the three anonymous reviewers for their helpful comments and suggestions.
Funding for this research was provided by "Central University Basic Research Service Fee" (2018ZD13), "111 Project" (B18021) and "Beijing Social Science Fund Project" (18GL042).

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1. State Grid Beijing Economic Research Institute, Beijing 102209, China; 2. School of Economics and Management, North China Electric Power University, Changping District, Beijing 102206 China.

References
[1] Li, Y. C., Fang, T. J., & Yu E. (2003). Study of support vector machines for short-term load forecasting. In Proceedings of the CSEE, 23(6), 55-59.
[2] Platt, J. (1998). Fast training of support vector machines using sequential minimal optimization. In B.Schölkopf, C.Burges, & Smola, A., editors, Advances in Kernel Methods-Support vector Learning, MIT Press.
[3] Zhou, J. C.; Wang, D., Quan, Y. U., & et al. (2012). An algorithm of solving the non-positive kernel SVR model based on the extended SMO method. Computer Engineering & Science
[4] Zhao, C. C., Jiang, X. A., & Jin, Y. H. (2014). Improved SMO algorithm of nonlinear regression support vector machine. Journal of Beijing University of Aeronautics and Astronautics, 40(1), 125-130.
[5] Men, C., & Wang, W. (2008). Selection of gaussian kernel parameter for SVM based on convex estimation. Advances in Neural Networks - ISNN 2008. Springer Berlin Heidelberg (pp. 709-714).
[6] Wang, G., Liu, L., & Wang, P. (2012). Soft sensor for oxygen content of flue gas based on differential evolution and the SMO algorithm. Electric Power Science and Engineering, 28(2), 71-74.
[7] Zhao, Z. D., Yuan, L., Wang, Y. X., & et al. (2007). A Novel Model of Working Set Selection for SMO Decomposition Methods[C]// iccai. IEEE Computer Society (pp. 283-290).
[8] Shao, X., Wu, K., & Liao, B. (2013). Single directional SMO algorithm for least squares support vector machines. Computational Intelligence & Neuroscience, 2013(4), 107-111.
[9] Wang, G., Liu, L., & Wang, P. (2012). Soft sensor for oxygen content of flue gas based on differential evolution and the SMO algorithm. Electric Power Science and Engineering, 28(2), 71-74.
[10] Li, Z. H., Bai, X. M., Zhou, Z. G., & et al. (2010). Method of power grid fault diagnosis based on
feature mining. Proceedings of the CSEE, 2010(10), 16-22.

[11] Niu, D. X., Wang, J. J., & Liu, J. P. (2011). Knowledge mining collaborative DESVM correction method in short-term load forecasting. Journal of Central South University of Technology, 18(4), 1211-1216.

[12] Sumithra, R., & Paul, S. (2010). Using distributed apriori association rule and classical apriori mining algorithms for grid based knowledge discovery. International Conference on Computing Communication and NETWORKING Technologies (pp. 1-5).

[13] Zheng, B. X., Xi, Y. G., Du, X. H. (2004). Outlier mining for time series data based on outlier index. ACTA AUTOMATICA SINICA, 30(1), 70-77.

[14] Niu, D. X., Wang, Y. L., Wu, D. D. (2010). Power load forecasting using support vector machine and ant colony optimization. Expert Systems with Applications, 37(3), 2531-2539.

[15] Song, X. F., Chen, W. M., Chen, Y. P., & et al. (2009). Candidate working set strategy based SMO algorithm in support vector machine. Information Processing & Management, 45(5), 584-592.