Heating load prediction based on particle swarm optimization support vector machine

Xiuchao Chen1,a*, Shenghui Wang1, Xing Jin1
1Changchun university of technology institute of electrical and electronic engineering, Changchun, Jilin, China
aemail:sjjxk@mail.ccut.edu.cn
*corresponding author: *email:exc1654056374@sina.com Telephone number:13134452292

Abstract: Heating load is affected by many uncertain factors, which makes it show certain randomness. To further improve the heating load forecasting accuracy, reduce the prediction error, using cross validation (CV) ideology in the choice of a model of performance evaluation and the superiority, combined with the advantages of particle swarm optimization (PSO), which is easy to implement and has stronger global optimization ability, the important parameters (penalty factor C and RBF kernel function parameter γ) are optimized, and the best parameters are automatically found in the training set, so as to obtain the best training model. Compared with other algorithms, the model precision of this method is improved a lot, and the prediction result is more accurate.

1.Introduction
Central heating system is a very complex thermodynamic system, which has strong nonlinear, great thermal inertia, great hysteresis and other characteristics. The operation regulation of heating system is also affected by the changes of external environment, the operation regulation mode of thermal pipe network, the differences between residents and the introduction of national policies and regulations.

In order to save energy and use energy efficiently, it is necessary to accurately predict the future heating load, so that the heat provided by the heating system matches the heat required by users. Short-term load prediction is the premise of reasonable arrangement of heat energy production, transportation and distribution, and also one of the important functions of the energy management system of heat network. The accuracy of heat supply load prediction directly affects the security and economy of the system operation. At present, common prediction methods include time series, support vector machine regression, neural network, grey system model.

The prediction accuracy of time series is relatively low, requiring high data [1-2]. SVM can better solve the problems of small sample size, nonlinear and over-fitting [3]. Although neural network has high accuracy, it has problems of over-fitting and falling into local minimum value [4]. The accuracy of grey system model depends on the exponential growth trend of original data. The selection of SVM parameters determines the performance of the prediction model, but the selection of SVM parameters lacks theoretical basis and effective methods, which need to rely on subjective experience and continuous trial calculation.

Considering the characteristics and requirements of heating system data, this paper uses the method of cross validation and particle swarm optimization algorithm to optimize the important parameters of
SVM model, and obtains the optimal parameters, which are applied to the prediction of heating load to observe the accuracy of the parameter model obtained by this method.

2. Related theories

2.1 The SVM theory

SVM can solve the problem of under-fitting and over-fitting well and has good generalization ability. SVR is a regression problem generalized on the basis of SVM theory, which has good performance in nonlinear and high-dimension problems. To estimate the regression function with SVM, the basic idea is to use a nonlinear mapping to map the data in the input space to a high-dimensional feature space, and then do linear regression in this high-dimensional feature space. For the training sample set \((x_i, y_i), x_i \in \mathbb{R}^n, x_i\) is the input variable; \(y_i\) is the output variable, \(y_i \in \mathbb{R}; i\) is the number of training samples, \(i = 1, 2, ..., n\).

By estimating function \(f(x) = \omega \times \phi(x) + b\) regression estimation is performed on the training sample set, where \(b\) is the threshold, \(\omega\) and \(b\) can be obtained by regularization risk functional minimization.

\[
R(f) = \sum_{i=1}^{n} C(\xi_i) + \lambda \|\omega\|^2
\]  

(1)

\(R(f)\) is the target function; \(n\) is the number of training samples; \(\lambda\) is the adjustment constant; \(C\) is the error penalty factor; \(\|\omega\|^2\) reflect the complexity of \(f\) flattening in higher dimensional space.

Loss function \([6]\):

\[
|y - f(x)|_c = \max\{0, |y - f(x) - \epsilon|\}
\]

(2)

The empirical risk function is:

\[
R_{\text{emp}}(f) = \frac{1}{n} \sum_{i=1}^{n} |y_i - f(x_i)|_c
\]

(3)

The regression function can be determined by using SVM to minimize the following objective functions:

\[
\min \left\{ \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} (\xi_i^* + \xi_i) \right\}
\]

(4)

\[
y_i - \omega \cdot \phi(x_i) - b \leq \epsilon + \xi_i^*
\]

(5)

\[
\omega \cdot \phi(x) + b - y_i \leq \epsilon + \xi_i
\]

(6)

\[
\xi_i^*, \xi_i \geq 0
\]

(7)

Where \(C\) is the weight parameter; \(\xi_i^*\) and \(\xi_i\) is relaxation factor; \(\epsilon\) is insensitive loss function. By introducing Lagrange multiplier and duality theory, the objective function is transformed into a duality problem:

\[
\max \left\{ -\frac{1}{2} \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) + \sum_{i=1}^{n} \alpha_i^* (y_i - \epsilon) - \sum_{i=1}^{n} \alpha_i (y_i - \epsilon) \right\}
\]

(8)

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

\(\alpha_i^*, \alpha_i \in [0, C]\)

(9)

By solving the above problems, the SVM regression function can be obtained:

\[
f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b
\]

(10)

2.2 Cross validation algorithm

The selection of SVM parameters has a great influence on the prediction accuracy of the regression model, and the optimal parameters can be obtained from the cross-validation method in a certain sense. The basic idea is to divide historical data into training sets and test sets. The training sets are used to train regression models, and the test sets are used to test models and evaluate models. A common CV method is k-fold Cross Validation (K-CV).

The main advantages of K-CV method are as follows:
1) This method can make all original data participate in training and testing, and fully mine limited sample data;
2) In the process of model parameter selection, the training set and test set change constantly and cross each other, which can be optimized in the whole sample space;
3) The final evaluation index of K-CV is to take the average value of all evaluation indexes of model test results, which can well prevent the occurrence of "under-learning" and "over-learning".

K-CV divides the historical data into K independent subsets, each of which is the test set, and the other K-1 subsets are the training set. In this way, K models can be obtained. The model performance index under K-CV is represented by the average value of the evaluation index of the test results of K models. The mean square error (MSE) is used as the evaluation index, and its expression is:

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (f(x_i) - y_i)^2
\]  

(11)

Type: \( f(x_i) \) is the predicted value, \( y_i \) is the actual value. The smaller the MSE, the higher the accuracy of the prediction model.

2.3 Basic principles of particle swarm optimization

Particle swarm optimization (PSO) algorithm is a new kind of stochastic optimization algorithm based on swarm intelligence.

PSO is initialized as a group of random particles (random solutions), and the ith particle is represented as \( X_i = x_{i1}, x_{i2}, ..., x_{id} \), the rate is \( V_i = v_{i1}, v_{i2}, ..., v_{id} \). The best position it's ever been in is \( p_{best,i} \), The best position experienced by all particles in the swarm is \( g_{best} \). Then the optimal solution is found by iterative calculation. With each iteration, the particle updates itself by tracking \( p_{best} \) and \( g_{best} \), the iteratively updated equation is:

\[
\begin{align*}
V_{id}(k + 1) &= V_{id}(k) + c_1 \cdot \text{rand}_1(k) \left( p_{best,id}(k) - x_{id}(k) \right) + c_2 \cdot \text{rand}_2(k) \left( g_{best,d}(k) - x_{id}(k) \right) \\
x_{id}(k + 1) &= x_{id}(k) + V_{id}(k + 1)
\end{align*}
\]  

(12)  

(13)

Type, \( V_{id}(k) \) is the velocity of the ith particle through k iterations in the d dimension, \( c_1, c_2 \) is the learning factor; \( \text{rand}_1, \text{rand}_2 \) is a random number \( \in (0,1) \). \( x_{id}(k) \) is the coordinates of the i-th particle in the d-th dimension after k iterations; \( p_{best,id} \) is the coordinates of the extreme point of the i-th particle in d-dimension, \( g_{best,d} \) is the coordinates of the global extreme point in the d-dimension of the group.

3. SVM model for heating load prediction based on PSO optimization

The steps of establishing the SVM heating load prediction model based on PSO optimization are as follows:

1) Initialize the population, the maximum evolutionary number is set as 200, the initial population size is set as 20, learning factor \( c_1 = 1.5, c_2 = 1.7 \), \( 0.01 \leq C \leq 150, 0.01 \leq \gamma \leq 1000, \varepsilon = 0.01 \), the cross-validation parameter \( k = 5 \), generate the initial value of each parameter randomly, and determine the flight speed and position of each particle;
2) Calculate the fitness function, take the mean square error as the calculation basis of the fitness function of this algorithm, and use the cross validation algorithm in the calculation process, so as to ensure that the best fitness function under the k-CV idea is obtained.
3) Let the current position of each particle be the individual extreme value \( p_{id} \), the global extreme value \( g_{id} \) of the current population is the individual extreme value of the particle with the best fitness.
4) If the calculated value of the optimal fitness function meets the set termination conditions, go to Step 6; otherwise, go to Step 2;
5) Update the position and velocity of particles with the iterative update formula, and then go to 3);
6) Generate a SVM heating load prediction model with optimized parameters and output test data.
4. Example analysis

4.1 Data selection and preprocessing
In this paper, the input variables are taken as the heat load of the three days before the forecast, the outdoor temperature of the forecast day, the temperature of supply and return water of the forecast day, the pressure of supply and return water, the water flow and return water flow, and the output variables are taken as the heat load of the forecast day, thus forming the prediction model. This data is derived from a heat source in Jilin, a total of 152 effective samples, 31 samples in December for heating load prediction.

Because of the instrument error, it is necessary to correct some abnormal data in the measured data. When a data in the measured data is compared with its adjacent data, if the relative error exceeds 150%, the interpolation method is used to replace the original data. In order to simplify the calculation, Mapminmax function in Matlab is used to normalize all the data into [-1, 1], so as to transform the dimensional expression into a dimensionless expression.

4.2 Result Analysis
The traditional SVM model and the currently commonly used neural network model were compared with the PSO model in this paper. The comparison between the model results and the real value is shown in Figure 1, the relative prediction errors of the three models are shown in Figure 2:

![Comparison between model results and real values](image.png)

FIG. 1 Comparison between model results and real values of the three models
As can be seen intuitively from the above two figures and Table 1, the RELATIVE error of the PSO-optimized SVM model for heating load prediction is much lower than that of the traditional SVM model and the model optimized by neural network. The selection of parameters is more accurate, and the fitting curve is closer to the actual load than the other two models, and the prediction error is smaller.

| model  | C    | $\gamma$ | MAPE  |
|--------|------|----------|-------|
| SVM    | 4    | 0.0625   | 0.0179|
| GASVM  | 100  | 256.2392 | 0.0104|
| PSOSVM | 110  | 460.3874 | 0.0087|

5. Conclusion
Based on the analysis of the important parameters of SVM and the performance of the model, this paper proposes a HEATING load prediction model of SVM optimized by PSO algorithm. The important parameters of SVM are carried out by PSO and cross validation algorithm, which overcomes the blindness of SVM parameter selection. Compared with the traditional SVM parameter selection method and neural network optimization method, this paper has more clear theoretical guidance for the selection of SVM model parameters. Example analysis shows that the prediction result of this method is closer to the actual load and has good prediction accuracy.

References
[1] Wei Y B, Lin H Q, Ma Z L,(2014) Load Prediction of district Heating System based on ARMAX Model [J].Automation & instrumentation.,29 : 1-4+21.
[2] Qi W G, Zhu X L.(2013)Application of Time Series Forecasting Method in Heating Control [J].Acta electronica sinica, 31 : 268-270.
[3] Vapnik V N.(1995)The Nature of Statistical Learning Theory[M]. Springer,New York.
[4] Hao Y Z, Li D Y, HAO B.(2003)Short-term Prediction of Heat Load in Heating Metering System Based on Neural Network [J].Hvac, 33 : 105-107.
[5] Zhang J, Tian Q, (2014) Wang M P. Prediction of Heating Load Based on Cross-validation Support Vector Regression [J]. Journal of north university of China (natural science edition). 35: 565-570.