Prediction of Cooling Load of An Energy Station based on GA-SVR

Dazhou Zhao¹,²,*, Weibo Zhang³, Zhongping Zhang¹,², Fan Yang³
¹Huadian Electric Power Research Institute Co., LTD, Hangzhou, China.
²Zhejiang Provincial Key Laboratory of Energy Storage and Building Energy-Saving Technology, Hangzhou, China.
³China Huadian Corporation LTD. Tianjin Company, Tianjin, China.

*Corresponding author e-mail: seudzz@126.com

Abstract. Taking an energy station as the research object, the external dry bulb temperature and load values at t-1, t-2, t-3 moments were selected as input parameters, and the load value at t moment was used as output parameters to establish the SVR(Support Vector Regress) cooling load prediction model, the key parameters of SVR are optimized by GA(Genetic Algorithm). The results show that the maximum absolute error between the predicted value and the actual value is 4.83 GJ/h, the maximum relative error is 9.2 %, the average absolute error is 1.25 GJ/h, and the average relative error is 2.4 %.

1. Introduction
An energy center is located in the northwest corner of a university in Shanghai, covers an area of approximately 5100m², the total construction area is about 6000m².

The project plans to build five sets of 4.4 MW gas internal combustion generators and one set of 1.25 MW small gas turbine. In the first phase, three sets of 4.4 MW gas internal combustion generator systems and one set of 1.25 MW small gas turbine were built.
The energy station supplies the surrounding buildings with cold and heat energy. High-grade heat energy used for power generation firstly, the middle-grade thermal energy of both the flue gas and the cylinder liner used for heating or cooling, according to the principle of "temperature matching, ladder utilization". Vacuum boiler as heating peak adjustment and backup equipment, electric refrigeration units as refrigeration and backup equipment. in addition, solar water heaters are also available to generate hot water.

In order to predict the energy load accurately to guide the operation of the system, the SVR cooling load prediction model is established through the energy station historical energy data, and the relevant parameters are optimized by GA, the best prediction results are obtained [1].

2. Mathematical method

2.1. SVM
In the mid-1990s, Vapnik et al. proposed a new machine learning method based on Statistical Learning Theory (SLT), that is supports vector machines (SVM) algorithms. The SVM algorithm based on the principle of minimization structural risk , has a strong generalization ability, and can ensure that the solution is the global optimal solution. These advantages make the SVM become one of the best modeling tools at present and is still in continuous development. stage.

SVM mapping the input space to a high-dimensional feature space through nonlinear transformation and fitting the sample data with a linear function in this feature space. Many nonlinear classification problems that are difficult to handle in low-dimensional space, but easy to obtain optimal classification by conversion to high-dimensional space. SVM does not directly calculate the value of a complex nonlinear transformation, but instead calculates the inner product of a nonlinear transformation, that is the kernel function k(x, y), which solves the high-dimensional calculation problem [2].

2.2. Input Parameters
The cooling load is related to meteorological information such as outdoor air temperature, relative humidity, solar radiation intensity, wind speed, wind direction, etc. At the same time, it is also affected by factors such as power price policy and user habits, the model is complex and non-linear, so the secondary factors must be ignored., Simplification of the model according to the main factors with greater load impact.

The results show that, outdoor temperature is strongly correlated with cooling load, also temperature is the most easily detected weather information [3]. Therefore, assumes that the cooling load is only related to the outdoor dry bulb temperature and load of past moments. So, the hourly cooling load is regressed and the load at the next moment is predicted by using SVM.

Figure 2. Annual Outdoor Temperature in Shanghai.
Six input parameters and one output parameter are defined in the model. The input parameters are actual load value \( Q_{t-1}, Q_{t-2} \) at \( t-2 \), \( Q_{t-3} \) at time at \( t-1, t-2 \) and \( t-3 \) respectively, and outdoor dry bulb temperature \( T_{t-1}, T_{t-2} \) and \( T_{t-3} \) at the time of \( t-1, t-2 \) and \( t-3 \) respectively, the parameters are expressed by a vector \( X \), \( X = [ Q_{t-1}, Q_{t-2}, Q_{t-3}, T_{t-1}, T_{t-2}, T_{t-3}] \) and the cooling load \( Q_t \) whose output parameter is \( T \) time is represented by \( Y \).

The relationship between input and output parameters is:

\[
F(x) = wT\phi(x) + b \tag{1}
\]

\( \phi(x) \) is a nonlinear mapping from input space to high-dimensional feature space, kernel function is introduced to get position parameters \( W \) (weight vector) and \( B \) (threshold), kernel functions mainly include:

1. Polynomial kernel function:

\[
K(x, x_i) = [(x \cdot x_i) + 1]^d \tag{2}
\]

2. Radial Basis kernel function:

\[
K(x, x_i) = \exp \left(-\frac{(x-x_i)^2}{\sigma^2}\right) \tag{3}
\]

3. Sigmoid kernel function:

\[
K(x, x_i) = \tanh\left(\gamma(x \cdot x_i) + c\right) \tag{4}
\]

The kernel function selected for the SVR model is radial basis function. In general, \( g \) has a greater impact on the generalization performance of the SVR model. If the value is small, the support vector connection will be relatively loose, and it will easily generate learning phenomena, so the generalization ability is poor. If the value is too large, the influence between support vectors is too strong, and the accuracy of the model is worse, and it is prone to under-learning phenomena. The penalty factor \( C \) determines the degree of influence of empirical risk on the model. If \( C \) becomes larger, empirical risk increases. If the value tends to infinity, minimization of structural risk tends to minimize empirical risk; If \( C \) is reduced and the empirical risk is reduced, the established model not truly reflect the characteristics of the object and lose the sense of modeling if \( C \) is too small [4].

2.3. Genetic Algorithm

Genetic Algorithm (GA) is a globally optimized search algorithm that simulates the evolution of organisms in nature. The algorithm mainly finds satisfactory optimization results through the steps of coding, selection, crossover and variation.

The genetic algorithm is used to optimize the parameters of the support vector machine. First, the initial population needs to be constructed after encoding the nuclear parameters and factor \( C \) and \( g \), then the genetic iteration is performed until the optimal solution is finally found. Its fitness value are the results of training the chromosome decoding as a training parameter of the support vector machine.

3. Calculation of the model

The entire campus cooling load were selected in August 2018 as the sample parameter. Time interval is 1 hour, a total of 744 sets of data were obtained. Some of the data is shown in the table below.
Table 1. Partial input and output data.

| Value of load at t-3(GJ/h) | Value of load at t-2(GJ/h) | Value of load at t-1(GJ/h) | Value of temperature at t-3(℃) | Value of temperature at t-2(℃) | Value of temperature at t-1(℃) | Value of load at t(GJ/h) |
|---------------------------|---------------------------|---------------------------|-------------------------------|-------------------------------|-------------------------------|---------------------------|
| 50                        | 49                        | 49                        | 28.3                          | 28.3                          | 28.4                          | 49                        |
| 49                        | 50                        | 49                        | 28.3                          | 28.4                          | 28.4                          | 55                        |
| 55                        | 49                        | 50                        | 28.4                          | 28.4                          | 28.1                          | 71                        |
| 71                        | 55                        | 49                        | 28.4                          | 28.1                          | 28.9                          | 69                        |
| 69                        | 71                        | 55                        | 28.1                          | 28.9                          | 29.7                          | 70                        |
| 70                        | 69                        | 71                        | 28.9                          | 29.7                          | 30.5                          | 72                        |
| 72                        | 70                        | 69                        | 29.7                          | 30.5                          | 32.5                          | 69                        |
| 69                        | 72                        | 70                        | 30.5                          | 32.5                          | 34.4                          | 72                        |
| 72                        | 69                        | 72                        | 32.5                          | 34.4                          | 36.4                          | 70                        |

The previous 600 sets of data were used as training values, and the next 144 sets of data were used as predictions in MATLAB. The best parameter $g = 7.0591$, $C = 9.5837$ using GA genetic algorithm.

Figure 3. Fitness Curve

Using the best parameters $C$ and $g$ as initial conditions, the predicted result of the load is compared with the actual value as shown in the figure below.
Figure 4. Comparison of predicted and actual values.

The results show that, the maximum absolute error of 4.83 GJ/h, the maximum relative error of 9.2 %, the average absolute error of 1.25 GJ/h, and the average relative error of 2.4 % can be obtained.

4. Conclusion
The GA-SVR cooling load prediction model was established using August energy supply data combined with historical weather data from an energy center. The first 600 data sets were used as training and the last 144 data sets were used for forecasting. The best C and g are optimized by the GA algorithm. The final prediction results show that the maximum absolute error between the predicted value and the actual value is 4.83 GJ/h, the maximum relative error is 9.2 %, the average absolute error is 1.25 GJ/h, and the average relative error is 2.4 %. The model can predict the short-term cooling load well.

References
[1] Cortes C,V. Support Vector Network[J]. Machine Learning, 1995, 20: 273-297.
[2] Florence L,M, Fred L. Vapnik learning theory applied to energy consumption forecasts in residential buildings[J]. International Journal of Computer Mathematics, 2008, 85(10): 1563-1588.
[3] TSE W L, SO A T P. A weather information-based intelligent ice storage control system[J]. ASHRAE Transactions, 2002, 108 (1): 18-29.
[4] Zhang Jiao, TIAN qi, WANG Meiping. Heating Load prediction based on support vector regression machine with parameters optimized by genetic algorithm[J]. HV&AC, 2017, 47 (2): 104-108.