Prediction of circulating water loss based on support vector machine and neural network

Aiming Yin¹, Fan Cao, Xuliang Jin, Lei Dong, Jinfeng Nie and Lin Ma

Thermal Power Technology Research Institute of China Datang Corporation Science and Technology Research Institute, Beijing 100040, China

¹ Email: yinaiming@cdt-kxjs.com

Abstract. Based on the operational data of the circulating water system in a thermal power plant, BP neural network and support vector machine regression were used to establish the prediction model of circulation water system evaporation and wind blow loss. The trail method was used to improve the BP neural network prediction model, and for the prediction model of support vector machine regression, the kernel function and the corresponding parameters were selected through optimization. The results showed that the mean square error of the simulation results of the two models were 0.071 and 0.070 respectively in summer and 0.046 and 0.047 respectively in winter, this meets the prediction requirements of project and demonstrates high prediction accuracy. With the evaluation index of neural network model, the simulation and prediction results of the two models were compared and analysed. The results showed that the simulation results of two model were basically the same, but the support vector machine model training sample time is shorter, the convergence speed is faster, and the overall network model performance is better.

1. Introduction

The circulating cooling water system is the largest water consumption system in a thermal power plant. According to the water balance diagram of the existing project, the water consumption of the circulating cooling water system can account for 70% to 90% of the total water consumption of the whole plant, and the larger the installed capacity, the greater the percentage of the water consumption. Referring to the water consumption classification of the cooling tower, the water loss of the circulating water system mainly includes evaporation loss, wind blow loss and sewage loss. The evaporation loss accounts for 30% to 50% of total loss, and wind blow loss is about 0.1% of total circulating water loss [1-4]. Evaporation loss and wind blow loss lead to increase ion concentration in water and pipeline equipment was corroded. In order to ensure the safe operation of the equipment, discharging a part of water and add new water periodically is necessary to maintain the ion concentration within the acceptable range of pipeline equipment. Therefore, accurately calculating amount of evaporation loss and wind blow loss of water and combined with the ion concentration to calculate the amount of sewage loss can significantly save the water replenishment of circulating water system, which can achieve efficient and stable operation of the circulating water drainage system, and is important for water saving and consumption reduction of power plants.

The evaporation loss and wind blow loss are affected by many factors, such as inlet tower temperature, inlet and outlet humidity, unit operating load and inlet air volume. At the same time, these parameters are interconnected in a complex way, so the amount of water loss cannot be
determined visually. In the actual operation process, the formula method is used to estimate the evaporation loss of the circulating water system. According to the specification, the wind blow loss is limited to 0.05%~0.1% of the total circulating water, so the water consumption cannot be accurately judged.

In this paper, a method based on machine learning was proposed to mine the historical operation data of circulating cooling water system in power plant. Back Propagation (BP) neural network and support vector machine are used to establish the prediction model for the circulating water system respectively. These two model methods were compared and analysed to explore the adaptability of the thermal power plant under two typical conditions (summer and winter) for supporting the intelligent monitoring of the subsequent circulating cooling water system and the scheduling and control of water intake and drainage.

2. Research object
The data of this study were taken from a coal-fired power plant with 2×300MW units in Hebei province, China. The cooling water system was a secondary circulation water supply system with a counterflow natural ventilation cooling tower. Unit one was put into operation in June 2003, and unit two in July 2003. The average annual temperature was 9.2 °C, the average annual relative humidity was 47%, the extreme maximum temperature was 40.9 °C, and the extreme minimum temperature was -26.2 °C. In summer or under the high load operation, each 300 MW unit was equipped with two pumps to connect the cooling towers at high speed. In winter or under the low load operation, each 300 MW unit was equipped with one pump at low speed.

The research team conducted a water balance test on the power plant in December 2017 and plotted the water balance map of the whole field. The result showed that the total amount of evaporation loss and wind blow loss can be calculated as a function of the amount of water replenished of circulating water system, the amount of discharged water, the amount of the circulating cooling water users water loss, and the change of the liquid level of the cooling tower pool. The formula is as follows:

\[ Q_e + Q_w = Q_f - Q_b - Q_m - H \times 3.14 \times 40.4162 \]  

Where \( Q_e \) is evaporation loss (t/h), \( Q_w \) is wind blow loss (t/h), \( Q_f \) is replenishment water volume (t/h), \( Q_b \) is sewage loss (t/h), \( Q_m \) is cooling user lost water volume (t/h), \( H \) is liquid level drop height (m).

3. Data processing
The Supervisory Information System (SIS) of thermal power plant was used to collect the operational data (the inlet temperatures, outlet temperatures, operating loads, vacuum degrees, condensate flow rates of two units, ambient temperature, replenishment water volume, sewage loss, cooling user lost water volume, liquid level drop height) from August 1, 2017 to July 31, 2018. The samples were collected every 1 hour. The data was pre-processed using R programing language and the abnormal data was filtered. Finally, the remaining high-quality data with 4637 sets of data were obtained. Considering the amount of water loss was quite different under different operation conditions, the winter and summer data were used to establish the simulation prediction model respectively.

According to the mechanism of evaporation and wind blow loss, the inlet temperatures, outlet temperatures, operating loads, vacuum degrees, condensate flow rates of two units and ambient temperature were selected as model input variables, and the sum of evaporation loss and wind blow loss were taken as output variables. The correlation between the parameters caused the information overlapped, which affected the accuracy of the results. In this paper, the Princomp function of the R language “psych” package was used for Principal Component Analysis (PCA) to reduce the input data. The principal components with a large contribution rate were selected as the input variables, and the total evaporation and wind blow loss of the two units were output variables. To eliminate the influence of dimensions between variables, the data was normalized, and the algorithm is shown in formula (2). 80% of the data was randomly selected for model training, and the others was used to validate the model.
BP neural network model
Artificial neural networks (ANN) are useful in many fields of power plants, such as NOx emissions prediction[5]. BP neural network is an algorithm of ANN, that is a computational model composed of a large numbers of neurons (nodes) and neurons (nodes) connected to each other. It is a multilayer feedforward neural network trained by error back propagation algorithm. The signal is transmitted from the input layer, processed by each hidden layer, and transmitted to the output layer. When the actual output value cannot match the expected output value, error begin the back-propagation phase. The error is used as the basis for correcting the weight of each unit. As the basis for correcting the weights of each unit, the error is fed back to the input layer from the hidden layer and distributed to all units of each layer. Therefore, the budget speed and accuracy of the model are related to the number of nodes and the number of hidden layers [6,7].

Eight principal components obtained by principal component analysis were input variables of BP neural network, and the total amount of evaporation and wind blow loss was the output variable. Using the “neuralnet” package of R language to train model, and adjusted the model calculation speed, the number of nodes and the hidden layer by trial method to obtain the optimal BP neural network model. Figure 1 showed that the simulation values of the BP neural network model were in good agreement with the real values, and both operation conditions can be well simulated.

The results showed that the optimal model contained two hidden layers under summer or high load operation. The first hidden layer of the model had four nodes, and the second hidden layer had two nodes, and the model structure was 8-4-2-1. The average error between the real value and the predicted value was 0.070, and the correlation was 0.740. In winter or under low load conditions, the optimal model contained one hidden layer. The hidden layer had four nodes, and the model structure is 8-4-1. The average error of the real value and predicted values of the BP neural network test sample is 0.046, with the correlation of 0.769, therefore, this BP neural network model had high prediction accuracy under different operating conditions.

\[ \bar{x}_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \]  

5. Support vector machine model
Support vector regression (SVR) is a kernel function to map linearly indivisible points in low-dimensional space into linearly separable points in high-dimensional feature space, to construct linear regression hyperplane in high-dimensional feature space. When the total deviation of all sample points from the hyperplane is the smallest, it is the optimal solution of the model [8-11].

In this paper, the kernel function was a radial basis function, as shown in formula (3). After the kernel function was determined, the corresponding optimal penalty parameter c and the kernel parameter g were determined, where c represented the penalty coefficient, that was, the error tolerance, and g represented the complexity. The cross-validation method is used to select the optimal penalty
parameter $c$ and the kernel parameter $g$, and the group $c$ and $g$ values with the smallest mean square error (MSE) were selected as the parameters of the model.

$$K(x_i, x_j) = \exp(-g\|x_i - x_j\|^2), g > 0 \quad (3)$$

In order to compare with BP neural network model prediction results, the two models used the same data samples, input and output data for model training and verification. As figure 2 showed that the support vector machine model could simulate well in summer and winter conditions. Under summer or high load operation conditions, the optimal compensation coefficients $c$ and $g$ of the support vector machine model are 0.001 and 1, respectively, and the model prediction mean square error is 0.070 with the correlation of 0.739. Under the winter or low load operation conditions, the optimal compensation coefficients $c$ and $g$ of the SVR model are 0.1 and 1, respectively, and the mean square error of the model prediction is 0.047, with the correlation of 0.759. Compared with the real value and prediction value of the SVR model under both operating conditions, both the mean square errors were less than 0.1, which met the prediction requirements and the model prediction accuracy were high.

![Figure 2. Validation result diagram of support vector machine prediction model.](image)

a) In summer or under high load conditions

b) In winter or under low load conditions

6. Model comparison

Both BP neural network and support vector machine can be used to deal with nonlinear regression problems, but the theory of the two is different, and the mechanism of regression is also different. In order to accurately compare the prediction effects, comprehensively judge the prediction ability of the two models, and introduce four indicators for model evaluation. The results are shown in Table 1.

1. Average absolute percentage error;

$$MAPE = \frac{1}{N} \sum \frac{|V_p - V_r|}{V_r} \quad (4)$$

2. Mean square error;

3. Training sample time;

4. Convergence speed.

Where $MAPE$ is average absolute percentage error, $V_p$ is training sample simulation value, $V_r$ is training sample raw value.

| Evaluation index                     | BP neural network model | Support vector machine model |
|--------------------------------------|-------------------------|------------------------------|
| operating condition                  | Summer | Winter | Summer | Winter |
| MAPE                                 | 0.065  | 0.052  | 0.064  | 0.053  |
| MSE                                  | 0.071  | 0.046  | 0.070  | 0.047  |
| Training time (seconds)              | 5.8    | 6.1    | 1.4    | 1.4    |
| convergence speed                    | Slow   | Slow   | Fast   | Fast   |
| Optimality                           | Partial | Partial | Global | Global |
Results in Table 1 implied that the prediction accuracy of the support vector machine model was slightly higher than that of the BP neural network model during summer or high load operation, but the opposite is true for winter or low load operation, but the difference between the two was small. Compared with the BP neural model, the support vector machine model is faster, which was more advantageous for the sample set with large data volume and does not fall into local convergence during the calculation process. It can obtain the global optimal solution and integrate various aspects. Consider that the overall performance of the support vector machine model is better.

7. Conclusions

The evaporation and wind blow loss of the circulating water system are affected by factors such as inlet temperature, outlet temperature, dry bulb temperature, wet bulb temperature, humidity, inlet and outlet air volume, and circulating water flow, and there is no accurate theoretical function model to calculate it. However, the complex relationship between operating parameters and water loss can be easily expressed by BP neural network model and support vector machine model. The developed model constitutes an easy-to-use and powerful optimization tool which allows estimating the evaporation and wind blow loss of the circulating water system. The flow chart of such model application is given in Figure 3. BP neural network model and support vector machine were used to establish the prediction model of evaporation and wind blow loss under different operating conditions. Compared with the measured values, the mean square error of two prediction results were 0.071 and 0.070 in summer, respectively, and were 0.046 and 0.047 in winter, respectively, which met the prediction requirements of project.

![Figure 3. Application of the SVR for calculating water loss.](image)

The neural network evaluation index was introduced to compare and analyze the two models. The comprehensive data showed that the support vector machine model and BP neural network training simulation results have strong simulation ability. The support vector machine modeling training time is short, the convergence speed is fast, and the overall network model performance is better than BP neural network.

References

[1] Dong C, Li J, Luo D 2015 *International Conference on Mechatronics & Automation*
[2] Hu D, Xu Z, Yang Y 2019 *Industrial Water Treatment* 1 39
[3] Lin Y, Wang Z, Hu D 2019 *Thermal power generation* 1 48
[4] Mingkun Z, Yinyi W, Jieyu L 2019 *Journal of Physics: Conference Series*
[5] Krzywański J, Nowak W 2017 *Journal of Power Technologies* 2 97
[6] Zhao Y 2015 *International Industrial Informatics and Computer Engineering Conference*
[7] Ding S, Su C, Yu J 2011 *Artificial Intelligence Review: An International Science and Engineering Journal* 2 36
[8] Yang S, Lu W, Chen N 2005 Journal of Molecular Structure Theochem 1 719
[9] Vapnik V N The nature of statistical learning theory 1995 New York: Springer
[10] Li B, Zheng D, Sun L 2007 Neurocomputing 4 70
[11] Hong W 2011 Neurocomputing 2 74