Non-Mechanism Model for Superheater Pollution Diagnosis of Waste Incinerator Based on BP Neural Network

Maodong Li¹, Bo Yang¹, Wei Zhai¹, Xiaoqian Ma²,³, Yuqing Xia²,³ and Yan Lin²,³

¹Guangzhou Special Pressure Equipment Inspection and Research Institute, 510663, Guangzhou, China
²School of Electric Power, South China University of Technology, 510640, Guangzhou, China
³Guangdong Province Key Laboratory of Efficient and Clean Energy Utilization, 510640, China

Abstract. Dust accumulation of boiler heating surface is one of the common problems in power plants, which will affect the unit's economic and safe operation in serious cases. In this paper, the factors that affect the ash deposition in the superheater of the garbage incinerator are analyzed, and a non-mechanism model for diagnosing the ash deposition pollution in the superheater of the garbage incinerator is established by using 3-layer BP neural network. Parameters such as garbage disposal amount, furnace temperature, primary air flow amount and secondary air flow amount are selected as input vectors, parameters such as secondary superheater inlet smoke temperature and primary superheater inlet temperature are taken as output values, and real-time data of units collected by a power plant DCS system are used as sample sets to train the network after normalization processing. The trained BP network has good effect in superheater pollution diagnosis.

1. Introduction
Ash deposition often occurs in superheaters of garbage incinerators, which seriously affects the safe and economic operation of boilers. Researchers at home and abroad usually use the heat balance method to calculate the heat transfer of the heating surface [1-4]. According to the change of heat, the internal relation between the heat transfer and the contamination of the heating surface is found. This indirect method is used to diagnose the ash deposition and slagging on the heating surface of the boiler [5].

For furnace ash deposition and slagging, it is generally necessary to arrange a large number of probes to monitor the change of water wall heat flow rate [6]; For convection-dominated heating surfaces such as high-temperature superheaters, the actual convection heat transfer coefficient is calculated by the heat balance method against flue gas flow to predict the pollution of the heating surface; The ash deposition of rotary air preheater is expressed in the form of reduced pressure difference.

Artificial neural network has distributed storage of knowledge, parallel processing, strong fault tolerance, non-linear mapping capability and strong self-learning and self-adaptive capability. It does not need to know the internal structure of the system, nor need to carry out any initial assumptions and dimension reduction processing. Its essence is to realize a non-linear mapping from multiple inputs to multiple outputs [6].
Neural network has strong non-linear fitting ability, self-study ability and self-organization ability. At the same time, it has strong parallel computing capability, fault tolerance capability and generalization capability [7]. The biggest characteristic and advantage of the neural network used for modeling is that the neural network itself is a nonlinear dynamic system and can represent any nonlinear mapping. Therefore, we can establish a non-mechanism model through several groups of input data and output data, and then optimize the model through a lot of training, and finally obtain a model within the allowable error range. Then, several groups of data are used to test the prediction effect of the model so as to obtain a neural network model which can reflect the internal mapping mechanism of input values and output values.

Scholars at home and abroad have done a lot of research on the application of neural network to prediction, but few people have applied it to pollution diagnosis of superheater of garbage incinerator [8-13]. Based on the above characteristics of artificial neural network, the author has established a BP neural network diagnostic model for superheater pollution prediction of garbage incinerator by carefully analyzing the factors affecting the dust accumulation of semi-convection and semi-radiation heating area, trained and optimized the model, then tested the prediction effect of the model, and finally got better results.

2. Method

2.1. Introduction of BP Neural Network
Multilayer feedforward neural network (MLP) is a widely used neural network at present, and BP algorithm is the most famous multilayer feedforward network training algorithm [14]. As early as 1974, Werbos described this algorithm in his doctoral thesis, which was then called "dynamic feedback." [15]. Although many excellent algorithms have been produced with the development of neural network science, and BP algorithm itself has some shortcomings such as slow convergence speed, easy falling into local minima and poor generalization ability [16], it is still one of the preferred algorithms for multi-layer feedforward network training at present due to its advantages of simplicity, small computation and strong parallelism, and has been widely applied to various practical problems. Hereinafter, the multilayer feedforward neural network using the BP algorithm as the network learning algorithm is simply referred to as the "BP network".

2.2. Establishment of BP Neural Network
BP network is a unidirectional multi-layer forward network, and its structure is shown in Figure 1. In addition to input and output nodes, the network also has one or more hidden layer nodes, and there is no coupling between nodes in the same layer. The input signal passes through each hidden layer node in turn from the input layer node to the output node, and the output of each layer node only affects the output of the next layer node [17].

![Figure 1. BP model with a hidden layer](image)
The cell characteristics (transfer functions) of the nodes are usually Sigmoid type $f(x) = 1/[1 + \exp(-Bx)] (B > 0)$, but in the output layer, the cell characteristics of the nodes are sometimes linear\[18\].

(1) Node output

Hidden layer node:
$$o_j = f\left(\sum w_{ij} \times input_i - b_{ij}\right)$$

Output layer node:
$$output_k = f\left(\sum w_{o\_jk} \times o_j - b_{o}\right)$$

In the formula: $f$ is a non-linear function; $w_{ij}$ is the weight value; $input_i$ is the input value; $w_{o\_jk}$ is the weight value; $b$ is nerve unit deviation.

(2) Excitation function

In the BP network established in this paper, the hyperbolic tangent function tansig is selected as the transfer function of the hidden layer:
$$f_1(x) = \frac{1-e^x}{1+e^x}$$

The transfer function of the output layer respectively selects the linear function purelin:
$$f_2(x) = x$$

Hyperbolic logarithmic function logsig:
$$f_3(x) = \frac{1}{1+e^{-x}}$$

Hyperbolic tangent function tansig:
$$f_4(x) = \frac{1-e^x}{1+e^x}$$

(3) Error calculation

A function of the magnitude of the error between the expected output and the actual output of the reactive neural network:
$$E_p = \frac{1}{2} \sum (t_{pi} - o_{pi})^2$$

Mean square error:
$$mse = \frac{1}{N} \sum \frac{1}{2} (t_{pi} - o_{pi})$$

In the formula: $t_{pi}$ is the expected output value of the node; $o_{pi}$ is the actual output value of the node.

(4) Learning Process of Network

The learning process of the network consists of forward propagation of information and backward propagation of errors. In the process of forward propagation, input information is processed layer by layer from the input layer through the hidden layer and transmitted to the output layer. If the actual output of the output layer does not match the expected output, the error will be transferred to reverse propagation. The learning process of the neural network, that is, the setting and error correction process of the weight matrix connecting the lower node and the upper node, is modeled as follows:
\[ \Delta w_{ji}(n) = \eta \times \phi_i \times o_j \] (9)

Among them: \( \eta \) is the learning factor; \( \phi_i \) is the calculation error of output node \( i \); \( o_j \) is the actual output of output node \( j \) [19].

The weights and thresholds of the network model are the system default values. A fixed momentum term \( mc \) and a fixed learning rate \( lr \) are added to the network. The training function of the network is \( \text{traingdm} \) and the performance function is \( \text{mse} \).

According to the analysis of factors affecting superheater pollution and considering the arrangement of measuring points during the operation of the waste power plant, a BP network is established with the structure shown in fig. 2.

![Figure 2. The building of BP network structure.](image)

Input layer: a total of 13 nodes, namely garbage disposal capacity, garbage heat value, furnace temperature, primary air flow, secondary air flow, feed water flow, feed water temperature, economizer outlet water temperature, primary desuperheating water flow, secondary desuperheating water flow, main steam flow, main steam pressure and main steam temperature.

Output layer: a total of 7 nodes, namely the secondary superheater inlet flue gas temperature, superheater inlet flue gas temperature, primary superheater outlet temperature, secondary superheater inlet temperature, secondary superheater outlet temperature, final superheater inlet temperature and final superheater outlet temperature.

Hidden layer: There are many methods to determine hidden layer nodes. In this paper, a BP network with variable hidden layer neurons is designed. Through error comparison, the optimal number of hidden layer neurons is determined, and the influence of the number of hidden layer neurons on network performance is tested.

Since the number of neurons in the input layer of the network is 13 and the number of neurons in the output layer is 7, according to the empirical formula:

\[ l = \sqrt{m + n + a} \] (10)

In the formula: \( m \) and \( n \) are the number of nodes in the input layer and the output layer respectively; the value range of \( a \) is \([1, 10]\). The number of neurons in the hidden layer of the neural network should be \(5 \sim 15\). According to Kolmogorov theorem [18], if there are \( n \) nodes in the input layer, the number of nodes in the hidden layer is \(2n + 1\). Therefore, the number of neurons in the hidden layer is selected as \([5, 27]\).
2.3. Pretreatment of Input and Output Data
When the proportion of test samples to total samples is about 25%, the generalization ability of neural network is better. According to the actual operation data of the garbage incineration power plant, 260 groups of data are selected, of which 200 groups of data are used as training samples and 60 groups of data are used as prediction samples for testing the network performance. Because the magnitude of each vector in the original sample is very different, in order to facilitate calculation and prevent some neurons from reaching supersaturation, the sample is normalized. In this paper, two normalization methods, mapminmax function and mapstd function, are used to normalize the sample data to [-1, 1].

3. Results and discussion

3.1. Influence of Different Output Functions on Training Results
In this paper, three different output functions are used to train the selected data, namely purelin linear function, logsig hyperbolic logarithmic function and tansig hyperbolic tangent function. Three different output functions have different influences on the training results, and the errors of the training results are also different. The most suitable output function type can be obtained by comparing the maximum training error and the maximum prediction error under different output functions. To compare the training effects of different output functions, according to the control variable method, it can be seen that the types of normalization functions must be the same, so the training and prediction errors of different output functions after Table 1 mapstd normalization and Table 2 mapminmax normalization are respectively compared below.

Table 1. Training results of different output layers normalized by mapstd.

| Output function | Maximum training error | Maximum prediction error | Location of occurrence | Mean square error | Implicit layer number | Cycles |
|-----------------|------------------------|--------------------------|------------------------|------------------|----------------------|--------|
| Purelin         | 57.23%                 | 3.87%                    | 176                    | 0.001            | 27                   | 92     |
| Logsig          | 60.29%                 | 5.34%                    | 111                    | 0.0013           | 24                   | 39     |
| Tansig          | 31.25%                 | 10.27%                   | 116                    | 0.0015           | 27                   | 36     |

As can be seen from Table 1, under mapstd normalization method, the maximum training errors of purelin, logsig and tansig are 57.23%, 60.29% and 31.25% respectively, the maximum prediction errors are 3.87%, 5.34% and 10.27% respectively, and the mean square errors are 0.001, 0.0013 and 0.0015 respectively. Among them, the maximum training error of tansig output function is the smallest, the maximum prediction error of purelin output function is the smallest, and its mean square error is also the smallest. Therefore, combining the above results, using purelin output function has the smallest error under mapstd normalization method.

Table 2. Training Results of Different Output Layers Normalized by mapminmax.

| Output function | Maximum training error | Maximum prediction error | Location of occurrence | Mean square error | Implicit layer number | Cycles |
|-----------------|------------------------|--------------------------|------------------------|------------------|----------------------|--------|
| Purelin         | 60.39%                 | 14.57%                   | 111                    | 1.32x10^-4       | 27                   | 31     |
| Logsig          | 63.93%                 | 13.45%                   | 111                    | 2.01x10^-4       | 25                   | 78     |
| Tansig          | 51.88%                 | 8.32%                    | 116                    | 1.99x10^-4       | 24                   | 65     |

As can be seen from Table 2, under the mapminmax normalization method, the maximum training errors of purelin, logsig and tansig are 60.39%, 63.93% and 51.88% respectively, the maximum prediction errors are 14.57%, 13.45% and 8.32% respectively, and the mean square errors are 1.32x10^-4, 2.01x10^-4 and 1.99x10^-4 respectively. Among them, the maximum training error of tansig output function is the smallest, its prediction error is also the smallest, and the mean square error of purelin output function is the smallest.
3.2. Influence of Different Normalization Methods on Training Results
This paper adopts two different normalization methods to preprocess the data, namely mapstd normalization method and mapminmax normalization method. After the data are preprocessed by two different normalization methods, the training results are quite different. Table 3 shows the training results of different normalization methods.

| Output function | Maximum training error | Maximum prediction error | Mean square error | Implicit layer number | Cycles |
|-----------------|------------------------|--------------------------|------------------|-----------------------|--------|
| purelin         | mapstd 57.23%          | mapminmax 60.39%         | 3.87%            | 14.57%                | 0.001  |
|                 |                       |                          |                  |                       |        |
| logsig          | mapstd 60.29%          | mapminmax 63.93%         | 5.34%            | 13.45%                | 0.0013 |
|                 |                       |                          |                  |                       |        |
| tansig          | mapstd 31.25%          | mapminmax 51.88%         | 10.27%           | 8.32%                 | 0.0015 |
|                 |                       |                          |                  |                       |        |

As can be seen from table 3, for purelin output function, the maximum training error and the maximum prediction error of mapstd normalization method are smaller than that of mapminmax normalization method, but the mean square error of mapminmax normalization method is smaller than that of mapstd normalization method. The mean square error is the most important measure, so the mapminmax normalization method is better for purelin output function. For logsig output function, the maximum training error and maximum prediction error of mapstd normalization method are smaller than that of mapminmax normalization method, but the mean square error of mapminmax normalization method is smaller. For tansig function, the maximum training error of mapstd is smaller, the maximum prediction error of mapminmax is smaller, and its mean square error is also smaller. Based on the above results, it can be found that the mean square error after mapminmax normalization is generally small, so the mapminmax normalization method is better than mapstd normalization method.

3.3. Probability Statistics of Optimal Implicit Layers
In this paper, 200 groups of data were trained for 100 cycles, in which the number of hidden layers was cycled from 5 to 27 in each training process. Finally, the number of hidden layers was counted when the mean square error of each cycle in 100 cycles was the minimum value, so as to find out the optimal number of hidden layers. The results are shown in fig. 3.

As can be seen from fig. 3, in 100 cycles, the number of hidden layers of 5-19 layers is not optimal, which indicates that the number of hidden layers in this range is not desirable. The probability of optimal training effect for hidden layers of 20 ~ 27 layers is 2%, 1%, 2%, 8%, 9%, 18%, 29%, 31%, respectively. The probability of hidden layers of 20 ~ 25 is lower, and the probability of hidden layers of 26 and 27
is not much different. However, the comparison shows that the probability of minimum mean square error is higher when the hidden layers are 27 layers, so the optimal number of hidden layers is 27 layers.

3.4. Probability statistics of error distribution
Figs. 4 and 5 below are error distribution diagrams of training data errors and prediction errors under different normalization methods and different output functions, respectively.

**Figure 4.** a) mapminmax&purelin Training data error; b) mapstd&purelin Training data error; c) mapminmax&logsig Training data error; d) mapstd&logsig Training data error; e) mapminmax&tansig Training data error; f) mapstd&tansig Training data error.
Figure 5. a) mapminmax&purelin Error of prediction data; b) mapstd&purelin Error of prediction data; c) mapminmax&logsig Error of prediction data; d) mapstd&logsig Error of prediction data; e) Mapminmax&tansig Error of prediction data; f) mapstd&tansig Error of prediction data.

As can be seen from figs. 4 and 5, the mapminmax normalization method is more concentrated in a small error range than the mapstd method, so the mapminmax normalization method is better in terms of the index of error distribution.

3.5. The working condition corresponding to the maximum error point

| input value                     | purelin | logsig | tansig |
|---------------------------------|---------|--------|--------|
|                                 | mapstd | mapminmax | mapstd | mapminmax | mapstd | mapminmax |
| Garbage disposal capacity, t/h   | 231.03  | 26.46  | 26.46  | 95.63    | 95.63  |
| Furnace temperature, °C         | 556     | 284    | 284    | 344      | 344    |
| Primary air flow, kNm3/h        | 48.5    | 14.1   | 14.1   | 18.4     | 18.4   |
| Secondary air flow, kNm3/h      | 3.8     | 0.1    | 0.1    | 1.2      | 1.2    |
| Feed water flow, °C             | 444.8   | 46.1   | 46.1   | 154.8    | 154.8  |
| Feed water temperature, °C      | 109     | 55.1   | 55.1   | 101.1    | 101.1  |
| Economizer outlet water temperature, °C | 196 | 180 | 180 | 198 | 198 |
| Primary desuperheating water flow, t/d | 5.55 | 0.01 | 0.01 | 0 | 0 |
| Secondary desuperheating water flow, t/d | 18.31 | 0 | 0 | 3.71 | 3.71 |
| Main steam flow, t/d            | 485.4   | 129.5  | 129.5  | 283.6    | 283.6  |
| Main steam pressure, MPa        | 3.88    | 2.22   | 2.22   | 2.65     | 2.65   |
| Main steam temperature, °C      | 326     | 224    | 224    | 246      | 246    |
| Calorific value of garbage, kJ/kg | 1980 | 6653 | 6653 | 3788 | 3788 |
The average values of working conditions corresponding to the 13 groups of input data are 507.21t/h, 951 °C, 60.1 kNm/h, 8.4 kNm/h, 994 °C, 185 °C, 11.36t/d, 42.49t/d, 1022.5t/d, 6.40MPa, 40MPa, 442 °C and 5913KJ/kg respectively. It can be seen from Table 4 that the values of input working conditions corresponding to the points with the largest errors are quite different from the average values, so these working conditions are abnormal, that is, the so-called extreme working conditions.

4. Conclusion
Through the research in this paper, the following conclusions can be drawn:
1. No matter which normalization method is used, when the output function adopts purelin function, the maximum training error and the maximum prediction error of a single point are both minimum;
2. When using the mapminmax normalization method, the overall error of the training process is the smallest, i.e. the mean square error is the smallest, so the mapminmax normalization method is more suitable;
3. The best prediction models are mapminmax normalization method and purelin output function. Because the probability of 27 hidden layers in 100 cycles is the highest, the best hidden layer number is 27 layers;
4. The points with the greatest prediction error are all extreme working condition points.

Acknowledgements
This study is supported by the General Administration of Quality Supervision, Inspection of Public Projects (2017QK152).

References
[1] Zhi-yong, L., F. Yu-xian, and L. Shu, Fuling Calculation of the 300 MW Boiler Convection Heat Surface and Application. ENERGY AND ENERGY CONSERVATION, 2012 (10): p. 116-118.
[2] Yuqin, Y., Research on Monitoring Method of Boiler Heating Surface Pollution Based on Cluster Analysis. 2009, North China Electric Power University (Beijing).
[3] Qiu, T., Use of Thermal Equilibrium in the Instruction of Heat Transfer. Journal of University for Staff and Workers of WISCGO, 2000 (1): p. 63-67.
[4] Jing, H., et al. Research on Heat Calculation Method of The Marine Supercharged Boiler. in Journal of Engineering Thermophysics. 2003.
[5] ONLINE ASH MONITORING FOR 300MW COAL-FIRED POWER BOILER. in Csp-e-jsme-asme International Conference on Power Engineering. 2001.
[6] Liqiang, W. and F. Zhao, On-line Monitoring of Ash Deposition in Superheater of Coal-fired Power Station Boiler. Sichuan Electric Power Technology, 2004. 27 (6): p. 9-11.
[7] Bao-hai, H., et al., Modeling a super-heater of power plant based on neural network using field data. Journal of North China Electric Power University, 2008. 35 (2): p. 66-69.
[8] DingJun-wei and SunYa-ming, Short-Term Load Forecasting Using Chaotic Learning Algorithm for Neural Network. Automation of electric power system, 2000. 24 (2): p. 32-35.
[9] Wei-ye, L., et al., Optimizing Strategy on Rough Set Neural Network Fault Diagnosis System. Proceedings of the CSEE, 2003. 23 (5): p. 98-102.
[10] Wenliang, L., et al., Multi-interval wind speed forecast model based on improved spatial correlation and RBF neural network. Electric Power Automation Equipment, 2009. 29 (6): p. 89-92.
[11] Tian-shu, B., et al., A Novel Neural Network Approach for Fault Section Estimation. Proceedings of the CSEE, 2002. 22 (2): p. 73-78.
[12] Gao-feng, F., et al., Wind Power Prediction Based on Artificial Neural Network. Proceedings of the CSEE, 2008. 28 (34): p. 118-123.
[13] Kai, C., et al., Short-Term Wind Speed Forecasting Combing Time Series and Neural Network Method. Power System Technology, 2008. 32 (8): p. 82-85.
[14] Wei, W., C. Wei-giang, and L. BO, Prediction of ups and downs of stock market by BP neural networks. Journal of Dalian University of Technology, 2001. 41 (1): p. 9-15.
[15] Werbos, P., Beyond Regression: "New Tools for Prediction and Analysis in the Behavioral Sciences. Ph. D. dissertation, Harvard University, 1974.
[16] Sarkar, D., Methods to Speed Up Error Back-Propagation Learning Algorithm. Acm Computing Surveys, 1995. 27 (4): p. 519-544.
[17] Juan, J., M. Xiao-qian, and Y. Zhao-sheng, Prediction models for incinerator slagging based on BP neural network optimized by genetic algorithm. Renewable Energy Resources, 2010. 28 (4): p. 80-84.
[18] MATLAB 6.5 assisted neural network analysis and design. 2003.
[19] Chun-yan, L., Realizing Star Pattern Recognition of Star Sensor with the Neural Network Technology. 2003, Liaoning Normal University.