

Prediction of $\text{Al}(\text{OH})_3$ fluidized roasting temperature based on wavelet neural network

LI Jie(李 劫)¹, LIU Dai-fei(刘代飞)¹, DAI Xue-ru(戴学儒)², ZOU Zhong(邹 忠)¹, DING Feng-qi(丁凤其)¹

1. School of Metallurgical Science and Engineering, Central South University, Changsha 410083, China;

2. Changsha Engineering and Research Institute of Nonferrous Metallurgy, Changsha 410011, China

Received 24 October 2006; accepted 18 December 2006

Abstract: The recycle fluidization roasting in alumina production was studied and a temperature forecast model was established based on wavelet neural network that had a momentum item and an adjustable learning rate. By analyzing the roasting process, coal gas flux, aluminium hydroxide feeding and oxygen content were ascertained as the main parameters for the forecast model. The order and delay time of each parameter in the model were deduced by F test method. With 400 groups of sample data (sampled with the period of 1.5 min) for its training, a wavelet neural network model was acquired that had a structure of $\{7\bar{2}1\bar{1}\}$, i.e., seven nodes in the input layer, twenty-one nodes in the hidden layer and one node in the output layer. Testing on the prediction accuracy of the model shows that as the absolute error $\pm 5.0\text{ }^\circ\text{C}$ is adopted, the single-step prediction accuracy can achieve 90% and within 6 steps the multi-step forecast result of model for temperature is receivable.

Key words: wavelet neural networks; aluminum hydroxide; fluidized roasting; roasting temperature; modeling; prediction

1 Introduction

In alumina production, roasting is the last process, in which the attached water is dried, crystal water is removed, and $\gamma\text{-Al}_2\text{O}_3$ is partly transformed into $\alpha\text{-Al}_2\text{O}_3$. The energy consumption in the roasting process occupies about 10% of the whole energy used up in the alumina production[1] and the productivity of the roasting process directly influences the yield of alumina. As the roasting temperature is the primary factor affecting yield, quality and energy consumption, its control is very important to alumina production. If some suitable forecast model is obtained, temperature can be forecasted precisely and then measures for operation optimization can be adopted.

At present, the following three kinds of fluidized roasting technology are widely used in the industry: American flash calcinations, German recycle calcinations and Danish gas suspension calcinations. For all these roasting technologies, most existing roasting temperature models are static models, such as simple

material and energy computation models based on reaction mechanism[2]; relational equations between process parameters and the yield and the energy consumption based on regression analysis[3]; static models based on mass and energy balance and used for calculation and analysis of the process variables and the structure of every unit in the whole flow and system[4]. However, all the static models have shortages in application because they cannot fully describe the characteristics of the multi-variable, non-linear and complex coupling system caused by the solid-gas roasting reactions. In the system, the flow field, the heat field, and the density field are interdependent and inter-restricted. Therefore, a temperature forecast model must have very strong dynamic construction, self-study function and adaptive ability.

In this study, a roasting temperature forecast model was established based on artificial neural networks and wavelet analysis. With characteristics of strong fault tolerance, self-study ability, and non-linear mapping ability, neural network models have advantages to solve complex problems concerning inference, recognition,

classification and so on. But the forecast accuracy of a neural network relies on the validity of model parameters and the reasonable choice of network architecture. At present, artificial neural networks are widely applied in metallurgy field[5–6]. Wavelet analysis, a time-frequency analysis method for signal, is named as mathematical microscope. It has multi-resolution analysis ability, especially has the ability to analyze local characteristics of a signal in both time and frequency territories. As a time and frequency localization analysis method, wavelet analysis can fix the size of analysis window, but allow the change of the shape of the analysis window. By integrating small wavelet analysis packet, the neural network structure becomes hierarchical and multiresolutional. And with the time frequency localization of wavelet analysis, the network model forecast accuracy can be improved[7–10].

2 Wavelet neural network algorithms

In 1980's, GROSSMANN and MORLET[11–13] proposed the definition of wavelet of any function $f(x) \in L2(\mathbf{R})$ in $a x + b$ affine group as Eqn.(1). In Eqn.(1) and Eqn.(2), the function $\psi(x)$, which has the volatility characteristic [14], is named as Mother-wavelet. The parameters a and b mean the scaling coefficient and the shift coefficient respectively. Wavelet function can be obtained from the affine transformation of Mother-wavelet by scaling a and translating b . The parameter $1/|a|^{1/2}$ is the normalized coefficient, as expressed in Eqn.(3):

$$wf(a, b) = 1/|a|^{1/2} \left\{ \int f(x) \psi[(x - b)/a] \right\},$$

$$a \in \mathbf{R}_+, b \in \mathbf{R} \quad (x = -\infty \rightarrow +\infty) \quad (1)$$

$$\int \psi(x) dx = 0 \quad (x = -\infty \rightarrow +\infty) \quad (2)$$

$$\psi_{a,b}(x) = 1/|a|^{1/2} \left\{ \psi[(x - b)/a] \right\} \quad (3)$$

For a dynamic system, the observation inputs $x(t)$ and outputs $y(t)$ are defined as

$$\begin{aligned} x^t &= [x(1), x(2), \dots, x(t)] \\ y^t &= [y(1), y(2), \dots, y(t)] \end{aligned} \quad (4)$$

By setting parameter t as the observation time point, the serial observation sample before t is $[x^t, y^t]$ and function $y(t)$, and the forecast output after t , is defined as

$$y(t) = g(x^{t-1}, y^{t-1}) + v(t) \quad (5)$$

If the $v(t)$ value is tiny, function $g(x^{t-1}, y^{t-1})$ may be regarded as a forecast for function $y(t)$.

The relation between input (influence factors) and output (evaluation index) can be described by a BP

neural network whose hidden function is Sigmoid type defined as Eqn.(6):

$$g(x) = \sum w_i S(x) \quad (i=0, \dots, N) \quad (6)$$

where $g(x)$ is the fitting function; w_i is the weight coefficient; S is the Sigmoid function; N is the node number.

The wavelet neural network integrates wavelet transformation with neural network. By substituting wavelet function for Sigmoid function, the wavelet neural network has a stronger non-linearity approximation ability than BP neural network. The function expressed by wavelet neural network is realized by combining a series of wavelet. The value of $y(x)$ is approximated with the sum of a set of $\psi(x)$, as expressed in Eqn.(7):

$$g(x) = \sum w_i \psi_i[(x - b_i)/a_i] \quad (i=0, \dots, N) \quad (7)$$

where $g(x)$ is the fitting function; w_i is the weight coefficient; a_i is the scaling coefficient; b_i is the shift coefficient; N is the node number.

The process of wavelet neural network identification is the calculation of parameters w_i , a_i and b_i . With the smallest mean-square deviation energy function for the error evaluation, the optimization rule for computation is that the error approaches the minimum. By making $\psi_0=1$, the smallest mean-square deviation energy function is shown in Eqn.(8). In this formula, K means the number of sample:

$$E = \frac{1}{2} \sum [g_j(x) - f_j(x)]^2 \quad (j=1, \dots, K) \quad (8)$$

At present, the following wavelet functions are widely used: Haar wavelet, Shannon wavelet, Mexican-hat wavelet and Morlet wavelet and so on[15]. These functions can constitute standard orthogonal basis in $L2(\mathbf{R})$ by scaling and translating.

In this study, a satisfactory result was obtained by applying the wavelet function expressed as Eqns.(9) and (10), which were discussed in Ref.[16].

$$\psi(x) = s(x+2) - 2s(x) + s(x-2) \quad (9)$$

$$s(x) = 1/(1 + e^{-2x}) \quad (10)$$

3 Roasting temperature forecasting model

3.1 Selection of model parameters

The roasting process includes feeding, dehydration, preheating decomposition, roasting and cooling, among which roasting temperature is the crucial operation parameter. When quality is good, low temperature is advantageous to increasing yield and decreasing consumption. The practice indicated that when temperature decreased by 100 °C, about 3% energy could

be saved[17]. There are many factors influencing on roasting, such as humidity, gas fuel quality, the ratio of air to gas fuel, feeding and furnace structure. All these factors are interdependent and inter-restricted.

By analyzing the roasting process, coal gas flux, feeding and oxygen content were ascertained as the main parameters of the forecast model. The model structure is shown in Fig.1. As the actual production is a continuous process, a previous operation directly influences the present conditions of the furnace, therefore, when ascertaining the input parameters, the time succession must be taken into consideration. The parameters whose time series model orders must be determined including: temperature, coal gas flux, feeding, and oxygen content. All these parameters except temperature must have their delay time determined.

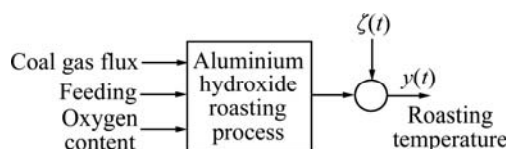


Fig.1 Logic model of aluminium hydroxide roasting

The model orders of the parameters were determined by the F test method[18], which is a general statistical method and is able to compute the remarkable degree of the variance of the loss function when the model orders of the parameters are changed. While an order increases from n_1 to n_2 ($n_1 < n_2$), the loss function $E(n)$ decreases from $E(n_1)$ to $E(n_2)$, as shown in the following equation:

$$t = [(E(n_1) - E(n_2)) / E(n_2)] [(L - 2n_2) / 2(n_2 - n_1)] \quad (11)$$

where t is in accord with F distribution named as $t-F[2(n_2 - n_1), L - 2n_2]$.

Assigning a confidence value to α , if $t \leq t_\alpha$, namely $E(n)$ does not decrease obviously, the order parameter n_1 is accepted; if $t > t_\alpha$, namely $E(n)$ decreases obviously, n_1 may not be accepted, the order must be increased and t must be recomputed until n_1 is accepted.

400 groups of sample data with a sampling period of 1.5 min were used to determine the orders of the model parameters. Through computation, the orders of temperature, coal gas flux, feeding and oxygen content were 3, 2, 1 and 1 respectively, and the delay time of coal gas flux, feeding and oxygen content were 3, 5, 1 respectively. The structure of the wavelet neural network model is shown in Fig.2, and its equation is defined as follows:

$$y(t) = \text{WNN}[y(t-1), y(t-2), y(t-3), u_1(t-3), u_1(t-4), u_2(t-5), u_3(t-1)] \quad (12)$$

where y is the temperature; u_1 is the coal gas flux; u_2 is

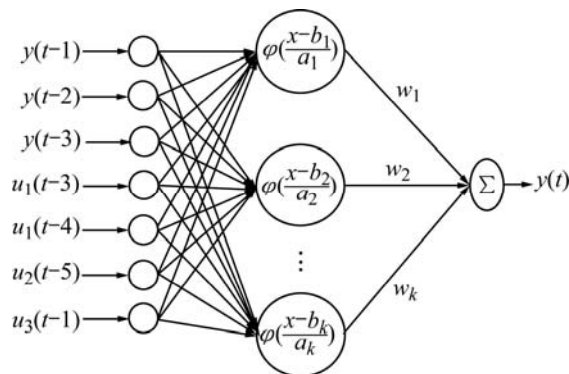


Fig.2 Structure of wavelet neural network model

the feeding; u_3 is the oxygen content; t is the sample time.

Then we can deduce the neural network single-step prediction model:

$$y_m(t+1) = \text{WNN}^1[y(t), y(t-1), y(t-2), u_1(t-2), u_1(t-3), u_2(t-4), u_3(t)] \quad (13)$$

And the multi-step prediction model is

$$y_m(t+d) = \text{WNN}^d[y(t+d-1), y(t+d-2), y(t+d-3), u_1(t+d-3), u_1(t+d-4), u_2(t+d-5), u_3(t+d-1)] \quad (14)$$

where $y_m(t+1)$ is the prediction result for time $t+1$ with the sample data of time t ; d is the prediction step; WNN^1 is the single-step prediction model; WNN^d is the d -step multi-step prediction model. For the input variable in the right of Eqn.(14) [y, u_1, u_2, u_3] whose sample time is remarked as $t+d-i$ ($i=1,2,3,4,5$), if $t+d-i \leq t$, their input values are real sample values. Whereas, if $t+d-i > t$, their input values as following $y(t+d-i), u_1(t+d-i), u_2(t+d-i)$ and $u_3(t+d-i)$ are substituted with $y_m(t+d-i), u_1(t), u_2(t)$ and $u_3(t)$, respectively. Consequently, the multi-step prediction model for time t can be constructed based on one-step prediction and multi-step recurrent computation.

3.2 Set-up of neural network model

At the end of the 20th century, the approximate representation capability of neural networks had been developed greatly[19–21]. It had been proved that single-hidden-layer forward-feed neural network had the characteristics of arbitrary approximation to any non-linear mapping function. Therefore, a single-hidden-layer neural network was adopted as the temperature forecast model in this work. As the training measure, the gradient decline rule was used, in which weightiness of neural network was modified according to the δ rule. The modeling process included forward computation and error back propagation. In forward computing, the information (neuron) was transmitted from the input layer neural nodes to the output nodes through the hidden neural nodes, with each neuron only

influencing the next one. If the expected error in output layer could not be obtained, error back propagation would be adopted and the weightiness of every node of the neural network would be modified. This process was repeated until the given precision was acquired.

3.2.1 Network learning algorithm

The number of hidden nodes was determined with the pruning method[22]. At first, a network with its number of hidden nodes much larger than the practical requirement was used; then, according to a performance criterion equation for network, the nodes and their weightiness that had no or little contribution to the performance of the network were trimmed off; finally a suitable network structure could be obtained. In view of existing shortcomings in BP algorithm, such as easily dropping into a local minimum, slow convergence rate, and inferior anti-disturbance ability, the following improved measures were adopted.

1) Attached momentum item

The application of an attached momentum item, whose function equals to a low-frequency filter, considers not only error gradient, but also the change tendency on error curved surface, which allows the change existing in network. Without momentum function, the network may fall into a local minimum. With the use of this method in the error back propagation process, a change value in direct proportion to previous weightiness change is added to present weightiness change, which is used in the calculation of a new weightiness. The weightiness modification rule is described in Eqn.(15), where β ($0 < \beta < 1$) is the momentum coefficient:

$$\Delta w_{ij}(t+1) = w_{ij}(t) - \eta \partial E(t) / \partial w_{ij} + \beta [w_{ij}(t) - w_{ij}(t-1)] \quad (15)$$

2) Adaptive adjustment of learning rate

In order to improve convergence performance in training process, a method of adaptive adjustment of learning rate was applied. The adjustment criterion was defined as follows: when the new error value becomes bigger than the old one by certain times, the learning rate will be reduced, otherwise, it may be remained invariable. When the new error value becomes smaller than the old one, the learning rate will be increased. This method can keep network learning at proper speed. This strategy is shown in Eqn.(16), in which SSE is the sum of output squared error in the output layer:

$$\begin{aligned} \eta(t+1) &= 1.05\eta(t) \quad [\text{SSE}(t+1) < \text{SSE}(t)] \\ \eta(t+1) &= 0.70\eta(t) \quad [\text{SSE}(t+1) > \text{SSE}(t)] \\ \eta(t+1) &= 1.00\eta(t) \quad [\text{SSE}(t+1) = \text{SSE}(t)] \end{aligned} \quad (16)$$

3.2.2 Results of network prediction

To set up the neural network model, 450 groups of sample data were used, in which 400 groups for training and 50 groups for prediction. When the training loop

times reached 22 375, the step-length-alterable training process was finished, with the network learning error $E=0.01$ and the finally determined structure of the network $\{7\bar{2}1\bar{1}\}$, i.e., seven nodes in the input layer, twenty-one nodes in the hidden layer and one node in the output layer. The trained network could accurately express the roasting process and would be applied for forecasting. The prediction results of the wavelet neural network are shown in Figs.3 and 4. Fig.3 indicates the change tendency of prediction error with the change of forecast step, from which it can be seen that with the forecast step increasing, the prediction error becomes bigger. And when the prediction step is lower than 6, namely, within 9 min after the last sample time, the average multi-step forecast error is less than 10°C . There is a satisfactory result shown in Fig.4: as an absolute error $\pm 5.0^\circ\text{C}$ is adopted, the single-step prediction accuracy of wavelet neural network can achieve 90%. Furthermore, from Fig.4 it can be seen that the prediction accuracy of 6 steps is worse, but the result of 5 steps is receivable. With the model prediction, the change tendency of the roasting temperature can be forecasted. If the prediction results showing the temperature may

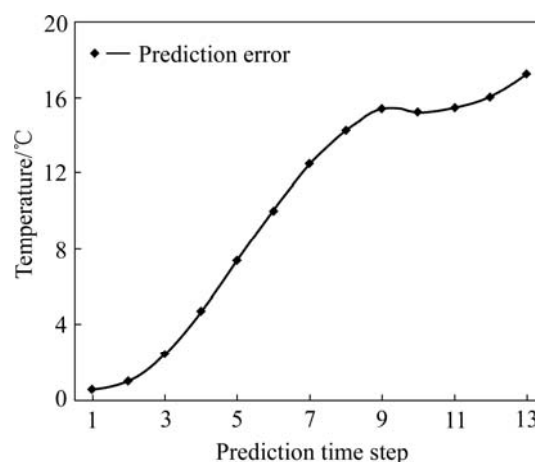


Fig.3 Change tendency of multi-step prediction error

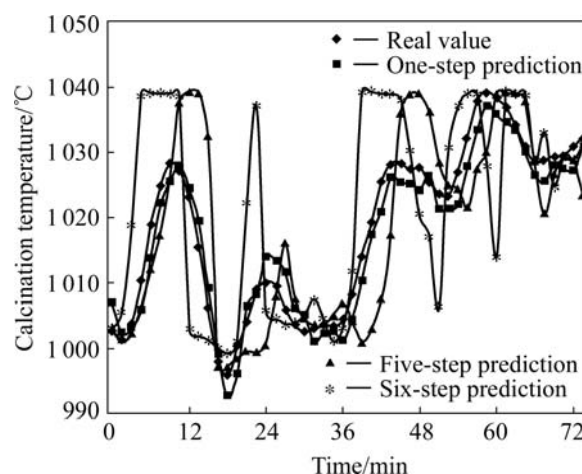


Fig.4 Result of wavelet neural network prediction

become high or low, the roasting operation parameters can be adjusted in advance, by which the roasting energy can be saved.

4 Conclusions

1) By analyzing the sample data, coal gas flux, feeding and oxygen content are ascertained as the main parameters for the temperature forecast model. The model parameter order and delay time are deduced from F test method. Then the wavelet neural network is used to identify the roasting process. The practice application indicates this model is good in roasting temperature forecast.

2) According to the process parameters analysis, the model has certain forecast ability. With forecast ability, the model provides a method for system analysis and optimization, which means that when influence factors are suitably altered, the change tendency of the roasting temperature can be analyzed. The forecast and the analysis based on the model have guiding significance for production operation.

References

- [1] YANG Chong-yu. Process technology of alumina [M]. Beijing: Metallurgy Industry Press, 1994. (in Chinese)
- [2] ZHANG Li-qiang, LI Wen-chao. Establishment of some mathematic models for $\text{Al}(\text{OH})_3$ roasting [J]. Energy Saving of Non-ferrous Metallurgy, 1998, 4: 11–15. (in Chinese)
- [3] WEI Huang. The relations between process parameters, yield and energy consumption in the production of $\text{Al}(\text{OH})_3$ [J]. Light Metals, 2003(1): 13–18. (in Chinese)
- [4] TANG Mei-qiong, LU Ji-dong, JIN Gang, HUANG Lai. Software design for $\text{Al}(\text{OH})_3$ circulation fluidization roasting system [J]. Nonferrous Metals (Extractive Metallurgy), 2004(3): 49–52. (in Chinese)
- [5] WANG Yu-tao, ZHOU Jian-chang, WANG Shi. Application of neural network model and temporal difference method to predict the silicon content of the hot metal [J]. Iron and Steel, 1999, 34(11): 7–11. (in Chinese)
- [6] TU Hai, XU Jian-lun, LI Ming. Application of neural network to the forecast of heat state of a blast furnace [J]. Journal of Shanghai University (Natural Science), 1997, 3(6): 623–627. (in Chinese)
- [7] LU Bai-quan, LI Tian-duo, LIU Zhao-hui. Control based on BP neural networks and wavelets [J]. Journal of System Simulation, 1997, 9(1): 40–48. (in Chinese)
- [8] CHEN Tao, QU Liang-sheng. The theory and application of multiresolution wavelet network [J]. China Mechanical Engineering, 1997, 8(2): 57–59. (in Chinese)
- [9] ZHANG Qing-hua, BENVENISTE A. Wavelet network [J]. IEEE Trans on Neural Networks, 1992, 3(6): 889–898.
- [10] PATI Y C, KRISHNA P S. Analysis and synthesis of feed forward network using discrete affine wavelet transformations [J]. IEEE Trans on Neural Networks, 1993, 4(1): 73–85.
- [11] GROSSMANN A, MORLET J. Decomposition of hardy functions into square integrable wavelets of constant shape [J]. SIAM J Math Anal, 1984, 15(4): 723–736.
- [12] GROUPELLAUD P, GROSSMANN A, MORLET J. Cycle-octave and related transforms in seismic signal analysis [J]. Geoprospection, 1984, 23(1): 85–102.
- [13] GROSSMANN A, MORLET J. Transforms associated to square integrable group representations (I): General results [J]. J Math Phys, 1985, 26(10): 2473–2479.
- [14] ZHAO Song-nian, XIONG Xiao-yun. The wavelet transformation and the wavelet analyze [M]. Beijing: Electronics Industry Press, 1996. (in Chinese)
- [15] NIU Dong-xiao, XING Mian. A study on wavelet neural network prediction model of time series [J]. Systems Engineering—Theory and Practice, 1999(5): 89–92. (in Chinese)
- [16] YAO Jun-feng, JIANG Jin-hong, MEI Chi, PENG Xiao-qi, REN Hong-jiu, ZHOU An-liang. Application of wavelet neural network in forecasting slag weight and components of copper-smelting converter [J]. Nonferrous Metals, 2001, 53(2): 42–44. (in Chinese)
- [17] WANG Tian-qing. Practice of lowering gaseous suspension calciner heat consumption coast [J]. Energy Saving of Non-ferrous Metallurgy, 2004, 21(4): 91–94. (in Chinese)
- [18] FANG Chong-zhi, XIAO De-yun. Process identification [M]. Beijing: Tsinghua University Press, 1988. (in Chinese)
- [19] CARROLL S M, DICKINSON B W. Construction of neural nets using the radon transform [C]// Proceedings of IJCNN. New York: IEEE Press, 1989: 607–611.
- [20] ITO Y. Representation of functions by superposition of a step or sigmoidal functions and their applications to neural network theory [J]. Neural Network, 1991, 4: 385–394.
- [21] JAROSLAW P S, KRZYSZTOF J C. On the synthesis and complexity of feedforward networks [C]// IEEE World Congress on Computational Intelligence. IEEE Neural Network, 1994: 2185–2190.
- [22] HAYKIN S. Neural networks: A comprehensive foundation [M]. 2nd Edition. Beijing: China Machine Press, 2004.

(Edited by LI Xiang-qun)