Research Article

An ELM Based Online Soft Sensing Approach for Alumina Concentration Detection

Sen Zhang, Xi Chen, and Yixin Yin

School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

Correspondence should be addressed to Sen Zhang; zhangsen@ustb.edu.cn

Received 19 August 2014; Accepted 30 October 2014

Academic Editor: Yi Jin

Copyright © 2015 Sen Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The concentration of alumina in the electrolyte is of great significance during the production of aluminum; it may affect the stability of aluminum reduction cell and the current efficiency. However, the concentration of alumina is hard to be detected online because of the special circumstance in the aluminum reduction cell. At present, there is lack of fast and accurate soft sensing methods for alumina concentration and existing methods can not meet the needs for online measurement. In this paper, a novel soft sensing method based on a modified extreme learning machine (MELM) for online measurement of the alumina concentration is proposed. The modified ELM algorithm is based on the enhanced random search which is called incremental extreme learning machine in some references. It randomly chooses the input weights and analytically determines the output weight without manual intervention. The simulation results show that the approach can give more accurate estimations of alumina concentration with faster learning speed compared with other methods such as BP and SVM.

1. Introduction

In the industrial aluminum reduction cells, the stability of the alumina concentration is the key issue to maintain high efficiency during the production of aluminum. It is easy to lead to the occurrence of the so-called “anode effect” if the alumina content in the electrolyte becomes too low (e.g., 1%-1.5%). When it occurs, cell voltage rises abruptly to 30v–50v, which directly affects the energy balance of aluminum reduction cells. It is important to avoid the high alumina concentration because alumina-rich sludge will accumulate at the base of the cell and the cell operation can be seriously disrupted [1]. Therefore, how to detect the distribution of alumina concentration in the cell in real time is the key problem to control the production of aluminum. The site environment of aluminum reduction cell, which has the characteristics of large current, strong magnetic field, high temperature, high humidity, and a lot of dust, is very severe and complex. Aluminum reduction cell is a severe nonlinear, multi-input multioutput, slow time-variety, long time-delay, and coupled process. Therefore, it is a challenge to seek an online soft sensing method for alumina concentration. Numerous investigations have been carried out on the soft measurement method for alumina concentration by the researchers. A prediction model based on wavelet neural network was proposed by Li et al. [2]. The prediction method based on linear regression and orthogonal transform is applied to improve the accuracy of the alumina concentration forecast by Lin et al. [3]. Yan and Liang proposed a predictive model of aluminum reduction cell based on LS-SVM [4]. Li et al. [5] proposed a new fuzzy expert control method based on smart identification, multicontrol mode, and decision-making mechanism to achieve the alumina concentration prediction and real time control. The GM(1, 1) model is introduced into the aluminum concentration estimate by Zhang et al. [6]. However, the computational burden of the above nonlinear predictive models is still large when the dimension of input variable increases; the learning speed and accuracy of networks are in general far slower and can not meet the requirement of real time detection.

Figure 1 shows the experimental environment in the ZunYi aluminium electrolysis factory in China. In industries, there often exist some crucial variables that can not be directly measured online due to the fact that no sensor is available in the certain complex environment or due to its high cost. Soft sensing technology is a way to solve the problems. The soft sensing technology is widely used and becomes...
2 Mathematical Problems in Engineering

Figure 1: Experimental environment.

one of the important developing directions of surveying and processing control area. In practice, the application and research of this technology have gotten more extensions and many related technologies based on it have emerged [7–10].

The main thought of the soft sensing technology is to build one model which uses variables that could be directly online measured as input and thus the variables to be estimated as output, to use many kinds of complex calculating and evaluation and to get the values of detecting variables by computer software [11–14]. So the main problem of soft sensing technology is how to build the relation model between detecting variables and other easy getting variables. By now, there are many methods to build the models, and many methods are in the trend of intersecting and mixing together. In all the methods, the artificial intelligence method is used often, such as the method based on model identification, the method based on artificial neural network, or the method based on fuzzy set theory. Some professors and scholars have done some research about the technology, such as global asymptotic stability of neural networks with multiple time-varying delays and fuzzy model-based robust control generalization performance and smaller training error [22, 23]. Figure 2 shows the single-hidden-layer feedforward neural network (SLFN) architecture.

Considering there are $N$ arbitrary distinct samples $(X_i, T_i)$, when $X_i = [x_{i1}, x_{i2}, \ldots, x_{im}]^T \in R^m$ and $T_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m$, standard SLFNs with $L$ hidden neurons and activation function $h(x)$ are mathematically modeled as

$$\sum_{i=1}^{L} \beta_i h(W_i \cdot X_i + b_i) = O_j, \quad j = 1, 2, \ldots, N,$$

where $W_i = [w_{i1}, w_{i2}, \ldots, w_{im}]^T$ is the weight vector connecting the $i$th hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the $i$th hidden neuron and the output neurons, and $b_i$ is the threshold of the $i$th hidden neuron.

The fact that standard SLFNs with $L$ hidden neurons with activation function $h(x)$ can approximate these $N$ samples with zero error means that $\sum_{j=1}^{N} \|O_j - T_j\| = 0$; there exist $\beta_j, W_j$, and $b_j$ such that

$$\sum_{i=1}^{L} \beta_i h(W_i \cdot X_i + b_i) = T_j, \quad j = 1, 2, \ldots, N. \quad (2)$$

The above $N$ equations can be written compactly as

$$H\beta = T, \quad (3)$$

where

$$H = \begin{pmatrix} h(W_1 \cdot X_1 + b_1) & \cdots & h(W_L \cdot X_1 + b_L) \\ \vdots & \ddots & \vdots \\ h(W_1 \cdot X_N + b_1) & \cdots & h(W_L \cdot X_N + b_L) \end{pmatrix}_{N \times L}, \quad \beta = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{pmatrix}_{L \times m}, \quad T = \begin{pmatrix} T_1^T \\ \vdots \\ T_N^T \end{pmatrix}_{N \times m}$$

$H$ is called the hidden layer output matrix of the neural network. ELM is to minimize the training error as well as the norm of the output weights. Minimize $\|H\beta - T\|^2$ and $\|\beta\|$. 

2. The Theory of Extreme Learning Machine

Huang proposed extreme learning machine (ELM) algorithm; ELM was originally proposed for the single-hidden-layer feedforward neural networks (SLFNs) and then it extends to the generalized SLFNs. ELM can randomly generate the input weights and the bias of hidden nodes. It uses the theory of least squares to get the output weights. The learning speed of ELM can be thousands of times faster than traditional feedforward network learning algorithms like backpropagation (BP) algorithm while obtaining better generalization performance and smaller training error [22, 23]. Figure 2 shows the single-hidden-layer feedforward network (SLFN) architecture.

The following sections are organized as follows. Section 2 shows the theory of the extreme learning machine. Section 3 proposes the modified ELM algorithm. Data selection and data preprocessing are shown in Section 4. Section 5 gives the simulation results. Section 6 summarizes the conclusions.
The minimal norm least squares method instead of the standard optimization method was used in the original implementation of ELM:

$$\tilde{\beta} = H^T T,$$

where $H^T$ is the Moore-Penrose generalized inverse of $H$. Several methods can be used to calculate the $H^T$; these methods may include orthogonal projection, orthogonalization method, iterative method, and singular value decomposition (SVD), and ELM algorithm makes use of SVD, where

$$H^T = (H^T H)^{-1} H^T.$$

The algorithm ELM can be summarized as Three-Step Learning Model. Given a training set $N = \{(X_i, T_i) \mid X_i \in \mathbb{R}^n, T_i \in \mathbb{R}^m, i = 1, 2, \ldots, N\}$, an activation function $h(x)$, and the hidden neuron number $L$, we have the following steps.

Step 1. Assign arbitrary input weighs $W_i$ and bias of hidden layer nodes $b_i$, $i = 1, 2, \ldots, L$.

Step 2. Calculate the hidden layer output matrix $H$.

Step 3. Calculate the output weights $\tilde{\beta}$.

3. The Modified ELM Algorithm

3.1. Ridge Regression Based ELM Algorithm. Ridge Regression, which was proposed by Horel and Kennard in 1970, proposes the idea to adding a small positive number on the main diagonal of the design matrix. Considering ELM algorithm, the estimation of $\tilde{\beta}^*$ can be obtained by employing the following formula:

$$\tilde{\beta}^* = (H^T H + kI)^{-1} H^T T,$$

where $k$ is the ridge parameter. Superficially, it is possible to get the inversion of design matrix $H^T H$. The following is the new results of (8) and (9):

$$E(\tilde{\beta}^*) = \sum_{i=1}^{N} \frac{H_i^2}{(\sum_{i=1}^{N} \lambda_i + k)^2} \beta \neq \beta,$$

which denotes that Ridge Regression is biased. Consider

$$(2) \quad V(\tilde{\beta}^*) = \frac{\sigma^2 \sum_{i=1}^{N} H_i^2}{(\sum_{i=1}^{N} \lambda_i + k)^2} = \frac{\sigma^2 \sum_{i=1}^{N} \lambda_i}{(\sum_{i=1}^{N} \lambda_i + k)^2} < V(\tilde{\beta}),$$

which denotes that Ridge Regression makes the estimation more stable.

Consider

$$(3) \quad \text{MSE}(\tilde{\beta}^*) = 1 \sum_{i=1}^{N} \frac{H_i^2}{(\sum_{i=1}^{N} \lambda_i + k)^2} + \beta^2 k^2$$

$$= \frac{1}{N} \left( \frac{d\gamma_1(k)}{dk} + \frac{d\gamma_2(k)}{dk} \right).$$
Mathematical Problems in Engineering

So, when \( k \to 0^+ \),
\[
\lim_{k \to 0^+} \frac{d y_1(k)}{d k} = -\frac{2\sigma^2}{(\sum_{i=1}^{\bar{N}} \lambda_i)^2} < 0,
\]
and
\[
\lim_{k \to 0^+} \frac{d y_2(k)}{d k} = 0.
\]

From the above equations, MSE(\( \hat{\beta}^* \)) is an increasing function of \( k \) when \( k \in (0, \delta) \), where \( (dy_1(k)/dk)|_{k=\delta} = (dy_2(k)/dk)|_{k=\delta} \). Therefore, the selection of parameter \( k \) is essential to the performance of Ridge Regression. In our ER-ELM algorithm, a method to determine the ridge parameter proposed by Huang [24] is used:
\[
k = \frac{\sigma^2}{\bar{\beta}},
\]
where \( \bar{\beta} \) is the ordinary ELM algorithm estimation and \( \sigma^2 = \sum_{i=1}^{N} (y_i - \hat{\beta} x_i)^2 / \nu \), \( \nu = n - 1 \).

3.2. To Select the Number of Hidden Nodes Based on the Improved ELM. The Error Minimized Extreme Learning Machine algorithm starts from a small size of ELM hidden layer and adds random hidden node (nodes) to the hidden layer, while the output weights are updated incrementally.

Suppose a SLFN, \( H_0 = H(w_1, ..., w_k, b_1, ..., b_k) \), denotes the hidden layer output matrix with \( l_0 \) hidden nodes and ridge parameter \( k_1 \) calculated by (19). Considering the poor performance due to lower number of hidden nodes, additional \( l \) hidden nodes are added to the SLFN. A new hidden layer output matrix \( H_2 \) is composed of \( H_1 \) and other \( l \) extra hidden nodes as
\[
H_2 = [H_1, H],
\]
where \( H = \begin{pmatrix} G(w_{11}, b_{11}, x_{11}, k_{11}) & ... & G(w_{1k}, b_{1k}, x_{1k}, k_{1k}) \\ ... & ... & ... \\ G(w_{l1}, b_{l1}, x_{l1}, k_{l1}) & ... & G(w_{lk}, b_{lk}, x_{lk}, k_{lk}) \end{pmatrix}_{N \times (l_0 + l_1)} \), \( l_1 = l \), and \( k_2 \) is ridge parameter of \( H^T H \).

Huang had proved \( E(H_2) = \min ||H_2\beta_2 - T|| \leq E(H_1) = \min ||H_1\beta_1 - T|| \), where \( E(H) \) denotes the output error function of SLFNs. We set a stopping criterion for the iterative algorithm as follows:
\[
|E(H_{n+1}) - E(H_n)| < \epsilon,
\]
where \( \epsilon > 0 \) is called the target error.

Consider
\[
H_2^+ = \begin{pmatrix} H_2^T H_2 + K \end{pmatrix}^{-1} H_2^T
= \begin{pmatrix} H_2^T & H_1 \end{pmatrix} \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix} \begin{pmatrix} H_1^T \cr H_2^T \end{pmatrix},
\]
where \( K = kI \).

In order to facilitate the calculation, the inversion is substituted as follows:
\[
\begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix}
= \begin{pmatrix} H_2^T & H \\ \gamma \end{pmatrix}^{-1} = \begin{pmatrix} K_1 & 0 & 0 & 0 \end{pmatrix}^{-1},
\]
where
\[
\varphi_{11} = (H_2^T H_2 + K)\begin{pmatrix} I & 0 \\ 0 & K_2 \end{pmatrix}^{-1}
+ (H_2^T H_1 + K_1)\begin{pmatrix} I & 0 \\ 0 & K_2 \end{pmatrix}^{-1}
+ (H_1^T H_1 + K_1)\begin{pmatrix} I & 0 \\ 0 & K_2 \end{pmatrix}^{-1},
\]
\[
\varphi_{12} = -(H_2^T H_1 + K_1)\begin{pmatrix} I & 0 \\ 0 & K_2 \end{pmatrix}^{-1},
\]
\[
\varphi_{21} = -(H_2^T H_1 + K_1)\begin{pmatrix} I & 0 \\ 0 & K_2 \end{pmatrix}^{-1},
\]
\[
\varphi_{22} = R^{-1},
\]
\[
R = H_2^T H_2 + K_2 - H_2^T H_1 (H_1^T H_1 + K_1)\begin{pmatrix} I & 0 \\ 0 & K_2 \end{pmatrix}^{-1} H_1^T H_1.
\]

So
\[
H_2^+ = \begin{pmatrix} U \\ D \end{pmatrix} = \begin{pmatrix} \varphi_{11} H_1^T + \varphi_{12} H_2^T \\ \varphi_{21} H_1^T + \varphi_{22} H_2^T \end{pmatrix},
\]
\[
D = R^{-1} H_2^T - R^{-1} H_1^T (H_1^T H_1 + K_1)\begin{pmatrix} I & 0 \\ 0 & K_2 \end{pmatrix}^{-1} H_1^T H_1.
\]

where \( M = I - H_1 H_1^+ \).

Similarly, we get \( U = H_1^+ - H_1^+ HD \).

Now, a new hidden layer output matrix is obtained which has less output error. Then, we can update the output weight matrix based on the new hidden layer output matrix.

4. Data Selection and Preprocessing

In the aluminum production, through referring to relative documents and soliciting experts opinion, there are many factors that affect the alumina concentration, such as alumina feeding speed, cell voltage, series current, current of anode rod, voltage between anode rod and cathode bar, and bath temperature [25]. Existing predictive method of alumina concentration is to get an average alumina content in the aluminum reduction cell; the distribution of alumina concentration in the cell is not known yet. At present, people want
to know the distribution situation of alumina content in the cell in order to know the process of production better and control the production of aluminum more accurately. Our experimental aluminum reduction cell has 24 anode rods; we can get the current signal of 24 anode rods, voltage signal between anode rods and cathode bars, and alumina concentration signal below the 24 anode rods. Through experiment and analysis, we find that the voltage signal between anode rod and cathode bar can reflect the alumina concentration below the anode rod more accurately. So we can use the voltage signal between anode rod and cathode bar in the aluminum reduction cell to reflect the distribution of alumina concentration. To use the ELM algorithm, we decide to select voltage between anode rod and cathode bar variable as model input according to experiments, and output variable is alumina concentration in the electrolyte below the anode rod.

It is necessary to do the data denoising preprocessing before using the algorithm, because the process of data collection may introduce noise, where data collection is affected by interference and influence of all kinds of noise signal. At present, there are two denoising methods, the traditional filtering method and the wavelet denoising method. Traditional denoising method is based on Fourier analysis and can always be used in the environment where signal and noise are very small. While wavelet analysis is known as the “microscope” of mathematical analysis, it is a time-frequency analysis method of signal, with the characteristics of multiresolution analysis [26]. Different signals may choose the different denoising methods. Through experiment and comparison, we decide to use the wavelet denoising method in view of the voltage signal between anode rod and cathode bar.

5. Simulative Results of Alumina Concentration

In this paper, we select voltage between anode rod and cathode bar as model input and alumina concentration in the electrolyte below the anode rod as model output.

First, ELM algorithm is applied to build soft sensing model of alumina concentration. In general, active functions play an important role in computing of neural networks. Widely used active functions in the ELM algorithm are \( \text{sig, sin, hardlim, and radbas} \). Through comparison, we find that the \( \text{sig} \) function has more outstanding performance than \( \text{sin, hardlim, and radbas} \). So we apply \( \text{sig} \) function as active function in the ELM algorithm. In order to make comparison, we also use the BP and SVM algorithm to build soft sensing models of alumina concentration [27, 28]. The BP parameters are chosen as 17 hidden layer neurons and 1 output layer neuron and transfer function of hidden layer is tan-sigmoid and transfer function of output layer is \( \text{linear} \).

Data in this paper came from a 350 kA prebaked aluminum reduction cell in the ZunYi aluminium electrolysis factory; Figure 3 shows experimental aluminum reduction cell and anode rod. We got the data of voltage between anode rod and cathode bar through the voltage measurement instruments. At the same time, we obtained the electrolyte below the experimental anode rod in the experimental aluminum reduction cell, and then we used the X fluorescence spectrometer in the laboratory to analyze the alumina concentration. Figure 4 shows the sampled electrolyte. We got the experimental data through this method. We select 100 pairs as samples to construct sample set \((X_i, T_i)\), where \(X_i \in R\) denotes voltage between anode rod and cathode bar and \(T_i \in R\) denotes alumina concentration obtained at the same time with the voltage signal.

The ELM, BP, and SVM soft sensing models are trained by the same sample set. All the data are preprocessed before training and simulating in the algorithm. For the purpose of comparing the three models quantitatively, we substitute ten \(X_i\), of new sample (out of the sample set) into ELM, BP, and SVM models, respectively, to calculate corresponding \(\hat{T}_i\); simulation results of ELM model are shown in Figure 5; the alumina concentration results of three models are shown in Table 1. To see the results more clearly, the actual alumina concentration values and simulation results are displayed under the same axis which is shown in Figure 6. There are five performance indicators to measure the quality of algorithm: Training Time, Testing Time, Training RMSE (root mean square error), Testing RMSE, and Average Relative Error (ARE), where

\[
\text{ARE} = \frac{1}{10} \sum_{i=1}^{10} \left| \frac{T_i - \hat{T}_i}{T_i} \right| \times 100\%.
\]

The simulation results of five performance indicators are shown in Table 2.

From Figure 7, the horizontal axis denotes the values of voltage between anode rod and cathode bar and the vertical axis denotes the alumina concentration below the experimental anode rod. The simulation results of ELM approach the sample data in a certain range. As is shown in Table 1 and Figure 8, the values of output of ELM model are closer to the actual values of alumina concentration. From Table 2, Training Time of ELM model is 0s whose
Table 1: Simulation alumina concentration results of three models.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual value (%)</td>
<td>1.94</td>
<td>2.37</td>
<td>2.03</td>
<td>1.95</td>
<td>2.16</td>
<td>2.02</td>
<td>1.99</td>
<td>1.94</td>
<td>2.64</td>
<td>2.94</td>
</tr>
<tr>
<td>Output of ELM (%)</td>
<td>1.8164</td>
<td>2.0641</td>
<td>2.0739</td>
<td>2.1544</td>
<td>1.9423</td>
<td>1.9065</td>
<td>1.9166</td>
<td>2.047</td>
<td>2.6761</td>
<td>2.6443</td>
</tr>
<tr>
<td>Output of BP (%)</td>
<td>1.7551</td>
<td>2.015</td>
<td>1.9953</td>
<td>2.4558</td>
<td>1.9594</td>
<td>1.9047</td>
<td>1.9510</td>
<td>2.0183</td>
<td>3.0956</td>
<td>2.9993</td>
</tr>
<tr>
<td>Output of SVM (%)</td>
<td>1.9176</td>
<td>2.003</td>
<td>2.006</td>
<td>2.0917</td>
<td>1.8828</td>
<td>1.8834</td>
<td>1.8979</td>
<td>1.8833</td>
<td>2.301</td>
<td>2.3042</td>
</tr>
</tbody>
</table>

Training RMSE is 0.0926% and Testing RMSE is 0.1784%, while Training Time of BP model is 28.3281 s whose Training RMSE is 0.1698% and Testing RMSE is 0.2626%. In the SVM model, Training Time is 0.0156 s and Training RMSE and Testing RMSE are 0.2635% and 0.2797%. In terms of Average Relative Error (ARE) performance indicator, ELM has the smallest ARE (6.78%) in all of the three algorithms. It is clear that SVM model has faster learning speed and less Average Relative Error than BP model, but SVM model is slower and less accurate than ELM model. So the soft sensing measurement model based on ELM algorithm has better performance than BP model and SVM model.

6. Conclusions

Alumina concentration is very important in the aluminum electrolysis, which may affect the performance of aluminum reduction cell. It is difficult to measure the alumina content due to the complicated environment of aluminum reduction cell. This paper proposes a novel soft sensing method of alumina concentration in the electrolyte based on extreme learning machine (ELM) and builds the relationship between...
Table 2: The comparison of performance indicators.

<table>
<thead>
<tr>
<th></th>
<th>Training Time (s)</th>
<th>Testing Time (s)</th>
<th>Training RMSE</th>
<th>Testing RMSE</th>
<th>ARE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELM</td>
<td>0</td>
<td>0</td>
<td>0.0926</td>
<td>0.1784</td>
<td>6.78</td>
</tr>
<tr>
<td>BP</td>
<td>28.3281</td>
<td>0.0156</td>
<td>0.1698</td>
<td>0.2626</td>
<td>9.17</td>
</tr>
<tr>
<td>SVM</td>
<td>0.0156</td>
<td>0</td>
<td>0.2635</td>
<td>0.2797</td>
<td>8.83</td>
</tr>
</tbody>
</table>

alumina concentration and voltage between anode rod and cathode bar. Through the simulation results and comparison of BP model and SVM model, we can see the validity and advantage of the proposed method. This method is able to effectively achieve rapid and reliable estimation of alumina concentration in a relatively short time.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has been supported by the National High Technology Research and Development Program (“863” Program) of China (Grant no. 2013AA040705).

References


