Neurocomputing 174 (2016) 232–237

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

An improved ELM algorithm for the measurement of hot metal temperature in blast furnace

Haigang Zhang, Yixin Yin, Sen Zhang *

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

A R T I C L E I N F O

Article history:
Received 17 September 2014
Received in revised form
13 April 2015
Accepted 14 April 2015
Available online 19 August 2015

Keywords:
Extreme learning machine
Principle component analysis
Blast furnace
Hot metal temperature
Soft measurement model

A B S T R A C T

This note proposes a modified ELM algorithm named P-ELM subject to how to get rid of the multicollinear problem in calculation based on PCA technique. By reducing the dimension of hidden layer output matrix (H) without loss major information through PCA theory, the proposed P-ELM algorithm can not only ensure the full column rank of newly generated hidden layer output matrix (H'), but also improve the training speed. In order to verify the effectiveness of P-ELM algorithm, this paper establishes a soft measurement model for hot metal temperature in the blast furnace (BF). Some comparative simulation results with other famous feedforward neural network and the ordinary ELM algorithm with its variants illustrate the better generalization performance and stability of the proposed P-ELM algorithm.

1. Introduction

Extreme Learning Machine (ELM), proposed by Huang et al., is a novel machine learning algorithm for single hidden layer feedforward neural networks (SLFNs) [1]. In contrast to other famous neural networks algorithms, the learning parameters, i.e., the input weights and the bias of hidden nodes in ELM are generated randomly without human tuning, while the output weights are determined based on the method of least squares [1,2,6]. Unlike the traditional feedforward neural networks learning algorithms, ELM has the fast training speed and gets rid of the opportunity to converge to local minima. Nowadays, because of its good generalization, ELM algorithm has been applied in many aspects like image segmentation [3], fault diagnosis [4,5], human action recognition, and human computer interface [7].

So far, ELM algorithm has drawn more and more attention. However, the multicollinear problem in ELM algorithm deteriorates its generalization performance in many complex industrial process. When to calculate the output weights, Huang just adds a threshold to every element in the main diagonal of matrix H^TH ($\hat{\beta} = \left(H^TH + I/C\right)^{-1} H^TT$) [1], which links the ELM theory with other mathematical theory such as matrix theory and ridge regression. In our previous work, LDL^T decomposition to H^TH is employed to overcome the multicollinear problem (R-ELM) [8]. One just needs to set a proper threshold to some singular elements to matrix D after the decomposition. Here, we present a novel approach to get rid of the multicollinear problem based on PCA theory. Compared with other related algorithms, the proposed algorithm can make sure that the estimation of output weights is unbiased. Simultaneously, the proposed ELM algorithm can also improve the training and testing time in contract to the ordinary ELM and its variant based on LDL^T decomposition.

Iron and steel industry is one of the pillar industries in the national economy of China, which is also an industry with high energy consuming and high emission. The blast furnace (BF) is the first and also the most important step in the manufacture of iron and steel [9,11]. Because of the great amount of iron production, the tiny improvement of automation technology in the blast furnace production process will bring huge benefits. However, due to lack of industrial data and the unknown mechanism in BF, the model is a typical 'black box' system [10]. For a long time, the operators of BF just speculate the process situation relying on the experience. The thermal situation in BF is the most typical and important performance indicator to maintain the normal operation. In the industrial process, the operators often know the thermal situation inside the BF by monitoring the temperature of hot metal [12]. Nowadays, the common used method for measuring the temperature of hot metal is artificial contact measurement. The operators insert the disposable thermocouple into the molten iron and then the temperature data will be uploaded to the host computer [11]. This ancient measurement operation is very inconvenient in the aspects of discontinuous data acquisition and unsafe operation. On the other hand, this measurement is costly because of the employment of one-time thermocouple. In this paper, a soft measurement model to the temperature of hot metal is established based on 5 performance

* Corresponding author.
E-mail address: zhangsen@ustb.edu.cn (S. Zhang).

http://dx.doi.org/10.1016/j.neucom.2015.04.106
0925-2312/

indexes and the proposed P-ELM algorithm is employed to train the sampling data and build a reliable feedforward neural network model.

In this paper, a modified ELM algorithm named P-ELM is proposed based on PCA technique. Compared with the previous methods, the proposed P-ELM algorithm can deal with the multicollinear problem under the condition that the estimation of output weights is unbiased. The improved P-ELM algorithm shows a better performance in dealing with interfered data obtained from complex industrial process. In order to verify the effectiveness of the proposed P-ELM algorithm in real industrial application, a soft measurement model subject to hot metal temperature in BF is presented. Finally, simulation results using the real industrial data show that P-ELM algorithm has better stability and generalization performance compared with other famous feedforward neural network. It is worth mentioning that [13] also employed PCA theory to the ordinary ELM algorithm. However, in this paper, PCA technique is used to deal with the hidden layer output matrix rather than the original sampling data in [13].

The following sections are organized as follows: Section 2 is the main part to describe the improved P-ELM algorithm. Some basic theory about BF will be presented in Section 3 and simulation results are given in Section 4. Section 5 summarizes the conclusion of this paper.

2. The improved extreme learning machine algorithm

In the last few years, ELM algorithm has received very wide range of applications and development because of its fast train speed and good generalization performance. In this section, a review of the ordinary ELM algorithm is presented and some properties of the solution will be discussed. And then the improved P-ELM algorithm is proposed to overcome the restrictions.

2.1. The theory of ordinary ELM algorithm

Suppose there are N arbitrary samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ denotes an n-dimensional feature of the ith sample and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}] \in R^m$ denotes the target vector. The mathematical model of SLFNs with $\tilde{N}$ hidden nodes is as follows:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_k) = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i, b_i, x_k) = t_k, \quad k = 1, 2, \dots, N \tag{1}$$

where w_i and b_i are the learning parameters which will be determined randomly. β_i is the output weight matrix connecting the ith hidden and the output nodes. $g(w_i, b_i, x_k)$ is a nonlinear piecewise continuous function which satisfies ELM universal approximation capability theorems.

The above $\tilde{N}$ equations can be written in matrix form as

$$H\beta = T \tag{2}$$

where

$$H(W,B,X) = \begin{pmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{pmatrix}_{N \times \tilde{N}}$$

called the hidden layer output matrix. According to the theory of Least Square, the output weight β can be estimated as

$$\hat{\beta} = H^+ T \tag{3}$$

where H^+ is the Moore–Penrose generalized inverse of H [19]. There are several methods to calculate the Moore–Penrose generalized inverse. Here singular value decomposition (SVD) method is widely used, where $H^+ = (H^T H)^{-1} H^T$. So

$$\hat{\beta} = (H^T H)^{-1} H^T T \tag{4}$$

It is apparent from (4) that one can not obtain correct and satisfactory result if the matrix $H^T H$ is singular despite that some mathematical softwares like Matlab have the corresponding methods to deal with the inverse of singular matrix. Next some properties of the solution of ELM algorithm will be discussed. In a complex industrial production environment, the data will often be interfered by external noise. Then the model (2) should be modified as

$$H\beta + \xi = T \tag{5}$$

where white noise is considered here and $\xi \in N(0, \sigma^2)$.

After adding the interference of external noise, the solution of ELM will also be modified as

$$\hat{\beta} = \left(H^T H\right)^{-1} H^T T = \frac{\sum_{i=1}^{\tilde{N}} H_i^T (H_i \beta_i + \xi_i)}{\sum_{i=1}^{\tilde{N}} H_i^T H_i} = \beta + \frac{\sum_{i=1}^{\tilde{N}} H_i^T \xi_i}{\sum_{i=1}^{\tilde{N}} H_i^T H_i} \tag{6}$$

Next we analyze the result from three aspects: expectation (E), variance (V) and mean square error (MSE). Then one can get

- $E\left(\hat{\beta}\right) = \beta$
- $V\left(\hat{\beta}\right) = E\left(\hat{\beta}^2\right) - E\left(\hat{\beta}\right)^2 = \dfrac{\sigma^2}{\sum_{i=1}^{\tilde{N}} H_i^T H_i} = \sigma^2 \sum_{i=1}^{\tilde{N}} \dfrac{1}{\lambda_i}$
- $MSE\left(\hat{\beta}\right) = \dfrac{1}{\tilde{N}} E\left[\left(\hat{\beta} - \beta\right)^T \left(\hat{\beta} - \beta\right)\right] = \dfrac{1}{\tilde{N}} \cdot \dfrac{\sigma^2}{\sum_{i=1}^{\tilde{N}} H_i^T H_i} = \dfrac{\sigma^2}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \dfrac{1}{\lambda_i}$ (7)

where λ_i is the ith eigenvalue of $H^T H$ [8].

As we all know, the learning parameters of hidden nodes in the model of ELM algorithm are generated randomly without human tuning. In most case, the number of hidden nodes is far less than that of the samples ($\tilde{N} \ll N$). So H is usually not a square matrix and $H^T H$ may not always be nonsingular. That is to say when $H^T H$ is multicollinear, some eigenvalues will tend to zero, while $V\left(\hat{\beta}\right)$ and $MSE\left(\hat{\beta}\right)$ will become larger. And the solution is not convincing.

2.2. The introduction for improved P-ELM algorithm

This subsection is the main part of the note. Here the improved P-ELM algorithm is presented to overcome the multicollinear problem using principle component analysis (PCA) technique. PCA is a useful statistical technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in high dimension data [14,16]. For a series of data with high dimension, PCA is a powerful tool for identifying patterns and expressing the data in such a way as to highlight their similarities and differences by reducing the number of dimensions [15].

Fig. 1 presents the distribution of sample data with two dimensions where we can see that most of the data distribute along the direction of w_1. Here w_1 is called the first principle component direction which can character the main information of data distribution. In addition, w_2 stands for another direction of the second principle component which means less important information like external disturbance. The main purpose of the theory of PCA is to represent the sample data using the main principle components. In the theory of mathematical statistics, the first principle component direction is the one with maximum variance of the data distribution [16]. So the sort of principle components is based on the size of variances of sample data in different directions. Next a brief theoretical derivation of PCA is presented [17].