A New Incremental Learning Modeling Method Based on Multiple Models for Temperature Prediction of Molten Steel in LF

Huixin TIAN,1) Zhizhong MAO1,2) and Anna WANG1)

1) School of Information Science & Engineering, P.O. Box 131, Northeastern University, Shenyang 110004, P. R. China. E-mail: icedewl@163.com 2) Key Laboratory of Integrated Automation of Process Industry, Northeastern University, Ministry of Education, Shenyang 110004, P. R. China.

(Received on June 30, 2008; accepted on October 14, 2008)

1. Introduction

With the fast development of secondary refining, ladle furnace (LF) has become one of the most important roles during secondary metallurgical process. Moreover, the development and improvement of LF refining technology could promote the advancement of secondary refining technology and the development of iron and steel industry. The one of main purposes of LF refining is to ensure that the temperature of molten steel can meet the continuous casting. Therefore, the temperature control of LF is essential for improving both productivity and product quality. The temperature predicting is a key factor for the temperature control successfully during LF refining process.

The studies for predicting the temperature of molten steel in LF can be mainly divided into two methods: mechanistic method and intelligent method. The mechanistic method is mainly based on thermodynamics and conservation of energy. Several mechanistic models were developed in earlier papers. For example, a thermodynamic analysis that based on the second law of thermodynamics was presented for molten steel by Camdali et al.1) Similarly, the physical energy during the steel production process such as loss energy, exergies going into and out of the LF were evaluated and the ways of minimizing these losses in practice are discussed by Camdali and Tunc.2,3) These models were extremely useful thermal analysis tool and were very accurate. However, these models cannot be used for on-line prediction as the parameters were hard to obtain. It is attributed to the harsh operating environment of ladle metallurgy especially the high temperatures and corrosive slag associated with the process.

With development and wide application of artificial intelligence in industry, many intelligent models based on the data have been proposed to overcome above limitations of mechanistic models. In these methods, the artificial intelligent algorithms are used to establish prediction models with production data. The intelligent methods obtain the information from production data, so there are not the limitations about parameters of mechanistic models during the production process of high temperatures. For example, Yan guang et al. used artificial intelligent technology to predict the temperature of LF.4) A genetic algorithm (GA) combined with BP network algorithm is presented for predicting the temperature of molten steel in Ref. 5).

However, in practical applications, acquisition of information in new production data is expensive and time consuming. Consequently, it is necessary to update the model in an incremental fashion to accommodate new data without compromising the performance on old data. That is to say, the model should learn new information without forgetting previously acquired knowledge. Practically, in some traditional intelligent prediction models of molten steel, the updates are neglected. Some of models use traditional up-
dating methods that retrain the models using all production data (involve old data and new data) or use the updated data (the older data is replaced by new data). They are not good enough because some good performance may have to be lost for learning new information. The retraining method might consume time for more and more data being used, so it does not adapt to the online real-time prediction. The using updated data method could not save the information of old data, so this information will disappear as time goes.

Against this background, we proposed a new incremental learning modeling method with multiple intelligent models to establish the prediction temperature model of molten steel in LF. The new prediction model can learn additional information from new data, and preserve previously acquired knowledge not requiring the original data that used to train the existing old model.

The rest of this paper is organized as follows. In Sec. 2, an intelligent prediction model based on ELM is established for predicting the temperature of molten steel by analyzing conversation of energy during refining process in LF. In Sec. 3, the new incremental learning modeling method based on multiple models is presented. In Sec. 4, the new prediction model based on the new incremental learning method is experimented by 300 t LF of Baoshan Iron & Steel Co. Ltd. Finally, the last section concludes this paper.

2. Intelligent Prediction Model Based on ELM

In this section an intelligent model based on ELM is presented by analyzing the production process of LF refining against the background of 300 t LF of Baoshan Iron & Steel Co. Ltd. The analysis is done based on conservation of energy. The model can be used to implement the on-line prediction of temperature if the alloy melting time is ignored.

2.1. Extreme Learning Machine

A novel learning algorithm for single-hidden-layer feedforward neural networks called extreme learning machine (ELM) was proposed by Huang et al. recently.\(^6\) In ELM, the input weights (linking the input layer to the hidden layer) and hidden biases are randomly generated, and the output weights (linking the hidden layer to the output layer) are analytically determined by using Moore–Penrose (MP) generalized inverse. Unlike traditional neural networks, the ELM algorithm needs to calibrate the parameters (learning rate, learning epochs, etc.) on more. The learning speed of ELM is extremely fast compared to other traditional methods. The traditional classic gradient-based learning algorithms may face several issues like local minima, improper learning rate and overfitting, etc. In order to avoid these issues, some methods such as weight decay and early stopping methods may need to be used often in these classical learning algorithms. The ELM tends to obtain the solutions straightforward without such trivial issues. The ELM learning algorithm looks much simpler than most learning algorithms for feedforward neural networks. Against above characteristics, ELM is used to establish generic intelligent model in this paper.

2.1.1. Single Hidden Layer Feedforward Networks (SLFNs)

For \(N\) arbitrary distinct samples \(\{(x_j, t_j)\}_{j=1}^{N}\), where \(x_j=[x_{j1}, x_{j2}, \ldots, x_{jm}]^T \in \mathbb{R}^m\) and \(t_j=[t_{j1}, t_{j2}, \ldots, t_{jm}]^T \in \mathbb{R}^m\), a standard SLFNs with \(N\) hidden nodes and activation function \(g(x)\) are mathematically modeled as

\[
\sum_{i=1}^{N} \beta_i g(x_i) = \sum_{i=1}^{N} \beta_i (w_i \cdot x_j + b_i) = o_j, \quad j = 1, \ldots, N
\]

(1)

where \(w_i=[w_{i1}, w_{i2}, \ldots, w_{im}]^T\) is the weight vector connecting the \(i\)th hidden node and the input nodes, \(\beta_i=[\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T\) is the weight vector connecting the \(i\)th hidden node and the output nodes, \(o_i=[o_{i1}, o_{i2}, \ldots, o_{im}]^T\) is the output vector of the SLFN, and \(b_i\) is the threshold of the \(i\)th hidden node. \(w_i \cdot x_j\) denotes the inner product of \(w_i\) and \(x_j\). The output nodes are chosen linear. The standard SLFNs with \(N\) hidden nodes and activation function \(g(x)\) can approximate these \(N\) samples with zero error means that \(\sum_{j=1}^{N} \| o_i - t_j \| = 0\). And these \(N\) equations can be written compactly as:

\[
H \beta = T
\]

(2)

where

\[
H = \begin{bmatrix}
g(w_1 \cdot x_1 + b_1) & \cdots & g(w_N \cdot x_1 + b_N) \\
\vdots & \ddots & \vdots \\
g(w_1 \cdot x_N + b_1) & \cdots & g(w_N \cdot x_N + b_N)
\end{bmatrix}_{N \times N}
\]

(3)

and

\[
\beta = \begin{bmatrix}
\beta_1^T \\
\vdots \\
\beta_N^T
\end{bmatrix}_{N \times m}
\]

(4)

\[
T = \begin{bmatrix}
t_1^T \\
\vdots \\
t_N^T
\end{bmatrix}_{N \times m}
\]

(5)

Where \(H\) is called the hidden layer output matrix.

2.1.2. ELM Algorithm

In most case, the number of hidden nodes is much less than the number of training samples \((N \ll N)\). It means that \(H\) is a nonsquare matrix and there may not exist \(H^{-1}\). In ELM, \(H\beta = T\) with zero error means that \(\sum_{j=1}^{N} \| o_i - t_j \| = 0\). And these \(N\) equations can be written compactly as:

\[
\hat{\beta} = H^\dagger T
\]

(6)

Where \(H^\dagger\) is the Moore–Penrose generalized inverse of matrix \(H\).

Thus, a simple learning method for SLFNs called extreme learning machine (ELM) can be summarized as follows:

Step 1: Randomly assign input weight \(w_i\) and bias \(b_i\) \(i=1, \ldots, N\).

Step 2: Calculate the hidden layer output matrix \(H\).
rigorously proved in an incremental method by Huang et al.

nace (for various metal alloy and slag addition in 300 t ladle furnace) can be calculated by the parameters of temperature change sum of heat exchanges and chemical reaction heats. So they are the energy change for additions, which includes the energy loss may be reflected by refining time. The second section is the energy loss of the ladle refractory wall and energy loss between ladle furnace and surroundings, which include the following three sections: the first section is the heat exchanges from the top surface. The energy loss in this section is relatively stable. It will increase with time. Therefore this energy loss may be reflected by refining time. The second section is the energy change for additions, which includes the sum of heat exchanges and chemical reaction heats. So they can be calculated by the parameters of temperature change for various metal alloy and slag addition in 300 t ladle furnace (Table 1). The third section is the energy loss by argon purging. To sum up, the main factors that affect the temperature can be obtained, which are the refining power consumption, the initial temperature, the ladle states, the heat effects of additions, the volume of argon purging, the weight of molten steel and the refining time.

2.3. Prediction Model of Temperature

On the basis of above analysis, the inputs of basic ELM prediction model can be determined. They are the 6 factors that mainly affect the temperature of molten steel in LF: the refining power consumption, the initial temperature, the ladle states, the heat effects of additions, the volume of argon purging, the weight of molten steel and the refining time. The output of basic ELM prediction model is the temperature of molten steel. There are 9 hidden nodes assigned for ELM model according to the empirical formulation.

3. Incremental Learning Modeling Method with Multiple Models

Aiming at the shortages of present method of modeling and updating, the idea of incremental learning is considered to develop a new modeling method with multiple models in this section for establishing the prediction temperature model of molten steel in LF. Firstly, the training data is divided into several training subsets. Secondly, several sub intelligent models are trained using the training subsets. Thirdly, according to the training error of every model, the weight of every model can be obtained. Finally, the prediction model is established by aggregating the sub intelligent models based on the weights. When the new production data accumulate enough, they are used to train a new sub intelligent model. Similarly, the weight of new sub model can be computed. Then the new sub model and the old sub models are aggregated based on the weights, and the prediction model is updated. The above new modeling method with incremental learning is described as following steps:

1) Input
- Training data $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where output $y \in R$.
- Choose the parameters of $L$ and $K$, where $K=\text{floor}(N/L)$, $(L$ is the size of subsets, and $K$ is the number of subsets.)
- Threshold $\varphi$ ($0 < \varphi < 1$).
- Initialize $t=1$.

2) Iterate while $t \leq k$.
- Train sub intelligent model $t$, and get a hypothesis $f_t: X \rightarrow Y$.
• Calculate absolute relative error for each training example as

\[ E_t(i) = \left| \frac{f_t(i) - y_i}{y_i} \right| \] ..........................(7)

• Calculate the error rate of \( f(x) \) as

\[ \epsilon_t = \frac{l_t}{L} \] ..........................(8)

where \( l_t \) is the number of training example which \( E_t(i) > \varphi \).

• Calculate the weight \( \beta_t \) as

\[ \beta_t = \frac{\epsilon_t}{1 - \epsilon_t} \] ..........................(9)

• Set \( t = t + 1 \).

3) Output
• Output the final hypothesis:

\[ F_{\text{final}}(x) = \frac{\sum_t \left[ \left( \frac{1}{\beta_t} \right) f_t(x) \right]}{\sum_t \left( \frac{1}{\beta_t} \right)} \] ..........................(10)

In the beginning the training data \( S = [\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}] \) is supplied. The parameters \( L, K \) and \( \varphi (0 < \varphi < 1) \) is choose, where \( L \) is the the size of subsets, \( K \) is the number of subsets, and \( \varphi \) is threshold. Then the training data \( S \) is divided into \( K \) training subsets \( S_i (i = 1, 2, \ldots, K) \) according to the parameters \( L \) and \( K \).

At each iteration \( t = 1, 2, \ldots, K \), the ELM is called to generate the hypothesis \( f_t : X \rightarrow Y \) with the training subset \( S_t \). The absolute relative error for each training example is calculated as Eq. (7). Then the error rate of \( f_t \) can obtained as Eq. (8) according to the threshold \( \varphi \). The normalized error \( \beta_t (0 < \beta_t < 1) \) is computed as Eq. (9) and will be used as the weight of \( f_t \) in the following combination section.

All hypotheses generated in the previous \( t \) iterations are then combined using weighted majority voting. The voting weights are computed as the logarithms of the reciprocals of normalized errors \( \beta_t \). Therefore, those hypotheses that perform well on their own training are given larger voting powers. The final composite hypothesis \( F_{\text{final}}(x) \) is then made based on the combined outputs of individual hypotheses as Eq. (10).

During practical production process, the prediction model can be established using above method by Eq. (10). Furthermore, with the continuous producing, the number of new production data will continuously increase. When new data accumulate enough (the number of new data is \( L \)), the old model will be updated by setting \( t = t + 1 \) and recomputing \( F_{\text{final}}(x) \). \textbf{Figure 3} is the scheme of the new modeling and updating method. The new modeling and updating method can save the information of old models that have been trained, and need not save the original data. That is to say, only a few new production data being saved is enough. It is efficient to save a mass of space. In addition, the new method also may save a lot of time as compared with traditional updating method. It is attributed to the conservation of old hypothesis of sub intelligent models.

4. Experiments

Five hundred and fifty data of production from 300t LF in Baoshan Iron & Steel Co. Ltd. from May to December in 2006 are used to test the new temperature predicted model of molten steel in LF based on new incremental learning modeling method. 50 data are randomly selected from this data set as testing data, and others are used as training and updating data. The experiments are performed using different parameters for analyzing the their effect on performance of new modeling method. Firstly, the experiments with different value of \( K \) are carried out for \( \varphi = 0.5 \). \textbf{Figure 4} shows the performance of new method for different value of \( K \). It can be observed that RMSE (root-mean-squared error) is decreasing significantly when \( K = 2 \) and 3, and increasing when \( K \) is bigger than 5. It is obviously that \( K = 4 \) is the best choice for above production data of 300t LF. Secondly, the experiments with different value of \( \varphi \) are done for \( K = 4 \). \textbf{Figure 5} shows the performance of new method for different value of \( \varphi \). It is easy to drawn a conclusion that the performance is sensitive to the value of \( \varphi \). With the increasing of value of \( \varphi \) from 0.05, the RMSE is decreasing and becomes stable in the range from 0.2 to 0.55. When the value of \( \varphi \) increases to the value around 0.7, the performance becomes unstable. Therefore, the best interval for choosing the value of \( \varphi \) should be 0.2–0.55. It is also obviously from the Fig. 5 that the minimum error corresponds to \( \varphi \) of 0.5, so 0.5 is choose as the value of \( \varphi \) in the following experiments.

Fig. 3. Scheme of modeling and updating method based on incremental learning.

Fig. 4. Performance of new method for different value of \( K \).
Then the experiments that predict the temperature are carried out according to the above analysis. The whole data set are divided into 6 sub datasets according to the refining time. The first 4 sub datasets are used as training datasets for establishing the prediction model. Every sub datasets includes 100 data. The 5th sub dataset also includes 100 data.

<table>
<thead>
<tr>
<th>Training</th>
<th>Updating</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3, 4</td>
<td>5, 6</td>
<td>1, 2</td>
</tr>
<tr>
<td>2, 3, 4, 5</td>
<td>6, 1</td>
<td>2, 3</td>
</tr>
<tr>
<td>3, 4, 5, 6</td>
<td>1, 2</td>
<td>3, 4</td>
</tr>
<tr>
<td>4, 5, 6, 1</td>
<td>2, 3</td>
<td>4, 5</td>
</tr>
<tr>
<td>5, 6, 1, 2</td>
<td>3, 4</td>
<td>5, 6</td>
</tr>
<tr>
<td>6, 1, 2, 3</td>
<td>4, 5</td>
<td>6, 7</td>
</tr>
</tbody>
</table>

Table 2. Orders of dataset for six-fold cross validation.

**Fig. 5.** Performance of new method for different value of $\varphi$.

**Fig. 6.** Comparison of real measured temperature and predicted temperature by different models.
data as updating dataset for updating the model. The last sub dataset includes 50 data as test dataset. That is to say, \( L=100, K=4, \) and \( \varphi=0.5. \) In order to do the experiment more accurately, a six-fold cross validation is done. In next experiment, the first 50 data are used as test dataset, and following data are used as training dataset and updating dataset. The sub dataset will change as this method in following experiments. That is, every data may be used as updating data one time, and the following 50 data are used as test dataset. Table 2 shows the orders of dataset for six-fold cross validation during the experiments. Here the number of sub dataset is defined according to the refining time. Figure 6 shows the comparison of real measured temperature, predicted temperature by the model before updating and predicted temperature by the model after updating. Obviously, the results of the model being updating are better than the one without updating. The comparison of performance of the model before updating and the model after updating can be obtained by calculating the prediction accuracy and the RMSE as Table 3 shown. The accuracy is calculated as follows:

\[
\text{accuracy} = \frac{N_a}{N_{w6}} \tag{11}
\]

Where, \( N_a \) is the number of heats with error \(<5^\circ C\), \( N_{w6} \) is the number of whole testing heats. According to the comparison, we can draw a conclusion that the prediction model based on incremental learning method of modeling and updating has the good ability of updating. The above mentioned problems of update in temperature predicting model of molten steel in LF are solved successfully by this new method.

For comparing the new method to other prediction methods, the BPNN (Back-Propagation Neural Network) and ELM are also carried out to predict the temperature as traditional method with the same data set. The structures of BPNN and ELM are also same as above proposed structure in Sec. 2.3. The experiments are also done by six-fold cross validation. The comparison of these method is described in Table 3. The above results of experiments shows that the new proposed method not only has the good ability of updating, but also has better performance and prediction accuracy compared with traditional methods. The results of predicted temperature by prediction model based on the new method are satisfied for production process in LF.

### 5. Conclusion

A new prediction model based on new incremental learning modeling and updating method is established for predicting the temperature of molten steel in LF in this paper. Firstly, the intelligent model is established based on the analysis of energy conversation during whole metallurgic process of LF using ELM algorithm. It is used as generic sub model in the following new modeling and updating method. Secondly, the incremental learning modeling method with multiple models is presented to aggregate the generic sub models. Then the new prediction model based on multiple models method is established. Finally, the new prediction model is used in 300 t LF of Baoshan Iron & Steel Co. Ltd. The experiments demonstrate that both updating ability and accuracy of new model are satisfied for production. As for the further research, some similar algorithm like RBF and SVM would be used based on the new proposed idea of incremental learning modeling and updating, and the new method might be used in more extensive field.

### Acknowledgement

This research is supported by National Natural Science Foundation of China (Grant No. 60674063), Key Laboratory of Process Industry Automation of Northeastern University of Ministry of Education and Liaoning Province.

### REFERENCES