ISIJ International, Vol. 52 (2012), No. 9, pp. 1585–1590

Prediction Model of End-point Manganese Content for BOF Steelmaking Process

Zhou WANG,[1,2] Jian CHANG,[3] Qi-Ping JU,[3] Fei-Ming XIE,[3] Bin WANG,[1,2] Hong-Wei LI,[3] Bao WANG,[1,2] Xin-Chun LU,[2,3] Guo-Qing FU[1,2] and Qing LIU[1,2]*

1) State Key Laboratory of Advanced Metallurgy, University of Science and Technology Beijing, No. 30 Xueyuan Road, Beijing, 100083 P. R. China. 2) School of Metallurgical and Ecological Engineering, University of Science and Technology Beijing, No. 30 Xueyuan Road, Beijing, 100083 P. R. China.
3) Fangda Special Steel Technology Co., Ltd., Nanchang, Jiangxi Province, 330012 P. R. China.

(*Received on January 4, 2012; accepted on April 9, 2012*)

Through analyzing the factors that influence end-point manganese content during BOF steelmaking process, multiple linear regression model for prediction of end-point manganese content was obtained on the basis of actual production data. Given the advantages of artificial neural network, it was used to predict end-point manganese content during BOF steelmaking process, and BP neural network model was established. By means of combining the characteristics of genetic algorithm and BP neural network completely, a combined GA-BP neural network model was established. The verification and comparison of the above three models show that the combined GA-BP neural network model has the highest prediction accuracy. The hit rate of the combined GA-BP neural network model is 90% and 84% respectively when predictive errors of the model are within ±0.03% and ±0.025%. Compared with two models aboved, the combined GA-BP neural network model could provide the most accurate prediction of end-point manganese content, and thus represents a good reference for real production.

KEY WORDS: BOF; end-point manganese content; multiple regression; BP neural network; genetic algorithm; prediction model.

1. Introduction

Currently, BOF steelmaking process plays an important role in conventional BF-BOF route which is the predominant steelmaking process around the world. The oxygen converter utilizes oxygen to react with the elements of steel to obtain qualified molten steel, which contains necessary chemical composition and bath temperature. Therefore, it is of great importance to maintain the stable blowing process, and control the end-point chemical composition and temperature of liquid steel accurately.[1] Until now, numerous research studies which involve prediction of BOF end-point carbon and temperature have been conducted in China,[2–6] however, there is relatively little research which is concerned with the prediction of end-point manganese content. Takawa *et al.*[7] analysed the material balance at the end of the blow and established the prediction model of BOF end-point manganese content based on the measurement of sublance. Due to the utilization of sublance, the cost of this method was high and the method was only appropriate for the converters that equipped with sublance. Yang *et al.*[8] established the adaptive fuzzy-neural net based prediction model of BOF end-point manganese content, and the model performed well on the prediction of end-point manganese. Liu *et al.*[9] analysed the factors that affect BOF end-point manganese and developed the prediction model of end-point manganese content based on measured values of bath carbon content and temperature at the end of blow.

During BOF steelmaking process, it would further improve the hit rate of end-point chemical composition of the molten steel if the end-point manganese content is predicted accurately. Furthermore, it would reduce the production cost and improve the quality of liquid steel. Thus, the authors carried out some research on the prediction model of BOF end-point manganese content, and established the algebra model, BP neural network model and combined GA-BP (Genetic Algorithm-Back Propagation) neural network model respectively based on the production data of Fangda special steel plants.

2. The Algebra Model

The algebra model is the model that could be expressed in the form of algebra. Generally speaking, the algebra model is obtained by using mathematical-statistics methods which involve the collection and analysis of production data.[10] As a kind of conventional model, the multiple linear regression model is also established on the basis of mathematical-statistics methods and it has a good applicability. Therefore, the multiple linear regression model was adopted to predict BOF end-point manganese content in this paper.

* Corresponding author: E-mail: qliu@metall.ustb.edu.cn
DOI: http://dx.doi.org/10.2355/isijinternational.52.1585

2.1. Establishment of the Multiple Linear Regression Model

The multiple linear regression model is based on the utilization of a large amount of production data; therefore, data from nearly 1 000 heats of the same campaign were collected from Fangda special steel plants. Before incorporating the production data into the model, the data were filtered and treated. The principles of filtration and treatment include removal of the variables which do not affect the model and omission of abnormal values of the variables so that the production data meets the actual requirements. Ultimately, the data from 529 qualified heats were chosen. For establishing and testing the model, the qualified data were divided into two parts: data from 429 heats were used to establish the model and data from the other 100 heats were applied to test the model.

The selection of independent variables plays a key role in establishing the model. The reactions that occur in the molten steel bath of a converter are very complex, and end-point manganese content is affected by numerous interacting factors. Therefore, in order to provide an adequate description of the entire melting process and clarify the model, the multiple linear regression model employs those factors which change dramatically and play a key role in the BOF steelmaking process as the basic variables. SPSS is one of the most popular statistical packages available that can be used to perform data entry and analysis and to create tables and graphs. Moreover, SPSS is so powerful that it is capable of handling large amounts of data, and can perform most of the analyses covered in the test. Based on the above consideration, SPSS was applied to complete the correlation analysis between end-point manganese content and other factors in this paper, and the following variables which have an obvious correlation with end-point manganese content were chosen and presented in **Table 1**. They were shown as follows: charged hot metal, charged scrap, hot metal temperature, manganese content in hot metal, phosphorus content in hot metal, sulphur content in hot metal, oxygen consumption, lime addition, dolomite addition and fluorite addition respectively.

SPSS was applied to complete the regression analysis with 429 heats data, and the regression model is shown in Eq. (1):

$$\begin{aligned}[\%Mn] = {} & -0.499 + 0.004 \cdot X_1 - 0.004 \cdot X_2 + 1.4896 \times 10^{-4} \cdot X_3 \\ & + 0.004 \cdot X_4 - 1.9883 \times 10^{-4} \cdot X_5 + 3.5804 \\ & \times 10^{-4} \cdot X_6 + 4.109 \times 10^{-5} \cdot X_7 - 0.001 \cdot X_8 \\ & - 0.019 \cdot X_9 - 0.110 \cdot X_{10} \end{aligned} \quad \text{..... (1)}$$

In the evaluation of properties of the model, not only the hit rate of the model was considered, but also the goodness of fit was proposaled in the paper.[11] The mathematical expression of the goodness of fit is shown in Eq. (2):

$$R^2 = 1 - \frac{\sum (y_i - y_i')^2}{\sum (y_i - y_i'')^2} \quad \text{..... (2)}$$

Where $y_i (i = 1, 2, 3 \cdots, n)$ is the target value of the model; $y_i' (i = 1, 2, 3 \cdots, n)$ is the output value of the model; $y_i'' (i = 1, 2, 3 \cdots, n)$ is the average value of y_i; and n is the sample size.

Goodness of fit (R^2) reflects the correlation degree between input variables and output ones. Specifically speaking, the more close to 1 the goodness of fit is, the more obvious the relation between input variables and output ones is, thus, the performance of the model is better. The goodness of fit (R^2) usually scales from 0 to 1.

2.2. Verification of the Multiple Linear Regression Model

The goodness of fit (R^2) of the multiple linear regression model was 0.555 when it was calculated with Eq. (2). In order to clarify the predictive accuracy of the model, test data (100 heats) were utilized to predict the end-point manganese content. The comparison of predicted values and actual values of end-point manganese content is shown in **Figs. 1** and **2** shows the distribution of predictive errors of the multiple linear regression model.

As illustrated in Fig. 1, there is a moderate discrepancy between predicted values and actual values. Figure 2 indicates that the predictive errors are mainly distributed within ±0.06%. The hit rate of the model is 74% when the predictive errors are within ±0.05%; however, there are only 44 heats when the predictive errors are within ±0.03%, which just account for 44% of the test heats. Based on above analysis and comparison, a conclusion can be made that the hit rate of the multiple linear regression model is unsatisfied and the predictive errors of some heats are relatively large. The reasons are as follows: BOF steelmaking is a very complex physico-chemistry process which involves multicomponent, multiphase and high temperature; meanwhile, end-point manganese content is affected by numerous interacting factors, but the multiple linear regression model simplified the relation between end-point manganese content and other interacting factors to be linear. Therefore, this model could only provide a limited reference for real production, and it is necessary to explore other methods to predict end-point manganese content with higher precision.

Table 1. Independent variables for the regression model.

Names of independent variables	Symbol	Units	Names of independent variables	Symbol	Units
Charged hot metal	X_1	t	Charged scrap	X_2	t
Hot metal temperature	X_3	°C	manganese content in hot metal	X_4	1%
Phosphorus content in hot metal	X_5	1%	Sulphur content in hot metal	X_6	1%
Oxygen consumption	X_7	Nm^3	Lime addition	X_8	t
Dolomite addition	X_9	t	Fluorite addition	X_{10}	t

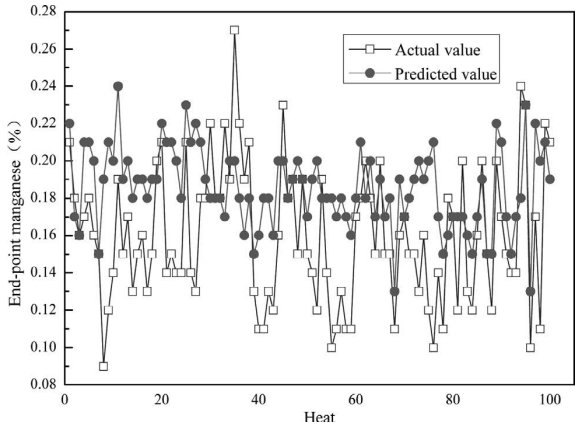

Fig. 1. Comparison of end-point manganese between predicted values and actual values of the regression model.

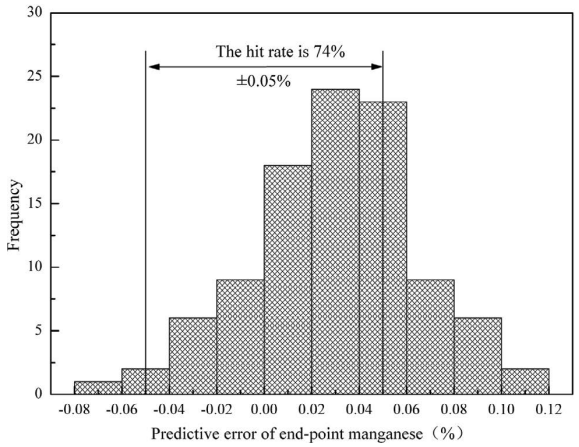

Fig. 2. Distribution of predictive errors of the regression model.

3. BP Neural Network Model

With the development of artificial intelligence, some control methods based on neural network or neural network combined with algorithms have been widely used in BOF end-point control.[12] Artificial neural networks (ANN) are the abstraction, simplification and simulation of biological neural networks, and they have the capability of self-study, self-adaptation, self-organizing and so on.[13] As a result, the static model which is established on the basis of artificial neural network could provide an effective reference for BOF steelmaking process, and it may have wide application in the field of metallurgy.[14] The utilization of artificial neural network could overcome the weakness of the regression model when dealing with nonlinear problems, hence it could greatly improve the prediction accuracy of the model. So the authors tried to established BP neural network model to predict end-point manganese content.

3.1. Establishment of BP Neural Network Model

The BP neural network usually refers to the multilayer feedforward neural network based on error back propagation algorithm, which was put forward by D. E. RUMELHART and J. L. MCCELLAND in 1986,[15] and the BP neural network has become the most widely used neural network at present. The learning process of BP neural network consists of signal forward propagation and error back propagation, and the continual loop of the two processes constitutes the training process of BP neural network.[16] The structure of BP neural network is shown in **Fig. 3**, which usually contains an input layer, an output layer and one or more hidden layers.

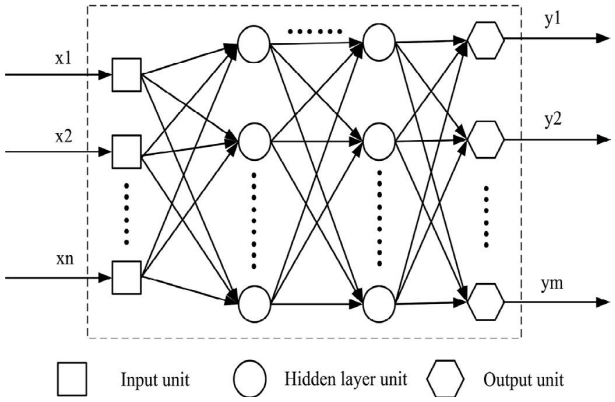

Fig. 3. Structure of BP neural network.

The selection of control variables has a tremendous influence on the establishment of BP neural network model. Thus, the authors analyzed the relation between end-point manganese content and other variables, and completed the correlation analysis between them with SPSS. Ultimately, 10 control variables were chosen and they were shown in Table 1. Consequently, the input variables of the BP neural network model are above 10 control variables, output variable of the BP neural network model is end-point manganese content and the BP neural network model employs three layers (one input layer, one output layer and one hidden layer).

As the values of control variables of BP neural network model may differ in several orders of magnitude, and the difference would greatly affect convergence speed and accuracy of BP neural network. Therefore, in order to improve the convergence speed and accuracy of BP neural network model, the input data and output data of the model were normalized. After the normalization with Eq. (3), the input data and output data all scaled from −1 to +1. Equation (3) is shown as follows:

$$x_{ij}' = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \quad \text{..... (3)}$$

Where x_{ij} is the value before normalization; x_{ij}' is the value after normalization; i is the sample point; j refers to the control variable; min (x_{ij}) and max (x_{ij}) are the minimum and maximum values of x_{ij} respectively.

The training process of BP neural network was proceeded with 429 heat data after confirmation of the fundamental structure of BP neural network. Then, the test data (100 heats) were used to verify the trained BP neural network. After the continual comparison of predicted values and actual values, the BP neural network with the highest prediction accuracy was saved. The BP neural network model achieved the highest prediction accuracy when it adopted the fundamental parameters presented in **Table 2**.

3.2. Verification of BP Neural Network Model

Similar to the multiple linear regression model, the BP neural network model was used to predict end-point manga-

nese content with the test data (100 heats). The comparison of predicted values and actual values of end-point manganese content is illustrated in **Fig. 4**.

As can be seen from Fig. 4, the predictive errors of this model are mainly within ±0.05% except a few heats. In other words, the predicted end-point manganese content is close to the actual end-point manganese content. **Figure 5** shows the distribution of predictive errors of BP neural network model. There are 80 heats when the predictive errors of the model are within ±0.03%, which account for 80% of the test heats; and the hit rate of the BP neural network model is 71% when the predictive errors are within ±0.025%. Compared with the multiple linear regression model, the BP neural network model obviously improves the prediction accuracy. However, the hit rate of the model still needs to be improved when the predictive errors are within ±0.025%. Thus, in order to meet the requirements of lean production in BOF steelmaking process, it is necessary to modify the BP neural network model to further improve the prediction accuracy. So the authors improved the BP neural network model with genetic algorithm, and the combined GA-BP (Genetic Algorithm-Back Propagation) neural network model was established ultimately.

Table 2. Fundamental parameters of BP neural network.

Names of parameters	Values	Names of parameters	Values
Nodes of input layer	10	Nodes of output layer	1
Nodes of hidden layer	6	Epochs	4 500
Learning rate	0.05	Momentum factor	0.9
Training function	trainlm	Number of hidden layer	1

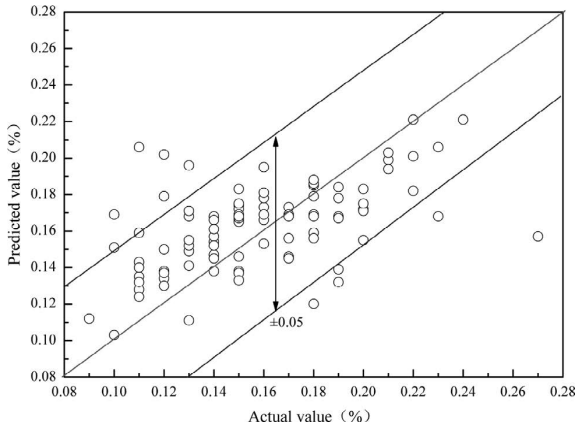

Fig. 4. Comparison of end-point manganese content between predicted values and actual values of BP neural network.

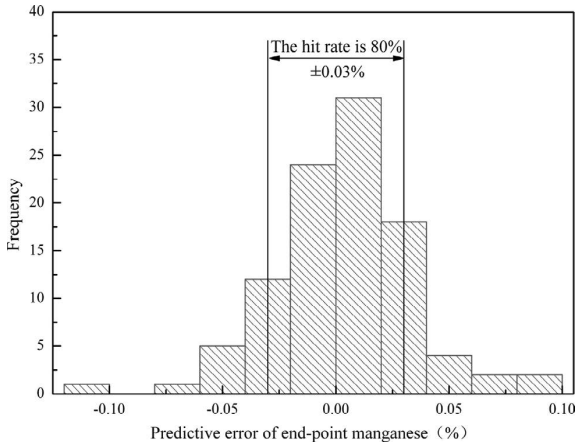

Fig. 5. Distribution of predictive errors of BP neural network.

4. Combined GA-BP (Genetic Algorithm-Back Propagation) Neural Network Model

Although the BP neural network has a powerful ability of studying, and it could approach the nonlinear functions with any precision, it could easily get stuck in a local minima.[17,18] Genetic algorithm is good at global optimization. Hence, based on the global optimization of genetic algorithm and powerful learning ability of BP neural network, BP neural network and genetic algorithm were amalgamated to establish a combined GA-BP neural network model to improve the prediction accuracy of end-point manganese content in this research. As a part of combined GA-BP neural network model, genetic algorithm was used to speed the learning process and optimize the parameters of BP neural network model.

4.1. A Brief Introduction to Genetic Algorithm

The GA (genetic algorithm), which is a non-numerical computing method based on the biological principle of natural selection and population genetics, is able to make an individual (each candidate solution) move to the optimal solution by means of a 'the survival of the fittest' mechanism. GA is an iterative search procedure, which has been successfully used for a variety of combinatorial optimization problems.[19–21] The main operations of GA involve changing the solutions from iteration to iteration by applying a crossover operator, which combines two parent chromosomes to obtain a new offspring chromosome, and a mutation operator, which modifies a single chromosome.

Compared with other optimization algorithms, the main characteristics of GA are as follows:[22]

(1) GA is easier to reach the global optimum because of population-based search strategy.

(2) GA has a wider application, and it is especially fit for the problems that are complex and nonlinear.

(3) The evolution process of GA is not blind, but heuristic.

(4) GA is general-purpose and robust.

Based on above comparison, GA was applied to optimize the initialized weights and thresholds of BP neural network in this paper, and **Fig. 6** illustrates the solution procedure of the GA. Just as Fig. 6 shows, the solution procedure of GA can be summarized as follows:

Step 1: Initialize a population, including its scale, selection probability, crossover probability and mutation probability. Generate the individual randomly. And each individual of population represents a group of initial weights of the network.

Step 2: Calculate the evaluation function corresponding to every group of weights, and determine the fitness of each individual.

Step 3: Judge whether the population reaches the optimization criterion.

Step 4: If yes, output the optimal solution; if no, generate

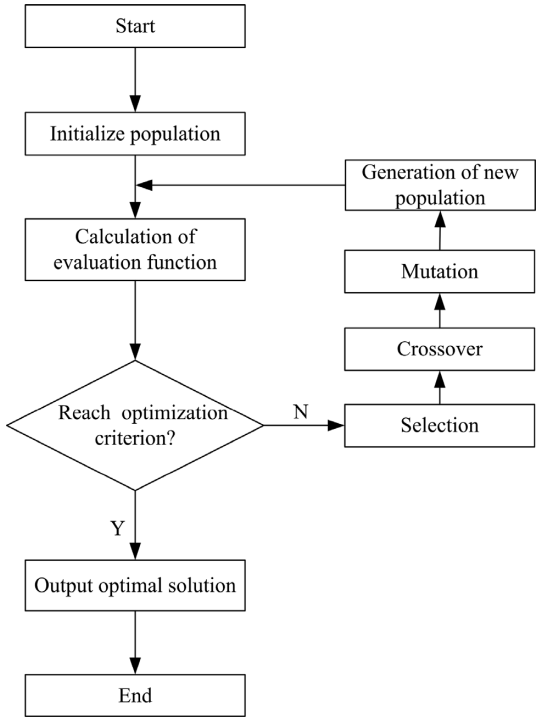

Fig. 6. Solution procedure of GA.

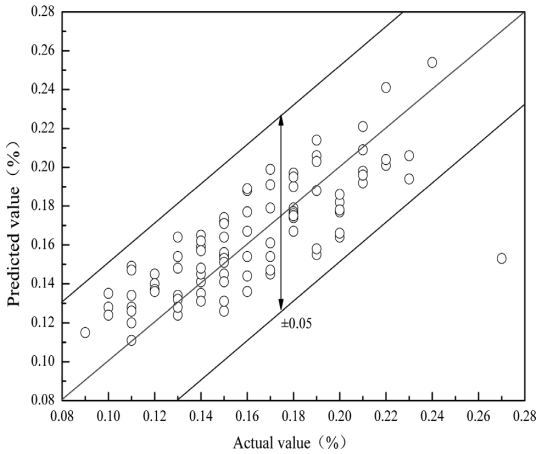

Fig. 7. Comparison of end-point manganese content between predicted values and actual values of GA-BP neural network.

the new population after the operation of selection, crossover and mutation. Then return to step 2, circulate and repeat the above procedure until fitness of the optimal individual reaches the demand criterion. Decode the optimal individual, and the optimum weights and thresholds of BP neural network are obtained.

4.2. Verification of Combined GA-BP Neural Network Model

Similar to the BP neural network model, 429 heats data were used to complete the training process of combined GA-BP neural network model, and the remaining 100 heats data (test data) were used to verify the model. The comparison of predicted values and actual values of end-point manganese content is illustrated in **Fig. 7**. It can be seen from Fig. 7 that the predicted values of end-point manganese content agree well with the actual values of end-point manganese content. Meanwhile, **Fig. 8** shows the distribution of

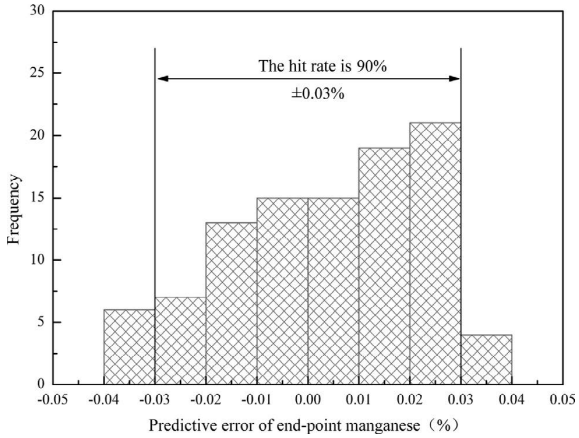

Fig. 8. Distribution of predictive errors of GA-BP network.

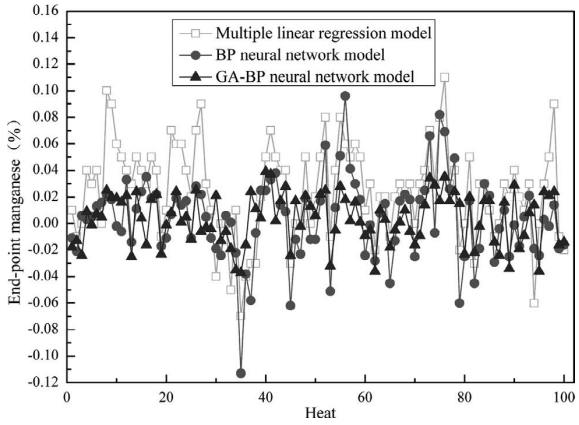

Fig. 9. Comparison of predictive errors of three models.

predictive errors of combined GA-BP neural network model. The predictive errors of this model are mainly distributed within ±0.03%, and the data from 90 heats (the total heats are 100) fall within this range, which represents 90% of the test data. Especially, the hit rate of combined GA-BP neural network model is 84% when the predictive errors of the model are within ±0.025%. Based on above analysis, it can be concluded that the combined GA-BP neural network model has high prediction accuracy. Moreover, **Fig. 9** compares the predictive errors of three models, and the results indicate that the combined GA-BP neural network model has the minimum relative errors. In other words, the combined GA-BP neural network model has the highest prediction accuracy among the three models, which could represent a good reference for real production.

5. Conclusions

(1) By analyzing the factors which affect end-point manganese content during BOF steelmaking process, a multiple linear regression model for prediction of end-point manganese content was established based on a large amount of production data. The test result shows that the hit rate of the model is 74% when the predictive errors of the model are within ±0.05%, and the hit rate is 44% when the predictive errors are within ±0.03%.

(2) Considering that artificial neural network has a powerful ability to deal with nonlinear problems, BP neural

network model was established to predict BOF end-point manganese content. The verification result reveals that this model has higher prediction accuracy compared with multiple linear regression model. Specifically, the hit rate of the model is 80% when the predictive errors of the model fall within ±0.03%, and the hit rate is 71% when predictive errors are within ±0.025%.

(3) In order to further improve the prediction accuracy, BP neural network and genetic algorithm were amalgamated to establish a combined GA-BP neural network model. Comparison of the predictive errors of the three models indicates that the combined GA-BP neural network model has the highest level of prediction accuracy. The test result indicates that the prediction accuracy of the combined GA-BP neural network model approaches 90% when the predictive errors of the model are within ±0.03%, and the prediction accuracy is 84% when the predictive errors are within ±0.025%. In summary, the combined GA-BP neural network model could provide the most accurate prediction of the end-point manganese content among the three models and is a good reference for real production.

Acknowledgements

The authors are grateful to the National Science Foundation Committee of China and the Ministry of Education of China for providing funding under the National Natural Science Foundation (No.50874014), the New Century Excellent Talents Program in University (NCET07-0067) and Fundamental Research Funds for Central Universities (FRF-BR-09-020B) for supporting this research.

REFERENCES

1) T. Takawa, K. Katayama, M. Hoteiya and N. Hirayama: *Trans. Iron Steel Inst. Jpn.*, **27** (1987), 951.
2) S. Y. Yun and K. S. Chang: *Iron Steelmaker*, **8** (1996), 37.
3) I. J. Coxa, R. W. Lewisb, R. S. Ransingb, H. Laszczewski and G. Berni: *J. Mater. Process. Technol.*, **120** (2002), 310.
4) S. M. Xie, J. Tao and T. Y. Chai: *Control Theor. Appl.*, **20** (2003), 903.
5) Y. F. Wang, X. P. Li, T. Y. Chai and S. M. Xie: *J. Northeastern Univ. (Natural Science)*, **24** (2003), 715.
6) L. Z. Chang and Z. B. Li: *Steelmaking*, **22** (2006), 41.
7) T. Takawa, M. Sato, T. Okada, M. Invi and T. Kvribayashi: *J. Iron Steel Inst. Jpn.*, **74** (1988), 664.
8) L. H. Yang, L. Liu and P. He: *Steelmaking*, **19** (2003), 10.
9) K. Liu, L. Liu and P. He: *Iron and Steel*, **43** (2008), 32.
10) Q. C. Peng, B. Zhang, J. X. Zou, L. C. Liu, Y. Zhao and W. B. Xiao: The 7th Proc. China Iron & Steel Annual Meeting, Metallurgical Industry Press, Beijing, China, (2009), 302.
11) B. Du and Y. Lin: *Comput. Technol. Autom.*, **22** (2003), 103.
12) L. G. Lee, A. R. Mcnamarka, K. C. Teh, H. M. Lie, B. J. Orenstein and D. J. Brown: *ISIJ Int.*, **30** (1990), 90.
13) K. J. Hunt and D. Sbarbaro: *IEE Proc-D: Control Theor. Appl.*, **138** (1991), 431.
14) I. Tone, N. Kitgawa and K. Ishida: The 74th Steelmaking Conf. Proc., Iron and Steel Society, Warrendale, PA, (1991), 375.
15) P. D. Wasserman: Neural Computing: Theory and Practice, Van Nostrand Reinhold Press, New York, (1989), 26.
16) C. R. Li, H. W. Zhao, X. Xie and Q. Yin: *Iron and Steel*, **46** (2011), 23.
17) K. Hornik, M. Stinchcombe and H. White: *Neural Networks*, **2** (1989), 359.
18) N. E. Cotter: *IEEE Trans. Neural Networks*, **1** (1990), 290.
19) M. Nasseri, K. Asghari and M. J. Abedini: *Expert Syst. Appl.*, **35** (2008), 1415.
20) M. Yang, Y. J. Wang and Y. P. Cheng: *Procedia Earth and Planetary Science*, **1** (2009), 1285.
21) F. Yin, H. J. Mao and L. Hua: *Mater. Des.*, **32** (2011), 3457.
22) S. F. Ding, C. Y. Su and J. Z. Yu: *Artif. Intell. Rev.*, **36** (2011), 153.