The Comparison of LSTM and XGB Prediction Model of Hot Metal Temperature in the Blast Furnace

KAI-FAN LO*, CHEN-YUAN LU* AND YUNG-CHANG KO**

** Green Energy & System Integration Research & Development Department*
*** Iron & Steel Research & Development Department*
China Steel Corporation

Thermal control of the blast furnace (BF) is very important. The BF operates at a high temperature in a closed environment, the physical and chemical reactions inside the BF are difficult to measure. Hot metal temperature (HMT) is an important parameter that can be directly measured during the operation. As a result, measuring of the HMT is the main method for monitoring the thermal state of the BF.

In this study, a long short-term memory (LSTM) model and XGBoost (XGB) model were used to train a HMT prediction model. The dataset included 66 selected features with 480 records per feature. An analysis was carried out with the data of an 8-hour operation to find out the critical reaction time of each feature that had the highest correlation with the HMT. The data format was converted from 480 records to 1 record per feature depending on the critical time. The XGB model was trained with the data at the critical time and the LSTM model was trained with the original data. The root mean square error (RMSE) was 15 for the XGB model and 14.6 for the LSTM model. The results showed a slight difference in the LSTM and XGB result. Both prediction models were able to predict the HMT in 2 hours and have been used as a tool for the adjustment of the operation to get better performance from the BF. The HMT variance could be decreased by 30% with the prediction model.

Keywords: AI model, Hot metal temperature, Prediction, XGB, LSTM

1. INTRODUCTION

The thermal state is an important factor to ensure the stable and efficient operation of the BF. But the thermal state is difficult to predict. The HMT is an important parameter of the thermal balance of the BF and is strongly related to the thermal state. Control of HMT is important to realize an efficient and stable operation of the BF. When the coke and iron ores are fed into the BF, the process leads to a long response time of approximately 6~8 hours. The HMT is a lagging indicator from when the temperature is detected, and the temperature is difficult to predict by physical model because of the complicated process dynamics. There are three challenges to predict the HMT. The first is that there are more than 2000 process variables in the BF, but not all affect the HMT. It is also difficult to fully control variables. The second is that the inner variables can’t be measured. The BF is a black box because of the high temperature and closed environment. The third is that reaction time delays can’t be determined. The BF is a continual device and it is difficult to determine when each reaction will be occurring.

In recent years, AI technology has developed rapidly. The AI model can handle more variables than people and analyze the previous data to replace a complicated physical model. In this research, the AI technology is introduced for training the prediction model. The BF is a continual device and the data is time-series. A deep learning model, LSTM[1, 2, 3], which is able to handle a contextual sequence input was used to train the prediction model. And a machine learning model, XGB[4], was used to compare with the LSTM model. The tuyere is the last operational feature and the operation of the tuyere corresponds to the HMT 2 hours after the operation. So, the models would predict the HMT in 2 hours. The predicted values would be used for controlling the tuyere to reduce the variance of the HMT.

2. EXPERIMENTAL METHOD

The model development process is shown in Figure 1. First, from the many features of the BF’s critical features that have a higher correlation with the HMT were found. The data was increased by the number of features, but this huge amount of data was a burden for training the model. Second, is the model training. In this research, the LSTM model and XGB model were used. Finally, the HMT prediction service of the trained model was published on the cloud server. The service supplied the predicted HMT in 2 hours for the operator every minute. And the operator would refer to the predicted data to control the tuyere.

Fig.1. The flow chart of the prediction method.

2.2 Feature Selection

The BF has more than 2000 features, but not all data had an effect on the HMT. The goal of the feature selection was to select the most influential features for increasing the model performance. The amount of data between the features and HMT were mismatched and difficult to calculate the correlation. Many feature selection methods were chosen to calculate the correlation scores between the features and HMT. Figure 2 shows the methods of the feature selection. The methods included Pearson, Spearman, Mutual information, Kendall, and Chi square. The features had different correlation scores from different methods. It was difficult to analyze the correlation. The correlation scores were finally ranked with each method in this study.

The correlation scores were ranked between the features and the HMT in 2 hours by these methods. The final correlation scores were obtained by the sum of the ranked features in all methods. Figure 3 is the correlation scores of these features. The correlation score rapidly increased at the 15th feature. It means the correlation between the features and the HMT were lower after 15th feature in every method. The first 14 features were chosen, and another 52 features similar to the 14 features were also chosen. Finally, 66 features were selected from more than 2000.

2.2 Trained model with LSTM

The production of iron in a BF is a continuous process. The data of the BF is time-series and continuous data. The current state of the BF would affect the state next time. A recurrent neural network (RNN) is an extension of a conventional feedforward neural network, which is able to handle a contextual sequence input. Unfortunately, it is difficult to train RNNs to capture long-term dependencies(1,2). The long short-term memory (LSTM)(3) was presented to solve the problem. The

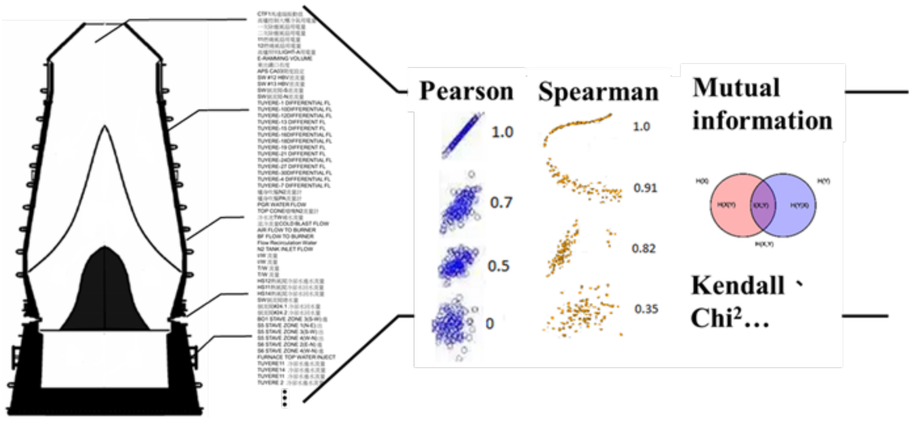

Fig.2. The methods of the feature selections.

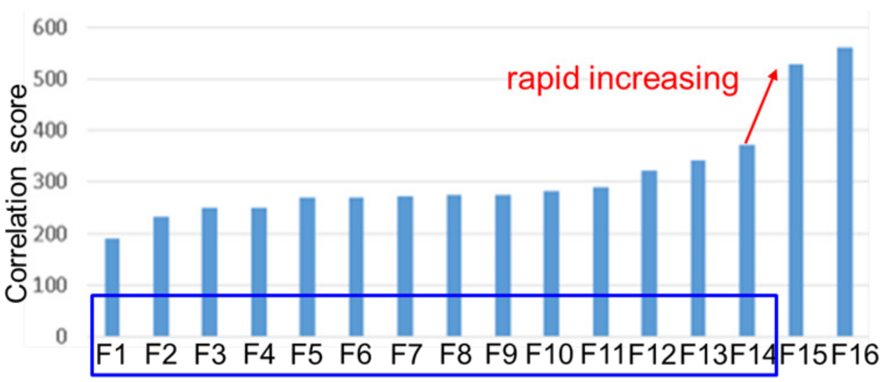

Fig.3. The correlation scores of the features.

models had been shown to perform well in tasks that require capturing long-term dependencies.

The time series data format is shown in Figure 4. An analysis was carried out with data from 8 hours of operation to predict the HMT 2hrs later. The structure of the LSTM is unlike the RNN which simply computes a weighted sum of the input signal and applies a nonlinear function, the LSTM unit maintains two states, ct (cell state) and ht (hidden state). Figure 5 is the structure of the LSTM model. The ht is the hidden layer parameter calculated from the previous state ht-1 and the feature Xt is from the current input. The ct-1 is the information passed from the previous state. The forget gate of the model disregards the unimportant information of ct-1. Then the other information of the ct-1 is calculated with ht-1 and the feature Xt of the current input to obtain the target value Yt+2. Then the cell pass ct and ht to the next time point for calculation.

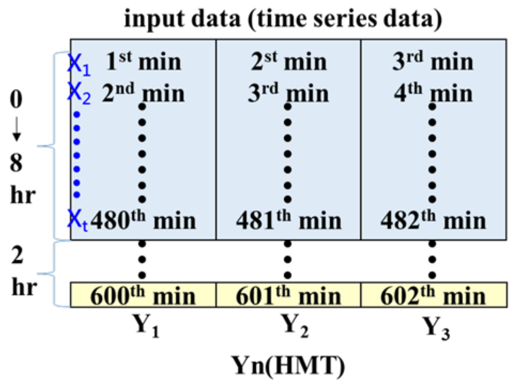

Fig.4. The time-series data format.

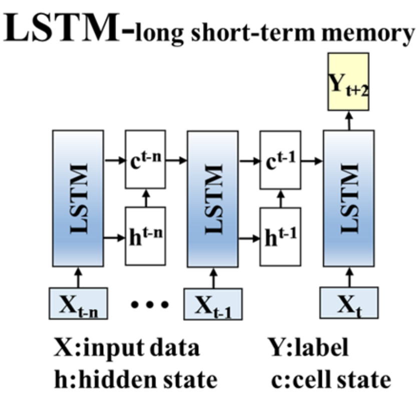

Fig.5. The structure of the LSTM method.

2.3 Trained model with XGB and Determined critical time

The machine learning model, Extreme Gradient Boosting (xgboost or XGB), was also used as training model. The XGB model is a parallel tree boosting model (fig.6) for machine learning. The input data format with XGB is cross-sectional data. Each feature can only provide a single value for the XGB model. The XGB model can't consider the past processes. The critical time needed before training the XGB model is to be determined. The critical time represented is when the reaction of the feature occurs. And the input data format was converted from time-series to cross-sectional depending on the critical times of the features.

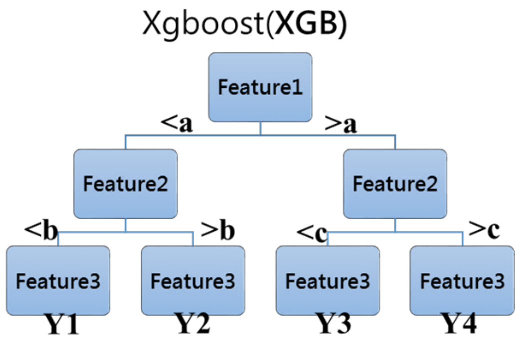

Fig.6. The structure of the XGB method.

Figure 7 shows the critical time determination. The times of the input data were set as features to train the XGB model. Critical times were analyzed by the feature importance of XGB model. It was to analyze every time point from the 1st to the 8th hour of each feature with the HMT and the critical times of the features being ranked and found. The most important hour was the critical time of the feature. The training data of the features were transformed from time-series data to cross-sectional data by critical time. Figure 8 is the data format. The data of each feature were based on the critical time which were during the previous 8 hours, and the data within the critical hour were taken.

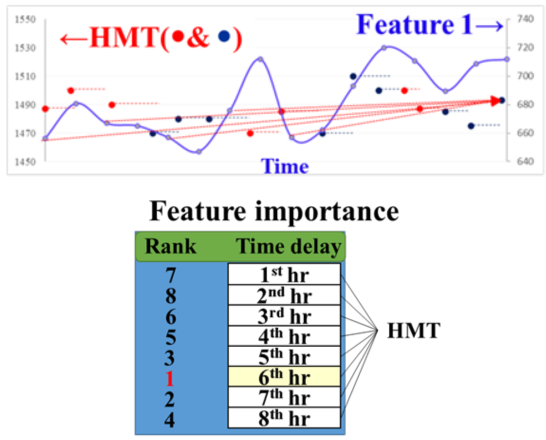

Fig.7. Critical Time determination.

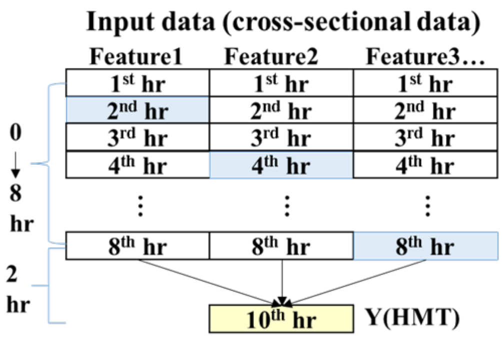

Fig.8. The cross-sectional data format.

3. RESULTS AND DISCUSSION

The AI model was trained with two methods, LSTM and XGB. The input data of the LSTM model had 66 features and the input data was a 66*480 matrix. The output was the HMT in 2 hours. Figure 9 is the predicted result of the LSTM model. There were 1400 records of the training data, and 700 records of the testing data. The blue line is the prediction data and the red points are the measured data from the tap hole of the BF. The RMSE was 14.6°C. The trend of the prediction was conformed to the HMT. The averaged HMT of the BF was about 1475°C. The mean temperature error rate was less than 1%.

Figure 10 is the predicted result of the XGB model. The input data had the same 66 features. But each feature only provided a single value within critical time. The input data was a 66 series data. The predicted data was similar with the LSTM model. The RMSE was 15 and more than LSTM slightly. The calculation resource of the XGB model was less than the LSTM model.

The LSTM model used 50 seconds and XGB model used less than 1 second to predict the HMT with the same calculation device. Both two models were able to predict the HMT in 2 hours and to provide the predicted HMT every minute. The predicted HMT was used to control the tuyere and the variance of the HMT was reduced. The HMT variance could be decreased by 30% with the prediction model.

4. CONCLUSION

1. Two prediction models were established. The prediction models were developed by XGB and LSTM. The root mean square error was 15 for XGB model and 14.6 for LSTM model. Both two prediction models were able to predict the HMT in 2 hours and to present continuous HMT data. The mean temperature error rate was approximately 1%.
2. The time series data of features and discontinuous data of the HMT were analyzed. Ranked correlation scores between the features and HMT were used in this study. There were 66 features successfully selected from more than 2000 features. The less data size increased the model training efficiency and accuracy.

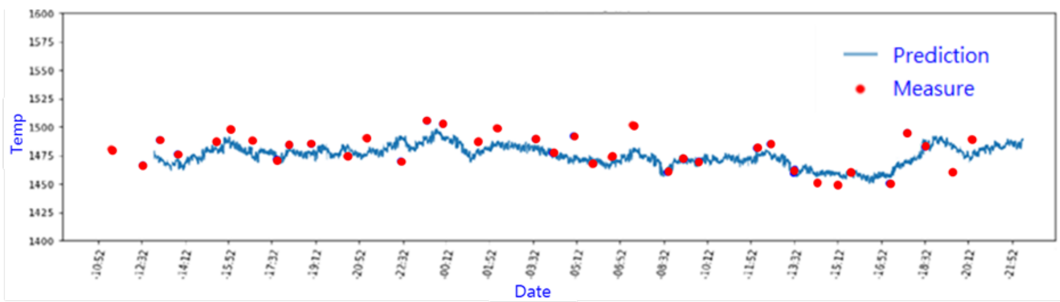

Fig.9. The Result of the LSTM model.

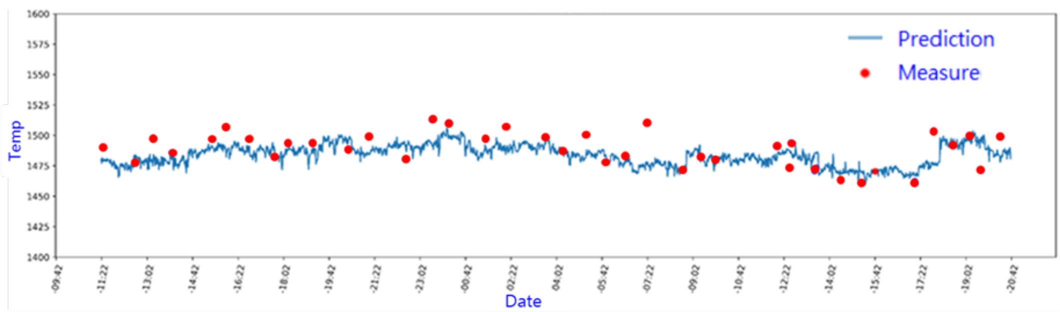

Fig.10. The Result of the XGB model.

3. The critical time of the features were found by feature importance of the XGB model. The training data of the features were transformed from time-series data to cross-sectional data by critical time. The prediction model could be trained with XGB model depending on the critical time.
4. There was a slight difference in LSTM and XGB results, and both two models could predict the HMT in 2 hours and provide the predicted HMT every minute. The predicted HMT was used to help the operation of the tuyere and the variance of the HMT was reduced. The HMT variance could be decreased by 30% with the prediction models.

REFERENCES

1. Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, Yoshua Bengio: Cornell University, Ithaca, New York, 2014.
2. Gulcehre K. Cho, R. Pascanu, Y. Bengio: ECML PKDD, 2014, pp. 530-546.
3. Sepp Hochreiter, Jürgen Schmidhuber: Neural Computation, 1997, vol. 9(8), pp. 1735-1780.
4. Tianqi Chen, Carlos Guestrin: KDD '16, 2016, pp. 785-794.