Real-Time Dynamic Carbon Content Prediction Model for Second Blowing Stage in BOF Based on CBR and LSTM

Maoqiang Gu 1, Anjun Xu 1,*, Hongbing Wang 2 and Zhitong Wang 3

1 School of Metallurgical and Ecological Engineering, University of Science and Technology Beijing, Beijing 100083, China; gmq910710@gmail.com
2 School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China; B0701131@ustb.edu.cn
3 Lingyuan Iron and Steel Co., Ltd., Lingyuan 122500, China; wazhitong18@gmail.com
* Correspondence: anjunxu@metall.ustb.edu.cn

Abstract: The endpoint carbon content is an important target of converters. The precise prediction of carbon content is the key to endpoint control in converter steelmaking. In this study, a real-time dynamic prediction of the carbon content model for the second-blowing stage of the converter steelmaking process was proposed. First, a case-based reasoning (CBR) algorithm was used to retrieve similar historical cases and their corresponding process parameters in the second blowing stage, based on the process parameters of the new case in the main blowing stage. Next, a long short-term memory (LSTM) model was trained by using process parameters of similar cases from the previous moment as the input and the carbon content for the next moment as the output. Finally, the process parameters of the new case were input into the trained LSTM model to produce a real-time dynamic prediction of the carbon content in the second blowing stage. Actual production data were used for the verification, and the results showed that the prediction errors of the proposed model within the ranges of (−0.005, 0.005), (−0.010, 0.010), (−0.015, 0.015) and (−0.020, 0.020) were 25%, 54%, 71%, and 91% respectively, which were higher than the prediction accuracies of the traditional carbon integral model, cubic model, and exponential model.

Keywords: converter; case-based reasoning; long short-term memory; real-time; dynamic

1. Introduction

Converter steelmaking is a very complex process, involving physical and chemical changes at high temperatures. It features high production efficiency, low energy consumption and low costs, and it is currently the main steelmaking method in China. Endpoint control refers to controlling the composition and temperature of molten steel within a reasonable range. However, existing detection methods cannot continuously monitor the composition and temperature in the converter bath because of its extremely high temperature during smelting. Thus, most steelworks still rely on manual experience to maintain endpoint control, and the control accuracy is low and unstable. Establishing an accurate prediction model would be of great significance for endpoint control.

At present, prediction models for the endpoint control of converter steelmaking can be divided into static and dynamic approaches. The former can be further divided into theoretical and data-driven models. Theoretical models are based on the material balance and heat balance, but they incorporate a large number of assumptions that do not truly reflect the actual production process, so their control accuracy is low. Data-driven models use historical production data for learning and modelling, so they can reflect the production process more effectively. Thus, many authors have used the data-driven approach for in-depth research into the converter steelmaking process. Han et al. used case-based reasoning (CBR) and combined different methods for case retrieval and reuse to establish a control model for oxygen supply [1]. Wang et al. proposed a CBR model...
for static parametric control based on causal relationships [2]. Park et al. performed a sensitivity analysis to eliminate irrelevant input parameters and established models based on an artificial neural network and least-squares support vector machine for endpoint temperature prediction [3]. Gao et al. proposed an improved twin support vector regression model for endpoint prediction [4,5]. Li et al. established an endpoint prediction model based on a backpropagation neural network and the improved particle swarm optimisation algorithm [6]. Han et al. proposed an endpoint prediction model based on a membrane algorithm for an improved extreme learning machine [7]. Yan et al. established a model for predicting the endpoint carbon content based on a genetic algorithm and kernel partial least-squares regression [8]. Cheng et al. proposed a data-driven multitask learning method for endpoint prediction [9]. Liang et al. proposed a two-step CBR method based on attribute reduction for predicting the endpoint phosphorus content [10]. Wang et al. established an integrated CBR model for predicting endpoint temperature of molten steel in argon oxygen decarburisation [11]. Other endpoint prediction models have been established based on the flame spectrum and flame images taken at the converter mouth [12–14]. Data-driven models are also widely used for other steelmaking processes. Iftikhar et al. integrated the grey-box model and bootstrap filter to establish a prediction model for the molten steel temperature that accounts for uncertainty [15]. Lv established a model for sensing the sulphur content during ladle furnace steel refinement [16]. Okura used the grey-box model to produce high prediction accuracy for the temperature of molten steel in a tundish [17]. Although the above data-driven models offer improved prediction accuracy, the following problems remain. Firstly, since the static models are only built with consideration for the initial conditions and static process data (a small dataset without a time-series feature cannot represent the actual production), their prediction accuracy is limited [18]. Secondly, the static model can only make endpoint predictions, so it cannot provide a reference for operators to adjust the operating parameters in the blowing process.

In order to solve the problems of the static model, dynamic models were established. Dynamic models can be divided into those based on sub-lance and off-gas analysis. Among sub-lance-based control models, Yue et al. established prediction models for the endpoint carbon, temperature, and phosphor and manganese contents based on the exponential model, heat balance and thermodynamic equations, respectively [19]. Min et al. used an adaptive network-based fuzzy inference system and relevance vector machine to establish a dynamic control model [20]. Wang et al. established a data-driven real-time prediction model for the endpoint carbon content [21]. Among the off-gas analysis-based control models, Hu et al. used off-gas analysis techniques to establish a carbon integral model for continuously calculating the carbon content and temperature [22,23]. Liu et al. proposed an algorithm based on off-gas analysis for dynamically calculating the carbon content of the converter bath at the second blowing stage [24]. Dofasco (Canada) stopped using the sub-lance after adopting an off-gas analysis-based blowing control system. All the heat sources were directly tapped, and the prediction accuracy was 100%, whereas the re-blowing rate was less than 1% [25]. Lin et al. proposed an improved exponential model that fits and updates the critical carbon content curve simultaneously to predict the carbon content in the second blowing stage [26].

The above dynamic control methods improve the prediction accuracy of endpoint control to some extent but still face several problems. The sub-lance cannot achieve continuous measurement and the previous models based on the sub-lance can only fit the relationship between oxygen supply amount or time and carbon content, without considering the impact of the lance position and bottom blowing flow rate on the carbon content. Furthermore, the off-gas detection equipment is far from the reaction zone in the converter bath, which induces a delay in the off-gas data and cannot make real-time prediction of the carbon content.

In this study, CBR and long short-term memory (LSTM) were used to establish a model for the real-time dynamic prediction of the carbon content in the second blowing stage.
The proposed model should offer improved endpoint control of the converter steelmaking process.

2. Converter Steelmaking Process

2.1. Smelting

Figure 1 shows the smelting process of converter steelmaking; the capacity of converter is 300 t. The process is performed to reduce impurities in the hot metal. The temperature of the molten iron is raised from 1350 °C to about 1650 °C. The smelting process includes the addition of scrap and hot metal, blowing oxygen, tapping, and alloying [27]. The blowing oxygen process can be divided into two stages: the main blowing stage and the second blowing stage. In the main blowing stage, the scrap steel and molten iron are loaded into the converter bath, and oxygen is blown into the bath through the oxygen lance while additives (e.g., lime, dolomite and sintered ore) are added. In the second-blowing stage (about 85% oxygen supply), a sub-lance is inserted into the converter bath for temperature, sample and carbon (TSC) measurements. The amounts of the oxygen supply and coolant are adjusted according to the measured carbon content and temperature. After the blowing is completed, the sub-lance is inserted into the bath again for temperature, sample and oxygen (TSO) measurements. If the requirements for endpoint control are met, the tapping and alloying processes are performed. Otherwise, the blowing process is repeated.

Figure 1. Diagram of the smelting process for converter steelmaking.

2.2. Decarburisation

During the converter steelmaking process, the decarburisation rate is slow at the beginning and end and fast in the middle. As shown in Figure 2, the blowing process can be divided into three stages according to the decarburisation rate: initial, intermediate, and final. In the initial stage, oxygen is blown into the bath and oxidises Si and Mn in the hot metal. Only part of the oxygen reacts with the carbon, and the decarburisation reaction is relatively gentle. As the Si and Mn contents in the hot metal decrease and the bath temperature increases, the decarburisation rate gradually increases. In the intermediate stage, the decarburisation reaction becomes violent, and the decarburisation rate remains high for a considerable time. In this stage, the decarburisation rate mainly depends on the oxygen supply. In the final stage, the decarburisation reaction causes the carbon content in the bath to decrease to a critical threshold. The limit on the decarburisation reaction becomes the mass transfer of carbon in the bath. Thus, the decarburisation rate decreases with the carbon content in the bath.

The decarburisation rate during the blowing process is described by:

\[
\begin{align*}
-\frac{dw_c}{dt} &= K_1 t & 0 \leq t < t_A \\
-\frac{dw_c}{dt} &= K_2 & t_A \leq t < t_B \\
-\frac{dw_c}{dt} &= K_3 w_c & t_B \leq t < t_E
\end{align*}
\]

where \( w_c \) is the carbon content in hot metal; \( K_1, K_2, \) and \( K_3 \) are undetermined coefficients; and \( t_A, t_B, \) and \( t_E \) are the ending times of the initial, intermediate, and final stages, respectively.
2.3. Process Parameters for Converter Steelmaking

As shown in Figure 3, the process parameters for converter steelmaking can be divided into two categories according to the data type: single values and time series.

(1) Single-value data

Single-value data include various elemental compositions (e.g., C, Si, Mn, P and S) in the hot metal, the weight of the hot metal, the temperature of the hot metal, the type and weight of scrap, the type and weight of the slagging agent (e.g., lime and dolomite), the iron ore, heating agent, and so on.

(2) Time-series data

According to the delay of automatic acquisition, time-series data can be further divided into non-delayed time-series data and delayed time-series data. The former mainly include changes in the lance position, oxygen flow, and bottom blowing gas flow. The latter mainly generally refer to gas data, including off-gas components and flow. As the detection equipment is away from the converter bath, so the off-gas data feature a certain delay. Therefore, off-gas analysis cannot directly participate in the control of the blowing process, and it is mainly used to evaluate the decarburisation stage.
3. Methodology

3.1. Case-Based Reasoning

CBR is a critical method in the field of artificial intelligence. Once a new problem occurs, similar problems that have been solved and corresponding solutions can be retrieved from the case library. By comparing differences in the backgrounds and times of the occurrence of present and previous problems, solutions to the latter may be adjusted and altered to solve the former, as shown in Figure 4.

![Figure 4](image-url)  
Figure 4. Process of the CBR model.

The CBR process consists of case description, retrieval, reuse, revision, and retention. Case description is also known as case representation, and it is the basis for CBR. A case is generally described in terms of its characteristics and solutions:

\[
\text{Case} = \{x_1, x_2, \cdots, x_n | s\}
\]  
(2)

where \(x_i\) is the \(i\)-th characteristic of the case, \(n\) is the number of characteristics, and \(s\) is the solution.

Case retrieval involves searching the case database for the cases closest to the description of the case in question. When a case has many attributes, the case database usually will not include an exact match. Therefore, a computational method is needed to locate the most similar case, such as the Euclidean distance. Suppose that a case has \(m\) influencing factors, the \(j\)-th influencing factor of a case in the case database is \(y_j\) and the \(j\)-th influencing factor for the new case is \(x_j\). Next, the Euclidean distance between the new case and the case from the case database is described by

\[
d(X, Y) = \sqrt{\sum_{j=1}^{m} (x_j - y_j)^2}
\]  
(3)

The similarity between the cases can be calculated as follows:

\[
G_{\text{sim}}(X, Y) = \frac{1}{1 + d(X, Y)}
\]  
(4)

3.2. Long Short-Term Memory Neural Network

LSTM was first proposed by Hochreiter and Schmidhuber in 1997. In contrast to traditional neural networks, its hidden layers consist of basic units, called memory blocks, in a structure, as shown in Figure 5.
Figure 5. Memory block structure of LSTM.

The memory block has three gates (forget gate $f$, input gate $i$, and output gate $o$) and a memory cell $c$; the value $\oplus$ represents the addition of two vectors, $\otimes$ represents the dot product of two vectors, $\sigma$ is the sigmoid activation function, and $\tanh$ is the hyperbolic tangent activation function. The calculation formula of the model is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$
$$\sim C_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (7)$$
$$C_t = f_t \ast C_{t-1} + i_t \ast \sim C_t \quad (8)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$
$$h_t = o_t \ast \tanh(C_t) \quad (10)$$

where $f_t$ is the output of the forget gate at time $t$, $\sigma$ is the sigmoid function, $h_{t-1}$ is the final output of the cell unit of the previous time and $x_t$ is the input of the present time $t$. The value $i_t$ is the output of the input gate at time $t$, $\sim C_t$ is the input cell state at time $t$, $C_t$ is the memory cell state at time $t$ and $h_t$ is the final output at time $t$. The values $W$ and $b$ are the matrix of coefficients and bias for each gate, respectively.

3.3. Prediction Model Based on off-Gas Analysis

(1) Carbon integral model

Based on the principle of mass balance, the carbon integral model first calculates the initial carbon amount of the molten steel according to the composition of steelmaking raw materials and then subtracts the amount of carbon overflowing from the off-gas in the
form of CO and CO$_2$; the remaining part is the amount of carbon in the molten steel in the molten pool.

According to the flow of off-gas and the percentage content of CO and CO$_2$ in the off-gas, the decarburization rate in the molten pool can be calculated by using the carbon balance in the converter smelting process:

$$V_c = \frac{12}{22.4} \times Q_{\text{smoke}} \times [\varphi(\text{CO}) + \varphi(\text{CO}_2)]$$ \hspace{1cm} (11)

where:

$V_c$ is the decarburization rate of molten pool;
$Q_{\text{smoke}}$ is the flow of off-gas;
$\varphi(\text{CO})$, $\varphi(\text{CO}_2)$ is the percentage content of CO and CO$_2$ in the off-gas;

Then the instantaneous carbon content in the molten pool can be expressed as:

$$w(C)(t) = \frac{W_{C_{ini}} - \int_0^t V_c \, dt}{W_{\text{steel}}}$$ \hspace{1cm} (12)

where:

$w(C)(t)$ is the carbon content in the molten pool at time $t$;
$W_{C_{ini}}$ is the carbon content in the molten pool at the initial conditions;
$W_{\text{steel}}$ is the weight of the molten steel in the molten pool;

(2) Exponential decay model

The exponential model is the most widely used decarburization characteristic model in the later stage of converter. It assumes that there is an exponential attenuation relationship between the decarburization rate and the carbon content in molten pool in the later stage of converter blowing.

$$V_c = k_1 \times (1 - e^{-k_2[w(C)(t) - w(C)_0]})$$ \hspace{1cm} (13)

Equation (14) is obtained from Equation (13):

$$w(C)(t) = w(C)_0 - \frac{\ln(1 - V_c/k_1)}{k_2}$$ \hspace{1cm} (14)

where $k_1$, $k_2$ is the undetermined coefficient of the model.

Cubic model

In addition to the exponential model, the curve used to describe the decarburization characteristics of the converter in the later stage also has a cubic equation:

$$w(C)(t) = b_0 + b_1 \cdot V_C + b_2 \cdot V_C^2 + b_3 \cdot V_C^3$$ \hspace{1cm} (15)

where $b_0$, $b_1$, $b_2$, $b_3$ is the undetermined coefficient of the model.

3.4. Principles of Model

The model framework can be divided into three parts: similar case retrieval, model training, and model validation, as shown in Figure 6.

Figure 7 shows a similar case retrieval process. The process parameters of the main blowing stage (e.g., the composition, temperature, and weight of the hot metal, the type and weight of scrap, and the measurement results of the TSC) of the new case are used to retrieve m similar cases in the historical case database and the corresponding time-series data (e.g., the lance position, the oxygen flow, the bottom-blowing gas flow, and variations in the off-gas and carbon content) in the second blowing stage. In this paper, m is the hyperparameter of the model, which is determined through experiment.
Figure 6. Process of the model.

Figure 7. Process of similar case retrieval.

Figure 8 shows the model training process, which involves training the LSTM models with time-series data for the second blowing stage of the retrieved similar cases. Among the time-series data, the lance position, oxygen flow, bottom-blowing gas flow and carbon content of the previous time are taken as the input, and the carbon content of the next time is taken as the output.
Figure 8. Process of model training.

Figure 9 shows the model verification process. The trained LSTM models take the inputs for the last moment of the new case (i.e., lance position, oxygen flow, bottom blowing gas flow, and carbon content) to obtain the change in carbon content for the next moment. The predicted carbon content is then used as the input for the next moment, and so on. Eventually, a curve is obtained for the variation in the carbon content for the new case.

![Figure 9: Process of model validation.](image)

**Figure 9.** Process of model validation. Where \( P(t_n) \): process parameters (oxygen flow, lance position, bottom blowing gas flow and carbon content) at the \( n \)-th moment. \( C(t_n) \): carbon content at the \( n \)-th moment.

4. Model Application Examples

4.1. Datasets

In order to verify the prediction accuracy of the model, 1209 heats of the actual production data collected from B steel plant were used. The statistical results of the single-process parameters of the converter blowing process are shown in Tables 1 and 2.

**Table 1.** Statistical results of influencing factors of single-value type.

| Influence Factors                        | Symbols | Maximum | Minimum | Mean   | Std.  |
|------------------------------------------|---------|---------|---------|--------|-------|
| \(W[C]_{\text{iron}}/\%\)                | \(X_1\) | 4.5016  | 3.9496  | 4.2273 | 0.0992|
| \(W[Si]_{\text{iron}}/\%\)                | \(X_2\) | 0.49606 | 0.03189 | 0.2644 | 0.08119|
| \(W[Mn]_{\text{iron}}/\%\)                | \(X_3\) | 0.14347 | 0.07155 | 0.10811| 0.01291|
| \(W[P]_{\text{iron}}/\%\)                 | \(X_4\) | 0.07953 | 0.0527  | 0.065721| 0.004375|
| temperature of hot metal/°C              | \(X_5\) | 1441    | 1267    | 1360.3 | 31.3 |
| Weight of hot metal/t                    | \(X_6\) | 282     | 263     | 272.57 | 3.39 |
| Weight of Scrap/t                       | \(X_7\) | 71      | 40      | 60.2621| 6.207|
| Amount of Lime/t                        | \(X_8\) | 20.605  | 2.063   | 9.4865 | 3.6389|
| Amount of Dolomite/t                    | \(X_9\) | 7.991   | 2.001   | 4.7244 | 1.0092|
| Oxygen consumption of main blowing stage/t | \(X_{10}\) | 15022   | 11298   | 12931  | 560  |
| \(TSC[C]/\%\)                           | \(X_{11}\) | 0.559   | 0.038   | 0.24437| 0.10683|
| \(TSC[C]/\circ C\)                      | \(X_{12}\) | 1690    | 1545    | 1617.3 | 24.9 |
| \(TSCO[C]/\%\)                          | \(Y\)   | 0.066   | 0.018   | 0.04225| 0.00782|
Table 2. Statistical results of the influencing factors of the time-series type.

| Influence Factors | Mean       | Maximum     | Maximum Length of Time-Series | Minimum Length of Time-Series |
|-------------------|------------|-------------|-------------------------------|------------------------------|
| Oxygen flow       | 66,282 Nm$^3$/h | 29,998 Nm$^3$/h | 2 min                         | 1 min                        |
| Lance position    | 5553 mm    | 1598 mm     | 2 min                         | 1 min                        |
| Argon flow        | 2155 Nm$^3$/h | 465 Nm$^3$/h | 2 min                         | 1 min                        |

Where $TSC_{[C]}$ means the carbon content of TSC detector and $TSC_{[T]}$ means the temperature of TSC detector. The value $TSO_{[C]}$ means the TSO measurement result of carbon content. The length of the time-series in Table 2 means the time interval between the end-point and the beginning-point of the time-series data. The role of argon bowing is to stir the molten pool to achieve a uniform chemical composition and temperature of the molten steel and to accelerate the chemical reaction.

4.2. Similar Case Retrieval

Because the converter steelmaking process is a system that is far from equilibrium, the sub-lance detection results cannot fully reflect the state of the converter bath. Therefore, the conditions of the raw materials and the operating process in the main blowing stage need to be comprehensively considered. The inputs of the CBR model are the carbon content of the hot metal, the silicon content of the hot metal, the manganese content of the hot metal, the phosphorus content of the hot metal, the temperature of the hot metal, weight of the hot metal, the weight of the scrap, the amount of lime, the amount of dolomite, the level of oxygen consumption in the main blowing stage, the carbon content of the TSC detector, and the temperature of the TSC detector. The Euclidean distance was used to measure similarity. The data were standardised to (0, 1), and the four cases with the highest similarity were selected for reuse. This paper uses an example when $m = 4$ to show the process of the model.

Table 3 presents the similar cases retrieved by CBR for a certain heat. Heat 1 is the new case, and heats 2, 3, 4, and 5 are similar cases retrieved from the case database.

Table 3. The similarity between the new case and the similarity cases.

| Heat | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | $X_{11}$ | $X_{12}$ | Similarity |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|------------|
| 1    | 4.13317 | 0.14899 | 0.09142 | 0.06122 | 1377 | 273 | 63 | 5.981 | 5.276 | 12,884 | 0.241 | 1584 | -          |
| 2    | 4.10874 | 0.09697 | 0.08969 | 0.06242 | 1376 | 275 | 63 | 4.219 | 4.755 | 12,659 | 0.207 | 1577 | 94.25%     |
| 3    | 4.05288 | 0.13592 | 0.0925 | 0.0605 | 1394 | 273 | 61 | 5.53 | 4.545 | 12,878 | 0.226 | 1603 | 93.18%     |
| 4    | 4.04647 | 0.15276 | 0.09448 | 0.05964 | 1381 | 272 | 65 | 6.081 | 4.906 | 13,373 | 0.299 | 1597 | 91.95%     |
| 5    | 4.1307 | 0.14709 | 0.0927 | 0.05653 | 1354 | 273 | 65 | 6.074 | 4.596 | 13,101 | 0.303 | 1586 | 91.41%     |

Figure 10 shows the lance position, oxygen flow, bottom-blowing gas flow and carbon content variation in the second blowing stage of similar cases; the change in carbon content is fitted to a curve according to the carbon integral model.
Figure 10. Cont.
Figure 10. Process parameters for the late blowing stage of similar cases. (a) Heat 2; (b) Heat 3; (c) Heat 4; (d) Heat 5.

4.3. Model Training and Validation

The lance position, oxygen flow rate, bottom-blowing gas flow and carbon content at the previous moment were used as the model inputs, and the change in carbon content
at the next moment was used as the model output. The model was set to 10 neurons, 1 batch and 500 epochs, with the mean absolute error as the loss function and Adam as the optimisation solver. Figure 11 shows the error curve of the trained model.

![Error Curve](image)

**Figure 11.** Training error curve of each model.

The trained model was used to verify the new case. The prediction results are shown in Figure 12:

![Prediction Result](image)

**Figure 12.** Prediction results of the application example.

As can be seen from the above Figure, the endpoint prediction result of the example was 0.0477 and the actual value was 0.046. The prediction error was 0.0011.

In order to further verify the model, all 1209 heat data were divided into training set and test set at random, of which 1109 heats were in the training set and 100 heats were in the testing set. The prediction results of the model when $m = 4$ is shown in Figure 13. The hit rate of the prediction error within the range of $(-0.02, 0.02)$ was 91%.
This paper also presents the influence of the hyperparameter on the model’s prediction accuracy. The prediction results are shown in Figure 14.

As can be seen from the above Figure, as the number of cases increases, the prediction accuracy of the model increases first and then decreases. The prediction accuracy of the model is highest when $m = 4$. This shows that increasing the number of cases when training the model helps to improve the prediction accuracy of the model, but as the similarity of the retrieved cases decreases, the reference value of the case decreases, which reduces the prediction accuracy of the model instead.

To further verify the prediction accuracy of the proposed model, its performance was compared against those of a carbon integral model, cubic model and exponential model. The prediction results are shown in Figure 15.
Figure 15. Prediction accuracies of each model.

The prediction accuracy of the proposed model \((m = 4)\) reached 25%, 54%, 74%, and 91% within the ranges of \((-0.005, 0.005)\), \((-0.010, 0.010)\), \((-0.015, 0.015)\), and \((-0.020, 0.020)\), respectively. Compared with the other three models, the proposed model improved the prediction accuracy in the range of \((-0.02, 0.02)\) by 22%, 14%, and 6%, respectively. This demonstrates that the established model can effectively predict the carbon content at the end of the converter steelmaking process and provide guidance to operators.

5. Conclusions

The precise prediction of carbon content is the key to endpoint control in converter steelmaking. In this study, a data-driven model based on CBR and LSTM was established for the endpoint carbon content by predicting the carbon content variation in the second blowing stage. The prediction accuracy of the proposed model \((m = 4)\) within the ranges of \((-0.005, 0.005)\), \((-0.010, 0.010)\), \((-0.015, 0.015)\), and \((-0.020, 0.020)\) reached 25%, 54%, 71%, and 91%, respectively. The success rate was higher than those of the traditional carbon integral model, the cubic model, and the index model.

In future studies, we will investigate the following problems in the endpoint control of the converter.

1. The dynamic prediction model to predict the temperature of molten steel in the second blowing stage.
2. The dynamic prediction model to predict the temperature and carbon content of molten steel in the whole blowing stage.

Author Contributions: Conceptualization, M.G. and A.X.; methodology, M.G.; validation, M.G.; formal analysis, M.G.; resources, A.X.; data curation, Z.W.; writing—original draft preparation, M.G.; writing—review and editing, H.W.; supervision, A.X.; project administration, A.X.; funding acquisition, A.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the National Key Technology R & D Program of China, grant number 2017YFB0304000&2017YFB0304001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Some of the data provided in this study are from third party and some are from my own research. All data has not been stored in the database.
Conflicts of Interest: The authors declare no conflict of interest.

References

1. Han, M.; Wang, X.Z. BOF oxygen control by mixed case retrieve and reuse CBR. IFAC Pro. Vol. 2011, 44, 3575–3580. [CrossRef]
2. Wang, X.; Han, M. Causality-based CBR model for static control of converter steelmaking. J. Dalian Univ. Technol. 2011, 51, 593–598. (In Chinese)
3. Park, T.C.; Kim, B.S.; Kim, T.Y.; Jin, B.; Yeo, Y.K. Comparative Study of Estimation Methods of the Endpoint Temperature in Basic Oxygen Furnace Steelmaking Process with Selection of Input Parameters. Korean J. Met. Mater. 2018, 56, 813–821. [CrossRef]
4. Gao, C.; Shen, M.G.; Wang, L.D. End-point Prediction of BOF Steelmaking Based on Wavelet Transform Based Weighted TSVR. In Proceedings of the 37th Chinese Control Conference, Wuhan, China, 25–27 July 2018.
5. Gao, C.; Shen, M.G.; Liu, X.P.; Wang, L.D.; Chen, M. End-point Prediction of BOF Steelmaking Based on KNNWTSVR and LWOA. Trans. Indian Inst. Met. 2018, 72, 257–270. [CrossRef]
6. Li, W.; Wang, X.C.; Wang, X.S.; Wang, H. Endpoint Prediction of BOF Steelmaking Based on BP Neural Network Combined with Improved PSO. Chem. Eng. Trans. 2016, 51, 475–480.
7. Han, M.; Liu, C. Endpoint prediction model for basic oxygen furnace steel-making based on membrane algorithm evolving extreme learning machine. Appl. Soft Comput. 2014, 19, 430–437. [CrossRef]
8. Zhou, M.C.; Zhao, Q.; Chen, Y.R. Endpoint prediction of BOF by flame spectrum and furnace mouth image based on fuzzy support vector machine. Optik 2018, 178, 575–581. [CrossRef]
9. Chen, J.; Wang, J. Endpoint prediction method for steelmaking based on multi-task learning. J. Comput. Appl. 2017, 37, 889–895. (In Chinese)
10. Liang, Y.R.; Wang, H.B.; Ju, A.J.; Tian, N.Y. A Two-step Case-based Reasoning Method Based on Attributes Reduction for Predicting the Endpoint Phosphorus Content. ISIJ Int. 2015, 55, 1035–1043. [CrossRef]
11. Wang, H.B.; Xu, A.J.; Al, L.X.; Tian, N.Y.; Du, X. An Integrated CBR Model for Predicting Endpoint Temperature of Molten Steel in AOD. ISIJ Int. 2012, 52, 80–86. [CrossRef]
12. Liu, H.; Wu, Q.; Wang, B.; Xiong, X. BOF steelmaking endpoint real-time recognition based on flame multi-scale color difference histogram features weighted fusion method. In Proceedings of the 35th Chinese Control Conference, Chengdu, China, 27–29 July 2016.
13. Luo, T.; Liu, H.; Wu, Q.S.; Wang, B. Prediction method of carbon content in BOF endpoint based on convolutional neural network. Inf. Tech. 2018, 42, 150–155. (In Chinese)
14. Yan, L.T.; Li, M.; Yang, D.Y. Prediction of Carbon Content at End Point Based on GA-KPLSR in Converters. Control Eng. China 2017, 24, 923–926.
15. Ahmad, I.; Kano, M.; Hasebe, S.; Kitada, H.; Murata, N. Prediction of Molten Steel Temperature in Steel Making Process with Uncertainty by Integrating Gray-Box Model and Bootstrap Filter. J. Chem. Eng. Japan 2014, 47, 827–834. [CrossRef]
16. Lv, W. A Novel Process Modeling Method for Steel Sulphur Content Soft Sensing during Ladle Furnace Steel Refining. ISIJ Int. 2019, 59, 1276–1286. [CrossRef]
17. Okura, T.; Ahmad, I.; Kano, M.; Hasebe, S.; Kitada, H.; Murata, N. High-Performance Prediction of Molten Steel Temperature in Tundish through Gray-Box Model. ISIJ Int. 2013, 53, 76–80. [CrossRef]
18. Yue, F.; Bao, Y.P.; Cui, H.; Gao, S.Y.; Li, B.H.; Zhang, J. Sub-lance control-based predication model for BOF end-point. Steelmaking 2009, 25, 38–59.
19. Han, M.; Zhao, Y. Dynamic control model of BOF steelmaking process based on ANFIS and robust relevance vector machine. Expert Syst. Appl. 2011, 38, 14786–14798. [CrossRef]
20. Wang, X.Z.; Xing, J.; Dong, J.; Wang, Z.S. Data driven based endpoint carbon content real time prediction for BOF steelmaking. In Proceedings of the 36th Chinese Control Conference, Dalian, China, 26–28 July 2017.
21. Hu, Z.G.; He, P.; Liu, L.; Tan, M.X. Continuous Determination of Bath Carbon in BOF by Off-gas Analysis. Res. Iron Steel 2003, 31, 12–15.
22. Hu, Z.G.; He, P.; Liu, L.; Tan, M.X. Continuous Determination of Bath Temperature in BOF by Off-gas Analysis. In Proceedings of the 5th CSM Annual Steelmaking Conference Proceedings, Congqing, China, 22–24 April 2003.
23. Liu, K.; Liu, L.; He, P.; Liu, W. A new algorithm of endpoint carbon content of BOF based on of off-gas analysis. Steelmaking 2009, 25, 33–37.
24. Wang, X.H.; Liu, J.Z.; Liu, F.G. Technological progress of BOF steelmaking in period of development mode transition. Steelmaking 2017, 33, 1–55.
25. Liao, D.S.; Sun, S.; Waterfall, S.; Boylan, K.; Pyke, N.; Holdridge, D. Integrated KOBM Steelmaking Process Control. In Proceedings of the 6th International Steelmaking Science and Technology Conference, Beijing, China, 12–14 May 2015.
26. Lin, W.H.; Jiao, S.Q.; Sun, J.K.; Liu, M.; Su, X.; Liu, Q. Modified exponential model for carbon prediction in the end blowing stage of basic oxygen furnace converter. Chin. J. Eng. 2020, 42, 854–861.
27. Sala, D.A.; Jalalvand, A.; Deyne, Y.D.; Manners, E. Multivariate Time Series for Data-Driven Endpoint Prediction in the Basic Oxygen Furnace; ICMLA/IEEE: Orlando, FL, USA, 2018.