Abstract: Oxygen volume control is very important in BOF (Basic Oxygen Furnace) steelmaking production. In this study, a case-based reasoning (CBR) model with mixed case retrieve and case reuse steps is adopted to calculate the oxygen volume. Two case retrieve methods are used here: KNN (k nearest neighborhood) and geometrical similarity. The geometrical similarity is proposed to make up the lack of KNN method when no case is retrieved by KNN. Correspondingly, two case reuse solutions are employed. Tests are implemented on a practical 180t converter and results show that this CBR system for BOF oxygen volume control is effective.

1. INTRODUCTION

BOF is a widely preferred and effective steelmaking method due to its high productivity and considerably low production cost. In practical steelmaking process, oxygen is blown onto the hot metal surface to lower the levels of impurities and raise the temperature. The criterion whether the molten steel is acceptable or not is often decided by the endpoint carbon content and temperature. So, the control of oxygen column in steelmaking production is very important. However, BOF steelmaking process is very complicated with high temperature and multiphase physical chemistry reactions. And also because of the lack of necessary measurement, oxygen control is also a hard task.

For converter steelmaking plant, mechanism model, statistical model and neural network model [1] are commonly used in recent years. A solution based on heat and mass balance from the static model [2] has been employed. But unfortunately, a lot of theoretical assumptions and too many parameters are involved in the traditional control methods, such as mechanism models based on heat balance and material balance, or statistic models based on regression analysis. Therefore, these models are often difficult on modeling precisely. Waste gas analysis based decarburization rate model [3] is another valid method but a waste gas analysis instrument is needed. With the development of artificial intelligence technology, neural network models and other intelligent methods have been widely applied in BOF steelmaking process. A fuzzy model was proposed [4] for the oxygen volume of BOF process. As a result of the application of the fuzzy model, acceptable levels of compatibility were achieved compared to the empirical BOF data in an integrated steel plant in Turkey. Three BP models were proposed [5] for the dynamic control of BOF process, these three models are used to predict the quantity of end-blow oxygen and end-blow coolant and determine whether the coolant is added or not. Other different intelligent models [6,7,8] were also adopted to describe the BOF steelmaking process. ICA and Greedy Kernel Components [9,10] are used to reduce the dimensions in steelmaking process. It is a main trend that artificial intelligence methods are applied in BOF process modeling.

Case based reasoning is an artificial intelligence approach for learning from the past experiences and very similar with the operator’s decision process. CBR is a proper method for complicated problems and has been applied in many fields, such as problem diagnosis [11], expression recognition [12], medical diagnosis [13] and chemical engineering mechanics [14] and so on.

In this research, CBR approach is applied in metallurgy engineering especially for BOF oxygen volume control. In proposed CBR model, a hierarchical framework is adopted here to represent the case. A case is represented as a vector space and contains discrete attributes and continuous attributes. Then, two retrieval techniques, ‘geometrically similar’ and ‘nearest neighbours’, are used to retrieve similar cases from a limited number of cases in case base. Correspondingly, two reused strategy are used to construct the final solution for current problem based on two case retrieve methods.

2. CBR MODEL FOR BOF STEELMAKING OXYGEN VOLUME CONTROL

2.1 BOF Steelmaking Description

The BOF comprises a vertical solid-bottom crucible with a vertical water-cooled oxygen lance entering the vessel from above. General view of BOF is given in Fig. 1. The vessel is tiltable for charging and tapping. The charge is normally made up of 85% molten pig iron (“hot metal”) and 15% scrap before blowing the oxygen. After that, the converter is rotated to a vertical position and then oxygen is turned on. During the blowing period, high-purity oxygen is blown onto the top of the hot metal at a speed of 16,000 cubic feet per minute. Assistant materials such as burnt-lime, dolomite and iron ore are added into vessel by two or three batches during the blowing time. BOF process aims to lower the levels of impurities which are carbon, silicon, manganese and phosphorus and raise the temperature from about 1350°C to
1680°C. The heat is from the oxidation reaction and no external heat is required. Carbon is oxidized to carbon monoxide and carbon dioxide, are taken out from the exhaust gas hood. Silicon, manganese and phosphorus are oxidized and combined with the assistant materials to form the slag. If the carbon level and temperature meet the requirements, the liquid steel is tapped from the tapping hole into the steel ladle. Slag is on the top of steel and be left in the converter.

The objective is to produce a desired amount of steel, which consists of specified chemical composition, at the proper tapping temperature. Control is difficult because the entire process takes only half an hour and it is hard to sample and analyze during this time. Generally, the real production adopts dynamic control practice. Before blowing, the static model estimates the oxygen consumption value of the total steelmaking process. When the blowing reaches about 85% of the total value, the sub-lance (i.e. a probe is dipped into bath) goes down into the liquid steel to detect carbon content and temperature. Based upon the on-line measured carbon content and temperature, the oxygen volume and the added coolant weight in second blow period are determined. The second blow period is also called dynamic period or end-blow period. Correspondingly, in metallurgy, the model of this period is called dynamic model. If the carbon content or temperature of second measurement does not hit the target, a re-blow step is needed until liquid steel carbon content and temperature are acceptable.

2.2 CBR Model

Case-based reasoning is an analogical reasoning method, mainly contains the following steps: case description, case retrieval, case reuse, case revise and retain. First of all, set up a description of the problem (case), according to the case description, case retrieval is occurred in case library to find a number of the most similar cases to current problem. The solution of current problem is constructed by the solutions of selected similar cases in case retrieve step. Finally, the solution is output to users. If the solution solves the problem successfully, then this case will be stored into the case base. The structure of case-based reasoning BOF oxygen volume control system is shown in Fig 2.

Fig. 1 General View of BOF

Fig. 2 BOF Oxygen Volume Control System

2.2.1 Case Description

Case description is the basis of case-based reasoning process. A felicitous representation guarantees the effectiveness of case retrieval and case reuse. A case is divided into two parts: problem part and solution part. For the problem of BOF oxygen volume control, solution attribute is oxygen volume. Problem attributes are built based on BOF production condition information, technics objectives and process data. There is a number of factors affect BOF oxygen volume control, such as requirements of technics objectives, the operation of blowing process, as well as raw materials component and so on. The case description of BOF oxygen volume contains two types:

(1) Discrete attributes. Endpoint aimed carbon content and temperature: According to operation rules, endpoint aimed carbon content and temperature of molten steel in BOF are vary with the steel species. They are the direction and object of the production and affect the oxygen volume greatly. Therefore, discrete attributes of a case include aimed molten steel carbon content AC and aimed molten steel temperature AT. Thus, a discrete attributes set can be expressed as DF = {AT, AC}.

(2) Continuous attributes. Hot metal information and Material information are considered when constructing the continuous attributes. Thus, continuous attributes set can be expressed as CF = {HM, MA}, where, HM represents hot metal information, including hot metal weight Fe_W, hot metal carbon content Fe_C, silicon content Fe_Si, manganese content Fe_Mn, phosphorus content Fe_P and hot metal temperature Fe_T; MA represents the amount of raw materials addition, including scrap SA, active lime LiA and dolomite DA;

Summarize above, a case can be described as a four-storey structure, shown in Fig 3.
2.2.2 Case Retrieval

Case retrieval is the key step in case-based reasoning process and it is the foundation of constructing an exact solution for current problem. The core of case retrieval is the similarity between current case and historical cases. By far, the most commonly used retrieval techniques are nearest neighbour retrieval, inductive approaches, knowledge guided approaches and validated retrieval. A number of the most similar cases to current problem can be obtained in this step and will be used in next step to construct a solution of current case. In this paper, a limited number of cases are selected as a case base for each current case. So a situation that can not search a similar enough case is existed. For this, two similarity calculation methods are adopted and correspondingly two case reuse schemes are used to construct the resolve of current problem. In case retrieve and reuse steps, two mixed case retrieve and reuse strategies are proposed here shown as Fig. 4.

**Fig. 4 Mixed Case Retrieve and Reuse Strategies**

The KNN case retrieve is a preferred method to search the similar cases, if there is no case satisfies the inequality $\text{Sim} > T$, then use the geometrical similarity to search the similar cases and use a different case reuse method to calculate the reused solution, where Sim is a similarity value between current case and a case in case base and T is a threshold value.

### a. KNN case retrieval

According to the case description, the similarity of discrete attributes and continuous attributes are calculated separately. And then the total similarity is established by considering discrete and continuous similarity synthetically. At the same time, the calculation is based on the three layer structure of its own.

Top layer similarity for the continuous attributes is expressed as the weighted summation form of four continuous attributes sets, shown as formula (1):

$$\text{Sim} = \sum_{i=1}^{4} w_i \cdot \text{Sim}_i$$

where, Sim is the total similarity between two cases; $\text{Sim}_i$ is the total similarity of a attribute set between the current case and a historical case, four continuous attributes sets are equipment information set, hot metal information set, raw materials addition information set and active lime component information set. $w_i$ is the weight of the corresponding attributes set, which is used to differentiate the importance of these four attributes sets.

Middle layer similarities are used to describe four attributes sets similarity respectively. For each attributes set, variables are continuous, using the following formula to calculate their similarities:

$$\text{Sim}_i = \frac{\sum_{j=1}^{m} w_j \cdot \text{sim}_j(f_j, f_i)}{\sum_{j=1}^{m} w_j}$$

(2)

where, $m$ is the number of attributes in the $i$th attribute set; $f_j$ is the value of the $j$th attribute in the $i$th attribute set of current case; $f_i$ is the value of the $j$th attribute in the $i$th attribute set of a historical case; $\text{sim}_j(f_j, f_i)$ is the similarity of the $j$th attribute in the $i$th attribute set between two cases, and also the bottom layer similarity, the calculation formula is shown as follows:

$$\text{sim}_j(f_j, f_i) = 1 - \frac{|f_j - f_i|}{\max(f_j, f_i)}$$

(3)

$w_j$ is the weight of the similarity of the $j$th attribute in the $i$th attribute set. Its value is established as follows:

Firstly, calculate the standard deviation of each attribute similarity value:

$$\sigma_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{sim}_j - \text{\overline{sim}_j})^2}$$

(4)

where, $n$ is the number of cases in case library; $\text{sim}_j^k$ is the similarity of the $j$th attribute between the $k$th case in case library and the $i$th case. The situation that $\text{sim}_j^k = \text{\overline{sim}_j}$ means that $j$th attribute of the $k$th case is the same as the $j$th attribute of the $i$th case.
library and current case; \( \bar{\text{sim}} \) is the average value of the \( j \)th attribute similarities between current case and cases in case library.

Then, the weight value is calculated as follows:

\[
w_{ij} = \frac{\sigma_j}{\sum_{j=1}^{n} \sigma_j}
\]

(5)

Total similarity of two cases is calculated as the form of the product of calculation of discrete attributes similarity and continuous attributes similarity:

\[
\text{SIM} = S_d \cdot \text{Sim}
\]

(6)

where, Sim denotes the similarity of continuous attributes. \( S_d \) denotes the similarity of discrete attributes, which calculated as below:

\[
S_d = \begin{cases} 
1 & \text{DF}_p = \text{DF}_c \\
0 & \text{DF}_p \neq \text{DF}_c 
\end{cases}
\]

(7)

Where, \( \text{DF}_p \) denotes discrete attributes of current case; \( \text{DF}_c \) denotes discrete attributes of a case in case base. If aimed temperature and aimed carbon content are the same as those of \( \text{DF}_c \), the similarity of these two cases \( S_d \) is set to 1, else \( S_d \) is set to 0.

For current problem, computing the total similarities between current case and every historical case and ordering them by descending. Then, a number of the most similar cases those similarities with current case are higher than a threshold value can be chosen and used in the next step.

b. Geometry similarity case retrieval

A shortage of using the KNN method to retrieve similar cases is that a similarity threshold must be set to select the similarity threshold. Those cases whose similarities are higher than the threshold value are used in reuse step. If the threshold is set too high, there is no similar case is retrieved. If the threshold value is too low, it will affect the accuracy of case reused. So, a geometry similarity is used to find cases for increment regress reused when none of cases are retrieved by KNN method.

In this paper, continuous attributes of a case is described as a set and be expressed as \( \text{CF} = \{\text{HM, MA}\} \). Further more, each cell contains their elements. Take hot metal information cell HM for example, it contains hot metal carbon content \( \text{Fe}_C \), silicon content \( \text{Fe}_Si \), manganese content \( \text{Fe}_Mn \), phosphorus content \( \text{Fe}_P \) and hot metal temperature \( \text{Fe}_T \). For this cell, we describe it as a pentagon and each element represents an apex, shown as Fig 5. Given a regular pentagon with a gravity center \( O \), make the lengths between point \( O \) and apexes (shown as the dotted line) equal to 1. Before describing the cases, normalize the values of attributes such as \( \text{Fe}_C, \text{Fe}_Si, \text{Fe}_Mn, \text{Fe}_P \) and \( \text{Fe}_T \) into \((0,1)\). Then, the hot metal information cell can be described as a pentagon shown as the small pentagon in Fig 5. The length of red line denotes the attribute value after normalization. Other cells are described as polygons.

![Fig.6 Two Types of Geometry Similarity](image)

In this paper, Type \( b \) is used to retrieve similar cases on geometry. The Calculation is shown as below:

\[
sim = \max (\text{sign}(\frac{x_i^k - x_j^k}{x_i^j - x_j^j})) \max (\text{abs} (\frac{x_i^k - x_j^k}{x_i^j - x_j^j}))
\]

(8)

where, \( x_i \) is a attribute and \( x_j \) is a neighbour attribute of current case. \( x_i^k \) is a attribute and \( x_j^k \) is a neighbour attribute of historical case in case base. Formula (8) can be divided into two items. First item is a sign item \( \text{sign}(\frac{x_i^k - x_j^k}{x_i^j - x_j^j}) \) which reflects the sign of similarity. If the sign item is positive, means the edge \( x_i, x_j \) of current case have no point of intersection with the edge \( x_i^k, x_j^k \) of historical case in case base. Else, if the sign item is negative, means the edge \( x_i, x_j \) of current case have a point of intersection with the edge \( x_i^k, x_j^k \) of historical case in case base. Second item is a value item \( \max (\text{abs} (\frac{x_i^k - x_j^k}{x_i^j - x_j^j})) \) which reflects the magnitudes of similarity. The larger the value the less similarity is. So, the retrieved cases’ similarity with current case must larger than zero and less than a threshold value.

2.2.3 Case Reuse
A proposal solution or a final solution will be achieved in case reuse step. The reused method affects the accuracy of the given solution to current problem. Corresponding two retrieved methods, two reused methods are adopted in case reuse step.

**a. Sum of weighted solutions**

For KNN case retrieve, the case reuse solution in this paper is obtained by following equation:

\[
S = \sum_{i=1}^{k} \left( w_i s_i \right) \sum_{i=1}^{n} w_i
\]

where \( S \) is the solution proposed by case-based reasoning, \( s_i \) is the solution of the \( i \)th similar case, \( k \) is the number of the similar cases, \( w_i \) is the weighted value to each similar solution, the magnitude is SIM.

**b. Incremental regression**

For geometry similarity case retrieval, an incremental model is adopted as the case reuse solution.

Given \( n \) similar cases are retrieved. Averaging the retrieved cases’ attributes values as a reference, shown as formula (10).

\[
x^{fr}_i = \frac{1}{n} \sum_{i=1}^{n} x^{f}_i; y^{fr} = \frac{1}{n} \sum_{i=1}^{n} y^i ; i = 1, 2, \ldots, n ; j = 1, 2, \ldots, l
\]

where, \( x^{fr}_j \) is the \( j \)th attribute value of the \( i \)th similar case. \( x^f_i \) is the reference value. \( y^i \) is solution of the \( i \)th similar case. \( y^{fr} \) is the reference value of solution.

To obtain the incremental model, independent variables and dependent variable are determined as formula (11).

\[
\Delta x^{f}_i = x^{fr}_i - x^{f}_i ; \Delta y^i = y^i - y^{fr} ; i = 1, 2, \ldots, n ; j = 1, 2, \ldots, l
\]

Then, the incremental model is constructed as below

\[
\Delta y = a_1 \cdot \Delta x_1 + a_2 \cdot \Delta x_2 + \ldots + a_i \cdot \Delta x_i
\]

Finally, the reused solution is calculated by formula (13)

\[
y^{fr} = y^{fr} + \Delta y
\]

This proposed solution output to operator, if be accepted, the solution will used to guide the production, else a further revise to the solution is carried out by operator.

**2.2.4 Case Retain**

The storage of case and maintenance of case base are very necessary to the performance of case-based reasoning system. With new cases store in the case base, the knowledge of case base will be enriched. However, if the number of cases in base is too large, the efficiency of the case-based reasoning will be descended. So, proper strategy for storage and maintenance is needed to perfect the case base and ensure the efficiency of the reasoning system. If the endpoint carbon content and temperature of current case hit the aims and the similarities between current case and cases in case base are not high, then current case will be added into the case base.

**3. APPLICATION**

Tests are carried out based on a 180 tons converter practical production data. 4000 sets of data are chosen as original case base. The endpoint carbon content and temperature of these heats are satisfied the technique aims. And calculate the oxygen volume of another 150 sets.

**3.1 Criteria**

Two criteria are employed to evaluate the results: hit ratio and root mean square error (RMSE). Hit is used to judge whether the prediction result satisfies the request of practical production or not. If the absolute value of distance \(|\delta|\) between practical oxygen volume and calculated value satisfies \(|\delta| < 500\), define that oxygen volume calculation hit. Hit ratio is the ratio of hit heats in all calculated heats that can be calculated as formula (14)

\[
Hr = \frac{N_{hit}}{N_{heats}}
\]

where \( Hr \) is the hit ratio, \( N_{hit} \) is the number of heats that hit the target, \( N_{heats} \) is the number of the calculated samples.

Root mean square error (MSE) is used to evaluate the precision of the model which can be calculated as follows

\[
\text{rmse} = \sqrt{\frac{\sum_{i=1}^{n} (p_i - r_i)^2}{n}}
\]

where \( p_i \) is the prediction value, \( r_i \) is the practical value, \( n \) is the number of the predicted samples.

**3.2 Tests**

Tests are carried out in accordance with the above method. The parameter \( T \) of similarity threshold is chosen as 0.95 that means the cases with similarity value more than \( T \) are searched as similar cases. The lower limit of similar cases number \( n \) is selected as 2 that means if there are no more than 2 similar cases, than geometrical similarity case retrieve and incremental regression are used.

The root mean square error between calculated values and practical values of oxygen volume is 282.25 cubic meters and with the error limit of 500 cubic meters the accuracy rate is 90.00%. The curves between practical values and calculated values of oxygen volume is shown as Fig.7.
3.3 Comparison with Single Case Retrieve and Reuse

To validate the mixed case retrieve and case reuse strategy, comparison between single case retrieve with KNN and mixed case retrieve methods as mentioned above is employed and the results are shown as Table 1.

Table 1 Comparison between single case retrieve and mixed case retrieve

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>Hit Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>single case retrieve CBR</td>
<td>303.50</td>
<td>88.00%</td>
</tr>
<tr>
<td>mixed case retrieve CBR</td>
<td>282.25</td>
<td>90.00%</td>
</tr>
</tbody>
</table>

The criterion of RMSE of oxygen volume is reduced from 303.50 to 282.25. The criterion of Hit Rate of oxygen volume is enhanced from 88.00% to 90.00%. Results show that mixed case retrieve strategy is effective to improve the BOF oxygen volume control precision on both RMSE and Hit Rate.

4. CONCLUSION

Case based reasoning method based on mixed case retrieve and case reuse can take full advantage of information on successful history furnace in case base to construct the solution of current problem. When KNN case retrieve can not obtain enough similar cases to current case, a proposed geometry similarity calculation algorithm is used to search similar cases on geometry and incremental regression is adopted as the reuse method for these cases. Test results show that the BOF oxygen volume control model built by mixed case retrieve and case reuse CBR is superior to single case retrieve and case reuse CBR. The precision of control can satisfy the request of practical production.

ACKNOWLEDGEMENT

This research is supported by the project (61074096) of the National Nature Science Foundation of China, the project (2006BAB14B05) of the National Key Technology R&D Program of China and the project (2006CB403405) of the National Basic Research Program of China (973 Program). All of these supports are appreciated.

REFERENCES