Research Article

Bin Shao, Hui Liu*, and Fu-gang Chen

Soft sensor method for endpoint carbon content and temperature of BOF based on multi-cluster dynamic adaptive selection ensemble learning

https://doi.org/10.1515/htmp-2022-0287
received April 26, 2023; accepted August 21, 2023

Abstract: The accurate control of the endpoint in converter steelmaking is of great significance and value for energy saving, emission reduction, and steel quality improvement. The key to endpoint control lies in accurately predicting the carbon content and temperature. Converter steelmaking is a dynamic process with a large fluctuation of samples, and traditional ensemble learning methods ignore the differences among the query samples and use all the sub-models to predict. The different performances of each sub-model lead to the performance degradation of ensemble learning. To address this issue, we propose a soft sensor method based on multi-cluster dynamic adaptive selection (MC-DAS) ensemble learning for converter steelmaking endpoint carbon content and temperature prediction. First, to ensure the diversity of the ensemble learning base model, we propose a clustering algorithm with different data partition characteristics to construct a pool of diverse base models. Second, a model adaptive selection strategy is proposed, which involves constructing diverse similarity regions for individual query samples and assessing the model's performance in these regions to identify the most suitable model and weight combination for each respective query sample. Compared with the traditional ensemble learning method, the simulation results of actual converter steelmaking process data show that the prediction accuracy of carbon content within ±0.02% error range reaches 92.8%, and temperature within ±10°C error range reaches 91.6%.

Keywords: BOF, endpoint carbon content and temperature prediction, ensemble learning, soft sensor, dynamic adaptive selection

1 Introduction

Converter steelmaking is a process that utilizes molten iron, scrap steel, and ferroalloy as the primary raw materials. The process involves heating and removing carbon through the physical and chemical reactions between the molten steel and various raw materials in the molten pool. The key aspect of this process is terminal control – ensuring that the carbon content, the temperature of the molten steel, and the content of various metal elements in the molten pool meet the requirements for steel production at the end of the oxygen-blowing stage. Therefore, achieving accurate real-time measurement and control of carbon content and temperature is crucial for improving the accuracy of primary furnace pouring, reducing the consumption of raw materials and energy, lowering production costs, and enhancing product quality [1].

Currently, there are several detection methods for measuring carbon content and temperature during converter blowing. These methods can be categorized into traditional sensor detection methods [2], spectral radiation detection methods [3], and flame image processing and identification detection methods [4], based on their measurement principles. However, the traditional sensor detection methods lack continuity in terminal judgment, have high equipment costs that are not conducive to large-scale implementation, and exhibit high levels of subjectivity due to operator experience. On the other hand, unfavorable working conditions during converter steelmaking make it challenging to obtain accurate spectral radiation and flame image information, leading to significant prediction errors caused by various interfering factors.

*** Corresponding author: Hui Liu,** Faculty of Information Engineering & Automation, Kunming University of Science and Technology, Kunming 650500, China; Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China, e-mail: liuhui621@126.com

Bin Shao: Faculty of Information Engineering & Automation, Kunming University of Science and Technology, Kunming 650500, China; Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China

Fu-gang Chen: Yunnan Kungang Electronic Information Technology Co., Ltd, Kunming 650500, China

In recent years, the soft sensor method has been commonly used to predict data that is difficult to measure in actual industrial processes [5]. The soft sensor method aims to tackle the issue of certain key variables being difficult to obtain in real-time due to harsh environmental conditions and high costs. To overcome these challenges, a mathematical relationship model is established using associated auxiliary variables, which can be used to predict the key variables. Essentially, the soft sensor method provides a means of estimating difficult-to-measure variables by relying on readily measurable parameters. Liu et al. [6] proposed a maximum correntropy criterion-based long-short-term-memory neural network to tackle the issue of industrial process data containing noise and anomalies, which can hinder the construction of accurate soft sensor models. In addition, the superior reliability and practicality of this method have been demonstrated through numerical and industrial examples. Jia et al. [7] built a soft sensor using a graph convolutional network by incorporating the concept of the graph into process modeling. The aim is to obtain localized spatial-temporal correlations that help in understanding the complex interactions among the variables included in the soft sensor. Aiming at the problem that the difference of time series samples between furnaces during steelmaking has a great influence on the model prediction, Zeng and Liu [8] proposed a secondary similarity measure instant learning soft sensing method, which effectively improves the prediction accuracy (PA) of the endpoint carbon content and temperature. At present, the soft sensor methods of the prediction of endpoint carbon content and temperature in converter steelmaking include traditional machine learning methods, just-in-time learning methods, and deep learning methods. Zhou et al. [9] proposed a data-driven hybrid method combining canonical correlation analysis and correlation analysis to predict and control the quality of molten iron in the blast furnace iron-making process by using the nonlinear subspace identification method based on the least square support vector machine. Liu and Zeng [10] established the just-in-time learning soft sensor model by combining the weighted grey correlation degree with the fuzzy C clustering method, which improved the PA of the just-in-time learning method for the endpoint of converter steelmaking. Yuan et al. [11] proposed an adaptive soft sensor modeling approach that combines the time difference model and local weighted partial least squares algorithm to address the issue of global models being inadequate in describing nonlinearity and time-varying process data effectively within the time difference framework. Liu et al. [12] applied the stacked auto-encoders to the problem of terminal prediction in steelmaking, and the accuracy of model prediction is improved by reducing the feature dimension in an unsupervised way. Cui et al. [13] proposed a mixed variational auto-encoder regression model, which enhances the PA of traditional VAE-based soft sensor modeling for complex multimodal industrial processes. However, during actual converter steelmaking production, the long-term use of sensors and varying qualities of raw materials often cause industrial process data to be nonlinear and highly volatile. This poses a significant challenge for traditional single machine learning models to accurately capture the fluctuations in data, which greatly affects the accuracy of endpoint carbon content and temperature predictions.

Considering the above factors, many researchers propose using the ensemble learning method to combine several single models to improve the PA and robustness of the model. Zhang et al. [14] proposed an ensemble model tree method to predict the hot metal temperature of the blast furnace, which proves that the ensemble model has better accuracy and robustness than a single model. Lv et al. [15] pruned the bagging ensemble model by negative correlation learning which improved the accuracy and efficiency of predicting molten steel temperature. Liu et al. [16] proposed a probability-weighted ensemble learning modeling algorithm based on the root mean squared error to predict the quality of molten iron in the blast furnace. Xiong et al. [17] proposed an improved local nearest neighbor peak density clustering weighted ensemble learning soft sensor method, which integrates with the grey relational degree weighted method to improve the PA of endpoint carbon content and temperature. Ahmad and Zhang [18] proposed the use of Bayesian selective combination to combine multiple models, which effectively improves the long-range PA in nonlinear process modeling. They also proposed the use of data fusion techniques to combine multiple neural networks and the combination weights change with the model input data, which can significantly improve model generalization especially in long range predictions [19]. These methods all contain the idea of ensemble learning. As the actual industrial process of converter steelmaking production is a closed dynamic process, the collected industrial process data also exhibit volatility. This volatility presents a challenge for traditional ensemble learning models, which are typically static and may not be able to accurately predict all query samples. Additionally, as the performance of sub-models can differ significantly, using all sub-models to predict every sample can result in an overall ensemble learning method that is unable to accurately predict the production process data.

Based on the above analysis, this study proposes a soft sensor method based on multi-cluster (MC) dynamic adaptive selection (MC-DAS) ensemble learning to predict carbon content and temperature at the endpoint of converter steelmaking. First, a variety of different clustering algorithms are used to cluster the training data, measuring the internal information of the data from different aspects, and forming different sample subsets. Second, a model adaptive selection strategy is proposed, through which a different number of base models are selected adaptively for each query sample as sub-models for subsequent integration. Finally, the local similarity measure was used to obtain the corresponding weight of each selected model. And the prediction results of the endpoint carbon content and temperature of the query sample by the weighted integration fusion were output. The simulation results of the process data of converter steelmaking show that the proposed method can effectively improve the PA of the endpoint carbon content and temperature.

2 Soft sensor method for endpoint carbon content and temperature based on MC-DAS ensemble learning

As converter steelmaking production involves extremely complex physical and chemical reactions and the raw materials added may vary significantly, the production process data can also differ greatly, with each feature having different dimensions of quantity. To eliminate the impact of dimension of quantity on subsequent experiments, the original data are first standardized after obtaining the actual industrial process data. The resulting dataset is denoted as $D' = \{(x_i, y_i); i = 1, 2, 3,...,n\}, (x_i, y_i) \in F^{h+1}, x_i \in F^h, y_i \in F$, where x_i represents the input characteristics of the entire dataset of converter steelmaking production process data, y_i represents the carbon content or temperature of the endpoint output, n is the total number of samples in the dataset, and h is the dimension of the input feature. By standardizing the raw data, the impact of dimensionality between data is removed, ensuring that subsequent experiments can be performed more accurately and effectively. This approach is crucial for achieving precise predictions of endpoint carbon content and temperature in converter steelmaking production.

2.1 Sample distribution of MC algorithm

One of the important factors affecting ensemble learning performance is the performance of a single model and the diversity among models. To achieve diverse base models, the input data of the base model are often modified to obtain different base models. Commonly used methods include bootstrap random distribution, random subspace algorithm, and others. In this study, due to the high volatility of converter steelmaking production process data and the tightness between samples, the similarity between data is measured through clustering methods. Samples with the same degree of similarity are classified into one class, and various clustering algorithms are used to measure the characteristics of data sample sets from different perspectives to generate different sample subsets. This approach ensures the diversity of the base model. Therefore, the study uses four different clustering methods to generate sample subsets, including:

(1) The self-organizing map (SOM) algorithm is a type of unsupervised learning algorithm that combines clustering and high-dimensional visualization. It is a type of neural network that simulates the human brain, where different regions of nerve cells have distinct characteristics that enable the clustering of data. The clustering effect of production process data D' is achieved in a specific manner by identifying the winning neurons s in the corresponding output layer and updating the weights of these neurons.

$$s = \arg\min_{c} \|(x_i, y_i) - W_c\|, \tag{1}$$

where c represents the current neuron, W_c denotes the weight vector, and ε stands for the learning rate, $\|\|$ represents a certain similarity measurement method. In the context of converter steel production process data, the SOM algorithm can map the clustering center of this complex data D' to a surface or plane using neural network techniques, while preserving the topology of the data. This allows for a better understanding and analysis of the data by converting high-dimensional data into a lower-dimensional representation.

(2) The fuzzy C-means (FCM) algorithm is a combination of fuzzy theory and clustering algorithms that produces more flexible clustering results. In the actual industrial process of converter steelmaking production, the samples in the dataset $D^{'}$ cannot be clearly divided into specific types of clusters. If forced to assign them to a specific cluster, the results may appear blunt or even incorrect. To address this issue, the FCM algorithm assigns weights to each data sample $D^{'}$, indicating the degree to which the sample belongs to the cluster. That is, the clustering centers are obtained by minimizing the following objective function:

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{C} u_{ij}^{m} \|(x_i, y_i) - c_j\|^2, 1 \le m < \infty, \tag{2}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{(x_i, y_i) - c_j}{(x_i, y_i) - c_k}\right)^{\frac{2}{m-1}}}, \tag{3}$$

where m represents the number of clusters, n stands for the number of samples, and C denotes the number of cluster centers, and u_{ij} indicates the degree of membership of the sample x_i to the cluster center c_j, which reflects a similarity measure of the data. This allows for better representation of data samples that do not easily fit into a single cluster and improves the accuracy of the clustering results.

(3) The spectral clustering (SPC) algorithm is a clustering algorithm that applies knowledge from graph theory. It regards all samples in the dataset $D^{'}$ as distinct points in space and connects these points with edges to form a graph. The graph is then cut in a way that minimizes the weight value of edges between different sub-graphs and maximizes the weight value of edges within the same sub-graphs, thereby achieving the purpose of clustering. In the context of converter steelmaking production process data, the SPC algorithm uses graph theory to measure the internal information of the data $D^{'}$, without needing to consider its specific structure. The algorithm only needs to consider the similarity matrix of the data and can achieve better clustering results as a result.

(4) The Gaussian mixture model (GMM) clustering algorithm is a probability-based clustering algorithm that describes the distribution of production process data $D^{'}$ using multiple Gaussian distribution models. The algorithm judges the probability of each sample in the dataset $D^{'}$ according to each Gaussian distribution and constantly updates until the model converges.

$$\begin{aligned} \gamma_{j,i} = p(z_j = i|(x_j, y_j)) &= \frac{p(z_j = i)*p((x_j, y_j)|z_j = i)}{p((x_j, y_j))} \\ &= \frac{\alpha_i * p((x_j, y_j)|\mu_i, \Sigma_i)}{\sum_{l=1}^{k} \alpha_l * p((x_j, y_j)|\mu_l, \Sigma_l)}, \end{aligned} \tag{4}$$

where $\gamma_{j,i}(i = 1, 2, 3, \ldots, k)$ denotes the probability of the i-th Gaussian distribution for the j-th sample, while z_j represents the random variable corresponding to a selected distribution such that $z_j \in \{1, 2, 3, \ldots, k\}$. The data samples are then divided into the distribution with the largest probability, achieving the effect of data clustering. In other words, the GMM clustering algorithm models the data using multiple Gaussian distribution models, allowing for flexible and accurate clustering of complex datasets, such as those encountered in converter steelmaking production processes. The algorithm considers the probability of each sample belonging to different distributions, thereby accounting for the uncertainties inherent in such complex datasets. By iteratively refining these probabilities, the GMM algorithm produces accurate and useful clustering results that reflect the underlying patterns and behaviors of the datasets.

In the data of the converter steelmaking process, there may not be a real label indicating which cluster a sample belongs to. As a result, the clustering results are evaluated by calculating the intra-class sample closeness or inter-class sample estrangement. This internal evaluation index is often referred to as the silhouette coefficient (SC), which determines the specific number of clusters for each clustering algorithm.

2.2 Construction of endpoint carbon content and temperature model based on gradient boosting decision tree (GBDT)

Assume the given dataset $D_m' = \{(x_i, y_i); i = 1, 2, 3, \ldots, n_m\}$, $(x_i, y_i) \in F^{h+1}, x_i \in F^h, y_i \in F$, where F represents the entire dataset, x_i represents the input dataset of the model, y_i represents the carbon content or temperature of the endpoint output, n_m is the number of samples in each cluster, and h is the dimension of the input feature.

The GBDT is a popular base model used in the ensemble learning framework for predicting the carbon content and temperature at the endpoint of converter steelmaking. It is an improvement over the boosting decision tree algorithm and uses the approximate method of gradient descent. In GBDT, the negative gradient of the loss function in the current model is used as the approximate value of residuals of the boosting tree algorithm in the regression problem. This allows the algorithm to fit a regression tree that can better handle the carbon content and temperature prediction problem in converter steelmaking. The decision trees used in GBDT are classification and regression trees (CART), which are established based on the previous tree in an iterative manner, resulting in a powerful ensemble learning model. Compared to other regression models, GBDT requires relatively less parameter adjustment time while still achieving high accuracy in predicting the endpoint parameters. Additionally, GBDT is robust to outliers, which is important given the complex and dynamic nature of the converter steelmaking production process. Overall, the GBDT algorithm is an effective base model for the ensemble learning framework used in converter steelmaking, allowing for accurate and efficient prediction of important endpoint parameters.

The process of building a GBDT base model starts with creating a subset of samples from the original data set D', which is formed through multiple clustering. This subset, denoted as D'_m, is then used to initialize the GBDT model.

$$F_0(x, \omega) = \underset{c}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, c), \tag{5}$$

where F represents the GBDT model, F_0 is the initial state of the model, N is the number of training samples, y_i is the initialized true value, c is the predicted value of initialization, $L(y_i, c)$ is the loss function of the model, and we want to find a minimum c to minimize the error of all samples. The loss function commonly used in the GBDT regression algorithm is the least squares regression.

Once the initialization is complete using the subset of clustered data, the GBDT algorithm proceeds to build m classification regression trees, where $m = 1, 2, 3, \ldots, M$. The building process involves performing the following circular steps in turn:

(1) The first step in building a GBDT model is to calculate the response value of the $m-$ th tree for each training sample in the subset D'_m. For a given tree m and sample i (where $i = 1, 2, 3, \ldots, n_m$) in the dataset D'_m, the response value can be calculated as follows:

$$r_{m,i} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x)}\right]_{F(x)=F_{m-1}(x)}. \tag{6}$$

(2) After calculating the response values for the m-th tree, the next step in the GBDT algorithm is to use the CART method to fit the residual of the current ensemble model to obtain the m-th regression tree. Once the regression tree is built using CART, the corresponding leaf node is $R_{m,j}$, where m represents the m-th tree, j (where$j = 1, 2, 3, \ldots, J$) is the number of leaf nodes of the tree. And the best fitting value for each leaf node is calculated:

$$c_{m,j} = \underset{c}{\operatorname{argmin}} \sum_{x_i \epsilon R_{m,j}} L(y_i, F_{m-1}(x_i) + c). \tag{7}$$

(3) After building the m-th regression tree and calculating the best-fitting values for each leaf node, the GBDT algorithm updates the model and uses m trees for prediction in the next iteration.

$$F_m(x) = F_{m-1}(x) + \delta \sum_{i=1}^{J} c_{m,j} I(x \epsilon R_{m,j}), \tag{8}$$

where δ is the learning rate, which is used to control the promotion degree of each boosting tree to the GBDT model.

Therefore, the GBDT model expression of the final M classification regression tree is

$$F_M(x) = F_0(x) + \delta \sum_{m=1}^{M} \sum_{i=1}^{J} c_{m,j} I(x \epsilon R_{m,j}), \tag{9}$$

where the former Σ represents the accumulation of M boosting trees, and the latter Σ is the summation of the best-fitting values of all leaf nodes of each boosting tree. That is, to obtain the final GBDT regression model, the initial regression $F_0(x)$ is taken, to add up the best-fitting values of all M tree leaf nodes continuously. So, the GBDT model establishment process is adding a new CART regression tree aimed at the sample subset D'_m to fit the residual between the predicted values c and the true value y_i of all the previous trees, whose predicted result is the sum of the predicted values of all the established trees.

2.3 Dynamic adaptive selection and fusion strategy of endpoint carbon content and temperature prediction sub-mode

The process of establishing each sub-model in GBDT involves constructing a hypothesis space based on the production process data of converter steelmaking. This hypothesis space represents a set of possible relationships between the predictor variables and target variables in the data. Once the hypothesis space is established, the parameters of each sub-model are optimized using the information contained in each sample subset. This optimization process involves finding the model spatial parameters that best match the sample subset corresponding to each model.

The principle of the method in this section is shown in Figure 1.

Another factor contributing to the improvement of ensemble learning performance is the integration and fusion strategy of the base model. However, not all base models in the pool would be suitable for all query samples, and it is necessary to choose the most appropriate model for each query sample to achieve optimal results. Therefore, diverse base models are built based on various clustering approaches to generate different subsets of training samples, and a base model pool M, $M = M_1, M_2, M_3, \ldots, M_m$, (where m represents the number of base models in the base model pool) is formed. To select the best sub-model corresponding to each query sample and carry out weighted fusion, this study adopts a strategy of dynamic selection and weighted fusion. First, for a new query sample x_q, a similar sample region is found in the total training sample set x_p by the standardized Euclidean distance. Then, the performance of each base model in the pool is evaluated in this similar region, and a certain number

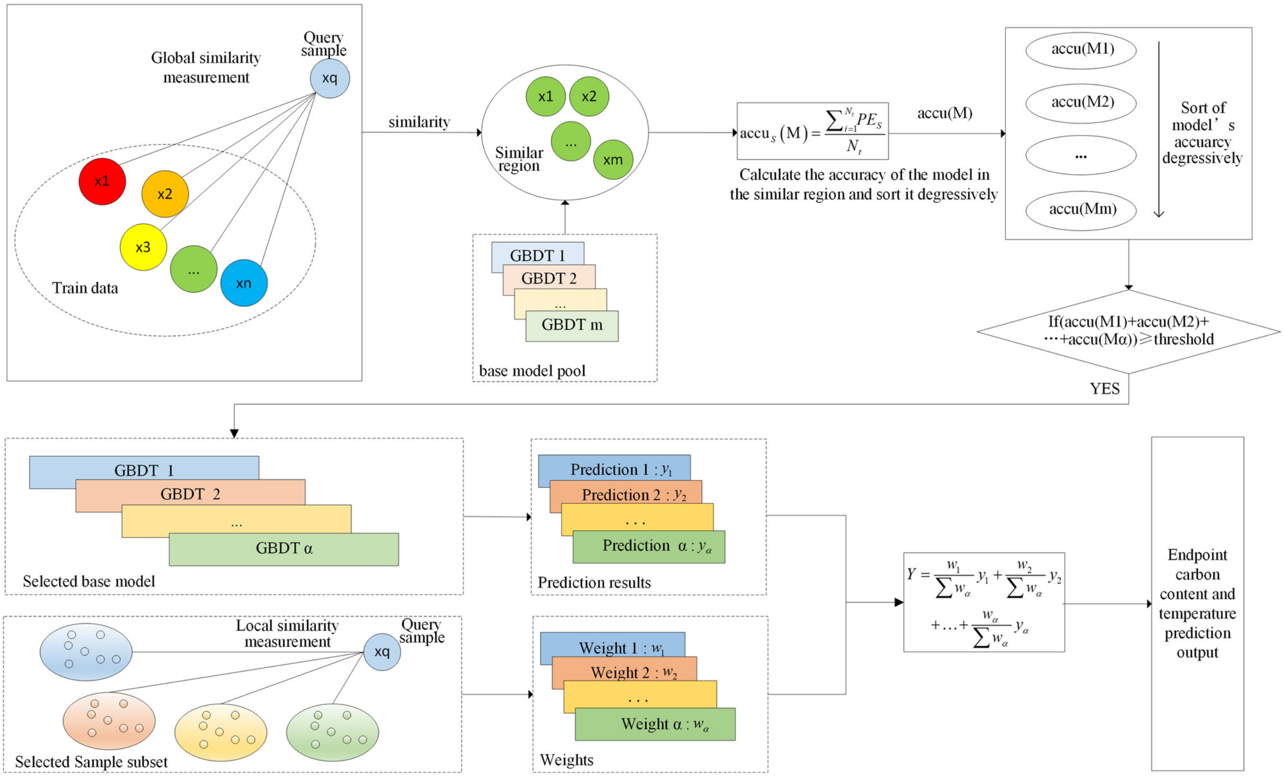

Global similarity measurement
Query sample
xq
x1
x2
x3
...
xn
Train data
similarity
x1
x2
...
xm
Similar region
GBDT 1
GBDT 2
...
GBDT m
base model pool
$\mathrm{accu}_s(\mathrm{M}) = \frac{\sum_{i=1}^{N_t} PE_s}{N_t}$
Calculate the accuracy of the model in the similar region and sort it degressively
accu(M)
accu(M1)
accu(M2)
...
accu(Mm)
Sort of model's accuarcy degressively
If(accu(M1)+accu(M2)+⋯+accu(Mα))⩾threshold
YES
GBDT 1
GBDT 2
...
GBDT α
Selected base model
Local similarity measurement
Query sample
xq
Selected Sample subset
Prediction 1 : y_1
Prediction 2 : y_2
. . .
Prediction α : y_α
Prediction results
Weight 1 : w_1
Weight 2 : w_2
. . .
Weight α : w_α
Weights
$Y = \frac{w_1}{\sum w_\alpha} y_1 + \frac{w_2}{\sum w_\alpha} y_2 + \ldots + \frac{w_\alpha}{\sum w_\alpha} y_\alpha$
Endpoint carbon content and temperature prediction output

Figure 1: Dynamic adaptive selection and fusion strategy.

of models are selected adaptively according to the query sample.

The specific methods are as follows: The evaluation index of each base model is based on its accuracy within a similar region that corresponds to the query sample. Specifically, the accuracy of each model is ranked from largest to smallest, and if the sum of the PA of the top $\alpha(\alpha \in m)$ models is greater than the set threshold value, then the top α models are selected as the sub-model involved in subsequent integration and fusion. The formula for model selection (MS) is shown in equation (10). This formula takes into account the PA of each model, the number of selected models, and the threshold value in the selection process.

$$M_\alpha = \begin{Bmatrix} 1, \mathrm{accu}(M_1) + \mathrm{accu}(M_2) + \ldots + \mathrm{accu}(M_\alpha) \\ \geq \mathrm{threshold} \\ 0, \mathrm{accu}(M_1) + \mathrm{accu}(M_2) + \ldots + \mathrm{accu}(M_\alpha) \\ < \mathrm{threshold} \end{Bmatrix}, \quad (10)$$

where accu(M) is the PA of the base model within the base model pool in the similar region, M_α indicates whether the top α base models in the base model pools are selected as a sub-model for subsequent ensemble learning, and if so, it indicates the sum of PA of the top α models and is recorded as 1 if it is greater than the threshold value, if not, 0.

This method measures the prediction ability of the model by obtaining different similar regions for each query sample x_q. By adaptively selecting the number of models, better prediction results can be achieved, which improves the overall prediction performance of the ensemble learning framework for endpoint carbon content and temperature in converter steelmaking. This approach is more consistent with the actual dynamic process of industrial operation. For a given query sample x_q, the output of the endpoint carbon content and temperature prediction on the GBDT base model is

$$y_{\mathrm{pre},\alpha} = F_\alpha(x_q, \omega), \quad (11)$$

where $\mathrm{F}_\alpha(\bullet)$ represents the functional relationship of the sub-model, and ω is the parameter of the model.

After the selection process, the base models that are retained are denoted as M_{new} (where $M_{\mathrm{new}} = [M_1, M_2, \ldots, M\alpha]$, α є $[1, m]$), and the corresponding sample subset of data is D_{new} (where $D_{\mathrm{new}} = \{(x_{1,i}, y_{1,i}), (x_{2,i}, y_{2,i}), \ldots, (x_{\alpha,i}, y_{\alpha,i})\}; \alpha$ є $[1, m]$, $i = 1, 2, 3, \ldots, n_m$). For a given query sample x_q, the distance to the samples in each sample subset is measured using the standard Euclidean distance (SED).

$$\mathrm{SED} = \sqrt{\sum_{i=1}^{n} \frac{(u_i - v_i)^2}{V[x_i]}}, \quad (12)$$

where u and v represent two n-dimensional feature vectors in converter steelmaking data, V is the variance of the two feature vectors. If the reciprocal of this variance is regarded as a weight, then this formula can be regarded as a weighted Euclidean distance.

Based on these distances, the mean value is then calculated to represent the similarity between the query sample and the corresponding subset of samples. This mean value is used as the corresponding weight w_α of each base model $M\alpha$.

$$w_\alpha = \mathrm{SED}(x_q, D_{\mathrm{new}}) = \frac{\sum_1^{n_m}\mathrm{SED}(x_q, (x_{\alpha,i}, y_{\alpha,i}))}{n_m}, \qquad (13)$$

where $\mathrm{SED}(x_q, D_{\mathrm{new}})$ represents the SED between the query sample and the sample subset.

The final predicted output for the endpoint carbon content and temperature, obtained by weighted fusion for all query samples x_q, is expressed as follows:

$$Y_{\mathrm{prediction}} = \frac{w_1}{\sum w_\alpha}y_{\mathrm{pre},1} + \frac{w_2}{\sum w_\alpha}y_{\mathrm{pre},2} + \ldots + \frac{w_\alpha}{\sum w_\alpha}y_{\mathrm{pre},\alpha}. \quad (14)$$

Although the SED is the same in this article, the two places of V used in this method are equivalent to measuring the similarity between the query sample and the global sample D', as well as the local sample D'_m, respectively. In the process of dynamic prediction of query samples in this method, the information of the global sample and the local sample is also taken into account. The pseudocode of the algorithm proposed in this section is as follows:

```
Algorithm: Dynamic adaptive selection and fusion strategy
Input: historical data $D'$, sample subset $D'_m$, base model pool $M$, query sample $x_q$
Output: the prediction results of endpoint carbon content and temperature of each query sample $Y_{\mathrm{prediction}}$
for $i = 1$ to $n$ do:
    SED $\leftarrow \sqrt{\sum_{i=1}^{n}\frac{(x_q - i)^2}{V[x_i]}}$
end for
Get $x_q$ with $D'$ global similarity GlobalSED
Sort GlobalSED from smallest to largest, taking the first 1,500 as similar region S
for $M = 1$ to $m$ do
    $\mathrm{accu}_S(M) \leftarrow \frac{\sum_{i=1}^{N_t}\mathrm{PE}_S}{N_t}$
end for
Get the PA $\mathrm{accu}_S(M)$ of the base model pool $M$ on $S$
Sort $\mathrm{accu}_S$(M) from largest to smallest
$M_\alpha \leftarrow \begin{cases} 1, \mathrm{accu}(M_1) + \mathrm{accu}(M_2) + \ldots + \mathrm{accu}(M_\alpha) \geq \text{threshold} \\ 0, \mathrm{accu}(M_1) + \mathrm{accu}(M_2) + \ldots + \mathrm{accu}(M_\alpha) < \text{threshold} \end{cases}$
After obtaining the $\alpha$ selected models, then get its prediction results to $x_q$ $y_{\mathrm{pre},\alpha} \leftarrow F_\alpha(x_q, \omega)$
$D'_{m\alpha}$←Choose the sample subset corresponding to the selected model $M_\alpha$ from $D'_m$
for $i = 1$ to $n_m$ do:
    SED $\leftarrow \sqrt{\sum_{i=1}^{n}\frac{(x_q - i)^2}{V[x_i]}}$
end for
Get $x_q$ with $D'_{m\alpha}$ Local similarity LocalSED
Obtain the corresponding model weight $w_\alpha \leftarrow \frac{\sum_1^{n_m}\mathrm{LocalSED}}{n_m}$
Output the prediction results of $x_q$ $Y_{\mathrm{prediction}} \leftarrow \frac{w_1}{\sum w_\alpha}y_{\mathrm{pre},1} + \frac{w_2}{\sum w_\alpha}y_{\mathrm{pre},2} + \ldots + \frac{w_\alpha}{\sum w_\alpha}y_{\mathrm{pre},\alpha}$
```

3 Soft sensor modeling of endpoint carbon content and temperature based on MC-DAS ensemble learning for basic oxygen furnace (BOF) process data

The production process of converter steelmaking involves large fluctuations in data, and a global single model is limited in its ability to effectively predict the production process data. Additionally, due to the closed and dynamic nature of the converter steelmaking process, traditional ensemble learning methods cannot effectively adjust for actual dynamic conditions, as they use the same base model to predict all samples. To address these issues and make predictions more consistent with the actual dynamic industrial process, a new method based on MC dynamic adaptive selection ensemble learning has been proposed for predicting the process data of converter steelmaking. This method improves PA by dynamically selecting a group of appropriate models to make predictions and adaptively adjust the weights of each model to enhance the overall prediction performance.

The following steps outline the endpoint carbon content and temperature soft sensor model based on MC-DAS ensemble learning, with the assumption that the collected production process data sample set after standardization is represented as $D' = \{(x_i, y_i); i = 1, 2, 3, \ldots, n\}$, where $(x_i, y_i) \in F^{h+1}, x_i \in F^h, y_i \in F$. Here x_i represents the input features of the entire dataset of converter steelmaking production process data, y_i represents the carbon content or

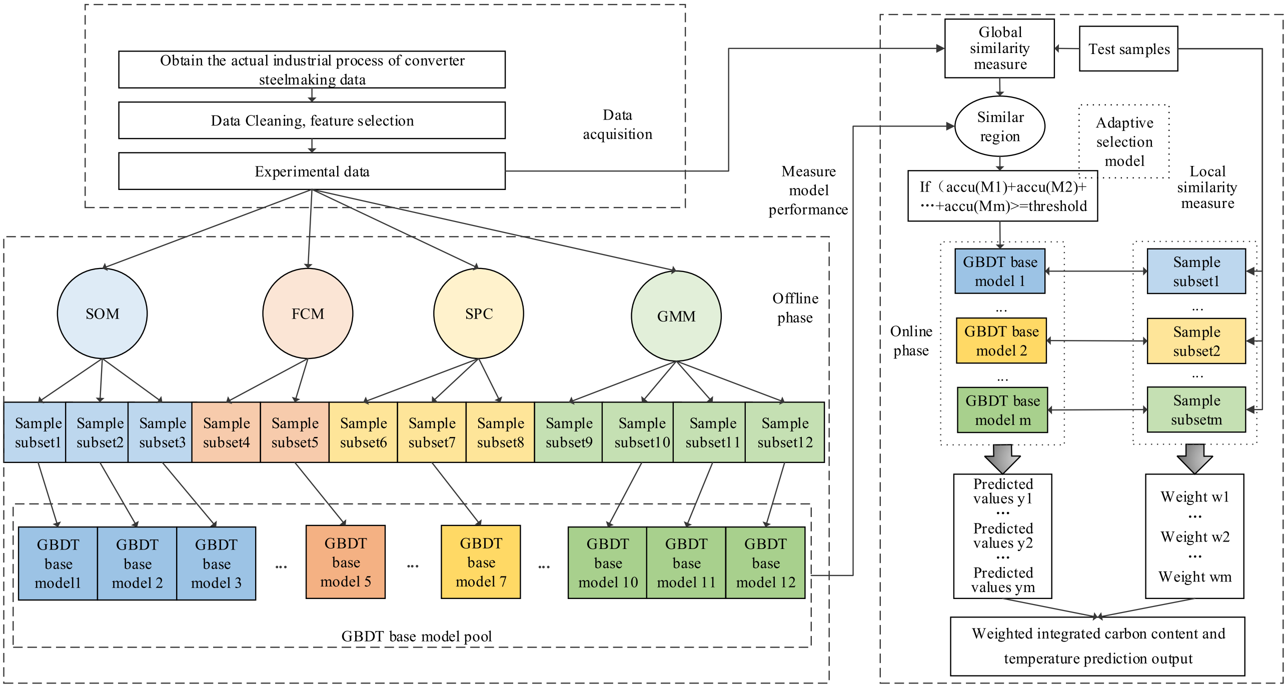


Figure 2: Soft sensor modeling flow chart of converter steelmaking process data endpoint carbon content and temperature based on MC-DAS ensemble learning.

temperature of the endpoint output, n is the total number of samples in the dataset, and h is the dimension of the input feature.

Step 1: In order to prepare the dataset for the experiment, the following operations are performed on the dataset D': cleaning, feature selection, and other relevant operations. The resulting sample dataset is then divided into two parts: training data D_1' and test data D_2';

Step 2: In the offline modeling section, the dataset D_1' is clustered using four different clustering algorithms: SOM, FCM, SPC, and GMM. These algorithms help to capture different aspects of the dataset and form various sample subsets D_m';

Step 3: After obtaining the different sample subsets D_m' through clustering, the next step is to use these subsets to build a GBDT base model for each subset. The resulting base models form a diverse base model pool M, where $M = M_1, M_2, M_3,\ldots, M_m$, and m represents the number of base models in the base model pool;

Step 4: In the online prediction section, for the query samples x_q of the test data D_2', use formula (15) to get some similar samples from D_1' which is similar to x_q as similar region S;

$$\text{GlobalSM} = \text{SED}(x_q, D_1'). \tag{15}$$

Step 5: To evaluate the performance of the base models in the base model pool M, the MC-DAS ensemble learning model selects $\alpha(\alpha \in m)$ base models adaptively as sub-models for subsequent integration based on the query sample x_q and the similar region S to which it belongs;

Step 6: According to the one-to-one correspondence between the selected model and the sample subset, measure the similarity between the query sample and the selected subset through the local similarity measure as the weight of the corresponding model w_α;

Step 7: According to the selected model and its weight, obtain the prediction results of carbon content and temperature of the query sample x_q by weighting fusion;

Step 8: Repeat Steps 4–7 to simulate the actual dynamic condition of converter steelmaking and obtain the prediction results of the endpoint carbon content and temperature of test data D_2'.

The soft sensor modeling process of endpoint carbon content and temperature of converter steelmaking process data based on MC-DAS ensemble learning is shown in Figure 2.

4 Experimental results and discussion

4.1 Experimental data and platform

This article uses Python for simulation experiments. The experimental data in this study are from the actual production process data of converter steelmaking in steel mills. In the process of converter steelmaking, accurate prediction of carbon content and temperature at the endpoint of steelmaking is the key to the control technology of the endpoint. In the converter steel production process data obtained by the sensor, there are metal materials and loading amounts, such as loaded iron water, loaded pig iron, loaded scrap metal materials, etc., as well as nonmetallic materials and loading amounts, including silicon manganese amount, slag dose, etc. Additionally, oxygen blowing time, average oxygen pressure, average oxygen pressure position, and gun position are also recorded, resulting in a total of 126-dimensional features. To establish the soft sensor model, a feature selection method was used to select six key features that affect the carbon content and temperature during tapping from the 126-dimensional feature data. These six selected features were used as auxiliary variables in the soft sensor modeling process, with the endpoint carbon content and temperature serving as the dominant variables. Table 1 shows the main variables and auxiliary variables selected for soft sensor modeling.

In this study, based on the abovementioned 6 dimensional features, 5,500 samples were collected from the actual industrial process of converter steelmaking, of which 5,000 samples were used randomly for training and 500 samples were used randomly for testing.

To evaluate the predictive performance of different model approaches, this study uses PA, root mean square error (RMSE), and mean absolute percentage error (MAPE). PA measures the accuracy of the prediction, with higher values indicating greater accuracy. RMSE measures the deviation between the predicted values and the actual values, with lower values indicating better performance. MAPE measures the percentage of error in the prediction, with lower values indicating better performance. The formulas for these evaluation metrics are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_t}\sum_{i}^{N_t}(y_{\mathrm{test},i} - y_{\mathrm{prediction},i})^2}, \quad i = 1, 2, \ldots, N_t, \tag{16}$$

$$\mathrm{MAPE} = \frac{1}{N_t}\sum_{i}^{N_t}\left|\frac{y_{\mathrm{test},i} - y_{\mathrm{prediction},i}}{y_{\mathrm{test},i}}\right| \times 100\%, \tag{17}$$

$i = 1, 2, \ldots, N_t$,

$$\mathrm{PE}_i = \begin{cases} 1, & |y_{\mathrm{test},i} - y_{\mathrm{prediction},i}| \le \mathrm{Th} \\ 0, & |y_{\mathrm{test},i} - y_{\mathrm{prediction},i}| > \mathrm{Th} \end{cases}, \quad i = 1, 2, \ldots, N_t, \tag{18}$$

$$\mathrm{PA} = \frac{\sum_{i=1}^{N_t}\mathrm{PE}_i}{N_t}, \tag{19}$$

where N_t represents the number of query samples, y_{test} represents the query samples, $y_{\mathrm{prediction}}$ represents the predicted value of the query sample, PE indicates whether the predicted accuracy of the endpoint carbon content and temperature is within the error range (Th), and if so, it indicates a hit and is recorded as 1, if not, 0. The prediction error range of carbon content is Th = 0.02%, and the prediction error range of temperature is Th = 10°C. The final PA is a measure of the proportion of the total query sample in the number of hits within the error range.

4.2 Comparison and analysis of the performance of this method

Based on the modeling flow described in Section 3, a soft sensor model was developed to predict the endpoint carbon content and temperature of BOF steelmaking for simulation experiments. In this section, the feasibility and innovation of this method are evaluated through a series of independent experiments.

4.2.1 Influence of the size of the similar region and the threshold of MS on the experiment

In this method, the performance of the model is evaluated using a similar region to determine the selection process for subsequent models. The threshold value for MS also

Table 1: Main variables and auxiliary variables of soft sensor modeling

Main variables	Auxiliary variables	Main variables	Auxiliary variables
Endpoint carbon content of molten steel (%)	Oxygen pressure 31	Endpoint temperature of molten steel (°C)	Amount of pig iron loaded
	Time of mixing iron		Molten iron P
	End of iron infusion to oxygen time		Gun position 22
	Temperature of molten iron		Oxygen pressure 29
	Start mixing iron with tapping time		Oxygen pressure 11
	Loading scrap quantity		Gun position 16

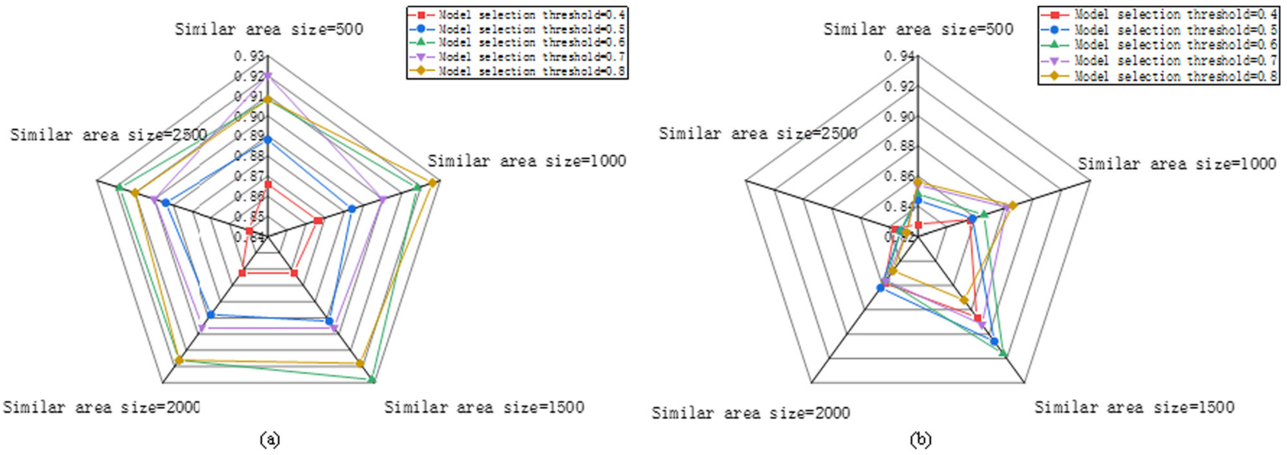


Figure 3: The effect of similar region size and MS threshold on (a) carbon content and (b) temperature accuracy.

determines the number of remaining models after selection and has a significant impact on the final experimental results. This section aims to determine the optimal PA and minimum RMSE by varying the size of the similar region and the threshold for MS. The size of the similar region varied from 500 to 2,500 with step intervals of 500, resulting in a total of five variables. The threshold for MS varied from 0.4 to 0.8, also resulting in a total of five variables. Figure 3 shows the effect of these variables on the PA of the whole algorithm framework. The figure shows that the optimal prediction accuracy and minimum RMSE were achieved when the endpoint carbon content and temperature prediction index were set to 1,500 in the similar region size and 0.6 in the MS threshold.

4.2.2 Determination of the number of clusters by multiple clustering methods

To obtain a diverse sample subset, this method determines the sample subset using several clustering algorithms based on their internal evaluation index – SC. The SC measures whether the current sample point is close enough to other sample points of the current class and far enough from the nearest other class, to obtain the SC of the sample for the dataset D'. For the sample subset generated by the dataset D', the SC of the subset is obtained using the mean value. Its range is between −1 and 1; the closer the samples in the same class, the farther the samples in different classes, the higher the score, the better the clustering result. The formula is as follows:

$$\mathrm{SC} = \frac{b - a}{\max(a, b)}, \tag{20}$$

where a represents the average distance between the sample and all other points in the same cluster, and b represents the average distance between the sample and all other points in the next closest cluster. The larger the SC is, the closer the samples belonging to the same class are and the farther the samples belonging to different classes are.

In the experiment, the contour coefficients for the number of clusters were calculated in the interval [2,9] with a step size of 1, and the final clustering number of each corresponding clustering algorithm was determined accordingly. Figures 4 and 5 show the results of using SOM, FCM, SPC, and GMM to evaluate the changes in indexes under different numbers of clusters. The SCs of the four clustering algorithms were found to be the best when the number of clusters is 3, 2, 3, and 4, respectively, for predicting the endpoint carbon content. For temperature prediction, the SCs of the four clustering algorithms were found to be the best when the number of clusters is 2, 2, 4, and 2, respectively. In addition, the corresponding clustering algorithms with the best clustering effect were identified.

4.2.3 Effects of different parts in this method

This section aims to validate the advantages of the MC algorithm in obtaining different subsets of samples by comparing it with the bootstrap random distribution method for partitioning data. In addition, the framework of dynamic adaptive selection ensemble learning consists of two structures: MS and Weight Fusion (WF). Thus, we compare the different structures of Bootstrap random distribution and MC data distribution to verify the advantages and effective-

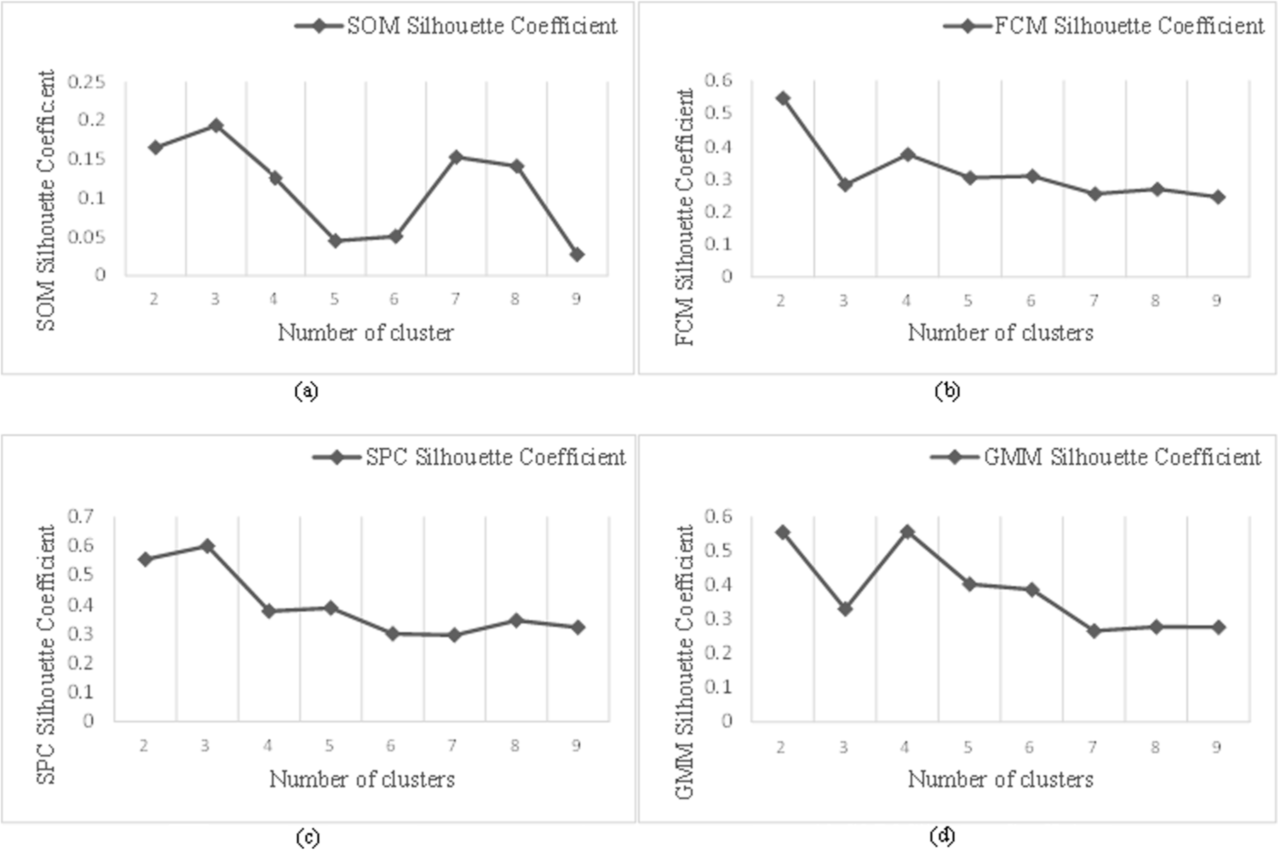


Figure 4: The determination of cluster number of carbon content under different clustering methods: (a) SOM SC, (b) FCM SC, (c) SPC SC, and (d) GMM SC.

ness of the dynamic adaptive selection ensemble learning framework. Figures 6 and 7 present the prediction results of carbon content and temperature, respectively. Table 2 provides the prediction performance evaluation indexes of carbon content and temperature for different structure combinations in the framework of dynamic adaptive selection ensemble learning. According to the analysis of Table 2 and Figures 6 and 7, it can be seen that:

(1) Compared with the traditional bootstrap random distribution, the PA of carbon content and temperature is improved by 8.6 and 13% respectively, and RMSE and MAPE are also reduced.
(2) Further, it also shows that when the data distribution method is bootstrap random distribution, due to the randomness of data distribution, the PAs of the endpoint carbon content and temperature of the overall ensemble learning model are changing, and the PAs of the endpoint carbon content and temperature are not stable enough. When the data distribution method is a variety of clustering methods, it measures the internal information of the data from different angles and forms a diverse sub-model, which is better than the sub-model obtained by randomly distributing the data. It is fully verified that a good data distribution method helps to improve the prediction performance of ensemble learning as a whole.
(3) Regardless of whether the current data distribution method is bootstrap random distribution or MC data distribution, the ensemble learning framework combined with MS has a significant improvement in the PA of carbon content and temperature compared with the simple weighted fusion strategy. It shows that by selecting some models with better performance related to the current query sample as sub-models for subsequent integration, the PA of the ensemble learning as a whole can be improved, and the superiority and necessity of the MS strategy are verified.
(4) After combining the operation of dynamic adaptive selection and weighted fusion, compared with the single MS strategy and the single weighted fusion strategy, the fused overall has a certain improvement in carbon content and temperature prediction, which verifies the superiority of the current method and is more in line with the actual dynamic industrial process.

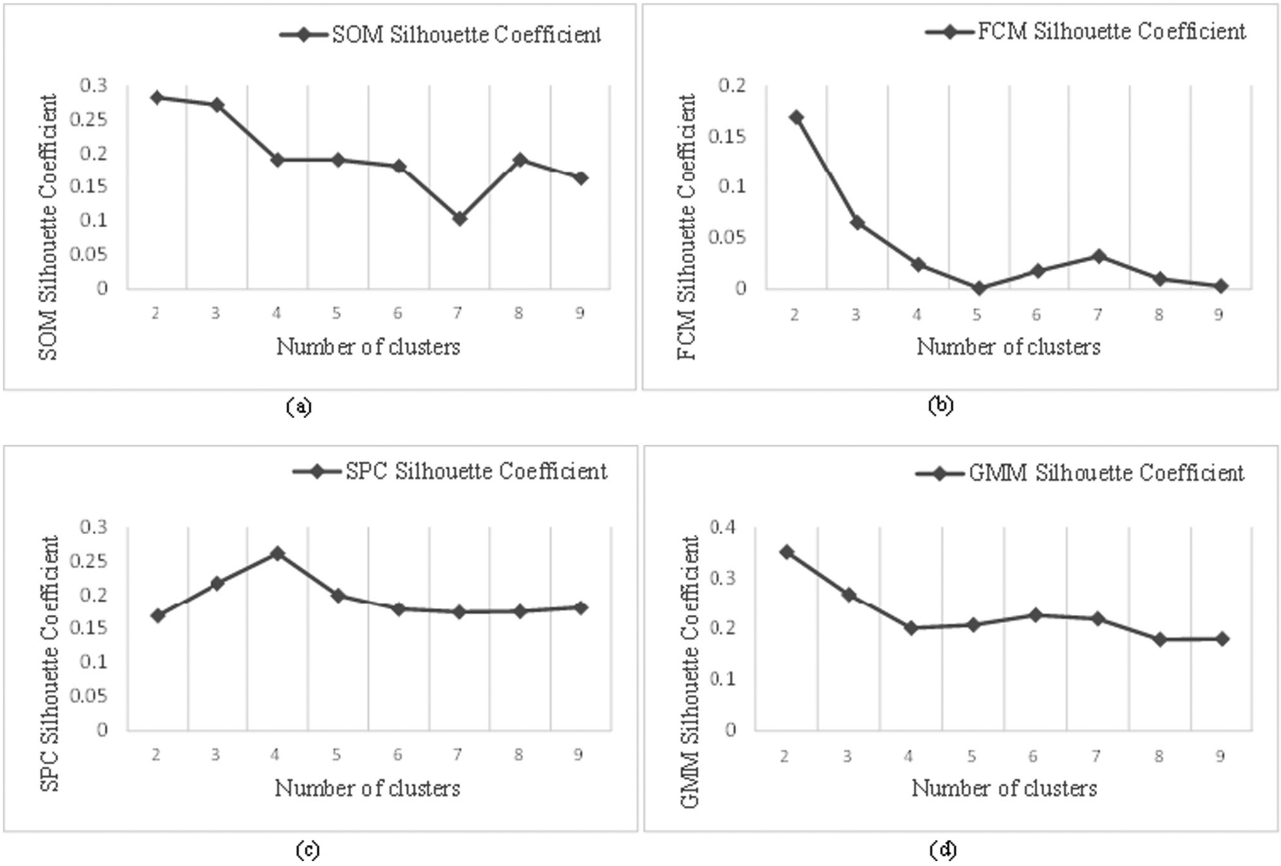


Figure 5: The determination of the cluster number of temperature under different clustering methods: (a) SOM SC, (b) FCM SC, (c) SPC SC, and (d) GMM SC.

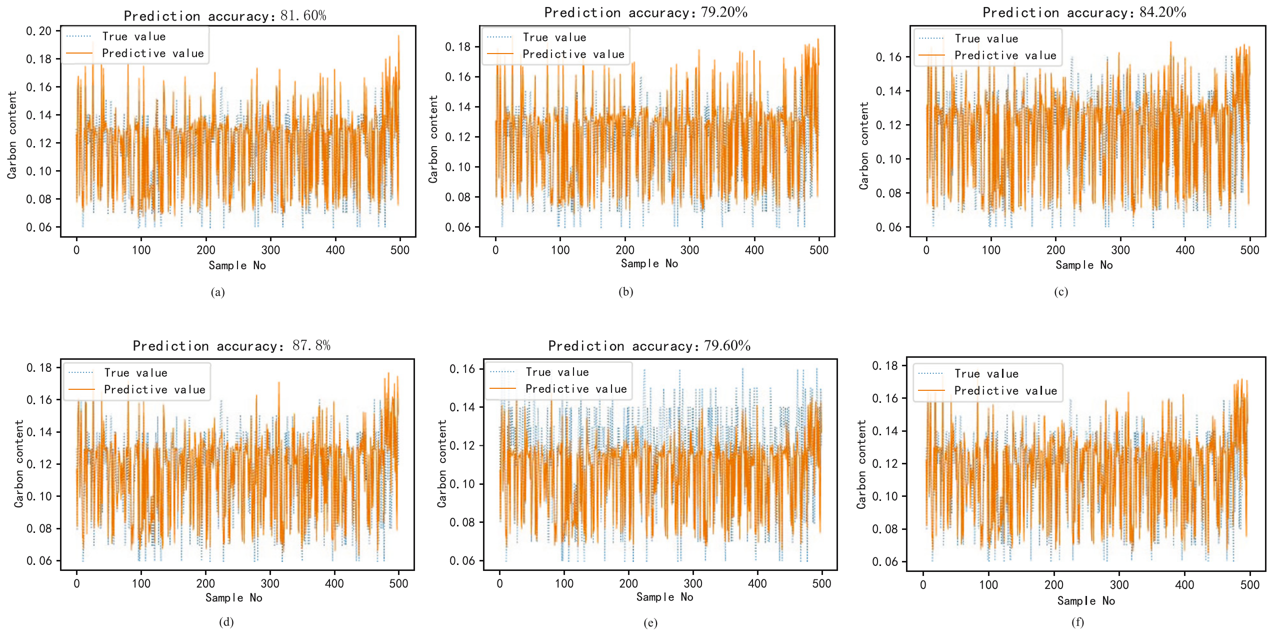


Figure 6: Performance comparison prediction results of this method under carbon content (Th = 0.02%). (a) Bootstrap + MS carbon content prediction results, (b) Bootstrap + WF carbon content prediction results, (c) Bootstrap + DAS carbon content prediction results, (d) MC + MS carbon content prediction results, (e) MC + WF carbon content prediction results, and (f) MC-DAS carbon content prediction results.

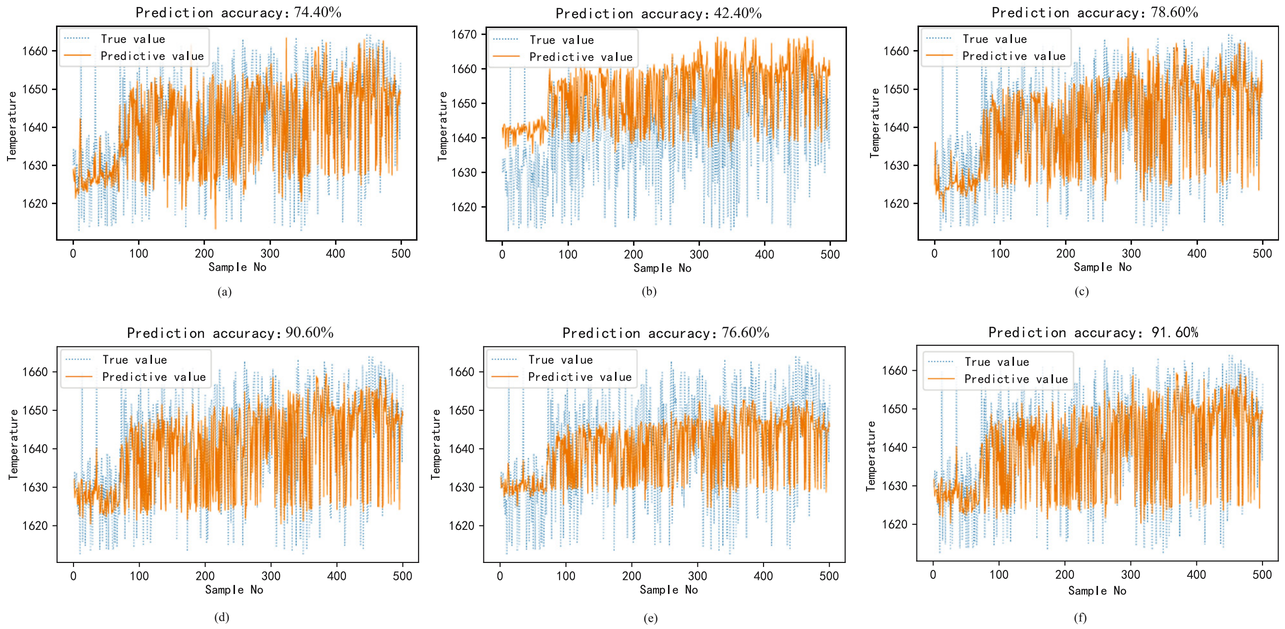


Figure 7: Performance comparison prediction results of this method at temperature (Th = 10°C). (a) Bootstrap + MS temperature prediction results, (b) Bootstrap + WF temperature prediction results, (c) Bootstrap + DAS temperature prediction results, (d) MC + MS temperature prediction results, (e) MC + WF temperature prediction results, and (f) MC-DAS temperature prediction results.

4.3 Comparison and analysis with other ensemble learning methods

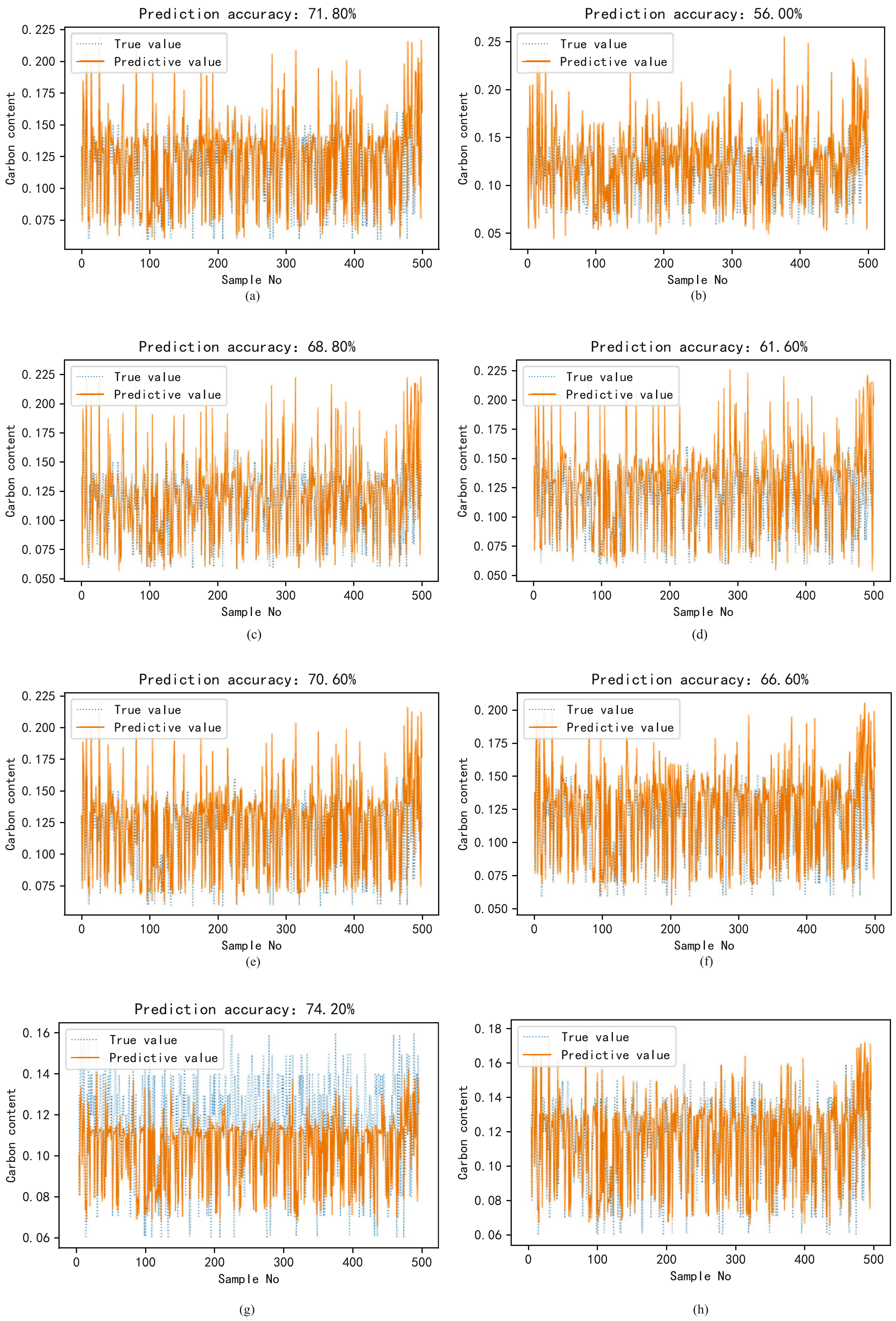
This section aims to compare the proposed dynamic ensemble learning method with many traditional ensemble learning methods to demonstrate its robustness and applicability. The comparison methods include GBDT global single model, eX-treme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Random Forest (RF), Bagging, and Adaboost. To verify the advantages of the framework of dynamic ensemble learning, we compare the static model (MC data distribution and Simple Average (SA) ensemble strategy) with the dynamic model of ensemble learning. Figures 8 and 9 present the prediction results of carbon content and temperature under different ensemble learning algorithms, respectively. Table 3 shows the performance evaluation indexes of different algorithms for predicting carbon content and temperature. From the analysis of Table 3 and Figures 8 and 9, it can be seen that

(1) Compared with the traditional ensemble learning method, the proposed method has better performance in the industrial process data of carbon content and temperature prediction at the end of converter steelmaking. The method in this chapter can generate a subset of samples with diversity through a variety of clustering methods and then construct a sub-model with diversity, so as to ensure the generalization ability of the ensemble learning model, and improve the robustness and PA of the model.

(2) Compared with the static ensemble learning model based on multi-clustering combined with average ensemble for the prediction of carbon content and temperature at the endpoint of converter steelmaking, dynamic ensemble learning can be more in line with the actual industrial operation conditions, and then obtain more accurate prediction results.

(3) In the process of model adaptive selection, through the different query samples, the model that is more consistent with the base model pool is selected for subsequent experiments, and the process of adaptive selection is realized. The model with poor performance on the query sample is selected, so that the constructed dynamic ensemble learning framework has a better model combination for the current query sample. Therefore, the method in this study performs better in the prediction of carbon content and temperature. At the same time, it can be seen from RMSE and MAPE that the method in this study performs better and more reasonable than other methods.

The simulation results demonstrate that the proposed method is capable of meeting the real-time requirements of converter steelmaking. Specifically, the prediction time for 500 query samples is 158.3 s, which translates to an average of 316.6 ms per query sample. In comparison, the subroutine detection method described in the literature [20] takes approximately 1 min per time (without "pouring"



(a)

(b)

(c)

(d)

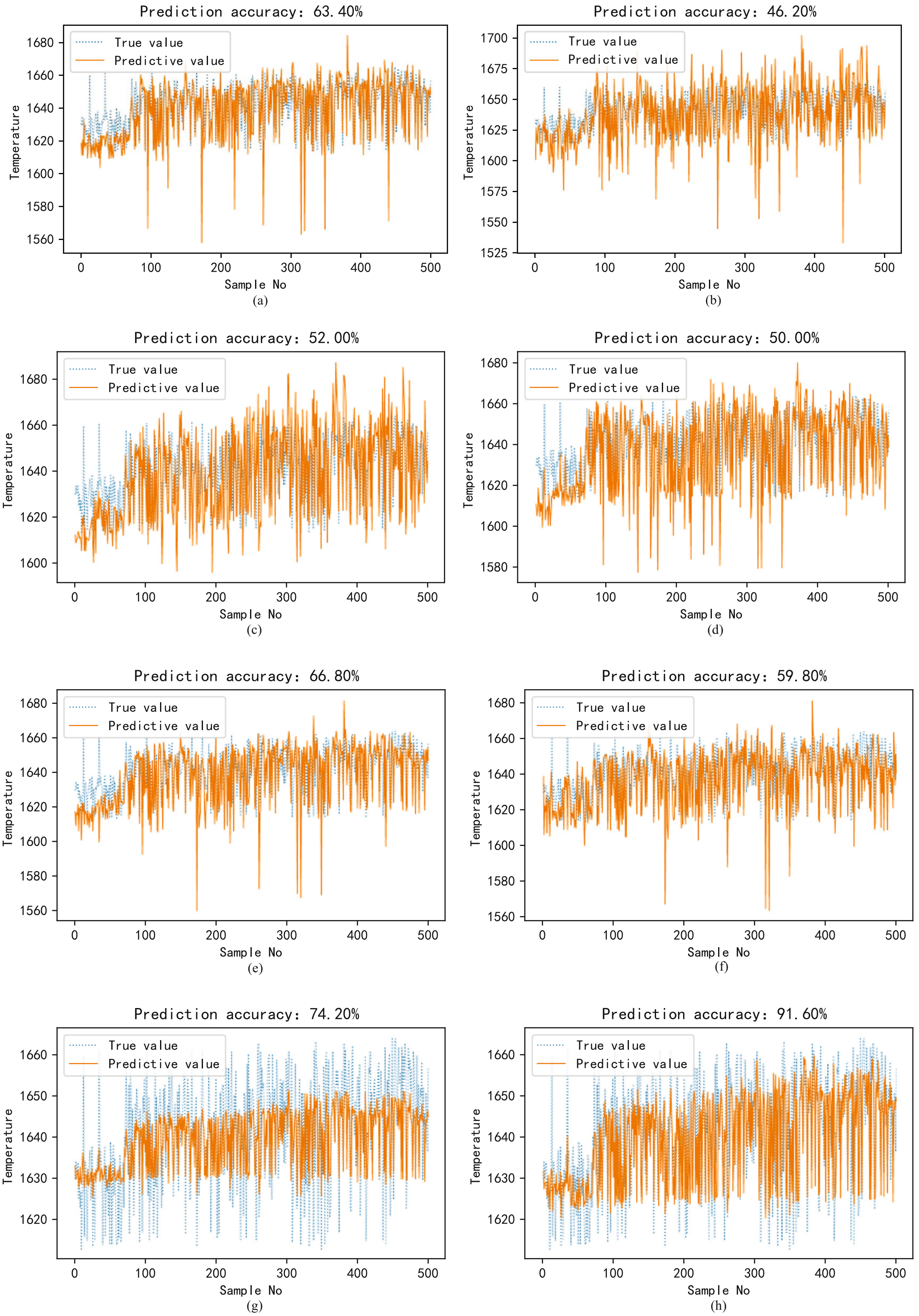
(e)

(f)

(g)

(h)

Figure 8: Prediction results of carbon content (Th = 0.02%) under different methods. (a) GBDT carbon content prediction results of global modeling, (b) XGBoost carbon content prediction results of global modeling, (c) LightGBM carbon content prediction results of global modeling, (d) RF carbon content prediction results of global modeling, (e) Bagging (GBDT) carbon content prediction results, (f) Adaboost (GBDT) carbon content prediction results, (g) Static ensemble learning carbon content prediction results, and (h) MC+DAS carbon content prediction results.



(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 9: Prediction results of temperature (Th = 10°C) under different methods. (a) GBDT temperature prediction results of global modeling, (b) XGBoost temperature prediction results of global modeling, (c) LightGBM temperature prediction results of global modeling, (d) RF temperature prediction results of global modeling, (e) Bagging (GBDT) temperature prediction results, (f) Adaboost (GBDT) temperature prediction results, (g) Static ensemble learning temperature prediction results, and (h) MC+DAS temperature prediction results.

Table 2: Comparison of performance indexes of this method at carbon content and temperature

Ablation experiment	Carbon content PA (Th = 0.02%)	RMSE	MAPE	Temperature PA (Th = 10°C)	RMSE	MAPE
Bootstrap-MS	81.6	0.0167	0.1129	74.4	10.2737	0.0047
Bootstrap-WF	79.2	0.0179	0.1137	42.4	14.9245	0.0076
Bootstrap-DAS	84.2	0.0153	0.1076	78.6	10.0963	0.0045
MC-MS	87.8	0.0146	0.1017	90.6	9.0265	0.0040
MC-WF	79.6	0.0154	0.1163	76.6	9.1213	0.0042
MC-DAS	92.8	0.0142	0.1047	91.6	9.0311	0.0040

Table 3: Comparison of performance indexes of different methods under carbon content and temperature

Different methods	Carbon content PA (Th = 0.02%)	RMSE	MAPE	Temperature PA (Th = 10°C)	RMSE	MAPE
GBDT	71.8	0.0229	0.1332	63.4	13.9487	0.0062
XGBoost [21]	56.0	0.0324	0.1832	46.2	19.5738	0.0091
LightGBM [22]	68.8	0.0258	0.1435	52.0	15.6460	0.0074
RF [23]	61.6	0.0283	0.1519	50.0	16.5279	0.0077
Bagging [24]	70.6	0.0223	0.1290	66.8	13.5034	0.0059
Adaboost [25]	66.6	0.0222	0.1343	59.8	14.1650	0.0065
MC-SA	74.2	0.0176	0.1373	74.2	9.2918	0.0044
MC-DAS	92.8	0.0142	0.1047	91.6	9.0311	0.0040

the furnace) to detect the endpoint carbon content and temperature in the furnace. Therefore, the proposed method is significantly less time-consuming than the subroutine detection method and can satisfy the real-time requirements of actual industrial processes.

5 Conclusion

This study focuses on the dynamic industrial process of BOF steelmaking, where data collected during the process of steelmaking can fluctuate significantly due to differences in the quality of raw materials. Traditional ensemble learning methods often ignore the performance differences between the query samples and the base model, resulting in poor prediction performance of the model as a whole. To address this issue, we propose a soft sensor modeling method based on MC dynamic adaptive selection ensemble learning.

(1) The different clustering algorithms are used to divide the original training data to form different sample subsets and also to build a diversity of the base model pool.
(2) A dynamic adaptive MS strategy is proposed, through which the model adapted to each sample is selected as a sub-model of subsequent integration, it improves the whole prediction precision and generalization performance of ensemble learning.

The simulation experiments conducted on the data of the converter steelmaking process prove that the method proposed in this study is effective in solving issues related to production process data. Additionally, compared to the global modeling model and traditional ensemble learning methods, the proposed method exhibits better prediction performance. Our results demonstrate that both the sample clustering-based distribution method and the dynamic selection ensemble-based ensemble learning modeling method can improve the PA of the ensemble learning framework. Furthermore, this approach provides a more reasonable and effective modeling method for actual dynamic industrial processes.

Acknowledgements: Thanks to all the authors for their contributions to this article and the editors for their valuable comments.

Funding information: This research is funded by the National Natural Science Foundation of China (Nos. 62263016 and 61863018) and the Applied Basic Research Programs of Yunnan Provincial Science and Technology Department of China (No. 202001AT070038).

Author contributions: Bin Shao: formal analysis, investigation, methodology, software, validation, visualization, writing – original draft, and writing – review and editing; Hiu Liu:

conceptualization, funding acquisition, project administration, resources, and supervision; Fugang Chen: experimental guidance.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: The data were obtained from the actual steelmaking plant, but the data are not available due to privacy.

References

[1] Zhang, C. J., Y. C. Zhang, and Y. Han. Industrial cyber-physical system driven intelligent prediction model for converter end carbon content in steelmaking plants. *Journal of Industrial Information Integration*, Vol. 28, 2022, id. 100356.

[2] Lu, C. Discussion on endpoint control technology of converter steelmaking. *Metallurgy and Materials*, Vol. 41, No. 2, 2021, pp. 87–88.

[3] Zhou, M. C., Q. Zhao, and Y. R. Chen. Endpoint prediction of BOF by flame spectrum and furnace mouth image based on fuzzy support vector machine. *Optik*, Vol. 178, 2019, pp. 575–581.

[4] Liu, X. C., H. Liu, F. G. Chen, and C. Li. A real-time prediction method of carbon content in converter steelmaking based on DDMCN flame image feature extraction. *Control Decis Mak*, 2021, pp. 1–9, id. 2166.

[5] Lin, B., B. Recke, J. K. H. Knudsem, and S. B. Jørgensen. A systematic approach for soft sensor development. *Computers & Chemical Engineering*, Vol. 31, No. 5–6, 2007, pp. 419–425.

[6] Liu, Q., M. W. Jia, Z. L. Gao, L. F. Xu, and Y. Liu. Correntropy long short term memory soft sensor for quality prediction in industrial polyethylene process. *Chemometrics and Intelligent Laboratory Systems*, Vol. 231, 2022, id. 104678.

[7] Jia, M. W., D. Y. Xu, T. Yang, Y. Liu, and Y. Yao. Graph convolutional network soft sensor for process quality prediction. *Journal of Process Control*, Vol. 123, 2023, pp. 12–25.

[8] Zeng, P. F. and H. Liu. Just-learning soft sensor method of endpoint carbon content and temperature in converter steelmaking based on quadratic similarity measure. *Computer Integrated Manufacturing System*, Vol. 27, No. 5, 2021, pp. 1429–1439.

[9] Zhou, P., H. Song, H. Wang, and T. Chai. Data-driven nonlinear subspace modeling for prediction and control of molten iron quality indices in blast furnace ironmaking. *IEEE Transactions on Control Systems Technology*, Vol. 25, No. 5, 2016, pp. 1761–1774.

[10] Liu, H. and P. F. Zeng. Endpoint carbon content and temperature measurement method based on WGRA-FCM for sample similarity measurement. *Control and Decision Making*, Vol. 36, No. 09, 2021, pp. 2170–2178.

[11] Yuan, X. F., Z. Q. Ge, and Z. H. Song. Process adaptive soft sensor modeling based on time difference and local weighted partial least squares algorithm. *Journal of Chemical Engineering*, Vol. 67, No. 03, 2016, pp. 724–728.

[12] Liu, C., L. Tang, and J. Liu. A stacked auto-encoder with sparse Bayesian regression for endpoint prediction problems in steelmaking process. *IEEE Transactions on Automation Science and Engineering*, Vol. 99, 2019, pp. 1–12.

[13] Cui, L. L., B. B. Shen, and Z. Q. Ge. Soft sensor modeling method based on mixed variational autoencoder regression model. *Journal of Automation*, Vol. 2, 2022, pp. 398–407.

[14] Zhang, X., M. Kano, and S. Matsuzaki. Ensemble pattern trees for predicting hot metal temperature in blast furnace. *Computers & Chemical Engineering*, Vol. 121, 2019, pp. 442–449.

[15] Lv, W., Z. Mao, P. Yuan, and M. Jia. Pruned bagging aggregated hybrid prediction models for forecasting the steel temperature in ladle furnace. *Steel Research International*, Vol. 85, No. 3, 2014, pp. 405–414.

[16] Liu, J. J., P. Zhou, and L. Wen. Probability weighted ensemble learning modeling of molten iron quality root mean square error. *Control Theory and Application*, Vol. 37, No. 5, 2020, pp. 987–998.

[17] Xiong, Q., H. Liu, and X. C. Liu. Soft measurement method of endpoint carbon content and temperature of converter steelmaking based on LNN-DPC weighted ensemble learning [J/OL]. *Computer Integrated Manufacturing System*, Vol. 28, No. 12, 2022, pp. 3866–3898.

[18] Ahmad, Z. and J. Zhang. Bayesian selective combination of multiple neural networks for improving long-range predictions in nonlinear process modelling. *Neural Computing & Applications*, Vol. 14, No. 1, 2005, pp. 78–87.

[19] Ahmad, Z. and J. Zhang. Combination of multiple neural networks using data fusion techniques for enhanced nonlinear process modelling. *Computers & Chemical Engineering*, Vol. 30, No. 2, 2006, pp. 295–308.

[20] Anonymous, Oxygen top blowing converter sublance test. *Angang Steel Technology*, Vol. 1, 1976, pp. 32–37.

[21] Wang, T. T., Y. J. Bian, Y. X. Zhang, and X. L. Hou. Classification of earthquakes, explosions, and mining-induced earthquakes based on XGBoost algorithm. *Computers & Geosciences*, Vol. 170, 2023, id. 105242.

[22] Wang, D. N., L. Li, and D. Zhao. Corporate finance risk prediction based on LightGBM. *Information Sciences: An International Journal*, Vol. 602, 2022, pp. 259–268.

[23] Fang, K. N., J. B. Wu, J. P. Zhu, and B. C. Xie. Review of random forest methods. *Statistics and Information Forum*, Vol. 26, No. 3, 2011, pp. 32–38.

[24] Tang, W. and Z. H. Zhou. Selective cluster integration based on bagging. *Journal of Software*, Vol. 4, 2005, pp. 496–502.

[25] Cao, Y., Q. G. Miao, and J. C. Liu. Research progress and prospect of Adaboost algorithm. *Journal of Automation*, Vol. 39, No. 6, 2013, pp. 745–758.