Classification of blast furnace internal state based on FLS and its application in furnace temperature prediction

WANG Gao-peng^{1*}, ZHAI Hai-peng¹, YAN Zhen-yu¹, ZHENG Rui-ji¹

¹ School of software engineer, Chongqing university of post and communication, Chongqing 400065, China

Abstract. The real-time and accurate prediction of the molten iron silicon content of the blast furnace plays an important role in regulating the temperature of the blast furnace and stabilizing the furnace condition. When the time is large, the accuracy and credibility of the forecast results decrease rapidly, which is not conducive to on-site operators to carry out production operations according to the forecast results. To this end, this paper adds a state variable to each piece of data through the flexible least square parameter estimation method, and selects the training set in a state similar to the test sample. This makes the selection of training data more accurate and reliable. Application examples show that the method proposed in this paper improves the accuracy of silicon content prediction results and has good guiding significance for actual production operations.

1 Introduction

The blast furnace reaction is a continuous process including complex physical and chemical changes, and various reactions are carried out in a high temperature environment. In addition, the blast furnace system itself has the characteristics of complex nonlinearity, large hysteresis, and large noise. The silicon content of molten iron is a reference variable to characterize the thermal state of the blast furnace and its change trend, and it is also an important production index to measure the stability of the furnace condition and the quality of the molten iron during the blast furnace smelting process. However, silicon content can not be detected on-line in real time, which results in too much reliance on expert experience in the regulation of blast furnace. Therefore, accurate forecasting of the silicon content of molten iron is of great significance to ensure the anterograde of the blast furnace and the high production and low consumption.^[1-3]

At present, silicon prediction models are mainly divided into three categories: mechanism models, reasoning models, and mathematical models. These predictive control models have achieved good results in the control of blast furnaces, and have provided blast furnace operators with an important reference basis for furnace temperature control, so that they can arrange and formulate blast furnace-related operations. With the rapid development of machine learning, mathematical models occupy a dominant position among the three models. It mainly includes linear regression model^[4], neural network model^[5-6], support vector machine model^[7]. Moreover, certain research progress has been made in the prediction of blast furnace temperature and practical operation of each model.

However, none of the above models evaluates the internal state of the blast furnace. The furnace temperature reflects the temperature level from the lower part of the furnace body to the main reaction zone of the hearth. In addition to the actual measured data, it is also closely related to the internal conditions of the blast furnace. Because the conditions in the furnace cannot be directly detected, and the observations only indirectly reflect a certain aspect of the conditions in the furnace, and the quantity is limited, resulting in the lack of direct, timely, and accurate information in the furnace when predicting the furnace temperature. There are defects in the actual application of the model, which are specifically manifested in:

1) The training data of the model cannot be selected effectively

a) If you choose a certain period of time data as the basis for modeling, after a long period of time, the internal state of the blast furnace has changed, the model is not adapted to the new internal state, and the prediction results are significantly deteriorated;

b) If you always choose the latest production data for modeling, the internal state changes may be small, but some items in the latest production data may change very little. If there are significant changes in these data items during forecasting, the model will not be able to adapt;

c) If you select data for a long enough period of time for modeling, because these data contain multiple internal states, the prediction results will not be accurately given.

2) When the furnace state changes, the furnace temperature prediction model is difficult to adjust accordingly

Current prediction models are generally trained based on matching input and output data. In the specific training,

*Corresponding author's e-mail: wanggp@cqupt.edu.cn

the output data is the furnace temperature. The value at which time point or several time points is selected for the matching input data is adjusted according to the production status of the blast furnace. The manually set value cannot accurately reflect the real-time dynamics. Changes are detrimental to the effect of the model; in addition, the prediction algorithm needs to be adjusted accordingly for different model structures.

In order to solve the above problems, this paper proposes a method based on the flexible least square parameter estimation method^[8] to add state variables to each data, and select the appropriate size training set which can hold the state information of test samples. The data of the training set is able to predict the silicon value under the fluctuation of furnace condition. Therefore, the accuracy of silicon prediction can be effectively improved. It provides an important basis for further research in the future.

2 The problem of division of blast furnace internal state

In the existing silicon prediction models, there is a lack of analysis on the division of the internal state of the blast furnace. The blast furnace reaction is a very complex reaction process that combines physical and chemical reactions. When the initial furnace conditions are different, even if the same input conditions are given, the output results obtained will be different. At present, in order to avoid the above situation, recent data is often used as the training set. Although this can avoid the interference of the multi-state data set on the prediction accuracy of the model to a certain extent, the recent data set lacks the evolution process of the furnace conditions compared with the historical data set, and it is difficult to predict the changes of the furnace conditions. Therefore, selecting the data under the same furnace condition as the test sample from the historical data as the actual training set is of great significance for improving the accuracy of silicon prediction. This paper is based on the flexible least squares parameter estimation method (FLS), adding state variables to each piece of data. Realize the division of the internal state of the blast furnace. When a certain sample needs to be predicted, the FLS algorithm is used to get the state variables closest to the sample, and then a group of training sets closest to the state variables are obtained by calculation. Compared with the general training set, the training set not only contains the state information of the predicted sample, but also contains the correlation of the furnace condition changes in the hours before and after the prediction sample information. It is helpful to improve the prediction accuracy of prediction samples. The technical route is shown in Fig 1.

Fig 1. Technical route

3 Flexible least square parameter estimation method

3.1 FLS Fundamental

Assuming a discrete time series $1, 2, \dots, n$, get a set of input values $x_1, x_2, \dots, x_n$ with a set of output values $y_1, y_2, \dots, y_n$, the corresponding regression coefficient can be calculated. And the linear measurement value meets:

$$y_i - x_i^T b_i \approx 0, i = 1, 2, \dots, n. \quad (1)$$

Due to the time series of the data, the regression coefficient of the model changes slowly in a continuous time period. So the coefficient dynamic stability satisfies:

$$b_{i+1} - b_i \approx 0, i = 1, 2, \dots, n-1. \quad (2)$$

Here: $x_i^T = (x_{i1}, \dots, x_{ik})$ is $1 \times k$ detectable vector, $b_i = (b_{i1}, \dots, b_{ik})^T$ is $k \times 1$ unknown regression coefficient.

Array of coefficients estimated by the model $b = (b_1, b_2, \dots, b_n)$, since neither will meet the measurement target (1), The dynamic index of the coefficient (2), So there will be two kinds of errors as follows:

Measurement residual:

$$r_M^2(b; n) = \sum_{i=1}^n [y_i - x_i^T b_i]^2. \quad (3)$$

Dynamic residual:

$$r_D^2(b; n) = \sum_{i=1}^{n-1} [b_{i+1} - b_i]^T [b_{i+1} - b_i]. \quad (4)$$

Make $C(b; \mu, n)$ equal to the sum of the weighted dynamic residual and the measured residual:

$$C(b; \mu, n) = \mu r_D^2(b; n) + r_M^2(b; n). \quad (5)$$

Combination (3)-(4) can be obtained

$$C(b; \mu, n) = \mu \sum_{i=1}^{n-1} [b_{i+1} - b_i]^T [b_{i+1} - b_i] + \sum_{i=1}^n [y_i - x_i^T b_i]^2. \quad (6)$$

If $\mu \geq 0$, Then the equation(6)use the FLS method can be used to find the minimum parameter sequence

$$b^{FLS}(\mu, n) = (b_1^{FLS}(\mu, n), \dots, b_n^{FLS}(\mu, n)). \quad (7)$$

If $\mu = 0$, Then the estimation of the parameter sequence b is equivalent to finding the minimum value of r_M^2 without considering r_D^2 . At this time, the value of r_D^2 will be very large, which is not conducive to effectively tracking the true parameter value. If the μ value tends to infinity, then the value of parameter b it is estimated that the minimum value of r_M^2 must also be subject to $r_D^2 = 0$.

For the FLS solution (7) of equation (6), the following method can be used: First make I as $k \times k$ identity matrix; Then we can see $X(n)^T = [x_1, \dots, x_n]$ is $k \times n$ regression matrix; $b(n) = (b_1^T, \dots, b_n^T)^T$ is $nk \times 1$ coefficient column vector; $y(n) = (y_1, \dots, y_n)^T$ is the column vector of the observations of $n \times 1$, Then stipulate

$$G(n) = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_n \end{pmatrix} = nk \times n \quad (8)$$

$$A_i(\mu) = \begin{cases} x_1 x_1^T + \mu I, i = 1; \\ x_i x_i^T + 2\mu I, i \neq 1, n; \\ x_n x_n^T + \mu I, i = n; \end{cases} \quad (9)$$

$$A(\mu, n) = \begin{bmatrix} A_1(\mu) & -\mu I & 0 & \dots & \dots & 0 \\ -\mu I & A_2(\mu) & -\mu I & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & -\mu I \\ 0 & \dots & \dots & 0 & -\mu I & A_n(\mu) \end{bmatrix} \quad (10)$$

It can be proved that equation (6) satisfies

$$C(b(n); \mu, n) = b(n)^T A(\mu, n) b(n) - 2b(n)^T G(n) y(n) + y(n)^T y(n) \quad (11)$$

Synthesize the previous analysis requirements to obtain the time-varying coefficient estimation sequence $b(n)$, so that the effective cost equation $\min_{b(n)} C(b(n); \mu, n)$ holds, then the necessary condition is $\frac{\partial C(b(n); \mu, n)}{\partial b(n)} = 0$, so the solution is

$$A(\mu, n)b(n) = G(n)y(n) \quad (12)$$

It can be proved that for any $\mu \geq 0$, matrix $A(\mu, n)$ is a positive semi-definite matrix, but when $\mu > 0$ and the rank of $X(n)$ matrix is k , matrix $A(\mu, n)$ is positive definite and equation (11) is a convex function of $b(n)$, so the minimization of equation (11) has The only solution

$$b^{FLS}(\mu, n) = A(\mu, n)^{-1} G(n) y(n) \quad (13)$$

At this time, formula (13) is called the estimated value of the time-varying coefficient linear regression equation at the current moment. That is, the state variable of each piece of data is obtained.

3.2 The Theory of Blast Furnace State Division Based on FLS

In traditional linear regression, based on the principle of least squares, the regression line obtained every time is a straight line, as shown in Figure (a) below. If the regression coefficient is regarded as the importance of the input value relative to the output value, then all the input values under the algorithm have the same importance. It is impractical to apply in the field of blast furnace, because the state of the blast furnace changes linearly. The importance of each input variable changes slowly on the timeline with the internal state. The FLS algorithm allows the regression coefficient to become a variable by increasing the constraint condition of $b_{i+1} - b_i \approx 0$ on the regression coefficient, and increases the degree of tolerance change $\frac{1}{\mu}$. When μ is smaller, the degree of tolerance is higher, that is, the importance of the input value is allowed to be in a short time The sharp change is shown in the figure as allowing the generation of acute angles, as shown in Figure (b); when μ is larger, the tolerance is lower, and the regression line at this time is close to the traditional regression line, as shown in Figure (c); according to the experience of blast furnace experts The actual situation of the blast furnace should be shown in Figure (d). It allows the importance of the input value to change, but the range of change in a short time is subject to certain constraints.

Fig 2. Comparison between traditional linear regression and FLS regression with different parameters

4 Experimental simulation

4.1 Raw data set acquisition

The data sample is divided into two parts. The first part contains all historical training sets, which are collected from real-time production data of a steel plant from January to December 2013. The second part is a recent data set with prediction samples, collected from the real-time production data of the steel mill from August to December 2014.

4.2 Data preprocessing

The preliminary data contains the problems of uneven time interval for obtaining data, unsmooth data, non-corresponding time point of input data, and ununiform data dimension. Time interval equalization, data smoothing, system identification, and data The normalization method preprocesses the preliminary data.

4.3 Add state variables to each piece of data through FLS algorithm

The regression coefficient of the FLS algorithm can be used as the state variable of the data because it adds dynamic residuals on the basis of linear regression only considering the measurement residuals. This makes the linear regression coefficient of the data change slowly. In the training process, the input data multiplied by the corresponding regression coefficient is approximately equal to the output value. At this time, the calculated regression coefficient can be understood as the importance of each input vector, and the importance of each input vector at a certain moment also reflects the situation in the furnace at this time. And the importance of each vector changes slowly with time, which is in line with the actual situation of blast furnace operation.

It can be seen from Table 1 that between adjacent samples, the variation range of the regression coefficient (RC) is between 0.001-0.01, indicating that the importance of each input variable (IV) only changes slightly in a short time, and on this basis, The fitting output value (FOV) is also very close to the actual output value (AOV). Therefore, the regression coefficient obtained based on the FLS algorithm can be regarded as the state variable of each sample.

Table 1. Comparison of fitting output value and actual output value

	SAMPLE 1	SAMPLE 2	SAMPLE 3	SAMPLE 4	SAMPLE 5
IV 1	-0.472	-0.416	-0.285	-0.200	-0.261
IV 2	-0.758	-0.954	-1.008	-1.028	-1.065
IV 3	0.538	0.579	0.665	0.786	0.939
IV 4	0.771	0.698	0.686	0.652	0.641
IV 5	-1.432	-1.413	-1.349	-1.294	-1.314
IV 6	-0.602	-0.466	-0.417	-0.411	-0.451
IV 7	0.375	0.350	0.342	0.342	0.342
RC 1	0.288	0.291	0.295	0.299	0.303
RC 2	-0.575	-0.569	-0.563	-0.556	-0.549
RC 3	-0.005	-0.009	-0.014	-0.018	-0.023
RC 4	0.317	0.311	0.305	0.298	0.292
RC 5	-0.067	-0.057	-0.045	-0.033	-0.021
RC 6	-0.092	-0.088	-0.083	-0.078	-0.073
RC 7	0.416	0.413	0.410	0.407	0.404
FOV	0.848	0.901	0.919	0.906	0.871
AOV	0.915	0.919	0.920	0.902	0.862

4.4 Choose the right training set

A state variable is added to each piece of data through the FLS algorithm, and the state variable closest to the test sample is used as the reference object. At this time, the accurate state variables of the test sample cannot be obtained, because the output result is unknown before the prediction, and the calculation of the state variable requires both input variables and output variables, so it can be obtained in two steps in a similar state to the test sample Training data.

The first step: The data which is closest to the test sample and satisfies the condition of computable state variable is obtained. Since the predicted data is three hours later, the input and output of the data before three hours are known relative to the test samples. The data can be used to calculate the state variables. It should be noted that because the state variables obtained at the end need to be compared with the state variables of the historical data, it should be ensured that the state variables calculated by the two are performed on the same basis, because the state variables are normalized input, The output data is calculated, if the two normalization does not use the same

vector average and standard deviation, then the state data obtained by the two will not be comparable. Therefore, it should be ensured that the normalized parameters of the current data are the same as the normalized parameters of the historical data. Because the state variables of the historical data are already fixed, the current data uses the vector average and standard deviation of the historical data to achieve normalization.

The second step is through the Euclidean distance formula:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (14)$$

In all training samples, the data that is closer to the Euclidean distance of the state data of the test sample is selected, but the currently obtained data does not well represent that it is in a similar state with the test sample, so the obtained data should be backward at the same time Delay for three hours (for example, get the data at 1 o'clock, 2 o'clock, and 4 o'clock, and delay the data for 4 o'clock, 5 o'clock, and 7 o'clock after three hours), the data at this time can be better represented It is in a similar state to the test sample. So take the final data as the training set of test samples. This process is shown in Figure 3 below.

Figure 3. Schematic diagram of training set selection

5 Experimental results and analysis

After the above experimental process is compiled and implemented by MATLAB, the second part is a recent data

set with prediction samples (real-time production data collected from a steel plant from August to December 2014) for training prediction. And two comparative experiments were added, and the experimental results are shown below.

Fig 4. Comparison of prediction results with different training sets

In Figure (a), the blue "+" sign is the actual measured silicon value. The training set of the red solid line is a group of data that is closest to the Euclidean distance of the input vector of the test sample from all the historical data. The training set of the blue solid line is the historical training set processed by FLS algorithm. It can be seen from the figure that the prediction effect of the former is far less than that of the latter in both prediction trend and prediction accuracy, which also shows that even with similar input, the output of blast furnace under different furnace conditions is often different.

In Figure (b), the blue "+" sign is the actual measured silicon value, the training set of red solid line adopts all historical data, and the training set of blue solid line adopts the historical training set processed by FLS algorithm. It is not difficult to see from the chart that although the former has certain effect in forecasting trend, it is far less accurate than the latter. This is because all historical data are selected as the training set, which contains too many kinds of furnace conditions, and each furnace condition is a kind of interference compared with other furnace conditions. Therefore, it can be concluded that if there are too many kinds of furnace conditions in the training set or there is a big difference between the furnace conditions in the training set and the predicted samples, the prediction effect is often not ideal.

In Figure (c), the blue "+" sign is the actual measured silicon value, the training set of red solid line is the recent training set, and the training set of blue solid line is the historical training set processed by FLS algorithm. In the first 1-50 sample points, it can be seen that the actual [Si] value changes slowly, which indicates that the furnace condition also changes slowly. At this time, the prediction effect of the model obtained by the recent training set is better than that of the model using the historical training set. When the sample points are around 100 and 250, it can be seen that the actual [Si] value is suddenly and violently changed, which indicates that the furnace condition has also changed dramatically. At this time, the prediction effect of historical training set is better than that of recent training set.

6 conclusion

(1) Based on FLS algorithm, by adjusting the super parameter μ , the change degree of the model is controlled. When it is consistent with the change degree of furnace condition, the regression coefficient obtained by FLS algorithm can be used as the state variable of the data.

(2) By matching the state variables of the data, the data group including the state variables of the prediction samples is selected as the training set, which can timely adjust the model parameters when the furnace conditions fluctuate, and improve the silicon prediction accuracy under the unstable furnace conditions.

The research work in this paper provides a potential new idea for the establishment of predictive control model of blast furnace smelting process. On this basis, the combined historical data and recent data can be further studied to make the prediction model perform well when the furnace conditions fluctuate and stabilize, and improve

the prediction accuracy and efficiency. It will be the author's research topic in the future to excavate the prediction advantages brought by the internal state division of blast furnace and study the mathematical model of smelting process more suitable for the requirements of big data era.

Reference

1. Chen, M., Yi, Y.X., Zhu, Q., Zhang, H.G.(2014) Prediction of blast furnace temperature based on distributed parameter model. *Control theory and Application*, 31(09):1232-1237.
2. Henrik, S., Gao, C.H., Gao, Z.W.(2013) Data-driven time discrete models for dynamic prediction of the hot metal silicon content in the blast furnace—a review. *IEEE Transactions on Industrial Informatics*, 9(4):2213-2225.
3. Barbasova, T.A., Zagorskina, E.V.(2019) Blast-Furnace Melting Blast Control. In: 2019 IEEE Russian Workshop on Power Engineering and Automation of Metallurgy. Russian. 13-17.
4. Wang, G.P.(2018) Silicon Prediction Model of Blast Furnace Based on ARX and PCR. In: 2018 13th World Congress on Intelligent Control and Automation (WCICA). China. 1214-1220.
5. Zhou, P., Zhang, L., Li, W.P., Dai, P., Chai, T.Y.(2018) Modeling of multiple hot metal quality in blast furnace based on stochastic weight neural network integrating self coding and PCA. *Journal of automation*, 44(10): 1799-1811.
6. Cui, J.M., Jiang, Z.G., Zhan, W.P., Gu, J.H.(2015) Prediction of blast furnace temperature based on time series neural network. *Metallurgical Automation*, 39(05):15-21.
7. Zhou, P., Li, R.F., Guo, D.W., Wang, H., Chai, T.Y.(2016) Multi output support vector regression modeling of multiple hot metal quality indexes in blast furnace ironmaking process. *Control theory and Application*, 33(06):727-734.
8. Jiang, C.H., Dong, M.L., Gui, W.H., Yang, C.H., Xie, Y.F.(2016) Two dimensional prediction of silicon content in hot metal of blast furnace based on Bootstrap. *Journal of automation*, 36(50):331-336.