RESEARCH ON APPLICATION OF SINTERING BASICITY BASED ON LS-SVM AND GRAY THEORY

Song Qiang¹, Yang Pu², Guo Xiaobo³
¹ School of Mechanical Engineering Anyang Institute of Technology, Anyang, Henan, 455004, China
² The University of Melbourne, Melbourne of Australia, 3010;
³ School of Mechanical Engineering Anyang Institute of Technology, Anyang, Henan, 455004, China
Email:13523323305@163.com; puyang@student.unimelb.edu.au

Abstract: Prediction of alkalinity in sintering process is difficult. So the recent new technology, gray least square support vector machine have been introduced. which can deal with nonlinear adaptive information, combination and gray support vector machine these advantages. The results show that, the basicity of sinter, can accurately predict the small sample and reference information using the model. The experimental results show that, the gray support vector machine model is effective and with practical advantages of high precision, less samples, and simple calculation.

Keywords: Basicity, In Sintering Process, gray Relation Analysis, grays Least Squares Support Vector Machine, Prediction, gray Model.

1. INTRODUCTION

In the modern steel enterprises, the sintering process for blast furnace materials is one of the most important production processes. The sintering alkalinity has a direct effect on production and economic benefits of the whole steel enterprise (K. Umer et al., 2017). Therefore, almost every steel factory is equipped with many instruments and automatic control systems in the sintering plant for the production process control. However, the complexity of the sintering process makes the process to be difficultly described by a set of mathematical models. Since this process often has large time delay and dynamic time variability, it is hard to perform control tasks of the whole sintering process using conventional control models.

Sintering process is a complex physical-chemistry process, which relates to a lot of characteristics including complicated mechanism, high nonlinearity, strong coupling, high time delay, and etc (Xie Q M, et al, 2015). The mathematical model for whole sintering process cannot be established, thus we can only construct mathematical model for one of performance indicators and the performance index of sintering process determines the policy of blending process. Because of the restriction of the detecting means, the chemical examination of sinter alkalinity generally needs 40 min. In the whole process, its time sometimes can even exceed 1 hour. Obviously, such a long time delay cannot meet the needs of actual productivity, and thus, the sinter alkalinity must be detected and a model for predicting the alkalinity should be established (S. Routray et al, 2018).

2. GRAY THEORY

2.1 Gray GM (1, 1) Model

Gray theory is a method to study the small sample, poor information, uncertainty, with partial information known, part information unknown small sample, poor information uncertainty problem as the research object, the known information through data mining, to extract valuable information, a correct description of system behavior, evolution and effective monitoring and prediction of the position information system. Statistical prediction method has many advantages compared with the traditional method, it does not need to determine whether the forecast variables subject to normal distribution, don't need large sample statistics, that is to say the research object specially for the small sample, poor information uncertainty, do not need to change according to the change of input variables forecast model, produced by the gray sequence the gray system theory, think, although the objective system of complex data representation, scattered, but it always has a whole function, it must contain the inherent law of some. The key is how to choose the appropriate way to tap it and use it. All the gray sequence can be generated by a weakening the randomness, reveal the regularity. A differential equation is a unified model; differential equation model has
higher prediction accuracy. The establishment of GM (1, 1) model is basically a cumulative generation of the original data, so that the generated sequence has certain regularity, by establishing the differential equation model, obtain the fitting curve, and thus to predict the unknown parts of the system. The GM model is first of the original data, an accumulated, generate 1-AGO, accumulated data through data mining have certain regularity, the original data is not obvious regularity, and its development trend is swinging. If the original data were accumulated generating, its regularity is obvious (Bere; Berce; and Nemes, 2012).

Assume that there is a time response sequence (which is called original time series),

\[ x^{(i)} = (x^{(i)}(0), x^{(i)}(0), \ldots, x^{(i)}(n)) \]  

(1)

Where \( x(0)(i) \) stand for the monitoring data at time \( i \), \( i=0,1,2,\ldots,n \).

The forecast value \( x^{(o)}(n+T) \), \( T=1,2,K,R. \) can be derived by the following three steps.

1. Build up the first-order accumulating generator operator (AGO)

This step is to weaken the indeterminacy in the original time series and get a more regular time series. Let \( x^{(i)} \) be the generated time series.

\[ x^{(i)} = \{x^{(i)}(1), x^{(i)}(2), \ldots, x^{(i)}(n)\}, x^{(i)}(k) = \sum_{i=0}^{k} x^{(i)}(k) \]

, \( k=1,2,\ldots,n \)  

(2)

Where \( x^{(i)} \) is the once Accumulated Generating Operation (1-AGO) sequence1-AGO

2. Construct first-order linear differential equation, the whitenization differential equation can be obtained:

\[ \frac{d x^{(1)}}{d t} + ax^{(1)} = u \]  

(3)

Where, \( a \) is a developing coefficient, whose value reflects the variation relation of data; and \( u \) is gray action quantity, which is the most important difference between gray and common model.

Using the least squares estimation, \( a \) and \( u \) can be obtained:

\[ \beta = \begin{bmatrix} a \\ u \end{bmatrix} = \left( B \cdot B^T \right)^{-1} B^T Y \]  

(4)

Where

\[ B = \begin{bmatrix} -0.5 x^{(1)}(2) + x^{(1)}(1) \\ -0.5 x^{(1)}(3) + x^{(1)}(2) \\ \vdots \\ -0.5 x^{(1)}(n) + x^{(1)}(n-1) \end{bmatrix} \quad Y = \begin{bmatrix} x^{(o)}(2) \\ x^{(o)}(3) \\ \vdots \\ x^{(o)}(n) \end{bmatrix} \]

According to linear first-order differential equation, Eqn.(5) can be derived

\[ \ddot{x}^{(i)}(k+1) = \left[ x^{(i)}(1) - \frac{n}{a} \right] e^{-ak} + \frac{n}{a} \]  

(5)

3. Inverse the accumulation generation

Let \( \dot{x}^{(i)} \) be the fitted and forecasted series

\[ \dot{x}^{(i)}(k) = \left\{ \dot{x}^{(i)}(1), \dot{x}^{(i)}(2), K, \dot{x}^{(i)}(n) \right\} \]  

(6)

Then, the predicted value can be calculated by Eqn.(7),

\[ x^{(o)}(i) = \dot{x}^{(i)}(i) - \dot{x}^{(i)}(i-1), i = 1,2, K,n \]  

(7)

Where, \( \dot{x}^{(i)}(1), \dot{x}^{(i)}(2), \ldots, \dot{x}^{(i)}(n) \) are fitted value of the original series; and \( \dot{x}^{(i)}(n+1), \dot{x}^{(i)}(n+2), \ldots \) are the forecast values.

4. Error examination

The relative error can be calculated by Eqn. (8),

\[ e(k) = \frac{x^{(o)}(k) - \dot{x}^{(o)}(k)}{x^{(o)}(k)} \times 100 \% \]  

(8)

Where, \( e \) is the error percentage.

2.2 Residual Forecasting Model

To evaluate modeling performance, we should do synthetic test of goodness

\[ \frac{s_2}{C} = s_1 \]  

(9)

Where

\[ S21 = \frac{1}{n} \sum_{k=1}^{n} \frac{(x^{(o)}(n) - x^{(o)}(k))^2}{S22 = \frac{1}{n} \sum_{k=1}^{n} (x(k) - e)^2} \]

\[ \varepsilon \varepsilon \varepsilon \{\varepsilon(1), \varepsilon(2), \ldots, \varepsilon(n)\} = \{x(0)(1)-\} \]

\[ \hat{x}^{(o)}(1), x^{(o)}(2) - \hat{x}^{(o)}(2), \ldots, x^{(o)}(n) - \hat{x}^{(o)}(n) \}

\[ P = P \{\left| k - \hat{e} \right| < 0.6745S1 \} \]  

(10)

The precision grade of forecasting model can be seen in Table 1.

Finally, applying the inverse accumulated generation operation (AGO), we then have prediction values \( \ddot{x}^{(i)}(k) = x^{(1)}(k) - x^{(i)}(k-1) \)

131
2.3 Gray Relation Analysis

Gray relational is uncertain correlation between things, or uncertain correlation between the system factors, between the factors to principal act. The fundamental mission of gray relational analysis is geometry approach to the micro or macro basing on the sequence of behavioral factors, to analyze and determine the influence degree between each factors or the contribution measure of factor to the main behavior, and its fundamental idea is judge whether the geometry of the sequence of curves is closely linked according to the level of similarity of it, and the closer the curve, the greater the correlation of the corresponding sequences, conversely, the smaller (Berce; Pacurar and Balc, 2008, C. Sun et al., 2016). The computational procedure of gray relational analysis is expressed below:

Assume

\[ X^i = \{x_0^i, x_1^i, x_2^i, x_3^i, x_4^i\} \]

Then

\[ \varepsilon_{ij} = \frac{1 + |s_i - s_j|}{1 + |s_i + s_j + s_i - s_j|}, \quad i=1, 2,...,9; \]

So we will obtain

\[ \varepsilon_{01} = 0.7429; \quad \varepsilon_{02} = 0.8571; \quad \varepsilon_{03} = 0.9348; \]
\[ \varepsilon_{04} = 0.8633; \]
\[ \varepsilon_{05} = 0.7333; \quad \varepsilon_{06} = 0.7081; \quad \varepsilon_{07} = 0.7717; \quad \varepsilon_{08} = 0.98 39; \quad \varepsilon_{09} = 0.9478; \]

Then we may find

\[ \varepsilon_{08} > \varepsilon_{09} > \varepsilon_{03} > \varepsilon_{04} > \varepsilon_{02} > \varepsilon_{07} > \varepsilon_{01} > \varepsilon_{05} > \varepsilon_{06} \]

\[ X_2 > X_3 > X_4 > X_5 > X_6 > X_7 > X_8 > X_9 \]

By the definition of the relational analysis, the allocation is sequenced by the relational coefficient. And the sequence of relational

<table>
<thead>
<tr>
<th>Table 1: Precision grade of forecasting model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision grade</td>
</tr>
<tr>
<td>Good</td>
</tr>
<tr>
<td>Qualified</td>
</tr>
<tr>
<td>Just</td>
</tr>
<tr>
<td>Unqualified</td>
</tr>
</tbody>
</table>
coefficient size is the order of ranking of the influence factor. Based on the above result, it is clear that $X_s = 0.8633$ is maximum, and it represents that the relational degree is biggest between the first allocation center and the ideal allocation center. Therefore, the first allocation center is the most optimal choice.

$X_s$ is the optimal factors, $X_s$ ranked second, $X_s$ is the worst in all factors. That is to say, the CaO ratio of sinter basicity of pulverized coal ratio, basicity on sinter influence is relatively large, the thickness of the material layer and mixed FeO content in ore has very little effect on the basicity of sinter, might as well put the two operating variables from outside.

3. LEAST-SQUARES SUPPORT VECTOR MACHINES ALGORITHM MODELING

3.1 Least-Squares Algorithm Support Vector Machine

Recently, least squares support vector machine (LS-SVM) has been applied to machine learning domain successfully. It is a promising technique owing to its successful application in classification and regression tasks (Fu et al., 2012). It is established based on the structural risk minimization principal rather than the minimized empirical error commonly implemented in the neural networks. LS-SVM achieves higher generalization performance than the neural networks in solving these machine learning problems. Another key property is that unlike the training of neural networks which requires nonlinear optimization with the danger of getting stuck into local minima, training LS-SVM is equivalent to solving a set of linear equation problem (Wang et al., 2018). Consequently, the solution of LS-SVM is always unique and globally optimal. In this study, the application of LS-SVM in the prediction of the alkalinity in sintering process was discussed (Mozer et al., 2017).

Giving a training set \( \{(x_i, y_i), i = 1, \ldots, N\} \), with $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$, where $x_i$ is the input vector of the first t samples, $y_i$ is the desired output value of the first t samples, and N is the number of samples, the problem of linear regression is to find a linear function $y(x)$, which is equivalent to applying a fixed non-linear mapping of the initial data to a feature space. In feature space, SVM models take the form:

$$y(x) = w^T \varphi(x) + b$$  \hspace{1cm} (11)

Where the nonlinear function mapping $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ maps the high-dimensional space into the feature space; and $w$ is not a pre-specified dimensional, possibly infinite dimensional, b is a real constant (C et al., 2016).

The least squares approach prescribes choosing the parameters ($w, b$) to minimize the sum of the squared deviations of the data, and the square loss function is described as:

$$\min J(w, \epsilon) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} \epsilon_i^2$$  \hspace{1cm} (12)

Where $\gamma$ is the trade-off parameter between a smoother solution, and training errors.

With constraints $y(x) = w^T \varphi(x) + b + \epsilon_i$, for $t = 1, \ldots, N$.

Important differences with standard SVM are the equality constraints and the squared error term, which greatly simplifies the problem (Xie et al., 2015).

Only equality constraints and the optimization objective function is the error loss, which will simplify the problem solving (Zhang X G, 2016).

To solve this optimization problem, one defines the following Lagrange function:

$$L(w, b, \alpha, \epsilon) = J(w, \epsilon) - \sum_{i=1}^{N} \alpha_i(w^T \varphi(x_i))$$  \hspace{1cm} (13)

Where, $\alpha_i$ is Lagrange multipliers? By Karush-Kuhn-Tucker (KKT) optimal conditions, the conditions for optimality are

$$\begin{align*}
\frac{\partial L}{\partial w} = 0 & \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \\
\frac{\partial L}{\partial b} = 0 & \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\
\frac{\partial L}{\partial \epsilon_i} = 0 & \rightarrow \alpha_i = \gamma \epsilon_i \\
\frac{\partial L}{\partial \alpha_i} = 0 & \rightarrow w^T \varphi_i + b + \epsilon_i - y_i = 0
\end{align*}$$  \hspace{1cm} (14)

Where, t = 1… N.

After elimination of $\epsilon_i$ and w, the solution is given by the following set of linear equations

$$\begin{bmatrix}
0 & I^T \\
I & \varphi(x_i)^T \varphi(x_i) + D
\end{bmatrix} \begin{bmatrix}
[0] \\
[\alpha]
\end{bmatrix} = \begin{bmatrix}
y
\end{bmatrix}$$  \hspace{1cm} (15)
Where, $y = [y_1, \ldots, y_N]^T$, $a = [a_1, \ldots, a_N]^T$, $D = \text{diag}([y_1, \ldots, y_N])$. Select $\gamma > 0$, and guarantee matrix $\begin{bmatrix} b \\ a \end{bmatrix} = \phi \begin{bmatrix} 0 \\ \alpha \end{bmatrix}$.

### 3.2 Selection of the Kernel Function

By the KKT-optimal conditions, $w$ is obtained, and thus the training set of nonlinear approximation is obtained too (Zheng and Qian, 2005).

$$y(x) = \sum_{i=1}^{N} a_i k(x, x_i) + b$$

Where $x, x_i$ denote training point and support vector respectively, $y$ is the output of network $a_i$, and $b$ are the solutions to Eq.(13).

$$k(x, x_i) = \phi(x_i)^T \phi(x_i), i, k = 1, \ldots, N$$

(17)

The Selection of the kernel function $\phi : R^d \rightarrow R^m$ has several possibilities. It is arbitrary symmetric functions which can the Mercer theorem. In this paper, the radial basis function (RBF) is used as the kernel function of the LS-SVM, because RBF kernels tend to give good performance under general smoothness assumptions, since Gaussian RBF (Radial Basis Function) function is usually used as a kernel function (Wang et al., 2013),

$$k(x, x_i) = \exp(-\frac{||x - x_i||^2}{2\sigma^2})$$

(18)

Where, $\sigma$ is a positive real constant.

By Eqn.(8) to Eqn.(10), the object nonlinear model is as follows:

$$y(x) = \sum_{i=1}^{N} a_i \exp(-\frac{||x - x_i||^2}{2\sigma^2}) + b$$

(19)

The LS-SVM prediction involves two parameters to be optimized, which are $\sigma$ (the width of the Gaussian kernels which cover the input space) and $\gamma$ is viewed as regularization parameters, which controls the tradeoff between complexity of the machine and the number of non-separable points.

LS-SVM is an efficient version of these improved SVM, Instead of a quadratic programming problem in standard SVM, a set of linear equations based on KKT optimization condition are solved in LS-SVM, which can reduce the computational complexity and time for training to a certain extent (Liang et al., 2017).

### 4. SINTER ALKALINITY FORECASTS AND SIMULATION BASED ON GRAY LEAST SQUARES SUPPORT VECTOR MACHINES

#### 4.1 System Input Parameters Selection

The gray least square support vector machine is used to predict the alkalinity, aiming at this important output index. In the whole process, the variables related to the alkalinity is synthesized and make sure ten important input variables as the input of gray neural network, such as the layer thickness, the trolley speed, adding water rate of the first mixture, the mixing temperature, the content of SiO2 in the mineral, the content of CaO in the mineral, the content of FeO in the mineral, adding water rate of the second mixture, the proportion of CaO, and the proportion of coal.

#### 4.2 Sample Data Processing

Because all the collected data is often not in the same order of magnitude, the collected data are normalized to [-1 1]; this will improve the training speed of neural network. We often use the following formula to cope with the initial data (Song et al., 2018):

$$x_0 = \frac{x_0 - x_{j_{min}}}{x_{j_{max}} - x_{j_{min}}} \times 0.8 + 0.1$$

(20)

Where $x_0$ and $x_0'$ were the old and new value of the variable for a sampling point respectively, $x_{j_{max}}$ and $x_{j_{min}}$ were the minimum and maximum value of the variable in the original dataset (Song et al., 2015).

After computing using the neural network, the ant normalization processing is done to obtain the actual value of the output value forecasting.

Ant normalization of the following formula:

$$x_0 = (x_{j_{max}} - x_{j_{min}}) \times (x_0' - 0.1) + x_{j_{min}}$$

(21)

In the sampling dataset, there were inevitably some anomalies, and these data would give an impact on a certain model, and even lead to misleading. Therefore, the model data used in training samples dataset and test samples dataset was carefully selected (Song et al., 2017).
Figure 1: The prediction of the alkalinity based on gray GM (1,1)

Figure 2: The prediction of the alkalinity based on gray least squares support vector machines

Figure 1 illustrates the prediction error and relative prediction error of GM(1,1) model of the alkalinity. Figure 2 presents the fitting curve bases on gray least squares support vector machine. It can clearly seen that the predicted values are in good agreement with the desired ones in the whole ranges of time step.

From Table 1, the accuracy of the gray support vector machine reaches 1.694%, whereas the accuracy of GM (1, 1) approach only is around 12.482%. It is not difficult to see that the forecast accuracy of gray least square support vector machine is higher than that of a single GM (1, 1) model or the model of SVM, and has better robustness. Therefore, we can conclude that the gray least squares support vector machine exhibits excellent learning ability with fewer training data, the generalization capability of LS-SVM is greatly improved.
Table 2: The comparison of the alkalinity based on gray GM (1, 1) and gray least support vector machine

<table>
<thead>
<tr>
<th>NO</th>
<th>Original data</th>
<th>GM(1,1) model</th>
<th>gray least support vector machine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model data</td>
<td>Relative error/%</td>
</tr>
<tr>
<td>1</td>
<td>2.02</td>
<td>2.02</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2.01</td>
<td>1.9878</td>
<td>+1.1</td>
</tr>
<tr>
<td>3</td>
<td>1.99</td>
<td>1.9942</td>
<td>-2.11</td>
</tr>
<tr>
<td>4</td>
<td>1.98</td>
<td>2</td>
<td>-1.01</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2.0070</td>
<td>-0.35</td>
</tr>
<tr>
<td>6</td>
<td>2.01</td>
<td>2.0135</td>
<td>-0.174</td>
</tr>
<tr>
<td>7</td>
<td>2.02</td>
<td>2.02</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2.04</td>
<td>2.0265</td>
<td>-0.662</td>
</tr>
<tr>
<td>9</td>
<td>1.93</td>
<td>1.979</td>
<td>2.538</td>
</tr>
<tr>
<td>10</td>
<td>1.93</td>
<td>1.973</td>
<td>2.22</td>
</tr>
<tr>
<td>11</td>
<td>1.93</td>
<td>1.967</td>
<td>1.917</td>
</tr>
<tr>
<td>12</td>
<td>1.92</td>
<td>1.956</td>
<td>1.562</td>
</tr>
<tr>
<td>13</td>
<td>1.94</td>
<td>1.949</td>
<td>0.46</td>
</tr>
<tr>
<td>14</td>
<td>1.93</td>
<td>1.961</td>
<td>1.606</td>
</tr>
<tr>
<td>15</td>
<td>1.94</td>
<td>1.949</td>
<td>5.179</td>
</tr>
<tr>
<td>16</td>
<td>1.93</td>
<td>1.944</td>
<td>0.206</td>
</tr>
</tbody>
</table>

Average relative error/%: +12.482, 1.694

5. CONCLUSIONS

This paper has proposed a mathematics model of the alkalinity, which is realized via gray least squares support vector machine. This algorithm combines the advantages of GM(1,1) and LS-SVM. The new model fully makes use of the advantages of accumulation generation of GM(1,1) method, and weakens the effect of stochastic disturbing factor in original data series, and strengthens the regularity of raw data, and avoids the theoretical defects existing in the gray forecasting model. Besides, SVM's ability to handle high-dimension and incomplete data allows successful extraction of information even when part of the data records was missing or unreasonable owing to the problems of instrument malfunction or maintenance, calibration and climate influences, so LS-SVMs method is suitable to simulate the alkalinity in an efficient and stable way.

These results fully demonstrate the prediction accuracy of new model is superior to a single model, the theoretical analysis and simulation results are fully presented the validity of the forecast model. This shows that the gray least squares support vector machine is available for the modeling of the alkalinity, and can get better performance.

Although the proposed gray LS-SVM-based model may be superior to other modeling methods in some aspects, it has some potential drawbacks such as the underlying Gaussian assumptions related to a least squares cost function. Some researchers have made some efforts to overcome these by applying an adapted form called weighted LS-SVM, so we will intend to continue the studies on the application of the alkalinity in sintering process.

ACKNOWLEDGMENTS

Anyang science and technology research project, Based on the application of gray support vector machine in the condition of small sample and poor information; Key Laboratory of rehabilitation robot of Anyang Institute of Technology; Young backbone teachers of colleges and universities in Henan Province (2016GGJS-151); game design and evaluation system design of rehabilitation robot based on multi-sensor information fusion technology.

REFERENCES


Song Qiang. (2017): Vector Control of Permanent Magnet Synchronous Motor Based on Sliding Mode Variable Structure Control, UPB