C-Mn yield forecasting model based on SVM

Jinghang Li1,*, Juanli Bai2 and Bozhong Yu1
1School of Information and Communication Engineering, North University of China, Taiyuan, China
2Building Environment and Energy Engineering, Xi’an University of Architecture and Technology, Xi’an, China

*Corresponding author: jinghangli2019@nuc.edu.cn

Abstract. With the rapid development of economy, how to increase steel output and reduce plant cost is of great significance. Deoxy alloying is an important link in iron and steel smelting. This paper aims at the optimization of the burdening scheme of deoxy alloying of molten steel. The aim is to establish the SVM optimization model based on the data and predict the yield of C/Mn. First of all, we established a time series model based on the obtained results of the problem and the furnace times as the time, and used the ARIMA time series method in SPSS software to predict the yield of C/Mn. Secondly, we improve the model, use the historical data to train and forecast the support vector machine model (SVM), and compare the predicted values of the SVM model and the time series model, and get the conclusion that the SVM model is more accurate than the time series model.

Keywords: ARIMA, SVM.

1. Introduction
As an important pillar industry of China's national economy, iron and steel smelting industry is the basis of a country's heavy industry and also an important part of national economic production. Iron and steel smelting plays an important role in China's economic development[1]. Because the steel output plays a very important role in the national economy, and the deoxidation alloying is an important link in the iron and steel smelting, it is very important to establish a reasonable model to predict the C/Mn yield in the process of steelmaking[2].

2. Model Establishment

2.1. ARIMA Model
ARIMA model is a kind of dynamic data processing model, which mainly studies the change of time and the law of development and change of things, and predicts the future trend[3]. In this paper, the historical harvest rates of C and Mn elements were predicted, and the results were shown in Figure 1 and 2. The time series is analyzed by randomness, stationarity and other factors.
2.2. SVM Model

Support Vector Machine (SVM) is a pattern recognition method based on statistical learning theory proposed by the research group of AT&T Bell Laboratory led by Vapnik in 1963. It can obtain good generalization ability based on limited sample training and has been widely used in pattern recognition, regression analysis and other fields. The emergence of support vector machine (SVM) provides a new method to solve the modeling problem of strongly nonlinear systems. At present, in the field of iron and steel metallurgy, support vector machine is also widely used gradually[4].

Let the input and output sample data set of a regression system be \( \{ (x_i, y_i) \}_{i=1}^N \). \( x_i \) is the input variable of \( n \)-dimensional system (\( x_i \in \mathbb{R}^n \)), and \( y_i \) is the output variable of system (\( y_i \in \mathbb{R} \)). Therefore, the SVM regression model can be expressed as:

\[
    f(m) = \phi(m)\omega^T + b
\]

In the above equation, \( \phi(m) \) is the mapping function that maps the input space to a high-dimensional feature space; \( \omega \) is the weight vector of the hyperplane (in \( \mathbb{R}^n \)); \( b \) is the offset quantity. The solution obtained through the minimum regularization risk functional is as follows:
The constant $C$ is called the penalty coefficient, $L$ represents the loss function. In order to obtain the sparse solution and improve the calculation speed of the test samples, the $\varepsilon$ insensitive loss function is selected:

$$L \times \varepsilon(f(m) - n) = \begin{cases} 0 & f(m) - n < \varepsilon \\ f(m) - n - \varepsilon & \text{other} \end{cases}$$

Minimizing $R(\omega)$ leads to $\omega = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \times \phi(m_i)$, where $\alpha_i$ and $\alpha_i^*$ are solutions to the minimizing $R(\omega)$ duality problem. Substituting $\omega$ into Equation (1), we get:

$$f(m) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \times \langle \phi(m_i), \phi(m) \rangle + b = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \times P(m_i, m) + b$$

Where $P(m_i, m) = \langle \phi(m_i), \phi(m) \rangle$ is called the kernel function, which is a symmetric function of positive real numbers and also satisfies Mercer's conditions. Common kernel functions include radial basis kernel function and polynomial kernel function. Here, we use radial kernel function:

$$P(m_i, m) = \frac{|m_i - m|^2}{\sigma^2}$$

In the model, the parameter penalty coefficient $C$ and the kernel width $\sigma$ can be obtained through cross test, and finally the support vector prediction model is obtained. Then we wrote the program of the model, and ran the program with Matlab, and finally got the predicted results and images.

Figure 3. SVM predicted the C yield of HRB400B
Figure 4. SVM predicted the Mn yield of HRB400B

According to the results obtained, we can see that the predicted value still changes within a stable range, the same as the time series model. However, we cannot intuitively see the advantages of the improved model, so we compare the two.

Figure 5. Comparison of the yield of C element predicted by support vector machine and time series

Figure 6. Comparison of the yield of Mn element predicted by support vector machine and time series
In the end, it can be seen that the results predicted by support vector machine are more accurate than those predicted by time series, and are closer to the actual value.

3. Conclusion
1. The determination of transcriptional information should be as accurate as possible, especially the transcriptional endpoint temperature, transcriptional focus C, transcriptional endpoint P, and transcriptional endpoint Mn, which will have a great impact on the predicted value of the yield rate.
2. The steel mill can use the model and algorithm provided in this paper to get a reasonable burdening scheme, but due to the complex deoxidation and evolution process, many factors affecting the yield can not be fully considered, so the method in this paper also has some errors.

References
[1] Zheng Zhiqiang. Analysis on the Development Status and Existing Problems of China's Iron and Steel Smelting Industry [J]. Guide of Sci-tech Magazine, 2013,(20) : 312,319.
[2] Fan Haiwen, Feng Ziteng, Yu Runle.Optimization of the Scheme Cost of "Deoxidized Alloying" in Steel and Water[J].Yunnan Chemical Technology,2020,47(2):45-46. DOI:10.3969/j.issn.1004-275X.2020.002.017.
[3] Chehelgerdi-Samani, M.a,b;Safi-Esfahani, F.a,b.PCVM.ARIMA: predictive consolidation of virtual machines applying ARIMA method[J].Journal of Supercomputing,2020.
[4] Zhao, T.;Chen, X.;Yang, L..IPCA-SVM based real-time wrinkling detection approaches for strip steel production process(Article)[J].International Journal of Wireless and Mobile Computing,2019,Vol.16(No.2): 160-165.