Prediction model for three rate values of raw materials for grinding based on xgboost

Yifan Wang [1], Xiaohong Wang [1] , Hongliang Yu [1],
Zhonghua Li [1], Liqing Deng [1]

[1] School of Electrical Engineering, University of Jinan, Jinan,China

Abstract. In cement production, the quality of raw meal formulation directly impacts cement quality and kiln system stability. Traditional methods often result in inconsistent raw meal quality due to nonlinearity, time delays, and material composition fluctuations.This study develops an XGBoost-based prediction model for three key parameters of raw meal (KH, SM, IM), using mill current, pressure differential, feed rate, and outlet temperature as inputs to accurately establish nonlinear relationships between process parameters and these quality indicators.Experimental results demonstrate the model's high prediction accuracy and strong generalization capability, providing reliable data support for real-time raw meal formulation adjustments. This approach significantly contributes to production stability enhancement and intelligent manufacturing advancement in the cement industry.

Keywords: XGBoost, three key parameters of raw meal, raw meal formulation.

1. Introduction

In the entire production process of cement, raw material batching is a very important part. Because cement production is a continuous process, each production link is interrelated, and the previous link will have a certain impact on the subsequent production links. Therefore, stable quality of cement raw materials is necessary to produce high-quality cement, and high-quality cement raw materials can also provide stable production conditions for kiln firing operations.

In the process of raw material preparation, many scholars have conducted relevant research to meet the demand of enterprises to improve the qualified rate of raw materials for grinding: In reference [1], an improved genetic algorithm was used to solve the optimal mixing scheme for cement raw materials, in order to achieve optimal control of raw material quality, in response to the many information un certainties caused by unstable raw material quality and significant lag in the detection process. Ref erence [2] established a raw material calcium oxide measurement model based on LS-SVM. A raw material batching optimization expert system based on an offline analyzer was established with the objective function of minimizing the difference between the measured values of the model and the production target values. Reference [3] established an RBF neural network clinker quality prediction model to predict the quality of clinker calcined under current kiln conditions, in order to solve the problem of adjusting the target value of raw material proportioning rate based on offline clinker qual ity. Reference [4] established an objective function based on the characteristics of the raw material rate control system, which minimizes the difference between the actual rate value and the ideal rate value. The taboo search algorithm of LINGO optimization software was used to solve it. The exper imental results showed that this scheme can quickly obtain the ratio adjustment value. Reference [5] addresses the problem of complex and fluctuating raw material composition in the raw material pro portioning process, and the inability of manual calculation of proportioning methods to meet produc tion needs. Based on the characteristics of solid waste raw materials, a non dominated sorting algo rithm was developed to optimize the proportioning of all waste cement raw materials.

Although the raw material preparation process only involves the mixing and grinding of raw ma terials, the raw material batching is carried out in a mill, which takes a long time to grind and also requires transportation on a belt scale for a period of time. Due to the nonlinearity and long-term lag of the system, as well as the influence of uncertain factors, many uncertain factors have been brought to the raw material quality control system. Therefore, this study proposes a grinding rate prediction method based on Bayesian optimization extreme gradient boosting (XGBoost). Through the efficient gradient boosting framework and regularization mechanism of XGBoost, the complex nonlinear relationship between multi-dimensional features such as raw material characteristics and process pa rameters and rate values is accurately modeled. Constructing prediction models for three key rate values of limestone saturation coefficient (KH), silica rate (SM), and alumina rate (IM), Provide di rect data-driven basis for real-time and precise adjustment of proportioning.

2. Analysis Of Raw Material Batching Process

The production of new dry process cement can be simply summarized as "two grinding and one burning". This production process is usually divided into three stages: raw material preparation, clinker calcination, and the addition of other auxiliary materials after the cement clinker cools down. After cement grinding, it is made into cement. The main process flow of cement production is shown in Figure 1:

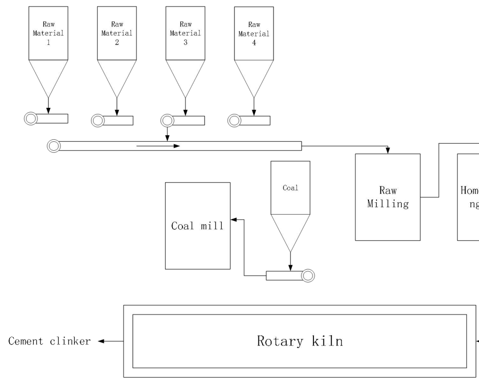

Fig. 1.Cement production process flow chart

During the raw material preparation stage, cement companies meet the quality requirements of raw materials by controlling the cutting ratio of raw materials. The feeding controller gives instruc tions to the belt scale to control the output of four types of raw materials in a certain proportion. Transported to the raw material mill through a belt scale for grinding, the ground cement raw material is transported to the homogenization silo through an air chute and an elevator.

3. Variable Selection and Data Feature Engineering

3.1 Selection of modeling variables for predicting the three rate values of raw materials for grinding

The main goal in the raw material preparation process is to ensure the qualified and stable quality of the ground raw materials. Although the raw material preparation process only involves the mixing and grinding of raw materials, the lag in testing and production has brought many uncertain factors to the quality control system of cement raw materials. Taking into account the above issues, key parameters that can characterize changes in raw material composition can be identified from the mill, and the three rate values of the raw material can be predicted. This not only reduces the lag effect of detection on the batching process, but also increases the frequency of batching adjustments, thereby achieving optimized control of the raw material batching process. The following are the selections of variables:

Mill current:The fluctuation of mill current directly reflects the amount of material inside the mill. When the grindability of raw materials deteriorates (such as high crystallinity of limestone and strong plasticity of clay), the load on the mill increases and the current rises, which is associated with KH fluctuations.

Grinding machine pressure difference:The grinding machine pressure difference (inlet pressure and outlet pressure difference) is mainly affected by the filling rate of materials inside the grinding machine and the ventilation rate. Affects silicon rate (SM) and aluminum rate (IM).

Raw material feeding amount: determines the stability of material ratio. If the feeding amount fluctuates (such as belt scale failure or material agglomeration), it will cause the actual proportion of raw materials entering the mill to deviate from the preset value, further exacerbating the fluctuation of the three rate value.

Grinding temperature: The fluctuation of grinding temperature can roughly reflect the quality of limestone and the content of calcium oxide in the raw material composition, thereby affecting the representativeness of KH

3.2 Feature engineering of data

Before conducting predictive modeling, it is necessary to perform feature processing on the data to meet the required data requirements. Its main tasks include feature alignment, filling missing values, and handling outliers.

Feature alignment: Due to the lag in detection and production, there is a time interval between the detection of the three rate value of the raw material and the related variables of the mill. In order to eliminate the impact of this lag, it is necessary to perform time offset, nearest neighbor matching, and feature extraction on the input variables.Avoid combining SI and CGS units, such as current in am peres and magnetic field in oersteds.This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.

Missing value filling: In response to the phenomenon of missing data in industrial sensor data acquisition systems caused by equipment noise, communication interruptions, or instantaneous faults, this article uses linear interpolation for processing and replaces missing values with the average be fore and after.

Exclude outliers: Due to the complex and ever-changing environment during the production pro cess, it may lead to abnormal data collection. For abnormal data, this article uses the IQR method (interquartile range method) to identify and eliminate outliers.

Through the preprocessing process of feature alignment, missing value filling, and outlier removal mentioned above, the data will be more complete and meet the modeling requirements, thereby im proving the accuracy and stability of predicting the three rate values of raw materials for subsequent grinding.

4. Prediction of Three Rate Values of Grinding Raw Materials Based on XGBoost

4.1 Principle of XGBoost Algorithm

XGBoost is an ensemble algorithm based on gradient boosting trees, belonging to supervised learning algorithms. This algorithm integrates the prediction results of multiple weak classifiers (de cision trees) through a series of training processes to achieve a strong classifier (a collection of mul tiple decision trees). This algorithm can avoid overfitting by adding regularization terms, terminating training in advance, and pruning, while improving algorithm efficiency through multi-threaded par allel computing. Compared with traditional decision tree models and SVM algorithms, it has higher accuracy. The specific algorithm steps are as follows:

In XGBoost, people need to be added to each tree one by one in order to improve the effect:

$$\hat{y}_i^{(0)} = 0; \hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i); ...; \tag{1}$$

$$\hat{y}_i^{(t)} = \sum_{k-1}^{t} f_k(x_i) = \hat{y}^{(t-1)} + f_t(x_i) \tag{2}$$

If there are too many nodes in the leaves, the risk of overfitting in the model increases. So adding a penalty term to the objective function to limit the number of leaf nodes:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^{\,2} \tag{3}$$

In the formula, represents the severity of punishment; The number of leaves; The weight of leaf nodes.

The complete objective function is:

$$Obj^{(t)} = \sum_{j=1}^{n} l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) + \Omega(f_t) \tag{4}$$

Find the optimal solution of the objective function:

$$\hat{L}^{(i)}(q) = -\frac{1}{2}\sum_{j=1}^{T} \frac{(\sum_{i\in I_j} g_i)^2}{\sum_{i\in I_j} h_i + \lambda} + \gamma T \tag{5}$$

Formula (5) can be used as the cotyledon score of a tree, and the structure of the tree becomes excellent as the score increases. Once the split result is less than the maximum value of the given parameters, the algorithm will stop increasing the depth of the cotyledons.

Therefore, in order to improve the accuracy of predicting the three rate values of raw materials, the XGBoost prediction model is adopted. Firstly, the data is preprocessed and divided into a training set and a testing set. Then, the objective function loss is minimized through training. Finally, the optimal parameters are used to predict the testing set.

4.2 Construction and Analysis of Three Rate Prediction Model for Grinding Raw Materials

Variable selection and data preprocessing: Based on the process analysis in the previous section, the selected input variables are mill current (A), mill pressure difference (P), raw material feeding amount (B), and outlet temperature (T), and the output variable is the three rate value of the raw material (R). Due to significant fluctuations in the detection data collected directly from the site. This is because the industrial site environment is complex, and various sensors are affected by interference signals, resulting in abnormal values in the detected data. We need to preprocess the data. This mainly includes filling in missing data values and removing outliers. And in order to eliminate dimensional differences, the minimum maximum normalization method is used to preprocess the raw data.

The processed data is shown in Figure 2

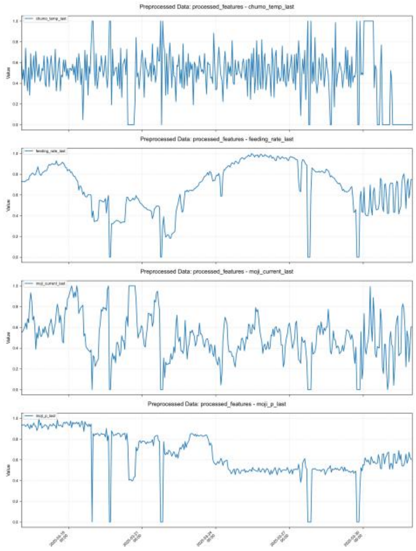

Fig. 2.Preprocessed data sample graph

Dataset partitioning and evaluation metrics: This article divides 80% of the training set and 20% of the testing set in chronological order; Evaluate performance using MAE (Equation 6), RMSE (Equation 7), and R ² (Equation 8):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| y_i - \hat{y}_i | \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)} \tag{7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{8}$$

Model parameter selection: For the prediction modeling of the three rate values of raw materials for grinding, if the designed network model is too complex, it may lead to overfitting, while insuffi cient training can easily cause underfitting. After experimentation, the selected hyperparameters are shown in Table 1.

Table 1. XGBoost model parameters

Parameter	Unoptimized
Learning Rate (eta)	0.1
n_estimators	100
max_depth	6
subsample	0.8
colsample_bytree	0.8
objective	'reg:squarederror'

Among them, Learning Rate (eta) controls the weight contribution of each tree, n_estimators is the number of trees, x_depth is the maximum depth of a single tree, subsample is the sample sampling ratio, colsample-bytree is the feature sampling ratio, objective is the selected objective function, and square error regression is used here, which is suitable for regression problems.

4.3 Prediction of Three Grinding Rates and Analysis of Feature Importance

Based on the parameter results shown in Table 1, an XGBoost model was constructed using the selected input and output variables to predict the three rate values of grinding materials. The predicted results are shown in the figure 3, which shows that the XGBoost model's predicted values are highly consistent with the true values, indicating that the model has excellent fitting ability in the task of predicting the wear rate values.

According to the importance display of the features in the figure, they are sorted in order of im portance from top to bottom, with higher values having a greater impact on the prediction results. It can be seen that the grinding temperature is directly related to the degree of raw material calcination reaction and is the core environmental factor for the formation of the three rate value. If the temper ature is too high or too low, it will change the proportion of mineral phase generation and significantly affect the grinding three rate value.

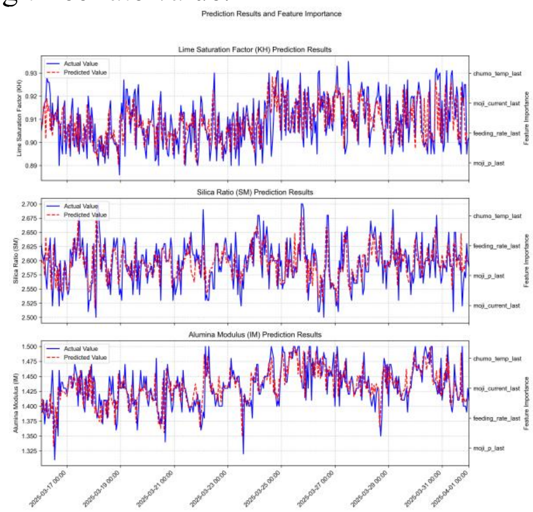

Fig. 3.Prediction Results of Grinding Three Rate Values

To verify the generalization ability of the XGBoost model, this study tested its performance in different scenarios through simulation experiments. The experiment used data from different time periods that did not participate in training as the validation set, and the prediction results are shown in the figure. The figure 5 shows that the XGBoost model's predicted values on the validation set are highly consistent with the true values, indicating that it not only performs well on the test set, but also has strong generalization ability, making it suitable for predicting the wear rate values in industrial environments.

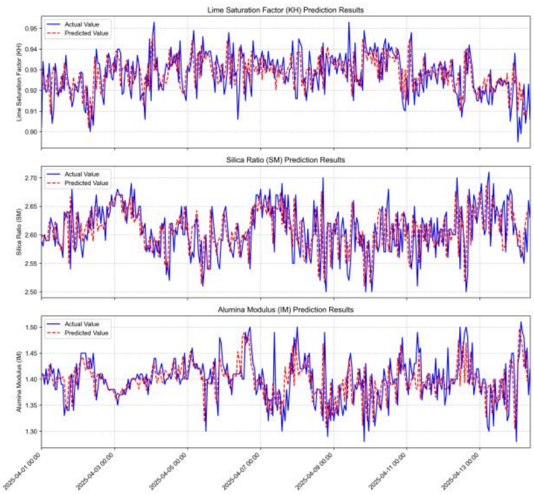

Fig. 4.Independent dataset validation result chart

5. Conclusion

This study proposes a dynamic prediction method for the output three rate values (KH, SM, IM) of cement raw materials based on the XGBoost algorithm. The core input variables are mill current, mill pressure difference, output temperature, and feed rate, and a closed-loop system of "prediction batching regulation" is constructed to solve the industrial problems of "parameter lag response and poor multi variable synergy" in traditional cement raw material batching. The experimental results show that the prediction model based on XGBoost achieves a relatively accurate prediction of the three rate values of raw materials on the test set, with the true values and test values being generally mild. To verify its generalization ability, this study conducted simulation experiments using data from different time periods. The results showed good consistency between the predicted values and the true values, indicating that the model is suitable for dynamic prediction tasks in industrial scenarios. Therefore, it proves that this method enhances the non-linear modeling ability of the gradient boosting tree, accurately captures the dynamic mapping relationship between the operating parameters of the mill and the three rate values, and provides data-driven and effective decision support for intelligent cement batching.

References

[1] Wan Xin, Yu Hongliang, Wang Xiaohong Method for Quality Control of Cement Raw Materials Based on Typical Operating Conditions [J]. Journal of Jinan University (Natural Science Edition), 2015, 29 (02): 88-93. DOI: 10.13349/j.cnki.jdxbn.2015.02.002

[2] Xu Lei Research on Cement Raw Material Proportioning System Based on Multi Objective Optimization [D]. Jinan University, 2020 DOI:10.27166/d.cnki.gsdcc.2020.000332.

[3] Lu Xinzhong Research on Optimization of Cement Raw Material Proportioning Based on Quality Prediction of Clinker[D]. Jinan University, 2019 DOI:10.27166/d.cnki.gsdcc.2019.000049.

[4] Zhang Zhigang, Lin Xiaofeng, Wang Zhangyun Optimization control system for cement raw material batching process [J]. Computational Technology and Automation, 2007, (01): 30-32

[5] Song Xiaoling, Xu Panpan, Hu Jingping, etc Research on Optimization of Raw Material Proportioning for Full Waste Cement Based on Biomimetic Algorithm [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2022, 50 (10): 45-50. DOI: 10.13245/j.hurst. 221007