International Conference on Applied Energy 2021
Nov. 29 - Dec. 5, 2021, Thailand/Virtual
Paper ID:464

Research on Data Preprocessing of Cement Clinker Firing Process Modeling

Zongliang Ma, Ping Jiang*, Dongyang Han*

University of Jinan

University of Jinan, Information Center of China Building Materials Industry

ABSTRACT

In order to reduce the energy consumption of cement clinker firing process and achieve green energy-saving production, a high-precision, strong and stable process model is urgently needed. However, in the process of modeling cement clinker firing process with data-driven modeling method, in addition to the objective difficulties of multivariable, nonlinear, large delay and strong coupling of the firing process itself, and the data collected from cement production sites can also make modeling more difficult due to its complexity, repeatability, and incompleteness., in order to make more effective use of the information contained in the data and obtain the cement clinker firing process model with higher accuracy and stability. This paper proposes a series of data preprocessing steps for the raw data, which can remove the redundant information in the data and make the modeling process more efficient and accurate. Data preprocessing includes time domain unification, abnormal data elimination, data filtering and principal component analysis.

Keywords: data preprocessing, abnormal data elimination, data filtering, principal component analysis, process modeling.

1. INTRODUCTION

Cement, as the most important basic building material, has a special position in the field of human production and life. Meanwhile， since the coal consumption cost accounts for 50% ~ 55% of the total cost of cement clinker firing, reducing the coal consumption in the process of cement clinker firing becomes an important direction of cement clinker firing process optimization under the condition of ensuring the quality of cement clinker.

In order to achieve the multi-objective optimization control of clinker quality and coal consumption in the process of cement clinker firing, a process model with high accuracy and good stability has become an important prerequisite and guarantee to achieve the optimization objectives. However, complex production sites and potentially problematic production equipment can affect the data collected from the site, which may contain invalid information and lead to a decrease in the accuracy and stability of the final model. Therefore, it is necessary to preprocess the raw data to reduce noise interference and improve the quality of data, so as to establish accurate and stable model. Finally, according to the model, the multi-objective optimization control of cement clinker firing process is completed, and the green, energy saving, efficient and high quality production is realized.

2. DATA PREPROCESSING

2.1 Unified time domain

In cement clinker production, the DCS (Distributed Control System) collects process parameters at a frequency of 1 second, while the change frequency of optimization objectives and decision variables selected in process modeling is hour level. Too high input frequency of sample data may lead to highly overlapping information contained in the input samples, which is not conducive to the establishment of the model. Therefore, it is necessary to unify the time domain of data into hours, that is, take the mean value of data collected every hour as the benchmark sample, and retain the most information with a smaller sample number.

2.2 Elimination of abnormal data

In the process of field data collection, improper operation of operators and abnormal field equipment will make the collected data appear gross error, which seriously distort the implied information and the reliability of the data. Therefore, only by eliminating gross error can we get better modeling effect. In this step, neither the gross error in the initial data can be

Selection and peer-review under responsibility of the scientific committee of the 13_{th} Int. Conf. on Applied Energy (ICAE2021).

retained, nor the normal data can be misjudged as gross error and removed by mistake, which will both increase the unreliability of the data manually. The commonly used methods with sufficient theoretical support and good elimination effect are Pauta criterion and Grubbs criterion.

2.2.1 Pauta *criterion*

The Pauta criterion is based on the confidence interval of 99.7%, using the value of three times the standard deviation as the determination boundary to separate the normal random error from the gross error[2].When using the Pauta criterion, the mean value $\bar{x}$ and residual error $\gamma_i = x_i - \bar{x}$ of sample data$X_i(i = 1，2，\cdots，n)$ should be calculated first, and the standard deviation S of the measurement column should be calculated by the Ambesier formula. If the residual error γ_i of the data X_i satisfy formula (2.1), X_i is considered to be an outlier containing a large error, which needs to be removed.

$$|\gamma_i| > 3S \qquad (2.1)$$

Although Pauta criterion is widely used in scientific research and engineering, it can only be used when the sample size N>10. Otherwise, the criterion may not be able to find the gross error data from the sample[3].

2.2.2 Grubbs criterion

The Grubbs criterion [4] gets the critical coefficient $G(n，a)$ according to the critical value table of the Grubbs criterion after determining the sample data capacity N and significance level a, and arranges the samples from small to large. Then , it calculates the Grubbs values of the maximum and minimum values according to the formula (2.2).

$$G_n = \frac{|\overline{X} - x_n|}{S} \qquad (2.2)$$

Where x_n is the value of the nth sample data in the data sample. $\overline{X}$ is the arithmetic average of data samples. S is the sample standard deviation. The discriminant rules for data elimination are as follows.

When $G_1 \geq G_n$ and $G_1 < G_0$, x_1 is an outlier.

When G_1<G_n and $G_n > G_0$, x_n is an outlier.

When $G_1 < G_n$, and $G_n < G_0$, there are no outliers.

The Grubbs criterion has a good discriminant effect on the data with small sample size[5]. However, it's limitations are also obvious. First, The Grubbs criterion requires that the sample be normally or approximately normally distributed. Second, when using The Grubbs criterion to judge data outliers, it is necessary to specify the statistical critical coefficient $G(n，a)$ in advance, which is obtained by querying table of Grubbs criterion critical values, but in the table clearly limits the sample size n up to 100, greatly limits the Mrs Grubbs criteria on the application of large-scale data.

2.2.3 Eliminate abnormal data summary

The data of secondary air temperature, one of the decision variables in the sintering process of cement clinker, is preprocessed by using two abnormal data elimination methods, and the discriminant effects of Pauta criterion and Grapbs criterion on outliers are analyzed by taking 50 and 100 sample size data as examples.

The original data with a sample size of 50, the data removed by using Pauta criterion and the data removed by using Grubbs criterion are shown in Fig 1.1. The original data with a sample size of 100, the data removed by using Pauta criterion and the data removed by using Grubbs criterion are shown in Fig 1.2.

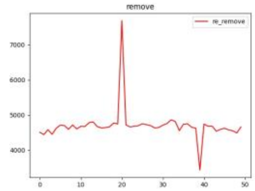

Fig. 2.1 Pre-processing data

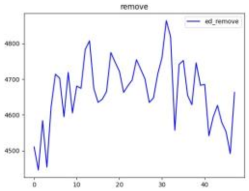

Fig. 2.1. Processed data(P)

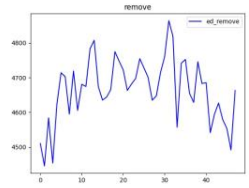

Fig 2.1 Processed data(G)

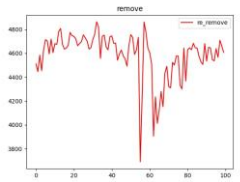

Fig. 2.2. Pre-processed data

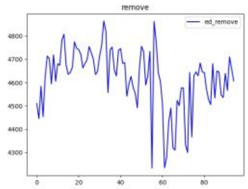

Fig 2.2 Processed data(P)

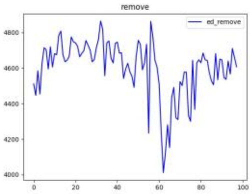

Fig. 2.2. Processed data(G)

Through the analysis of the error discrimination results, it can be concluded that when the sample size is small, the Pauta criterion and the Grubbs criterion have similar accuracy, and both can monitor abnormal data to ensure the reliability of data. However, when the sample size increases, the accuracy of Grubbs criterion begins to decrease, while the accuracy of Pauta criterion can still maintain a high accuracy. In the process of data-driven modeling, thousands of samples need to be preprocessed, which may reduce the accuracy and effectiveness of the Grubbs criterion for abnormal data monitoring. In conclusion, when outliers are removed from samples with large sample size, the Pauta criterion can provide better removal effect of gross error.

2.3 Data filtering

In the process of digitization of decision variables in cement production site, periodic or aperiodic interference may be superimposed on the analog signal due to mutual interference between circuits, which will be attached to the data, resulting in error values, redundant values and missing values appear in the sample data. These problems will blur the data characteristics of sample data and reduce the security and reliability of data. Therefore, it is necessary to use data filtering method to eliminate the mixed noise data in the original data, obtain the sample data containing more real information, and improve the representativeness of data[6].

In model fitting of sample data, the deterministic components and the uncertain components in the data are generally separated, and the random fluctuation in the data is eliminated by adjusting the hyperparameter of the model to fit the data set. However, considering the complex characteristics of modeling objects such as nonlinear, strong coupling and large delay, in order to improve the accuracy and reliability of data-driven modeling, smoother data is needed, so data filtering is required to process the data. Considering the fluctuation degree of decision variables and data characteristics in the data sample, the limiting filtering method and the moving mean filtering method are generally used.

2.3.1 Limiting filtering

According to the production experience, determine the possible maximum deviation Δ of adjacent sampling data, and then subtract the two adjacent data sampling values x_i and x_{i+1} in the data that needs to be filtered, compare the absolute value Δ_i of the difference with the maximum deviation Δ.

If $\Delta_i \leq \Delta$, take x_{i+1} as the sampling value.

If $\Delta_i > \Delta$, take x_i as the sampling value.

It can effectively overcome the pulse interference caused by accidental factors. But it can not suppress periodic disturbances.

2.3.2 Moving average filtering

For a continuous sample of N values, consider it as a fixed queue of length N. New data from each sample is placed at the end of the queue and the original data at the head of the queue is discarded[7]. According to the first-in and first-out principle, the new filtering values can be obtained by arithmetic average operation of N data in the queue. The moving average *filtering* is simple and has a small amount of calculation, especially in the form of recursive calculation, which can save storage units and facilitate real-time data processing [8].

However, due to low sensitivity and poor suppression of occasional pulse interference, it is not suitable to eliminate the sampling value deviation caused by pulse interference, and it is not suitable for the occasion where pulse interference is more serious.

2.3.3 Summary of data filtering

The data with a sample size of 1000 were selected, and two filtering methods were used to filter the samples respectively. The sample processed by amplitude limiting filtering is shown in Fig 2.3. The data processed by moving average filtering are shown in Fig 2.4.

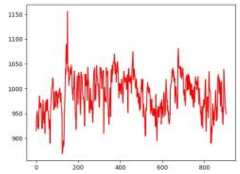

Fig.2.4 Pre-processing data

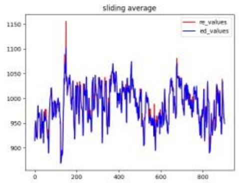

Fig. 2.3. Contrast figure

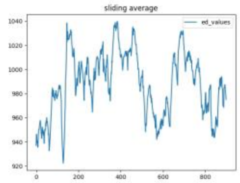

Fig. 2.4. Processed data

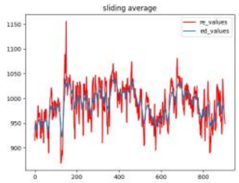

Fig. 2.4. Contrast figure

Although we can optimize the filtering effect by adjusting the maximum deviation Δ and overcome the pulse interference caused by accidental factors, the periodic interference on the sample is difficult to be suppressed by means of amplitude limiting filtering.

The moving average filter has a good filtering effect on periodic interference and has better smoothness. The processed data will improve the confidence and accuracy of the training network, and get better generalization effect when the real data is fitted. However, the numerical deviation caused by pulse interference should not be eliminated.

In conclusion, we can analyze the main interference types according to the characteristics of the samples, and flexibly use amplitude limiting filter and moving mean filter to preprocess the data, so as to obtain more reliable data.

2.4 Data analysis

Due to the complex working conditions and many process parameters in the calcination process of cement clinker, the complexity of modeling problems is increased when modeling data based on sample data. In view of this problem, principal component analysis is proposed to obtain the main information in the sample data, reduce the dimension of input data and the calculation difficulty in data-driven modeling, so that a more stable and accurate model will be obtained.

2.4.1 Principal component analysis

As a common data analysis method, *Principal component analysis* is often used for dimensionality reduction of high-dimensional data and can be used to extract the main feature components of data[9]. Singular value decomposition covariance matrix is usually used to reduce data dimension. That is, the eigenvalues and eigenvectors of the covariance matrix of the sample data are calculated by SVD, and the eigenvalues are sorted from large to small, the corresponding eigenvectors of the first K eigenvalues are selected to form the eigenvector matrix, so as to realize the transformation of the data from the original data to the new data. At the same time, the main information is extracted and the dimension of sample data is reduced [10].

2.4.2 Summary of principal component analysis

A 23-dimensional sample composed of 23 parameters in the firing process of cement clinker was selected for principal component analysis, and the contribution rate parameter was determined to be 0.95, that is, the new data should contain 95% of the original data. The original data and new data are shown in Fig 2.5.

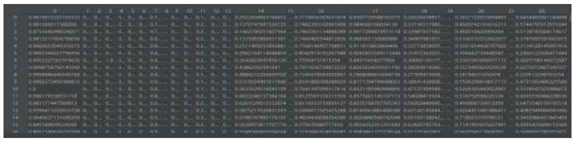

Fig. 2.5. original data(a)

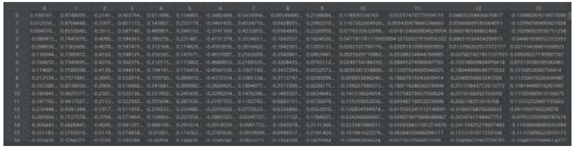

Fig. 2.5. new data (b)

As can be seen from the figure, after the contribution rate is determined, the 23-dimensional data is converted into 13-dimensional data, which greatly reduces the dimension of data and the computational complexity of process modeling.

2.5 Conclusion

First of all, unified time domain can process the initial data and make the data conform to the actual engineering requirements. Secondly, The Pauta criterion can eliminate the gross errors in the samples. Then, the periodic and aperiodic interference in the data sample can also be offset by moving mean filtering and amplitude limiting filtering. Finally, principal component analysis can reduce the computational complexity of modeling. In summary, data peprocessing can make the data smoother, the reliability and authenticity of the data higher, and the model fitting effect better, so as to obtain a model with high accuracy and good stability.

ACKNOWLEDGEMENT

I would like to thank my supervisor, Ms. Jiang Ping, for her careful guidance on my thesis and help in my life, so that I can concentrate on my study and research.

REFERENCE

[1] Lin Honghua. Robust Processing Method for Eliminating Abnormal Data [J]. Journal of China Jiliang University,2004(01):22-26

[2]Zhang Min, Yuan Hui. PauTa Criterion and Outlier Elimination [J]. Journal of Zhengzhou University of Technology,1997(01):87-91.

[3] Xu Chenhui, Ma Minghui. Journal of Shanghai university of engineering science,2018,32(01):64-67

[4] Shao Ting-ting, Zhang Shui-shui, Zhang Yong-bo. Modern electronic technique,2008,31(24):148-150

[5] Xiong Yanyan, Wu Xianqiu. Comparison and application of four criteria for gross error [J]. University physics experiment,2010,23(01):66-68

[6] Wang Qinghe, Wang Qingshan. Several Common Digital Filtering Algorithms in Data Processing [J]. Metrology Technology,2003(04):53-54

[7] Pei Yixuan, GUO Min. Basic Principle and Application of Moving Average Method [J]. Journal of Artillery Launch and Control,2001(01):21-23.

[8] Jiang Yanhua, WANG Yanwen. Application of moving average filter in harmonic and reactive current detection [J] Journal of lntu. Natural science,2014,33(12):1685-1688.

[9] Zhang Peng. Research on Comprehensive Evaluation Based on Principal Component Analysis [D]. Nanjing University of Science and Technology,2004.

[10] Xiaoxiao Han, Yaohui Zhang, Fujun Sun, Shaohua Wang. Journal of sichuan ordnance engineering,2012,33(10):124-126.