

Composition analysis of glass relics based on a clustering model

Zhuo Liu¹, Xinxin Di², Qi Wang², Lei Wang^{1,3,*}

¹College of Artificial, Tianjin Normal University, Tianjin, China

²School of Mathematical Sciences, Tianjin Normal University, Tianjin, China

³Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin, China

*Corresponding author: wanglei2014@tju.edu.cn

Abstract: The analysis and identification of the chemical composition of ancient glass are of great significance for the evolution of history through time and space. The chemical content of ancient glass will change due to weathering, so a comprehensive evaluation model for the chemical content of the ancient glass is proposed in this paper. In this paper, the chi-square test is used to analyze the four classification variables, and it is found that there are significant differences between the ornamentation and colour, and the chemical composition content before weathering is predicted. Then the system clustering is used, and a subclass division model is proposed for the clustering results. Finally, the sensitivity analysis of the classification results is carried out, and different degrees of disturbance items are added respectively to obtain different model accuracy rates. The results show that the sensitivity of the model is better. The results show that the model can be used to analyze the composition of ancient glass relics and achieve good results.

Keywords: Chi-square test, cluster analysis, entropy weight prediction model, sensitivity analysis of subclass division.

1. Introduction

To analyze and identify the composition and classification of glass cultural relics[1], since the opening of the Silk Road, glass has flowed into China from West Asia, and the production process of glass in China has been developing continuously at various times while retaining the characteristics of various times, which is an important source of our national self-confidence and cultural self-confidence. Therefore, the analysis and identification of the chemical composition of ancient cultural relics are of great significance for inheriting the excellent traditional culture of the nation and cultivating the foundation of cultural self-confidence. Against this background, this paper discusses the correlation among the subclassification, surface weathering and chemical composition of ancient cultural relics, and analyses and identifies the composition and classification of glass cultural relics. Based on this, this paper mainly uses the methods[2] of statistical analysis, cluster analysis and correlation analysis to answer the above questions.

Firstly, the chi-square test was used to analyze the four categorical variables[3], and it was found that there were significant differences between surface weathering and ornamentation and colour. To study the statistical law of the change of chemical composition content of different types of glass before and after weathering, the chemical composition content figures of high-potassium glass and lead-barium glass with and without weathering were drawn for comparative analysis, and it was found that the main chemical composition content of high-potassium glass decreased after weathering, while that of lead-barium glass increased after weathering. Then, this paper establishes the prediction model of chemical composition content before weathering based on the entropy weight method, uses the entropy weight method to weigh, uses the increase (decrease) ratio of each chemical composition before and after weathering, and finally predicts the chemical composition content before weathering through Matlab solution.

Then, the hierarchical clustering method is used to establish the subclass division model[4] based on standard deviation, and the K-means algorithm is used to cluster analysis for different types,

different weathering States and a K value of 2. The clustering results were tested for differences, and the appropriate indicators were determined according to the significant P value for subclassification. Finally, the sensitivity analysis of the classification results is carried out, and different degrees of disturbance items are added to obtain different model accuracy rates, and the results show that the sensitivity of the model is better. By setting 1%, 2%, 5%, 10%, 20% and 30% disturbance to the data value of the variable, the data after disturbance is compared with the undisturbed data, and finally, the accuracy of the model is 100%, which shows that the model has good stability and universality.

The absolute weathering rate of different glass types[5], different ornamentation types and different colours. Secondly, the entropy-TOPSIS model is established to analyze the relationship between the surface weathering of glass relics and their glass types, ornamentation and colours.

2. A prediction model based on chi-square test

2.1 Data preprocessing

We take the absolute value of the difference between the content of iron oxide and copper oxide corresponding to the cultural relic[6] with unknown colour and the content of iron oxide and copper oxide corresponding to other cultural relics with weathered lead-barium glass and find the colour of the cultural relic with the minimum difference as the filling data, that is, the requirement.

$$\min\{|x_{m1} - x_{n1}| + |x_{m2} - x_{n2}|\} \tag{1}$$

Where 1 represents iron oxide, 2 represents copper oxide, $x \in \{\text{Prediction model based on chi - square test}\}$, $x \in \{\text{Lead - barium weathered glass without missing other colors}\}$.

It should be noted that 11 groups of cultural relics in Table 2 have been sampled and tested twice. For the two test results of the same cultural relic, we take the average value as the content of iron oxide and copper oxide corresponding to the cultural relic:

$$x_{ij}^* = \frac{1}{2} \left(\sum_{k=1}^2 x_{ijk} \right), i = 1, \dots, 11, j = 1, \dots, 14, k \in \{1, 2\} \tag{2}$$

Where i is the i cultural relic, j is the j chemical element, and k is the k detection.

Therefore, the sum of differences between No.19 and No.43, between No.40 and No.28, between No.48 and No.44, and between No.58 and No.43 are the smallest, so the corresponding colours of 19, 40, 48, and 58 are supplemented as their corresponding colours, which are all light blue.

2.2 Chi-square test of weathering on the surface of cultural relics

The surface weathering, glass type, texture and color of the glass relics involved in this question are all categorical variables, so we use chi-square test. We imported the processed data into SPSS software for interactive analysis, and obtained the following results. As shown in Figure 1.

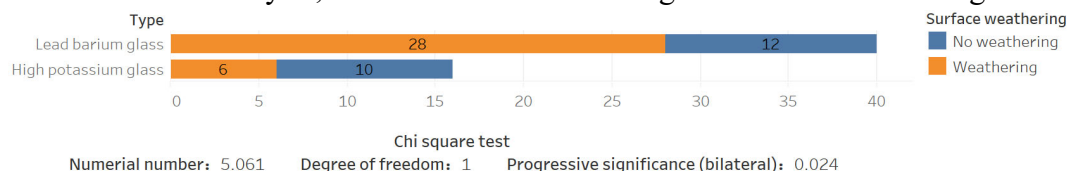


Fig.1 Type and surface weathering

2.3 Analysis of Chemical Composition Content of Weathered Surface of Cultural Relics Samples

The chemical composition maps of high-potassium glass and lead-barium glass before and after weathering are drawn respectively[7], as shown in Figures 2 and 3 below.

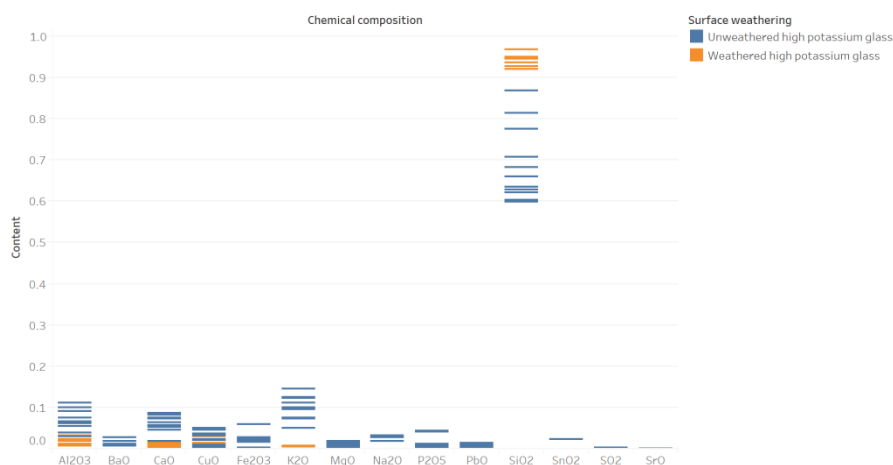


Fig. 2 Chemical composition of high-K glass before and after weathering

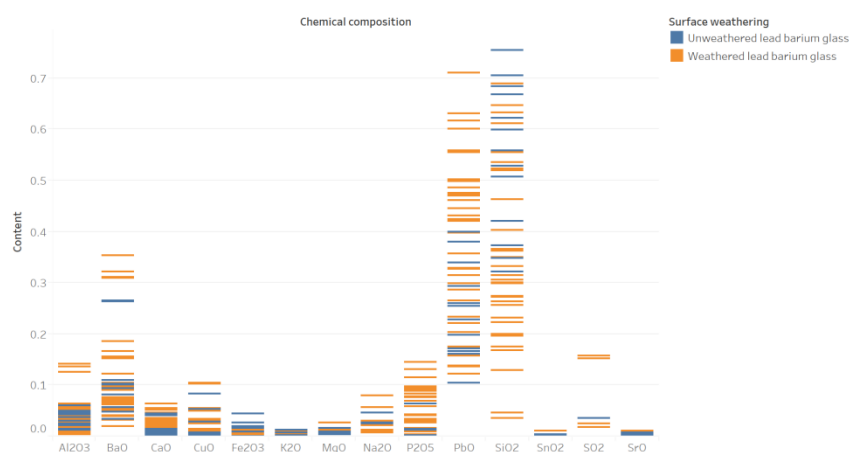


Fig. 3 Chemical composition of Pb-Ba glass before and after weathering

For high-potassium glass, the content of weathered silicon dioxide is higher and the content of non-weathered silicon dioxide is lower; the components contained before weathering but not contained after weathering include sulfur dioxide, barium oxide, sodium oxide, lead oxide, strontium oxide and tin oxide; in general, the components with higher content of non-weathered silicon dioxide include phosphorus pentoxide, calcium oxide, potassium oxide, magnesium oxide, iron oxide and copper oxide.

For lead-barium glass, the content of silica and most other components is lower after weathering and higher before weathering; the components that are not present before weathering but are present after weathering include magnesium oxide, strontium oxide, and tin oxide.

Considering that the number of samples is small and there may be particularity, within the allowable range of error, the content of the main chemical components of high-potassium glass after weathering is generally decreasing, while the content of main chemical components of lead-barium glass after weathering is generally increasing.

2.4 Prediction model of chemical composition content before weathering based on entropy weight method

Calculate the weighted average of the contents of 14 chemical components before and after weathering.

The weighted mean value \bar{m}_j of the content of the j chemical component before weathering and the content of the j chemical component after weathering.

$$\bar{m}_j = \sum m_{ij} \cdot w_i \quad (3)$$

$$\overline{m}_j^* = \sum m_{ij}^* \cdot w_i^* \tag{4}$$

Define the increase (decrease) ratio of each chemical component before and after weathering

$$q_i = \frac{\overline{m}_j^* - \overline{m}_j}{\overline{m}_j} \tag{5}$$

The predicted chemical composition of glass relics before weathering is shown in the Table 1.
 Table 1 Prediction Results of Chemical Composition Content of Glass Cultural Relics before Weathering

| Ornamentation | Type | Cultural relics sampling point | SiO2 | Na2O | K2O | CaO |
|---------------|----------------|--------------------------------|---------|--------|---------|--------|
| B | High potassium | 7 | 76.2009 | 0.0000 | 0.0000 | 6.8515 |
| B | High potassium | 27 | 72.6782 | 0.0000 | 0.0000 | 5.9923 |
| B | High potassium | 9 | 73.2309 | 0.0000 | 11.3317 | 3.7726 |
| B | High potassium | 22 | 60.3056 | 0.0000 | 12.4145 | 8.8293 |
| B | High potassium | 12 | 66.1975 | 0.0000 | 18.5043 | 4.1336 |
| B | High potassium | 10 | 76.3004 | 0.0000 | 18.5170 | 1.2888 |
| C | Lead-barium | 54 Severe weathering point | 48.7565 | 0.0000 | 0.0000 | 0.0000 |
| C | Lead-barium | 43 Part 2 | 57.2487 | 0.0000 | 0.0000 | 3.1655 |
| A | Lead-barium | 49 | 63.7411 | 0.0000 | 0.0000 | 1.9225 |
| C | Lead-barium | 11 | 67.2376 | 0.0000 | 0.3635 | 1.3822 |
| C | Lead-barium | 58 | 62.9482 | 0.0000 | 0.6214 | 1.4230 |
| A | Lead-barium | 19 | 61.7256 | 0.0000 | 0.0000 | 1.2593 |
| C | Lead-barium | 51 Part 2 | 56.6746 | 0.0000 | 0.0000 | 2.7631 |
| C | Lead-barium | 51 Part 1 | 56.8212 | 0.0000 | 0.0000 | 1.6488 |
| C | Lead-barium | 08 Severe weathering point | 17.0140 | 0.0000 | 0.0000 | 2.4197 |
| C | Lead-barium | 41 | 46.8072 | 0.0000 | 1.0330 | 2.5192 |
| A | Lead-barium | 50 | 49.3477 | 0.0000 | 0.0000 | 1.7458 |
| ... | | | | | | |

3. Classification of glass types based on hierarchical clustering

3.1 Establishment and solution of hierarchical clustering model

This paper first classifies the data[8] into weathered group data and unweathered group data. The method of system clustering is used to group according to whether it is weathered or not, and the types obtained by clustering are compared with the real types, to judge the rationality and feasibility of the model. As shown in Table 2.

Table 2 Flow chart of hierarchical clustering algorithm

The algorithm flow of hierarchical clustering is as follows:

- Step1: Consider each object as a class, and calculate the minimum distance between each two classes.
- Step2: The two classes with the smallest distance are divided into a new class.
- Step3: Recalculate the distance between the new class and all class
- Step4: Repeat Step 2 and Step 3 until all classes are finally merged into one class

The data of the weathered group and unweathered group are brought into SPSS to solve, and the corresponding vertical icicle diagram and pedigree diagram are obtained as follows:

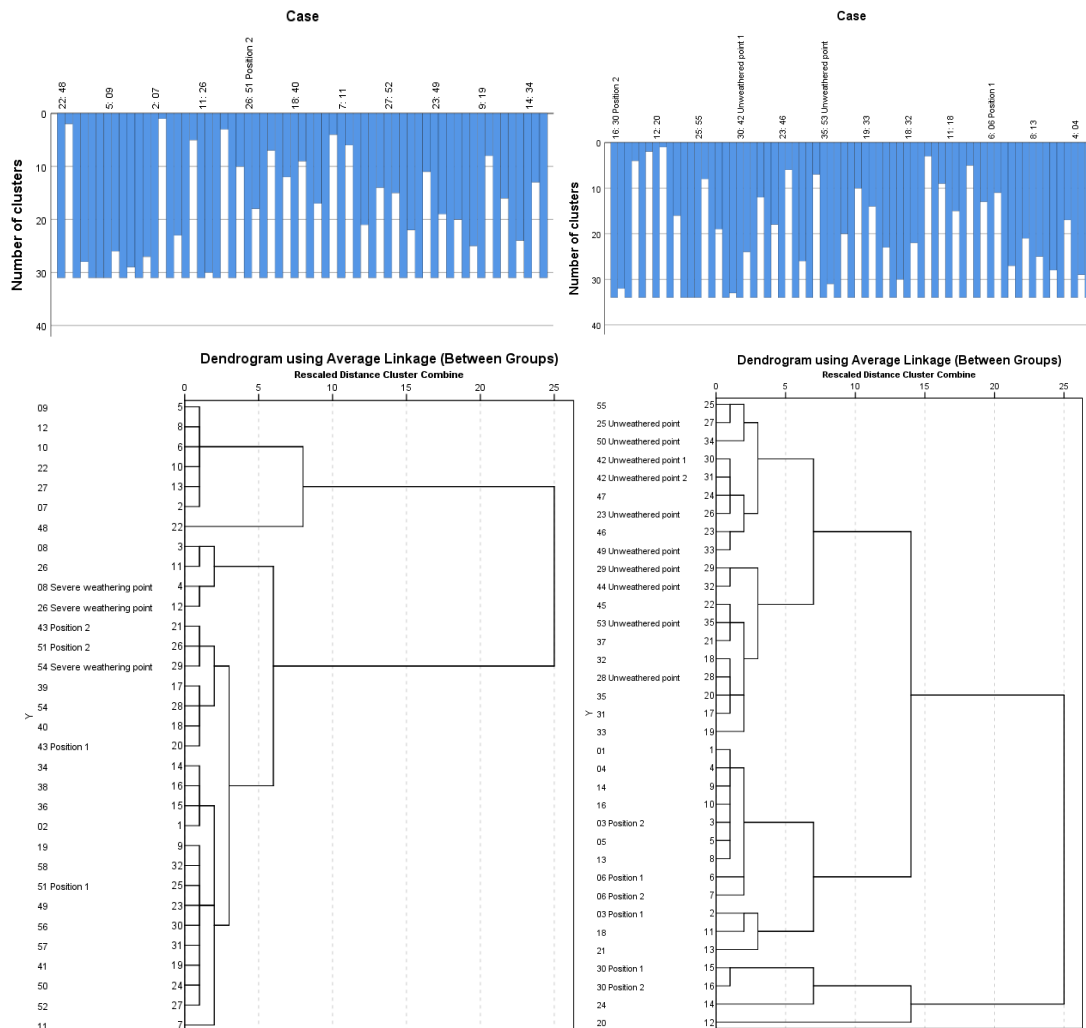


Fig. 4 Vertical icicle plot versus pedigree plot for weathered (left) and unweathered (right)

The final result is that 4 of the 35 data in the unweathered group are wrong, and the correct rate of the group is 88.57%. Among the 32 data of the group, only the classification results of No.48 cultural relics are wrong, and the correct rate is 96.875%. Therefore, the rationality and feasibility of the model are high. As shown in Figure 4.

3.2 Subclass division model based on standard deviation

Considering the above classification results[9], we compared the chemical composition of lead-barium glass and high-potassium glass and took the sample standard deviation S as the standard to judge the dispersion degree of 14 chemical composition indicators.

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (6)$$

The standard deviation of each chemical composition of high potassium weathered, high potassium unweathered, lead and barium weathered and lead and barium unweathered can be obtained. As shown in Figure 5.

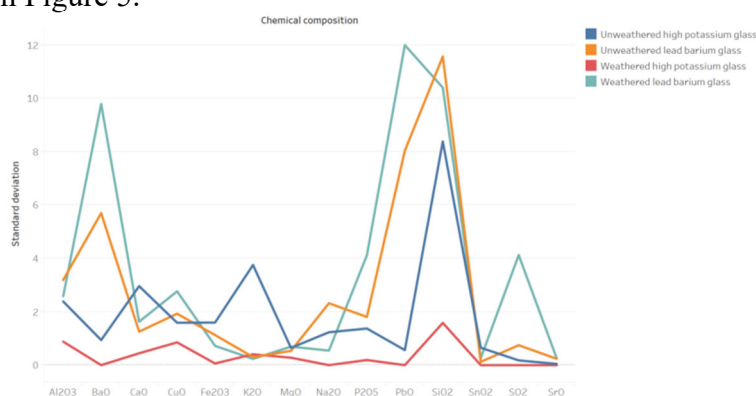


Fig. 5 Standard deviation diagram of chemical composition of different types of glass

After the weathering of potassium glass, there is a significant difference between magnesium oxide and aluminium oxide, so they are selected to distinguish and classify, and other classifications are the same. Finally, sensitivity analysis is carried out to disturb the data.

4. Prediction model based on K-means algorithm

In this paper, the chemical composition of the unknown type of glass relics was identified to determine the type[10]. On the basis of the second model, we use the K-means model again to select the appropriate indicators as the basis for classification according to the differences of 14 chemical components. As shown in Table 3.

Table 3. Predicted classification of cultural relics

| Cultural relic number | Surface weathering | Predictive classification |
|-----------------------|--------------------|---------------------------|
| A1 | No weathering | High potassium subclass 2 |
| A2 | weathering | Lead-barium 2 subclass |
| A3 | No weathering | Lead-barium 2 subclass |
| A4 | No weathering | Lead-barium 2 subclass |
| A5 | weathering | Lead-barium 1 subclass |
| A6 | weathering | High potassium subclass 1 |
| A7 | weathering | High potassium subclass 1 |
| A8 | No weathering | Lead-barium 1 subclass |

We perturb the data and set the perturbation of 1%, 2%, 5%, 10%, 20% and 30% for the data value of the variables respectively. The data with perturbation is compared with the unperturbed data, and the results show that the model is accurate[11].

The K-means and Igorithm is used to cluster the data into K categories, where the K value is the best number of clusters selected according to the C-H value, and the results are as Table 4.

Table 4 Classification of chemical composition between different categories

| | High potassium, no weathering | High potassium weathering | No weathering of lead and barium | Lead-barium weathering |
|--------------------------------|----------------------------------|---------------------------------|-------------------------------------|---------------------------|
| SiO ₂ | 2 | 2 | 2 | 3 |
| Na ₂ O | 1 | 1 | 1 | 2 |
| K ₂ O | 3 | 1 | 1 | 2 |
| CaO | 3 | 4 | 1 | 2 |
| MgO | 1 | 1 | 1 | 2 |
| Al ₂ O ₃ | 3 | 3 | 5 | 2 |
| Fe ₂ O ₃ | 1 | 1 | 1 | 2 |
| CuO | 1 | 4 | 1 | 2 |
| PbO | 1 | 1 | 3 | 1 |
| (BaO | 1 | 1 | 4 | 4 |
| P ₂ O ₅ | 1 | 1 | 1 | 2 |
| SrO | 1 | 1 | 1 | 2 |
| SnO ₂ | 1 | 1 | 1 | 2 |
| SO ₂ | 1 | 1 | 1 | 2 |

It can be seen from table 4 that for high-potassium glass, silicon dioxide belongs to a single category before and after weathering. In the chemical composition category after weathering, calcium oxide and copper oxide belong to the same category, aluminium oxide belongs to a single category, and the categories of other chemical components do not change; For lead-barium glass, silica, alumina, lead oxide, and barium oxide are all individually in one category, and the remaining chemical components are in another category. Silicon dioxide, lead oxide, and barium oxide are all in a separate group, with the remaining chemical components in another group. On the premise of the above clustering results, we select the DBI index and profile coefficient as the indicators of different analyses. By calculating the DBI and profile coefficients, we find that the DBI index increases after the weathering of high-K glass-like, After weathering, the DBI index of barium glass decreases, the profile coefficient of high-potassium glass is farther away from 1, and the profile coefficient of lead-barium glass is closer to 1. It is concluded that the correlation between the chemical compositions of the weathered Pb-Ba glass is strong, while the correlation between the chemical compositions of the weathered high-K glass is weak.

5. Conclusion

In this paper, when analyzing the relationship between the surface weathering of glass relics and glass type, texture and colour, the variables are classified or quantitatively judged first, and then the chi-square test suitable for determining the correlation of classified variables in mathematical statistics is selected, and it is obtained that there is no significant difference between surface weathering and glass type at the 95% confidence level. There are significant differences in surface weathering, colour and ornamentation. In the prediction, the entropy weight method is used to determine the weight of each chemical composition for the change rate of each chemical composition before and after weathering, which makes full use of the data. Compared with the subjective weight and the default weight of the chemical composition, the entropy weight method used in this paper is more scientific and rigorous. For the second model, when analyzing the classification rules of high-potassium glass and lead-barium glass, it is noticed that the chemical composition of different glass types is very sensitive to whether there is surface weathering, so when studying the classification

rules, the data are divided into weathered group and unweathered group, and the two groups of data are analyzed separately. When the unknown cultural relics of the third model are classified, the model of the second model with higher accuracy can be used for direct classification, and better classification results can be obtained while data and disturbance are increased, which reflects the universality of the application of the model and the accuracy and reliability of the classification judgment model.

However, due to the individualization of data processing, the accuracy of prediction may be relatively low.

References

- [1] Fu Weizhong, Chu Liuping. Research on the Evaluation of Manufacturing High Quality Development from the Perspective of Yangtze River Delta Integration -- TOPSIS Evaluation Model based on the improved CRITIC- entropy weight method combined weight [J]. *Industrial technology and economics*,2020,39(09):145-152.
- [2] L Syafaah,D Arisnanda,M Irfan,R H Hendaryati. Correlation of risk factors for coronary heart disease in type-2 diabetes mellitus using the chi-square method[J]. *IOP Conference Series Materials Science and Engineering*,2020,821(1).
- [3] li xingqi, gao xiaohong. Dimensionless method of effectiveness evaluation [J]. *Journal of statistics and decision*, 2021 ((15) : 24 to 28, DOI: 10.13546 / j.carol carroll nki tjyc. 2021.15.005.
- [4] Shen Ying. T test and F Application of inspection in Trend Analysis of chemical test quality control data [J]. *Chinese journal of inspection and quarantine*,2019,29(02):105-107.
- [5] he xiuli. Multivariate Linear model and Ridge regression analysis [D]. *Huazhong University of Science and Technology*,2005.
- [6] Yan Chun, Li Yaqi, Sun Haitang. Research on Auto Insurance Fraud Identification based on Ant Colony Optimization Random Forest Model [J]. *Insurance research*, 2017 (6) : 114-127. The DOI: 10.13497 / j.carol carroll nki is. 2017.06.011.
- [7] Deng Yurui, xu-dong cheng, Tang Fang, zhou yong. Risk control of Rice mildew in storage based on multiple linear regression analysis and random forest algorithm [J]. *Journal of University of Science and Technology of China*,202,52(01):47-54+72+55.
- [8] Tian Bing, Huang Shan, Sun Ye, Yan Yujie, Jin Kangkang. Rockburst Prediction based on SOM neural network clustering and gray TOPSIS evaluation method [J]. *China mining*,2021,30(01):188-192.
- [9] guo wenjuan. K-means clustering algorithm based on Optimized initial clustering Center [J]. *Wind technology*, 2022 (4) : 63-65. The DOI: 10.19392 / j.carol carroll nki. 1671-7341.202204021.
- [10] Wu Huijie. Study on Application and Optimization Method of random machine forest Algorithm [D]. *Jiangnan great learning*, 2021. DOI: 10.27169 / , dc nki. Gwqgu. 2021.001718.
- [11] Zhang Weifeng. Statistical analysis of Spearman's simple correlation coefficient and Gini Gamma correlation coefficient [D]. *wide East university of technology*, 2020. DOI: 10.27029 / , dc nki. Ggdgu. 2020.001333.