

Analysis and Study on Chemical Composition of Ancient Glass Products

Sen Hou ^{1, *, #}, Hui Wang ^{1, #}, Xinyi Huang ^{2, #}, Zhongyi Mo ^{1, #}, Yuling Long ^{3, #}

¹ School of Computer and Software Engineering, Xihua University, Chengdu, China

² School of Science, Xihua University, Chengdu, China

³ School of comprehensive health management, Xihua University, Chengdu, China

* Corresponding author: 1309210812@qq.com

#These authors contributed equally.

Abstract. Based on the artificial neural network, K-Means clustering and decision tree algorithm, this paper constructs a variety of models to study the relevant factors of the surface weathering of glass relics and the changes of various chemical components. In this paper, the statistical results are obtained by processing the data with glass type, texture and color respectively, and chi square test is carried out with weathering type respectively. At the same time, a logical regression classification model is established to analyze the relationship. Secondly, the data are divided into four groups based on the glass type and weathering degree, and the statistical rules are obtained by analyzing the average, median, maximum, minimum and change degree of each group of data. Finally, a multi-stage weathering backtracking model is built based on multihead self-attention to predict the chemical composition content of the detection point before weathering. The accuracy of the test is 99%, and the effect is good.

Keywords: K-Means clustering, Decision tree algorithm, Chi square test.

1. Introduction

The Silk Road, which originated in China and connects Eurasia and Southeast Asia, plays a vital role in the commodity trade economy and is the crystallization of the wisdom of the Chinese nation. Glass products were one of the important trade commodities at that time. Various glass relics unearthed in China had high historical research value, which could help us understand the production level at that time. According to literature review [1] According to the research on the ancient Silk Road and ancient glassware, the glass independently developed in China usually contains high lead oxide and barium oxide. This glass has high fragility and cannot withstand frequent changes in cold and heat. With the high potassium glass coming from the Silk Road, glass products can enter the public life.

The weathering of glass will be affected by many factors, and its weathering process is also the result of the exchange of internal materials and external materials [2]. Professional archaeologists will judge the types of glass relics according to the content of weathering materials. A number of glass relics have been detected and their glass types are known. We need to analyze the weathering materials, patterns, colors and other data.

2. data processing

2.1. Encoding processing

In order to facilitate subsequent programming and calculation, the corresponding variables are coded in this paper. The coding rules are as follows:

Pattern type code:

$$W = \begin{cases} 1, & \text{the pattern type is A} \\ 2, & \text{the pattern type is B} \\ 3, & \text{the pattern type is C} \end{cases} \quad (1)$$

Glass type code:

$$G = \begin{cases} 1, & \text{the glass type is high potassium} \\ 2, & \text{the glass type is lead barium} \end{cases} \quad (2)$$

Color Type Code:

$$C = \begin{cases} 1, & \text{the color is blue-green} \\ 2, & \text{the color is light blue} \\ 3, & \text{Color is purple} \\ 4, & \text{the color is dark green} \\ 5, & \text{the color is dark blue} \\ 6, & \text{the color is light green} \\ 7, & \text{Color is black} \\ 8, & \text{Color is green} \end{cases} \quad (3)$$

Surface weathering code:

$$F = \begin{cases} 1, & \text{the weathering type is weathering} \\ 2, & \text{the weathering type is non weathering} \end{cases} \quad (4)$$

G represents the glass type, W represents the decoration type, C represents the color type, and F represents the weathering type.

3. Model establishment

3.1. Data Mapping

This section is about data processing. It is conducive to the analysis of subsequent problems to describe the data in advance, so the general situation of the data should be obtained through preliminary statistics [3]. First, the data contained in the text are divided into two categories based on the classification of weathering types. Secondly, the number of eligible glass relics is counted according to the types of ornamentation, glass and color. The statistical results are shown in the following Fig.1:

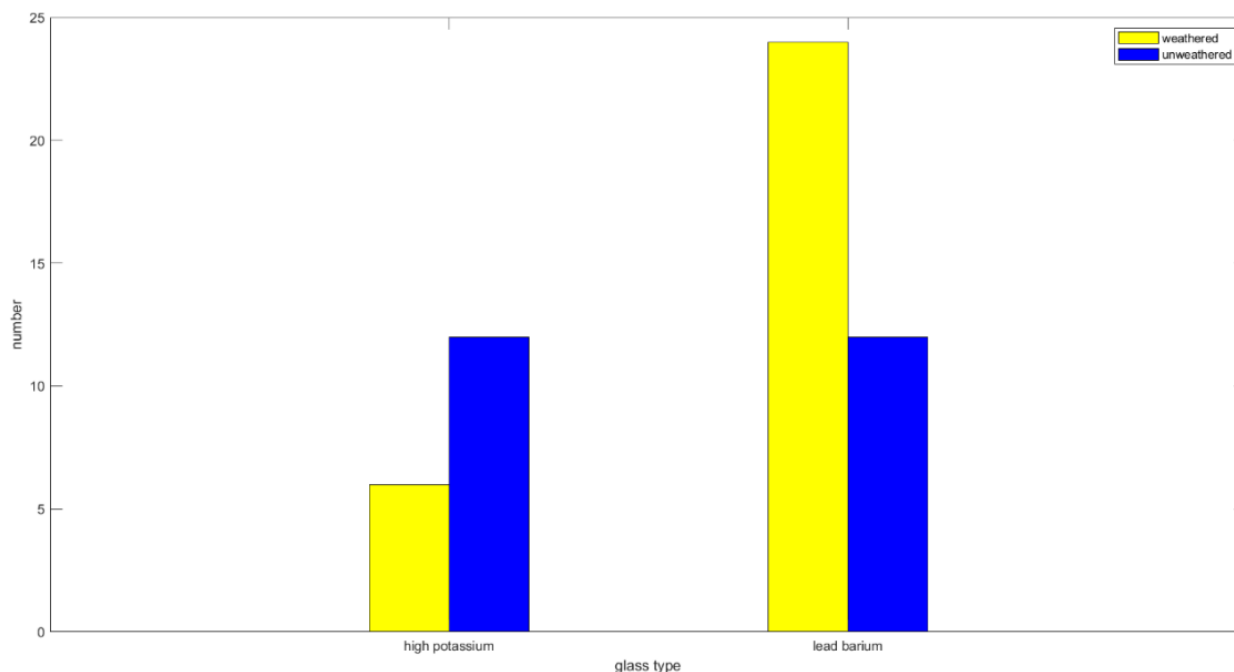


Figure 1. Bar chart of glass type

It can be seen from the bar chart of glass types that the quantity of high potassium type glass that is not weathered is significantly higher than the quantity of glass that is weathered [4], while the quantity of lead barium type glass that is weathered is significantly higher than the quantity of glass that is not weathered. It indicates that the type of glass is likely to be related to the type of weathering, as shown in Fig 2.

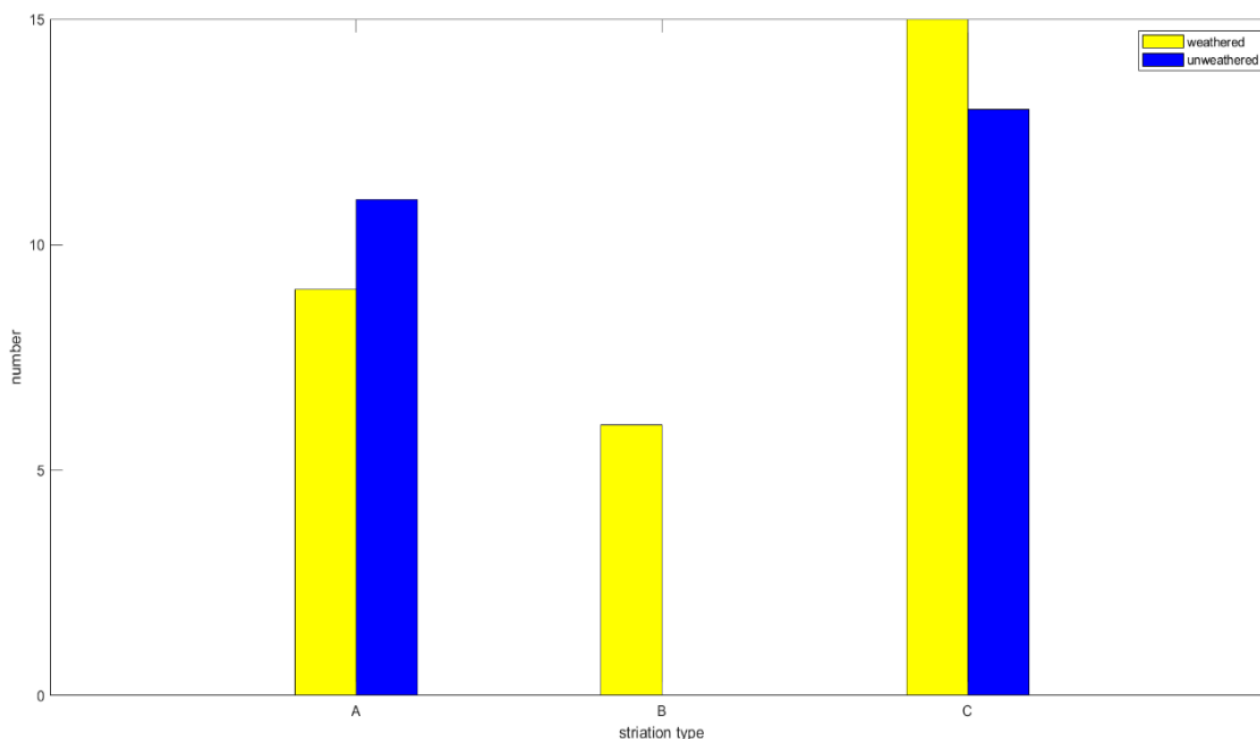


Figure 2. Bar Chart of Pattern Types

It can be seen from the bar chart of decoration type that when the decoration type is A or C, the amount of weathering and non weathering is similar, but when the decoration type is B, only weathering type appears, indicating that there may be a relationship between the decoration type and weathering type, as shown in Fig 3.

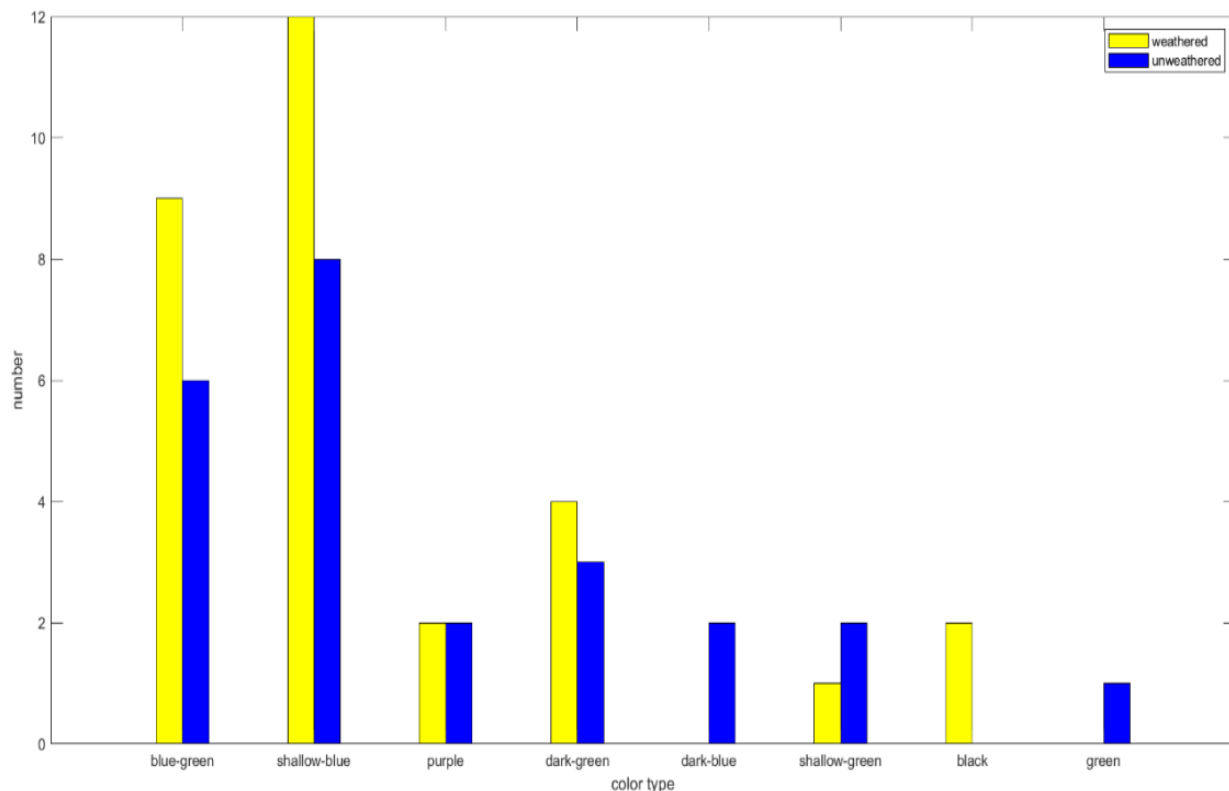


Figure 3. Bar chart of color type

It can be seen from the bar graph of color types that the number of weathered and unweathered types corresponding to different colors is not large, indicating that there may be no obvious relationship between color types and weathering types.

3.2. Chi square test

Weathering type, glass type, decorative pattern type and color type are categorical variables, and chi square test should be used to explore the relationship between the two categorical variables^[5]. This paper first imports the data processed in the data description part into SPSS for chi square test. Due to the space limitation of the article, this paper draws the analysis results into a table, and the analysis results are shown in the following Table 1:

Table 1. chi square test results

Variable Type	Pearson Chi square value	Significance value
G	5.400	0.02
W	5.747	0.056
C	6.287	0.507

When the significance value is less than 0.05, it can be considered that there is an obvious relationship between the two variables. According to the results of chi square test, the correlation between the three variables and weathering type is as follows: glass type, decoration type and color type. This result is consistent with the preliminary analysis of data description.

3.3. The construction of logical regression classification model

Logistic regression classification is to summarize and summarize the marked data through machine learning algorithms and classification models or classification decision functions to predict the output of the data, and classify the output limited discrete values through input discrete or continuous variables [6]. The model in this paper uses the method of logistic regression.

Assumptions of logical regression model:

$$h(x) = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^n a_i x_i = a^T x \quad (5)$$

$$h(x) = g(a^T x)$$

Where x represents the eigenvector, g represents a common logic function Sigmoid function:
 $g(z) = \frac{1}{1 + e^{-z}}$, where z represents a scalar or array of any size.

In this paper, only three independent variables are considered: decoration type x_1 , glass type x_2 , and color type x_3 . The values of each variable are calculated according to the code.

The sigmoid function image is shown in Fig 4 below:

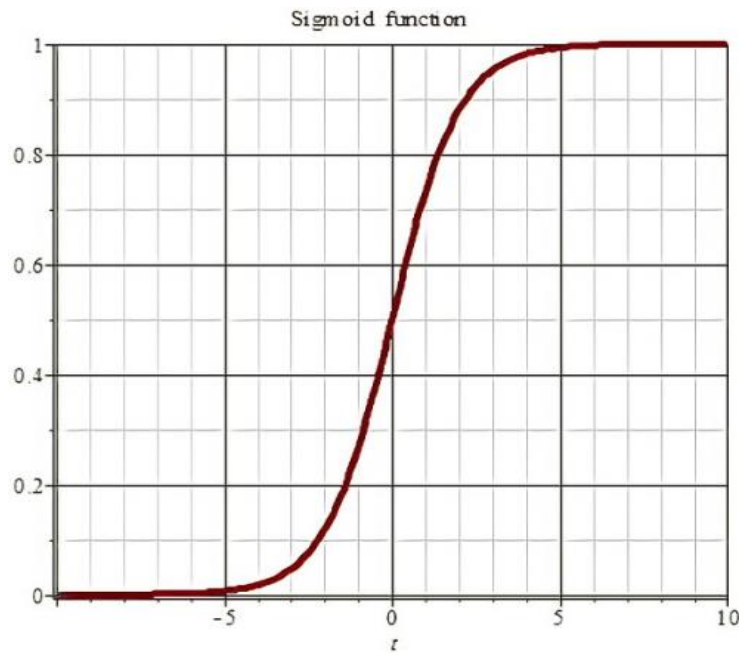


Figure 4. Sigmoid Function Image

According to the function image:

$$\begin{cases} g(z) = 0.5, z = 0 \\ g(z) < 0.5, z < 0 \\ g(z) > 0.5, z > 0 \end{cases} \quad (6)$$

Logical regression can be seen as a probability estimation. To implement a logical regression classifier, you can multiply each feature by a regression coefficient, add all the result values, and substitute the sum into the sigmoid function to obtain a value between 0-1.^[7] All data greater than 0.5 are classified as Category 1, and those less than 0.5 are classified as Category 0. At this time, $h(x)=g(z)$, namely:

$$y = \begin{cases} 0, h(x) < 0.5 \\ 1, h(x) > 0.5 \end{cases} \quad (7)$$

loss function:

$$H(h(x), y) = \begin{cases} -\log(h(x)), y = 1 \\ -\log(1 - h(x)), y = 0 \end{cases} \quad (8)$$

$$J(a) = \frac{1}{m} \sum_{i=1}^m H(h(x_i), y_i) = -\frac{1}{m} \left[\sum_{i=1}^m y_i \log h(x_i) + (1 - y_i) \log (1 - h(x_i)) \right] \quad (9)$$

Gradient descent method

Use gradient descent method to update parameters:

$$a_j = a_j - \frac{a}{m} \sum_{i=1}^m (h(x_i) - y_i) x_i^j \quad (10)$$

At this time, use L2 norm regularization:

$$\|x\|_2 = \sqrt{\sum x_i^2} \quad (11)$$

Update loss function:

$$J(a) = \frac{1}{2m} \sum_{i=1}^m (h(x_i) - y_i)^2 + \lambda \sum_{j=1}^m a_j^2 \quad (12)$$

Model results

Based on the logical regression algorithm, 70% data is selected as the training data set, and 30% data as a test set, the results obtained by running the program are shown in the following Table 2 :

Table 2. Table of Logistic Regression Coefficients

Pattern type	Glass type	Color Type	Constant term
-0.008	-1.556	0.450	0.011

Get the final equation:

$$h(x) = 0.011 - 0.008x_1 - 1.556x_2 + 0.45x_3 \quad (13)$$

Where, x_1 represents the decoration type, x_2 represents the glass type, and x_3 represents the color type. The influence on weathering degree from large to small is: glass type, color type and decorative pattern type.

3.4. Statistical laws and analysis

According to the glass type and weathering type, the data are classified into four groups: weathered lead barium glass, unweathered lead barium glass, weathered high potassium glass, and unweathered high potassium glass [8]. The 14 chemical compositions of each group of data were statistically analyzed, and their maximum, minimum, average and median values were recorded. Due to space limitations, only the relevant data of weathered lead barium glass were displayed, as shown in Table 3 below:

Table 3. Results of Statistical Laws

chemical composition	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	BaO	P2O5	SrO	SnO	SO2
Statistical value														
average value	23.53	1.61	0.40	2.81	1.17	2.55	0.98	2.39	43.55	14.74	5.83	0.50	0.47	8.88
median	23.45	1.38	0.32	3.01	1.15	2.22	0.62	1.15	44.44	10.00	5.88	0.46	0.47	8.81
minimum value	3.72	0.00	0.00	0.00	0.00	0.50	0.00	0.19	25.39	0.00	0.00	0.00	0.00	0.00
Maximum	39.57	2.22	1.05	6.40	2.73	5.73	2.74	10.57	61.03	35.45	14.13	1.12	0.47	15.95

It can be seen from the table that the average and median of each chemical composition are similar, and the floating range is large.

After obtaining a series of statistical values, the change rate can be calculated to explore the influence of weathering on the chemical composition content of the same glass type. The rate of change is calculated as follows:

$$P(i) = \frac{b(i) - a(i)}{b(i)} \quad (14)$$

Where, $P(i)$ represents the change rate of the i th chemical component, $b(i)$ represents the content of the i th chemical component before weathering, and $a(i)$ represents the content of the i -th chemical component after weathering.

3.5. Establishment of multi-stage weathering backtracking model

Compared with the traditional BP neural network, the network based on self attention mechanism can train all vectors at one time, and can accurately capture the correlation between different inputs. Self Attention is to filter a small amount of important information from a large amount of information, and ignore the unimportant information, so as to focus on the important information.

In this paper, a weathering backtracking model is proposed based on the self attention mechanism [9]. The model first encodes the data and records the i th pre weathering chemical composition data item as a^i . Secondly, a^i is multiplied by the coefficient matrix W^k , W^q , W^v to obtain k , q and v of a^i . Then use the obtained k and q as the inner product to obtain the self attention matrix A . Finally, normalize the matrix A to get A' , and multiply A' and v to get the final output vector. This vector represents the chemical composition after weathering. The network obtains the relationship between pre weathering and post weathering data by continuously updating W^k , W^q , W^v values.

3.5.1. Vector construction

Data encoding is the first step to build the model, which converts the original data into data that can be input into the model. In this paper, the first group of data before weathering is recorded as a^i , and then matrix A is defined as the input matrix and recorded as $A = \{a^i, i \in L\}$. Where A represents the set of all input vectors, and L represents the number of input vectors.

3.5.2. Calculate weight coefficient

Multiply the vector a^i by the three coefficient matrices W^k , W^q , W^v to get three vectors, called Key, Query and Value (hereinafter referred to as k , q and v). The calculation formula is as follows:

$$k^i = a^i \quad q^i = W^q a^i \quad (15)$$

$$k^i = W^k a^i \quad (16)$$

$$v^i = W^v a^i \quad (17)$$

After calculating three new vectors, we can calculate the score of self attention. This score represents the degree of attention paid to other vectors when in the i -th vector, as shown in the following fig 5:

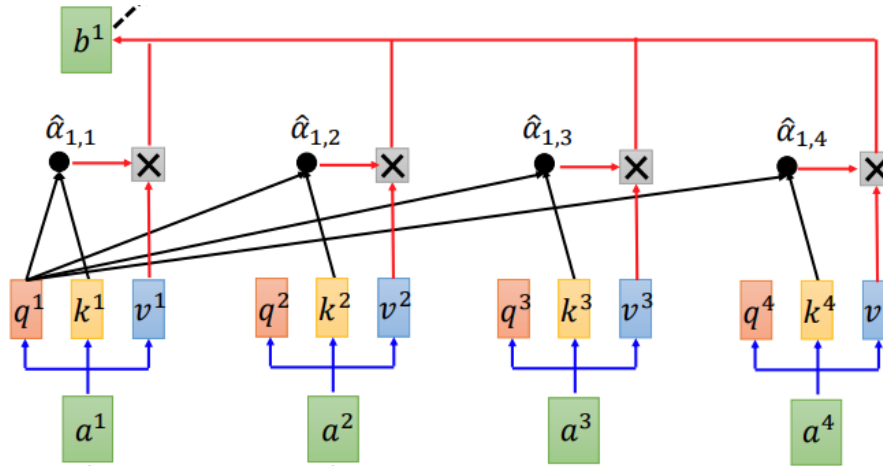


Figure 5. Relationship between score and attention

Fig 5 above is an example. First, we need to calculate the score $a_{1,1}, a_{1,2}, \dots, a_{1,L}$ between the first word and other words. The formula for calculating the score between the first word and the i th word is as follows:

$$a_{1,i} = q^1 \cdot k^i \quad (18)$$

$a_{1,i}$ represents the fraction between the first vector and the i th vector, expressed by the dot product of q of the first vector and k of the i -th vector. Similarly, the scores between any vector and other vectors can be obtained, and the attention score matrix A can be constructed [10].

3.5.3. Weighted Sum

The matrix A is obtained by normalizing the softmax function. Then multiply $a_{i,j}$ and corresponding v_j to get the final result b_i . The calculation formula of b_i is as follows:

$$b_i = \sum_{j=1}^L v_j \cdot a_{i,j} \quad (19)$$

The results of the model established in this paper are almost consistent with the true value, and the difference can be seen only after taking 5 decimal places, and the accuracy of the data obtained from the reverse calculation based on the change rate is greatly improved.

4. Conclusion

According to the analysis, the glass type is most related to the weathering type. In view of the inconsistent ranking of the correlation between the color type and the decorative pattern type, this paper consulted the literature, and through the research on the weathering results of the lead barium glassware, it can be seen that during the weathering process of the glass, the internal elements and external elements exchange a lot, the main color developing element Cu flows outward, and the Fe element gathers in the outer layer. Cu^{2+} and Fe^{2+} are green in alkaline environment and blue in acidic environment. It can be inferred that the color of the weathered glass is different from that of the glass after firing. The current technology cannot directly measure the accurate color of the glass before weathering, the results of the model established in this paper are almost consistent with the true value, and the difference can be seen only after taking 5 decimal places, and the accuracy of the data obtained from the reverse calculation based on the change rate is greatly improved.

References

- [1] Chen Shuyu, Hou Zhili An Analysis of the Ancient Silk Road and Ancient Chinese Glass [J] China National Expo, 2019, (05): 87 – 88.
- [2] Sun Dongmei Study on weathering of Huangshan rocks by 8 fungi [D] Anhui Normal University, 2021.
- [3] Wang Heng Research on influencing factors of logistics industry in Henan Province based on Pearson correlation analysis [J] Journal of Taiyuan City Vocational and Technical College, 2019, (09): 29 - 31. DOI: 10.16227/j.cnki.tycs.2019.0646.
- [4] Wang Jie, Li Mo, Ma Qinglin, Zhang Zhiguo, Zhang Meifang, Wang Julin A weathering study of an octagonal lead barium glassware in the Warring States Period [J] Glass and enamel, 2014, 42 (02): 6 - 13. DOI: 10.13588/j.cnki.g.e.1000 - 2871.2014.02.002.
- [5] Wang Zhengcun, Xiao Zhongjun, Yan Zhiguo Research on optimization of classification and recognition of logical regression [J]. Journal of Qilu University of Technology, 2019, 33 (05): 47 - 51. DOI: 10.16442/j.cnki.qlydxxb.2019.05.008.
- [6] Wang Jun, Zhang Haiyang, Zhao Kaixuan, Liu Gang Cow motion behavior recognition based on optimal binary decision tree classification model [J] Journal of Agricultural Engineering, 2018, 34 (18): 202 - 210.
- [7] Wang Yong, Liu Jianping, Cai Changxia An improved K-means clustering algorithm [J] Industrial Control Computer, 2010 (8): 3.
- [8] Huang Wenzheng Application of Artificial Neural Network in Cultural Relics Protection Engineering [C]// Proceedings of 2011 AASRI Conference on Information Technology and Economic Development (AASRI-ITED 2011 V1)., 2011: 302 - 305.
- [9] Cui Jing, Hu Zewen, Ren Ping. Research on the identification of potential "boutique" papers in the field of artificial intelligence based on decision trees and logical regression models [J]. Information Science, 2022,40 (05): 90-96. DOI: 10.13833/j.issn.1007-7634.2022.05.012.
- [10] Guo Wenjuan. K-means clustering algorithm based on optimized initial clustering center [J]. Science and Technology Wind, 2022 (04): 63 - 65. DOI: 10.19392/j.cnki.1671 - 7341.20204021.