Prediction and sub-classification of glass original components based on statistical learning algorithm

Jinyan Ye [1,*], Ziheng Guo [1], Xinyu Yu [2], Yinghua Song [1]

[1] School of Information Science and Engineering, Hohai University, Changzhou, Jiangsu, 213000

[2] School of Mechanical and Electrical Engineering, Hohai University, Changzhou, Jiangsu, 213000

* Corresponding Author Email: yjyan0415@163.com

Abstract. To address the problem that it is difficult for people to identify the information of artifacts affected by weathering, this paper takes several groups of weathered and unweathered silicate glass as the research object, and extracts the data of their chemical composition content and surface decoration, thus analyzing the correlation between decoration, glass type and weathering degree using chi-square test and KS test and studying the proportion of chemical composition of weathered glass before weathering. On this basis, L1 regularized logistic regression was used to roughly classify the initial classification, a subcategorization model was established by hierarchical clustering, and the entropy weight method was used to determine the significant indicators and the optimal number of categories for subcategorization. The results show that glass type has a significant effect on whether weathering occurs; the chemical components that have a significant effect on the determination of the type of lead-barium glass are SiO_2, PbO and BaO, and the chemical components that have a significant effect on the determination of the type of high-potassium glass are SiO_2 and K2O; high-potassium glass is divided into three subclasses, and lead-barium glass is divided into three subclasses.

Keywords: Silicate Glass, Chi-square Test, Logistic Regression, Hierarchical Clustering.

1. Introduction

According to the analysis of available documents, the ancient homemade glass in China can be traced back to as early as the eleventh century B.C [1], and many classic products emerged. Among them, silicate glass, with its excellent heat resistance and chemical stability, has been the most widely made and used from ancient times to the present. Silicate glass is made from silica, lime and alkali as raw materials, properly blended, then melted, cooled and cured, and is a non-crystalline inorganic substance [2] According to the different fluxes used, ancient silicate glass can be divided into lead-barium glass and high-potassium glass. However, due to intrinsic factors such as the composition and structural defects of the glass, as well as external factors such as the burial environment, the glass may be affected by weathering, resulting in a change in the ratio of chemical composition within the glass, which affects the judgment of the relevant workers on the category of silicate glass. In this paper, we hope to study the proportional changes of chemical composition of different categories of silicate glass before and after weathering, and derive the weathering law as well as the classification law of silicate glass, so as to provide assistance to related workers for digital and scientific research.

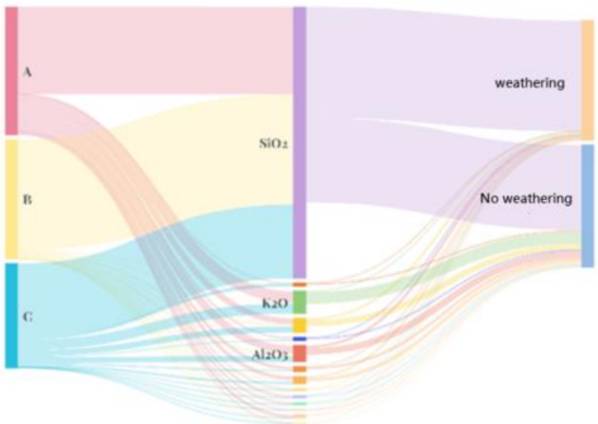
2. Data pre-processing and statistical analysis

2.1. Data sources and pre-processing

Data on the chemical composition content of ancient glass artifacts before and after weathering were compiled [3], and the data contained the number of the sample glass artifact, the type of decoration, the type of glass, and the proportion of the 14 chemical compositions. Some of the data are missing due to the limitations of the detection technique, indicating that the chemical was not detected in this sample. It is assumed that the non-detected i.e., the content is zero, and the complementary zeroes are processed for these data. In addition, the sum of each test data of the sample is not exactly 100%, so the content sum in the range of 85% to 105% is considered as limited data and kicked out the invalid data.

2.2. Descriptive statistics

The Sankey diagram consists of nodes and their connecting lines. Nodes represent the sources, intermediate nodes and sinks of various information, resources and energy flows, which are generally represented by squares; the information, material and energy flows between nodes are represented by connecting lines [4] As shown in Figure 1, Sankey diagrams were used to analyze the weathering of the two types of glass artifacts in relation to the type of ornamentation and chemical composition, allowing visual visualization of the data composition and flow state. The glass types are high potassium and lead-barium, and the glass samples are classified at the first level of the ornamentation of the artifacts, and the various chemical compositions are assigned to 14 chemical substances according to the total number of units.1 finally, all the glass can be divided into weathered and unweathered.



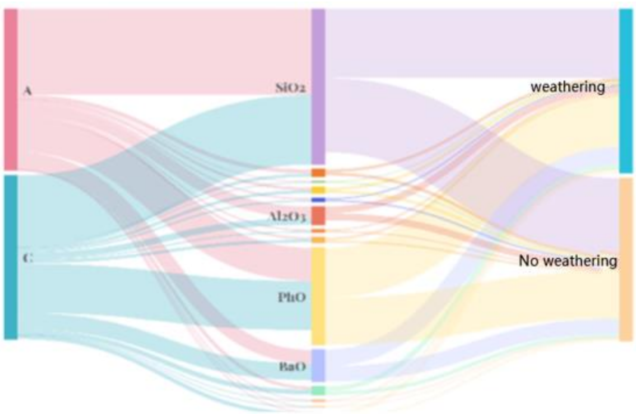


Figure a. High potassium glass artifacts Figure b. Lead barium glass artifacts

Figure 1. Chemical composition and weathering flow sankey diagram.

As shown in Table 1 and 2, the relationship between weathering and chemical composition ratio of two kinds of glass cultural relics was analyzed by descriptive statistical charts. The results show that, whether it is weathered or not, silica has the highest proportion of components in silicate glass; for high-potassium glass, the contents of SiO_2, K_2O and Ca_2O have changed significantly before and after weathering. For lead-barium glass, SiO_2 and PbO have changed significantly.

Table 1. Chemical composition description statistics (lead barium glass).

	Weathered				Unweathered			
	Mini-mum	Maxi-mum	Aver-age	Standard deviation	Mini-mum	Maxi-mum	Aver-age	Standard deviation
SiO_2	11.7	92.72	36.09	19.04	31.94	75.51	54.94	14.15
Na_2O	0	7.92	0.88	1.861	0	0.66	0.835	1.588
K_2O	0	1.05	0.147	0.22	0	0.712	0.22	0.249
CaO	0	5.8	2.123	1.459	0	4.37	0.971	1.162
MgO	0	2.73	0.674	0.68	0	1.67	0.447	0.553
Al_2O_3	0.45	14.34	3.843	3.64	1.44	5.45	3.11	1.408
Fe_2O_3	0	2	0.555	0.636	0	4.59	0.806	1.425
CuO	0	7	1.739	1.868	0	8.46	1.687	2.555
PbO	0	70.21	36.84	16.74	9.3	28.5	22.36	8.268
BaO	0	34	9.118	7.808	3.42	26.23	10.52	7.258
P_2O_5	0	14.13	3.987	3.992	0	5.75	0.92	1.611
SrO	0	1.12	0.355	0.262	0	0.91	0.287	0.324
SnO_2	0	1.31	0.061	0.247	0	0.44	0.036	0.127

Table 2. Chemical composition description statistics (high potassium glass)

	Weathered				Unweathered			
	Minimum	Maximum	Average	Standard deviation	Minimum	Maximum	Average	Standard deviation
SiO_2	92.35	96.77	94.21	1.815	59.01	79.64	67.77	6.928
Na_2O	0	0	0	0	0	3.38	0.834	1.377
K_2O	0	1.01	0.652	0.399	0	14.52	9.567	3.932
CaO	0.21	1.66	0.856	0.544	0	8.7	5.734	2.841
MgO	0	0.64	0.128	0.286	0	1.86	1.054	0.623
Al_2O_3	0.81	3.5	1.814	1.03	3.05	10.6	6.406	2.284
Fe_2O_3	0.17	0.35	0.278	0.069	0	4.22	1.788	1.309
CuO	0.55	3.24	1.566	1.045	0	4.73	2.415	1.517
PbO	0	0	0	0	0	1.62	0.384	0.571
BaO	0	0	0	0	0	1.97	0.458	0.761
P_2O_5	0	0.61	0.264	0.231	0	4.34	1.181	1.198
SrO	0	0	0	0	0	0.12	0.034	0.04
SnO_2	0	0	0	0	0	2.36	0.236	0.746

2.3. Correlation analysis

A chi-square test is a hypothesis test in which the distribution of a statistic approximately follows a chi-square distribution when the null hypothesis holds. The chi-squared test describes the degree of correlation between two categorical variables. A larger chi-squared value indicates a greater correlation and less independence between actual and expected, while a smaller chi-squared value indicates a smaller correlation and greater independence between actual and expected [5]. The formula for the chi-square test is as follows:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

Where χ^2 is the cardinality, E_i is the expected frequency, and O_i is the actual frequency.

Hypothesis testing. Original hypothesis: Glass type and decoration are not related to weathering or not. Alternative hypothesis: glass type and decoration are related to weathering or not.

Let the significance level α = 0.5 and the degrees of freedom be n-1. The estimates of the overall mean and the chi-square values are shown in equations (2) and (3):

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2}$$

$$\chi^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{{\sigma_0}^2} \tag{3}$$

The results of the chi-square test were shown in Table 3, according to the significance level of α=0.05 of the original hypothesis. The following conclusions can be drawn: the asymptotic significance of both ornament type and glass type meet $p<0.05$, which falls into the rejection domain of the original hypothesis, i.e., the original hypothesis of "ornament type and glass type are not related to surface weathering" is rejected.

Table 3. Cardinality test table for decoration type, glass type and weathering.

	Decoration Type	Glass Type
	Significance	Significance
Person Chi-Square	0.084	0.009
Likelihood Ratio	0.028	0.009

3. Prediction of pre-weathering chemical composition content

Due to the small number of data samples for the experiments, a large number of machine learning algorithms for classification prediction were not applicable. In order to predict the chemical composition content of the glass samples before weathering more efficiently and accurately, the method of normal distribution interval estimation is used for prediction. Therefore, in this paper, it is necessary to check whether the relevant data conform to the normal distribution before performing the normal distribution interval estimation.

The KS test is a nonparametric statistical method often used to compare whether the distributions of two aggregates are the same [6]. It is used to solve the goodness of fit between the empirical and normal distributions of finite samples with compatibility for finite samples, and its core idea is to determine whether the samples obey normal distribution by comparing the maximum distance between the empirical cumulative density distribution function and the theoretical cumulative density distribution function.

Hypothesis testing. The original hypothesis H_0: The sample data obeys normal distri-bution. Alternative hypothesis H_1: The sample data do not obey normal distribution (significance level α=0.05).

The KS test statistic is D_N:

$$D_N = \max_{1<k<n} \{|F_n(x_k) - F_0(x_k)|, |F_n(x_{k+1}) - F_0(x_k)|\} \tag{4}$$

Rejection domain of D_N: If $D_N > D_{N,1-\alpha}$, then H_0 is rejected; conversely, H_0 is accepted.

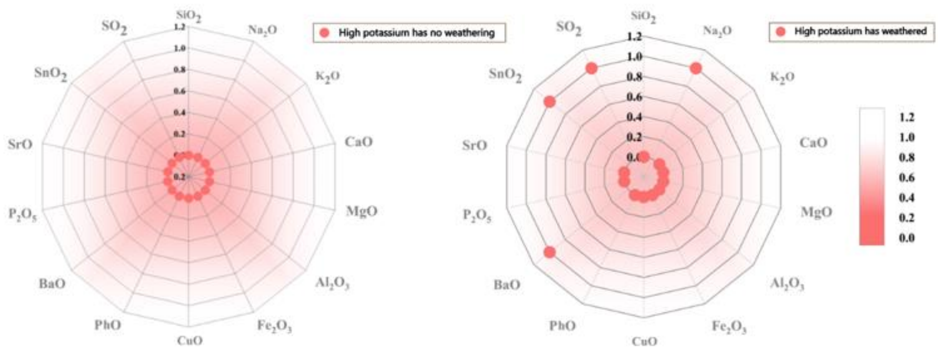


Figure 2. KS test value distribution (high potassium glass).

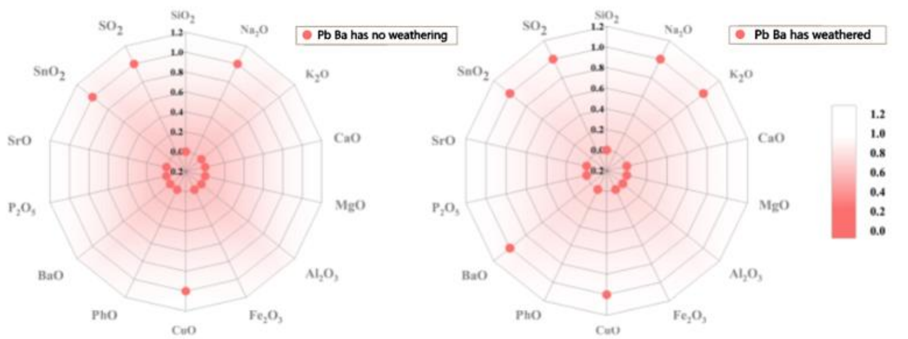


Figure 3. KS test value distribution (lead barium glass).

According to the heat maps of the distribution of detection values in Figures 2 and 3, it can be seen that the KS test values of the two types of glass artifact samples are mostly 0, i.e., they mostly obey a normal distribution. Therefore, for chemical compositions that pass the KS test, i.e., those that obey a normal distribution, the unweathered confidence interval can be used to estimate their pre-weathering compositions; for chemical compositions that are KS-tested, i.e., those that do not obey a normal distribution, a binary estimate with random values between zero and the minimum and maximum values is used because the proportions are small, random, and mostly zero. Therefore, the normal distribution interval estimation was used to predict the pre-weathering information as follows Table 4:

Table 4. Confidence interval table for estimation of normal distribution intervals

	High potassium glass		Lead barium glass	
	Lower confidence interval	Upper confidence interval	Lower confidence interval	Upper confidence interval
SiO_2	62.42	73.55	44.63	62.26
Na_2O	0.00	3.38	0.00	4.66
K_2O	6.84	11.82	0.02	0.50
CaO	3.37	7.30	0.35	2.11
MgO	0.65	1.51	0.16	0.82
Al_2O_3	5.04	8.20	2.36	4.03
Fe_2O_3	0.87	2.99	0.06	1.81
CuO	1.40	3.51	0.00	8.46
PbO	0.04	0.79	18.10	29.09
BaO	0.00	2.86	6.30	14.70
P_2O_5	0.49	2.31	-0.05	1.85
SrO	0.01	0.07	0.11	0.49
SnO_2	0.00	2.36	0.00	0.44
SO_2	0.00	0.47	0.00	3.66

4. Glass artifact classification model

4.1. Primary classification of glass artifacts

The binarized identification data sequences are closely related to the specific gravity parameters of each chemical component of the sample, and the machine learning algorithm logistic regression is considered to build the classification model. Among common machine learning algorithms, logistic regression model is a nonlinear binary classification model based on probability distribution, which is widely used in a variety of scenarios. As a common machine learning algorithm for qualitative variable analysis, the logistic regression model does not require too many restrictions on the normality of the data and the type of independent variables, and has shown good predictive ability in practical applications [7].

Given that glass artifacts can be divided into two major categories, lead-barium glass and high-potassium glass, and the predicted values of the linear probability model $\hat{y}_i = \omega_0 + \omega_1 x_{1i} + \omega_2 x_{2i} + \cdots + \omega_k x_{ki}$ may appear unrealistic such as $\hat{y}_i > 1$ or $\hat{y}_i < 0$, (the predicted value of $\hat{y}_i$ means the probability of occurrence of the event y=1). Therefore, to ensure that the probability takes a range of $0 \le \hat{y}_i \le 1$, the sigmoid function is introduced as shown in equation (5):

$$\hat{y} = S(\boldsymbol{x}_i'\boldsymbol{\omega}) = \frac{exp(\boldsymbol{x}_i'\boldsymbol{\omega})}{1 + exp(\boldsymbol{x}_i'\boldsymbol{\omega})} \tag{5}$$

Where x_i denotes the ith explanatory variable; ω_i denotes the regression coefficient corresponding to each variable.

Next, the parameters are solved using the method of great likelihood estimation (MLE), and the great likelihood function L (ω) is shown in equation (6):

$$L(\omega) = \sum_{i=1}^{N} \left[y_i(\vec{\omega} \cdot \vec{x}) - lg\left(1 + exp(\vec{\omega} \cdot \vec{x})\right) \right] \tag{6}$$

The weight vector $\vec{\omega}$ can be obtained by solving the parameters using the gradient descent method, i.e., by making $\frac{\partial L(\bar{\omega})}{\omega_i} = 0$. Finally, the surface weathering probability of the glass artifact sample is predicted according to equation (4), and if the predicted probability is greater than the truncation threshold (0.5), the glass artifact is judged to be a lead-barium glass; otherwise, it is a high-potassium glass.

However, without the inclusion of regularization, it is prone to overfitting due to the large number of features and the negligible effect of some of them on the prediction results. L1 regularization and L2 regularization are often used to minimize the output weights, which are sparser compared to L2 regularization [8]. Therefore, using L1 regularization, the regularization term is introduced in the above great likelihood function equation (7) as follows:

$$J(\omega) = L(\omega) + \partial \sum_{i=1}^{N} |\ \omega_i\ | \tag{7}$$

L1 regularization allows the coefficients of some of the features that have a small impact on the result to be reduced to 0. Therefore, L1 is suitable for cases where features are correlated with each other. L1 regularization can produce a sparse weight matrix, and there are many features with weight 0 in the sparse matrix, and these features are filtered out, which can achieve the effect of feature selection and prevent overfitting [9].

Table 5. Regularized regression coefficients

Parameters	Regression coefficient	Parameters	Regression coefficient
CHEMICAL1- SiO_2	-0.014	CHEMICAL8-CuO	0.000
CHEMICAL2- Na_2O	0.000	CHEMICAL9-PbO	0.421
CHEMICAL3- K_2O	-0.427	CHEMICAL10-BaO	0.000
CHEMICAL4-CaO	-0.090	CHEMICAL11- P_2O_5	0.000
CHEMICAL5-MgO	0.000	CHEMICAL12-SrO	0.000
CHEMICAL6- Al_2O_3	0.000	CHEMICAL13- SnO_2	0.000
CHEMICAL7- Fe_2O_3	0.000	CHEMICAL14- SO_2	0.000
		WEATHERING	0.000

After L1 regularization, the final binary classification model of glass artifacts is constructed as shown in equation (8):

$$TYPE_i = \alpha + \sum \omega_n \times CHEMICAL_n + \lambda \times WEATHERING_i + \varepsilon_i \tag{8}$$

Where $TYPE_i$ denotes the glass type to which the ith sample belongs, and ω_n denotes the regression coefficient corresponding to the nth chemical element. The calculated regression coefficients for each regression are shown in Table 5.

Table 6. Predicted classification results.

Code of artifact	A1	A2	A3	A4
Result	K-high potassium glass	Lead-barium glass	Lead-barium glass	Lead-barium glass
Code of artifact	A5	A6	A7	A8
Result	Lead-barium glass	Lead-barium glass	K-high potassium glass	K-high potassium glass

The classification criteria of lead-barium glass and high-potassium glass have been obtained in the previous work, based on which the species prediction of unclassified silicate glass was carried out, and the relevant data were brought into the prediction model, and the classification results were obtained as shown in Table 6.

4.2. Subclasses of glass artifacts

Lead barium glass is mainly homemade in China, while soda lime glass is closely related to the exchange between China and foreign countries [10] If only the glass artifact samples are divided into two categories, lead-barium glass and high-potassium glass, it is not conducive to further study and analysis of glass artifacts. Therefore, subclassifications were made for lead-barium glass artifacts and high-potassium artifacts glass separately.

(1) Entropy method of selecting indicators

Therefore, we need to analyze the fourteen chemical substances, select the information-rich and significant components as classification indicators, and use these indicators to classify the subclasses.

The entropy power method is related to the concept of entropy in information theory. In information theory, the degree of dispersion of a sample can be judged by calculating the magnitude of information entropy. The greater the degree of sample dispersion, the greater the influence of the index on the overall, and the use of entropy weighting method can effectively reduce the influence of human subjective factors on the weights. In a classification system of glass artifact samples, the greater the degree of variation in the weight of a chemical component, the greater the amount of information contained in that component, and the greater the degree of influence on the classification results. Therefore, entropy can measure the amount of information contained in each indicator in the system, and use it to determine the weight of each indicator.

First, the matrix X is built based on the values of the jth indicator of the ith sample, and the matrix is normalized to obtain the matrix Z, as shown in equations (9) and (10), where x_{ij} denotes the value of the jth indicator of the ith sample.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{9}$$

$$z_{ij} = \frac{x_{ij} - min\,\{x_{1j}, x_{2j}, \cdots, x_{nj}\}}{max\{x_{1j}, x_{2j}, \cdots, x_{nj}\} - min\{x_{1j}, x_{2j}, \cdots, x_{nj}\}} \tag{10}$$

Then, normalization is performed according to the standardization matrix Z to generalize the statistical distributivity of the uniform sample and obtain the matrix P, as shown in equation (11), where z_{ij} denotes the value of the jth indicator of the i-th sample after standardization.

$$p_{ij} = \frac{z_{ij}}{\sqrt{\sum_{i=1}^{n} z_{ij}}} \tag{11}$$

The entropy weight value of each indicator can be calculated according to the normalization matrix P, as shown in equations (12) to (14). Where, p denotes the value of the jth indicator of the ith sample after normalization, e_j denotes the information entropy of the jth indicator, and W_j denotes the entropy weight value of the jth indicator.

$$e_j = -\frac{1}{ln\,n}\sum_{i=1}^{n} p_{ij}\,ln(p_{ij}) \qquad (i = 1,2,\ldots,m) \tag{12}$$

$$d_j = 1 - e_j \tag{13}$$

$$W_j = \frac{d_j}{\sum_{j=1}^{m} d_j (j = 1,2,\ldots,m)} \tag{14}$$

The entropy weights of each chemical component content in the high-potassium glass artifacts and lead-barium glass artifacts were calculated separately, and the results are shown in Figure 4. The overall trend tends to level off after the second indicator, which indicates that the first two indicators are significant and informative, so the two indicators with the highest scores,$r_{SiO_2} = 0.70$ $r_{K_2O} = 0.09$, are selected as the indicators for clustering of high potassium artifacts; Figure b shows the descending entropy weight of chemical composition of lead-barium glass artifacts, and the overall trend tends to level off after the third indicator. The overall trend tends to level off after the third indicator, indicating that the first three indicators have significant entropy weight and rich information, so the three indicators with the highest scores are selected,$r_{SiO_2} = 0.36$ $r_{PbO_2} = 0.30$ $r_{BaO} = 0.12$ i.e., SiO2, PbO and BaO are used as the indicators of lead-barium artifacts the indicators of clustering.

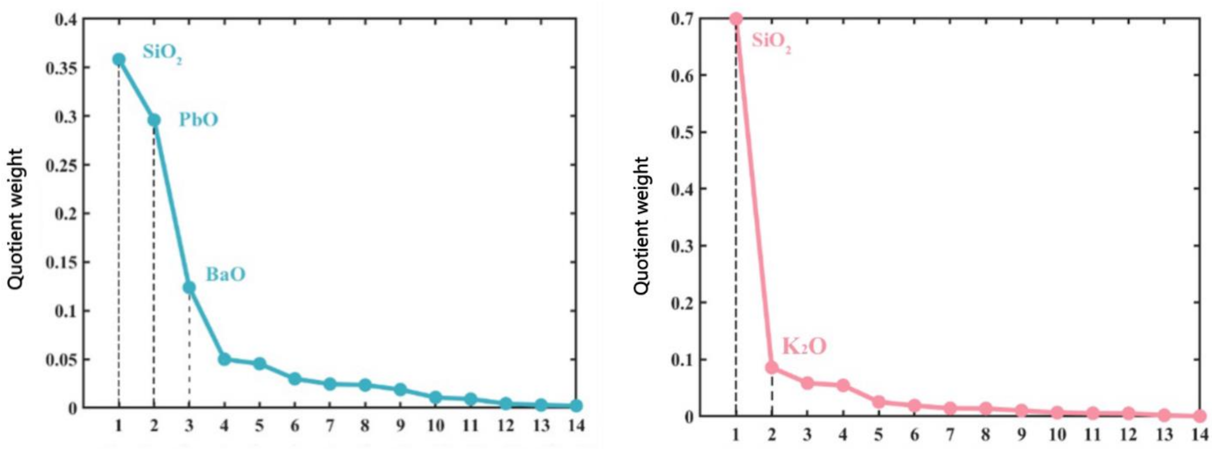


Figure a. High potassium glass artifacts Figure b. Lead barium glass artifacts

Figure 4. Entropy weight of chemical composition of glass artifacts.

(2) Hierarchical clustering for subclassing

The purpose of using clustering algorithm is to divide similar glass artifacts into a number of different class clusters according to certain characteristics thus achieving subclass division, so that the similarity between glass artifacts in the same class cluster is as large as possible, and the similarity between objects in different class clusters is as small as possible. At present, common clustering methods, such as K-means and K-means++, need to determine the number of clusters first before class classification. For the subcategory classification of glass artifacts, the hierarchical clustering method is used for subcategory classification because the number of subcategories is difficult to determine and the amount of data is small. The advantage of the hierarchical clustering method is that the number of subclasses does not need to be specified in advance, the hierarchical relationship of clusters can be observed through the drawn genealogical map, and the objectivity of the algorithm is stronger. The hierarchical clustering method is a greedy algorithm, i.e., it gradually finds the local optimal solution to achieve the overall optimal solution.

The algorithm process of hierarchical clustering is to first construct n classes, each containing one data sample; calculate the Euclidean distance $d(x, y)$ between two of the n samples as shown in equation (15), where m represents the dimensionality of the data sample; merge the two classes with the smallest inter-class distance, eliminate the two classes and construct a new class, and the merging rule uses the central distance; so on and so forth, keep looking for the two classes with the smallest class distance in the class cluster for merging until the number of classes is left at one.

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_m - y_m)^2} = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \tag{15}$$

The aggregation coefficients were calculated based on the number of clusters, and the aggregation coefficients were arranged in descending order to obtain dotted line plots, and the degree of aberration of the scatter plot at each point was observed, and the number of clusters was taken as i when the aberration degree of the ith point was significantly flat. The aggregation coefficients of high potassium artifacts and lead-barium artifacts are shown in dotted line Figure 5.In Fig. a, it can be observed that the degree of aberration of the aggregation coefficients tapers off after the number of clusters K=3, therefore, for the high potassium artifacts, they should be divided into three subclasses; in Fig. b, it can be observed that the degree of aberration of the aggregation coefficients tapers off after the number of clusters K=3, therefore, for the lead-barium artifacts, they should be divided into three subclasses.

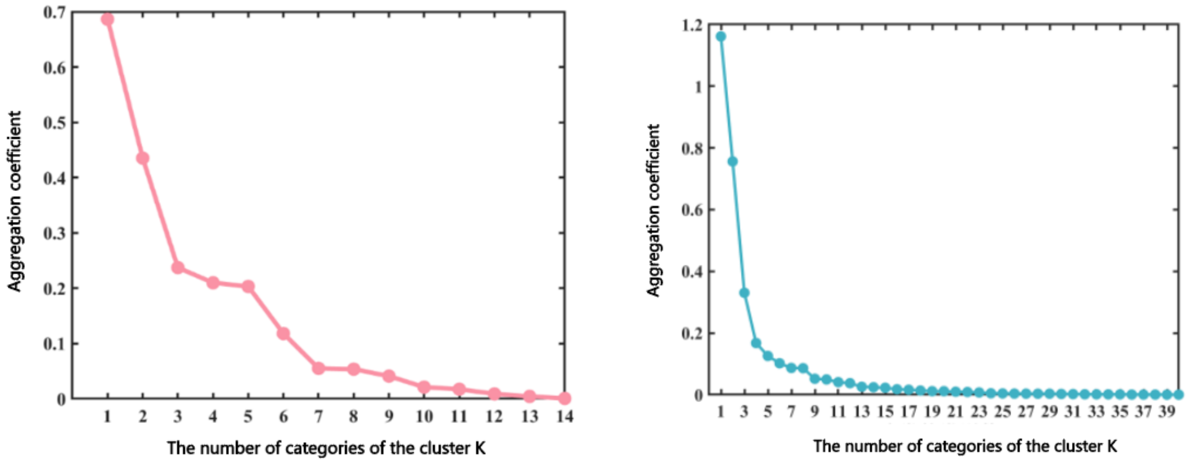


Figure a. High potassium glass artifacts. Figure b. Lead barium glass artifacts.

Figure 5. Clustering coefficient chart for two types of glass artifacts

Systematic clustering was carried out with a known number of clusters of high potassium artifact samples of 3 and lead-barium artifact samples of 3. The members of each cluster were obtained, and the final sub-classification results are shown in Figure 6. It can be observed that the samples in the same subclass have small differences, and the samples in different subclasses have significant differences.

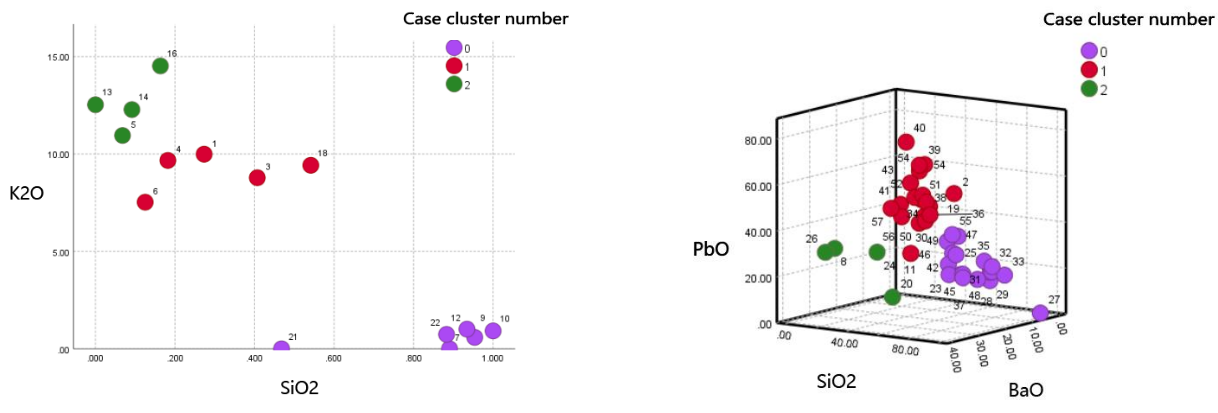


Figure a. High potassium glass artifacts. Figure b. Lead barium glass artifacts.

Figure 6. Two types of glass artifacts sub-classification results chart.

5. Conclusions

To determine the specific gravity of the main chemical components and the type of glass decoration in the detected glass artifacts, this paper first used the available data and found that whether the glass artifacts were weathered or not was significantly correlated with the type of glass and the type of displayed decoration by chi-square test analysis, and after KS test, the sample data were found to conform to a normal distribution, so the normal distribution interval estimation could be used to predict the chemical composition of the glass artifacts before weathering percentage information. Secondly, L1 regularized logistic regression was used to classify the glass products into two categories, lead-barium glass and high-potassium glass, and by entropy weight calculation, it was found that SiO2, PbO and BaO were selected as significant indicators for lead-barium glass, and SiO2 and K2O were selected as significant indicators for high-potassium glass. Finally, the hierarchical clustering method was used to divide the high-potassium glass pseudo-products and lead-barium glass pseudo-products into subclasses, and the number of subclasses for high-potassium glass pseudo-products and lead-barium glass pseudo-products was determined to be three by using the trend of the clustering coefficient graph. Since the difference between the model prediction results and the actual one is small, the model accuracy is considered to reach the actual analysis level and has some generality, which can provide some reference for the fields of heritage analysis, heritage conservation and chemistry.

References

[1] The cleaning of five Western Zhou tombs in Pangjiagou, Luoyang [J]. Cultural relics, 1972(10):20-31+70.Fangfang. Research on power load forecasting based on Improved BP neural network [D]. Harbin Institute of Technology, 2011.

[2] Yang Boda. A few questions about the study of ancient glass history in China [J]. Cultural relics, 1979(05): 76-78.

[3] Chinese Society of Industrial and Applied Mathematics [EB/OL]. [2022-09-15]. http://www.mcm.edu.cn./html_cn/node/5267fe3e6a512bec793d71f2b2061497.html.

[4] Cheng L, Lv X.Y., Xiao L.M., et al. The process of water supply, consumption, and discharge and its Sankey diagram [J]. Hydropower Energy Science, 2021, 39(06): 37-41.

[5] Li Meng, Bao Lei, Hu Yi, Cheng Song, Hu Xiaobo, GAO Ying. Implementation of random number online detection method based on chi-square test [J]. Microelectronics, 2022, 52(03): 388-392. DOI: 10.13911/j.cnki.1004-3365.210329.

[6] Fang Xiangyu, Lei Yiju, Ou Zujun. Homomorphism discrimination of U--type design combinations based on KS test [J]. Journal of Lanzhou College of Arts and Sciences (Natural Science Edition), 2022, 36(01):6-9+69. DOI: 10.13804/j.cnki.2095-6991.2022.01.005.

[7] JI Li-Ming, XIONG Xing-Wang, YANG Zi-Rong. A logistic regression-based classification model for diesel engine operating conditions [J]. Small Internal Combustion Engine and Vehicle Technology, 2023, 52(02): 6-9+20.

[8] Li Xin. Research on olfactory perception and its prediction model based on chemical molecular descriptors [D]. Guangdong University of Technology, 2021. DOI: 10.27029/d.cnki.ggdgu.2021.0000 46.

[9] Chen C. L1 regularization with pinball loss function for limit learning machine [J]. Information Technology and Informatization, 2023(03): 37-40.

[10] Zeng Weisheng. Ancient Chinese and Western glassware and its cultural differences from excavated glass of Han Dynasty [J]. Journal of Gemology and Gemmology (in English and Chinese), 2022, 24(06): 84-92. DOI: 10.15964/j.cnki.027jgg.2022.06.008.