Analysis of different glass compositions based on Spearman's correlation coefficient

Liwen Zhang [1,*,#], Huaikun Xiao [2,#], Yuxiang Su[3,#]

[1]College of management, Hainan University, Haikou, Hainan, 570208 China

[2]College of Information and communication engineering,Hainan University, Haikou, Hainan, 570208 China

[3]College of cyberspace security, Hainan University, Haikou, Hainan, 570208 China

*Corresponding author: 1563034339@qq.com

[#]These authors contributed equally.

Abstract: Glass is a valuable physical evidence of our early trade exchanges, and the analysis and identification of the composition of ancient glass products is of great significance to the study of our ancient history. However, ancient glass is highly susceptible to weathering by the burial environment, and during the weathering process, the internal elements of the glass and the environmental elements inevitably exchange in large quantities, resulting in changes in its composition ratio and color, thus affecting the correct judgment of its category. In this paper, Fisher's linear discriminant analysis is established to predict, analyze and identify the composition of ancient glass using mathematical and statistical methods. The study of this problem is important for the research and identification of ancient historical relics in China.

Keywords: Ancient glass; Fisher linear discriminant analysis; Spearman correlation analysis.

1. Introduction

The history of glass in China is long, and glass is also a valuable physical evidence of our early trade exchanges. The analysis and identification of the composition of ancient glass products are of great significance to the study of our ancient history[1]. However, ancient glass is highly susceptible to weathering by the burial environment, and during the weathering process, the internal elements of glass and environmental elements inevitably exchange in large quantities, resulting in changes in its composition ratio and color, thus affecting the correct judgment of its category[2].

Archaeologists now have to analyze and identify the chemical composition of a batch of ancient glass artifacts in China based on the relevant data. Some of these artifacts are marked as having no weathering on the surface, but shallow local weathering cannot be excluded; in some of the weathered artifacts, there are also unweathered areas on their surfaces[3]. Therefore, it is necessary to analyze and identify the composition of ancient glass objects by establishing mathematical models.

In this paper, considering the influence of weathering and unweathering on the chemical composition content of artifacts, the artifacts are divided into weathered and unweathered and judged separately[4]. Firstly, the unweathered known artifacts and unknown artifacts are screened out by data pre-processing, and the linear discriminant function is determined by using Fisher's linear discriminant analysis algorithm, and then the test data are tested according to the linear discriminant function to get the categories of the test data. Finally, the weathered unknown artifacts are tested according to the same method. From this, the types of all unknown artifacts can be identified. After that, the validation and sensitivity analysis is carried out by using the classify function, and all known category samples are used as training data and all unknown samples are classified to obtain the same solution as Fisher's linear discriminant analysis algorithm with less error. Finally, by reclassifying the training samples using the classify function after downscaling the overall data, we observe whether their classification results change and find out the range of sensitivity change by experiment.

Since weathering will change the chemical composition and increase the chance of analysis, this paper firstly carries out data preprocessing and selects the unweathered data, and the unmeasured chemical composition is taken as 0[5]. The normality test of the data is found not to satisfy the normal distribution. Therefore, the correlation coefficients between the chemical compositions were analyzed by the Spearman correlation coefficient analysis in the nonparametric test for two categories of glass artifact samples, namely, high potassium glass and lead-barium glass, and the correlation coefficients and significance of each chemical composition were calculated. Finally, the results of the analysis were summarized to conclude the differences between the correlations of the chemical components of high potassium glass and lead-barium glass.

2. Model assumptions and notation

2.1 Assumptions[6]

1. It is assumed that all changes in the chemical composition of the glass are due to weathering and are not influenced by other environmental factors.
2. It is assumed that a data blank indicates that the composition is detected.
3. it is assumed that the difference in composition and content between the two parts of Site 1 and Site 2, which are different in shape, is not due to weathering, i.e., the weathering properties of the two sampling sites are the same for Site 1 and Site 2.
4. it is assumed that the only unknown category of glass artifact type is potassium high glass or lead-barium glass.
5. For components with a chemical content of 0 after weathering, it is assumed that their pre-weathering content is also.

2.2 Notations

Important notations used in this paper are listed in Table 1.

Table 1 Notations

Symbols	Description
a	Content of a component after weathering
b	Content of the component before weathering
μ	Component weighted mean value
S	Profile coefficient of the sample
K	Number of clusters
C_i	Clusters to which the glass artifact belongs
μ_i	Mean vector, center of mass
$d(x,y)$	Distance of the sample to the center point
a_i	Average distance from sample i to other samples in the same cluster
b_{ij}	Average distance of sample i to all samples of the nearest cluster C_j
SSE	internal error sum of squares
S_i	intra-class dispersion matrix
S_w	Total intra-class dispersion matrix
N_i	Number of samples in class w
y_0	Segmentation threshold on the projection space
S_b	Sample interclass dispersion matrix
$J_F(W)$	Fisher criterion function

3. Model construction and solving

3.1 Identification of artifacts based on Fisher's linear discriminant analysis algorithm

3.1.1 Model building

$A1, A3, A4, A8$ is an unweathered artifact; $A2, A5, A6, A7$ is a weathered artifact. In this paper, the Fisher linear discriminant analysis algorithm is used to identify artifacts with unknown properties.

The basic idea of Fisher linear discriminant analysis algorithm is to choose a projection direction (linear transformation, linear combination), reduce the high-dimensional problem to a one-dimensional problem to solve, while the transformed one-dimensional data meet the samples within each class as much as possible to gather together, and the samples of different classes as far apart as possible. W and threshold w0, i.e., determine the linear discriminant function, and then, according to this linear discriminant function, test the test data to obtain the class of the test data. One of Fisher's principles for choosing the projection direction W is to make the projection of the original sample vector in that direction take into account the requirement that the distribution between classes be as separate as possible and the projection of samples within classes be as dense as possible[7].

Since the identified artifacts have two properties: high potassium glass and lead-barium glass, this paper uses the biclassification Fisher linear discriminant analysis to analyze and discriminate. The steps of using the binary Fisher linear discriminant analysis method are as follows.

(1) Calculate the mean vector of the two types of samples m_i , N_i is the number of samples of ω_i .

$$m_i = \frac{1}{N_i} \sum_{X \in \omega_i} X \quad i = 1, 2 \tag{1}$$

(2) Calculate the sample intra-class dispersion matrix S_i and the total intra-class dispersion matrix S_w

$$S_i = \sum_{X \in \omega_i} (X - m_i)(X - m_i)^T, i = 1, 2 \tag{2}$$

$$S_w = S_1 + S_2 \tag{3}$$

(3) Calculate the sample inter-class dispersion matrix S_b

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \tag{4}$$

(4) Find the vector w^*

$$w^* = S_w^{-1}(m_1 - m_2) \tag{5}$$

(5) thus defining Fisher's criterion function

$$J_F(W) = \frac{\omega^T S_b \omega}{\omega^T S_w \omega} \tag{6}$$

(5) Projection of all samples within the training set

$$y = (w^*)^T X \tag{7}$$

(6) Calculate the segmentation threshold on the projection space . The selection of the threshold value can have different schemes, and the method used in this paper is.

$$y_0 = \frac{N_1 \bar{m}_1 + N_2 \bar{m}_2}{N_1 + N_2} \tag{8}$$

(7) For a given sample X, compute its projection point y on the vector w^*

$$y = (w^*)^T X \tag{9}$$

(8) According to the classification of decision rules can be discerned

$$\begin{cases} y > y_0 \Rightarrow X \in w_1 \\ y < y_0 \Rightarrow X \in w_2 \end{cases} \tag{10}$$

From this, the unknown artifacts can be identified as belonging to high potassium glass or lead-barium glass.

Then this paper uses the classify function for validation and sensitivity analysis, using all known categories of samples as training data and classifying all unknown samples, the format of the classify function is

$$[class, err] = classify(sample, training, group) \tag{11}$$

The function of the classify function is to discriminate each column of the sample into the class specified by training[8]. The results of the classification function are consistent with those of Fisher's linear discriminant analysis, and the error ratio is 0.0352, so the results are more accurate.

Finally, the sensitivity test was performed by reclassifying the training sample after downward adjustment of the overall data, and observing whether the classification results were consistent in order to obtain the range of sensitivity change.

3.1.2 Model solving

In this paper, the weathered artifacts and unweathered artifacts are firstly separated by data preprocessing, and the unweathered artifacts data are stored in the appendix wfhh.mat. The MATLAB code is written according to the above model to discriminate the unknown category of weather-free artifacts by Fisher's linear discriminant analysis algorithm.

The running results are shown in Table 2.

Table 2 No weathering unknown artifact identification results table

Artifact Number	Running results	Properties
A1	1	High potassium glass
A3	0	Barium lead glass
A4	0	Lead barium glass
A8	1	High potassium glass

From the table, we can see that A1 and A8 are identified as high potassium glass, and A3 and A4 are identified as lead-barium glass.

Similarly, the MATLAB code was written to discriminate the weathered artifacts of unknown category by Fisher's linear discriminant analysis algorithm according to the above model. The results are shown in Table 3.

Table 3 Weathering unknown artifact identification results table

Artifact Number	Running results	Properties
A2	0	Barium lead glass
A5	0	Lead barium glass
A6	1	High potassium glass
A7	1	High potassium glass

From the table can be seen A2, A5 identified as high potassium glass, A6, A7 identified as lead barium glass.

In summary, the results of the analysis of the chemical composition of the unknown category glass artifacts of the base data species are A1, A2, A5, and A8 identified as high potassium glass, and A3, A4, A6, and A7 identified as lead-barium glass.

Finally, this paper uses the classify function for validation and sensitivity analysis, and uses all known category samples as training data and classifies all unknown samples. The results obtained through simulation are shown in Table 4.

Table 4 classify function identification results table

Unknown artifact serial number	Running results	Properties
A1	1	High potassium glass
A2	0	Barium lead glass
A3	0	Lead barium glass
A4	0	Lead barium glass
A5	0	Lead barium glass
A6	1	High potassium glass
A7	1	High Potassium Glass
A8	1	High Potassium Glass

In agreement with the results obtained by Fisher's linear discriminant analysis, A1, A6, A7, and A8 are high potassium glasses, and A2, A3, A4, and A5 are lead-barium glasses. And the error ratio is 0.0352, which is very small and the obtained results are accurate.

The sensitivity was tested by reclassifying the training samples after the overall data was adjusted downward and using the classify function to observe whether the classification results were consistent. First, the overall chemical element content was adjusted downward by 4%, and the classification results showed that the 4% downward adjustment of the data did not affect the classification. Then the overall chemical element content was adjusted downward by 5%, and the reclassification results are shown in Table 5.

Table 5 The classify function reclassifies the results

Unknown artifact serial number	Running results	Properties
A1	1	High potassium glass
A2	0	Barium lead glass
A3	0	Lead barium glass
A4	0	Lead barium glass
A5	0	Lead barium glass
A6	1	High potassium glass
A7	1	High Potassium Glass
A8	1	High Potassium Glass

From the results, it can be seen that its classification discrimination results show errors and differences in the discrimination of A8 artifacts, which shows that the sensitivity perturbation range of this model should be within 4%.

3.2 Chemical composition analysis of high potassium glass and lead-barium glass samples using Spearman correlation

3.2.1 Model building

In this paper, the unweathered data were firstly selected by data preprocessing, and the data were divided into two parts, high potassium glass and lead-barium glass, for analysis separately. The unmeasured chemical composition was taken as 0 according to the question.

In this paper, we first used the normality test on the screened data and found that the data did not satisfy the normality, so we used the Spearman correlation coefficient method in the nonparametric test to investigate the linear correlation between the chemical components. Spearman's rank correlation is a method to investigate the correlation between two variables based on rank information. It is calculated based on the difference between two pairs of ranks, and the requirements of Spearman's rank correlation for data conditions are not as strict as those of the product-difference correlation coefficient. correlation to conduct the study. The closer the correlation coefficient is to +1 or -1, the stronger the correlation between the two variables[9].

The Spearman's correlation coefficient formula is as follows

$$\rho=\frac{\sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)\left(y_{i}-\bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2}\sum_{i=1}^{n}\left(y_{i}-\bar{y}\right)^{2}}} \tag{12}$$

where n is the sample size, ρ is the correlation coefficient, and x, y are the corresponding elements of the two variables.

In practical applications or in specific topics where the link between variables is irrelevant and the corresponding elements of the two observed variables are subtracted to obtain a difference d, the above formula can also be transformed into.

$$\rho=1-\frac{6\sum d_{i}^{2}}{n\left(n^{2}-1\right)} \tag{13}$$

From this, the Spearman correlation test was conducted for high potassium glass and lead-barium glass respectively, and the elements with strong chemical composition correlations were derived based on the significance level and correlation coefficients, and the differences in the chemical composition correlations between different categories were inferred by analyzing their correlation coefficients separately.

3.2.2 Model Solving

The Spearman correlation coefficients and P-values for different chemical compositions of high potassium glasses were obtained using MATLAB, and the results of the solution are shown in

	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	BaO	P2O5	SrO	SnO2	SO2
SiO2	1.000(0.000***)	-0.422(0.172)	-0.622(0.031**)	-0.662(0.019**)	-0.074(0.820)	-0.448(0.145)	-0.718(0.009***)	-0.403(0.194)	-0.029(0.929)	-0.067(0.837)	-0.091(0.779)	-0.269(0.398)	0.393(0.206)	-0.064(0.843)
Na2O	-0.422(0.172)	1.000(0.000***)	0.661(0.019**)	0.736(0.006***)	-0.598(0.040**)	0.119(0.712)	-0.074(0.820)	-0.156(0.628)	0.143(0.658)	-0.393(0.206)	-0.404(0.193)	-0.358(0.254)	-0.172(0.593)	-0.325(0.302)
K2O	-0.622(0.031**)	0.661(0.019**)	1.000(0.000***)	0.827(0.001***)	-0.389(0.212)	-0.126(0.697)	0.004(0.991)	0.210(0.512)	-0.167(0.604)	-0.333(0.290)	-0.420(0.175)	-0.052(0.872)	-0.131(0.685)	0.101(0.755)
CaO	-0.662(0.019**)	0.736(0.006***)	0.827(0.001***)	1.000(0.000***)	-0.432(0.161)	0.095(0.770)	0.177(0.582)	0.221(0.490)	-0.167(0.604)	-0.452(0.140)	-0.459(0.134)	-0.404(0.193)	-0.437(0.155)	0.257(0.419)
MgO	-0.074(0.820)	-0.598(0.040**)	-0.389(0.212)	-0.432(0.161)	1.000(0.000***)	0.553(0.062*)	0.447(0.145)	0.025(0.940)	-0.113(0.728)	0.379(0.224)	0.546(0.066*)	0.628(0.029**)	0.131(0.684)	0.322(0.308)
Al2O3	-0.448(0.145)	0.119(0.712)	-0.126(0.697)	0.095(0.770)	0.553(0.062*)	1.000(0.000***)	0.588(0.044**)	-0.046(0.888)	0.326(0.301)	0.291(0.358)	0.154(0.633)	0.317(0.315)	-0.480(0.114)	-0.018(0.955)
Fe2O3	-0.718(0.009***)	-0.074(0.820)	0.004(0.991)	0.177(0.582)	0.447(0.145)	0.588(0.044**)	1.000(0.000***)	0.618(0.032**)	-0.004(0.991)	0.433(0.159)	0.557(0.060*)	0.335(0.288)	-0.437(0.155)	0.120(0.711)
CuO	-0.403(0.194)	-0.156(0.628)	0.210(0.512)	0.221(0.490)	0.025(0.940)	-0.046(0.888)	0.618(0.032**)	1.000(0.000***)	-0.058(0.858)	0.465(0.128)	0.200(0.534)	-0.009(0.977)	-0.481(0.113)	0.244(0.445)
PbO	-0.029(0.929)	0.143(0.658)	-0.167(0.604)	-0.167(0.604)	-0.113(0.728)	0.326(0.301)	-0.004(0.991)	-0.058(0.858)	1.000(0.000***)	0.608(0.036**)	-0.290(0.360)	0.135(0.675)	-0.317(0.315)	-0.600(0.039**)
BaO	-0.067(0.837)	-0.393(0.206)	-0.333(0.290)	-0.452(0.140)	0.379(0.224)	0.291(0.358)	0.433(0.159)	0.465(0.128)	0.608(0.036**)	1.000(0.000***)	0.283(0.373)	0.500(0.098*)	-0.208(0.517)	-0.393(0.206)
P2O5	-0.091(0.779)	-0.404(0.193)	-0.420(0.175)	-0.459(0.134)	0.546(0.066*)	0.154(0.633)	0.557(0.060*)	0.200(0.534)	-0.290(0.360)	0.283(0.373)	1.000(0.000***)	0.463(0.130)	0.306(0.334)	-0.018(0.955)
SrO	-0.269(0.398)	-0.358(0.254)	-0.052(0.872)	-0.404(0.193)	0.628(0.029**)	0.317(0.315)	0.335(0.288)	-0.009(0.977)	0.135(0.675)	0.500(0.098*)	0.463(0.130)	1.000(0.000***)	0.233(0.466)	-0.220(0.491)
SnO2	0.393(0.206)	-0.172(0.593)	-0.131(0.685)	-0.437(0.155)	0.131(0.684)	-0.480(0.114)	-0.437(0.155)	-0.481(0.113)	-0.317(0.315)	-0.208(0.517)	0.306(0.334)	0.233(0.466)	1.000(0.000***)	-0.172(0.593)
SO2	-0.064(0.843)	-0.325(0.302)	0.101(0.755)	0.257(0.419)	0.322(0.308)	-0.018(0.955)	0.120(0.711)	0.244(0.445)	-0.600(0.039**)	-0.393(0.206)	-0.018(0.955)	-0.220(0.491)	-0.172(0.593)	1.000(0.000***)

Table 6.

	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	BaO	P2O5	SrO	SnO2	SO2
SiO2	1.000(0.000***)	-0.422(0.172)	-0.622(0.031**)	-0.662(0.019**)	-0.074(0.820)	-0.448(0.145)	-0.718(0.009***)	-0.403(0.194)	-0.029(0.929)	-0.067(0.837)	-0.091(0.779)	-0.269(0.398)	0.393(0.206)	-0.064(0.843)
Na2O	-0.422(0.172)	1.000(0.000***)	0.661(0.019**)	0.736(0.006***)	-0.598(0.040**)	0.119(0.712)	-0.074(0.820)	-0.156(0.628)	0.143(0.658)	-0.393(0.206)	-0.404(0.193)	-0.358(0.254)	-0.172(0.593)	-0.325(0.302)
K2O	-0.622(0.031**)	0.661(0.019**)	1.000(0.000***)	0.827(0.001***)	-0.389(0.212)	-0.126(0.697)	0.004(0.991)	0.210(0.512)	-0.167(0.604)	-0.333(0.290)	-0.420(0.175)	-0.052(0.872)	-0.131(0.685)	0.101(0.755)
CaO	-0.662(0.019**)	0.736(0.006***)	0.827(0.001***)	1.000(0.000***)	-0.432(0.161)	0.095(0.770)	0.177(0.582)	0.221(0.490)	-0.167(0.604)	-0.452(0.140)	-0.459(0.134)	-0.404(0.193)	-0.437(0.155)	0.257(0.419)
MgO	-0.074(0.820)	-0.598(0.040**)	-0.389(0.212)	-0.432(0.161)	1.000(0.000***)	0.553(0.062*)	0.447(0.145)	0.025(0.940)	-0.113(0.728)	0.379(0.224)	0.546(0.066*)	0.628(0.029**)	0.131(0.684)	0.322(0.308)
Al2O3	-0.448(0.145)	0.119(0.712)	-0.126(0.697)	0.095(0.770)	0.553(0.062*)	1.000(0.000***)	0.588(0.044**)	-0.046(0.888)	0.326(0.301)	0.291(0.358)	0.154(0.633)	0.317(0.315)	-0.480(0.114)	-0.018(0.955)
Fe2O3	-0.718(0.009***)	-0.074(0.820)	0.004(0.991)	0.177(0.582)	0.447(0.145)	0.588(0.044**)	1.000(0.000***)	0.618(0.032**)	-0.004(0.991)	0.433(0.159)	0.557(0.060*)	0.335(0.288)	-0.437(0.155)	0.120(0.711)
CuO	-0.403(0.194)	-0.156(0.628)	0.210(0.512)	0.221(0.490)	0.025(0.940)	-0.046(0.888)	0.618(0.032**)	1.000(0.000***)	-0.058(0.858)	0.465(0.128)	0.200(0.534)	-0.009(0.977)	-0.481(0.113)	0.244(0.445)
PbO	-0.029(0.929)	0.143(0.658)	-0.167(0.604)	-0.167(0.604)	-0.113(0.728)	0.326(0.301)	-0.004(0.991)	-0.058(0.858)	1.000(0.000***)	0.608(0.036**)	-0.290(0.360)	0.135(0.675)	-0.317(0.315)	-0.600(0.039**)
BaO	-0.067(0.837)	-0.393(0.206)	-0.333(0.290)	-0.452(0.140)	0.379(0.224)	0.291(0.358)	0.433(0.159)	0.465(0.128)	0.608(0.036**)	1.000(0.000***)	0.283(0.373)	0.500(0.098*)	-0.208(0.517)	-0.393(0.206)
P2O5	-0.091(0.779)	-0.404(0.193)	-0.420(0.175)	-0.459(0.134)	0.546(0.066*)	0.154(0.633)	0.557(0.060*)	0.200(0.534)	-0.290(0.360)	0.283(0.373)	1.000(0.000***)	0.463(0.130)	0.306(0.334)	-0.018(0.955)
SrO	-0.269(0.398)	-0.358(0.254)	-0.052(0.872)	-0.404(0.193)	0.628(0.029**)	0.317(0.315)	0.335(0.288)	-0.009(0.977)	0.135(0.675)	0.500(0.098*)	0.463(0.130)	1.000(0.000***)	0.233(0.466)	-0.220(0.491)
SnO2	0.393(0.206)	-0.172(0.593)	-0.131(0.685)	-0.437(0.155)	0.131(0.684)	-0.480(0.114)	-0.437(0.155)	-0.481(0.113)	-0.317(0.315)	-0.208(0.517)	0.306(0.334)	0.233(0.466)	1.000(0.000***)	-0.172(0.593)
SO2	-0.064(0.843)	-0.325(0.302)	0.101(0.755)	0.257(0.419)	0.322(0.308)	-0.018(0.955)	0.120(0.711)	0.244(0.445)	-0.600(0.039**)	-0.393(0.206)	-0.018(0.955)	-0.220(0.491)	-0.172(0.593)	1.000(0.000***)

Table 6 High potassium glass Spearman correlation coefficient table

Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively

According to the significance levels and correlation coefficients, the following elemental chemical compositions were more significantly correlated, as shown in Table 7.

Table 7 High potassium glass correlation higher element correlation coefficient table

	Potassium oxide	Calcium oxide	Iron oxide	Magnesium oxide	Barium oxide	Sulfur Dioxide

Silicon dioxide	-0.622	-0.662	-0.718			
Sodium oxide	0.661	0.736		-0.598		
Potassium oxide		0.827				
strontium oxide				0.628		
Aluminum oxide		0.588				
copper oxide			0.618			
Lead oxide					0.608	-0.6

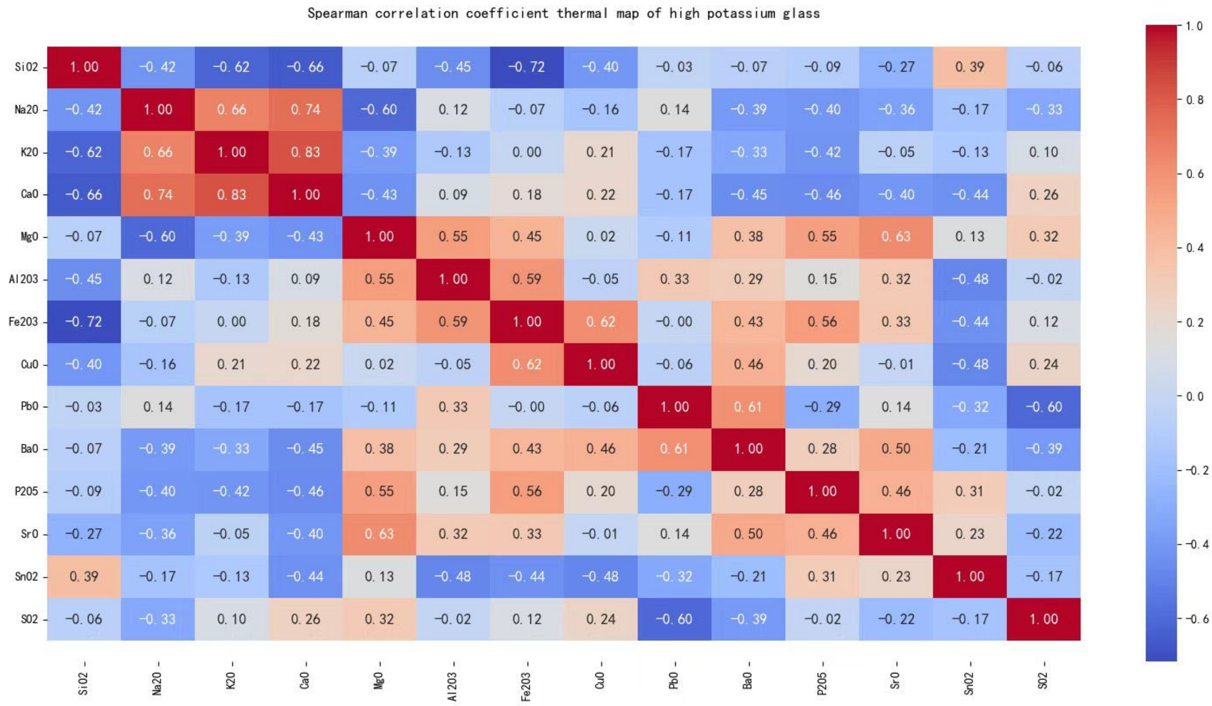

Figure 1 High potassium glass correlation heat map

Where a positive number indicates a positive correlation and a negative number indicates a negative correlation. The larger the absolute value of the correlation coefficient, the stronger the correlation. The heat map of correlations derived by visual analysis in Python is shown in Figure 1.

Similarly, MATLAB can be used to find the Spearman correlation coefficients and P-values for different chemical compositions of lead-barium glass, and the results of the solution are shown in Table 8.

Table 8 Lead barium glass Spearman correlation coefficient table

Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively

	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	BaO	P2O5	SrO	SnO2	SO2
SiO2	1.000(0.000***)	-0.071(0.818)	-0.136(0.658)	-0.262(0.388)	0.166(0.587)	-0.239(0.431)	-0.087(0.778)	-0.264(0.383)	-0.613(0.026**)	-0.740(0.004***)	0.033(0.915)	-0.600(0.030**)	-0.507(0.077*)	0.077(0.802)
Na2O	-0.071(0.818)	1.000(0.000***)	-0.011(0.970)	0.201(0.510)	-0.033(0.914)	-0.037(0.904)	-0.473(0.103)	0.201(0.510)	0.100(0.744)	0.011(0.971)	-0.400(0.176)	0.046(0.882)	-0.230(0.449)	-0.157(0.610)
K2O	-0.136(0.658)	-0.011(0.970)	1.000(0.000***)	-0.153(0.618)	0.273(0.366)	0.560(0.047**)	-0.043(0.889)	0.068(0.826)	-0.107(0.727)	0.175(0.567)	-0.241(0.429)	0.070(0.821)	0.072(0.815)	0.159(0.605)
CaO	-0.262(0.388)	0.201(0.510)	-0.153(0.618)	1.000(0.000***)	0.187(0.542)	0.006(0.986)	0.323(0.282)	-0.419(0.154)	0.481(0.096*)	-0.091(0.768)	-0.121(0.693)	0.297(0.324)	0.621(0.023**)	0.155(0.614)
MgO	0.166(0.587)	-0.033(0.914)	0.273(0.366)	0.187(0.542)	1.000(0.000***)	0.439(0.133)	0.016(0.959)	-0.339(0.258)	0.000(1.000)	-0.266(0.380)	-0.417(0.157)	0.241(0.428)	0.221(0.469)	-0.284(0.347)
Al2O3	-0.239(0.431)	-0.037(0.904)	0.560(0.047**)	0.006(0.986)	0.439(0.133)	1.000(0.000***)	0.287(0.342)	-0.030(0.922)	-0.264(0.384)	0.429(0.144)	0.006(0.986)	0.203(0.505)	0.280(0.355)	-0.077(0.802)
Fe2O3	-0.087(0.778)	-0.473(0.103)	-0.043(0.889)	0.323(0.282)	0.016(0.959)	0.287(0.342)	1.000(0.000***)	-0.613(0.026**)	0.084(0.786)	-0.149(0.626)	0.416(0.158)	-0.086(0.780)	0.560(0.046**)	-0.252(0.407)
CuO	-0.264(0.383)	0.201(0.510)	0.068(0.826)	-0.419(0.154)	-0.339(0.258)	-0.030(0.922)	-0.613(0.026**)	1.000(0.000***)	-0.278(0.358)	0.451(0.122)	0.208(0.495)	0.102(0.741)	-0.625(0.022**)	0.309(0.304)
PbO	-0.613(0.026**)	0.100(0.744)	-0.107(0.727)	0.481(0.096*)	0.000(1.000)	-0.264(0.384)	0.084(0.786)	-0.278(0.358)	1.000(0.000***)	0.093(0.762)	-0.256(0.399)	0.531(0.062*)	0.620(0.024**)	-0.154(0.615)
BaO	-0.740(0.004***)	0.011(0.971)	0.175(0.567)	-0.091(0.768)	-0.266(0.380)	0.429(0.144)	-0.149(0.626)	0.451(0.122)	0.093(0.762)	1.000(0.000***)	-0.041(0.894)	0.447(0.126)	0.236(0.438)	0.154(0.615)
P2O5	0.033(0.915)	-0.400(0.176)	-0.241(0.429)	-0.121(0.693)	-0.417(0.157)	0.006(0.986)	0.416(0.158)	0.208(0.495)	-0.256(0.399)	-0.041(0.894)	1.000(0.000***)	-0.235(0.440)	-0.105(0.733)	0.309(0.304)
SrO	-0.600(0.030**)	0.046(0.882)	0.070(0.821)	0.297(0.324)	0.241(0.428)	0.203(0.505)	-0.086(0.780)	0.102(0.741)	0.531(0.062*)	0.447(0.126)	-0.235(0.440)	1.000(0.000***)	0.359(0.228)	0.079(0.797)
SnO2	-0.507(0.077*)	-0.230(0.449)	0.072(0.815)	0.621(0.023**)	0.221(0.469)	0.280(0.355)	0.560(0.046**)	-0.625(0.022**)	0.620(0.024**)	0.236(0.438)	-0.105(0.733)	0.359(0.228)	1.000(0.000***)	-0.123(0.690)
SO2	0.077(0.802)	-0.157(0.610)	0.159(0.605)	0.155(0.614)	-0.284(0.347)	-0.077(0.802)	-0.252(0.407)	0.309(0.304)	-0.154(0.615)	0.154(0.615)	0.309(0.304)	0.079(0.797)	-0.123(0.690)	1.000(0.000***)

According to the significance levels and correlation coefficients, the following elemental chemical compositions were more significantly correlated, as shown in Table 9.

Table 9 Lead barium glass correlation higher element correlation coefficient table

	Lead oxide	Barium oxide	Strontium oxide	Aluminum oxide	Tin oxide	Copper oxide
Silicon dioxide	-0.613	-0.740	-0.600			
Potassium oxide				0.560		
calcium oxide					0.621	
Iron oxide					0.560	-0.613
copper oxide					-0.625	
Lead oxide					0.620	

Where a positive number indicates a positive correlation and a negative number indicates a negative correlation. The larger the absolute value of the correlation coefficient, the stronger the correlation. The heat map of correlations obtained through visual analysis in Python is shown in Figure 2.

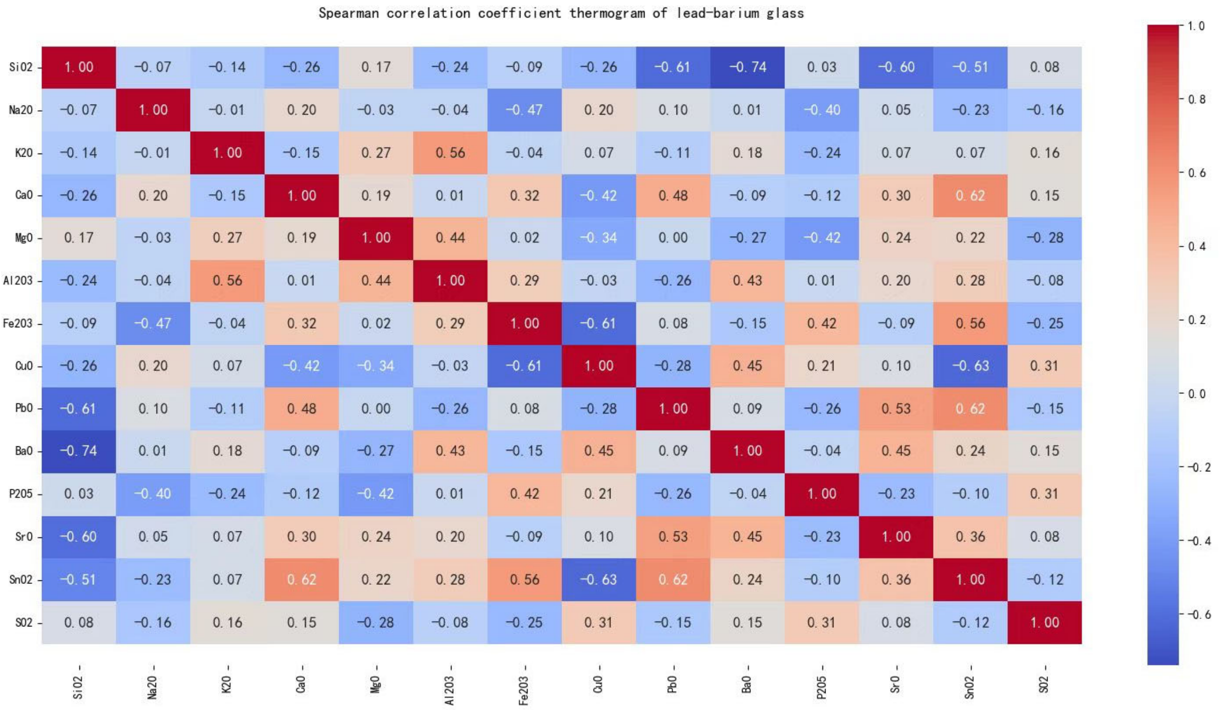

Figure 2 Lead barium glass related relationship heat map

From the above obtained Spearman's coefficient correlation coefficient, it can be seen that the chemical elements with strong correlation between high potassium glass and lead barium glass are copper oxide and iron oxide at the same time, except that the chemical components with significant correlation between high potassium glass and lead barium glass are different. The correlation coefficient between copper oxide and iron oxide in high-potassium glass is 0.618, while the correlation coefficient between copper oxide and iron oxide in lead-barium glass is -0.613. It can be seen that copper oxide in high-potassium glass has a strong positive correlation with iron oxide, while copper oxide in lead-barium glass has a strong negative correlation with iron oxide.

4. Conclusion

Fisher's linear discriminant analysis model was established to determine the linear discriminant function to test the unknown weathered artifacts and unweathered artifacts separately and identify the type of all unknown artifacts.

In this paper, the correlation between the chemical composition of two types of glass artifacts, high potassium glass and lead-barium glass, was analyzed separately using Spearman correlation analysis, and a heat map was created using Python software to visualize the analysis.

In this paper, the data are carefully screened and pre-processed before building the model to ensure the reliability and availability of the data, and also to improve the accuracy of the model.

By comparing and selecting the conditions of use and accuracy of results for several different models and methods, the models in this paper all have high stability and sensitivity.

Due to the large amount of data and the minimal content of some chemical components, only the chemical components that have an impact on the analysis and modeling are considered for analysis in this paper when clustering and identification of unknown artifacts, and not all of them are analyzed[10].

Model improvement can continue to improve the accuracy of the algorithm, reduce the time complexity of the model, and analyze and calculate all chemical components more comprehensively.

This paper classifies, identifies, and predicts the chemical composition of ancient glass products and analyzes the chemical composition correlation, which is important for the study of ancient historical relics in China. The prediction, classification, identification analysis and correlation test model established in this paper is not only applicable to the composition analysis and identification of ancient glass products, but also can be used to analyze systems with more diverse compositions and more complex relationships. The model is universal and has a good generalization value, considering a combination of various factors.

References

[1] J. P. Jia, X. Q. He, Y. Jin. Statistics (4th ed.) [M]. People's University of China Press, 2009.

[2] Mao Shisong, Wang Jinglong, Pu Xiaolong, et al. Advanced mathematical statistics (2nd ed.) [M]. Beijing: Higher Education Press, 2006.

[3] Sun JG, Liu J, Zhao LY. A study of clustering algorithms [J]. Journal of Software, 2008(1): 14.

[4] Xie JG, Qiu ZD. Fisher's linear discriminant model for unbalanced data sets [J]. Journal of Beijing Jiaotong University (Natural Science Edition), 2006.

[5] Han C-G. Mathematical modeling methods and their applications [M]. Higher Education Press, 2009.

[6] Green L R, Hart F A. Colour and chemical composition in ancient glass: an examination of some Roman and Wealden glass by means of ultraviolet-visible-infra-red spectrometry and electron microprobe analysis[J]. Journal of Archaeological Science, 1987, 14(3): 271-282.

[7] Artusi R, Verderio P, Marubini E. Bravais-Pearson and Spearman correlation coefficients: meaning, test of hypothesis and confidence interval[J]. The International journal of biological markers, 2002, 17(2): 148-151.

[8] Noushath S, Kumar G H, Shivakumara P. Diagonal Fisher linear discriminant analysis for efficient face recognition[J]. Neurocomputing, 2006, 69(13-15): 1711-1716.

[9] Mulcahy-Stanislawczyk J . Linear Discriminant Analysis and Fisher's Linear Discriminant.

[10] J Fröhlich, and C. E. Pfister . "Semi-infinite Ising model. I. Thermodynamic functions and phase diagram in absence of magnetic field." Communications in Mathematical Physics (1965-1997) (1987).