Exploration And Analysis of Weathering Factors of Glass Artefacts Based on Cluster Analysis and Grey Correlation Analysis

Mingquan Wei [a], Jinglei Xu [b] and Xusheng Xing [c]

School of Northeastern University, Northeastern University, Shenyang 110000, China.

[a] 1757973428@qq.com, [b] 2432741729@qq.com, [c] xingxs2022@163.com

Abstract. The aim of this paper is to investigate the relationship between the chemical composition of ancient glass artefacts after weathering and to construct a model for classifying and predicting the class of artefacts based on some of the characteristics of the ancient glass artefacts obtained. Cardinality tests, information entropy, cluster analysis, support vector machines and grey correlation are mainly used to conduct the study. Through chi-square test, the relationship between weathering of glass relics and glass type, glass color and glass pattern is explained. Finally, the relationship between weathering and glass type is closer. Information entropy was used to explain the relationship between the weathering of cultural relics and the content of chemical components contained in cultural relics. Among them, potassium oxide ($\boldsymbol{K_2O}$) and calcium oxide (**CaO**) had the largest information gain. The effective information gain of lead barium is silicon dioxide ($\mathbf{Si}\boldsymbol{O_2}$), potassium oxide (**PbO**), copper oxide (**CuO**) and sodium oxide ($\boldsymbol{Na_2O}$). By using cluster analysis, the law analysis and corresponding subcategories of high-potassium cultural relics and lead-barium cultural relics are made, and 3-5 subcategories are formed. Support vector machine was used to judge the type of glass relics, and the final result was that A1, A6 and A7 were high potassium. A2, A3, A4, A5, A8 are lead barium; The correlation coefficient matrix was used to analyze the correlation between the chemical components in different kinds of cultural relics, and the differences were explained based on the results of grey correlation degree analysis. Finally, the conclusion was drawn: in the firing process of high-potassium glass, the substance with high potassium content as the co-solvent did not have an essential effect on the formation of high-potassium glass. However, lead barium glass is different. As lead ore (high content of lead oxide and barium oxide) is added as co-solvent in the firing process, lead barium glass produced by firing contains a large amount of lead oxide and barium oxide.

Keywords: Chi-square test, Information entropy, Cluster analysis, Support vector machine, Grey correlation degree.

1. Background

The Silk Road was a trade channel for communication between China and the West in ancient times, among which jewelry glass was an important commodity for communication and trade between China and the West in ancient times. Early glass was often made into bead-shaped ornaments in West Asia and Egypt and introduced to China. After absorbing the technology, ancient Chinese glass was made with local materials, so the appearance was similar to foreign glass products, but the chemical composition was different.

In order to lower the melting point of the raw materials, the ancients added different chemicals to the glass to lower its melting point. Ancient commonly used fluxes are grass and wood ash, natural natrofoam, saltpeter and lead ore, and add limestone as a stabilizer, limestone will be converted into calcium oxide (**CaO**)) after calcination. The chemical composition of the resulting glass varies depending on the flux added. Based on this characteristic, modern archaeologists divide ancient glass into lead barium glass and high potassium glass. However, due to its age, glass can undergo weathering. The chemical composition of glass can react with the surrounding materials, resulting in changes in the amount of certain chemical elements, which makes it difficult to identify glass. Therefore, this paper hopes to determine the characteristics of different glass by analyzing the existing data of cultural relics, so as to provide help for archaeological research.

1.1. Data prescription

This article explores this using the data given in the attached excel sheet, which contains three forms.

Accessory Form 1 gives information on the classification of these objects, dividing them into two categories, high potassium and lead-barium, according to their internal chemical composition, A, B and C, according to their different decorations, and eight categories, blue-green, light blue, purple, dark green, dark blue, light green, black and green, according to their colour, and two categories, high potassium and lead-barium, according to their type.

Accessory Form 2 then gives the results of the determination of the chemical content of the different artefacts, while it is important to note that for some weathered artefacts, it is possible to sample twice, with some sampling points sampling the unweathered parts of the weathered artefacts for analysis. Examples include: artefact 42, artefact 49, artefact 50, while some sampling points do not indicate whether the sampling point was weathered or not. For example: No. 06, No. 43, No. 51, and the weathered artefacts themselves were sampled and more weathered points were sampled for their chemical content, for example: No. 08, No. 26, No. 54. There are also cases where multiple points were sampled for objects that were not weathered themselves, e.g., Artefact 03, Artefact 30. Accumulation and non-100% of the chemical composition ratios of some artefacts may occur as a result of testing methods etc. In this paper, data with a cumulative sum of composition ratios between 85% and 105% are considered valid. Therefore, the test data for artefacts No. 15 and 17 are discarded.

The data given in Form 3 in the Accessory is information on whether the artefacts are weathered and the content of the various chemical components in the artefacts, but the specific category to which the artefacts belong is not known; later in this paper we explain how to use the data in Forms 1 and 2 to model and use the model to predict the categories of the artefacts in Form 3.

2. Model assumptions and symbolic description

In order to avoid the influence of unpredictable factors on our research model, we have made the following assumptions to support the soundness of our model.

1. Cultural relics sampling point is random sampling.
2. There are no errors in the ornamentation, type or colour of the artefacts identified.
3. All artefacts are sampled in a consistent manner and the effects of weathering during the collection process are ignored.
4. Consistent accuracy of approximate results from sampling
5. The same type of glass artefacts is fired in the same way and the same fluxes are added.
6. The artefacts do not vary greatly from the time span of the collection, and the objects collected belong to the same historical period.

3. The relationship between weathering and surface ornamentation, type of glass, and color of glass

The data in Table 1 in the Accessory gives the following statistical chart

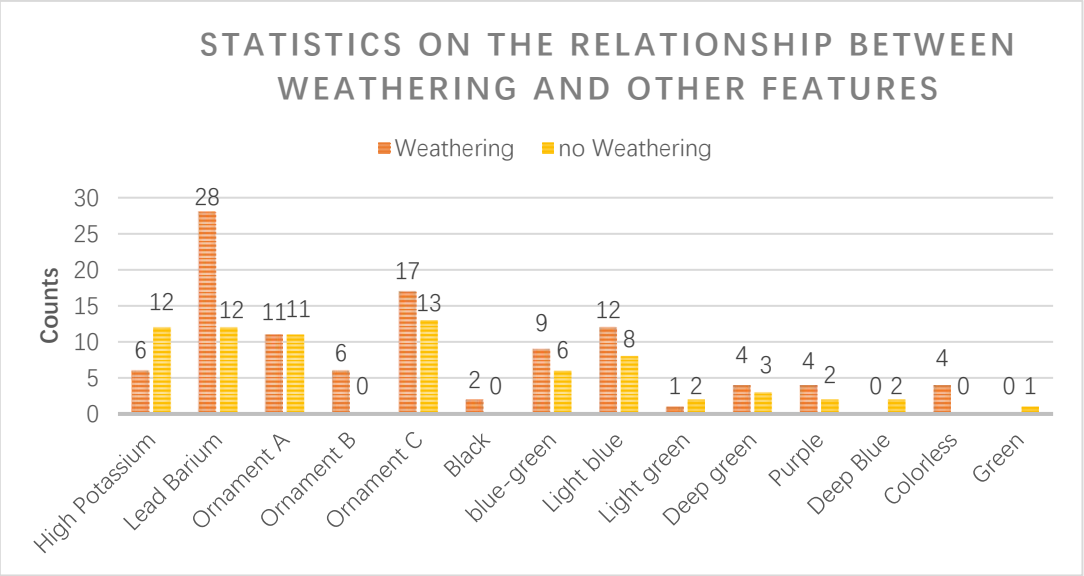


Fig. 1 Statistics on the relationship between weathering and other features

This paper uses a chi-square test to explore whether weathering and other characteristics are related to each other. A chi-square test is a method of hypothesis testing that is usually used to analyse comparisons between the composition of two variables and to analyse the relationship between two variables in terms of correlation.

The chi-square test is carried out by means of a hypothesis that first sets up the hypothesis between two variables: the two variables are unrelated, then based on this hypothesis the value of χ^2s derived, which indicates the degree of deviation between the observed and the theoretical values, and according to the χ^2distribution and degrees of freedom the probability value P of obtaining the current statistic and its more extreme cases, provided that the H_0 hypothesis holds.

If P is small, then this indicates a large deviation between the observed and theoretical values, and hypothesis H_0 is rejected, indicating a significant difference between the hypothesis and the actual situation. If the calculation yields a relatively large P, then the original hypothesis is valid [1].
Cardinality Test Formula

$$\chi^2 = \sum \frac{(V-T)^2}{T} \tag{1}$$

Where V denotes the observed value of a category and T is the theoretical value obtained based on assumptions.

In order to better analyse the required assumptions, this paper first calculates the weathering rate of glass artefacts for different properties

Table. 1 Weathering rates of glass artefacts under different characteristics

		Number of weathering	Number of unweathered	Weathering rate
Ornamentation Type	A	11	11	0.5000
	B	6	0	1.0000
	C	17	13	0.5667
Ornamentation Type	High Potassium	6	12	0.3333
	Lead Barium	28	12	0.7000
Colour	Blue and green	9	6	0.6000
	Pale blue	12	8	0.6000
	Purple	2	2	0.5000
	Dark green	4	3	0.5714
	Deep Blue	0	2	0.0000
	Light green	1	2	0.3333
	Black	2	0	1.0000
	Green	0	1	0.0000
Total by category		30	24	

A chi-square test was performed based on statistical regularity and 0.05 was used as the level of significance. Several hypotheses can be obtained as follows:

H_1:Assuming that the decoration is not related to weathering,Calculate the weathering rate as 58.62%

H_2:Assuming that the type is independent of weathering, Calculate the weathering rate as 58.62%

H_3:Assuming that colour is independent of weathering, Calculate the weathering rate as 5.55%

A table of the theoretical values is obtained as follows:

Table. 2 Theoretical values for decoration and weathering or not

Ornamentation	Weathering	Unweathered
A	12.90	9.10
B	3.51	2.48
C	17.59	12.41
Total	34.0	24.00

Table. 3 Theoretical values for type and weathering or not

Type	Weathering	Unweathered
High Potassium	10.55	7.45
Lead Barium	23.44	16.55
Total	34.00	24.00

Table. 4 Theoretical values for colour and weathering

Color	Weathering	Unweathered
Blue and green	8.33	6.67
Pale blue	11.11	8.89
Purple	1.11	0.89
Dark green	3.89	3.11
Deep Blue	0	0.89
Light green	1.67	1.33
Black	1.11	0
Green	0	0.44
Total	34.00	24.00

Table. 5 Diagram of the relationship between degrees of freedom and probabilities

Freedom\P value (Probability)	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52

After calculating the theoretical value, χ^2 can be calculated according to the theoretical value and observed value:

For the original hypothesis H_1, we calculated that $\chi^2 = 0.7211$ and the degree of freedom is $(2-1)*(2-1)=1$. According to the table, the cardinality distribution at the degree of freedom k=1 and p=0.05 is 3.84. The cardinality value is 0.7211, which is less than 3.84, so we accept the original hypothesis. That is, it is considered that the decoration has no significant effect on weathering.

For the original hypothesis H_2, $\chi^2 = 6.8804$ was calculated with the degree of freedom is $(2-1)*(2-1)=1$. According to the table, the cardinality distribution at the degree of freedom $k = 1, p = 0.05$ is 3.84. The cardinality value obtained by the calculation is 6.8804, which is greater than 3.84, so the original hypothesis is rejected. That is, the type is considered to have a significant effect on weathering.

For the original hypothesis H_3, the calculated $\chi^2 = 0.2879$ and the degree of freedom is$(3-1)*(2-1)=2$. According to the table, the cardinality distribution at the degree of freedom $k = 2, p = 0.05$ is 5.99. The calculated cardinality value is 0.2879, which is less than 5.99. Therefore, the original hypothesis is accepted, i.e., there is no significant effect of colour on weathering.

In summary, it can be concluded that the type of glass has a significant effect on weathering. It is clear from the data that lead-barium glass is more susceptible to weathering, and the calculations are supported by the literature reviewed, as the lead oxide in lead-barium glass reacts readily in the presence of carbon dioxide to form lead carbonate, which causes the glass to weather.

3.1. The law of chemical composition content of glass weathering

In order to study more specifically the relationship between different types of glass and whether they are weathered and their chemical elemental content, the data in Table II of the Accessoryes were first combined with those in Table I. The two types of glass were studied separately according to the different categories of glass, and for the two artefacts 30 and 03, which were not weathered and for which two test points were taken, the average of the chemical content of the corresponding elements of both was taken as the valid value.

This paper uses information entropy to investigate the relationship between different types of glass and whether they are weathered and their chemical element content. Information entropy represents a measure of the uncertainty of a random variable. The higher the entropy, the more information there is; the lower the entropy, the less information there is. Information gain represents the degree to which the uncertainty of information about a class is reduced by learning information about the feature [2]. In this paper, we use information gain to measure the extent to which individual chemical elements influence whether or not the glass eventually weathered.

Defining the information entropy in the data set of the chemical element content to be analysed as K, the expression is:

$$K = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i) \quad (2)$$

Probabilities can be expressed using statistical data on the number of individuals in each category:

$$\mathrm{p}(x_i) = \frac{|C_i|}{|D|} \tag{3}$$

Where D represents the entire data set, and C_i represents the number of samples in class i.

After determining the influence of one thing on the result, the influence of another thing on the result can be defined as the condition: $\mathrm{H}(Y|X)$, which expresses the uncertainty of random variable Y under the condition of known variable X, and its expression is as follows:

$$\mathrm{H}(Y|X) = -\sum_{i=1}^{n} p_i H(Y|X = x_i) \tag{4}$$

The expression of information gain can be obtained:

$$\mathrm{g}(\mathrm{D}, \mathrm{A}) = \mathrm{K}(\mathrm{D}) - \mathrm{H}(\mathrm{D}|\mathrm{A}) \tag{5}$$

The above formula can calculate the quantitative degree of influence of various chemical elements in different categories of cultural relics on whether they are weathered, that is, information gain.

In order to select the optimal information gain, set the information gain ratio to screen out the chemical elements with the greatest influenc.

$$\mathrm{F} = \frac{g(D,A)}{IV(a)} \tag{6}$$

$$IV(a) = -\sum_{v=1}^{V} \frac{D^v}{D} \log_2 \frac{D^v}{D} \tag{7}$$

Chemical elements with relatively low information gain were screened by information gain ratio.

In order to better run the model, it is necessary to process the content value of chemical elements. According to the average value of various samples of chemical elements, the element 75% lower than the average value is defined as low content, which is defined as label 0 in the actual calculation process. Values between 75% and 125% of the mean are defined as moderate content, which is defined as label 1 in the calculation, and values above 125% are defined as high content, which is defined as label 2 in the calculation.

After calculation, the following results can be obtained:

Table. 6 Information gain on weathering by chemical elements in high potassium type glasses

Chemical elements	Information gain
K_2O	0.693
CaO	0.592

Table. 7 Gain of information on the chemical elements of lead-barium type glass on weathering

Chemical elements	Information gain
SiO_2	0.661
PbO	0.504
K_2O	0.420
CuO	0.918
Na_2O	0.192

It can be seen from the above two tables that, during the weathering of high-potassium glass, the contents of potassium oxide and calcium oxide will change significantly compared with other chemical elements, and the weathering effect is more significant in the content of potassium oxide than in the content of calcium oxide. Combined with the content changes of chemical elements in the Accessory, it can be seen that the weathering effect will reduce the contents of potassium oxide and calcium oxide.

The weathering of lead-barium glass will affect more chemical elements, and the degree of influence on element content from high to low is as follows: silicon dioxide (SiO_2), lead oxide (PbO), potassium oxide (K_2O), copper oxide (CuO). According to the changes of element content in the data table, it can be concluded that weathering makes the content of silicon dioxide (SiO_2) decrease

significantly, the content of lead oxide (PbO), copper oxide (CuO) and other elements decrease slightly, while the content of potassium oxide (K_2O) increases during weathering.

According to the existing literature, the silicon is lost along the regolith from the inside out in the process of glass relics weathering, and the content in the outermost crust is very low. barium accumulates in the outermost crust to form a sandwich structure, and a large number of insoluble barium salt is formed in the outermost crust, which is easy to fall off and cause element loss. The contents of magnesium, calcium and other elements in the innermost unweathered parts are higher, indicating that these elements may come from the exchange process with the external burial environment, and gradually accumulate to the central unweathered parts. The content of Plumbum decreased gradually from outside to inside, but it was not obvious [3].

3.2. Predicting the chemical composition of glass when it is not weathered

Use the available data to predict the data when the weathering detection point is not weathered. The chemical composition content of weathered cultural relics can be described according to the glass cultural relics of the same type, color and pattern, because the chemical composition content of glass cultural relics of the same type and color is very similar. In this paper, the average value of each chemical component of the cultural relics is used to predict the content of chemical elements in the cultural relics.

The number of the detection points of weathered cultural relics in form 2 is 08 severely weathered point, 26 severely weathered point and 54 severely weathered point.

Among them, the type, color and pattern of the relics No. 08 and No. 26 are the same, so it is predicted that the chemical composition of the two relics is very similar when they are not weathered. Cultural relics of the same type, color and pattern include cultural relic No. 24 and cultural relic No. 31, which can be predicted according to the chemical composition values of these two cultural relics when they were not weathered.

For cultural relic No. 54, there are cultural relic No. 45, No. 46, and No. 47 with the same color, type, and decoration. The value of this cultural relic can be predicted by the value of these three cultural relics.

The results are as follow:

Table. 8 Analysis of the chemical elemental composition of artefacts before weathering

Heritage number	54After weathering	54Before weathering	26After weathering	27Before weathering	08After weathering	08Before weathering
SiO_2	17.11	56.01	3.72	48.93	4.61	48.93
Na_2O	0	2.44	0	0	0	0
K_2O	0	0.22	0.4	0	0	0
CaO	0	0.57	3.01	1.04	3.19	1.04
MgO	1.11	1	0	0.45	0	0.45
Al_2O_3	3.65	4.28	1.18	2.35	1.11	2.35
Fe_2O_3	0	0	0	2.3	0	2.3
CuO	1.34	0.65	3.6	4.45	3.14	4.45
PbO	58.46	22.21	29.92	22.85	32.45	22.85
BaO	0	10.08	35.45	14.825	30.62	14.825
P_2O_5	14.13	0.1	6.04	0.88	7.56	0.88
SrO	1.12	0.5	0.62	0.61	0.53	0.61
SnO_2	0	0	0	0	0	0
SO_2	0	0	15.95	0	15.03	0

4. Classification and subdivision of high potassium glass and lead barium glass

Statistical results were used to analyze the classification rules of high-potassium glass and lead-barium glass. The chemical composition given in the obtained data was statistically analyzed in this paper, and the corresponding classification rules of chemical elements were obtained. After certain classification rules are obtained, each category needs to be subdivided. In this paper, the method of cluster analysis is adopted. Firstly, R-type clustering is carried out for the given chemical components, 14 indicators are divided into 8 categories, and then a factor is extracted from each category as a representative factor, and then Q-type clustering is carried out. According to the 8 indexes selected by the R-type clustering, the cultural relics of each type are divided into subcategories.

Due to the small amount of content, data 15 and 17 in form 2 were removed. Since the data in form 2 has multiple detection points for the same cultural relic, it is necessary to integrate these data. The integration requirements are as follows:

1) If multiple detection points are of the same type (weathered points or unweathered points), the average value of the detection data is taken as the final detection result of the cultural relic.

2) If multiple detection points are not of the same type, the data of weathering points will be taken as the final detection result of the cultural relic.

This paper uses cluster analysis to classify cultural relics. Cluster analysis is a quantitative method, which can provide a more accurate and detailed classification tool from the point of view of data analysis.

Cluster analysis, also known as group analysis, is a multivariate statistical analysis method for quantitative classification of multiple samples (or indicators). The classification of samples is called Q-type cluster analysis, and the classification of indicators is called R-type cluster analysis [3].

4.1. Q-type clustering

Q-type clustering analysis needs to measure the similarity of samples and the similarity between classes, and the similarity of samples can be measured by distance.

Denoting as Ω sample point set, distance $d(\cdot,\cdot)$ is a function of $\Omega\times\Omega\to R^{+}$, satisfying the conditions:

$$\begin{aligned} &d(x,y)\geqslant 0, x,y\in\Omega \\ &d(x,y)=0 \text{ if and only if } x=y \\ &d(x,y)=d(y,x), x,y\in\Omega \\ &d(x,y)\leqslant d(x,z)+d(z,y), x,y,z\in\Omega \end{aligned} \tag{8}$$

This distance definition is widely used by the public. It satisfies positive quality, symmetry and triangle inequality. In cluster analysis, for quantitative variables, Minkowski Distance is most commonly used, i.e

$$d_q(x,y)=\left[\sum_{k=1}^{p}|x_k-y_k|^q\right]^{\frac{1}{q}}, q>0 \tag{9}$$

Where: $x=[x_1,x_2,\cdots,x_p]^T$; $y=[y_1,y_2,\cdots,y_p]^T$. When $q=1,2$ or $q\to+\infty$, we get:

1)Absolute distance

$$d_1(x,y)=\sum_{k=1}^{p}|x_k-y_k| \tag{10}$$

2)Euclidean distance

$$d_2(x,y)=\left[\sum_{k=1}^{p}|x_k-y_k|^2\right]^{\frac{1}{2}} \tag{11}$$

3)Chebyshev Distance

$$d_\infty(x,y)=\max_{1\leqslant k\leqslant p}|x_k-y_k| \tag{12}$$

There are four methods to measure similarity between classes.

If you have two sample classes $G_1$and G_2, you can measure the distance between them in the following four ways:

1)Shortest distance method

$$D(G_1,G_2)=\min_{\substack{x_i\in G_1\\ y_i\in G_2}}|d(x_i,y_i)| \tag{13}$$

2)Longest distance method

$$D(G_1,G_2)=\max_{\substack{x_i\in G_1\\ y_i\in G_2}}|d(x_i,y_i)| \tag{14}$$

3)Center of gravity method

$$D(G_1,G_2)=d(\bar{x},\bar{y}) \tag{15}$$

In the above equation: $\bar{x},\bar{y}$ is the centre of gravity of G_1,G_2 respectively

4)Class average method

$$D(G_1,G_2)=\frac{1}{n_1n_2}\sum_{x_i\in G_1}\sum_{x_j\in G_2}d(x_i,x_j) \tag{16}$$

It is equal to the average of the distance between the two sample points in $G_1,G_2$and $n_1,n_2$is the number of sample points in G_1,G_2, respectively.

4.2. R-type clustering

R-type clustering method needs to consider variable similarity measure and variable clustering method.

In the process of cluster analysis, the similarity measure of variables should be determined first. There are two commonly used similarity measures.

1)Correlation coefficient

Let's call the variable x_j, $[x_{1j},x_{2j},\cdots,x_{mj}]^T\in R^m(j=1,2,\cdots,p)$. Then, the sample coefficients of two variables x_j and x_k can be used as their similarity measurement, i.e

$$r_{jk}=\frac{\sum_{i=1}^{m}(x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k)}{\left[\sum_{i=1}^{m}(x_{ij}-\bar{x}_j)^2\sum_{i=1}^{m}(x_{ik}-\bar{x}_k)^2\right]^{\frac{1}{2}}},j,k=1,2,\cdots,p \tag{17}$$

Where: $\bar{x}_j=\frac{1}{m}\sum_{i=1}^{m}x_{ij}$

2)Cosine of angle

$$r_{jk}=\frac{\sum_{i=1}^{m}x_{ij}x_{ik}}{\left(\sum_{i=1}^{m}x_{ij}^{2}\sum_{i=1}^{m}x_{ik}^{2}\right)^{\frac{1}{2}}},j,k=1,2,\cdots,p \tag{18}$$

The variable clustering method of R-type cluster analysis is similar to Q-type cluster analysis. This paper uses Q-type cluster analysis and R cluster analysis method to analyze the glass type of combining and classification, the result is more reasonable.

For the data in form 2, after data preprocessing, Q-type clustering method is adopted as the input matrix, and class average method is adopted for distance measurement. Eight categories are selected as the main factors according to the dimension of factors, and then R-type clustering analysis is carried out for classification.

The same procedure is required for high potassium and lead barium artifacts respectively. Take high-potassium cultural relics as an example:

Firstly, data standardization is carried out to eliminate the influence of other factors such as dimension. Then, the correlation coefficient matrix is solved to cluster by class average method. According to the clustering results, the clustering diagram is drawn and the R-type clustering analysis results are obtained. See table below.

Table 9 Results of Q type classification of high potassium

Category	Chemical composition contained in clusters
Category 1	9(PbO) 11(P_2O_5) 12(SrO)
Category 2	5(MgO) 6(Al_2O_3) 7(Fe_2O_3)
Category 3	8(CuO)
Category 4	14(SO_2)
Category 5	10(BaO)
Category 6	2(Na_2O) 3(K_2O) 4(CaO)
Category 7	1(SiO_2)
Category 8	13(SnO_2)

After the results are obtained, the actual data are used to select appropriate representative factors for each category. Considering the incomplete data, more complete data is used for Q - type clustering. That is, Class 1 takes No. 11 (phosphorus pentoxide), Class 2 takes No. 6 (alumina), and Class 6 takes No. 4 (calcium oxide).

After the selection is completed, the selected 8 factors are used as the basis for R-type cluster analysis for classification, and the classification results are shown in the table below:

Table. 10 Results of R-type clustering of high potassium

category	Cultural relic number
category1	3、7
category2	1、4、5、6、9、10、12、13、14、16、18、22、27
category3	21

Table. 11 R-type clustering of high potassium was divided into four categories

category	Cultural relic number
category1	1、5、6、14、16、18
category2	4、9、10、12、13、22、27
category3	3、7
category4	21

Table. 12 High potassium R-type clustering into five categories

category	Cultural relic number
category1	4、22
category2	9、10、12、13、27
category3	1、5、6、14、16、18
category4	3、7
category5	21

The same is true for the classification of Pb and barium cultural relics. Q-type cluster map and results, R-type cluster map and results are in the appendix.

Cluster analysis only analyzes the hidden correlation between data from the perspective of data analysis, and then obtains the results without the participation of human factors, which has a certain degree of objectivity. At the same time, Q-type clustering was carried out before classification. While reducing the data dimension, some factors of low data record quantity could also be removed by artificial selection without affecting the result.

According to the final classification results, it can be seen that cultural relics divided into three categories independently will not be divided into other groups when the number of classifications is increased, which maintains the stability of the grouping. See Tables 7 to 9 for details.

In order to further verify the sensitivity of the classification method, changes occurred when the content of each chemical component was changed by 5% in turn to observe the classification results. According to the experimental results, a change of 5% in each chemical composition will not affect the final classification result. This shows that the classification method has a certain range of application, the sensitivity is not high, but has a certain degree of fault tolerance, can adapt to more situations, the model is relatively stable.

5. Classification of glass artefacts of unknown category

Analyze the chemical composition in the list and identify the category to which it belongs. Use the existing data to generate a model that can predict the unknown data, have the model learn the classification features in the existing data, and classify the artifacts in Form 3. The support vector machine model in machine learning is used to analyze the chemical composition of glass relics of unknown class.

The data set was divided into training set and test set, and the marked results were as follows: high potassium: 1; Lead barium: -1.

In this paper, the method of support vector machine is used to classify the unknown type of glass relics. Support vector machine (SVM) is a kind of linear classifier for binary classification of data according to supervised learning. The boundary of its decision is the hyperplane of the maximum margin to solve the learning sample, which can transform the problem into a convex quadratic programming problem. It has a wide range of applications in the forecasting field. When the original data is linearly separable, the support vector machine can generate the optimal classification hyperplane in the original space. When the linear is not separable, the relaxation factor is added to map the low-latitude data to the high-dimensional space to make it linearly separable. In this hyperplane, the optimal classification hyperplane is sought to achieve the purpose of classifying unknown data [4][5].

It is a typical binary classification problem to divide cultural relics into two categories according to their characteristics. Glass of unknown type is divided into two categories: high potassium and lead barium. Therefore, the ideal result can be obtained by using support vector machine.

It is assumed that data can be separated by a hyperplane, and the expression of the partitioned hyperplane is:

$$w^T x + b = 0 \tag{19}$$

Where $w = (w_1, w_2, \ldots, w_n)$ is the normal vector, determining the direction of the hyperplane, and n represents the dimension of the data. In this study, $n = 14$ is chosen because there are fourteen chemical components. b is the displacement term that determines the distance between the hyperplane and the origin of the coordinates. Similar to the equation of a line, this hyperplane can be determined by the normal vector w and the displacement term b.

According to the extension of Euclidean distance in hyperplane, the distance d from samples in n-dimensional space to hyperplane can be expressed as:

$$d = \frac{|w^T x_i + b|}{||w||} \tag{20}$$

So, the distance from the sample set to the hyperplane:

$$\rho = \min_{(x_i, y_i) \in S} \frac{y_i(w^T x_i + b)}{||w||} = \frac{a}{||w||} \tag{21}$$

In the above equation, y_i represents the classification label of data set i, where high potassium is set as 1 and lead barium as -1.

The optimization objective is:

$$\max_{w,b} \frac{a}{||w||} \text{ s.t. } y_i(w^T x_i + b) \geq a, \forall i \tag{22}$$

It is assumed that the hyperplane $w^T x + b = 0$ can correctly classify the training samples, and a=1 is taken as the following formula for $(x_i, y_i) \in D$, where D is the known data set.

$$\begin{cases} w^T x_i + b \geq +1, & y = 1 \\ w^T x_i + b \leq -1, & y = -1 \end{cases} \tag{23}$$

The values of the sample points closest to the hyperplane that make the above inequality medium valid are called support vectors, and the distance between two different classes of support vectors and the hyperplane is $\frac{2}{||w||}$

In order to maximize the interval, the equation for solving the convex quadratic programming problem becomes:

$$\min_{w,b} \frac{1}{2}||w||^2 \text{ s.t. } y_i(w^T x_i + b) \geq 1, i = 1,2,\ldots,14 \tag{24}$$

The classification of unknown glass is based on the hyperplane equation trained by the known data set. If 14 chemical components of unknown glass are input into the decision formula as x_i vectors, a value greater than 1 or less than -1 will be obtained. If it is greater than 1, it will be judged as high potassium; if it is less than -1, it will be judged as lead barium.

Considering that weathering and unweathering can have different effects on the results, weathering and unweathering data were trained separately and forecasts were made separately.

Using the hyperplane equation obtained, the final results are calculated in the following table:

Table. 13 Heritage judgment results

(Weathered on the left, unweathered on the right)

Cultural relic number	Decision result	Cultural relic number	Decision result
A2	Lead and barium	A1	High Potassium
A5	Lead and barium	A3	Lead and barium
A6	High Potassium	A4	Lead and barium
A7	High Potassium	A8	Lead and barium

According to the previous criteria for high potassium and lead barium glass. If the content of lead or barium is high, the glass relic can be basically judged as lead barium glass; if the content of lead

or barium is low, and the content of potassium is higher than the average, it can be roughly judged as high-potassium glass. The contents of chemical elements in cultural relics of unknown types are calculated as follows:

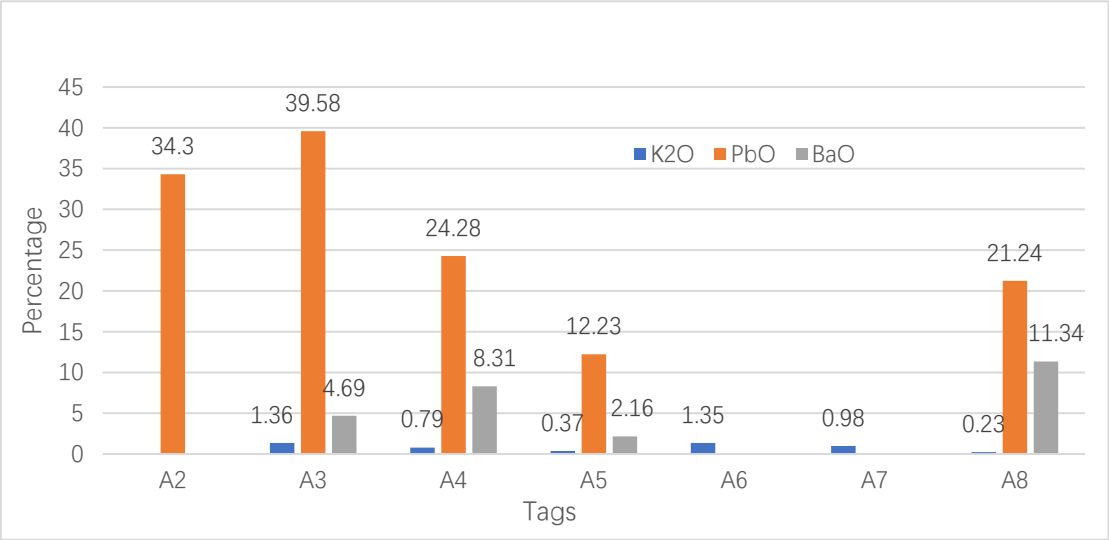


Fig. 2 Content values of important chemical elements in cultural relics to be tested

It can be seen intuitively from the chart that the content of lead in cultural relics A2, A3, A4, A5 and A8 is relatively high. At the same time, for the two cultural relics A6 and A7, which almost do not contain lead barium but contain potassium, the result obtained by support vector machine is high potassium glass, so the result obtained by support vector machine is consistent with the statistical law, and there is a certain rationality.

Secondly, from the principle of support vector machine, the expression of hyperplane obtained is the result of learning more than 50 groups of known data, and it is an extraction of the features of more than 50 groups of data. The feature vector contains the feature extraction of 14 kinds of chemical elements, and each chemical element has a corresponding weight coefficient. So, support vector machines can optimally partition all possible situations.

Finally, from the perspective of classification, this paper not only focuses on the chemical elements contained in glass cultural relics, but also considers the influence of cultural relics weathering on the content of chemical elements. Therefore, two independent hyperplane equations are obtained by independent training of weathering and unweathering for different prediction objectives. The classification of unknown types of cultural relics is also carried out according to whether they are weathered cultural relics, so the data finally obtained is reasonable.

To explore the sensitivity of the model. The content of each chemical element can be gradually increased, and then the judgment continues. The results are as follows:

Table. 14 Sensitivity analysis

increment	5%	10%
Previous results (unweathered)	1,-1,-1,-1	1,-1,-1,-1
Now results (unweathered)	1,-1,-1,-1	1,-1,-1,-1
change	No	No

According to the results, it can be seen that when the chemical content changes by 10%, the classification result does not change, which shows that the classification method has a certain stability, so it has a certain range of application.

6. Chemical component correlation

This paper intends to use the correlation coefficient matrix to calculate the chemical element content of each type of cultural relic before and after weathering to explain the relationship between different chemical elements of the same type of cultural relic and the gray correlation degree analysis to get the chemical elements with strong correlation of each type of cultural relic, and explain the differences between different types of cultural relic through the results.

Correlation coefficient is a measure that studies the degree of linear correlation between variables and is generally denoted by the letter r. Due to different research objects, correlation coefficient can be defined in various ways, and Pearson correlation coefficient is more commonly used. If each variable has N scalar observations, Pearson correlation coefficient is defined as:

$$\rho(A,B)=\frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{A_i-\mu_A}{\sigma_A}\right)\left(\frac{B_i-\mu_B}{\sigma_B}\right) \tag{25}$$

Where: μ_A and σ_A are respectively the mean and standard deviation of A, and μ_B and σ_B are the mean and standard deviation of B.

Grey correlation analysis method is a method to measure the degree of correlation among factors according to the similarity or dissimilarity of the development trend among factors, which is also called "grey correlation degree".

According to the concept of correlation coefficient, the correlation coefficient matrix of each element content before and after weathering of different cultural relics was calculated respectively.

The difference between different types of cultural relics is solved by grey correlation analysis. The conclusion can be drawn by using the content matrix of chemical elements of different cultural relics as input and grey relational degree analysis.

According to the classified data, the correlation coefficient matrix of high-potassium weathering before, high-potassium weathering after, lead-barium weathering before and lead-barium weathering after were respectively solved.

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
SiO_2	1.000	−0.405	−0.619	−0.678	−0.187	−0.558	−0.571	−0.473	−0.133	−0.097	−0.168	−0.287	0.413	−0.181
Na_2O	−0.405	1.000	0.541	0.585	−0.593	0.187	−0.225	−0.132	0.273	−0.359	−0.371	−0.385	−0.170	−0.323
K_2O	−0.619	0.541	1.000	0.526	−0.206	−0.001	−0.157	0.053	−0.067	−0.330	−0.312	0.096	0.007	0.143
CaO	−0.678	0.585	0.526	1.000	−0.352	0.081	0.185	0.348	0.164	−0.209	−0.461	−0.400	−0.543	0.313
MgO	−0.187	−0.593	−0.206	−0.352	1.000	0.437	0.433	0.004	−0.095	0.307	0.574	0.641	0.210	0.302
Al_2O_3	−0.558	0.187	−0.001	0.081	0.437	1.000	0.551	0.009	0.225	0.205	0.609	0.440	−0.451	−0.138
Fe_2O_3	−0.571	−0.225	−0.157	0.185	0.433	0.551	1.000	0.499	−0.063	0.341	0.709	0.464	−0.365	0.087
CuO	−0.473	−0.132	0.053	0.348	0.004	0.009	0.499	1.000	0.063	0.501	0.081	0.067	−0.465	0.248
PbO	−0.133	0.273	−0.067	0.164	−0.095	0.225	−0.063	0.063	1.000	0.576	−0.216	0.011	−0.220	−0.418
BaO	−0.097	−0.359	−0.330	−0.209	0.307	0.205	0.341	0.501	0.576	1.000	0.273	0.488	−0.192	−0.364
P_2O_5	−0.168	−0.371	−0.312	−0.461	0.574	0.609	0.709	0.081	−0.216	0.273	1.000	0.681	−0.009	−0.181
SrO	−0.287	−0.385	0.096	−0.400	0.641	0.440	0.464	0.067	0.011	0.488	0.681	1.000	0.184	−0.229
SnO_2	0.413	−0.170	0.007	−0.543	0.210	−0.451	−0.365	−0.465	−0.220	−0.192	−0.009	0.184	1.000	−0.173
SO_2	−0.181	−0.323	0.143	0.313	0.302	−0.138	0.087	0.248	−0.418	−0.364	−0.181	−0.229	−0.173	1.000

Fig. 3 Correlation coefficient matrix of high potassium weathering

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
SiO_2	1.000	0.000	−0.385	−0.819	−0.825	−0.883	−0.990	0.478	−0.997	0.000	−0.993	−0.997	0.000	0.000
Na_2O	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
K_2O	−0.385	0.000	1.000	0.165	0.207	0.209	0.495	−0.714	0.426	0.000	0.304	0.426	0.000	0.000
CaO	−0.819	0.000	0.165	1.000	0.897	0.978	0.787	−0.390	0.780	0.000	0.809	0.780	0.000	0.000
MgO	−0.825	0.000	0.207	0.897	1.000	0.962	0.806	−0.639	0.799	0.000	0.789	0.799	0.000	0.000
Al_2O_3	−0.883	0.000	0.209	0.978	0.962	1.000	0.856	−0.500	0.853	0.000	0.865	0.853	0.000	0.000
Fe_2O_3	−0.990	0.000	0.495	0.787	0.806	0.856	1.000	−0.563	0.995	0.000	0.974	0.995	0.000	0.000
CuO	0.478	0.000	−0.714	−0.390	−0.639	−0.500	−0.563	1.000	−0.499	0.000	−0.384	−0.499	0.000	0.000
PbO	−0.997	0.000	0.426	0.780	0.799	0.853	0.995	−0.499	1.000	0.000	0.988	1.000	0.000	0.000
BaO	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
P_2O_5	−0.993	0.000	0.304	0.809	0.789	0.865	0.974	−0.384	0.988	0.000	1.000	0.988	0.000	0.000
SrO	−0.997	0.000	0.426	0.780	0.799	0.853	0.995	−0.499	1.000	0.000	0.988	1.000	0.000	0.000
SnO_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
SO_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Fig. 4 Correlation coefficient matrix of high potassium weathering

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
SiO_2	1.000	−0.021	−0.333	−0.315	0.203	0.141	−0.128	−0.501	−0.613	−0.706	−0.288	−0.456	−0.388	0.101
Na_2O	−0.021	1.000	−0.170	−0.300	0.117	−0.015	−0.362	0.117	−0.145	0.071	−0.393	0.088	−0.271	−0.155
K_2O	−0.333	−0.170	1.000	0.334	0.201	0.220	0.184	−0.048	0.114	0.166	0.078	−0.045	0.449	0.008
CaO	−0.315	−0.300	0.334	1.000	0.155	0.352	0.455	−0.401	0.474	−0.287	0.200	0.168	0.710	−0.073
MgO	0.203	0.117	0.201	0.155	1.000	0.350	−0.011	−0.295	−0.129	−0.313	−0.109	0.072	0.105	−0.255
Al_2O_3	0.141	−0.015	0.220	0.352	0.350	1.000	0.047	−0.202	−0.413	−0.259	0.056	−0.058	−0.027	−0.116
Fe_2O_3	−0.128	−0.362	0.184	0.455	−0.011	0.047	1.000	−0.230	0.154	−0.150	0.192	0.005	0.420	−0.139
CuO	−0.501	0.117	−0.048	−0.401	−0.295	−0.202	−0.230	1.000	−0.089	0.859	0.092	0.366	−0.264	0.175
PbO	−0.613	−0.145	0.114	0.474	−0.129	−0.413	0.154	−0.089	1.000	0.077	−0.036	0.329	0.548	−0.129
BaO	−0.706	0.071	0.166	−0.287	−0.313	−0.259	−0.150	0.859	0.077	1.000	0.132	0.391	−0.002	0.050
P_2O_5	−0.288	−0.393	0.078	0.200	−0.109	0.056	0.192	0.092	−0.036	0.132	1.000	−0.153	−0.051	0.048
SrO	−0.456	0.088	−0.045	0.168	0.072	−0.058	0.005	0.366	0.329	0.391	−0.153	1.000	0.135	0.037
SnO_2	−0.388	−0.271	0.449	0.710	0.105	−0.027	0.420	−0.264	0.548	−0.002	−0.051	0.135	1.000	−0.080
SO_2	0.101	−0.155	0.008	−0.073	−0.255	−0.116	−0.139	0.175	−0.129	0.050	0.048	0.037	−0.080	1.000

Fig. 5 Correlation coefficient matrix of Pb barium before weathering

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
SiO_2	1.000	0.456	0.272	−0.250	0.128	0.603	0.272	−0.295	−0.262	−0.440	−0.274	−0.503	0.519	−0.597
Na_2O	0.456	1.000	−0.055	−0.362	−0.178	0.075	−0.104	−0.240	−0.074	−0.089	−0.359	−0.152	0.177	−0.131
K_2O	0.272	−0.055	1.000	0.059	0.355	0.363	0.339	−0.193	−0.139	−0.137	−0.081	−0.120	0.112	0.065
CaO	−0.250	−0.362	0.059	1.000	0.617	0.223	0.540	−0.043	−0.096	−0.196	0.475	−0.013	0.052	0.039
MgO	0.128	−0.178	0.355	0.617	1.000	0.589	0.660	−0.325	−0.028	−0.525	0.460	0.099	0.298	−0.311
Al_2O_3	0.603	0.075	0.363	0.223	0.589	1.000	0.511	−0.305	−0.402	−0.379	0.099	−0.125	0.849	−0.235
Fe_2O_3	0.272	−0.104	0.339	0.540	0.660	0.511	1.000	−0.283	−0.112	−0.459	0.337	−0.173	0.175	−0.268
CuO	−0.295	−0.240	−0.193	−0.043	−0.325	−0.305	−0.283	1.000	−0.391	0.660	−0.021	0.038	−0.179	0.241
PbO	−0.262	−0.074	−0.139	−0.096	−0.028	−0.402	−0.112	−0.391	1.000	−0.539	−0.120	0.353	−0.456	−0.343
BaO	−0.440	−0.089	−0.137	−0.196	−0.525	−0.379	−0.459	0.660	−0.539	1.000	−0.199	−0.002	−0.108	0.715
P_2O_5	−0.274	−0.359	−0.081	0.475	0.460	0.099	0.337	−0.021	−0.120	−0.199	1.000	0.273	−0.147	0.086
SrO	−0.503	−0.152	−0.120	−0.013	0.099	−0.125	−0.173	0.038	0.353	−0.002	0.273	1.000	−0.131	0.173
SnO_2	0.519	0.177	0.112	0.052	0.298	0.849	0.175	−0.179	−0.456	−0.108	−0.147	−0.131	1.000	−0.086
SO_2	−0.597	−0.131	0.065	0.039	−0.311	−0.235	−0.268	0.241	−0.343	0.715	0.086	0.173	−0.086	1.000

Fig. 6 Correlation coefficient matrix of Pb barium after weathering

For high potassium glass, the correlation coefficient between phosphorus pentoxide and iron oxide is high before weathering, while the correlation coefficient between many chemical elements is high after weathering, indicating that a lot of chemical reactions occurred in the differentiation process of high potassium glass, which changed the chemical composition ratio of the original high potassium glass to a certain extent.

Before weathering, the relationship between tin oxide and calcium oxide, barium oxide and copper oxide are close, while after weathering, the relationship between tin oxide and alumina, sulfur dioxide and barium oxide are close. Comparing before and after weathering, it can be seen that the proportion of chemical composition content of lead-barium glass does not change too much, indicating that lead-barium cultural relics have fewer chemical reactions during weathering and have higher chemical properties, and potassium is more stable.

The difference between the two kinds of cultural relics can be explained according to the results of grey correlation analysis. The concrete steps of grey correlation analysis are as follows:

Determine the comparison sequence and the reference sequence

Let there be m evaluation objects and an evaluation indicator variable, and each indicator variable $x_j(j=1,2,\cdots,n)$ is a benefit-type indicator variable. The comparison sequence is

$$a_i=\{a_i(j)\,|j=1,2,\cdots,n\},i=1,2,\cdots,m \tag{26}$$

Where: $a_i(j)$ is the value of the ith evaluation object regarding the jth index variable x_j.

The reference number is listed as $a_0=\{a_0(j)|j=1,2,\cdots,n\}$, which is usually $a_0(j)=\max\limits_{1\leqslant i\leqslant m}|a_i(j)|$, that is, the reference series is equivalent to each index value of a virtual best evaluation object.

1) Determine the corresponding weight of each indicator

Analytic hierarchy process can be used to determine the corresponding weight $\omega=[\omega_1,\omega_2,\cdots,\omega_n]$ of each index, where $\omega_j,j=1,2,\cdots,n$ is the corresponding weight of the jth evaluation index.

2) Calculate the grey correlation coefficient

$$\xi_i(j)=\frac{\min\limits_{1\leqslant s\leqslant m}\min\limits_{1\leqslant t\leqslant n}|a_0(t)-a_s(t)|+\rho\max\limits_{1\leqslant s\leqslant m}\max\limits_{1\leqslant t\leqslant n}|a_0(t)-a_s(t)|}{|a_0(j)-a_i(j)|+\rho\max\limits_{1\leqslant s\leqslant m}\max\limits_{1\leqslant t\leqslant n}|a_0(t)-a_s(t)|} \tag{27}$$

$i=1,2,\cdots,m,j=1,2,\cdots,n$

$\xi_i(j)$ is the correlation coefficient between the comparison series a_i and the reference series a_0 on the jth index, where $\rho\in[0,1]$ is the resolution coefficient. Among them, $\min\limits_{1\leqslant s\leqslant m}\min\limits_{1\leqslant t\leqslant n}|a_0(t)-a_s(t)|$ and $\max\limits_{1\leqslant s\leqslant m}\max\limits_{1\leqslant t\leqslant n}|a_0(t)-a_s(t)|$ are respectively called two-stage minimum difference and two-stage maximum difference.

Generally speaking, the larger the resolution coefficient ρ, the greater the resolution; The smaller the ρ, the smaller the resolution [6].

3) Calculate grey weighted correlation degree

The calculation formula of grey weighted correlation degree is:

$$r_i=\sum_{k=1}^{n}\omega_k\xi_i(k) \tag{28}$$

Where: r_i is the gray weighted correlation degree between the ith evaluation object and the ideal object. If the weight of each indicator variable is not given, the weight of each indicator variable is equal.

4) Evaluation analysis

According to the grey weighted correlation degree, the ranking of evaluation objects can be established to establish the correlation order of evaluation objects. The greater the correlation degree, the better the evaluation result [7].
The results of grey correlation analysis on the two types of cultural relics are as follows:

Table. 15 Grey correlation table

	High Potassium	Lead and barium
SiO_2	0.9889	0.8416
Na_2O	0.3974	0.4381
K_2O	0.4193	0.4343
CaO	0.4104	0.4439
MgO	0.3982	0.4365
Al_2O_3	0.4139	0.4525
Fe_2O_3	0.4002	0.4369
CuO	0.4022	0.4447
PbO	0.4253	0.7665
BaO	0.3963	0.5122
P_2O_5	0.3994	0.4513
SrO	0.3955	0.4352
SnO_2	0.3958	0.4337
SO_2	0.3957	0.4377

According to the results of the above table, it is easy to know that the correlation degree of potassium-containing chemicals in high-potassium glass is not high, indicating that in the firing process of high-potassium glass, substances with high potassium content as a co-solvent do not have an essential effect on the formation of high-potassium glass; However, lead barium glass is different. As lead ore (high content of lead oxide and barium oxide) is added as co-solvent in the firing process, lead barium glass produced by firing contains a large amount of lead oxide and barium oxide.

7. Model evaluation

7.1. Model Benefits

(1) Chi-square test can objectively determine whether a factor has a significant impact on the results from the perspective of data.

(2) The combination of Q-type cluster analysis and R-type cluster analysis can not only reduce the data dimension, but also obtain a reasonable result.

(3) Support vector machine can map low-dimensional data into high-dimensional space and classify it into binary categories, which is robust and sparse.

(4) Grey relational degree analysis can find the numerical relation among subsystems in the system through certain methods. Due to the small amount of calculation, it is equally applicable to the quantity and regularity of samples, and there will be no inconsistency between quantitative results and qualitative analysis results.

7.2. Inadequate models

(1) Q type cluster analysis needs to manually select a certain number of categories for clustering.
(2) Support vector machine is only suitable for binary classification and cannot deal with multivariate classification.
(3) With the continuous expansion of the application field of grey correlation analysis theory, the shortcomings of some existing models make it unable to solve some practical problems well. At the same time, the whole theoretical system is not perfect at present, and its application is subject to some limitations.

References

[1] FANG Xiangzhong. Chi-square distribution and Chi-square Test [J]. Chinese Statistics,2022(05):29-31.

[2] Wang Jingxiang. The principle of decision tree algorithm research and actual application [J]. Computer programming skills and maintenance, 2022 (8) : 54-56 +72.doi:10.16184/j.cnki.com.prg. 2022.08.043.

[3] Wang Jie, Li Mo, Ma Qinglin, Zhang Zhiguo, Zhang Meifang, Wang Julin. A warring states period of eight prism lead barium glass weathering research [J]. Journal of glass and enamel, and 2014 (02) : 6-13. DOI: 10.13588 / j.carol carroll nki g.e. 1000-2871.2014.02.002.

[4] Sun Jigui, Liu Jie, Zhao Lianyu. Research on clustering Algorithm [J]. Journal of Software,2008(01):48-61. (in Chinese)

[5] Ding Shifei, Qi Bingjuan, Tan Hongyan. Review on Theory and Algorithm of Support Vector Machine [J]. Journal of University of Electronic Science and Technology of China,2011,40(01):2-10.

[6] QI Hengnian. Review of Support Vector Machines and Their Applications [J]. Computer Engineering,2004(10):6-9.

[7] SUN F F. Grey Relational analysis method and its application [J]. Science and Technology Information,2010(17):880-882. (In Chinese)