# Study on Composition Analysis and Identification of Ancient Glass Products

Nian Yang*

Xi'an Jiaotong University, Xi'an, China

*Corresponding author: wsynzs@stu.xjtu.edu.cn

**Abstract:** Glass is the precious material evidence of the early trade of the ancient Silk Road, but the ancient glass is easily weathered by the influence of burial, and its composition ratio changes, which affects the correct judgment of its category. The study of the composition analysis and identification of ancient glass products is of great help to understand the social culture and foreign trade civilization at that time. This paper mainly studies the composition analysis and identification of ancient glassware, to evaluate, predict and classify the ancient glassware, this paper establishes a comprehensive evaluation model, using the chi-square test, K-means clustering model, decision tree model, Lasso regression and grey correlation degree test. It helps archaeologists to analyze and predict the correlation between the weathering degree of cultural relics and their attributes and chemical composition content, and according to the existing classification standards of cultural relics. A labelled subclassification scheme is formulated to identify the types of unknown cultural relics. At the same time, the correlation between the chemical components of different types of cultural relics was analyzed.

**Keywords:** Ancient glass relics, correlation analysis, clustering algorithm, decision tree.

## 1. Introduction

Glass is one of the earliest silicate materials invented by human beings, which was introduced into China as jewellery from the Western Regions in ancient times and became a precious cultural relic through the precipitation of time. However, due to the physical and chemical properties of glass, it is easy to react with moisture and gases in the air, which makes it difficult to confirm the types of cultural relics and other information, thus affecting more detailed archaeological work. Therefore, it is urgent to analyze the weathering mechanism of glass cultural relics and find an effective classification method.

Cao Caixia and other scholars used the theory of fuzzy mathematics to study the classification of glass cultural relics. [1]Yang Yun and other scholars used the method of BiLSTM-CRF to study the extraction of knowledge points of glass cultural relics, which provided a solid mathematical foundation for subsequent researchers.[2]Chen Min used atomic absorption spectrophotometer, infrared reflectance spectroscopy and other methods to study the influencing factors of lead-barium glass weathering with detailed experimental data, and constructed conditions for further analysis of the mechanism. [3]Wang Chengyu and other scholars have made a detailed explanation of the weathering mechanism of silicate glass and provided a research direction for improving the weathering of glass.[4]In a word, the weathering composition analysis and classification identification of glass cultural relics has gradually become a hot research topic, which has broad research space and great value of dual significance of technology and culture, and more scholars will participate in the research of this issue in the future.

However, the above research methods are only aimed at analyzing the relationship between the type, decoration and colour of glass cultural relics and whether the surface is weathered or not, and then exploring the existing classification rules of glass cultural relics through data, and dividing them into subcategories with explicable and clear rules. Finally, the chemical composition of the unknown type of glass relics was analyzed to identify its type, and the correlation analysis of chemical elements was carried out by grey correlation analysis.

# 2. Establishment and solution of the model

## 2.1 Correlation analysis

To explore the relationship between the type, decoration and colour of glass relics and whether the surface is weathered or not, we use the method of calculating the correlation coefficient. Since the glass relic "weathered or not" and the other three variables are all categorical variables, the Pearson coefficient used for continuous variables is no longer applicable, and we consider using the Spearman rank correlation coefficient (Spearman's rank correlation coefficient). We select the four factors of ornamentation, type, colour and surface weathering, and get relatively reliable results through quantitative analysis.

We intend to explore the correlation coefficient between the surface weathering of glass relics and the glass type, ornamentation and colour . The results are shown in Table 1.

Table 1 Spearman Coefficient

| | Ornamentation | Type | Color | Surface weathering |
|---|---|---|---|---|
| Ornamentation | 1.000 (0.000*** ) | 0.190 (0.176 ) | 0.470 (0.000*** ) | -0.095 (0.504 ) |
| Type | 0.190 (0.176 ) | 1.000(0.000***) | 0.470(0.000***) | -0.272(0.051*) |
| Color | 0.470(0.000***) | 0.470 (0.000*** ) | 1.000(0.000) | 0. 142(0.314) |
| Surface weathering | -0.095 (0.504 ) | -0.272(0.051*) | 0. 142 (0.314 ) | 1.000 (0.000*** ) |

Based on the above, there is no correlation between the ornamentation and the weathering of the surface. There is no direct connection between the decoration of glass relics and its weathering, but there may be some indirect connection. There is a certain correlation between glass type and surface weathering, and the correlation coefficient is large, 90% of the confidence is that the glass type is directly related to the weathering of the cultural relics themselves. There is no correlation between color and surface weathering. There is no direct connection between the color of glass relics and its weathering, but there may be some indirect connection.

## 2.2 Prediction of chemical composition content before weathering

The erosion of glass by atmospheric action is called weathering of glass, which is essentially a long chemical reaction process. This process and its products are closely related to the properties of glass itself and environmental conditions. Environmental factors mainly refer to the temperature, humidity and time of glass storage[5], while the nature of glass itself is mainly reflected in the category of glass cultural relics (affecting the chemical composition of atmospheric reaction), decoration (affecting the contact area with air), colour (physical characterization of chemical composition). Here, we assume that each relic sample undergoes the same weathering process and external environment, so that the weathering process is determined only by the nature of the glass relic itself. We can select the weathered and unweathered cultural relics which are similar to the nature of the sample to be tested, calculate the rate of change of each chemical component in the weathering process, and approximately think that the weathering process of the sample to be predicted can also be described by this rate of change[6].

In particular, we take the artefact sample numbered 54 as an example to illustrate our prediction method.

The decoration of the cultural relic numbered 54 is category C, which is a light blue lead-barium glass relic. Therefore, first of all, we find 9 groups of data of weathered lead-barium glass relics 2, 11 and 43, which are both light blue, and 11 groups of data of unweathered lead-barium glass relics 25, 28 and 29, which are both light blue (or taken from the unweathered points of weathered samples). Assuming that the mass of weathered and unweathered samples is 1g, then the mass change rate of each chemical component is calculated by the formula (1).

$$\xi = (C_{after} - C_{before}) \frac{1}{C_{before}} \quad (1)$$

Where $\xi$ denotes the rate of change of the mass of the chemical composition.

Then, we set the mass of the chemical composition of the sample to be tested as Cx, and deform the formula (1.1) to obtain the formula (1.2).

$$Cx = \frac{Cy}{1+\xi} \quad (2)$$

Where $Cy$ is the mass of the chemical composition of the sample to be predicted. The mass of each component of the sample to be tested before weathering can be obtained by calculating according to the change rate obtained in Formula (2) and Table 3, and then the content ratio of each component can be obtained by normalizing. The results of the calculation are shown in Figure 1.

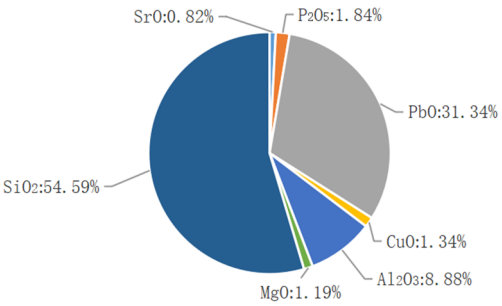



Fig. 1 Prediction of chemical composition content of No.54 cultural relic before weathering

Similarly, we can also get the predicted values of the contents of various chemical components of other weathered cultural relics.

## 3. Study on classification of high-potassium and lead-barium glass based on decision tree

Decision trees are widely used models for classification and linear regression tasks. Compared with other classification algorithms, the decision tree has two advantages: one is that the resulting model is easy to visualize, and the other is that the algorithm[7] is completely independent of data scaling.

From the decision tree, we can know that the PbO content alone can completely distinguish the two: the glass with a PbO content of less than 5.46% belongs to high potassium glass, otherwise, it belongs to lead-barium glass.

### 3.1 Use the K-means algorithm to determine the number of subclasses

After using the decision tree model to divide the glass relics into high-potassium and lead-barium, the title requires us to select the appropriate chemical composition to further divide them according to the physical and chemical properties of the samples and plans to divide them into more subcategories. Since the classification criteria and the number of classes are unknown, we can first use the K-means algorithm to determine the number of subclasses, and then further analyze the classification criteria.

To find the most appropriate number of subclasses, we use the "elbow rule", that is, with the number of clusters K as the abscissa and the sum of squared errors SSE (or distortion distance) as the ordinate, SSE will gradually decrease with the increase of the number of clusters. The number of clusters K corresponding to the inflexion point where the decline rate suddenly slows down is the most appropriate choice.

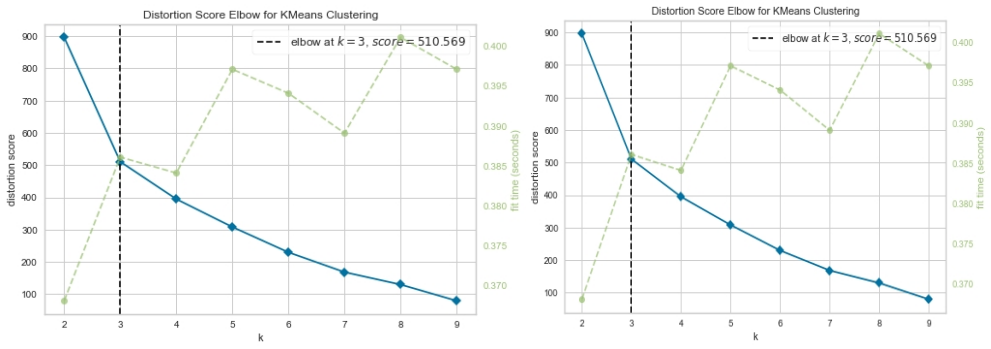



Fig. 2 Elbow diagram for finding the optimum K

We write Python code , select the number of clusters in the range of [2,9], and find the most suitable number of clusters for high-potassium glass and lead-barium glass respectively.

According to Figure 2, we can find that the decline rate of the distance score of high-potassium glass suddenly slows down when K = 3, while the corresponding inflexion point of lead-barium glass comes later, and the decline rate slows down significantly when K = 5. According to the elbow rule, we can confirm that the most suitable number of subclasses of high-potassium glass is 3, while the most suitable number of subclasses of lead-barium glass is 5.

### 3.2 Use decision tree to explore the criteria and basis for subclassification

K-means clustering algorithm can control the granularity of clustering and is very good at finding the most suitable number of subclasses. However, because the classification of the algorithm is more abstract and can not be interpreted well, it can not be directly used as our target division.

The decision tree can give us a scientific and clear way of classification after the number of classifications is given, so we decided to establish a decision tree model based on K-means clustering analysis to explore the different sub-classification criteria of the two kinds of glass,which is shown in Figure 3.

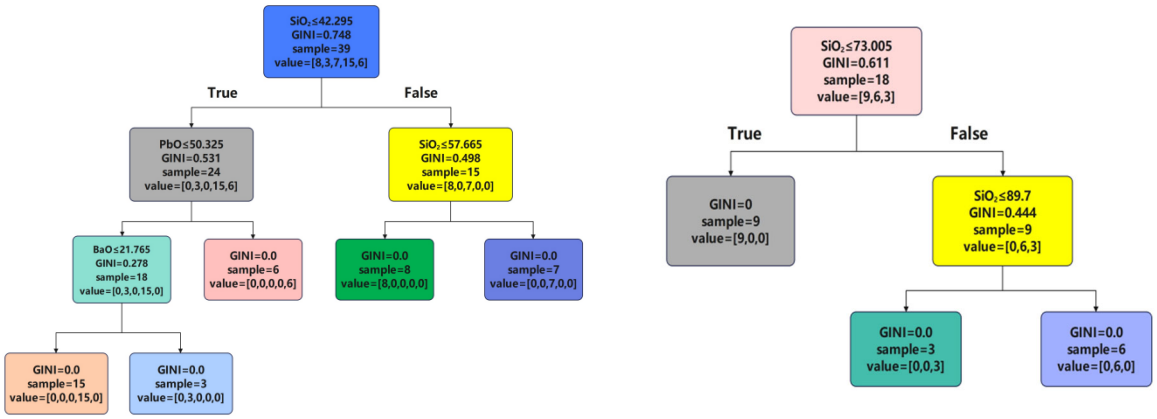



Fig. 3 Decision tree for subclasses of high-K lead and barium glasses

From the decision tree, we can find that the content of $SiO_2$ plays a vital role in the classification of high-potassium glass and lead-barium glass, and it is the only criterion in the classification of high-potassium glass. Finally, the classification model of glass cultural relics can be obtained, as shown in formula (3).

$$\text{Glass artifacts}\begin{cases}\text{High potassium}(PbO \geq 5.46)\begin{cases}A(SiO_2 \leq 73.005)\\ B(73.005 \leq SiO_2 \leq 89.7)\\ C(SiO_2 \geq 89.7)\end{cases}\\ \text{Lead and barium}(PbO \geq 5.46)\begin{cases}A(SiO_2 \leq 42.95, PbO \leq 50.325, BaO \leq 21.765)\\ B(SiO_2 \leq 42.95, PbO \leq 50.325, BaO \geq 21.765)\\ C(SiO_2 \leq 42.95, PbO \geq 50.325)\\ D(42.95 \leq SiO_2 \leq 57.665)\\ E(SiO_2 \geq 57.665)\end{cases}\end{cases} \quad (3)$$

## 4. Establishment and solution of prediction model

After cross-validation, we select the appropriate regularization parameter α and use the data in the datasheet as the training set to write the Python source program to construct the Lasso regression model. Run the program to obtain the regression expression as shown in Formula (3). Where L represents the Lasso regression score because we convert the two qualitative variables of high-potassium glass and lead-barium glass into quantitative variables. It is specified that 1 is high-potassium glass and 2 is lead-barium glass, so the value range of the Lasso regression score is [1,2]. We believe that a score below 1.5 is high potassium glass, and vice versa is lead-barium glass.which is shown in formula (4).

$$L = [-0.004\ 0.025\ -0.042\ 0.007\ 0.004]\begin{bmatrix}x1\\ x2\\ x3\\ x4\\ x5\end{bmatrix} + 1.755 \quad (4)$$

### 4.1 Prediction of unknown cultural relic categories

We take A1 in Table 3 as an example to illustrate the specific steps of prediction, and other samples are also carried out in the same way.First of all, we substitute the chemical indicators of A1 cultural relic into formula (4), and calculate the Lasso score of A1 cultural relic L = 1.441. Considering the

Table 2 Classification prediction of glass cultural relics of unknown category

| Cultural relic number | Surface weathering | Lasso regression score | Lasso regression prediction type | Classification rule and prediction type |
|---|---|---|---|---|
| A1 | Unweathered | 1.441 | High potassium | High potassium |
| A2 | weathering | 1.844 | Lead and barium | Lead and barium |
| A3 | Unweathered | 1.866 | Lead and barium | Lead and barium |
| A4 | Unweathered | 1.783 | Lead and barium | Lead and barium |
| A5 | weathering | 1.607 | Lead and barium | Lead and barium |
| A6 | weathering | 1.326 | High potassium | High potassium |
| A7 | weathering | 1.351 | High potassium | High potassium |
| A8 | Unweathered | 1.735 | Lead and barium | Lead and barium |

above division basis, A1 cultural relics should belong to high-potassium glass cultural relics.In order to verify the rationality of Lasso regression, we can use the classification rules found in the second question to verify. The scores and prediction results of each cultural relic sample in Table 2.

It can be seen from Table 2 that the classification results obtained by using the Lasso regression model are completely consistent with the classification results under the guidance of the classification rules found in the previous text, which highlights the rationality of Lasso regression and achieves the effect of mutual confirmation.

### 4.2 Correlation analysis of chemical composition

To explore the correlation between the chemical composition of different types of glass cultural relics, we adopt the idea of the first classification and then correlation.

From the analysis of the prediction model, it can be seen that the type and weathering of glass relics will have a significant impact on their chemical composition. Therefore, we first divide the data of glass cultural relics into four categories according to the above two elements, namely, high potassium weathering, high potassium non-weathering, lead-barium weathering and lead-barium non-weathering. Because the Pearson coefficient and Spearman coefficient do not apply to the sample, we consider using the method of grey correlation analysis to analyze the correlation among the cultural relics of the same category. The procedure is as follows[8].

Step 1: Determine the comparison sequence and the reference sequence.

Step 2: Determine the corresponding weight of each indicator.

Step 3: Calculate the grey correlation coefficient. The calculation formula is shown in formula (5).

$$\xi_i(j)=\frac{\min_{1\le s\le m}\min_{1\le t\le n}\left|a_0(t)-a_s(t)\right|+\rho\max_{1\le s\le m}\max_{1\le t\le n}\left|a_0(t)-a_s(t)\right|}{\left|a_0(j)-a_i(j)\right|+\rho\max_{1\le s\le m}\max_{1\le t\le n}\left|a_0(t)-a_s(t)\right|} \tag{5}$$

Here $\rho = 0.5$ is chosen.

Step 4: Calculate the grey weighted correlation degree. The calculation formula of grey weighted correlation degree is shown in formula(6).

$$r_i=\sum_{k=1}^{n} w_k\xi_i(k) \tag{6}$$

## 5. Analysis of results

### 5.1 Analysis of classification results of high-potassium and lead-barium glasses based on decision tree

It can be seen from the above that for the given data sets of several relevant indicators affecting the classification of glass cultural relics, 3 characteristics are selected in the subclassification of high-potassium glass, and 5 characteristics are selected in the subclassification of lead-barium glass. To analyze the sensitivity and rationality of the classification algorithm, we carry out the following operations:

Firstly, the training set of samples is classified by the CART decision tree algorithm, and the optimal decision tree is generated. According to the characteristics of the subset divided by nodes, we can know that in high-potassium glass, the chemical composition content of $SiO_2$ is the most sensitive to its subset division; The chemical compositions of $SiO_2$, PbO and BaO in Pb-Ba glass are more sensitive to the classification of their subclasses, while the rest of the chemical compositions not mentioned in the decision tree are less sensitive to the classification of their subclasses.

When dealing with the data, we select 80% of the data, use the CART algorithm to divide the sub-categories from two aspects of high-potassium glass and lead-barium glass and construct decision trees to describe the characteristics of different categories of customer groups. For the new decision tree, using the remaining 20% of the data to test, it is found that the correct rate of classification rules of the decision tree obtained by the algorithm is 90.7% and 93. It shows that the decision tree

constructed by the algorithm is reliable and effective, so more detailed classification types can be induced from the obtained classification rules. At the same time, we also fully refer to the previous literature[9] to analyze the division rules of glass cultural relics. By comparison, it is found that our classification rules are similar to them, so it can be concluded that our decision tree model is reasonable and accurate.

### 5.2 Use the K-means algorithm to determine the number of subclasses for result analysis

Variance-based sensitivity analysis, also known as Sobol analysis, is a global sensitivity analysis method. In the Sobol sensitivity analysis, we first determined the range of input parameters according to the results of Lasso regression, selected five relatively important independent variables in the regression model as input parameters for sensitivity analysis[10], and then carried out model construction simulation tests for 1536 times. The sensitivity of each input parameter to the target is obtained. The results are shown in Figure 4.

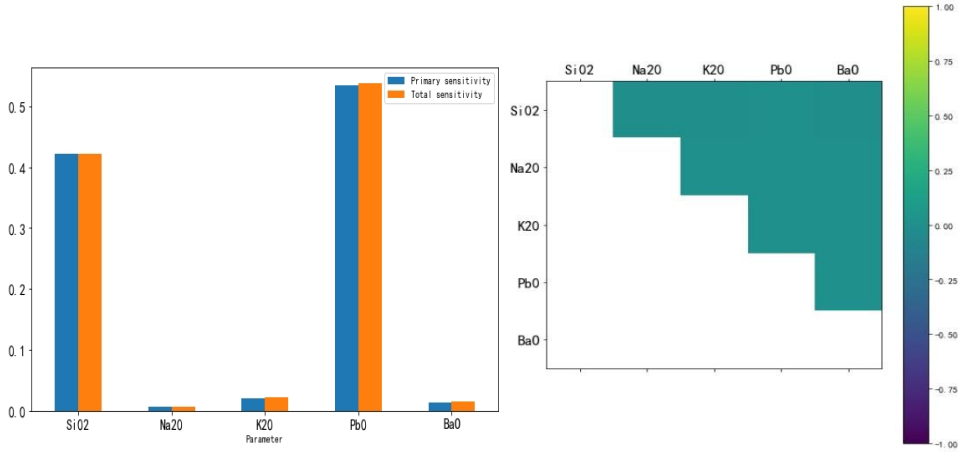



Figure 4 Class Prediction and Second-order interactive heat map

## 6. Conclusion

The classification and prediction model of glass cultural relics established in this paper has the characteristics of clear division and simple practice. For the newly unearthed glass relics, we can use physical means to accurately determine the chemical composition of the relics, and then we can get their classification according to the model established in this paper, which lays the foundation for the study of the age and culture of glass relics and has important value and significance for the development of archaeological and cultural undertakings in China.

In this paper, the elbow rule is used to determine the number of subclasses of high-potassium and lead-barium glass, which avoids the blindness and subjectivity caused by artificial selection of the number of categories, and the decision tree is used to further divide, which ensures the interpretability of the division results. In the final prediction of the unknown sample category, we use the classification law and Lasso regression to compare the prediction results to ensure that the Lasso regression results are reasonable and the classification law is reliable. When predicting the chemical composition content of pre-weathering samples, the rigour of the method of converting content to quality is still open to question, and the accuracy of decision tree division may be affected due to the small number of samples. Based on this paper, we can use the appropriate algorithm to carry out the difference value of the vacancy value, reduce the amount of data elimination, and further optimize the existing model to get better results.

## References

[1] Cao Caixia, Guo Hong. Classification of disease types of stone, ceramics and glass cultural relics in the library by fuzzy mathematics [J]. Journal of Beijing Union University (Natural Science Edition), 2009, 23 (4): 58-60.

[2] [Yang Yun, Song Qingyi, Yun Xinyu, et al. Knowledge extraction of glass relics based on BiLSTM-CRF [J]. Journal of Shaanxi University of Science and Technology, 2022, 40 (3): 179-184.

[3] Chen Min. Factors affecting the weathering of lead glass [D]. Dalian University of Technology, 1987.

[4] Wang Chengyu, Tao Ying. Weathering of silicate glass. Journal of Silicate, 2003, 31 (1): 78-85.

[5] Wang Yinghong. Laboratory simulation study on the influence of atmospheric pollutants (SO2, NOx) on the weathering of Yungang Grottoes [D]. Chinese section University of Technology, 2016.

[6] Wang Xiaoping. Research on Decision Tree Optimization Algorithm Based on Rough Set [D]. Sichuan Normal University, 2013.

[7] Wikipedia contributors. Variance based sensitivity analysis. Wikipedia[EB/OL]. (2022-09-04).

[8] Si Shoukui, Sun Xijing. Mathematical Modeling Algorithm and Application [M]. Version 3. National Defense Industry Press, 2021: 424-425.

[9] Wang Mei, Yu Jing, Liu Guang. A decision tree algorithm based on attribute sensitivity. The computer system shall Used, 2010, 19 (11): 52.

[10] Zhang Bin. Study on the Application of PIXE in Ancient Ceramics and Glass Producing Areas [D]. Fudan University, 2004.