

Analysis of the composition of ancient glass products based on Spearman's correlation coefficient

Yibo Hou *

College of Information and Electrical Engineering, China Agricultural University, Beijing, China

* Corresponding author: tp10101001@163.com

Abstract. Ancient glass stored underground for a long time has undergone different degrees of weathering, which has led to changes in its chemical composition and affected the correct judgment of its category. Based on this, this paper uses various algorithms such as Spearman correlation coefficient, chi-square test and cluster analysis to establish a comprehensive model for glass composition evaluation. The results show that: (1) The surface weathering degree of glass artifacts is significantly correlated with their types; (2) The content of Lead oxide can be used to make a general distinction between high-potassium glass and lead-barium glass; (3) Among the known glass samples, high-potassium glass can be divided into three categories and lead-barium glass can be divided into four categories; (4) According to the reasonableness test, it is known that the model has good reliability and can provide guidance for the study of ancient glass.

Keywords: Glasses, Subclass division, Spearman correlation coefficient, Decision Trees, K-means++ clustering analysis.

1. Introduction

The Silk Road was an ancient conduit connecting China and the West for trade and cultural interaction, and as a good example of trade, glass has witnessed part of the history of Silk Road trade exchanges. The process of making glass first originated in the West. It was made somewhere in Mesopotamia or Egypt about 3,500 years ago and later brought to China via the Silk Road. Then, we studied the manufacturing process and used local resources and materials to make similar glass products, similar in appearance to the imported foreign products, but different in composition [1].

In recent years, many scholars in China have conducted studies on ancient glass. Gan [2] determined the chemical composition of glass by proton-excited X-ray fluorescence technique, energy-dispersive X-ray fluorescence technique and inductively coupled plasma emission spectroscopy; Cui [3] performed a multivariate statistical analysis of the chemical composition of glass using laser exfoliation inductively coupled plasma emission spectroscopy.

However, the above research methods were only analyzed for individual types of glass. In this paper we focus on two glass types, namely, high potassium glass and lead-barium glass. Firstly, the Spearman coefficient was used to determine the relationship between the degree of surface weathering and type, decoration and color at the correlation level. The strength of the correlation was judged based on the absolute value of the correlation coefficient obtained from the formula. Next, a chi-square test was conducted to analyze the variability of each factor with the degree of weathering and to find a factor that significantly affects the degree of weathering to further verify the correctness of the previous conclusions.

After that, we analyzed the classification patterns of the two types of glass and built a decision tree model. The known data [4] were classified by glass type, and the classification law was obtained based on the model diagram by comparing the differences in the chemical composition of the two types of glass and selecting the element with the greatest difference. According to the idea of the above problem, when subclassing for glass artifacts, the K-means++ algorithm of the clustering model is used to select the appropriate K value as the number of clusters with the help of the elbow rule. The rationality of subclassing is subsequently verified by comparing the obviousness of the difference between the factors used for classification of a certain glass among that type of glass. The research method is of guidance for the classification of glass types.

2. Model building

2.1. Assumptions

- (1) For glass artifacts with no detectable color, to distinguish them from other known colors, we tentatively consider them to be white, which has less impact on the correctness of the results
- (2) Since the weathering process is very slow, we assume that the measured chemical composition content is accurate and does not change within a short period of time.
- (3) The sample information of the glass artifacts is assumed to have a high degree of correctness.

2.2. Correlation analysis among the attributes of glass products

X and Y are two sets of independent and identically distributed data, the number of elements of which are N. The $i(1 \leq i \leq N)$ values obtained by the two sets of random variables are represented by X_i, Y_i , respectively. The two sets of X and Y are simultaneously arranged in ascending or descending order to obtain two sets of elements arranged x and y, where the elements x_i and y_i are sorted X_i in X and Y_i in Y, respectively. The corresponding elements in the sets x and y form the ranking difference set d and the Spearman correlation coefficient between X and Y can be calculated from x, y, and d [5]. r_s is calculated as follows.

$$r_s = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (1)$$

However, the link between variables is not necessary normally, so it can be calculated

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2)$$

In the above equation, d_i is the equivalence difference between X_i and Y_i ; the Spearman correlation coefficient takes values in the range of [-1,1], and the larger the absolute value of r_s , the stronger the correlation. Therefore, when $r_s > 0$, it can be considered that there is a positive correlation between the two groups of variables involved in the discussion; when $r_s < 0$, it can be considered that there is a negative correlation between the two groups of variables involved in the discussion [6].

The Spearman correlation coefficients between the attributes about known data are shown in Table 1.

Table 1. Correlation coefficient table

	Ornamentation	Colors	Typies	Surface weathering
Ornamentation	1.000 (0.000***)	-0.433 (0.001***)	-0.357 (0.006***)	0.116 (0.384)
Colors	-0.433 (0.001***)	1.000 (0.000***)	0.557 (0.000***)	-0.019 (0.885)
Typies	-0.357 (0.006***)	0.557 (0.000***)	1.000 (0.000***)	0.344 (0.008***)
Surface weathering	0.116 (0.384)	-0.019 (0.885)	0.344 (0.008***)	1.000 (0.000***)

In Table 1, ***, **, *represent 1%, 5%, and 10% significance levels, respectively. It can be seen that the degree of surface weathering of the glass artifacts is significantly correlated with the type of artifacts, while there is no significant correlation with color and ornamentation. To ensure the accuracy of the results, we conducted a variance analysis of each factor and the degree of weathering using the chi-square test. The chi-square test is a common nonparametric test that can be used to explore the correlation between two or more variables of a definite type, the essence of which is to compare the degree of agreement between theoretical and actual frequencies. The test results are shown in Table 2.

Table 2. Results of chi-square test analysis of variance

Attributes	Name	Degree of surface weathering		Total	X ²	P
		No	Yes			
Ornamentation	A	11	11	22	4.957	0.084*
	B	0	6	6		
	C	13	17	30		
Typies	Lead barium	12	28	40	6.88	0.009***
	High Potassium	12	6	18		
Colors	light green	2	1	3	9.432	0.307
	Light blue	8	12	20		
	Dark Green	3	4	7		
	Deep Blue	2	0	2		
	White	0	4	4		
	Purple	2	2	4		
	Green	1	0	1		
	Blue-Green	6	9	15		
Black	0	2	2			

It can be seen that for the degree of surface weathering, different glass types show significance at the level, indicating that there is a significant difference between the degree of surface weathering and types, which is consistent with the results of correlation analysis.

In summary, it can be seen that the degree of weathering of glass artifacts is mainly related to its type, with lead-barium glass being more susceptible to weathering while high-potassium glass is relatively less susceptible to weathering.

2.3. Analysis of the classification rules of high potassium and lead-barium glass

The decision tree algorithm [7] is a currently used data mining classification algorithm. It first learns the sample set to build a decision tree classification model, and then classifies unknown types of samples based on the generated model for the purpose of data integration. A huge advantage of the decision tree algorithm over other neural network-based machine learning algorithms is its interpretability. Decision tree diagrams can be visualized through a tree structure to show us a clear decision process.

Since there are differences in the chemical content in different glasses, we differentiate the chemical components according to those with significant differences and build a decision tree model using SPSSPRO. Considering the size of the generated decision tree, Figure 1 shows the decision tree diagram for classifying the glass artifacts.

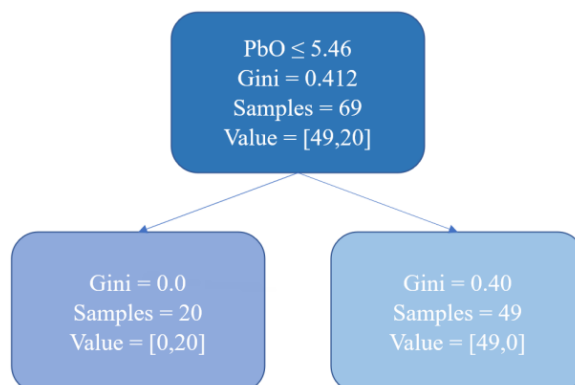


Figure 1. Decision tree model for glass artifact classification

We use the difference in lead oxide content to make the distinction. Those with PbO (lead oxide) content below 5.46 are high potassium glasses and those above that value we consider as lead-barium glasses. From the known data we can also conclude that the different types of PbO content have

relatively significant differences, while other chemical components such as barium oxide, silica, etc. do not have significant differences, so it is reasonable to use PbO content for differentiation to some extent.

2.4. Subclassification of glass objects

Clustering is the process of dividing a sample into multiple classes consisting of similar objects, and is a classical task in the field of data mining.

In order to subclass two classes of glass, we used the K-means++ algorithm to first select the appropriate clustering center and then calculate the distance of each element to the clustering center, which is iterated several times until the position converges. In the process, we used SPSS for implementation and obtained the specific classification results.

For the selection of K values, we use the elbow rule and take the K value when the decline in the aggregation coefficient folding graph tends to be significantly slow as the number of clusters.

The folding graph of aggregation coefficients is shown below.

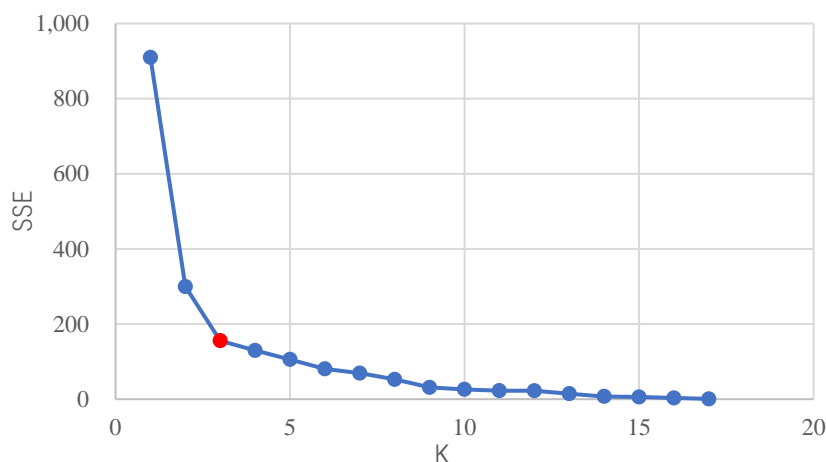


Figure 2. Folding graph of polymerization coefficient of high potassium glass

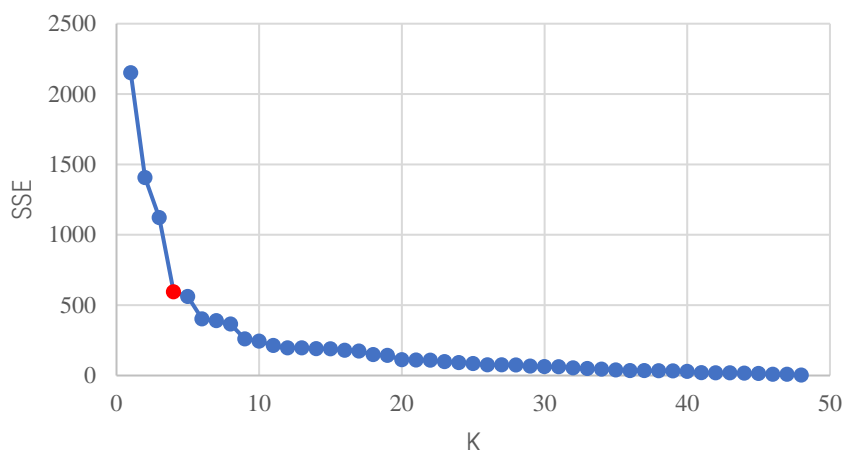


Figure 3. Folding graph of polymerization coefficient of lead barium glass

From Figure 2, it can be seen that the polymerization coefficient of high potassium glass turns significantly at K=3, and the change tends to be flat. Similarly, it can be seen from Figure 3 that the polymerization coefficient of lead-barium glass shows an inflection point at K=4, so we classify high potassium glass into 3 categories and lead-barium glass into 4 categories when we make subcategories.

We then performed a differential analysis of the chemical composition of the classification results and extracted the main chemical components.

For high potassium glasses, it is clear from the subclass classification results that silica is the main differentiating element, followed by potassium oxide, the specific content is shown in Figure 4 and Figure 5.

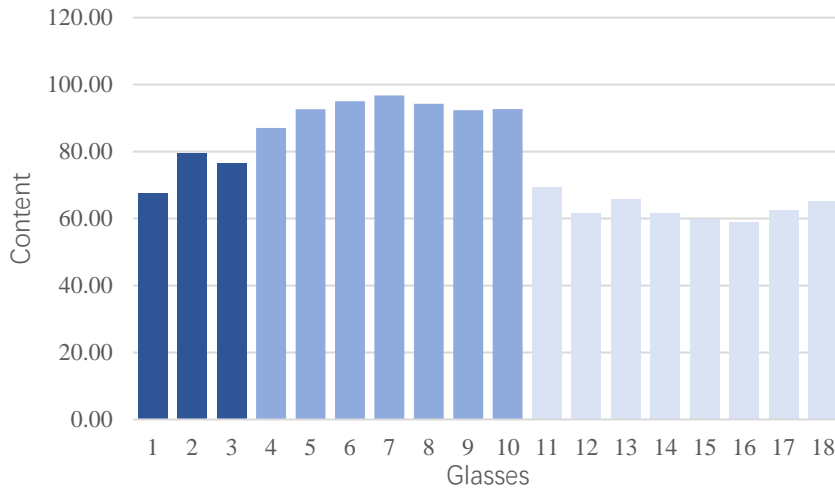


Figure 4. Content of silica in high potassium glass

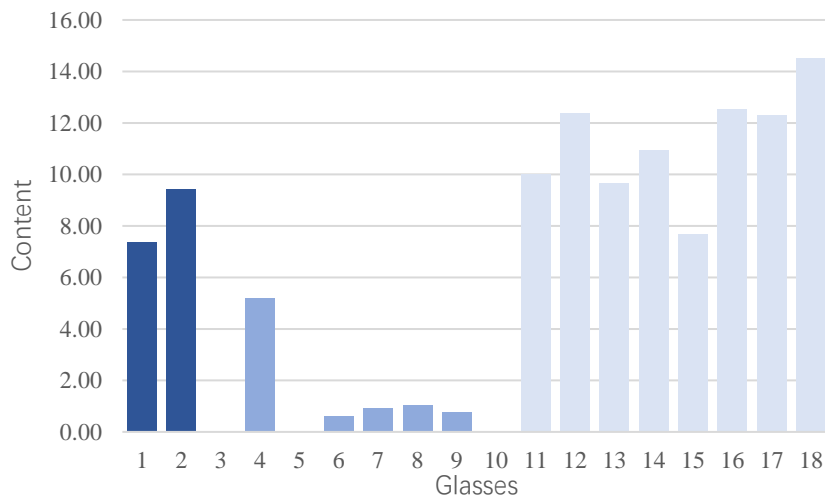


Figure 5. Potassium oxide content in high potassium glass

It can be seen that the three color categories represent the three categories divided under high potassium glass, which have more significant differences in silica content and potassium oxide content, but not much difference in the content of other elements, so the classification has some reasonableness and sensitivity.

For lead-barium glass, the differences in chemical composition are mainly reflected in the contents of silica, barium oxide and lead oxide. Since there are more lead-barium glasses in the sample, we use a line graph to display them.

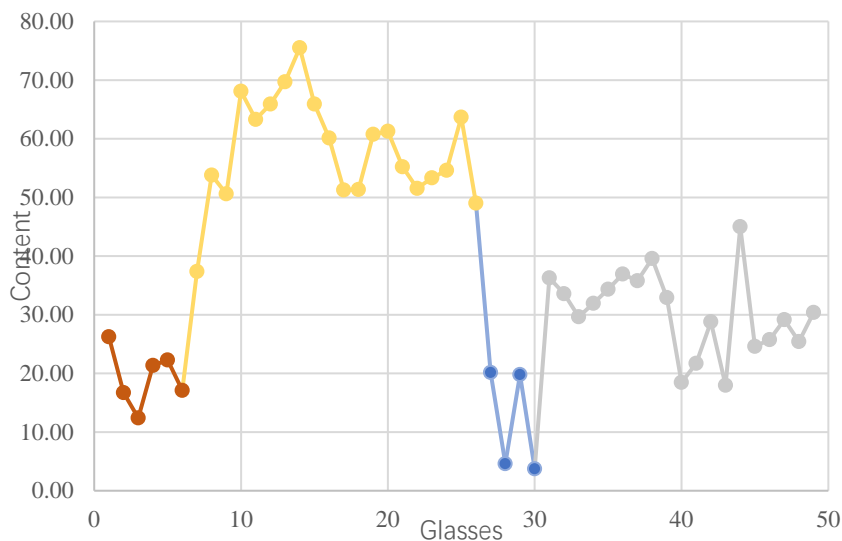


Figure 6. The content of silica in lead barium glass

The four different categories of lead-barium glass are indicated by orange, yellow, blue and gray colors in Figure 6, respectively, and it can be seen that the silica content of the four categories is significantly different. In terms of silica content, the category shown in yellow is higher than the other categories, while the category shown in blue is significantly lower than the other categories.

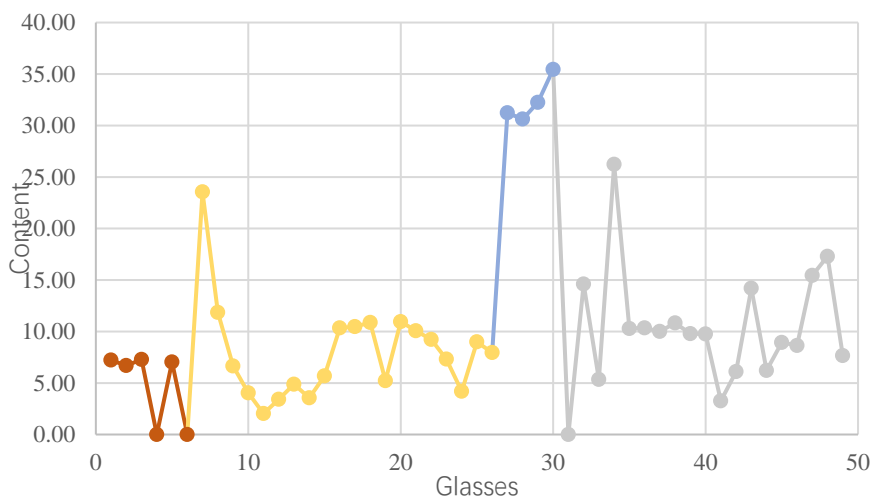


Figure 7. The content of barium oxide in lead barium glass

As can be seen in Figure 7, in terms of barium oxide content, the category shown in blue has the highest content, while the category shown in red has the lowest content, and the categories shown in yellow and gray have comparable content levels, but are classified into different categories due to differences in the content of other chemical components.

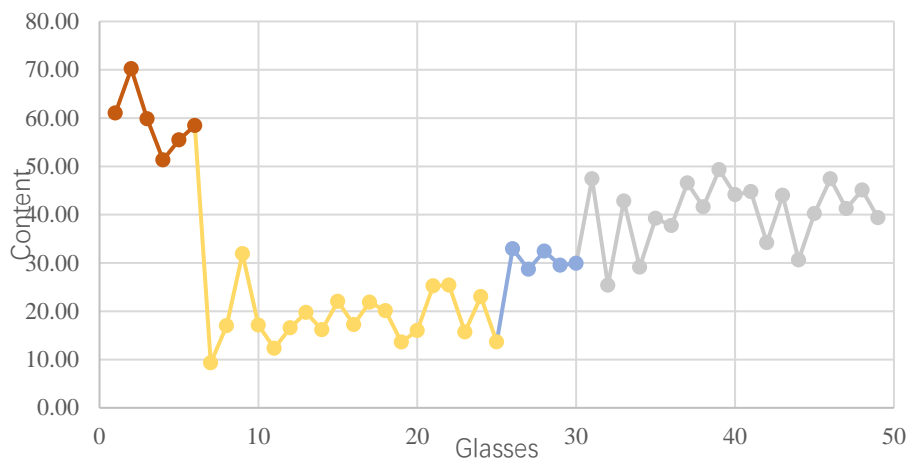


Figure 8. Lead oxide content in lead barium glass

As can be seen from Figure 8, the category shown in red has the highest lead oxide content, and in comparison, the category shown in yellow has the lowest lead oxide content.

In summary, the different categories we have classified for lead-barium glass have significant differences in chemical composition, so the classification of this subcategory is reasonable and sensitive.

3. Conclusion

(1) The degree of surface weathering of glass artifacts is significantly correlated with the type, but not with the decoration and color of the glass, etc.;

(2) High potassium glass and lead-barium glass have significant differences in the content of PbO, which can be used to broadly classify the glass.

(3) The typical chemical compositions of high potassium glass are silica and potassium oxide. The typical chemical compositions of lead-barium glass are silica, lead oxide, and barium oxide. The two types of glass can be further classified according to their typical chemical compositions.

References

- [1] Chen Shuyu, Hou Zhi Li. An analysis of the ancient Silk Road and ancient Chinese glass [J]. China Ethnic Expo, 2019, (05): 87 - 88.
- [2] Gan Fuxi, Cheng Huan Sheng, Li Qinghui. The origin of ancient Chinese glass - a study of the earliest ancient glass in China [J]. Chinese Science (Series E: Technical Sciences), 2007, (03): 382 - 391.
- [3] Cui Jianfeng, Wu Xiaohong, Tan Yuanhui, et al. Composition analysis of ancient glassware excavated from Chu tombs of the Warring States period in the Yuanshui basin, Hunan [J]. Journal of Silicates, 2009, 37 (11): 1909 - 1913.
- [4] Higher Education Society Cup National Student Mathematical Modeling Competition. 2022.
- [5] Shoukui Si, Zhaoliang Sun, editors, Xijing Sun, et al. Mathematical modeling algorithms and applications [M]. Beijing: National Defense Industry Press, 2015.
- [6] Ste. Spielman [J]. Statistics and Forecasting, 1999, (05): 53 - 54.
- [7] Zhou Zhihua. Machine Learning [J]. Aeroport, 2018, (02): 94.