Composition Analysis and Identification of Ancient Glass Objects Using Regression and Clustering Algorithms

Wangxiao Mao

North China Electric Power University, Baoding, Hebei, China

2628763305@qq.com

Abstract. Among the artifacts excavated today, many are glass-products. Archaeologists have classified archaeological artifacts into two main types, high potassium glass and lead-barium glass, based on the chemical composition of the glass artifacts and other testing methods. For the protection of cultural relics, the chemical composition of the artifacts needs to be further analyzed and sub-categorized in order to achieve different protection measures for different categories. In this paper, firstly, based on the data, we analyze the classification pattern of high potassium glass and lead-barium glass, use logistic regression, test the significance of the results, and evaluate and analyze the regression results with the statistical distribution of the data. Then for different categories select suitable chemical indicators for subclass classification, suitable chemical indicators imply their relatively large amount of information, so use the entropy weight method to process chemical composition indicators to derive their information entropy weights, select chemical indicators according to the weights, and finally perform clustering classification according to the selected indicators. Finally, for the analysis of the rationality and sensitivity of the classification results, different classification algorithms may lead to different classification results, so the rationality and sensitivity can only be analyzed by comparing and analyzing the classification results based on different clustering algorithms for the selected chemical composition indicators.

Keywords: Logistic-Regression; Clustering; Reasonableness and Sensitivity Analysis.

1. Introduction

The Silk Road was an important channel for commodity and cultural exchanges between China and the West in ancient times, and glass was an important evidence of prosperous commodity trading. Early Western glass was usually made into jewelry to trade into China in the form of commodities, from the results of the analysis of the chemical composition of ancient glass in the world, the chemical composition of ancient glass in the West is dominated by nacre, while China is dominated by lead barium, lead barium glass can not be found in ancient glass in the West. It can be seen that the chemical composition of ancient glass in China and ancient glass in the West but very different. Therefore, the analysis and identification of the composition of ancient glass products in China is of great significance.

Glass is made of quartz sand, whose main component is silica, and in order to make the melting temperature lower during the calcination process, it is necessary to add fluxes. Different glass added fluxes are different, for lead-barium glass, fluxes for lead ore; for potassium glass, to grass ash and other substances with high potassium content as a flux. Ancient glass is highly susceptible to weathering after burial, resulting in changes in the proportion of glass composition, but for glass samples without weathering on the surface there may be localized weathered areas, and for glass samples with weathered surfaces, there are also unweathered areas.[1,2]

2. Approaches

Data from the official website of the National Student Mathematical Modeling Competition 2022 C Question “Analysis and Identification of the Composition of Ancient Glass Products”

1)Using a classification model, dummy variables are created for the content of chemical components in the artifacts and whether the surface is weathered, and a logistic regression model is used, with the dependent variable being binary variable, representing the category of the artifact. 1

indicates lead-barium glass and 0 indicates high-potassium glass, while setting the training group and test group during the training makes the accuracy of model prediction greatly improved.

2)For each category of artifacts, the entropy weighting method is performed to select the chemical components that have a greater impact on the artifacts in that category for further cluster analysis. Use the entropy method to ensure as much as possible that the results are objectivity and accuracy.[3,4]

3)Systematic clustering was carried out for the two categories of high potassium and lead-barium, respectively, and the clustering results were analyzed by combining the characteristics of the specific chemical composition of each subcategory with the differences in chemical composition. [5,6,7]

4)For the sensitivity of subclass delineation, the K-means++ clustering method is used to compare the differences between two different clustering methods for subclass delineation and further determine whether the results of clustering are sensitive or not.[6]

3. Experimental Study

A.Solving the classification law of high potassium, lead barium glass

1)Binary logistic regression

The logistic regression model is expressed in the form:

$$logit\ P = ln\left(\frac{P}{1-P}\right) = a + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m$$

a is a constant term that represents the estimate of logit P when all independent variables are zero. b_i is the partial regression coefficient, which indicates the amount of change in logit P for each unit change in x_i, with other independent variables held constant.For this problem, the value of a is not meaningful as a constant term, while x_i denotes each chemical composition, and its b_i coefficient indicates the magnitude of the effect of chemical composition x_i on the glass type.[3]

For the category of glass artifacts, it belongs to the fixed type data and only two categories, so the SPSS software was used to import the data after the anomaly processing for binary logistic regression. When the result is 1, it means its category is lead barium; when the result is 0, it means its category is high potassium.

The exact accuracy table of the regression is shown in the following table.

Table 1. Logistic Regression Exact Table

	0	1	Correct Percentage
0	14	0	100%
1	0	38	100%
Overall percentage			100%

From the table, we can see that the overall accuracy is 100%, which is very high. The regression results obtained are shown in the following table.

Table 2. Binary logistic regression results

Ingredients	Coefficient	Ingredients	Coefficient
SiO_2	-0.94	PbO	-0.227
Na_2O	-0.004	BaO	1.065
K_2O	-3.301	P_2O_5	-0.237
CaO	-2.247	SrO	10.569
MgO	-9.922	SnO_2	-9.267
Al_2O_3	1.354	SO_2	-3.919
Fe_2O_3	1.217	Surface weathering	3.285
CuO	-5.681	Constant	79.142

The results of the significance test are shown in the following table.

Table 3. Binary logistic significance test results

Ingredients	Significance	Ingredients	Significance
SiO_2	0	PbO	0
Na_2O	0.447	BaO	0
K_2O	0	P_2O_5	0.032
CaO	0.002	SrO	0
MgO	0.416	SnO_2	0.345
Al_2O_3	0.088	SO_2	0.365
Fe_2O_3	0.021	Surface weathering	0.061
CuO	0.79	Constant	0

According to the above figure, it can be seen that the p-values of the seven chemical components of silica, potassium oxide, calcium oxide, iron oxide, barium oxide, phosphorus pentoxide and strontium oxide are less than 0.05, so they can be considered to be significantly related to the classification of high potassium, lead and barium glass.

2)The data were statistically analyzed using Excel software to average each chemical composition of glass in the high potassium and lead-barium categories and visualized using histograms as follows.

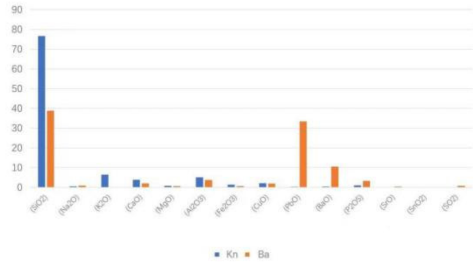

Figure 1. Comparison of the average value of each chemical composition of glass

3) Combined with the chemical background described, for lead-barium glass, the flux is lead ore; for potassium glass, substances with high potassium content, such as grass ash, are used as fluxes. Therefore, for high potassium glass, the potassium content in its chemical composition should be significantly larger than that of lead-barium glass, and for lead-barium glass, the lead content in its chemical composition, the barium content should also be significantly larger than that of high potassium glass. Combined with the above graph and the binary logistic regression results, it can be concluded that the classification results are scientifically and accurately. That is, the classification of high potassium, lead and barium glass depends mainly on the seven chemical components of silica, potassium oxide, calcium oxide, iron oxide, barium oxide, phosphorus pentoxide and strontium oxide.[8]

B.Select chemical composition for subclass classification

Firstly, it is necessary to select the appropriate chemical composition for each of the two types of high potassium and lead-barium, and this paper uses the entropy weighting method to select the appropriate chemical composition by calculating the information entropy of each chemical composition.

1)Entropy method is an objective assignment method. In the specific use process, the entropy weight method uses the information entropy to calculate the entropy weight of each index according to the degree of variation of each index, so as to arrive at a more objective index weight. The steps to calculate the information entropy weights of each chemical component are as follows.

Table 4. The steps to calculate the information entropy weights of each chemical component

Step
a.Calculate the weight of the i evaluation object about the j index value using the original matrix
b.Calculate the entropy value of the jth index value
c.Calculate the coefficient of variation of the jth indicator
d. Calculate the weight of the jth indicator

Based on the information entropy weights and cumulative sums of the two types of glass chemical compositions, the chemical composition indexes with cumulative sums over 80% and weights greater than 0.05 were selected. Also referring to the conclusion of related information, the results of the selection of chemical composition indexes are as follows.

For high potassium glass, tin oxide, sodium oxide, sulfur dioxide, barium oxide, lead oxide, strontium oxide are selected. Since the weight of phosphorus pentoxide is only 0.0423, phosphorus pentoxide is discarded even if the cumulative sum is 0.7827; for lead-barium glass, sulfur dioxide, tin oxide, sodium oxide, potassium oxide, iron oxide, phosphorus pentoxide, and copper oxide are selected.

2)Sub-classification of the two glasses using systematic clustering.

After suitable chemical indicators have been selected for each of the two classes of glass, the two glasses are then clustered separately to obtain the classification of sub-classes.

The two glasses are systematically clustered separately using the selected chemical components. The merging algorithm of systematic clustering combines the data points of the two closest classes by calculating the distance between them, and iterates this process repeatedly until all data points are combined into one class and a clustering spectrum map is generated. The clustering process is shown in the following figure.

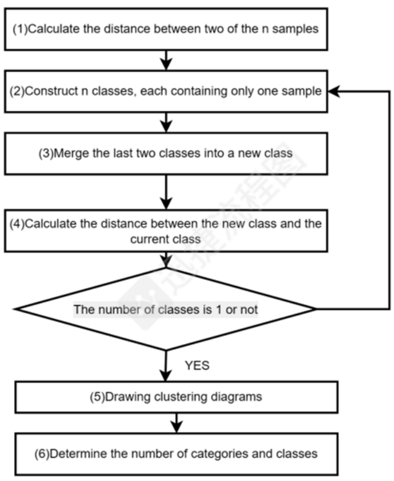

Figure 2. The clustering process

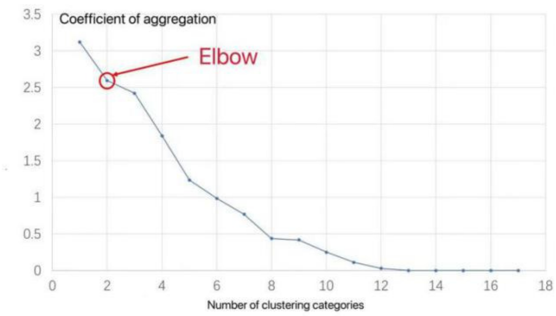

Figure 3. Systematic clustering folding diagram for high potassium types

In order to estimate the number of clusters through the clustering graph, this paper refers to the elbow rule, which means that the optimal number of clusters is roughly estimated through the graph.

According to the above figure, it can be seen that: when the number of categories from 1 to 2, a clear trend change occurs, and at the same time the number of samples is only 18, so it is more reasonable to divide into two categories. Therefore, it can be considered that the elbow is K=2, so the category can be set to 2. That is, the high potassium type glass is sub-classified into two categories.

Similarly, the same treatment for lead-barium type glass, the obtained K=2.

3)According to the data, for high potassium type glass, class 1 has no sodium oxide component, while class 2 has relatively high sodium oxide component; class 2 has no barium oxide, tin oxide, class 1 has barium oxide, tin oxide component content is relatively high. For lead barium glass, 1 type of sodium oxide, magnesium oxide, iron oxide composition is significantly higher than 2 types, 2 types of barium oxide, sulfur dioxide composition is significantly higher than 1 type.

According to the above analysis results and reference to relevant literature, the high potassium glass is classified as: $Sn_2O-BaO-SiO_2$ glass system (1 category), Na_2O-SiO_2 glass system (2 category). The lead-barium glass is classified as: $Na_2O-MgO-Fe_2O_3-SiO_2$ glass system (Class 1), $BaO-SO_2-SiO_2$ glass system (Class 2).[8]

Table 5. Reasonableness and sensitivity of analysis results

Steps
Step1: Randomly select a sample as the first clustering center
Step2: Calculate the shortest distance between each sample and the currently existing cluster center
Step3: Repeat Step2 until K clustering centers are selected

In order to analyze the reasonableness and sensitivity analysis of the classification results derived from systematic clustering, this paper uses another clustering algorithm, the K-means++ algorithm, to classify sample data from two types of glass, respectively. k-means++ algorithm is described as follows.

The basic principle of the k-means++ algorithm for selecting the initial clustering centers is that the initial clustering centers should be as far away from each other as possible. The description of the algorithm for selecting the initial clustering centers is shown in the following table.

The value of K for clustering is the number of sub-classes of the systematic clustering, i.e., K=2. The data of the two types of glass were clustered using SPSS with the K-means++ algorithm with K=2. The specific classification comparison results are shown below.

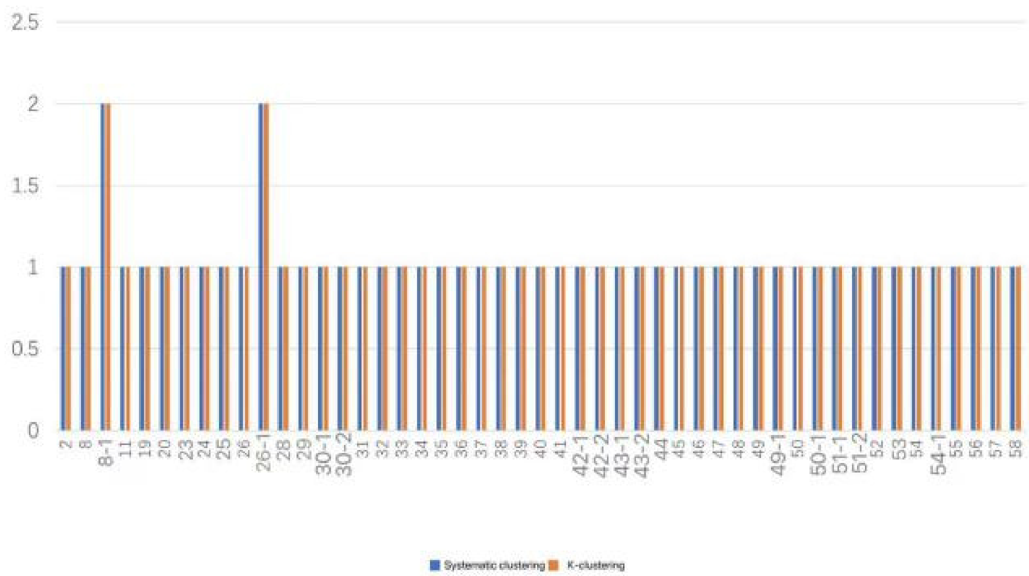

Figure 4. Comparison of the results of two classifications of lead and barium types

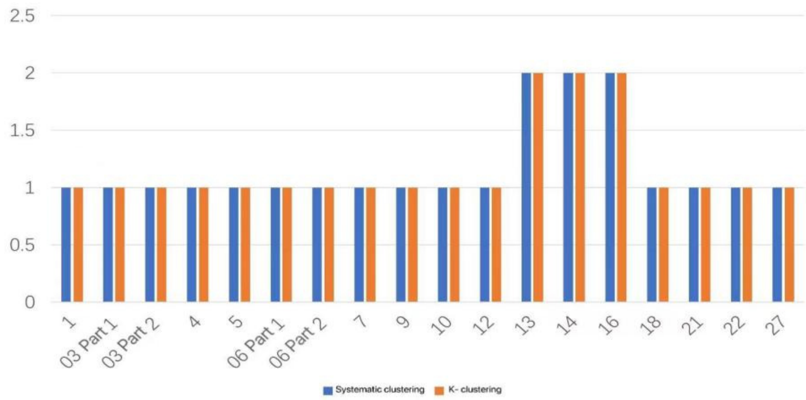

Figure 5. Comparison of the results of two classifications of high potassium types

According to the above figure, it can be seen that the classification results of the two classification methods are exactly the same, so it can be considered that the system clustering model has strong rationality and strong stability, i.e., the model is insensitive.

References

[1] Li QH, Huang KJ, Li F, Gan FX. Chemical composition of a group of ancient Warring States glass samples excavated in China[J]. Conservation of cultural relics and archaeological science,2006(02):8-13.DOI:10.16334/j.cnki.cn31-1652/k.2006.02.002.

[2] Zhao Yutang, Zhang Ming, Zheng Zhong. Study on the relationship between glass fiber corrosion resistance and chemical composition[J]. Power Technology,1995(05):25-28.

[3] Yu, Qiuju. Research on the implementation effect of bilingual course teaching for college students based on logistic regression analysis[J]. Science and Technology Wind,2020, No.410(06):73-74.DOI: 10. 19392/ j.cnki.1671-7341.202006066.

[4] Tao Xiangrong. Multiple logistic regression analysis of readers' characteristics and needs[J]. Modern Intelligence,2006(06):144-148.

[5] DU Jun-Peng, Lv Jun, Wu Jisheng, Zhao L-Y, Wei Chun-Feng, Zhang Y. Research on river water quality evaluation based on principal component analysis and systematic cluster analysis[J]. Water Resources Technology Supervision,2022,No.182(12):216-220.

[6] Tang F.H., Fang X.N., Kang Z.M., Yang W.Q., Li D.. Deep mining of systematic cluster analysis method in the application of student performance evaluation[J]. Journal of Jiamusi University(Nature Science Edition),2022,40(05):164-167.

[7] Gao Huixian, Xue Yulian, Fang Zhong. A comparison of the current situation of financial development in each province based on system clustering[J]. Journal of Fujian Normal University Fuqing, 2021, 39 (02): 189-195.

[8] GB/T 33503-2017, Method for analysis of chemical composition of lead-containing glass [S].