Composition analysis and identification of ancient glass products based on K-Means clustering analysis algorithm

Yijing Zhong[1, *, #] and Yuhang Pei[2, #]

[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China

[2]College of Science, Wuhan University of Science and Technology, Wuhan, China

*Corresponding author: 15171626524@163.com

#These authors contributed equally to this work

Abstract. Ancient glass is a valuable physical evidence of early trade exchanges, and it is very susceptible to weathering caused by the environmental impact of burial. Therefore, the establishment of a glass cultural relics identification model and the analysis of the chemical composition of glass cultural relics are extremely important for the classification and protection of cultural relics in our country. In this paper, we first use differential analysis to find out the chemical composition with significant differences before and after weathering, and obtain the classification laws of high potassium glass and lead-barium glass. Then the 14 chemical components of high potassium glass and lead-barium glass were analyzed by K-Means clustering, and finally 8 subclasses were divided. After testing, the model has low sensitivity, strong stability and good robustness. On this basis, the improved PYH-Topsis solution distance method is used to construct a judgment matrix, analyze the chemical composition of glass cultural relics of unknown categories, and identify the types to which they belong.

Keywords: PYH-Topsis method; K-Means clustering; High potassium glass; Lead barium glass.

1. Introduction

Ancient glass is very susceptible to weathering due to the influence of the buried environment. During the weathering process, internal elements are exchanged with environmental elements in large quantities, resulting in changes in their composition ratio. The surface of the severely weathered glass has been completely covered by weathered objects, and its original appearance is almost unrecognizable, thus affecting the correct judgment of its category [1]. In this paper, by analyzing the differences in the chemical composition of weathered and non-weathered high-potassium glass and lead-barium glass respectively, the classification laws of high-potassium glass and lead-barium glass are obtained, and some of the chemical components are classified into subclasses, and finally the rationality and sensitivity of the classification results are analyzed.

In addition, this paper also analyzes the chemical composition of glass cultural relics of unknown categories, identifies the types of glass cultural relics according to the above-mentioned subclasses, and analyzes the sensitivity of the classification results.

2. Glass composition analysis based on K-Means clustering analysis algorithmt

2.1. Classification law of glass

For high-potassium glass and lead-barium glass, Spss software is used in this paper to analyze the differences in the chemical composition of weathered glass and non-weathered glass [2], and the differences between silica ($SiO2$), potassium oxide ($K2O$), barium oxide (BaO), and lead oxide (PbO) before and after weathering are relatively significant. Moreover, the mean content of these four chemical components in high-potassium glass was 76.6438889, 6.401666667, 0.398888889, 0.274444444, and the mean content of these four chemical components in lead-barium glass was 34.42307692, 0.175128205, 11.37128205, 36.74051282.

Since silica (SiO2) is the main component of glass, the content in high-potassium glass and lead-barium glass is very high, but the potassium oxide content of high-potassium glass is much higher than that of lead-barium glass, and the barium oxide and lead oxide content of lead-barium glass are much higher than that of high-potassium glass. Therefore, it is reasonable to use the content of potassium oxide, lead oxide and barium oxide to classify high-potassium glass and lead-barium glass.

2.2. Subclasses of glass

First, the 14 chemical components of high potassium glass and lead-barium glass were analyzed by K-Means clustering [3].The steps are as follows:

1. Randomly select k points from the chemical composition sample points as the initial cluster center;
2. Divide each sample point into the cluster represented by the center point closest to it;
3. Replace the original center point with the center point of all samples in each cluster;
4. Repeat steps 2 and 3 until the center point remains unchanged or the predetermined number of iterations is reached, the algorithm terminates;

The objective function of the K-Means algorithm is the sum of squares of errors in the cluster:

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{k} w^{(i,j)} \parallel x^i - \mu^j \parallel^2 \tag{1}$$

The results of the K-Means clustering analysis are shown in Table 1 and Table 2.

Table 1. Analysis of the differences of high potassium glass and lead-barium glass

Analysis on the Difference of High Potassium Glass				
	Cluster category (mean ± standard deviation)		F	P
	Category 1 (n=9)	Category 2 (n=9)		
Silicon dioxide (SiO2)	63.624±3.558	89.663±7.119	96.333	0.000***
Sodium oxide (Na20)	0.927±1.427	0.0±0.0	3.796	0.069*
Potassium oxide (K20)	10.818±2.37	1.986±3.22	43.917	0.000***
Sulfur dioxide (S02)	0.136±0.205	0.0±0.0	3.923	0.065*
Tin oxide (Sn02)	0.0±0.0	0.262±0.787	1	0.332
Phosphorus pentoxide (P205)	1.523±1.652	0.533±0.451	3.007	0.102
Strontium oxide (SrO)	0.048±0.051	0.008±0.023	4.516	0.050**
Calcium oxide (CaO)	6.363±2.64	1.327±1.419	25.408	0.000***
Barium oxide (BaO)	0.579±1.001	0.219±0.657	0.814	0.38
Lead oxide (PbO)	0.41±0.64	0.139±0.333	1.271	0.276
Magnesium oxide (MgO)	1.133±0.672	0.437±0.593	5.435	0.033**
Copper oxide (CuO)	2.819±1.565	1.492±1.136	4.233	0.056*
Iron oxide (Fe203)	2.312±1.643	0.44±0.735	9.737	0.007***
Alumina (A1203)	7.349±2.346	2.764±1.67	22.809	0.000***
Note: *, ** and *** represent the significance level of 1%, 5% and 10% respectively				

Analysis of the difference of lead barium glass				
	Cluster category (mean ± standard deviation)		F	P
	Category 1 (n=7)	Category 2 (n=4)		
Silicon dioxide (SiO2)	39.03±9.402	12.065±9.13	21.344	0.001 ***
Sodium oxide (Na2O)	1.461±2.975	0.0±0.0	0.921	0.362
Potassium oxide (K20)	0.281±0.427	0.1±0.2	0.62	0.451
Sulfur dioxide (S02)	0.0±0.0	8.88±7.646	10.3	0.011**
Tin oxide (SnO2)	0.0±0.0	0.0±0.0		NaN
Strontium oxide (SrO)	0.313±0.289	0.493±0.107	1.381	0.27
Phosphorus pentoxide (P205)	3.98±4.101	5.08±2.089	0.243	0.634
Lead oxide (PbO)	28.994±13.446	30.145±1.622	0.028	0.871
Barium oxide (BaO)	12.607±9.635	32.388±2.15	15.702	0.003***
Copper oxide (CuO)	3.721±2.719	6.93±4.116	2.478	0.15
Iron oxide (Fe203)	0.893±0.85	0.0±0.0	4.217	0.070*
Alumina (A1203)	3.193±1.794	1.082±0.273	5.223	0.048**
Magnesium oxide (MgO)	0.456±0.465	0.0±0.0	3.674	0.088*
Calcium oxide (CaO)	1.483±1.406	2.28±0.95	1	0.344
Note: *, ** and *** represent the significance level of 1%, 5% and 10% respectively				

According to the results of clustering analysis, the analysis of the differences in high-potassium glass shows that the differences in silica (SiO2) are significantly higher than those of other chemical components. Therefore, the first division is based on the content of silica (SiO2) and is divided into two categories, named high-potassium high-silicon glass and high-potassium low-silicon glass. By calculating the mean and standard deviation of each chemical composition of high-potassium and low-silicon glass, and comparing the Z values obtained after standardization, Z=standard deviation/mean.

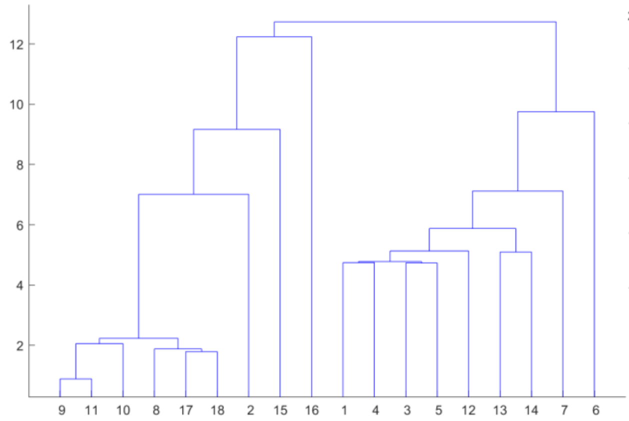

Fig.1 High potassium glass cluster diagram

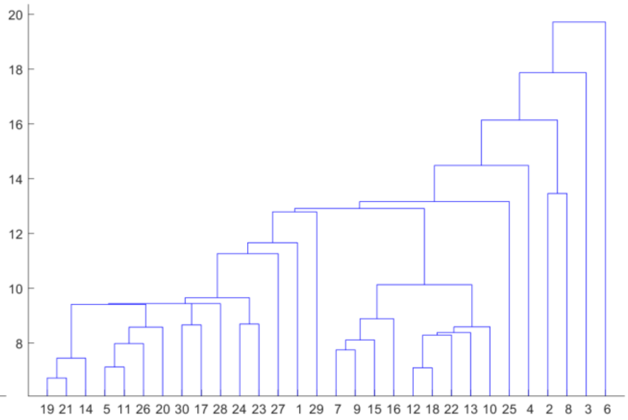

Fig.2 Lead-barium glass cluster diagram

From the data, it can be seen that the Z value of barium oxide (BaO) is the largest, but the specific content is continuous, the center distance is small, and the scatter plot is more discrete. As shown in Figure 5-10, there is no significant difference and it should be discarded. The Z value of sodium oxide (Na2O) is second only to barium oxide (BaO), the specific content is polarized, the center distance is large, and the scatter plot is more concentrated, which has significant differences. Therefore, the second division is based on the content of sodium oxide (Na2O), named high potassium, low silicon, high sodium glass and high potassium, low silicon and low Sodium glass.

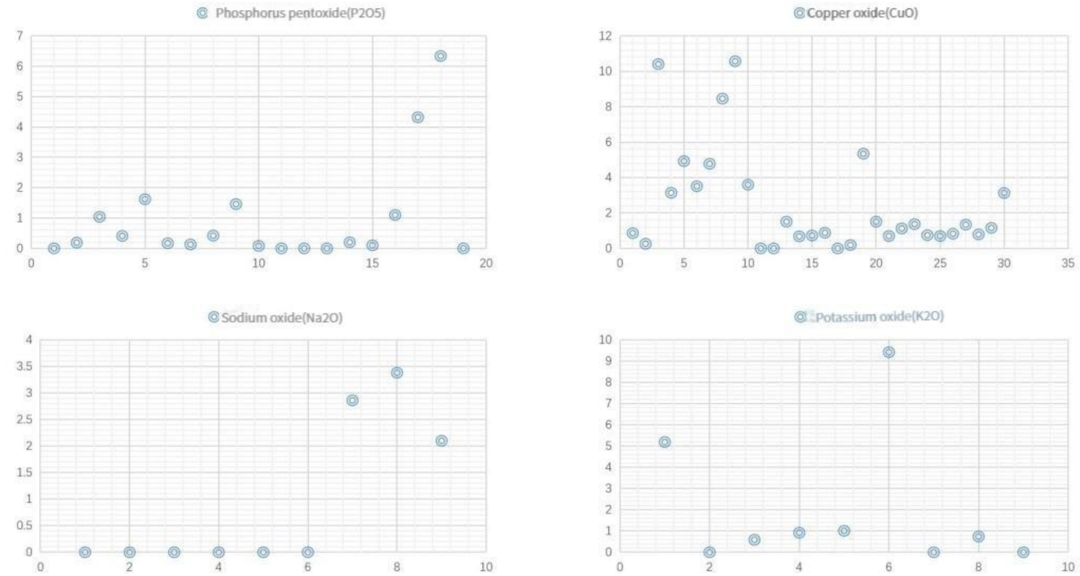

Fig.3 Scatter chart of four chemical components

By analogy, the second division of high-potassium and high-silicon glass is based on the content of potassium oxide (K2O), and is named high-potassium, high-silicon and high-potassium glass (that is, the content of potassium oxide (K2O) is higher in high-potassium glass) and high-potassium, high-silicon and low-potassium glass (that is, the content of potassium oxide (K2O) is lower in high-potassium glass).

From the analysis of the differences in lead-barium glass, it can be seen that the differences in silica (SiO2) are still significantly higher than those of other chemical components. Therefore, the first division is based on the content of silica (SiO2) and is divided into two categories, named lead-barium high silicon glass and lead-barium low silicon glass.

By comparing the Z-value of the chemical composition of lead-barium high-silicon glass and lead-barium low-silicon glass, the distribution of specific content, the length of the center distance, and the concentration of the scatter plot, it can be seen that the second division of lead-barium high-silicon glass is based on the content of phosphorus pentoxide (P2O5), named lead-barium high-silicon high-phosphorus glass and lead-barium high-silicon low-phosphorus glass.

By analogy, the second division of lead-barium low-silicon glass is based on the content of copper oxide (CuO), which is divided into lead-barium low-silicon and high-copper glass and lead-barium low-silicon and low-copper glass.

In summary, after two subcategories of each category are divided in this paper, 8 classification forms are obtained, namely high potassium, high silicon, high potassium glass, high potassium, high silicon, low potassium glass, high potassium, low silicon, low sodium glass, high potassium, low silicon, low sodium glass, lead-barium, high silicon, high phosphorus glass, lead-barium, high silicon, low phosphorus glass, lead-barium, low silicon, high copper glass, lead-barium, low silicon, low copper glass.

2.3. Sensitivity analysis

For the mathematical model constructed above, this paper gives it 85 by changing the decision-making chemical composition content of the subcategory 85%, 90%, 95%, 105%, 110%, 115% The disturbance factor is obtained, and the judgment matrix is obtained. The results are shown in Table 2, where “0” represents no change in the sub-classification, and “1” represents a change in the sub-classification.The experimental results prove that the mathematical model has low sensitivity, strong stability and good robustness.

Table 2. Subclasses sensitivity analysis

Silicon dioxide	Potassium oxide	Phosphorus pentoxide	Sodium oxide	Copper oxide
0	0	0	1	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	1	0	0	0
0	0	0	0	0

3. Type identification based on PYH-TOPSIS method

3.1. Model establishment and solution

This paper regards the type identification of glass cultural relics as a system, establishes a hierarchical structure model according to the overall goals, evaluation criteria, and decision-making layers, compares the importance of various elements in the evaluation criteria, and constructs a corresponding judgment matrix. Relying on the judgment matrix, the relative importance weight value of each factor at the same level is calculated. After passing the consistency test of the judgment matrix, the relative importance of each factor is finally determined based on the comprehensive judgment of the decision maker.

Let the overall research object have m evaluation objects, and compare these evaluation objects in pairs to obtain a judgment matrix Z[4].

$$Z = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mm} \end{bmatrix} \tag{2}$$

Where x_{ij} is the comparison result of the relative importance of the i-th evaluation object and the j-th evaluation object, (i, j=1,2,3,..., m).

Through differential analysis, the four chemical components with the greatest heterogeneity were selected from the high potassium and lead-barium glass artifacts, namely silica (SiO2), potassium oxide (K2O), lead oxide (PbO) and barium oxide (BaO).Assuming that $E_i(i = 1,2,3,4)$ corresponds to the four chemical components of high potassium glass and $E_I'(i = 1,2,3,4)$ corresponds to the four chemical components of lead-barium glass, then:

$$P_j = \sum_{i=1}^{4} \frac{E_i - x_{ij}}{E_i - E_i'} w_i \quad (i, j = 1,2,3, \dots, m) \tag{3}$$

Where P_j is the determination value and w_j is the weight value.

When $P_j > 0.5$, the unknown category of glass cultural relics is close to the chemical composition of E_i', it belongs to lead-barium glass;

When $P_j \leq 0.5$, the unknown category of glass is close to the chemical composition of E_i, it belongs to high potassium glass.

In summary, it is learned through Excel solution: A_1, A_5, A_6, A_7 is high potassium glass, and A_2, A_3, A_4, A_8 is lead-barium glass.

3.2. Sensitivity analysis

For the identification and classification results, this paper gives a certain disturbance to the established mathematical model and observes the change of its P_j-value, as shown in Figure 4.In addition, its sensitivity is also verified through the analysis of the model formula, as follows:

$$\Delta P_j = \sum_{i=1}^{4} \frac{E_i - \Delta\alpha \cdot x_{ij}}{E_I - E_i'} w_i \leq \sum_{i=1}^{4} \frac{\Delta\alpha \cdot (E_i - x_{ij})}{E_i - E_i'} w_i \leq \Delta\alpha \cdot w_i \tag{4}$$

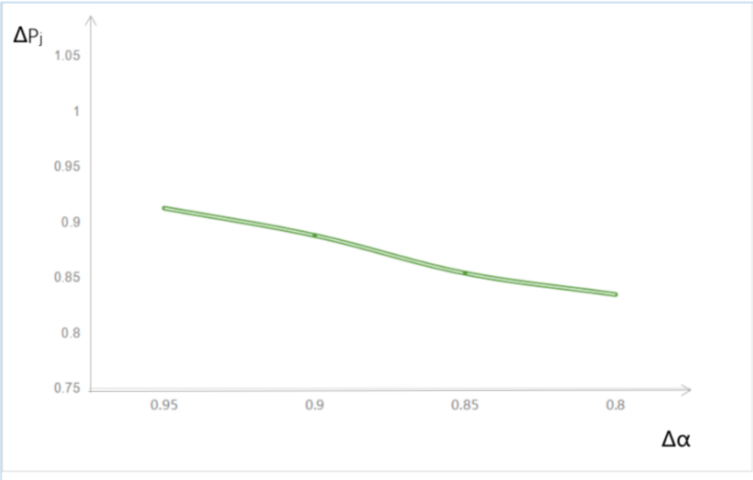

Fig.4 Model sensitivity analysis

It can be seen from the figure that after the mathematical model is disturbed, its ΔP_j-value change range is small, its sensitivity is low, its stability is strong and it has good robustness.

4. Summary

When establishing a subclasses division model, this paper conducted two K-means clustering analyses to reasonably divide the original high-potassium glass and lead-barium glass into 8 types of glass, and the model has low sensitivity and good stability. When establishing the PYH-Topsis model, this paper regards the type identification of glass cultural relics as a system. A hierarchical structure model is established according to the overall goal, evaluation criteria, and decision-making layer. The determination threshold is set very cleverly, which can accurately identify the type of unknown glass. However, the model in this paper fails to have sufficient data, and there are a small number of identification accuracy rates that need to be further improved. Moreover, the chemical composition, which accounts for almost 0%, is not included in the modeling considerations. At the same time, some decision-making indicators do not take into account influencing factors other than natural factors.

The model established in this paper is reasonable and effective for the division of subclasses, has a high identification accuracy rate, and accurately analyzes the correlation and differences between the chemical composition of glass cultural relics. Therefore, the model in this paper can be popularized and applied to cultural relics identification, glass chemical composition analysis and other work. This paper innovates the PYH-Topsis method on the basis of the Topsis method. Although it is only applied to the sorting of the proximity of two-dimensional objects to the target in this article, it can be tried to apply it to three-dimensional objects in the future. At the same time, when the amount of data increases, the identification effect of the model in this paper can be further improved.

References

[1] Wang Jie, Li Mo, Ma Qinglin, etc. Research on the weathering of an octagonal prismatic lead-barium glassware during the Warring States Period [J]. Glass and enamel, 2014, 42(2): 6-13. DOI:10.13588/j.cnki.g.e.1000-2871.2014.02.002.

[2] Zhao Qi, Luo Qing, Chen Lipei, etc. Analysis of the difference in income of agricultural business entities in the central region-based on SPSS clustering, principal component analysis and gray correlation analysis [J]. Anhui Agricultural Science, 2021, 49(9): 221-224. DOI: 10.3969/j.issn.0517-6611.2021.09.060.

[3] Wang Sen, Liu Chen, Xing Shuaijie.Review of research on K-means clustering algorithm [J/OL]. Journal of East China Jiaotong University: 1-8[2022-09-18]. DOI: 10.16749/j.cnki.jecjtu.20220914.001.

[4] Zhou Tao, Wang Yongchao, Li Wenjun, etc. Decision-making of green process plan based on interval number TOPSIS method [J]. Combined machine tools and automated processing technology, 2022(1): 160-163. DOI: 10.13462/j.cnki.mmtamt.2022.01.036.