Classification and study of glass based on cluster analysis

Yueyang Gao [1,*,#], Fengrui Su [1,#], Junjun Xiong [2,#]

[1] School of Economics and Management, Beijing Jiaotong University, Beijing, China

[2] School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China

* Corresponding Author Email: gyy20726051@163.com

[#] These authors contributed equally.

Abstract. As one of the earliest human materials invented by mankind, glass products play a great role in the study of history. Because the ancient glass is easily affected by the buried environment, resulting in the weathering of the surface of the cultural relics, this paper classifies the cultural relics in order to better protect and study the ancient glass. In this paper, some data of weathered cultural relics are collected to establish a model for subclassification of cultural relics.This paper first conducted principal component analysis of two different types of glass and screened out different principal components of two types of glass. Then, according to the number of cluster categories (K value) obtained by the contour algorithm, combined with the obtained principal components, the cluster analysis is used to divide the subclasses, and the specific subclass division method and division results are obtained. The subclass characteristics are obtained by analyzing for each subclass data. For the established subclass classification, multiple cluster analysis was performed using changed K values to perform the sensitivity and rationality analysis of the model. The results show that the model established in this paper finds a good classification rule for some chemical substances of weathered cultural relics, gives a logical division method, and explores and compares the similarities and differences between different chemical components.

Keywords: Principal Component Analysis, Cluster Analysis, Glass Classification.

1. Introduction

The Silk Road connected the ancient trade circulation and cultural exchanges between China and the West, with an endless stream of merchants. Glass products were the most valuable objects in the early days of trade transactions. At the end of the Spring and Autumn Period, glassware was made into zhu-shaped ornaments and imported to China from the West. China also learned the process of glass making, and made from local materials. Therefore, although the appearance was roughly similar, the composition was different. Glass can be identified in the archaeological record by its composition (usually sodium-calcium barium glass). However, these precious glass relics are easily weathered under the influence of buried environment, dark, reduced transparency, halo, or the formation of weathering product crust [1]. Glass is the main raw material of silica (SiO_2), because in the combustion process of flux, and used to change the annealing temperature of additives, cause the chemical composition and content of glass, we choose lead ore as flux can fire form lead barium glass, select higher potassium content of plant ash firing as flux will burn form potassium glass. Oxides will affect the researchers to correctly classify the glass cultural relics. The seriously weathered glass surface has been completely covered by the weathered objects, and it is basically impossible to manually identify [2,3]. In this paper, we will analyze the chemical composition of ancient glass products and identify the correlation between their various components, so as to study and predict the chemical composition of unknown categories of glass relics. Finding practical and effective methods to help the classification of glass cultural relics is of great significance to the implementation of targeted protection measures [4].

As shown in Figure 1, in this paper, the cluster analysis, based on the existing glass chemical composition data, looked for the correlation relationship and internal structure between the data (chemical composition), and specifically analyzed the classification rules of high potassium glass and lead-barium glass. In order to highlight the influence of the main chemical components on the different categories, we precipitated the characteristic elements of each category, and also to give specific sub-classification criteria later. Finally, the accuracy of the model prediction under different parameters is compared.

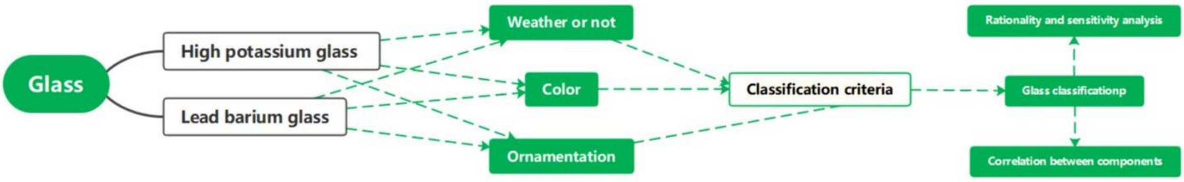

Figure 1. Article Structure

2. Establishment and solving of the glass classification model

In order to establish a model that can effectively classify and subclass of high potassium glass and lead barium glass, we first use principal component analysis to classify the two types of glass, and then establish a cluster model to classify subclasses. Figure 2 is the modeling idea of the overall classification method.

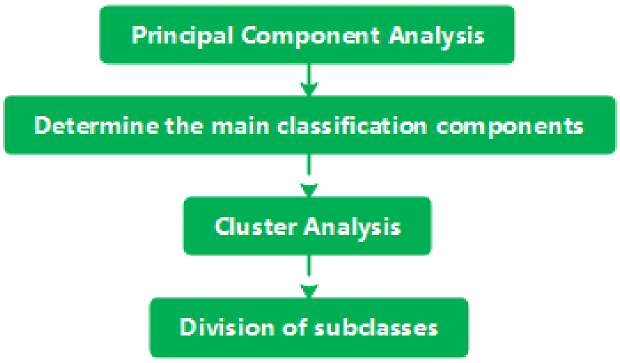

Figure 2. Modeling Process

2.1. Principal Component Analysis

For the collected data, the principal component analysis of high potassium glass and lead barium glass was carried out to obtain the classification of main components [5]. In order to reduce the complexity of the model, we need to reduce the number of variables and convert 14 chemical substances with possible correlation into a set of linear unrelated variables through principal component analysis. The following is the data processing step [6]:

Step 1: Establish a multidimensional vector matrix

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix} = (X_1, X_2, \ldots, X_p) \tag{1}$$

Step 2: Standardized treatment of the matrix and the calculation of the covariance matrix of the standardized samples:

$$R = \frac{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{ki} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)^2 \sum_{k=1}^{n}(x_{kj} - \bar{x}_j)^2}} \tag{2}$$

Step 3: Calculate the eigenvalue and the vector of R:

Characteristic Value $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, (R is positive semidefinite matrix, and $tr(R) = \sum_{k=1}^{p} \lambda_k = p$)

$$\text{Feature Vector:} a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix}, a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix}, \dots, a_p = \begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{bmatrix} \tag{3}$$

Step 4: Calculate the principal component sharing rate and the cumulative contribution rate:

$$\text{Contribution Rate} = \frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k} \tag{4}$$

$$\text{Cumulative Contribution Rate} = \frac{\sum_{k=1}^{i} \lambda_k}{\sum_{k=1}^{p} \lambda_k}, (i = 1, 2, \dots, p) \tag{5}$$

Step 5: Find out the principal components:
The i_{th} principal component:

$$F_i = a_{1i}X_1 + a_{2i}X_2 + \cdots + a_{pi}X_p \tag{6}$$

After the above steps, the cumulative variance interpretation rate reaches more than 80%, and it is identified as the main component. These four main components are Al_2O_3, Na_2O, SnO_2 and PbO. Table 1 Is the variance explained.

Table 1. Principal Component Analysis of High Potassium Type Glass

Total Variance Explained			
Ingredient	Characteristic Root		
	Characteristic Root	Variance Interpretation Rate (%)	Cumulative Variance Interpretation Rate (%)
1	5.871	41.938	41.938
2	2.581	18.436	60.374
3	1.803	12.878	73.252
4	1.789	12.776	86.029
5	0.965	6.892	92.921
6	0.403	2.881	95.801
7	0.236	1.684	97.486
8	0.157	1.123	98.608
9	0.093	0.667	99.275
10	0.072	0.511	99.786
11	0.021	0.148	99.934
12	0.008	0.054	99.988
13	0.002	0.011	99.999
14	0.000	0.001	100

Similarly, for the lead barium type glass, we obtain the principal component of the lead barium type glass. According to Table 2, Based from the cumulative variance interpretation rate and related data, the principal components of lead barium glass are CaO, BaO, SrO, P_2O_5, K_2O, SO_2 and SiO_2. According to the significance analysis of the subsequent cluster analysis, strontium oxide and potassium oxide were removed, and only calcium oxide, barium oxide, phosphorus pentoxide, silicon dioxide and sulfur dioxide were used as the main components of lead and barium glass.

Table 2. Principal Component Analysis of Lead-barium Type Glass

Total Variance Explained			
Ingredient	Characteristic Root		
	Characteristic Root	Variance Interpretation Rate (%)	Cumulative Variance Interpretation Rate (%)
1	3.576	25.54	25.54
2	2.949	21.062	46.602
3	1.665	11.89	58.492
4	1.085	7.748	66.24
5	0.908	6.483	72.723
6	0.841	6.004	78.727
7	0.747	5.338	84.065
8	0.618	4.416	88.482
9	0.564	4.03	92.512
10	0.365	2.605	95.117
11	0.350	2.499	97.615
12	0.206	1.472	99.087
13	0.123	0.88	99.968
14	0.005	0.032	100

2.2. K-Means cluster analysis

According to the classification rules and principal components of high potassium glass and lead barium glass obtained from the principal component analysis as the division basis, the sampling point data of the two types of glass were processed by K-Means clustering analysis through the clustering analysis algorithm. According to the contour algorithm, the K-Means cluster analysis is 8 and 4, respectively. The principal components obtained from the principal component analysis were taken as variables, and the cluster analysis test was conducted according to the K value to obtain the results of subclass partitioning. According to the chart and specific data, the specific subclass division of high potassium glass and lead barium glass, the subclass characteristics and the results of the collected samples are shown in Figure 3.

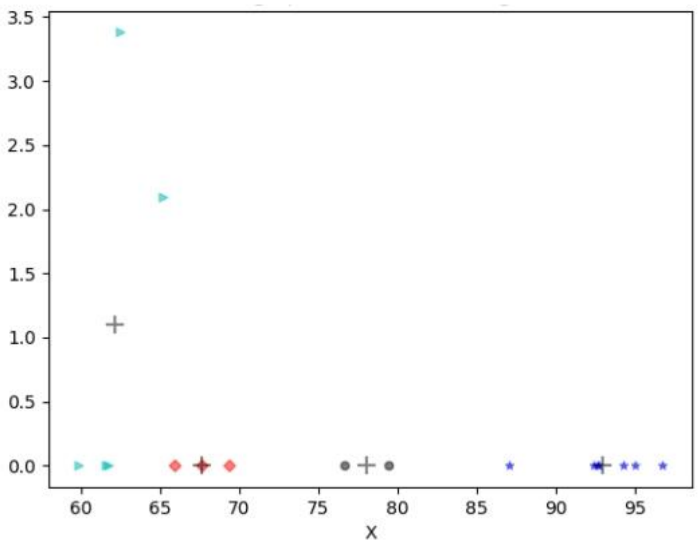

Figure 3. High potassium glass clustering results

For high potassium glass, we divided the following main components, including alumina, sodium oxide (Na_2O), tin oxide (SnO_2), lead oxide (PbO). According to the different contents of the above main components, we get different subcategories and their characteristics. The following are the classification results:

High potassium glass subclass 1 is sulfur rich high potassium, characterized by silica content> 60%; sulfur dioxide content is 0.3%~0.5%; strontium oxide content is about 0.05%; basically no sodium oxide, lead oxide, barium oxide, strontium oxide, tin oxide and other substances.

High potassium glass subclass 2 is silicon rich less potassium and high potassium subclass, characterized by silica content> 90%; potassium oxide content about 1%; calcium oxide content about 1%; basically does not contain sodium oxide, magnesium oxide, barium oxide, lead oxide, strontium oxide, tin oxide, sulfur dioxide and other substances[7].

High potassium glass subclass 3 is sodium-rich lead and high potassium subclass, characterized by sodium oxide about 3.5%; alumina about 1.5%; basically no barium oxide, strontium oxide, tin oxide and sulfur dioxide.

High potassium glass subclass 4 is aluminum-rich phosphorus, characterized by silica content> 90%; potassium oxide content about 1%; calcium oxide content about 1%; basically without sodium oxide, magnesium oxide, barium oxide, lead oxide, strontium oxide, tin oxide, sulfur dioxide[8].

High potassium glass subclass 5 is silicon rich in sulfur, which is characterized by alumina content of about 4%; it basically contains no sodium oxide, barium oxide, strontium oxide and strontium oxide.

High potassium glass subclass 6 is sodium rich calcium and high potassium subclass, with about 2% sodium oxide, about 0.05% strontium oxide; basically no barium oxide, phosphorus pentoxide, tin oxide and sulfur dioxide.

High potassium glass subclass 7 is tin-rich less iron and high potassium, characterized by no iron oxide 2. Tin oxide about 2% 3. Basically no sodium oxide, calcium oxide, iron oxide, copper oxide, lead oxide, barium oxide, sulfur dioxide.

High potassium glass subclass 8 is lead-rich barium high potassium subclass characterized by lead oxide content about 1%; barium oxide content about 2%~3%; basically without sodium oxide, tin oxide and sulfur dioxide.

For the classification of HK glass, the division obtained by the images may be biased due to the small sample size.

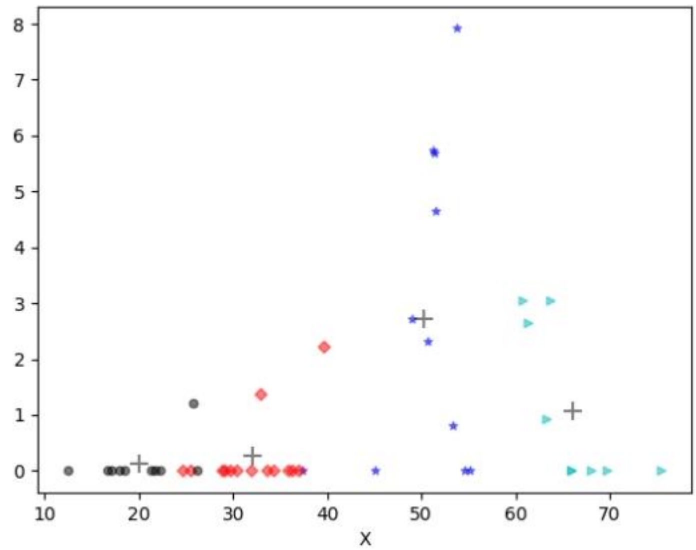

Figure 4. Lead-barium glass clustering results

For lead and barium glass, we divided the following main components, including calcium oxide (CaO), barium oxide (PbO), phosphorus pentoxide (P_2O_5), silica (SiO_2), sulfur dioxide (SO), and obtained different subclasses and their characteristics[9]. As shown in Figure 4, the classification results are presented below:

Lead barium glass subclass 1 is rich silicon high lead less barium lead barium subclass, characterized by silica, lead oxide accounted for about 40%; basically no nano oxide, potassium oxide, iron oxide, strontium oxide, tin oxide, sulfur dioxide.

Lead-barium glass subclass 2 is silicon-rich lead-barium subclass, characterized by silica content over 50%; lead oxide content 15% -35%; basically free of potassium oxide, phosphorus pentoxide, strontium oxide, tin oxide and sulfur dioxide.

Lead barium glass subclass 3 is lead-rich, low-silicon lead barium subclass, characterized by about 20%, 30%, about 40%, potassium oxide, strontium oxide, tin oxide, sulfur dioxide.

Lead-barium glass subclass 4 is a high lead-rich barium subclass, characterized by lead oxide, barium oxide content of about 30%, potassium oxide, magnesium oxide, iron oxide, strontium oxide and tin oxide.

3. Plausibility and sensitivity analysis of the glass classification results

The rationality of the classification results of glass subclass was analyzed, and the relationship between the patterns, types and colors of the subclass was analyzed by images, and the relationship between different detection points in each subclass of high potassium subclass was analyzed. For subclass 1, they are all blue-green, and decorated is A sampling point without weathering. For subcategory 2, all blue-green, with B and weathered sampling sites. For subclass 4, all blue-green with A and no weathering. For subcategory 5, the colors of all three are blue-green, and no weathering is relatively high. For subclass 8, all are blue-green with unweathered sampling sites in A. For subclass 3,6,7, there is only one sampling point, and no detailed analysis is done.

Next, we will analyze the rationality of the subclassification of lead and barium glass. Through the analysis, the following analysis conclusions are obtained: For subclass 1, most of the colors of the sampling points are deep, and the patterns A and C are half. For subclass 2,95% of the sampling points were unweathered, and the colors of the sampling points were mostly light colors, with A and C having half of the two patterns respectively. For subclass 3, the sampling points were all already weathered locations, and 85% of the sampling points were light-colored. For subclass four, the sampling points are all weathered positions, the colors are purple and all are C decorative [10].

4. Conclusions

In conclusion, the subclassification results of the two kinds of glasses are highly reasonable.

For the sensitivity analysis of the subclass classification results, the sensitivity test was performed by changing the K value of the K-Means cluster analysis. Because the significant P value can reflect the variables in the categories of the cluster analysis, so according to the P value of as a variable material sensitivity in high potassium glass, we choose potassium oxide as the main analysis components, found from the figure of potassium oxide and most of the chemical composition are negative correlation for weathering degree, but potassium oxide and silica present larger positive correlation, judgment. After the test, when the number of cluster categories of high potassium glass was 8 or 9, the significant effect was the best, and the P value of all variables was 0.000 * * *, that is, the classification effect was the most obvious, and there were significant differences. Similarly, through the above way, when the cluster category of lead barium glass is selected as 4, the classification effect is best.

References

[1] Wang, J, Li M, Ma Q.L. A study on weathering of an octagonal lead braium glassware in the Warring States Period[J]. Glass and Enamel, 2014,42(2): 7-8.

[2] Feltz A.Encyclopedia of Materials: Science and Technology (Second Edition)[M].Elsevier Ltd, 2001: 6614-6624.

[3] Arun K&C. John.Fundamentals of Inorganic Glasses (Third Edition) [M]. Elsevier Inc,2019: 1-18.

[4] Yi, Q., Wu, S. L., Liu, Y. J., Chan, T., Lu, Y., Saha, N., Southam, G., Huang, L. B. Mineral weathering of iron ore tailings primed by Acidithiobacillus ferrooxidans and elemental sulfur under contrasting pH conditions[J]. Science of The Total Environment, 2023, 856(1):159078.

[5] PedroMartín-Sanz, J., Santiago-Martín, A., Valverde-Asenjo, I., Quintana-Nieto, J. R., González-Huecas, C., López-Lafuente, A. L. Comparison of soil quality indexes calculated by network and principal component analysis for carbonated soils under different uses[J]. Ecological Indicators, 2022, 143:109374.

[6] Serrão, R. G., Oliveira, M. R., Oliveira, L. Theoretical derivation of interval principal component analysis[J]. Information Sciences, 2023, 621:227-247.

[7] Hou, Y. C., Li, H., Zhang, L. Y. Genome-based approach to new Nd: Phosphate glass: Composition-structure-property statistical modeling and validation. Journal of Non-Crystalline Solids[J]. 2023, 600(15):121988.

[8] Das, B., Mandal, A. K. A comparative study on copper doped sodium alumina-phosphate glass with conventional and microwave heating[J]. Optical Materials, 2022, 134(A):113146.

[9] Wu, J. S., Gross, T. M., Huang, L. P., Jaccani, S. P., Youngman, R. E., Rzoska, S. J., Bockowski, M., Bista, S., Stebbins, J. F., Smedskjaer M. M. Composition and pressure effects on the structure, elastic properties and hardness of aluminoborosilicate glass. Journal of Non-Crystalline Solids[J], 2020, 530:119797.

[10] Rossano, S., Khomenko, V., Bedidi, A., Muller, C., Loisel, C., Ferrand, J., Sarrasin, L., Bertin, A. Glass colourations caused by Mn-Fe redox pair: Application to ancient glass technology. Journal of Non-Crystalline Solids[J], 2022, 594:121710.