ISSN 2616-5775 Vol. 5, Issue 13: 38-43, DOI: 10.25236/AJCIS.2022.051306

Composition identification of ancient glass products based on cluster analysis

Zhenghu Pang*

Institute Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, China, 233030
**Corresponding author: pzh17856418906@163.com*

Abstract: *The Silk Road was a passage for the exchange of Chinese and Western cultures in ancient times, and glass was a valuable physical evidence of early trade exchanges. The main raw material of glass is quartz sand, and the main chemical composition is silicon dioxide (SiO2). Due to the high melting point of pure quartz sand, in order to reduce the melting temperature, it is necessary to add flux during refining. Since different materials can cause glass products to have different properties, the classification of glass products is worth studying. In this paper, a composition analysis and identification model of glass products is developed. First of all, the relationship between surface weathering and glass type, ornamentation and color of glass cultural relics is studied, and after the chemical composition content is counted, the chemical composition content before weathering is predicted according to the weathering point detection data.*

Keywords: *Glass composition identification, K-means, Systematic clustering*

1. Introduction

The Silk Road was a passage for the exchange of Chinese and Western cultures in ancient times, and glass was a valuable physical evidence of early trade exchanges. [1]Early glass in West Asia and Egypt is often made into bead-shaped jewelry into China, China's ancient glass to absorb its technology after the local local production, so with the appearance of foreign glass products similar, but the chemical composition is not the same[2]. The main raw material of glass is quartz sand, and the main chemical composition is silicon dioxide (SiO2). Due to the high melting point of pure quartz sand, in order to reduce the melting temperature, it is necessary to add flux during refining. Different fluxes are added, and their main chemical components are also different [3]. For example, lead barium glass is added to lead ore as a flux during the firing process[4], and its lead oxide (PbO) and barium oxide (BaO) content are higher; Potassium glass is fired with a high potassium content such as grass and wood ash as a flux, and its potassium oxide (K2O) content is high[5]. Ancient glass was highly susceptible to weathering due to the influence of the buried environment. In the weathering process, the internal elements and environmental elements are exchanged in large quantities, resulting in changes in the proportion of their composition, which affects the correct judgment of their categories, and this paper studies the analysis and identification of the composition of ancient glass products[6].

2. The basic funamental of Classification models

2.1 The process of clustering a model

In this paper, the data samples are selected to analyze the classification rules of high-potassium glass and lead-barium glass, and the appropriate chemical composition is selected for each category to divide the subclass, the specific division method and the classification result are given, and the rationality and sensitivity of the classification results are analyzed. We classify high-potassium glass and lead-barium glass and subclass divisions, and analyze the rationality and sensitivity of classification models. Therefore, we establish a systematic clustering model of glass species, on this basis, divide the two classes according to the standard deviation, and analyze the rationality and sensitivity of the model[7]. Firstly, we conduct statistics on the values of different chemical components of high-potassium glass and lead-barium glass, and find the change of their representative chemical indicators as the basis for classification [8]. Then, on this basis, subclass division is carried out, the changes of chemical

ISSN 2616-5775 Vol. 5, Issue 13: 38-43, DOI: 10.25236/AJCIS.2022.051306

composition, color changes, texture changes, etc. before and after weathering are observed, and the corresponding classification basis is given; Finally, on this basis, the sensitivity test of the data is carried out, and the relevant rationality basis is given.

The steps for system clustering are as follows:

1) Treat each sample as a class;

2) Calculate the interclass distance matrix and merge the two closest classes into one new class;

3) Calculate the distance between the new class and the current category. If the number of classes is equal to 1, proceed to the next step, otherwise go to 2;

4) Decide on the number and category of clusters.

According to the above calculation steps, different glass types in the system clustering are obtained. For the weathering of high-potassium glass, the K-means algorithm is used for analysis, and the algorithm flow is as follows:

1) Cluster category difference analysis according to the field;

2) Analyze the frequency of each cluster category according to the cluster summary;

3) According to the data aggregation class labeling, you can know which category each sample data is divided into;

4) The cluster center coordinates can be used to analyze the distance between each sample and the center point;

5) Review the analysis.

2.2 The specific process of principal component analysis

We classify high-potassium glass and lead-barium glass and analyze the relationship between chemical composition, so we establish a subclass classification model by principal component analysis to analyze the differences in the relationship between chemical composition associations between different categories [9].

1) In order to eliminate the influence of the dimensions of different variables, it is first necessary to standardize the variables and deal with them;

2) Calculate the correlation coefficient array of the standardized data, and find the eigenvalues and eigenvectors of the correlation coefficient matrix.

3) KMO test and Bartley spherical test;

4) Determine the p principal components and conduct statistical analysis.

5) Find the factor score formula.

3. Results

3.1 The establishment of simulation model

This model mainly studies the classification of ancient glass products, using cluster analysis and principal component analysis to study glass products, and determine their main components after classifying them, providing reference for future glass product identification [10].

3.2 Analysis of experimental results

Different glass types in system clustering. We set the number of clusters to 2, calculated the category to which each class Chinese number, and analyzed whether it actually belonged to high-potassium glass and lead-barium glass, and calculated the number of correct and wrong classifications, and the classification results are shown in Table 1 below. Among them, there are two error data, numbers 11, 48.

ISSN 2616-5775 Vol. 5, Issue 13: 38-43, DOI: 10.25236/AJCIS.2022.051306

Table 1: A table of system clustering results that compare to the actual values

Type of artifacts	High potassium glass	Lead-barium glass
Unweathered	4、5、1、3、13、14、16、6、18、21	30、46、47、25、55、50、29、44、32、35、28、33、31、49、42、23、53、37、20、24
weathered	7、9、10、11、12、22、27、48	8、26、34、38、52、56、57、36、51、58、19、41、50、43、49、39、54、40、2

Combining the above classification results, we cross-compare the chemical composition of high-potassium glass and lead-barium glass, and summarize the types of high-potassium glass subclass classification in the selection of different chemical components. We use the standard deviation as a benchmark to reflect the degree of dispersion of individual chemical components, and the results are shown in Table 2. For high-potassium glass weathering, the K-means algorithm was used to analyze the results, as shown in Table 3 below.

Table 2: Comparison data before and after weathering of standard deviations for different glass types

chemical composition	High potassium weathered	High potassium unweathered	Lead and barium weathered	Lead and barium unweathered
SiO_2	1.582568	8.382369	10.39953	11.56859
Na_2O	0	1.232129	0.545838	2.319506
K_2O	0.406393	3.753404	0.235306	0.303248
CaO	0.445271	2.960822	1.627547	1.25646
MgO	0.279623	0.647359	0.692715	0.534737
Al_2O_3	0.880454	2.385438	2.583112	3.190738
Fe_2O_3	0.063443	1.595738	0.722231	1.12935
CuO	0.853374	1.589319	2.765751	1.926569
PbO	0	0.563912	11.99272	8.034573
BaO	0	0.940291	9.7845	5.697241
P_2O_5	0.191659	1.372911	4.115203	1.806457
SrO	0	0.046338	0.259699	0.238102
SnO_2	0	0.65227	0.264222	0.124538
SO_2	0	0.177615	4.124391	0.746388

Table 3: Differential analysis of high-potassium glass

Component	Cluster category (mean ± standard deviation)		F	P
	Category 2 (n=4)	Category 1(n=2)		
Na_2O	0.0±0.0	0.0±0.0		NaN
K_2O	0.63±0.457	0.37±0.523	0.4	0.561
CaO	0.655±0.354	1.3±0.509	3.495	0.135
MgO	0.0±0.0	0.59±0.071	371.307	0.000
Al_2O_3	1.392±0.481	3.005±0.7	11.709	0.027
SO_2	0.0±0.0	0.0±0.0		NaN
SnO_2	0.0±0.0	0.0±0.0		NaN
SrO	0.0±0.0	0.0±0.0		NaN
P_2O_5	0.277±0.264	0.285±0.106	0.001	0.972
Fe_2O_3	0.26±0.065	0.275±0.106	0.05	0.834
BaO	0.0±0.0	0.0±0.0		NaN
PbO	0.0±0.0	0.0±0.0		NaN
CuO	1.82±1.013	1.045±0.7	0.898	0.397

From the above table, it can be intuitively seen that the significant P-values of magnesium oxide (MgO) and alumina (Al2O3) are less than 0.05, so we choose these two indicators as suitable variables. Similarly, the calculation of the degree of weathering of the other different types of glass is the same. We then come up with a classification for each type of glass. Although each indicator is different in genesis, there is often a correlation between different indicators, which arises because different chemical components play a different role in the classification results. In order to find out what these suitable

ISSN 2616-5775 Vol. 5, Issue 13: 38-43, DOI: 10.25236/AJCIS.2022.051306

chemical composition factors are and the corresponding dominant roles, we use the principal component analysis method to solve these problems, and analyze the data after the weathering of lead-barium glass.

In order to eliminate the influence of the dimensions of different variables, it is first necessary to standardize the variables in the processing; According to the distribution of the standard deviation, a total of 6 suitable chemical indicators are finally selected, 27 sample objects, and the i th index value of the j sample is F_{ij}, and the standardized values are standardized as follows:

$$F_{ij} = \frac{F_{ij} - \overline{F}_i}{S_i} \tag{1}$$

Where F_i and S_i are the mean and standard deviation of the indicator respectively. The purpose of standardization is to eliminate the effects of the dimensions of different variables, and the transformation of standardization does not change the correlation coefficient of the variables. Then we calculate the correlation coefficient array of the standardized data and find the eigenvalues and eigenvectors of the correlation coefficient matrix. Note that the correlation coefficient between the ith indicator and the ith indicator is $r_{ii'}$, which is calculated as

$$r_{ii'} = \frac{\sum_{k=1}^{27} \overline{F}_{ik} \overline{F}_{i'k}}{27-1}, i, i' = 1, 2, \cdots, 27 \tag{2}$$

The correlation coefficient matrix is $R = \left(r_{ii'}\right)_{27\times27}, r_{ii} = 1, r_{ii'} = r_{i'i}$. Then we perform the KM0 test and the Bartley spherical test. The KMO test statistic is used to compare the correlation coefficient with the partial correlation coefficient between the variables, and the statistics are:

$$KMO = \frac{\sum\sum_{i\neq j} r_{ij}^2}{\sum\sum_{i\neq j} r_{ij}^2 + \sum\sum_{i\neq j} p_{ij}^2} \tag{3}$$

Where r_{ij} is the correlation coefficient between variable x_i and x_j, and the partial correlation coefficient p_{ij} between variable x_i and x_j. The range of values for statistics is [0,1]. The closer the KMO is to 1, the more correlated the variables are, and the more suitable they are for factor analysis.

The Bartley spherical test statistics are obtained from the correlation coefficient matrix of the original variables and approximate the chi-square distribution. If the chi-square value is significant and P is less than 0.05, the null hypothesis is rejected to indicate that there is a correlation between the variables, that is, the original variable is suitable for factor analysis. Conversely, if it is about 0.05, it means that factor analysis is not suitable.

Table 4: KMO and Bartlett tests

KMO		0.476
Bartlett sphericity test	d.f.	45.053
	N	28
	p	0.000

From Table 4, it can be seen that the value of KMO is 0.476, indicating that the correlation between the variables is general, and factor analysis can be basically performed. In order to see if the data come from a population that follows a multivariate normal distribution, the significant value in the table is 0, indicating that the data comes from a normally distributed population and is suitable for further analysis. According to the above steps, the correlation coefficient table of each indicator can be found, and it can be found that some indicators have a strong correlation, and the total variance is explained in the following table5.

ISSN 2616-5775 Vol. 5, Issue 13: 38-43, DOI: 10.25236/AJCIS.2022.051306

Table 5: Explanation of the total variance

ingredients	eigenvalue	Variance percentage	Cumulative %
1	2.162	36.039	36.039
2	1.600	26.659	62.698
3	1.153	19.221	81.918
4	0.537	8.951	90.869
5	0.365	6.081	96.950
6	0.183	3.050	100.000

As can be seen from the above table, the cumulative contribution rate reached 81.917% when the three principal components were extracted, and the composition score coefficient matrix after the weathering of lead barium glass was shown in Table 6.

Table 6: Table of factor load factors

Index	Ingredient1	Ingredient2	Ingredient3
CaO	-0.187	0.389	0.361
Al2O3	-0.223	0.302	-0.578
CuO	0.369	0.185	0.198
PbO	-0.200	-0.440	0.401
BaO	0.419	0.106	0.041
P2O5	-0.160	0.379	0.448

Finally, the factor scoring formula and the weight of the factor are obtained.

Formula for lead-barium weathered glass model:

$$\begin{cases} F1=-0.16x_4-0.2x_6+0.375x_8-0.226x_9+0.428x_{10}-0.137x_{11} \\ F2=0.382x_4+0.328x_6+0.13x_8-0.431x_9+0.069x_{10}+0.358x_{11} \\ F3=0.37x_4-0.585x_6+0.209x_8+0.349x_9+0.021x_{10}+0.482x_{11} \\ F=\left(0.363/0.835\right)\times F1+\left(0.283/0.835\right)\times F2+\left(0.19/0.835\right)\times F3 \end{cases} \tag{4}$$

Formula for lead-barium unweathered glass model:

$$\begin{cases} F1=-0.076x_2-0.545x_6+0.494x_9+0.328x_{10} \\ F2=0.748x_2-0.09x_6-0.309x_9+0.489x_{10} \\ F=\left(0.382/0.659\right)\times F1+\left(0.277/0.659\right)\times F2 \end{cases} \tag{5}$$

Formula for high potassium weathered glass model:

$$\begin{cases} F1=-0.286x_3+0.108x_4+0.088x_6-0.235x_7+0.257x_8+0.289x_{11} \\ F2=-0.015x_3+0.435x_4+0.446x_6+0.225x_7-0.183x_8+0.032x_{11} \\ F=\left(0.544/0.896\right)\times F1+\left(0.352/0.896\right)\times F2 \end{cases} \tag{6}$$

Formula for high-potassium unweathered glass model:

$$\begin{aligned} &F1=-0.124x_3-0.011x_4+0.345x_5+0.358x_6+0.426x_7+0.192x_8 \\ &F2=0.369x_3+0.497x_4-0.219x_5+0.025x_6+0.135x_7+0.316x_8 \\ &F3=0.526x_3+0.109x_4+0.207x_5+0.504x_6-0.158x_7-0.615x_8 \\ &F=\left(0.346/0.818\right)\times F1+\left(0.303/0.818\right)\times F2+\left(0.169/0.818\right)\times F3 \end{aligned} \tag{7}$$

After the weathering of lead barium, the glass is divided into three principal components, and the chemical components of principal component 1 are copper oxide and barium oxide, and their weights are 43.429%; The chemical composition of principal component 2 is calcium oxide and alumina, and its weight is 33.827%; The chemical composition of principal component 3 is lead oxide and phosphorus pentoxide, and its weight is 22.744. Lead-barium unweathered glass is divided into two principal components, the chemical composition of principal component 1 is alumina and lead oxide, and its weight is 57.976%; The chemical composition of principal component 2 is sodium oxide and barium oxide, and its weight is 42.024%. It can be seen that the content of the main chemical components of lead-barium glass after weathering is on the rise. After high potassium weathering, the glass is divided into two principal components, and the chemical components of principal component 1 are barium oxide and phosphorus pentoxide, and their weights are 60.736%; The chemical composition of the principal component 2 is potassium oxide, calcium oxide, alumina, iron oxide, and its weight is 39.624%. The high potassium unweathered glass is divided into three principal components, the chemical components of principal component 1 are magnesium oxide and iron oxide, and their weights are 42.231; the chemical components of principal component 2 are calcium oxide and copper oxide, and their weights are 37.058%; The chemical composition of principal component 3 is potassium oxide and alumina, and its weight is 20.621%. It can be seen that the content of main chemical components of high-potassium glass after weathering shows a downward trend. It is concluded that the correlation relationship between chemical composition after weathering of lead-barium glass is strong, while the correlation relationship of chemical composition after weathering of high-potassium glass is relatively weakened.

4. Conclusions

This paper uses systematic clustering to study all glass types before and after differentiation, when the number of clusters is selected as 2, it meets the two cases of only high potassium glass and lead-barium glass, and the accuracy of the classification is analyzed with real data as a reference value, and the two-class division is carried out under the condition of good accuracy, and the k-means clustering algorithm is used to obtain a better subclass division effect. The analysis of this problem can also be applied to the identification problems of other modern craft products, such as gold jewelry, etc., which can be solved by adjusting the parameters on the basis of referring to this model. In addition, this model also has a certain reference significance for the production process of handicrafts.

References

[1] Li Aiying. Prediction of particulate matter concentration in Urumqi based on RF-KMEANS-LIBSVM [J].Environmental Protection Science, 2022, 48(04):118-124.
[2] Gu Lijuan, Si Shoukui, Sun Huijing, Dong Chao. Fuzzy Clustering Analysis Applied to Artillery Precision Strike Effectiveness Assessment [J].Ordnance Automation, 2015, 34(12):1-3.
[3] Qin Risheng, Guan Hua, He Xin, Duan Ruimin. Clustering Analysis of Actual Daily Load Curve Based on Improved PSO-Kmeans Algorithm [J].Electrotechnical Techniques, 2022(11):1-6.
[4] Zhu Yifei. Application of kmeans-MOTE algorithm of clustering fusion in personal credit risk assessment [D].Shanghai Normal University, 2022.
[5] Chen Chenpeng, Zhao Xin, Bi Guihong, Chen Shilong, Xie Xu. Short-term wind speed prediction based on Kmeans-VMD-LSTM[J].Motor and Control Applications, 2021, 48(12)
[6] Li Zipeng, Wang Decai, Zhang Yu, Tan Guohui, Sun Hua, Qiao Mu, Mei Shuqi, Peng Xianwen. Principal component analysis and cluster analysis of slaughtering and body size traits of selenium black pigs[J].Genetic Breeding, 2022, 58(08)
[7] Zhang Minjuan, Hu Minmin, Bian Zhenhua. Identification of Chinese herbal ingredients and quantitative analysis of simferlin in regulating spleen and suppressing pancreas combination [J]. Jiangsu Medicine, 2017, 43(13):912-914.
[8] Wang Rongshen, Zhang Shoujun, Tao Huimin, Liu Dongmei, Li Wanzhong. Study on the identification of chemical constituents of Aconite seeds [J]. Guangzhou Chemical Industry, 2016, 44(02):93-95.
[9] Liu Mengtian. New research progress on the identification method of animal-derived ingredients in meat foods [J]. Science and Technology Perspectives, 2015, (25):157-158.
[10] Guo Zhongkun, Wang Kezhou, Ji Guoxia, Ji Chuanliang. Research progress on the composition, identification methods and pharmacological effects of Colla Corii Asini [J]. Journal of Liaoning University of Traditional Chinese Medicine, 2015, 17(04):71-74.