Subclassification Determination and Weathering Composition Prediction of Ancient Glass Artifact Compositions Based on Clustering Algorithms

Jingqi Zhang *, Zhiyuan Wu, Yi Lin

Aviation Engineering School, Air Force Engineering University, Xi'an, China, 710038

* Corresponding Author Email: 18144757088@163.com

Abstract. Ancient glass is susceptible to environmental weathering, and the content of chemical elements varies with the degree of weathering. Through the correlation analysis of weathering degree and color, pattern and composition by Chi-square test, it is found that the weathering degree of glass surface changes based on composition and color, and pattern has little correlation. The principal component analysis of the glass is based on different chemical compositions and whether the glass is weathered or not. Through the supervised learning k-means algorithm, four types are obtained: unweathered high potassium glass, weathered high potassium glass, unweathered lead barium glass, weathered lead barium glass; Further, the glass is classified into 9 subclasses based on the content of other elements to facilitate the study. The cloud model was used to predict the composition of cultural relics before weathering, and the relationship between the main chemical composition of high-potassium glass and lead-barium glass before weathering was obtained, which was convenient for subsequent research and restoration of cultural relics.

Keywords: Chi-square Test, K-means, Xgboost, Cloud Modeling.

1. Introduction

Glass has a very long history of development, and art glass products were collected and loved by ancient countries because of their unique colors, patterns and shapes[1]. Glass is also valuable evidence of the early Silk Road trade. In modern daily life, glass is the most common common material, in fact, glass has historically been one of the most expensive materials. Although China's glass products and foreign glass products have similar appearance and similar color, their chemical composition is completely different[1]. Identifying the types of glass is conducive to further research on the development history of glass in China, so as to better study the development of the ancient Silk Road[2]. Ancient glass is very susceptible to weathering under the influence of the environment, and its composition ratio changes. Therefore, it is necessary to analyze the composition of glass and explore the relationship between its weathering degree and characteristics as well as the correlation degree between components [3].

2. Correlation analysis of weathering degree and properties of glass artifacts

2.1. Correlation analysis of degree of weathering with color, composition and ornamentation

The chi-square test formula[3]:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \tag{1}$$

Where f_o represents the observation frequency, f_e indicates the expected frequency, and the smaller the Chi-square value, the more independent the two variables are [4]. The data in this paper are from references [5].

According to the characteristics of different types of glass described in literature [4] and [5], a set of original data of chemical composition content was sorted out, and correlation analysis was conducted between surface weathering and glass characteristics. Table 1 shows the analysis results.

Table. 1. Weathering and difference test of three characteristics

trait	X^2	P
color	9.432	0.307
ingredient	6.88	0.009***
figure	4.957	0.084*
Annotation： ***. **. * They represent the significance level of 1%, 5% and 10% respectively		

Among the three characteristics, only the p-value of composition and surface weathering is less than 0.05, which presents significance at the level, and the original hypothesis is rejected, and it is considered that there is a significant correlation between the degree of weathering and composition.

3. k-means clustering to get the characteristics of glass relics

3.1. The contents of chemical components of different kinds of glass relics were analyzed

Through the literature [5] can be obtained, high potassium glass and lead-barium glass material content than whether the weathering of the variability, the initial classification of the glass artifacts will first be high potassium glass and lead-barium glass to classify, Figure 1 shows the proportion of chemical composition data of glass relics in literature [5].

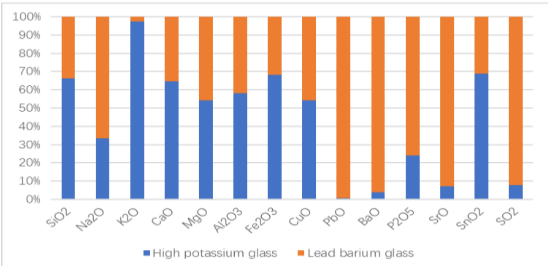

Figure. 1. Comparison of chemical content of different glass types

Comparison of the contents of defined glass types reveals that high potassium glass and lead-barium glass differ greatly in silica, NaO, K_2O, CaO, Fe_2O_3, PbO, BaO, P_2O_5, SrO, and SO_2.

3.2. Optimization of k-means algorithm

Table 2 was obtained by principal component analysis of the chemical composition of the glass:

Table. 2. Principal component analysis table

Factor Load Factor Table		
Chemical composition	factor loading factor principal component	Commonality (common factor variance)
SiO_2	-0.83	0.689
Na_2O	-0.114	0.013
K_2O	-0.676	0.457
CaO	-0.314	0.099
Al_2O_3	-0.548	0.3
Fe_2O_3	-0.421	0.177
P_2O_5	0.474	0.225
PbO	0.827	0.683
SO_2	0.43	0.185
SrO	0.738	0.545
BaO	0.731	0.534
SnO_2	-0.171	0.029

Using the k-means algorithm employing Euclidean distance:

$$d_1(x_i,c_k)=\sum_{d=1}^{D}d\left(x_{id},c_{kd}\right) \tag{2}$$

The original glass products were classified by unsupervised learning[6].Through classification, it was found that when the clustering number *k=2* was adopted by K-means algorithm, part of the lead barium glass was still classified as high potassium glass. Figure 2 shows the average chemical composition of the four glass artifacts.

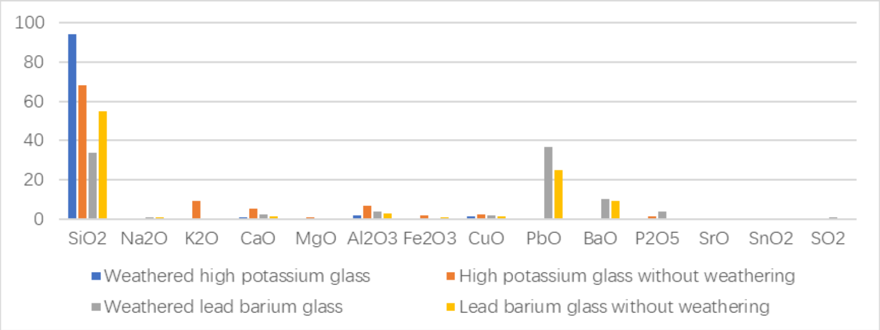

Figure. 2. Average sorted glass content

According to *Xgboost* algorithm, the importance of material content of high-potassium glass and lead-barium glass was sorted, and Table 3 was obtained.

Table. 3. High potassium and lead barium glass importance coefficient

Importance factor			
SiO_2	0.00	CuO	0.00
Na_2O	0.00	PbO	0.95
K_2O	0.00	BaO	0.00
CaO	0.00	P_2O_5	0.00
MgO	0.00	SrO	0.05
Al_2O_3	0.00	SnO_2	0.00
Fe_2O_3	0.00	SO_2	0.00

The weighted distance of the improved *k-means* algorithm[6] is:

$$d_2\left(x_i,c_k\right)=\sum_{d=1}^{D}t_d\times d\left(x_{id},c_{kd}\right) \tag{3}$$

The sum of the weighted distances from the data points to the clustering center[6] is:

$$E=\sum_{k=1}^{K}\sum_{i=1}^{N_i}\sum_{d=1}^{D}t_d\times d\left(x_{id},c_{kd}\right) \tag{4}$$

Further clustering of high potassium glass and lead-barium glass in glass artifacts based on species resulted in a more favorable classification with a silhouette coefficient[7] of: S=0.804.

3.3. Classification of chemical substances in glass artifacts under weathering and non-weathering

For high-potassium glass in the conduct of artifacts whether the weathering state is divided, the division of the results are better,by calculating the silhouette coefficient[7] is: S=0.706.Lead-barium glass artifacts are not easily classified into weathered and unweathered states, and the content of lead-barium glass artifacts before and after weathering was visualized and analyzed, Figures 3 and 4 are obtained:

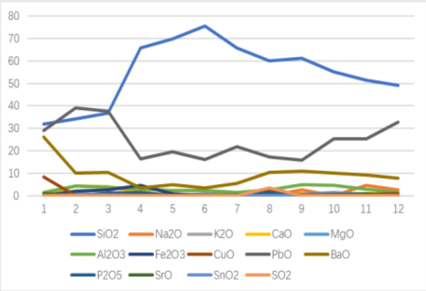

Figure. 3. Unweathered lead barium glass

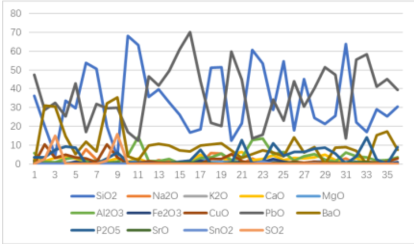

Figure. 4. Weathered lead barium glass

Obtain the coefficient of importance for the weathering of lead-barium glass in Table 4.

Table. 4. Coefficient of weathering importance of lead-barium glass

Importance factor			
SiO_2	0.1592	CuO	0.0375
Na_2O	0.0428	PbO	0.1422
K_2O	0.0024	BaO	0.0225
CaO	0.0899	P_2O_5	0.1796
MgO	0.0423	SrO	0.035
Al_2O_3	0.1485	SnO_2	0
Fe_2O_3	0.0982	SO_2	0

Lead barium glass weathering classification found that there are still a small number of lead barium glass weathering degree is not clear, guessing that the degree of weathering is too low or the surface has just begun to weather, and the chemical composition without weathering is not much difference.

According to the algorithm analysis can be obtained weathered high-potassium glass, no weathered high-potassium glass, weathered lead-barium glass, no weathered lead-barium glass a total of four categories. The data in Table 5 are the information of four types of clustering centers.

Table. 5. Cluster center value

Type	High potassium glass		Lead barium glass	
chemical component/ Cluster center	Pre-weathering	After weathering	Pre-weathering	After weathering
SiO_2	67.98±8.76	93.96±1.73	54.78±14.38	33.61±17.22
Na_2O	0.7±1.29	0±0	0.84±1.59	0.95±1.92
K_2O	9.33±3.92	0.54±0.41	0.22±0.39	0.14±0.21
CaO	5.33±3.09	0.87±0.45	1.33±1.47	2.35±1.61
MgO	1.08±0.68	0.2±0.28	0.53±0.55	0.7±0.66
Al_2O_3	6.62±2.49	1.93±0.88	3.01±1.26	3.84±3.41
Fe_2O_3	1.93±1.67	0.27±0.06	0.89±1.5	0.56±0.69
CuO	2.45±1.66	1.56±0.85	1.29±2.4	2.00±2.49
PbO	0.41±0.59	0±0	24.79±8.37	36.87±15.16
BaO	0.6±0.98	0±0	9.41±5.99	10.49±8.87
P_2O_5	1.4±1.43	0.28±0.19	0.5±0.61	4.16±4.15
SrO	0.04±0.05	0±0	0.32±0.31	0.37±0.25
SnO_2	0.2±0.68	0±0	0.07±0.16	0.06±0.23
SO_2	0.1±0.19	0±0	0.31±1.06	0.99±3.61

Based on this archaeological staff can after the artifacts are unearthed, all the glass artifacts can be identified by their chemical composition, and the general classification of the glass artifacts and whether they are weathered or not, which can further reduce the time and energy wasted on the verification of the chemical composition.

4. Classification of glass subclasses

Since there are many subtle differences in each glass cultural relic, these chemical composition differences have high research value and significance for the research of glass cultural relic. Therefore, we need to conduct subclass analysis and research on glass cultural relic based on four categories, and determine the characteristic conditions and boundary conditions of each subclass. In this paper, *K-means* cluster analysis is adopted. The subclasses are divided by input cluster parameter[7] *k*.

All glass relics sampling sites are divided into four categories: (1) weathered high-potassium glass; (2) non-weathered high-potassium glass; (3) weathered lead-barium glass; (4) non-weathered lead-barium glass[7].

The four glass classifications above were further subdivided by combining the selected oxide decision variables; based on *k-means* clustering, a classification of 2 classes was concluded to be the most appropriate based on the elbow rule[8].

4.1. Non-weathered high-potassium glass

Since the classification effect is not obvious when the two classifications of *k=2* are made, based on this the weathering-free high-potassium glass can be divided into three subclasses: weathering-free high-alumina, high-copper, high-potassium glass[8] (alumina content interval: (3.05%,7.5%), copper-oxide content interval: (0%,5.09%), and phosphorus pentoxide content interval: (0%,1.27%)); weathering-free high-alumina, high-copper, high-potassium glass (alumina content range: (10.05%,11.15%), copper oxide content range: (2.18%,2.51%), phosphorus pentoxide content range: (4.18%,4.5%)).

4.2. Weathered high-potassium glass

Based on this, the weathered high-potassium glass can be divided into two subclasses: weathered low phosphorus high-potassium glass (alumina content range: (0.81%,1.98%), copper oxide content range: (0.84%,3.24%), phosphorus pentoxide content range: (0%,0.61%)); weathered high-alumina high-potassium glass (alumina content range: (2.51%,3.5%), copper oxide content range: (0.55%,1.54%), phosphorus pentoxide content range: (0.21%,0.36%)); and weathered high-alumina high-potassium glass (alumina content range: (2.51%,3.5%), copper oxide content range: (0.55%,1.54%), phosphorus pentoxide content range: (0.21%,0.36%)).

4.3. Non-weathered lead-barium glass

Due to the better classification, the classification of weathering-free lead-barium glass only takes the classification of *k=2* based on which the weathering-free lead-barium glass can be divided into two subclasses: weathering-free ordinary lead-barium glass **(**alumina content interval: (0%,3.86%), copper oxide content interval: (0%,0.86%), and special case: 8.64%, phosphorus pentoxide content interval: (0%,1.62%)**);** weathering-free High-alumina lead-barium glass **(**alumina content range: (4.35%,5%), copper oxide content range: (0%,0.77%), phosphorus pentoxide content range: (0%,0.2%)**).**

4.4. Weathered lead-barium glass

The subcategories are as follows: weathered common lead barium (alumina content range: (0%,1.45%); copper oxide content range: (0.45%,2.69%), phosphorus pentoxide content range: (0%,0.76%)); Weathered high aluminum lead barium (alumina content range: (0.79%,2.73%); copper oxide content range: (3.33%,13.65%) and special cases: 8.64%; phosphorus pentoxide content range: (0.86%,2.74%)).

5. Prediction of the composition of glass artifacts before weathering

In order to further study and restore ancient artifacts glass needs to predict the composition of the artifacts prior to weathering[9]; a cloud model is used to make the prediction.

5.1. Introduction to Cloud Modeling

Cloud models are uncertainty transformation models between qualitative concepts and their numerical representations[10]:

a) *Expectation*: *E(x)*, is the expectation of the spatial distribution of cloud drops in the domain of the argument;

b) *Entropy*: *En*, represents the measurable granularity of a qualitative concept, the higher the entropy the more macroscopic the concept usually is;

c) *Hyperentropy*: *He*, is an uncertainty measure of entropy.

The overall characterization is noted as: $C = \left(Ex, En, He\right)$

Define the sets respectively C, π ; satisfy the mapping relation:

$$\pi = A^{Forward}\left(C\right) \tag{5}$$

Mapping conditions:

$$\begin{cases} 1) \quad \theta = \left\{ t_i \middle| N_{rnd}\left(E_n, He\right) \right\} \\ 2) \quad X = \left\{ x_i \middle| N_{rnd}\left(Ex, t_i\right), t_i \in \theta, \right\} \\ 3) \quad \pi = \left\{ \left(x_i, y_i\right) \middle| x_i \in X, y_i = e^{\frac{-\left(x_i - Ex\right)^2}{2t_i^2}} \; t_i \in \theta \right\}, i = 1, 2 \cdots N \\ N_{rnd}\left(\mu, \sigma\right) = \left(\frac{1}{\sigma \cdot \sqrt{2\pi}}\right) \exp\left(\frac{-\left(x - \mu\right)^2}{2\sigma^2}\right) \end{cases} \tag{6}$$

Expanding Cloud Computing from One Dimension to Higher Dimension[10].

5.2. Data forecasting

The dataset is first classified into four categories: no weathering, mild weathering, moderate weathering, and severe weathering, giving the corresponding subclassifications, and the subclassifications that are similar before and after weathering are used as the cloud droplet mapping set for the modeling operation,The prediction of glass chemical composition content based on cloud model is shown in Table 6 and Table 7.

Table. 6. Prediction of composition of high potassium glass before weathering (%)

Chemical composition/ type of glass	Weathered high potassium glass	A1	A2
SiO_2	After weathering	96.77	87.96
	Pre-weathering	63.44	65.44
K_2O	After weathering	0.92	1.02
	Pre-weathering	8.38	8.46
CaO	After weathering	0.21	0.46
	Pre-weathering	5.46	5.44
Fe_2O_3	After weathering	0.26	0
	Pre-weathering	3.54	3.48
CuO	After weathering	0.84	1.21
	Pre-weathering	3.24	3.54
P_2O_5	After weathering	0	0.26
	Pre-weathering	0.34	0.38
Al_2O_3	After weathering	0.81	0.46
	Pre-weathering	8.67	8.46

Table. 7. Prediction of composition of lead-barium glass before weathering (%)

Chemical composition	Weathered lead barium glass	B1	B2
SiO_2	After weathering	3.72	45.01
	Pre-weathering	65.41	65.41
CaO	After weathering	3.01	3.21
	Pre-weathering	0.64	0.89
MgO	After weathering	0	1.2
	Pre-weathering	1.04	1.24
CuO	After weathering	3.6	3.41
	Pre-weathering	0.41	0.45
PbO	After weathering	29.92	29.21
	Pre-weathering	13.21	14.01
BaO	After weathering	35.45	35.14
	Pre-weathering	6.09	5.98
SrO	After weathering	0	0.11
	Pre-weathering	0.46	0.41

Redicted chemical compositions with high coefficients of importance in some glass artifacts are given in the table, which are higher before and after weathering and predicted more accurately.

6. Conclusion

Different types of glass have their own specific characteristics. By using supervised cluster analysis, we can extract similar features to facilitate classification research and further analysis. By clustering the glasses with different weathering degrees, we can predict them separately, and find that the chemical composition ratio before weathering can be predicted well. The application of these methods can help us further distinguish different types of glass relics, understand their origin and retention time, and provide assistance in the process of restoration of cultural relics. However, the shortcoming of the current model is the small amount of data. To solve this problem, we can establish a time series of the artifacts that have been traced and infer the original composition ratio of the glass artifacts by learning the degree of weathering and the time series.

References

[1] Chen Shuyu. The origin and development of ancient glass in China [J]. Cultural Relics Identification and Appreciation,2019(04):44-45.

[2] Yin Yulong. Compositional analysis of ancient glassware by correlation prediction [J]. Contemporary Chemical Research, 2023(1):122-126.

[3] Fang Xiangzhong. The chi-square distribution and chi-square test[J]. China Statistics,2022(05):29-31.

[4] LI Qinghui, Huang Jiaozhen, Li Fei et al. Detection of chemical constituents of a batch of ancient glass samples from the Warring States Period unearthed in China [J]. Science of Cultural Relics Conservation and Archaeology,2006(02):8-13.

[5] LI Fei, LI Qinghui, GAN Fuxi et al. Proton excited X-ray fluorescence analysis of the chemical composition of a group of Chinese ancient glasses[J]. Journal of Silicates,2005(05):581-586.

[6] Song Jian, LI Yanfang, Chen Zhanfang. K-means text clustering based on weighted dense Canopy [J]. Journal of Central South University for Nationalities (Natural Science Edition),2023,42(05):636-642.

[7] Zhang Qian, LV Kunshan, Zhao Xianqing et al. Composition analysis of ancient glass products based on random forest classification model [J]. Henan Science and Technology,2023,42(09):19-22.

[8] Long Wenjia, Zhang Xiaofeng, Zhang Chain. Business process clustering method based on k-means and Elbow rule [J]. Journal of Jianghan University (Natural Science Edition), 2019,48(01):81-90.

[9] Wang Zi Dong. Mathematical model for composition analysis and identification of ancient glass products [J]. Modern Information Technology, 2019,7(14):88-93+98.

[10] Liu Ke, Zhang Ran, Cui Zhibin, et al. Research on Radial basis function neural network algorithm for cloud model optimization [J]. Journal of Henan University of Science and Technology (Natural Science Edition),2023,44(05):49-55+7.