Research on The Classification Models of Ancient Chinese Glass Based on K-means Clustering Algorithm

Peiwen Luo #, Xiaoyu Xiong *,#, Zhaochun Zhuang #

College of Electronic Engineering(College of Artificial Intelligence),South China Agricultural University, Guangzhou 510642, China

* Corresponding Author Email: ymtksy@stu.scau.edu.cn

#These authors contributed equally.

Abstract. Ancient Chinese glass is mainly potassium glass and lead-barium glass. Based on the main chemical composition of potassium glass and lead-barium glass, a K-means clustering algorithm was used to establish classification coefficients to classify glass types at one level, with the classification condition being whether the classification coefficient ε was greater than 5/16, with greater than being potassium glass and less than being lead-barium glass. a clustering algorithm with a K value of 2 was then used to classify the same type of glass into subclasses, with potassium glass being classified as whether it contained Na2O and lead-barium glass being classified as the BaO content is greater than 20% and smaller.

Keywords: K-means, Glass Classification, Subcategories.

1. Introduction

As for glass, it has been recorded very early in ancient Chinese historical data. In early human civilization, pottery, bronze, and glass were the three earliest materials used by ancient humans, and they were the product of a combination of art and craft.[1] Therefore, many scholars have also studied ancient glass from various angles, including scientific and technological archaeology of ancient glass in China. Some scholars have found that analyzing the composition of glass can help improve traditional processing techniques for glass products and further develop the application value of glass.[2] Therefore, analyzing the evolution of the chemical composition of ancient Chinese glass has important research value. In the study of the composition of ancient glass, Zeng Weisheng and others discovered that the Chinese glass system from the middle and late Warring States to the Han Dynasty was dominated by lead barium glass [3]. As early as the Spring and Autumn Period, it was discovered that lead oxide (lead oxide) and nitrate (potassium nitrate) were good combustion promoters for making glass [4]. Ma Hongjiao and others analyzed four pieces of glass unearthed from Boshan in the late Yuan and early Ming dynasties, all of which were potassium calcium glass, with a high K2O content[5], so high-potassium glass was also a common category of ancient glass; Fu Xiufeng [6] and others subdivided more than 100 ancient glass samples (mainly composed of lead barium and high potassium glass) into five system categories by multivariate statistical method: K_2O $CaO(^{1}0wt\%)SiO_2$ system, K_2O SiO_2 system, PbO BaO SiO_2 system, $PbO(^{2}5wt\%)BaO$ SiO_2 system, CaO $PbO(^{4}0wt\%)BaO$ SiO_2; With the progress of technology, Dong Junqing and others proposed a reference standard for classifying glaze flux systems based on the ratio of $Na_2O/(Na_2O+K_2O)$ by using techniques such as Raman spectroscopy to analyze cultural relics unearthed in Zibo, Shandong Province[7].

K-means++ clustering is an algorithm to select the initial value for K-means clustering. It uses a special probability distribution to select the cluster center to ensure the dispersion of the cluster. Cheng Junzhao et al. [8] cluster the external regions with the same conditions through K-means++, and use the reliability deviation degree as the evaluation index to evaluate the power supply reliability level of the distribution network. Zeng Jianbang et al. [9] determined the outlier factor of non-warning vehicles through K-means++clustering based on dynamic k-value, and advanced the time to identify abnormal units with inconsistent power battery voltage by six days. Zhang Yuqin et al. [10] combined K-means with local search strategy to initialize and partition the data set, and then separately cluster the data in each group, which has a good performance in clustering quality and processing time. Chang Hongwei [11] uses K-means to calculate the transaction flow information score value through the database for internal grading of customer credit.

It can be seen from the above that K-means++ clustering is mostly used to divide research objects according to specific conditions, and the data processing effect is good and the processing time is fast. Therefore, this paper uses K-means++ clustering to establish a glass classification model, and uses the same method to classify the glass after the initial classification, which is convenient for further research.

Based on the above literature analysis, for the classification of ancient glass samples, some scholars use statistical methods, some scholars use relevant instruments to draw conclusions, but few scholars use machine learning algorithms to build mathematical models for the classification of ancient glass samples. Moreover, K-means++ clustering is mostly used to divide research objects according to specific conditions, with good data processing effect and fast processing time. Therefore, this paper uses K-means++ clustering to establish a glass classification model, and uses the same method to classify the glass after the initial classification, providing a reference for the subsequent study of the origin and development of ancient glass.

2. Glass classification

2.1. Classification models for potassium glass and lead-barium glass

Potassium glass and lead-barium glass were used as the two main classes of K-means clustering, which means K=2, and the shortest distance was calculated through continuous iterations to optimise the classification model to the greatest extent.

K_2O, PbO and BaO were selected as the main components, and by combining them, four groups of criteria were obtained, as shown in the Table 1.

Table 1. Clustering algorithm criteria

Chemical composition	K_2O	PbO	BaO
Group 1	√	\	\
Group 2	\	√	√
Group 3	√	√	\
Group 4	√	√	√

And for each criterion the clustering algorithm steps are shown below.

1) A randomly selected sample from the chemical composition dataset as the initial clustering centre c1.;

2) Calculate the shortest distance S(x) between each sample and the currently existing cluster centre;

$$S(x)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+...+(x_n-y_n)^2} \tag{1}$$

3) Calculate the probability of each sample point being selected as the next cluster centre h

$$h=\frac{S(x)^2}{\sum_{x\in X}S(x)} \tag{2}$$

4) Selection of 2 clustering centres according to the roulette wheel method;

5) Repeat the above steps until a total of 2 clusters have been selected

The error rates for the categorised combinations were calculated and are shown in the Table 2.

Table 2. Classification error rates for different group

Chemical composition	Group 1	Group 2	Group 3	Group 4
Error rate (%)	11.94	10.45	22.38	23.88

The combination with the lowest error rate, 2: PbO +BaO, was chosen as classification condition one.

Based on the fact that ancient glass products mainly consist of lead-silicate glass and potassium-silicate glass with Chinese characteristics [2,12], an attempt was made to use K_2O, $CaO+K_2O$ and $CaO+K_2O+Na_2O$ as classification condition two, respectively, for cluster analysis and to draw scatter plots, as shown in Figure 1.

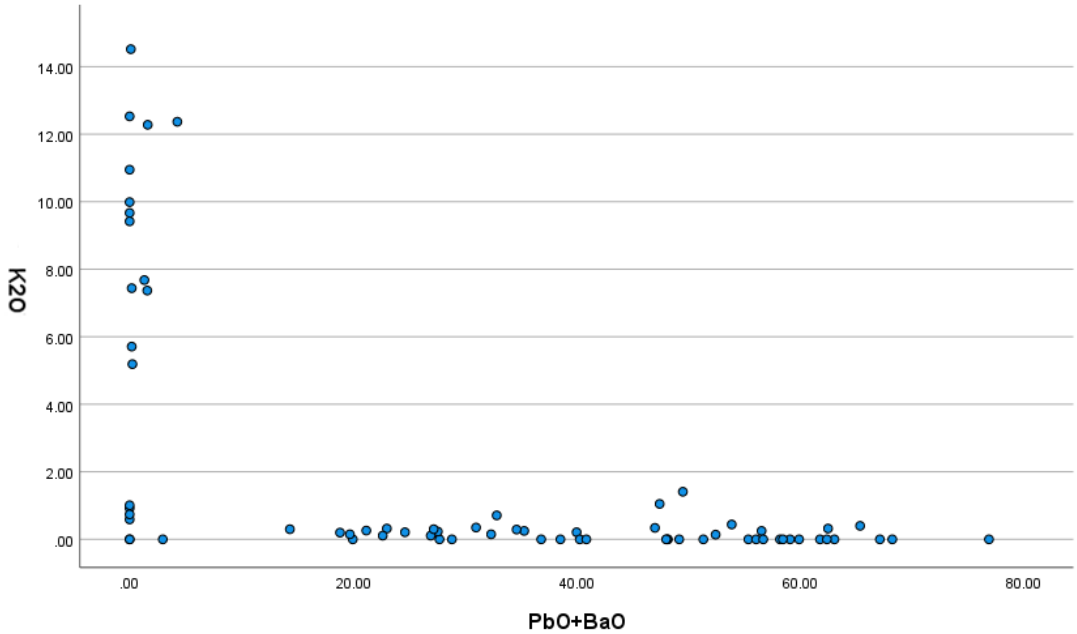

Figure 1. Scatter plot of the K_2O standard

As the Figure 1 yields, when potassium oxide is used as a criterion, the lead-barium glass is best classified in combination with classification condition one, as is the potassium glass.

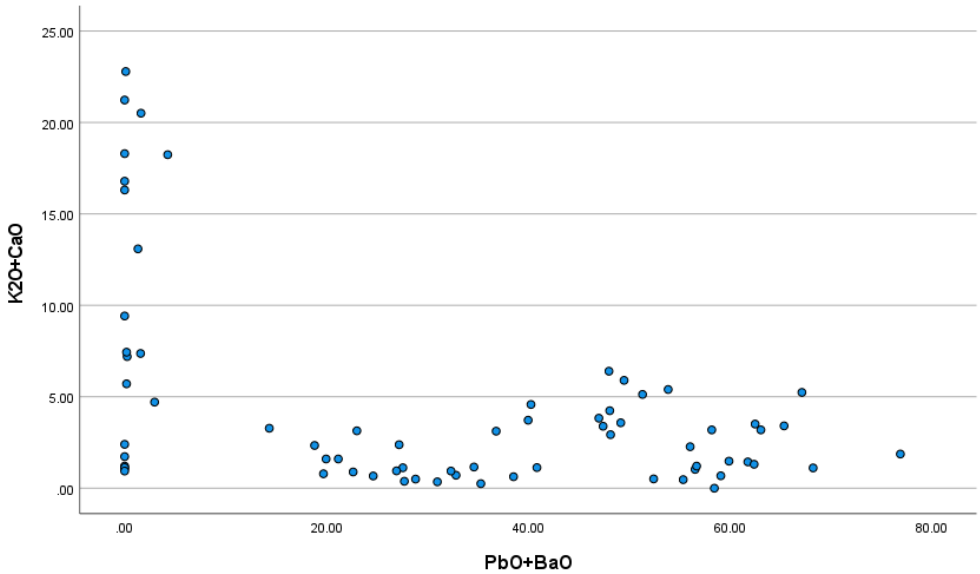

Figure 2. Scatter plot of the K_2O+CaO standard

K_2O+CaO as a classification condition is less effective for lead-barium glass. As shown in Figure 2.

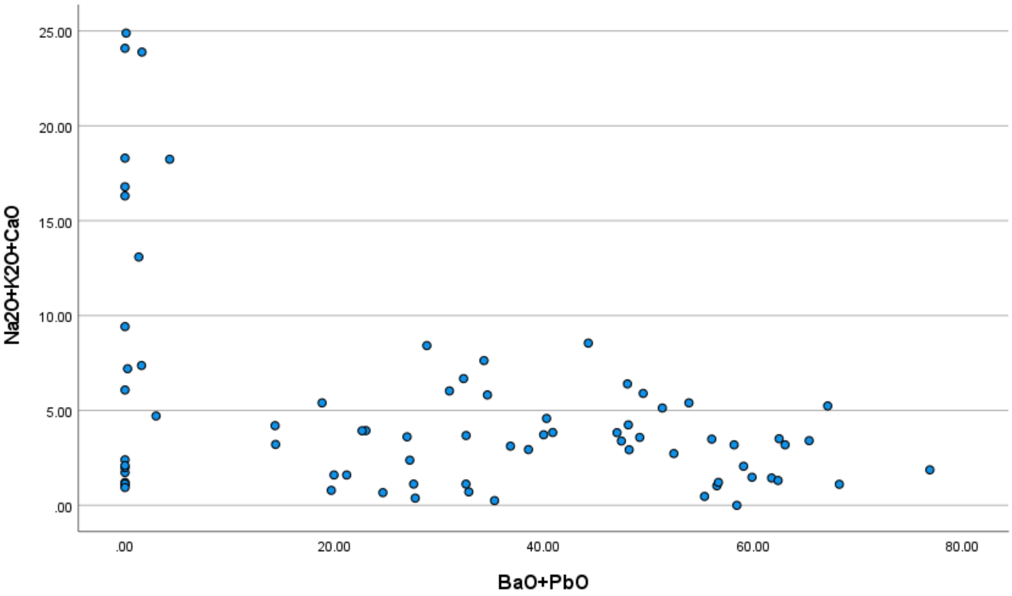

Figure 3. Scatter plot of the $CaO+K_2O+Na_2O$ standard

$CaO+K_2O+Na_2O$ is also less effective as a classification condition for lead-barium glass. As shown in Figure 3.

2.2. Glass subcategories

The classification conditions were confirmed based on the chemical composition data and the classification conditions, and spatial scatter plots of potassium glass and lead-barium glass were plotted based on the classification conditions, as shown in Figures 4 - 5.

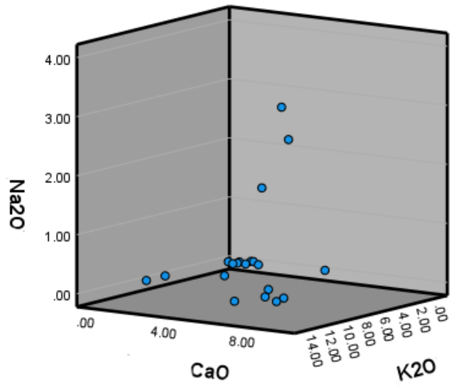

Figure 4. Spatial scatter diagram of potassium glass

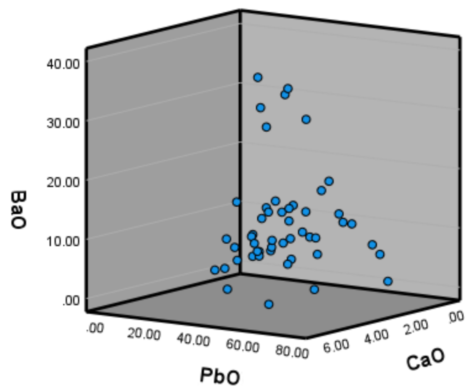

Figure 5. Spatial scatter diagram of lead-barium glass

3. Results

3.1. Glass Class I classification criteria

Notation of glass type classification coefficient $\varepsilon = \frac{BaO + PbO}{K_2O}$ (Content in glass)

When $\varepsilon > \frac{5}{16}$,the glass is potassium glass;

When $\varepsilon < \frac{5}{16}$,the glass is barium lead glass.

3.2. Classification criteria for glass subclasses

The clustering centres were calculated as shown in Table 3.

Table 3. Subclass analysis of glass

Glass	Subcategory	
	1	2
Potassium	contain Na_2O	without Na_2O
Lead-Barium	BaO content less than 20%	BaO content more than 20%

4. Conclusion

In this paper, the influencing factors of ancient glass are studied, and the main chemical composition of ancient glass is modelled by using the K-means clustering algorithm. The dispersion of the scatter diagram produced by the model is small and fits the actual situation well enough to distinguish potassium glass from lead-barium glass. The model has a unit change of 1/16 of the data, which yields a strong model sensitivity.

References

[1] Di Jin, Zhang Wenli, Du Ying Pursuing the "Colorful Light" spanning thousands of years [N] Science and Technology Daily, 2021-09-03 (008).

[2] Liu Yanyan. A Study on the Modeling Language of Boshan Coloured Glaze from the Perspective of Traditional Crafts [D]. Shandong Academy of Arts, 2022.

[3] Zeng Weisheng. Looking at Ancient Chinese and Western Glassware and Its Cultural Differences from the Unearthed Glass of the Han Dynasty [J]. Journal of Gemology and Gemology (Chinese and English), 2022-24 (06): 84-92. DOI: 10.15964/j.cnki.027jgg.2022.06.008

[4] Gan Fuxi. The origin and development of ancient Chinese glass [J]. Nature Journal, 2006 (04): 187-193+184.

[5] Hongjiao Ma, Julian Henderson, Jianfeng Cui, Kunlong Chen, Glassmaking of the Qing Dynasty: A Review, New Data, and New Insights,

Advances in Archaeomaterials, Volume 1, Issue 1,2020, Pages 27-35, ISSN 2667-1360, https://doi.org/10.1016/j.aia.2020.11.001.

[6] Fu Xiufeng, Gan Fuxi. Research on the composition of a batch of ancient glass in southern and southwestern China based on multivariate statistical analysis [J]. Cultural Relics Protection and Archaeological Science, 2006, (04): 6-13.

[7] Dong Junqing. Restudy on the origin and development of ancient Chinese glass technology [D]. Shanghai: University of the Chinese Academy of Sciences, 2020.

[8] Cheng Junzhao, Xu Min, Feng Lei, Liu Hongyin, Li Chen, Guo Zuogang. Evaluation of power supply reliability level of distribution network based on GRA-Kmeans++ algorithm. [J/OL]. China Southern Power Grid Technology:1-9[2023-03-06].

[9] Zeng Jianbang, Zhang Yueya, Zhang Zhuang, Shan Fengwu, Shen Zuying, Liu Xing. Fault identification method of electric vehicle power battery voltage inconsistency based on dynamic k-value K-means++clustering [J]. China Science: Technical Science, 2023,53 (01): 28-40

[10] Zhang Yuqin, Liang Li, Zhang Jianliang, Feng Xiangdong. Big data clustering method based on improved K-means++and DBSCAN [J]. Foreign electronic measurement technology, 2022,41(09):40-46.

[11] Chang Hongwei. A k-means based internal rating model for customer credit [J]. Information Systems Engineering, 2023 (03): 32-36.

[12] Lu Jiajia. A Classification Model for Ancient Glass Based on Integrated Feature Selection and Random Forest [J/OL]. Journal of Silicate: 1-6 [2023-03-28]. DOI: 10.14062/j.issn.0454-5648.20220790.