Research On the Classification Patterns of High-Potassium Glass and Lead-Barium Glass Based On K-Means Clustering Analysis

Shirong Guo *

Leicester International Institute, Dalian University of Technology, Dalian, China, 116081

* Corresponding Author Email: sg645@student.le.ac.uk

Abstract. This study employs the K-means clustering analysis algorithm to conduct a comprehensive investigation into the classification patterns of high-potassium glass and lead-barium glass. Initially, by selecting two cluster centers for analysis, significant differences in composition between high-potassium and lead-barium glass are revealed through the merging and simple processing of data samples from forms one and two. Subsequently, extensive analysis and statistics with multiple cluster centers reveal a suitable cluster ratio of 1:3, leading to the determination of four cluster centers for further analysis, resulting in the identification of four subcategories. Detailed analysis indicates one as a high-potassium type and three as lead-barium types. Through algorithmic computations and validation using the silhouette coefficient model, we summarize the composition characteristics of each category, demonstrating the rationality of the classification. Furthermore, fitting classification is applied to eight cultural relic samples, and the validity and sensitivity of the classification results are further verified through the silhouette coefficient model, ensuring the reliability of the classification.

Keywords: K-means clustering analysis, High-potassium glass and lead-barium glass, Cultural relic sample classification.

1. Introduction

Glass, as a valuable material in the evidence of ancient cultural exchanges and trade between the East and West, possesses unique artistic value and a diverse range of types. The primary chemical component of glass is silica, and during refining, fluxing agents and stabilizers need to be added. The different additives result in variations in their main chemical composition. For instance, glass from the Chu culture is primarily lead-barium glass, where the addition of lead ore as a fluxing agent during the firing process leads to higher content of lead oxide and barium oxide. On the other hand, potassium glass, prevalent in regions such as Lingnan in China, Southeast Asia, and India, is made by firing substances with high potassium content, like wood ash.

Glass is highly susceptible to weathering in burial environments. Due to chemical reactions between internal elements and the external environment, the proportions of glass components can change. Various external factors can collectively impact the chemical composition, color, and patterns of artifacts, making the accurate classification of glass artifacts challenging. This necessitates a detailed analysis of these influencing factors. This series of papers focuses on the study of the compositional analysis and classification of ancient glass objects, an important topic in the fields of archaeology and conservation. By adopting different methods and techniques, researchers have worked to improve the accuracy of identification and classification of ancient glass.

This series of texts brings together in-depth studies on the compositional analysis and classification of ancient glass objects, providing valuable information for the fields of archaeology and conservation. The following is a comprehensive review of these 10 literatures:

Firstly, in a study published in Practice and Understanding of Mathematics, Jiang Yongsheng et al [1]. discuss in depth the classification of glass based on the method by introducing a probabilistic judgement function. This method provides a new way of thinking about the classification of glass products and offers a powerful tool for a deeper understanding of the differences in glass composition. Yang Qi's research in Digital Technology and Applications, on the other hand, focuses on the

application of machine learning techniques in the analysis and identification of the composition of ancient glass products [2]. By leveraging machine learning, the researchers expect to improve the accurate classification of ancient glass and explore new research possibilities. Dong Han et al.'s study, published in the Journal of Xianyang Normal College, is dedicated to the classification prediction of ancient glass compositions [3]. Through an in-depth study of ancient glass compositions, they attempted to predict the classification of glass in advance, providing a more accurate guide for the classification of cultural relics. In Modern Information Technology, Hui Jiaojiao and Cao Hongbao collaborated to discuss the identification of glass artefacts based on machine learning classification algorithms [4-5]. The application of this method is expected to improve the efficiency of rapid identification of glass artefacts and provide support for heritage conservation work. In addition, Du Shiqiang et al. conducted a study on the composition analysis and identification of excavated glass artefacts in the Journal of Northwest University for Nationalities (Natural Science Edition) [6], which provided important clues for glass artefacts in archaeological excavations. Tanyu Deng et al. conducted a study on the composition analysis and classification model of glass artifacts using the CLR method in Mathematical Modelling and its Applications, which provided a new path to improve the accuracy of classification. The study by Peiying Ma et al, published in the same journal, Mathematical Modelling and its Applications [7], focuses on the analysis and classification of ancient glass artefacts based on compositional data, which provides more detailed information for classification through in-depth analyses of the correlation of different components. In Chemical Management, Shuifan Zhang et al. conducted a study on the chemical composition analysis and correlation of ancient glass products [8], emphasising the importance of a comprehensive understanding of ancient glass for heritage conservation. Wang Zhihao et al. proposed a machine learning-based prediction and subclassification method for the original composition of weathered silicate glass in Journal of Silicates, which provides a new direction for the classification study of specific types of glass artefacts. Finally, Yulong Yin provided an in-depth analysis of ancient glass artefact compositions through correlation prediction in Contemporary Chemical Research, emphasising the role of attention to correlation in classification studies [9-10].

Together, these studies reveal new advances in the field of compositional analysis and classification of ancient glassware, providing useful references and insights for conservation and archaeological studies. The combined application of these methods is expected to.

This paper focuses on glass artifacts of high-potassium and lead-barium types, analyzing the statistical patterns of surface weathering on different types of artifacts and predicting the chemical composition before weathering. By selecting appropriate mathematical models, the study aims to identify classification patterns and subcategories for the two types of glass artifacts. Finally, at the application stage, the team's established classification rules are employed to classify unknown artifacts, addressing practical issues in the classification of glass artifacts.

2. Modelling and solving glass classification models

2.1. Establishment of Cluster Models under the K-means Algorithm

We integrated the (material) type of the artefacts and artefact numbers, filled all the vacancies with the natural number 0, and deleted invalid data where the content of each component lay outside the range of 85% to 105%. A sample was obtained that was ready for data analysis. A total of 69 samples with different artefact sampling points but the same artefact number were included.

And each sample has 14 dimensions of the attribute (is the content of different oxides), then these 69 artefacts of different oxides of the composition of the content of the data, can be seen as a vector from the origin of 14 dimensions of the vector pointing to the point, now we use kmeans cluster analysis algorithm, the data samples for processing. The goal of the kmeans algorithm is to classify the 69 objects into the specified k class clusters based on the degree of similarity of the main components between the objects, where each object belongs to, and only belongs to, the centre of the class cluster closest to the object (point).

Suppose there are two cluster centers $\{C_1, C_2\}$, then calculate the Euclidean distance from each object to each cluster center using the following formula:

$$\mathrm{dis}(X_i, C_j) = \sqrt{\sum_{t=1}^{m} (X_{it} - C_{jt})^2} \tag{1}$$

In the above formula, X_i denotes the ith object $1 \le i \le 69$, C_j denotes the jth cluster centre with $1 < j < 2$, X_{it} denotes the tth attribute of the ith object with $1 \le t \le 14$, and C_{jt}t denotes the tth attribute of the ith cluster centre. The formula for calculating the class cluster centre is as follows:

$$C_t = \frac{\sum_{X_i \in S_1} X_i}{|S_l|} \tag{2}$$

where C_1 denotes the centre of the 1st cluster, $1 \le 1 \le k$, $|S_1|$ denotes the number of objects in the 1st class cluster, and X_i denotes the ith object in the centre of the 1st class cluster, $1 \le i \le |S_1|$. It follows that the class cluster centre is actually the point represented by the average of the individual attributes of that point.

In the pursuit of classifying the major classes of glass, namely high-potassium and lead-barium, the k-means cluster analysis algorithm was employed. However, the challenge arose concerning the unknown number of subclasses required for each of these major classes. To address this, an approach involving the assumption of 4, 8, and 12 class cluster centers for the classification of all 69 elements was adopted. The well-crafted script of the k-means algorithm was executed after importing the data into MATLAB.

Given the predominant components of lead-barium glass, as elucidated in the literature, mainly consisting of lead oxide and barium oxide, and recognizing that the primary component of high-potassium glass is potassium oxide, a comprehensive analysis of the results from the three algorithms was conducted. The intriguing revelation emerged that the ratio of the number of subclasses pertaining to artifacts of high-potassium glass to the number of artifacts belonging to lead-barium glass consistently stood at 1:3. This finding suggested the optimal configuration of establishing a subclass under high-potassium glass and three subclasses under lead-barium glass.

Conducting sensitivity analysis necessitated awareness of the sample type (i). Employing the Euclidean distance formula, the average distance to the center of a class of clusters (a(i)) was calculated. The smaller the value of a(i), indicative of the proximity to the center of the clusters, the more conducive it was for subdivision into this class of clusters. a(i) was denoted as the degree of dissimilarity within the class of clusters.

This nuanced approach not only provided a systematic means of classifying high-potassium and lead-barium glass but also laid the foundation for sensitivity analysis, offering insights into the dissimilarity within class clusters. The findings contribute to the refinement of classification models, enhancing the robustness of glass classification methodologies.

Then define an inter-cluster dissimilarity b(i). We use b(i)1, b(i)2, b(i)3, b(i)4 to denote the distances of sample i to the cluster centres of other class clusters, then the inter-cluster dissimilarity is the minimum of the distances of the samples to the cluster centres of all the other clusters, i.e., b(i) = min{ b(i)1, b(i)2, b(i)3 .....b(i)k}. a smaller value of b(i) means that it is less likely to belong to other clusters.

Then the profile coefficient s(i) of sample i is related to the within-class cluster dissimilarity a(i) and between-cluster dissimilarity b(i) of sample i as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} s(i) = \begin{cases} 1 - \dfrac{a(i)}{b(i)}, & a(i) < b(i) \\ 0, & a(i) = b(i) \\ \dfrac{b(i)}{a(i)} - 1, & a(i) > b(i) \end{cases} \tag{3}$$

Then contour coefficient of all data = sum of contour coefficients of all samples / number of samples.

The size of s(i) lies in the closed interval [-1,1]. For s(i) of the whole sample the closer s(i) is to 1, the more reasonable the cluster analysis is, and the closer it is to -1 the less reasonable the cluster analysis is.

2.2. Solving the k-means algorithm.

Cluster analysis was carried out using kmeans algorithm in the software spss pro, which requires two cluster centres and the results obtained by the system are as follows table 1:

Table 1 Means of the components of the two clustering centres

type of clustering	SiO2	K2O	CaO	MgO	Al2O3	PbO	BaO
1	27.490	0.165	2.542	0.543	2.698	41.290	12.216
2	68.310	3.557	2.387	0.817	5.069	8.599	3.521

The insightful examination of the clustering results revealed a significant disparity in the potassium oxide content between cluster center 2 and cluster center 1, indicating a distinguishing factor between high-potassium glass and lead-barium glass. Specifically, cluster center 2 exhibited a markedly higher potassium oxide content, while cluster center 1 demonstrated elevated levels of lead oxide and barium oxide. This pivotal observation led to the unequivocal identification of cluster 1 as representing lead-barium glass and cluster 2 as indicative of high-potassium glass.

To validate the accuracy of this cluster analysis, a subsequent sensitivity test of the k-means algorithm for the two cluster centers was conducted. Leveraging an online program, the profile coefficient obtained was an impressive 0.97, affirming the high precision of the cluster analysis.

Building on this success, the investigation extended to setting the number of clustering centers to 4, 8, and 12 using the SPSS Pro software. This endeavor sought to derive the average components of the clustering centers, refining the understanding of the distinctive characteristics of high-potassium and lead-barium glass. Aligning with prior knowledge that high-potassium glass exhibits a notably higher potassium oxide content, while lead-barium glass is distinguished by increased levels of lead oxide and barium oxide, the classification process gained further clarity.

A key revelation emerged as the ratio of the number of cluster centers belonging to high-potassium glass to the number of cluster centers belonging to lead-barium glass consistently maintained a 1:3 proportion. This finding informed the decision not to subdivide high-potassium glass into subclasses but to establish three subclasses for lead-barium glass. The nuanced results of this sub-classification endeavor are meticulously detailed in Table 2.

This intricate analysis not only solidifies the differentiation between high-potassium and lead-barium glass but also underscores the robustness of the classification model. The strategic utilization of sensitivity tests and refined subclassification methodologies contributes to the overarching goal of enhancing the accuracy and reliability of glass classification processes.

Table 2 Subclassification of high-potassium and lead-barium glasses

Central value	SiO2	Na2O	K2O	CaO	MgO	Al2O3	PbO	BaO	P2O5
High potassium	76.584	0.439	6.073	3.676	0.796	4.914	1.111	0.565	0.981
Lead barium a	26.537	0.219	0.180	2.880	0.734	2.944	47.338	8.007	4.901
Lead barium b	57.601	2.079	0.162	1.229	0.725	5.186	20.637	7.338	0.937
Lead barium c	21.593	1E-16	0.189	1.871	0.101	2.009	26.344	27.708	5.084

A sensitivity test was conducted with a result of 0.92, indicating a high degree of feasibility for this classification.

We classified the three types of lead-barium glass as lead-barium glass a, lead-barium glass b, and lead-barium glass c with respect to their characteristic values (a component that is much more or less than the other two types of lead-barium glass). Lead-barium glass a is characterised by a higher

content of calcium oxide and a higher content of iron oxide. Lead-barium glass b is characterised by a higher content of sodium oxide, a higher content of aluminium oxide, a lower content of phosphorus pentoxide and a lower content of strontium oxide. Lead-barium glass c is characterised by a lower content of magnesium oxide, a higher content of copper oxide, a higher content of tin oxide and a higher content of sulphur dioxide.

Glass type identification

Following the four classifications, the kmeans algorithm was used to classify the eight unknown types of artefacts, and after importing the data, we obtained the following predictions, which include a sensitivity analysis using the profile coefficients (table 3).

Table 3 Identification results

Heritage Number	A1	A2	A3	A4	A5	A6	A7	A8
Prediction Type	high potassium	lead and bariuma	lead and bariuma	lead and bariuma	lead and bariumb	high potassium	high potassium	lead and bariumc
Probability of Predicted Outcome	0.88	0.8	0.86	0.86	0.88	0.99	0.92	0.97

3. Conclusions

The research explores the classification patterns of high-potassium glass and lead-barium glass using the K-means clustering analysis algorithm. Initially, the study selects two cluster centers, revealing significant compositional differences through the analysis of data samples. Further analysis with multiple cluster centers suggests a suitable ratio, leading to the identification of four cluster centers and four subcategories. Detailed analysis differentiates one high-potassium type and three lead-barium types. Algorithmic computations and validation using the silhouette coefficient model confirm the rationality of the classification. Applying fitting classification to eight cultural relic samples, the study verifies the validity and sensitivity of the results.

The modeling and solving section discuss the establishment of cluster models under the K-means algorithm. The formula for Euclidean distance calculation is presented, and the cluster center calculation formula is detailed. Sensitivity analysis using silhouette coefficients ensures the reliability of the classification. The k-means algorithm is applied to classify glass artifacts into subclasses, with detailed analysis revealing characteristic values for lead-barium glass subclasses.

In glass type identification, the k-means algorithm is used to classify unknown artifacts, providing predictions with a sensitivity analysis using profile coefficients. The results demonstrate the effectiveness of the classification methodology.

Overall, the research contributes valuable insights into classifying high-potassium and lead-barium glass artifacts, providing a framework for understanding their composition and aiding in the categorization of cultural relics.

References

[1] Y.S. Jiang, J.F. Zhang, T.T. Xue et al. Exploration of classification based on probabilistic judgement function - an example of glass classification [J]. Practice and Understanding of Mathematics, 2023, 53 (10): 116-125.

[2] Yang Q. Composition analysis and identification of ancient glass products based on machine learning [J]. Digital Technology and Applications, 2023, 41 (07): 111-114.

[3] Dong Han,Zou Minghua,Li Lu et al. A study on classification prediction of ancient glass compositions [J]. Journal of Xianyang Normal University, 2023, 38 (04): 31-37.

[4] Hui Jiaojiao, Cao Hongbao. Identification of glass artefacts based on machine learning classification algorithm [J]. Modern Information Technology, 2023, 7 (13): 101-104.

[5] Du SQ, Le JW, Zhou HL, et al. Compositional analysis and identification study of excavated glassware [J]. Journal of Northwest University for Nationalities (Natural Science Edition), 2023, 44 (02): 8-16.

[6] T.Y. Deng, Y.H. Qiu, J.H. Chi. CLR-based compositional analysis and classification model for glass artefacts [J]. Mathematical Modelling and its Applications, 2023, 12 (02): 41-51.

[7] Ma Peiying,Han Yanlai,Li Delan et al. Analysis and classification of ancient glassware based on compositional data [J]. Mathematical Modelling and its Applications, 2023, 12 (02): 63-73.

[8] ZHANG Shuafan,SUN Ruijuan,REN Hanzhuo. Chemical composition analysis and correlation study of ancient glass products [J]. Chemical Management, 2023, (08): 138-141+164.

[9] Wang Zhihao,Christina Zhao,Li Zhiqun et al. Machine learning-based prediction and subclassification method for the original composition of weathered silicate glass [J]. Journal of Silicates, 2023, 51 (02): 416-426.

[10] Yin YL. Compositional analysis of ancient glassware by correlation prediction [J]. Contemporary Chemical Research, 2023, (01): 122-126.