

Research on glass classification and identification based on heritage samples

Zhiyuan Liu *

Institute of electric power, South China University of Technology, Guangzhou, Guangdong, 510640

* Corresponding Author Email: 317645676@qq.com

Abstract. The study of ancient glass has been intensified in archaeological circles today, but its susceptibility to weathering can affect archaeologists' ability to classify it. In order to classify glass artefacts, this paper establishes a 'subclass classification model based on the sum of the absolute values of relative changes' and uses the K-means clustering algorithm to classify the subclasses. In order to identify different classes of glass objects, the chemical composition of the glass is extracted from the centroids of the different subclasses, and the Euclidean distance between the class centres is solved for, and the class represented by the class centre with the lowest Euclidean distance is taken as the class corresponding to the unknown object.

Keywords: K-means clustering algorithm, Euclidean distance, glass artefacts.

1. Introduction

With the excavation of ancient Chinese glass artefacts, the archaeological community has continued to study ancient Chinese glass in depth [1]. Quartz sand, the main raw material for glass, has a main chemical composition of silica [2]. However, as pure quartz sand has a high melting point, it is not easily refined and requires the addition of fluxes to lower the melting temperature, such as grass wood ash, lead ore, etc [3]. Limestone is also added as a stabiliser, which changes to calcium oxide when it is calcined. In addition, fluxes can also affect the main chemical composition of the glass [4]. For example, the high content of lead oxide (PbO) and barium oxide (BaO) in lead-barium glass is due to the use of lead ore as a flux; potassium glass is made by firing a substance with a high potassium content, such as grass wood ash, as a flux. However, ancient glass is highly susceptible to weathering by the environment in which it is buried, which can lead to changes in its various chemical compositions and thus affect the archaeologist's judgement of its type [5-6].

In this paper, in order to statistically and analyse the classification patterns of high-potassium and lead-barium glass, the appropriate chemical composition was selected to sub-classify each category [7], giving detailed classification methods and results, and analysing their reasonableness and sensitivity [8]. Then, in order to identify the glass composition of unknown samples, this paper draws on the classification model to classify samples of unknown composition [9-10].

2. Glass classification model building and solving

In order to study the classification patterns of high potassium glass and lead-barium glass, to select the appropriate chemical composition for each category and to analyse the reasonableness and sensitivity of the results, a 'subclass classification model based on heritage samples' can be developed.

2.1. Model building

Step1: The raw data are first pre-processed. The mean value of the j th chemical component before weathering is calculated as A_j , the mean value of the j th chemical component after weathering is calculated as B_j , and the relative rate of change of the mean value of each chemical component after weathering compared to that before weathering is defined as the "degree of weathering change" C_j . The relative rate of change of each chemical composition of high potassium glass before and after weathering t_1 , the relative rate of change of each chemical composition of lead-barium glass before and after weathering t_2 , the relative rate of change of high potassium glass after weathering compared

with that of lead-barium glass before weathering t_3 , and the relative rate of change of lead-barium glass after weathering compared with that of high potassium glass before weathering t_4 , i.e.

$$t_{i,j} = \frac{(B_{kj}-A_{kj})}{A_{kj}} \times 100\% \quad (1)$$

where $i=1, 2, 3, 4$, denotes four relative rates of change; $j=1, 2, \dots, 14$, denotes 14 chemical compositions; $k=1, 2$, where 1 denotes high potassium glass and 2 denotes lead-barium glass, the specific formula and the values of i, j and k are determined by the above definitions. The higher the absolute value of each relative rate of change, the more significant the change in the chemical composition during the weathering process, and therefore the more representative it is in the classification of glass types. Combining the four parameters above, the 'significance' of the chemical element is defined as t_0 .

$$t_0 = \sum_{i=1}^4 |t_i| \quad (2)$$

By calculating the four parameters above and comparing the values of t_0 , the classification of high potassium glass and lead-barium glass can be summarised.

The classification rules for high potassium glass and lead-barium glass can be summarised.

Step2: In order to classify the glass into subclasses, it is necessary to select a representative chemical composition. By calculating the value of t_0 , the "significance" of each chemical element can be compared. The 8 chemical components with the highest significance were selected from the 14 chemical components, and 57 of the 67 valid artefact samples were selected (the remaining 10 valid artefact samples were used to test for reasonableness). Using the K-means clustering algorithm, two types of glass from each of the 57 selected samples were

Multidimensional clustering analysis was carried out for each of the 57 samples, based on the following principles.

(1) Initialisation. (1) Initialisation. k (the number of categories) was chosen randomly as the centre of the initial clustering.

(2) Clustering of the samples. For the cluster centres selected during initialisation, the Euclidean distance of all samples to each centre is calculated and each sample is clustered.

Euclidean distances are calculated for all samples to each centre, and each sample is clustered into the class of its nearest centre to form a cluster.

(3) Recalculate the class centre (centre of mass) after clustering, i.e. the mean value of the samples in each class, as the new class centre.

(4) Steps (2) and (3) are then re-executed until the clustering results no longer change.

For each glass, the number of subclasses was set to p , and the values of p were set to 3, 4, 5 and 6. Considering the two glasses together, there were a total of 16 subclasses for both, which were two-by-two combinations of four values: 3 classes for high potassium glass and 3 classes for lead-barium glass, 3 classes for high potassium glass and 4 classes for lead-barium glass. The results of the K-means clustering analysis were used to compare the correct rates of the 16 combinations and the specific values of the respective Euclidean distances, and the best solution was taken as the result of the subclass classification, the reasonableness of which was expressed in terms of the correct rate.

Step3: Finally, the sensitivity of the model was measured. After determining the combination of the subclasses, we modified the number of chemical components from the original 8 to 7 and 9, solved the model again, compared the resulting clustering results, and analysed the sensitivity.

2.2. Solving the model

(1). In order to analyse the classification pattern of the glass types, it is first necessary to use Excel to calculate the four relative rates of change and the 'significance' of the chemical elements based on the definitions in the model building section.

The four relative rates of change and the "significance" of the chemical elements were calculated as shown in the following table (only some of the data are shown)

Table 1. Chemical elements values

	SiO ₂	Na ₂ O	K ₂ O	CaO	MgO	Al ₂ O ₃	Fe ₂ O ₃
t ₁	38.5%	-100.0%	-94.1%	-83.4%	-82.1%	-71.9%	-86.4%
t ₂	-53.3%	-85.5%	-51.9%	107.7%	6.7%	-32.6%	-21.3%
t ₃	71.9%	-100.0%	148.4%	-34.1%	-69.3%	-56.7%	-64.0%
t ₄	-62.4%	-67.8%	-98.9%	-47.7%	-37.8%	-56.3%	-70.2%
t ₀	226.1%	353.2%	393.3%	272.9%	195.9%	217.5%	241.9%

From the table 1, it can be seen that the content of lead oxide (PbO) varies significantly between the different types of glass before and after weathering, followed by barium oxide (BaO), sulphur dioxide (SO₂) and strontium oxide (SrO); the least significant variation is in copper oxide (CuO); the rest of the chemical components are within normal limits.

We can thus summarise the classification of the two types of glass: for the classification of glass types, the most representative of the 14 chemical components is lead oxide (PbO), followed by barium oxide (BaO), sulphur dioxide (SO₂) and strontium oxide (SrO), and the least representative chemical component is copper oxide (CuO). Therefore, when classifying glass types, the focus is on the magnitude and variation of lead oxide (PbO), barium oxide (BaO), sulphur dioxide (SO₂) and strontium oxide (SrO) as criteria for classification, with the remaining chemical elements used as references.

(2). In order to classify glass types into subclasses, a representative selection of chemical elements is required. Based on the values of t₀ calculated in the table above

Based on the values of t₀ calculated in the table above, the 14 chemical components are listed in descending order of t₀ as follows

(only the top 7 are shown):

Table 2. Chemical components listed in descending order

Rank	1	2	3	4	5	6	7
Chemical components	PbO	BaO	SO ₂	SrO	P ₂ O ₅	K ₂ O	Na ₂ O

From the full table 2, the top 8 chemical constituents in terms of 'significance' were selected for the K-means clustering analysis: PbO, BaO, SO₂, SrO, P₂O₅, K₂O, Na₂O and SnO₂.

Next, 57 samples were selected from all the samples (the total number of valid samples was 67, including 16 high potassium glass classes and 41 lead-barium glass classes) and the K-means clustering analysis was carried out in Matlab based on the values of the above eight chemical classes detected in the two classes. The two separate glass classification criteria were then combined two by two, giving a total of 16 results for the classification scheme.

The reasonableness test was then performed. For each result, the remaining 10 sample data (where samples 1 and 2 are high potassium glass classes, and samples 3, 4, 5, 6, 7, 8, 9 and 10 are lead-barium glass classes) are substituted into the model, and the Euclidean distance between the sample data and each clustering centroid is calculated, and when the Euclidean distance takes the minimum value, the sample data is classified into the class represented by that centroid, thus obtaining the glass class of that sample Validation results. The validation results were compared with the known true categories and the correctness of each validation was counted for each of the 16 combinations.

On the one hand, this demonstrates that the K-means clustering analysis is reasonable, and on the other hand, it demonstrates that the best subcategorisation scheme cannot be chosen from the point of view of correctness alone.

Next, we used Python to calculate the Euclidean distance to further investigate and analyse the optimal criteria for subclassification.

We then used Python to calculate the Euclidean distance to further investigate and analyse the optimal criteria for subclass classification. The minimum Euclidean distance from the centroid of each of the ten sample points in the two-class cluster analysis (referred to as 'high 3 lead 3' for 3 high potassium and 3 lead barium, and so on for the rest of the cases) was calculated in the following table 3 and 4. (only some of the data are shown).

Table 3. The minimum Euclidean distance
(from the center point of the high potassium glass subclass)

Sample	1	2	3	4	5	6	7
high 3 lead 3	0.505	0.481	48.331	16.372	55.873	59.943	33.773
high 3 lead 4	0.505	0.481	48.331	16.372	55.873	59.943	33.773
high 3 lead 5	0.505	0.481	48.331	16.372	55.873	59.943	33.773
high 3 lead 6	0.505	0.481	48.331	16.372	55.873	59.943	33.773
high 4 lead 3	0.552	0.675	48.243	16.270	55.795	59.724	33.694

Table 4. The minimum Euclidean distance
(from the center point of the lead barium glass subclass)

Sample	1	2	3	4	5	6	7
high 3 lead 3	12.825	12.794	8.346	6.546	15.789	22.340	8.192
high 3 lead 4	1.509	1.334	8.346	6.925	15.789	22.340	8.192
high 3 lead 5	5.409	5.348	21.229	5.294	28.970	34.128	7.872
high 3 lead 6	3.768	3.700	28.714	5.868	36.552	41.029	13.892
high 4 lead 3	12.825	12.794	8.346	6.546	15.789	22.340	8.192

For samples 1 and 2, both of which belong to the high potassium glass category, the minimum Euclidean distances from the centroids of the lead-barium glass subclasses are much greater than the minimum Euclidean distances from the centroids of the high potassium glass subclasses, as opposed to samples 3-10, which is consistent with the correctness rates mentioned above and demonstrates the soundness of the model. A cross-sectional and vertical comparison of the tables also shows that in the case of 'High 6 Pb 3', the Euclidean distances are the smallest of the 16 cases for samples 1, 2, 3, 5, 6, 8, 9 and 10, which means that for this subclass the fit is the highest compared to the other cases and the classification results are the most appropriate. The most appropriate classification result. The clustering centroids for this classification are shown in the table 5.

Table 5. The clustering centroids

Subclass	The clustering centroids
Subclass 1	(0.00, 2.06, 0.24, 0.48, 0.12, 0.02, 0.00, 0.00)
Subclass 2	(0.00, 0.00, 0.17, 0.33, 0.18, 0.00, 0.00, 0.00)
Subclass 3	(0.00, 9.18, 0.09, 0.39, 2.16, 0.06, 0.39, 0.20)
Subclass 4	(0.00, 0.42, 0.00, 0.00, 0.199, 0.00, 0.00, 0.00)
Subclass 5	(0.00, 0.87, 0.04, 0.00, 0.11, 0.00, 0.00, 0.00)
Subclass 6	(1.39, 6.56, 0.29, 0.00, 0.24, 0.01, 0.00, 0.00)
Subclass 7	(1.78, 0.18, 20.37, 7.45, 1.42, 0.25, 0.08, 0.19)
Subclass 8	(0.19, 0.17, 39.71, 6.32, 3.80, 0.34, 0.07, 0.00)
Subclass 9	(0.00, 0.06, 8.37, 9.44, 1.38, 0.15, 0.00, 1.87)

For each sample, the Euclidean distance is calculated for each of the nine categories mentioned above and the

The classification of the glass categories was completed by taking the smallest value and assigning it to that category. The results of the above classification method for the 67 samples in the annex are shown in the table 6.

Table 6. Subclass segmentation results

Subclass	Sample
Subclass 1	03 part 2
Subclass 2	21, 27
Subclass 3	01, 04, 05, 06 part 1, 06 part 2, 18
Subclass 4	07, 09, 10, 12
Subclass 5	03 part 1, 22
Subclass 6	13, 14, 16
Subclass 7	11, 23(not weathered), 25(not weathered), 28(not weathered), 29(not weathered), 31, 32, 33, 35, 37, 42(not weathered 1), 42(not weathered 2), 44(not weathered), 45, 46, 47, 48, 49(not weathered), 50(not weathered)
Subclass 8	02, 19, 30 part 1, 30 part 2, 34, 36, 38, 39, 40, 41, 43 part 1, 43 part 2, 49, 50, 51 part 1, 51 part 2, 52, 54, 54(severe weathering), 55, 56, 57, 58
Subclass 9	08, 08(severe weathering), 20, 24, 26, 26(severe weathering), 53(not weathered)

Under the above classification method, the data from the 10 validation samples mentioned above were examined and the validation results were compared with the actual categories.

The correct classification rate was 100%, which proved that the classification results were very good.

3. Glass identification model development and solution

3.1. Model building

In order to analyse the chemical composition of the unknown category of glass artefacts and identify their type, a 'model for identifying unknown artefacts' can be developed by referring to the calculation process used in the previous section to classify the subcategories of high potassium and lead barium.

In order to obtain the classification of unknown artefacts, the characteristic chemical composition of the unknown artefacts (Na₂O, K₂O, PbO

PbO, BaO, P₂O₅, SrO, SnO₂, SO₂), and then calculate the Euclidean distances between the chemical content and the cluster centroids of the PbB and K₂O classes respectively. The characteristic chemical composition of the sample to be identified is given as V_j (which can be

(which can be viewed as an 8-dimensional coordinate value), $j = 1, 2, \dots, 8$, and the subclassification data centre point in Problem 2 is U_i (which can also be considered as

an 8-dimensional coordinate value), $i=1, 2, \dots, 9$, for each sample to be identified, the Euclidean distance d_i is obtained from the centroid of each cluster.

The Euclidean distance d_i is obtained and the minimum value is taken as the class of the artefact to be identified, i.e.

$$\min d_i = |V_j - U_i| \quad (3)$$

where $j=1, 2, \dots, 8$, $i = 1, 2, \dots, 9$. The order in which i is obtained when d_i takes its minimum value, based on the order of i , is

The specific category of the artefact sample can be indicated. The results were obtained by calculating the mean Euclidean distance between each sample and the centre of each cluster

The mean, variance and minimum of the Euclidean distance between each sample and each cluster centre.

3.2. Solving the model

Based on the model building, after extracting the characteristic chemical content of the unknown artefacts, the Euclidean distance between each artefact and each cluster centre was calculated via Python.

The Euclidean distance between each artefact and each cluster centre was calculated, as shown in the following table 7:

Table 7. Euclidean distance for unknown artifacts

	Subclass	2	3	4	5	6	7	8	9
A1	2.38	1.08	9.27	1.10	1.38	6.78	21.77	44.30	12.69
A2	36.95	36.93	37.44	37.08	37.08	37.40	20.44	13.38	30.51
A3	39.66	39.76	40.49	39.95	39.90	40.01	19.57	2.33	31.68
A4	26.66	26.72	27.53	26.96	26.95	27.34	8.31	16.25	17.56
A5	12.30	12.28	15.28	12.49	12.46	13.63	9.82	28.04	8.64

A6	0.89	1.40	8.09	0.93	0.50	5.40	21.83	40.38	12.87
A7	1.21	1.05	8.47	0.57	0.17	5.75	21.82	40.38	12.83
A8	23.87	23.92	25.54	24.22	24.19	24.82	4.83	19.41	13.02

The minimum value of the Euclidean distance for A1 is found at the cluster centroid sub-2, which is one of the subclasses of the high potassium class.

A1 belongs to subclass 2 of the High Potassium category. The rest of the unknown artefacts, and so on, can be identified in their categories, as shown in the table below. The table 8 below shows:

Table 8. Classification results of unknown artifacts

Number	Category
A1	Subclass 2(High potassium)
A2	Subclass 8(Lead barium)
A3	Subclass 8(Lead barium)
A4	Subclass 7(Lead barium)
A5	Subclass 9(Lead barium)
A6	Subclass 5(High potassium)
A7	Subclass 5(High potassium)
A8	Subclass 7(Lead barium)

4. Conclusions

To classify the glass, a 'subclass classification model based on heritage samples' was developed. The 'significance' is defined as the sum of the absolute values of the four relative changes in the samples according to their type and whether they are weathered or not. Based on this parameter, the classification pattern of the glass can be summarised. The 8 chemical components with the highest significance and 57 samples were selected and sub-classified using the K-means clustering algorithm using Matlab software. The number of subclasses was set for different categories of glass, with categories 3, 4, 5 and 6 respectively, and the number of subclasses for two categories of glass was combined to obtain 16 classification results. (9), (10). The solution with the highest "Euclidean clustering" was chosen as the final subclass classification method, and the subclass classification results for the known samples were derived. For the reasonableness of the model, the classification results were validated using the remaining 10 known samples and the results showed good reasonableness.

To identify the glass of unknown components, an 'unknown artefact identification model' was developed. The Euclidean distance between the unknown samples and the nine class centres was solved based on the cluster centroids of the different subclasses obtained in the previous section, and the class represented by the class centre with the lowest Euclidean distance was taken as the class of the unknown artefact.

References

- [1] Wu Changfeng. Isotope archaeology traces out ancient Silk Road glass production and distribution patterns [N]. Science and Technology Daily,2022-11-25(003). DOI: 10.28502/n.cnki.nkjrb.2022.006650.
- [2] Wang Dong, Wen Rui, Zhu Yingpei, Hu Xingjun, Li Wenying. Study on glass beads with metal foil layers excavated from Yingpan Cemetery, Yuli County, Xinjiang[J]. Archaeology and Cultural Relics,2022(04):117-122.
- [3] Ji Luoyuan, PJ Cherian. Peacock blue-glazed pottery specimens excavated from Kerala, India [J]. Journal of the Palace Museum, 2022(06):55-67+148. DOI: 10.16319/j.cnki.0452-7402.2022.06.004.
- [4] Do Khai Ly, Su Miao, Yang Yuanyuan. Propagation and communion - the influence of Chinese culture on decorative arts along the Silk Road [J]. Journal of Donghua University (Social Science Edition),2022,22(02): 37-46.DOI:10.19883/j.1009-9034.2021.0319.
- [5] Wang Liyuan. The foreign exchange of Qi from the unearthed artifacts in and around Linzi, the capital of Qi[J]. Hai Dai Journal, 2021(02):148-159.
- [6] Pan Ling, Tan Wenyu. The cultural factors of the west in the Xianbei remains of Hulunbuir--and the "grassland silk road" in the Han period[J]. Archaeology,2022(05):110-120+2.
- [7] Cui Ning, Wang stack. Chinese and foreign exchanges on the grassland silk road in Tongliao during the Liao Dynasty[J]. Journal of Inner Mongolia University for Nationalities (Social Science Edition), 2022,48(01):23-28. doi: 10.14045/j.cnki.nmsx.2022.01.001.
- [8] Duck-shaped glass notes: a testimony of the Grassland Silk Road [J]. Communist Party Member,2022(02):2.
- [9] Shang Yue. A grassland silk road connects east and west cultures [N]. Liaoning Daily, 202201-05(012). doi: 10.28534/n.cnki.nlnrb.2022.000037.
- [10] Jiang Bo. Ports, shipwrecks and trade goods: archaeological findings and research on the Maritime Silk Road [J]. Studies in the history of maritime transport, 2021(04):8-22. DOI: 10.16674/j.cnki.cn35-1066/u.2021.04.002.