

Fisher Linear Discriminant and Systematic Cluster Analysis based on analysis and identification of the composition of ancient glass products

Jiankun Fang *, Shihao Cao, Zhaoyang Li

Department of Computer and Information Technology, Beijing Jiaotong University, Beijing, China

* Corresponding Author Email: 20722004@bjtu.edu.cn

Abstract. The Silk Road was a channel of cultural exchange between China and the West in ancient times, of which glass was a valuable physical evidence of early trade exchanges. Therefore, it is important to analyze the ancient glass products for the research related to ancient times. In this paper, in order to get the classification law of high potassium and lead-barium glass, firstly, the discriminant relationship function was obtained by using Fisher Linear Discriminant Analysis. Secondly, the data were downscaled by Principal Component Analysis to get five principal components, and then Fisher Linear Discriminant Analysis was used to get the discriminant function, and the results all passed the cross-check, and the model accuracy was 95.29% and 91.43%, respectively. The subclasses were classified by Systematic Clustering Model and Elbow Method, and it was concluded that the high potassium glass was mainly divided into two subclasses based on whether the silica content was greater than 80%; the lead-barium glass was mainly divided into four subclasses based on the content of three different chemical components: silica, barium oxide and phosphorus pentoxide. And through the reasonableness and sensitivity analysis, this paper got that different chemical compositions have different size influence on the classification of glass.

Keywords: Ancient glass products, Principal Component Analysis, Fisher Linear Discriminant Analysis, Systematic Clustering Model.

1. Introduction

As one of the trade commodities of the Silk Road, glass was introduced to China from West Asia and Egypt in the early days. By absorbing and developing its technology, China has produced high potassium glass and lead-barium glass with Chinese characteristics ^[1]. The analysis of glass types is very important for the conservation of the glass artifacts and research related to ancient times. However, because ancient glass is highly susceptible to weathering caused by environmental factors, its internal elements can exchange heavily for elements contained in the buried environment, thus changing the ratio of the chemical composition of the glass and interfering with the determination of its glass type ^[2].

2. Exploring the classification rules of ancient glass products

There are 67 valid samples in this paper, including 16 high potassium glasses and 51 lead-barium glasses. Weathered glass accounts for 59% of this. Unweathered glass accounted for 41%. For these samples there are 15 indicators, including the presence or absence of weathering and 14 chemical composition contents. Silicon dioxide (SiO_2), sodium oxide (Na_2O), potassium oxide (K_2O), calcium oxide (CaO), magnesium oxide (MgO), aluminum oxide (Al_2O_3), iron oxide (Fe_2O_3), copper oxide (CuO), lead oxide (PbO), barium oxide (BaO), phosphorus pentoxide (P_2O_5), strontium oxide (SrO), sulfur dioxide (SO_2), and tin oxide (SnO_2), respectively. The sum of these fourteen chemical content components is approximately equal to 100. However, some samples have some missing indicators. Some of the data are shown in Table.1.

Table 1. Some data of ancient glass samples

Number	1	2	3	4	5
Type	High Potassium	Lead Barium	High Potassium	High Potassium	High Potassium
Whether weathering	No	Yes	No	No	No
SiO ₂	69.33	36.28	87.05	61.71	65.88
Na ₂ O	/	/	/	/	/
K ₂ O	9.99	1.05	5.19	12.37	9.67
CaO	6.32	2.34	2.01	5.87	7.12
MgO	0.87	1.18	/	1.11	1.56
Al ₂ O ₃	3.93	5.73	4.06	5.5	6.44
Fe ₂ O ₃	1.74	1.86	/	2.16	2.06
CuO	3.87	0.26	0.78	5.09	2.18
PbO	/	47.43	0.25	1.41	/
BaO	/	/	/	2.86	/
P ₂ O ₅	1.17	3.57	0.66	0.7	0.79
SrO	/	0.19	/	0.1	/
SnO ₂	/	/	/	/	/
SO ₂	0.39	/	/	/	0.36

2.1. Fisher Linear Discriminant analysis of glass classification law

To analyze the classification pattern of high potassium glass and lead-barium glass based on the data, we intend to perform Fisher Linear Discriminant analysis on the data ^[3-4].

This paper proposes to perform Fisher linear discriminant analysis on all valid samples for surface weathering, silicon dioxide (SiO₂), sodium oxide (Na₂O), potassium oxide (K₂O), calcium oxide (CaO), magnesium oxide (MgO), aluminum oxide (Al₂O₃), iron oxide (Fe₂O₃), copper oxide (CuO), lead oxide (PbO), barium oxide (BaO), phosphorus pentoxide (P₂O₅), strontium oxide (SrO), sulfur dioxide (SO₂), and tin oxide (SnO₂). Among all valid samples, 50 samples were randomly selected as the training set, and the remaining 17 samples were used as the test set.

In order to eliminate the influence of chance, this paper intends to repeat the training for Fisher Linear Discriminant analysis five times, so as to achieve the purpose of cross-validation, and finally obtain the mean value for all models, and obtain Table.2 of classification coefficient functions after processing.

Table 2. Classification coefficient function after sample processing

	Type = Lead barium glass	
	0.00	1.00
SiO ₂	30.595	31.247
Na ₂ O	33.525	35.706
K ₂ O	30.609	27.865
CaO	38.567	39.796
MgO	40.357	41.821
Al ₂ O ₃	26.497	28.363
Fe ₂ O ₃	28.588	26.777
CuO	15.462	14.926
PbO	32.076	33.824
BaO	40.164	42.312
P ₂ O ₅	41.149	43.326
SrO	-66.732	-82.072
SnO ₂	5.498	11.724
SO ₂	11.715	11.399
Surface weathering/ no weathering	38.626	63.288
(Constants)	-1522.002	-1618.660

Based on Fisher Linear Discriminant analysis, the classification function relationship of Function (1) (Y_1 (lead-barium glass)), Function (2) (Y_2 (high-potassium glass)) was established. Using the Functions (1) and (2) to discriminate the sample categories, the 15 indicators were brought into Y_1 (lead-barium glass) and Y_2 (high-potassium glass) in turn, and the type corresponding to the larger value was the glass type of the sample.

$$\begin{aligned}
 Y_1(\text{HighPotassium}) &= -1522.002 + 30.595 \cdot SiO_2 + 33.525 \cdot Na_2O + 30.609 \cdot K_2O \\
 &+ 38.567 \cdot CaO + 40.357 \cdot MgO + 26.497 \cdot Al_2O_3 + 28.588 \cdot Fe_2O_3 \\
 &+ 15.462 \cdot CuO + 32.076 \cdot PbO + 40.164 \cdot BaO + 41.149 \cdot P_2O_5 \\
 &- 66.732 \cdot SrO + 5.498 \cdot SnO_2 + 11.715 \cdot SO_2 + 38.626 \\
 &\cdot \text{SurfaceWeathering/NoWeathering}
 \end{aligned} \quad (1)$$

$$\begin{aligned}
 Y_2(\text{LeadBarium}) &= -1618.660 + 31.247 \cdot SiO_2 + 35.706 \cdot Na_2O + 27.865 \cdot K_2O \\
 &+ 39.796 \cdot CaO + 41.821 \cdot MgO + 28.363 \cdot Al_2O_3 + 26.777 \cdot Fe_2O_3 \\
 &+ 14.926 \cdot CuO + 33.824 \cdot PbO + 42.312 \cdot BaO + 43.326 \cdot P_2O_5 \\
 &- 82.072 \cdot SrO + 11.724 \cdot SnO_2 + 11.399 \cdot SO_2 + 63.288 \\
 &\cdot \text{SurfaceWeathering/NoWeathering}
 \end{aligned} \quad (2)$$

Table.3 shows the results of Fisher discrimination of glass types, as can be seen from the table that the recognition rate of the training set is 100% and the recognition rate of the test set is 95.29%. Due to the high recognition rate of this classification model, it passed cross-validation.

Table 3. Glass type Fisher discriminant set

Prediction Group Information		Measurement Group Information			
		High Potassium	Lead Barium	Total	Recognition Rate
Training Set	High Potassium	65	0	65	100
	Lead Barium	0	185	185	100
	Recognition Rate				100
Test Set	High Potassium	22	3	25	88
	Lead Barium	1	59	60	98.33
	Recognition Rate				95.29

2.2. Glass classification law based on principal component analysis and Fisher Linear Discriminant analysis

To further support the conclusion, the data were processed by principal component analysis and judged again using Fisher Linear Discriminant analysis^[5-7].

Table 4. Total Variance Explanation

Ingredients	Initial eigenvalue			Extraction of the sum of squares of loads		
	Total	Percentage of variance	Cumulative%	Total	Percentage of variance	Cumulative%
1	4.373	29.155	29.155	4.373	29.155	29.155
2	2.418	16.122	45.278	2.418	16.122	45.278
3	1.970	13.132	58.410	1.970	13.132	58.410
4	1.416	9.442	67.851	1.416	9.442	67.851
5	1.161	7.740	75.591	1.161	7.740	75.591

Based on the total variance explained in Table.4, the five eigenvalues were set as principal components because their cumulative sum exceeded 75%.

The composition matrix was also analyzed accordingly and five principal components were interpreted: the first principal component represents the content of lead-barium silicate and high-potassium silicate, the second principal component represents the chemical components that are more easily oxidized and have higher content, the third principal component represents the products of the reaction of copper ions in glass with oxygen and sulfur dioxide in air, the fourth principal component

represents the oxides in glass that react easily with water, and the fifth principal component represents the products of oxidation of tin in glass with low oxidizability.

Based on Fisher Linear Discriminant analysis, the classification function relationship of Function (3) (Y_3 (High Potassium)), Function (4) (Y_4 (Lead Barium)) was established. Using the Functions (3) and (4) to discriminate the sample categories, the five indicators were brought into Y_3 (High Potassium), Y_4 (Lead Barium) in turn, and the type corresponding to the larger value was the glass type of the sample.

$$Y_3(\text{HighPotassium}) = -3.918 + 1.993 \cdot F1 - 0.126 \cdot F2 + 1.072 \cdot F3 - 2.113 \cdot F4 + 0.056 \cdot F5 \quad (3)$$

$$Y_4(\text{LeadBarium}) = -1.495 - 1.011 \cdot F1 + 0.065 \cdot F2 - 0.556 \cdot F3 + 1.003 \cdot F4 + 0.073 \cdot F5 \quad (4)$$

Table.5 shows the results of glass type Fisher discrimination, and from Table.5, the recognition rate is 95.71% for the training set and 91.43% for the test set. The classification model has a high recognition rate and passes cross-validation.

Table 5. Glass type Fisher discriminant set

Prediction Group Information		Measurement Group Information			
		High Potassium	Lead Barium	Total	Recognition Rate
Training Set	High Potassium	60	0	60	100
	Lead Barium	9	141	150	94
	Recognition Rate				95.71
Test Set	High Potassium	18	2	20	90
	Lead Barium	4	46	50	92
	Recognition Rate				91.43

2.3. Systematic clustering models solve subclass delimitation problems

Based on the chemical composition of lead-barium and high-potassium glass types, a systematic cluster analysis was performed to determine the optimal number of clusters by applying the elbow rule, and finally the optimal number of clusters was obtained as $K = X$. The results allowed the classification of the types into X classes.

First, a systematic cluster analysis was performed on the high potassium glass to obtain the shortest clustering graph of the high potassium classification in Figure 1. The high potassium classification at each scale can be obtained from Figure 1, and to obtain the optimal number of clusters, the elbow rule is used to estimate their optimal number by graphs [8-9]. The coefficients obtained from the clustering results of Figure 1 allow drawing the line graph of high potassium aggregation coefficients of Figure 2.

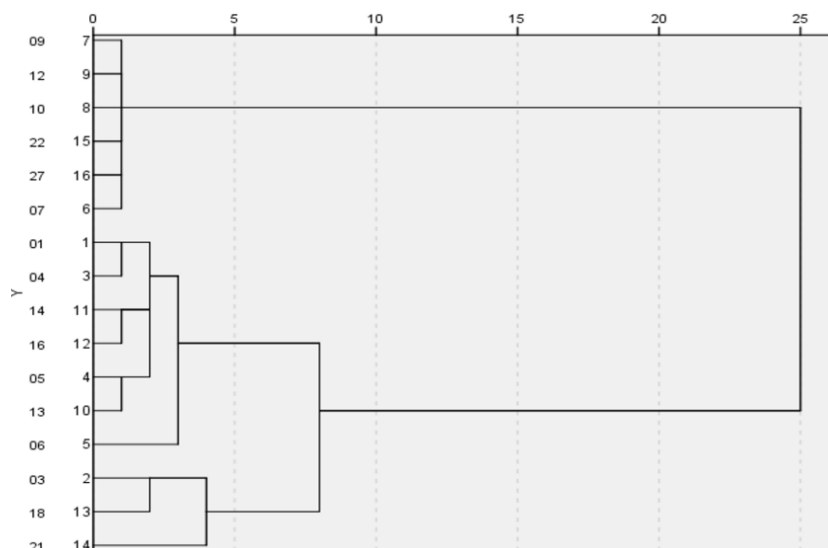


Figure 1. Shortest clustering diagram for high potassium classification

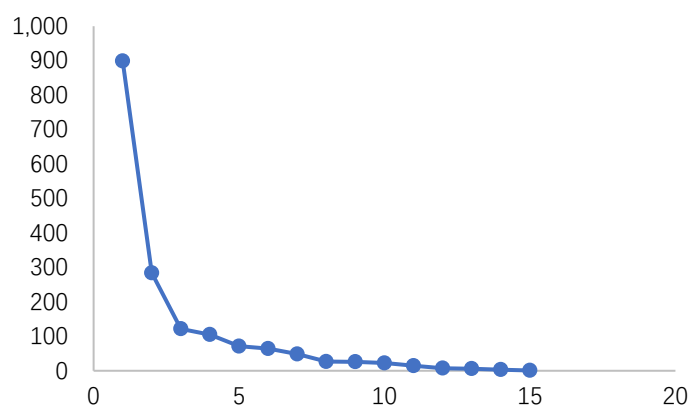


Figure 2. High potassium polymerization coefficient line graph

According to Figure 2, when the number of categories is 2, the decreasing trend of the fold tends to slow down, indicating that the point is at the inflection point where the number of categories is high and the boundaries of the categories are clearer, so the category can be set to 2. In turn, the elbow value $K = 2$. Therefore, all high potassium glasses are classified into 2 categories, as shown in Figure 3 for the classification.

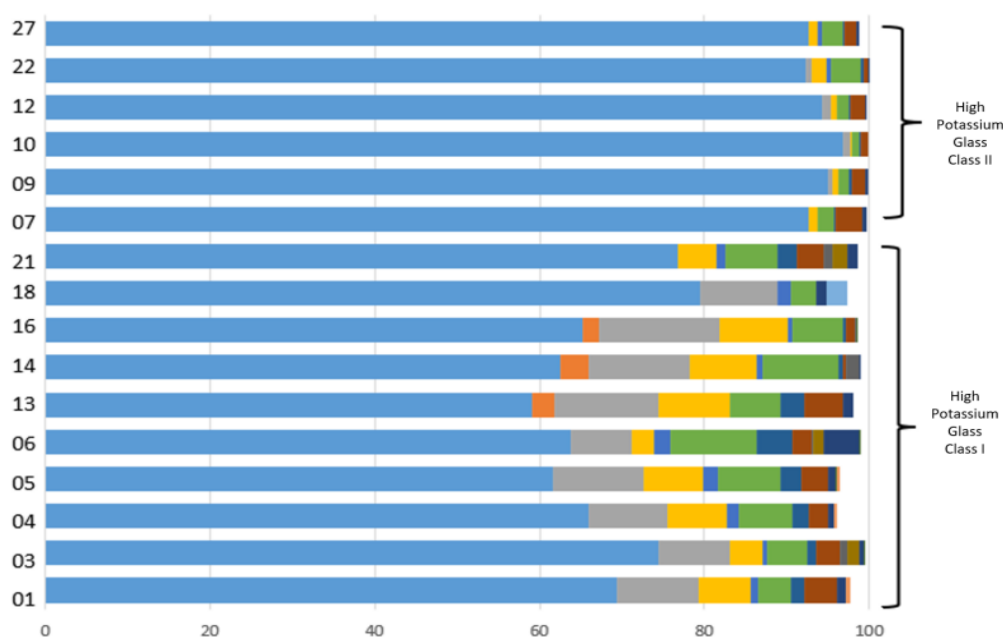


Figure 3. Classification of high potassium glass based on cluster analysis model

The 2 subclasses of high potassium glass, named high potassium I and high potassium II, are divided by the following method: based on the content of silica, if the chemical content of silica in the glass is less than 80% and the content of potassium oxide is greater than 5%, it can be considered high potassium glass I. If the content of silica in the glass is greater than 80% and the content of potassium oxide is almost 0, it can be considered high potassium glass II. The results of the delineation are artifacts 1, 3, 4, 5, 6, 13, 14, 16, 18, 21, i.e., high potassium glass with ornament A or ornament C, which are judged as high potassium glass class I; artifacts 7, 9, 10, 12, 22, 27, i.e., high potassium glass with ornament B, which can be judged as high potassium glass class II.

Reasonable analysis: It is found that the high potassium glass with decoration B is weathered glass, and the high potassium glass with decoration A and C are unweathered glass, and the difference in chemical composition between the two types of glass, silica and potassium oxide, respectively, is large, so this classification method is more reasonable.

The same systematic clustering analysis and elbow-law analysis for lead-barium glass yields that the decreasing trend of the fold tends to slow down when the number of categories is 4, which in turn

yields the elbow value $K = 4$. Therefore, all lead-barium glass is classified into 4 categories, as shown in Figure 4.

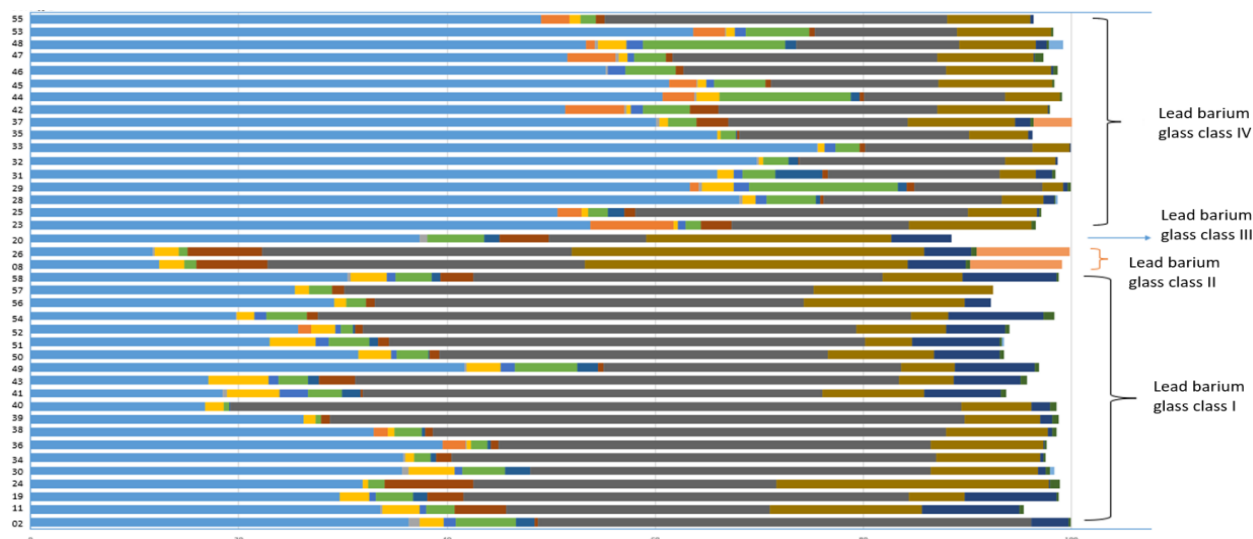


Figure 4. Classification of lead barium glass based on cluster analysis model

The four subclasses of lead barium glass are clearly named as lead barium I, lead barium II, lead barium III and lead barium IV, which are divided as follows: if the silica content of lead barium glass is greater than 40% and the phosphorus pentoxide content is less than 5%, the glass can be considered as lead barium IV glass; if the silica content of lead barium glass is less than 40% and the phosphorus pentoxide content is greater than 5%, the glass can be considered as glass is lead-barium III glass; if the silica content of lead-barium glass is less than 20% and the barium oxide content is greater than 30%, the glass can be considered as lead-barium II glass; if the silica content of lead-barium glass is greater than or approximately equal to 20%, less than 40%, and the lead oxide content is greater than 30%, the glass can be considered as lead-barium glass I class. The results of the classification are: heritage number 02, 11, 19, 24, 30, 34, 36, 38, 39, 40, 41, 43, 49, 50, 51, 52, 54, 56, 57, 58 of the glass is lead-barium glass class I, heritage number 08, 26 of the glass is lead-barium glass class II, heritage 20 of the glass is lead-barium glass class III, heritage number 23, 25, 28, 29, 31, 32, 33 The glass with artifact numbers 23, 25, 28, 29, 31, 32, 33, 35, 37, 39, 42, 44, 45, 46, 47, 48, 53, 55 is lead-barium glass category IV.

Rational analysis: the classification method is mainly based on three chemical components silica, phosphorus pentoxide and barium oxide content of different lead barium glass into four types, each type can be judged by the amount of content of at least two of the above-mentioned chemical components to improve the rationality of classification, reduce the content of a certain chemical composition of a certain artifact and the possibility of misjudging the type of glass, according to which the classification method is known to be more reasonable.

2.4. Sensitivity Analysis

The five indicators F1-F5 obtained from the principal component analysis were made to fluctuate up and down continuously by 200% for each of these variables individually, keeping the other parameters constant, and the probabilities of the subclass types of lead-barium and high-potassium glass were predicted by the predictive system cluster analysis model, respectively. The results are plotted as Figure 5 and Figure 6 shown below.

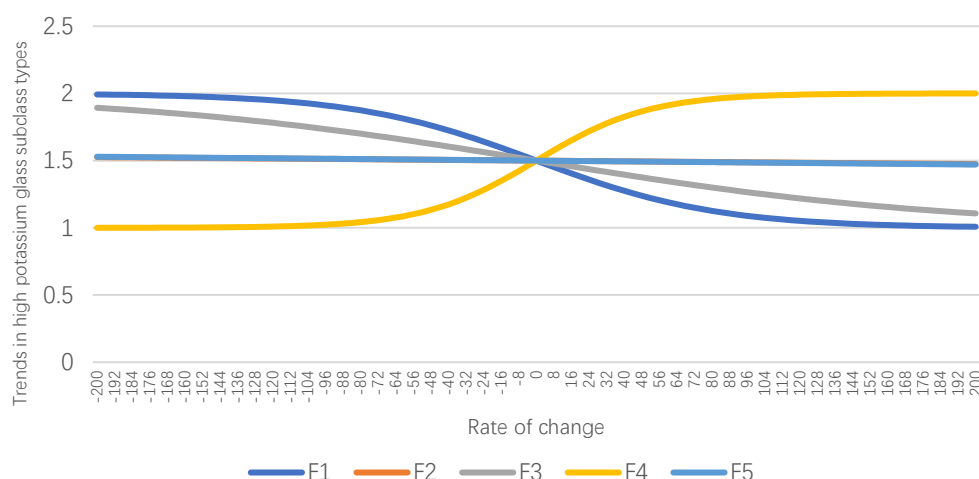


Figure 5. Sensitivity analysis of high potassium glass subclasses

From Figure 5, for the classification of high potassium glass subclasses, the content of F4, F5, and F3 had a large effect on the classification of their subclasses, while the content of F2 and F5 had almost no effect on the classification of their subclasses.

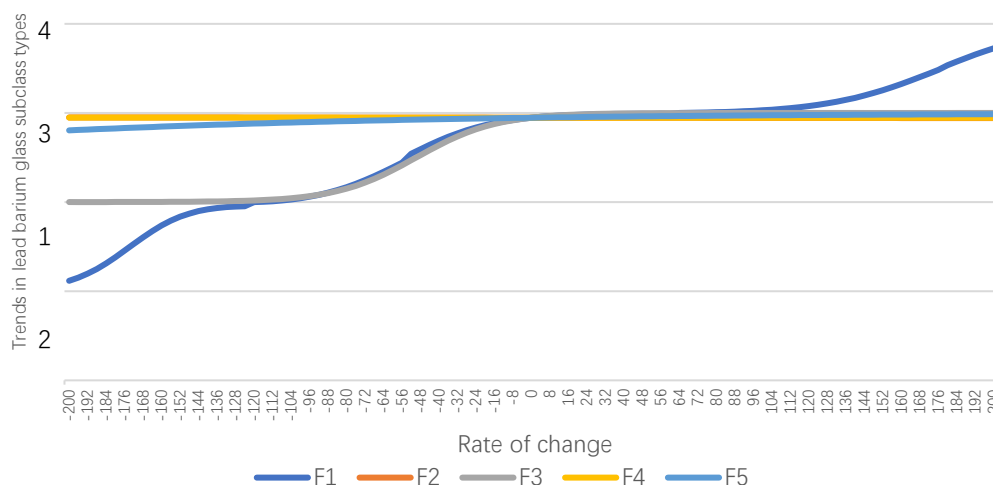


Figure 6. Sensitivity analysis of lead barium glass subclasses

From Figure 6, it can be obtained that for the classification of lead-barium glass subclasses, the contents of F1 and F3 have a large influence on the classification of their subclasses, while the contents of F2, F4 and F5 have almost no influence on the classification of their subclasses.

3. Forecast of ancient glassware categories

3.1. The establishment of simulation model

For the chemical composition of the unknown category of glass artifacts, after entering the relevant data into Fisher Linear Discriminant Function (1) and Function (2), a collated analysis leads to Table.6.

Table 6. Unknown category of glass artifact type identification results

Artifact Number	Y1	Y2	max(Y1, Y2)	Type
A1	1530.967	1527.002	1530.967	High Potassium
A2	1615.649	1632.471	1632.471	Lead Barium
A3	1693.693	1717.516	1717.516	Lead Barium
A4	1619.93	1644.137	1644.137	Lead Barium
A5	1483.304	1485.518	1485.518	Lead Barium
A6	1463.546	1433.804	1463.546	High Potassium
A7	1475.846	1451.003	1475.846	High Potassium
A8	1463.538	1478.933	1478.933	Lead Barium

As can be known from Table.6, A1, A6 and A7 artifact types are high potassium glass, and A2, A3, A4, A5 and A8 artifact types are lead-barium glass.

The eight sample data were then processed according to the weights to obtain the following Table.7.

Table 7. Input data after weight processing

Artifact Number	F1	F2	F3	F4	F5
A1	1.58	0.13	0.11	-0.13	0.13
A2	-1.42	0.84	-1.41	-1.31	-1.26
A3	-0.38	1.91	-0.28	1.14	-1.26
A4	-0.47	1.46	-0.17	0.89	-0.28
A5	1.11	-0.16	-1.44	2.39	2.26
A6	1.08	-1.69	0.12	-1.74	-0.27
A7	1.08	-1.61	-0.04	-1.48	-0.2
A8	-2.59	-0.87	3.1	0.24	0.88

After inputting the data into the classification model discriminant Function (3) and Function (4), a collation analysis leads to Table.8.

Table 8. Unknown category of glass artifact type identification results

Artifact Number	Y1	Y2	max(Y1, Y2)	Type
A1	-0.38555	-3.26599	-0.38555	High Potassium
A2	-5.66795	-0.62673	-0.62673	Lead Barium
A3	-7.69554	0.22045	0.22045	Lead Barium
A4	-7.11716	0.04182	0.04182	Lead Barium
A5	-8.1528	0.73518	0.73518	Lead Barium
A6	2.23752	-4.52838	2.23752	High Potassium
A7	1.51046	-4.16833	1.51046	High Potassium
A8	-6.10489	-0.3517	-0.3517	Lead Barium

As can be seen from Table.8, the results coincide with the solution method in Table.6, the test results are correct, and the conclusions can be used.

3.2. Sensitivity analysis

The five indicators F1-F5 obtained through Table.7 making these variables individually fluctuate up and down by 300% continuously respectively, keeping the other parameters constant, and predicting the probability of the type of glass at each case through the predictive system cluster analysis model. The results are plotted in Figure 7 as follows.

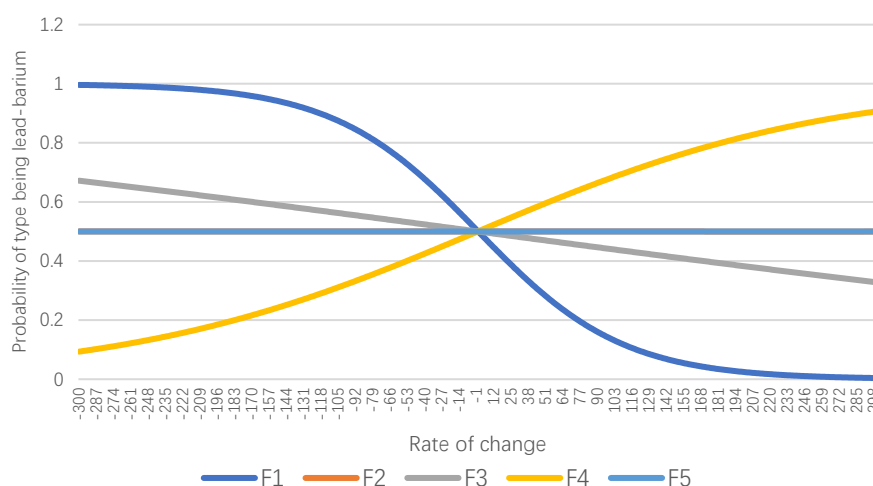


Figure 7. Sensitivity analysis of glass subclasses

As can be known in Figure 7, F1 is the most sensitive, with changes in glass type is more affected, followed by F4 is also more sensitive; F2 and F5 almost do not affect the glass type. Because of this of F1 in the prediction of glass type, more attention should be paid to the chemical composition of F1 and F4 related content.

4. Conclusions

In this paper, in order to study the classification pattern of ancient glass products. Fisher Linear Discriminant analysis, Principal Component analysis, and Systematic Cluster analysis models were used to analyze 67 samples and predict 8 unknown samples^[10]. Among them, 5 principal components were obtained by principal component analysis, and the accuracy of the classification model was obtained by this with 95.29% and 91.43%, respectively. Then the artifacts were classified into subclasses by systematic clustering and elbow method, and the high potassium glass was divided into two classes, and the lead-barium glass was finally divided into four classes. Finally, the classification model predicted the glass classes of eight unknown samples.

References

- [1] Yin Yulong. Composition analysis of ancient glass products by correlation prediction [J]. Contemporary Chemical Research, 2023, No.126(01): 122-126.
- [2] Wang Zhihao, Zhao Xiangwei, Li Zhiquan. A machine learning-based method for prediction and subclassification of original composition of weathered silicate glass [J]. Journal of Silicates, 2023, 51(02): 416-426. DOI: 10.14062/j.issn.0454-5648.20220985.
- [3] Shen Haijun, Zhang Tanglei, Xu Zhenxing, et al. Research on the identification of apple freshness based on Fisher Discriminant analysis [J/OL]. Food Industry Science and Technology: 1-11 [2022-09-18]. DOI:10.13386/j.issn1002-0306.2022050023.
- [4] Shan Qiufu, Zhang Tao, Li Chao. Construction of cigarette mainstream smoke quality prediction model based on Fisher Linear Discriminant analysis method [J]. Food and Machinery, 2021, 37(02): 78-84+92. DOI: 10.13652/j.issn.1003-5788.2021.02.014.
- [5] Ma Fengyan. Application of principal component analysis and linear discriminant analysis in classification problems [J]. Science and Technology Perspectives, 2015, No.136(13): 52+54. DOI: 10.19694/j.cnki.issn2095-2457.2015.13.032.
- [6] Xiao Lijing, Wang Dongsheng, Wei Yuan. Research on chestnut nut quality evaluation based on principal component analysis [J]. Hebei Fruit Tree, 2023, No.173(01): 18-23. DOI: 10.19440/j.cnki.1006-9402.2023.01.007.

- [7] Peng Yongtao, Chen Zhongzhi, Zhao Chen, Wang Xinnan. Cluster analysis of urban rail transit system based on principal component analysis [J]. Sino-Foreign Highway, 2021,41(05): 322-326. DOI: 10.14048/j.issn.1671-2579.2021.05.068.
- [8] Mei Xiaoren, Xia Qing. Performance evaluation model and empirical research of university teachers based on systematic cluster analysis [J]. Financial Theory and Teaching, 2020(04): 89-90. DOI: 10.13298/j.cnki.ftat.2020.04.016.
- [9] Zhang Yang. A comparative analysis of Tokyo Bay Area and Bohai Bay Area-Based on entropy value method and systematic cluster analysis [J]. National Circulation Economy, 2021(36): 106-110. DOI: 10.16834/j.cnki.issn1009-5292.2021.36.026.
- [10] Sichin Bilig, Li Qinghui, Gan Fuxi. Analysis of ancient Chinese potassium glass fractions by laser exfoliation-inductively coupled plasma-atomic emission spectrometry/mass spectrometry [J]. Analytical Chemistry, 2013, 41(09): 1328-1333.