Analysis of glass artifacts based on hierarchical clustering and regression models

Tao Yun*, Bingzhao Chen, Zexin Bi

School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou, Guangdong, China, 510006

* Corresponding Author Email: 13178508308@163.com

Abstract. Glass is a valuable evidence of the early friendly trade between China and the Western countries. In this study, the rationality and sensitivity of the developed model were first evaluated by analyzing the definitive and quantitative data of each type of glass artifacts and establishing a model to classify the glass artifact categories into subcategories, as well as categorizing unknown types of glass. Considering the discrete and discontinuous nature of the data, the chi-square test was chosen to investigate the correlation between glass type, decoration and color and surface weathering. Secondly, descriptive statistics of high potassium and lead-barium glass artifacts before and after weathering were conducted to obtain statistical patterns about the chemical composition content, and then the chemical composition content of glass artifacts before weathering was predicted based on the mean rate of change of the chemical composition content of high potassium and lead-barium glass before and after weathering, respectively. Finally, this study further explored the relationship between glass artifacts and the content of each chemical composition using Spearman's correlation coefficient to derive the basis for the classification of glass types, based on the hierarchical clustering method to classify the subclasses of high potassium glass and lead-barium glass with and without weathering, respectively, and used the elbow rule to derive the most appropriate number of subclasses for the above four types of glass, which endowed the model with rationality.

Keywords: Chi-square test, Spearman correlation coefficient, Hierarchical clustering, Glass artifacts.

1. Introduction

China had good cultural exchanges with the Chinese and Western countries through the Silk Road in the early days, among which glass was a valuable proof of friendly trade exchanges [1-2]. In the early days, glass was made into jewelry and introduced into China. After learning its techniques, our traditional glass artisans made a new generation of glass with similar appearance but different chemical composition by taking local materials. At the same time the chemical composition of different influencing factors are mainly the addition of different accelerants, for example, lead barium glass with an accelerant of aluminum ore, which has a high content of lead oxide, barium oxide, which is usually our own research and development of glass varieties [3]. In contrast, high potassium glass uses substances such as grass ash, which has a higher content of potassium, as an accelerant, and is mainly popular in Lingnan, Southeast Asia and India.

To this end, we explore the shallow characteristics of glass artifacts, analyze the relationship between whether the glass surface is weathered or not and each definite class of variables, and predict the chemical composition content of artifacts before weathering based on the statistical pattern of each chemical composition content and weathering point data [4-6]. It is possible to understand the factors related to weathering on the surface of glass artifacts, and to analyze the relationship between each category of variables in order to reach a preliminary understanding of the characteristics of glass artifacts [7]. In addition, this study further explores the deeper characteristics of glass artifacts by analyzing the basis for classification of glass types (high potassium, lead-barium) in the annexes based on the content of each chemical component of the glass, and then classifying each type of glass into subclasses before and after weathering, and analyzing the rationality and sensitivity of the classification [8]. The relationship between glass types and their respective constituent contents is explored, and subclasses are classified according to this relationship, and quantitative variation is analyzed to achieve a deeper understanding of the characteristics of glass artifacts [9-10].

2. Model building and solving

2.1. Modeling and solution of the chemical composition content prediction model

2.1.1. Flowchart of ideas

The modeling process of the glass chemical composition content prediction model is shown in Figure 1.

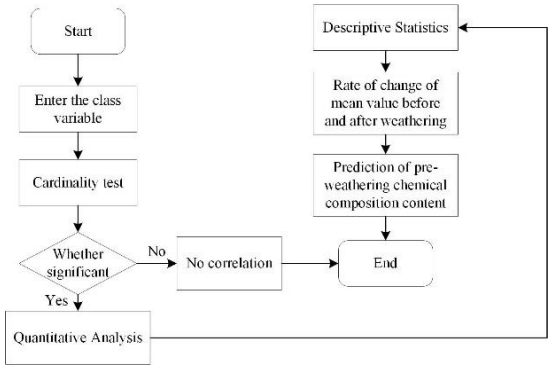

Figure 1. Flow Chart.

2.1.2. Data pre-processing

The given condition shows that$85\% < Z_i < 105\%$, excluding the data outside this interval, and $Z_i$is the cumulative sum of component proportions.

According to

$$Z_i = \sum_{j=1}^{j} z_{ij} \tag{1}$$

Obtains

$$Z_{15} = 79.47\% < 85\% \tag{2}$$

$$Z_{17} = 71.89\% < 85\% \tag{3}$$

Since the average potassium oxide content of unweathered high potassium glass is 10.179 in the data in row 21, no potassium oxide content was detected in this row, which would cause interference in the subsequent analysis. In summary, the data in rows 15, 17 and 21 were excluded. Finally, all vacant data were filled with zeroes.

2.1.3. Cardinality test

The chi square test results for surface weathering, type, decoration, and color are shown in Table1.

Table.1. Cardinality test

Properties	Name	Whether the surface is weathering		Total	X^2	Correction X^2	P
		Weather-free	Weathering				
Type	Lead Barium	12	28	40	4.16	2.986	0.041**
	High potassium	9	6	15			
Ornament	A	10	11	21	4.513	4.513	0.105
	B	0	6	6			
	C	11	17	28			
Color	nan	4	0	4	10.111	10.111	0.257
	Light green	1	2	3			
	Light blue	12	6	18			
	dark green	4	3	7			
	Dark blue	0	2	2			
	Violet	2	2	4			
	Green	0	1	1			
	Blue-Green	9	5	14			
	Black	2	0	2			

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively.

According to the results of the chi-square test, based on the surface weathering and glass type, the p-value is 0.041**, which is less than 0.05, showing significance at the level, and the original hypothesis is rejected, indicating that whether the surface weathering of glass artifacts and its glass type data have a greater correlation; based on the surface weathering and glass decoration, the p-value is 0.105, which is greater than 0.05, showing no significance at the level, and the original hypothesis cannot be rejected, indicating that whether the surface weathering of glass artifacts and its decoration data have a greater correlation. It means that whether the surface of glass artifacts is weathered or not is not correlated with its decoration data; based on surface weathering and glass color, the p-value is 0.257, which is greater than 0.05, and does not present significance at the level, and the original hypothesis cannot be rejected, which means that whether the surface of glass artifacts is weathered or not is not correlated with its color data.

2.1.4. Statistical laws and forecasts

Descriptive statistics were performed on the data before and after weathering for high potassium and lead-barium glass, respectively, and the statistical results are shown in Table 2. The prediction model was also established for the data of each chemical composition of high potassium glass and lead-barium respective glass without weathering and after weathering.

Table 2. Descriptive statistics of the content of each chemical component of high potassium glass after weathering.

Variable Name	Sample size	Maximum value	Minimum value	Average value	Standard deviation	Median	Variance
Silicon dioxide	6	96.77	92.35	93.963	1.734	93.505	3.005
Sodium oxide	6	0	0	0	0	0	0
Potassium oxide	6	1.01	0	0.543	0.445	0.665	0.198
calcium oxide	6	1.66	0.21	0.87	0.488	0.83	0.238
magnesium oxide	6	0.64	0	0.197	0.306	0	0.094
Aluminum oxide	6	3.5	0.81	1.93	0.964	1.72	0.93
Iron oxide	6	0.35	0.17	0.265	0.069	0.275	0.005
copper oxide	6	3.24	0.55	1.562	0.935	1.545	0.874
Lead oxide	6	0	0	0	0	0	0
barium oxide	6	0	0	0	0	0	0
phosphorus pentoxide	6	0.61	0	0.28	0.21	0.28	0.044
Strontium oxide	6	0	0	0	0	0	0
Tin oxide	6	0	0	0	0	0	0
Sulfur dioxide	6	0	0	0	0	0	0

Using equation (4), the average value of each chemical composition content for high potassium glass and lead-barium glass without weathering was found.

$$M_i = \frac{1}{T}\sum_{j=1}^{14} z_{ij} \tag{4}$$

Using equation (5), the average value of the content of each chemical composition of high potassium glass and lead-barium glass after weathering was found.

$$N_i = \frac{1}{T}\sum_{j=1}^{14} z_{ij} \tag{5}$$

Using equation (6), the rate of change before and after weathering was obtained separately.

$$A_i = \left(\frac{M_i - N_i}{M_i}\right) \times 100\% \tag{6}$$

Based on the model algorithm established above, the following table of mean change rates is derived, as shown in Table 3.

Table 3. High potassium and lead barium glass before and after weathering mean change rate table.

Variable Name	High Potassium Glass			Lead barium glass		
	No Weathering Mean	Weathering Mean	Change Rate	No Weathering Mean	Weathering Mean	Change Rate
Silicon dioxide	67.194	93.963	-39.84%	53.444	33.615	37.10%
Sodium oxide	0.758	0	100.00%	0.772	0.953	-23.45%
Potassium oxide	10.179	0.543	94.67%	0.258	0.143	44.57%
calcium oxide	5.389	0.87	83.86%	1.232	2.346	-90.42%
magnesium oxide	1.066	0.197	81.52%	0.492	0.701	-42.48%
Aluminum oxide	6.659	1.93	71.02%	3.195	3.838	-20.13%
Iron oxide	1.892	0.265	85.99%	0.933	0.556	40.41%
copper oxide	2.377	1.562	34.29%	1.557	1.996	-28.20%
Lead oxide	0.358	0	100.00%	23.594	36.872	-56.28%
barium oxide	0.474	0	100.00%	410.499	10.487	0.11%
phosphorus pentoxide	1.43	0.28	80.42%	0.904	4.155	-359.62%
Strontium oxide	0.045	0	100.00%	0.297	0.366	-23.23%
Tin oxide	0.215	0	100.00%	0.065	0.056	13.85%
Sulfur dioxide	0.111	0	100.00%	0.282	0.987	-250.00%

According to the rate of change sought above, a set of data was taken in high potassium and lead-barium glass respectively to predict their chemical composition content before weathering, and the results are shown in Table 4.

Table 4. Predicted chemical composition content before weathering table.

	High Potassium Glass			Lead barium glass		
	After weathering	Rate of change	Pre-weathering prediction	After weathering	Rate of change	Pre-weathering prediction
Silicon dioxide	92.35	-0.3984	55.56	36.28	0.371	49.74
Sodium oxide	0	1	0	0	-0.2345	0
Potassium oxide	0.74	0.9467	1.44	1.05	0.4457	1.52
calcium oxide	1.66	0.8386	3.05	2.34	-0.9042	0.22
magnesium oxide	0.64	0.8152	1.16	1.18	-0.4248	0.68
Aluminum oxide	3.5	0.7102	5.99	5.73	-0.2013	4.58
Iron oxide	0.35	0.8599	0.65	1.86	0.4041	2.61
copper oxide	0.55	0.3429	0.74	0.26	-0.282	0.19
Lead oxide	0	1	0	47.43	-0.5628	20.74
barium oxide	0	1	0	0	0.0011	0
phosphorus pentoxide	0.21	0.8042	0.38	3.57	-3.5962	-9.27
Strontium oxide	0	1	0	0.19	-0.2323	0.15
Tin oxide	0	1	0	0	0.1385	0
Sulfur dioxide	0	1	0	0	-2.5	0

2.2. Modeling and solving the glass type classification model

2.2.1. High potassium, lead barium glass classification law

The Spearman correlation coefficient is calculated as follows.

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^{\,2}}{n(n^2-1)} \tag{7}$$

The Spearman's correlation coefficients between the type and the content of each chemical component were calculated to analyze the classification pattern of high potassium and lead-barium glasses, and the results were obtained as shown in Tables 5 and 6.

Table 5. Spearman's correlation coefficient table for weathered glass artifacts.

	Type variables	K_2O	PbO	BaO
Type	1.000(0.000***)	0.315(0.042**)	-0.607(0.000***)	-0.558(0.000***)
K_2O	0.315(0.042**)	1.000(0.000***)	-0.407(0.007***)	-0.233(0.138)
PbO	-0.607(0.000***)	0.407(0.007***)	1.000(0.000***)	0.209(0.185)
BaO	-0.558(0.000***)	-0.233(0.138)	0.209(0.185)	1.000(0.000***)

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively.

Table 6. Table of Spearman's correlation coefficients for unweathered glass artifacts.

	Type variables	K_2O	PbO	BaO
Type	1.000(0.000***)	0.868(0.000***)	-0.868(0.000***)	-0.880(0.000***)
K_2O	0.868(0.000***)	1.000(0.000***)	-0.775(0.000***)	-0.750(0.000***)
PbO	-0.868(0.000***)	-0.775(0.000***)	1.000(0.000***)	0.819(0.000***)
BaO	-0.880(0.000***)	-0.750(0.000***)	0.819(0.000***)	1.000(0.000***)

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively.

Correlation; based on type and barium oxide content, the P value is less than 0.01, showing significance at the level, and the original hypothesis is rejected, indicating a strong correlation between type and barium oxide content data; meanwhile, the correlation coefficients of potassium oxide, lead oxide, and barium oxide are 0.868, -0.868, and -0.880, respectively, indicating a positive correlation between type and potassium oxide content; type and lead oxide and barium oxide content are negative correlation.

In summary, the classification of glass types is based on the following: high potassium oxide glass has higher potassium oxide content and lower lead oxide and barium oxide content; lead barium glass has higher lead oxide and barium oxide content and lower potassium oxide content. This is also consistent with the narrative in the background of the question: high potassium glass adds grass ash with high potassium content as a flux, and lead-barium glass adds aluminum ore with high lead and barium content as a flux.

2.2.2. Flow chart of hierarchical clustering algorithm

The flow of the algorithm about glass type classification is shown in Figure 2.

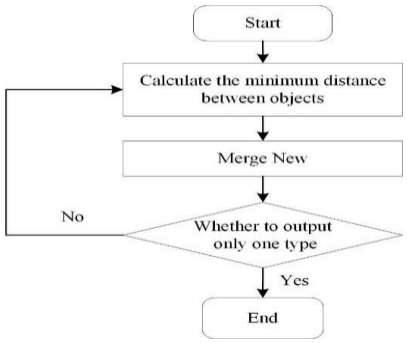

Figure 2. Flow chart of hierarchical clustering algorithm.

2.2.3. Classification of glass subclasses based on hierarchical clustering

The classification of high potassium glass and lead-barium glass has been described in the previous section, and the next section is a subcategory for high potassium glass and lead-barium glass. For the convenience of classification, high potassium glass and lead-barium glass are discussed separately according to whether they are weathered or not, and if the sampling point of the weathered glass is an unweathered area, the glass is considered as unweathered glass.

(1) The unweathered high-potassium glass was divided into subclasses as shown in Figure 2. The unweathered high-potassium glass could be classified into type A glass (silica content higher than 75%) and type B glass (silica content lower than 75%) by cluster analysis, and the classification results are shown in Table 7.

Table 7. Subclasses of unweathered high-potassium glass.

Heritage Sampling Sites	SiO_2	K_2O	PbO	BaO	Classification
03 Part 1	87.05	5.19	0.25	0	A
18	79.46	9.42	0	0	A
13	59.01	12.53	0	0	B
06 Part 2	59.81	7.68	0.35	0.97	B
03 Part2	61.71	12.37	1.41	2.86	B

(2) The unweathered lead-barium glass is divided into subclasses, and the unweathered lead-barium glass can be divided into C-type glass (lead oxide content lower than 10%), D-type glass (silica content higher than 40%), and E-type glass (which does not meet the two conditions of appeal) by hierarchical cluster analysis as shown in Table 8.

Table 8. Subcategories of unweathered lead barium glass.

Heritage Sampling Sites	SiO_2	Al_2O_3	PbO	BaO	Classification
20	37.36	5.45	9.3	9.3	C
24	31.94	1.59	29.14	29.14	E
30 Part 1	34.34	4.35	39.22	39.22	E
30 Part 2	36.93	3.86	37.74	37.74	E
23 Unweathered spots	53.79	1.42	16.98	16.98	D
25 Unweathered spots	50.61	1.9	31.9	31.9	D

(3) The weathered high potassium glass was sub-classified, and the weathered high potassium glass could be classified into F-type glass (silica content higher than 93%) and G-type glass (silica content lower than 93%) by hierarchical clustering analysis the results of the classification are shown in Table 9.

Table 9. Subcategories of weathered high-potassium glass.

Heritage Sampling Sites	SiO_2	K_2O	PbO	BaO	Classification
07	92.63	0	0	0	G
22	92.35	0.74	0	0	G
27	92.72	0	0	0	G
09	95.02	0.59	0	0	F
10	96.77	0.92	0	0	F
12	94.29	1.01	0	0	F

(4) The weathered lead-barium glass is divided into subclasses as shown in the figure, and the weathered lead-barium glass can be divided into H-type glass (alumina content higher than 10%), I-type glass (sulfur dioxide content higher than 1%) by hierarchical cluster analysis, and J-type glass that does not meet the two conditions of appeal, and the division results are shown in Table 10.

Table 10. Subcategories of weathered lead barium glass.

Heritage Sampling Sites	SiO_2	CaO	Al_2O_3	SO_2	Classification
48	53.33	2.82	13.65		H
08	20.14	1.48	1.34	2.58	I
08 Severe weathering point	4.61	3.19	1.11	15.03	I
26	19.79	1.44	0.7	1.96	I
26 Severe weathering point	3.72	3.01	1.18	15.95	I
11	33.59	3.51	2.69		J
43 Part 1	12.41	5.24	2.25		J

2.2.4. Analysis of the rationality and sensitivity of classification results

The weathered lead-barium glass artifacts are selected for the rationality and sensitivity analysis of their subclass delineation results, and all other kinds of glass artifacts are verified in the same way.

The rationality analysis is carried out to verify the rationality of subclass delineation by hierarchical cluster analysis through the calculation of k values by the elbow rule. The results are shown in Figure 3. The core idea of the elbow method is that as the number of clusters k increases, the sample division will become finer and finer, then the degree of aggregation of each cluster will become higher and higher, and then the sum of squared errors will naturally become smaller and smaller. When k is less than the true number of clusters, the degree of aggregation of each cluster increases substantially as k increases, and the degree of distortion S decreases significantly. When k reaches the true number of clusters, increasing the value of k decreases the rate of aggregation, and the decrease of S decreases abruptly, and finally leveling off as the value of k continues to increase. At this time, the graph of the relationship between S and k will take the shape of an elbow, and the elbow point, i.e., the point of inflection, is the point where k takes the best value. The inflection point exists at K=3, and the number of division categories is chosen to be 3 on comprehensive consideration.

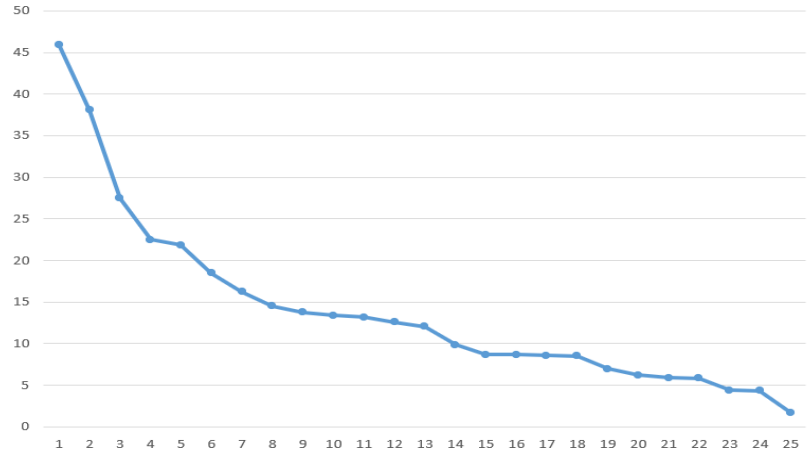

Figure 3. Elbow Law.

For the analysis of sensitivity, the contour coefficient $s(i)$ takes into account both tightness and separation and is calculated as shown.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{8}$$

$a(i)$ Denotes the average distance of sample i to other samples in its category C. The smaller $a(i)$ is, the more sample ishould belong to the category. $b(i)$ denotes the average distance of sample i to all samples in some other classC_i, the smallest value of b_{ij}, the larger $b(i)$ is, the less sample iC_i belongs to other classes.Therefore, the value of $s(i)$ is between -1 and 1. The more $s(i)$ converges to 1, the more reasonable the cluster is, and vice versa, the more unreasonable it is. Calculated by MATLAB, the contour coefficients$s(i) = 0.866$, which is close to 1 and has good sensitivity.

Subsequent reasonableness and sensitivity analyses of the other three types of glass artifacts were conducted, and good reasonableness as well as sensitivity were obtained.

3. Conclusions

The correlation between glass type, decoration and color and surface weathering was obtained by chi-square test for the mining of glass artifacts definite class data; for the classification of glass types, Spearman correlation coefficient analysis applicable to discrete data was used to obtain a more accurate classification basis. For subclass classification, a hierarchical clustering classification algorithm was chosen, and a contour coefficient of 0.866 was obtained with good sensitivity. However, due to the small sample size, there is a certain chance in the analysis of the data, which has a certain impact on the prediction and analysis of the model, and further in-depth research will be conducted by expanding the sample size and choosing a suitable improved model in the future.

References

[1] Shiwei X, Ye T, Zizheng W, et al. [J]. International Journal of Electrical Power and Energy Systems, 2023,152.

[2] Jinwon S Seonghyun J, Min Y C, et al. [J]. Computational Statistics and Data Analysis, 2023,185.

[3] Xiangyu S, Min X, Jiang D. Max-sum test based on Spearman's footrule for high-dimensional independence tests [J]. Computational Statistics and Data Analysis, 2023,185.

[4] Xiangwen L ,Yiwen H ,Yuxue X , et al. Whole-tumor histogram analysis of diffusion-weighted imaging and dynamic contrast-enhanced MRI for soft tissue sarcoma: correlation with HIF-1alpha expression.[J]. European radiology, 2023, 33(6).

[5] C M S, M X J H, Eugenia M, et al. Evidence for inefficient contraction and abnormal mitochondrial activity in sarcopenia using magnetic resonance spectroscopy. [J]. Journal of cachexia, sarcopenia and muscle, 2023.

[6] Zeng Y. Classification and identification of glass artifacts based on high-dimensional clustering and XGBoost algorithm [J]. Advances in Computer, Signals and Systems, 2022, 6(6).

[7] Brigitte B, Laure D. Exceptional Potash Glass Artifacts Excavated at Tissamaharama (Sri Lanka) [J]. Journal of Glass Studies, 2022, 64.

[8] Carter A, Dussubieux L, Polkinghorne M, et al. Glass artifacts at Angkor: evidence for exchange [J]. Archaeological and Anthropological Sciences, 2019, 11(3).

[9] Colomban P. Rocks as blue, green and black pigments/dyes of glazed pottery and enamelled glass artefacts? A review [J]. European Journal of Mineralogy, 2014, 25(5).

[10] Ufuoma U J, M. M Y, Kabiru A M, et al. Environmental degradation of structural glass systems: A review of experimental research and main influencing parameters [J]. Ain Shams Engineering Journal, 2023, 14(5).