Classification and identification of glass artifacts based on fuzzy clustering and binary logit regression analysis

Tong Zhang, Mingyu Cui, Nina He *

School of Management, Qufu Normal University, Rizhao, China, 276827

* Corresponding Author Email: he1505091645@163.com

Abstract. This paper establishes a mathematical model based on the existing data of glass types, summarizes the classification rules of high potassium glass and lead-barium glass, and selects the appropriate chemical composition and subclasses for high potassium glass and lead-barium glass. And based on the classification law, the chemical composition data of the unknown category of glass artifacts are used to identify the type to which they belong.

Keywords: In(p/1-p) Type discrimination model, fuzzy cluster analysis, glass weathering.

1. Introduction

At present, with the in-depth promotion of heritage conservation work, the restoration of cultural relics has gradually become a social consensus, and plays an important role in cultural heritage and national development. Among them, cultural relics represented by ancient glass are highly susceptible to environmental and other factors that interfere with subsequent conservation work such as category definition. Therefore, by establishing a reasonable mathematical model to clarify the impact factors of cultural relics conservation and the interrelationship between the factors by scientific means, it is conducive to meeting the cultural needs of the masses and promoting the long-term development of cultural relics conservation.

2. Glass classification model building and solving

2.1. Analysis of the classification law of high potassium glass and lead-barium glass

Firstly, based on the analysis of the data of decoration, color and chemical content, the t-test was used to find out the factors that have a significant influence on the classification of glass, and further clarify the content of the role of these factors on the classification of glass, and finally get the law of glass classification. The results are shown in the figure 1.

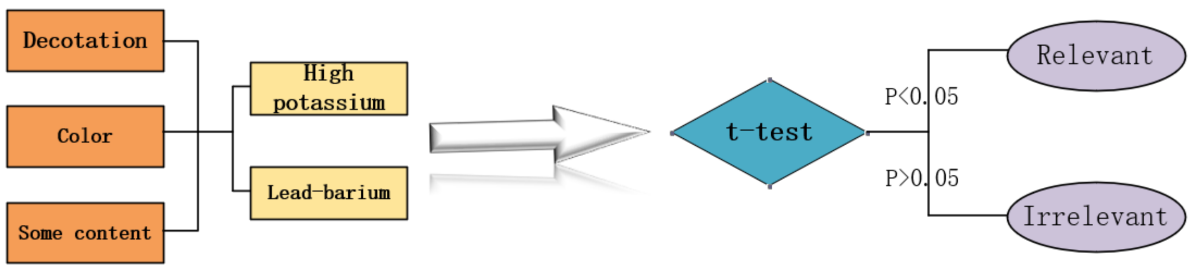

Figure 1. Flow chart of classification law of high potassium glass and lead-barium glass

(1) The influence of decoration and color on glass classification

First, we discussed the ornamentation and color, and the data were filtered by Excel to obtain the following table (Table 1), which shows the statistical frequency of the categories:

Table 1. Ornamentation, color frequency table

		Ornamentatio			Color							
Type		A	B	C	Dar	ligh	blue-	dark	gree	light	purpl	blac
High	1	6	6	6	1	4	13	1	0	0	0	0
Lead Barium	4	16	0	24	1	16	3	6	1	3	4	2

The t-test for ornamentation and color respectively yielded,

Table 2. Test analysis results

	Type (mean ± standard deviation)		t	p
	Lead Barium (n=40)	High Potassium (n=18)		
Decoration	2.20±0.99	2.00±0.84	0.792	0.433
Color	3.33±2.06	5.50±2.23	-3.550	0.001**

From Tables 2, it can be seen that: different types of samples do not show significance (p>0.05) for ornamentation total 1 item, which means that different types of samples show consistency for ornamentation all, and there is no difference. In addition, the type samples showed significance (p<0.05) for the total 1 item of color, which means that the different types of samples have differences for color. The specific analysis shows that the type shows a 0.01 level of significance for color (t=-3.550, p=0.001), and the specific comparison of differences shows that the mean value of lead barium (3.33) is significantly lower than the mean value of high potassium (5.50).

(2)The influence of chemical composition on glass classification

Second, using t-tests, each chemical composition was analyzed to investigate whether there was a significant difference between the two glass types at each chemical composition. Also weathering was not considered, only the glass type was considered.

The confidence interval was set at 0.95, and if p<0.05, the difference between different types of glass in this composition is significant, and the classification law of high potassium glass and lead-barium glass can be obtained by comparing the mean value of this chemical composition in both types of glass; conversely, there is no significant difference between different types of glass in this composition, and the change of composition has no effect on glass classification.

Taking SiO_2 as an example, an independent sample t-test using SPSS yields.

Table 3. SiO_2 Independent sample test results

	F	Significance	t	Degree of freedom	Significance (two-tailed)	Mean Difference	Standard error value
Assuming equal variance	.358	.553	-7.436	42	.000	-39.95481	5.37341
Not assuming equal variance			-7.919	29.901	.000	-39.95481	5.04542

As shown in Tables 3, p=0.000<0.05, SiO_2 has a significant effect on the classification of glass types. Meanwhile, the mean value of lead-barium glass is 39.47, which is significantly lower than the mean value of high potassium glass (83.68). Therefore, the SiO_2 composition of lead-barium glass is low and the composition of high-potassium glass is high.

The remaining 13 chemical compositions can be determined by t-test synoptically to determine whether they are significantly different from the glass types, and if significantly different, the glass types with higher composition mean values are compared, and the results are as follows in table 4:

Table 4. Table of variability of each chemical composition

	ingredients	whether there is a significant difference	Type with higher means
1	SiO_2	Yes (p=0.000)	High potassium
2	Na_2O	No (p=0.483)	—
3	K_2O	Yes (p=0.000)	High potassium
4	CaO	Yes (p=0.004)	High potassium
5	MgO	No (p=0.570)	—
6	Al_2O_3	No (p=0.156)	—
7	Fe_2O_3	No (p=0.015)	High potassium
8	CuO	No (p=0.588)	—
9	PbO	Yes (p=0.000)	Lead barium
10	BaO	Yes (p=0.000)	Lead barium
11	P_2O_5	Yes (p=0.016)	Lead barium
12	SrO	Yes (p=0.000)	Lead barium
13	SnO_2	No (p=0.253)	—
14	SO_2	No (p=0.329)	—

In summary, the chemical components that affect the classification of glass are SiO_2、K_2O、CaO、Fe_2O_3、PbO、BaO、P_2O_5 and SrO.Glass composition K_2O、CaO、Fe_2O_3 Higher than lead barium glass;PbO、BaO、P_2O_5、SrO Lower than lead barium glass.

2.2. Subdivision of two types of glass

The chemical composition of high-potassium glass and lead-barium glass was subdivided, and some chemical components with more missing values and the above classification of glass were screened out, and fuzzy cluster analysis was used to obtain a pedigree chart.

(1) High potassium glass

Use $i = 1,2,\cdots,9$ to represent nine chemical componentsI、II、III、IV、V、VI、VII、VIII、IX;use $x_j(j = 1,2,3\cdots 56)$ to represents the 56 valid indicator variables in Table 2;The value of the j indicator variable of the i chemical component unit is denoted as a_{ij}, then the Manhattan distance between the i and j chemical component units is:

$$d_{ij} = \sum_{k=1}^{4}\left|a_{ik} - a_{jk}\right| \tag{1}$$

Take the similarity factor:

$$r_{ij} = 1 - 0.1d_{ij} \tag{2}$$

This then yields the F similarity matrix.transitive closure is obtained by the flat method, that is, the equivalence matrix found.Plot the high-potassium glass lineage diagram with SPSS as follows:

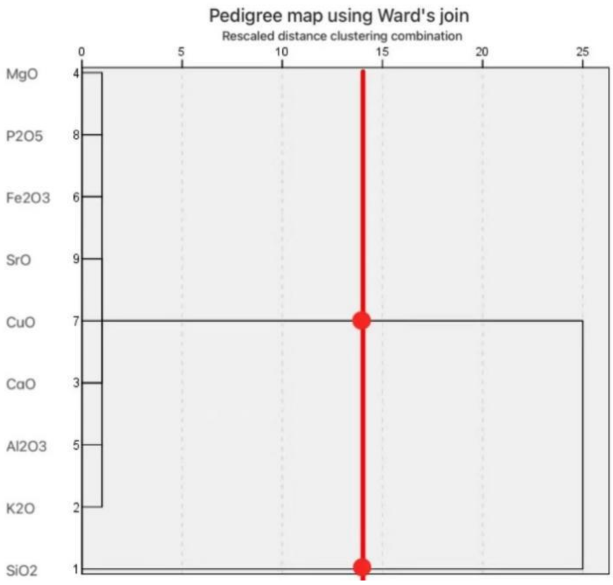

Figure 2. Cluster lineage of chemical components of high-potassium glass

From Figure 2, high-potassium glass can be divided into SiO_2-type high-potassium glass and other chemical composition high-potassium glass according to chemical composition.

(2) Lead barium glass

The subclassification process of lead-barium glass is similar to that of high-potassium glass, using fuzzy clustering, and the lineage chart is drawn as figure 3:

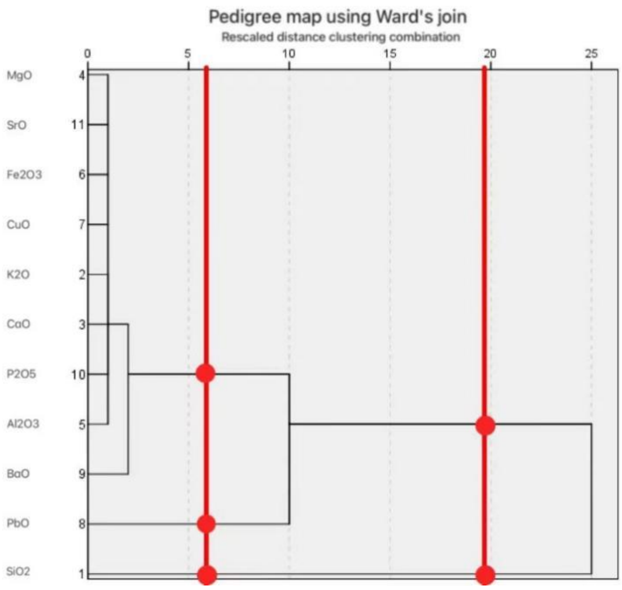

Figure 3. Lineage diagram of chemical composition of lead-barium glass

If divided into two categories, SiO_2 become a separate category, other chemical components are the second category, lead barium glass can be subdivided into SiO_2 type lead barium glass and other chemical composition type lead barium glass; If it is divided into three categories, $SiO_2$is a class, PbO is the second category, other chemical components are the third category, lead barium glass can be subdivided into SiO_2 type lead barium glass and other chemical composition type lead barium glass. The greater the distance between the two red lines, the greater the change in the middle distance of their classes, and the better their clustering effect.

3. Identification of glass artifacts of unknown composition

3.1. Identify the type of unknown artifact

Firstly, the data of the detected glass is cleaned, the correlation coefficient between each two chemical components is calculated by the CORREL function, and the correlation coefficient matrix and correlation coefficient heat map are obtained after correlation screening, and in order to ensure the accuracy of the data, the chemical composition column data of correlation r≥0.8 is temporarily ignored, that is, SiO_2.

After that, a 0-1 variable is introduced to data encode the glass type, setting high potassium to 0 and lead barium to 1. Binary Logit regression analysis was performed using 14 chemical components as independent variables and glass type as dependent variables. The results are shown in the table 5.

Table 5. Likelihood ratio test results of binary Logit regression model

model	-2x log-likelihood	chi-square value	df	p	AIC value	BIC value
Intercept only	83.079					
Final model	0.000	83.079	14	0.000	30.000	63.512

From Table 5, it can be seen that the original assumption of the model test here H_0 the quality of the model is the same when the independent variables are put in. The results show that the p-value is less than 0.05, rejecting the original assumption that the independent variables put in this model are valid, and the model construction is meaningful.

Table 6. R side analysis table

Dependent variable	$McFadden\ R^2$	$Cox\,\&\,Snell\ R^2$	$Nagelkerke\ R^2$
type	1.000	0.700	1.000

From Table 6, it is reasonable to take each chemical composition as the independent variable and the type as the dependent variable, that is, the glass type can be reflected by the chemical composition content. As can be seen from the table above, the model formula is:

$$\begin{gathered} ln(p/1-p) = -27.889 + 0.086 * SiO_2 - 1.303 * Na_2O - 0.132 * K_2O - 0.542 * CaO \\ +2.950 * MgO + 1.049 * Al_2O_3 + 0.886 * Fe_2O_3 - 2.877 * CuO + 1.642 * \\ PbO + 2.543 * BaO - 2.770 * P_2O_5 + 2.382 * SrO - 5.328 * SnO_2 - 1.122 * SO_2 \end{gathered} \quad (3)$$

Bringing chemical composition data into the model predicts its type, and the prediction results are shown in Table 7:

Table 7. Table of prediction results

Artifact number	Surface weathering	intercept	$ln(p/1-p)$	type
A1	No weathering	-21.89	-12.97	High potassium
A2	weathering	-21.89	-3.541	High potassium
A3	No weathering	-21.89	58.628	Lead barium
A4	No weathering	-21.89	31.218	Lead barium
A5	weathering	-21.89	22.367	Lead barium
A6	weathering	-21.89	-17.51	High potassium
A7	weathering	-21.89	-13.14	High potassium
A8	No weathering	-21.89	16.169	Lead barium

From Table 7, the inspection results of glass cultural relics of unknown categories are A1, A2, A6, and A7 are high-potassium glass, and the rest of the cultural relics are lead barium glass. The model results are shown in Figure 4:

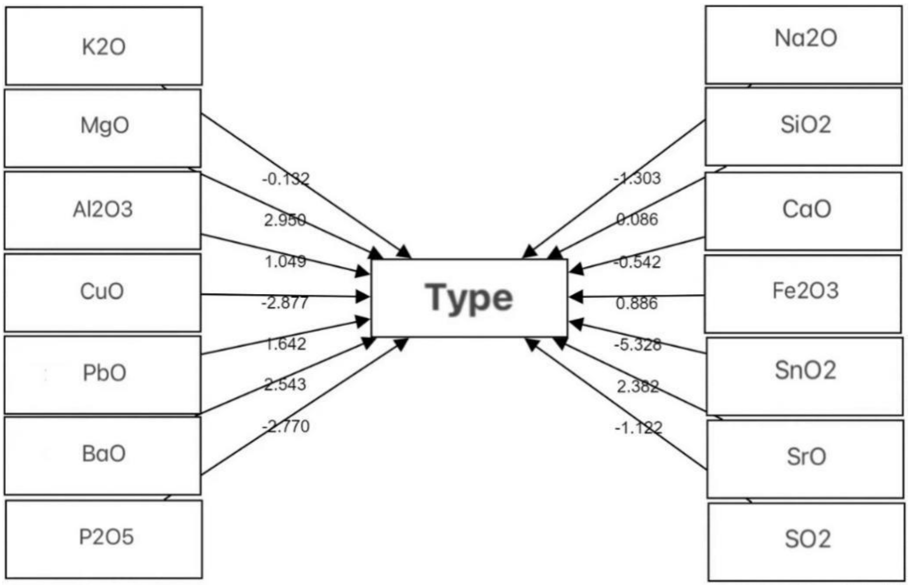

Figure 4. Binary Logit regression results

The OR value is chosen to reflect the strength of the correlation between the chemical composition and the glass type:

$$p(y=1)=\frac{e^{z}}{(1+e^{z})}$$
$$p(y=0)=\frac{1}{(1+e^{z})} \tag{4}$$

3.2. Sensitivity Analysis

To further test the sensitivity of the model, 1/20 of the mean is set to the random perturbation term, that is, the random deviation $\varepsilon = 1/20 * \bar{X}$

Averaging each chemical component gives random deviations, as shown in Table 8:

Table 8. Table of random deviations of chemical composition

chemical composition	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3
mean	60.38	0.15	0.64	3.51	0.78	5.13	2.12
Random deviation	3.02	0.01	0.03	0.18	0.04	0.26	0.11
chemical composition	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
mean	56.90	0.19	0.63	1.55	0.63	4.84	1.43
Random deviation	2.84	0.01	0.03	0.08	0.03	0.24	0.07

The random disturbance terms of each element in Table 5-17 are added to the chemical composition content to calculate whether there is a deviation of the glass type, that is, whether the random disturbance term affects its accuracy after adding the random disturbance term to test its sensitivity.

The chemical content of each element after adding the random perturbation term is substituted into the model after adding the random perturbation term, and the feedback results are shown in Table 5-18:

Table 9. Feedback results

Artifact number	Surface weathering	intercept	$ln(p/1-p)$	type
A1	No weathering	-21.889	-11.20	High potassium
A2	weathering	-21.889	-1.77	High potassium
A3	No weathering	-21.889	60.40	Lead barium
A4	No weathering	-21.889	32.99	Lead barium
A5	weathering	-21.889	24.13	Lead barium
A6	weathering	-21.889	-15.74	High potassium
A7	weathering	-21.889	-11.38	High potassium
A8	No weathering	-21.889	17.94	Lead barium

Comparing this result with the results of Table 9, it can be seen that adding random perturbation terms has no effect on the test results of glass type, that is, the model sensitivity is low and the model robustness is high.

4. Conclusions

Firstly, the t-test is used to study the decoration, color and various chemical components of the two glasses, and it is concluded that the factors that have a significant influence on the classification type are chemical composition. Secondly, fuzzy clustering is used to divide the two kinds of glass, and the high potassium glass can be further divided into two categories according to the chemical composition. Lead barium glass can be divided into two or three categories, in line with the law of the composition required for glass production, that is, the model is reasonable. Finally, the sensitivity of the model is detected by comparing the clustering results under the square Euclidean distance and Euclidean distance algorithms, respectively.

Based on the attached data, this paper establishes a category discrimination model to identify cultural relics of unknown categories. Firstly, the correlation coefficient between each chemical component is calculated by Correl, and the data is cleaned according to the correlation, and the correlation coefficient matrix and correlation coefficient heat map are established. Secondly, 0-1 variables were introduced to code the glass type, 14 chemical components were used as independent variables, and the glass type was used as the dependent variable for binary Logit regression analysis, and the ln(p/1-p) type discriminant model was constructed, and A1, A2, A6, and A7 were detected as high-potassium glass, and the rest of the cultural relics were lead-barium glass. Finally, by setting random perturbation terms to test the sensitivity of the model, when the chemical composition content parameters in the model change by $1/20 * \bar{X}$, the accuracy is still 99.9%, and the model is robust.

References

[1] Mei Jinying. The application of analytical procedures in auditing [D]. Jingdezhen Ceramics University, 2022. DOI: 10.27191/d.cnki.gjdtc.2022.000059.

[2] Sheng, Lin-Li. Clinical study on predicting the growth trend of pure ground glass nodules in lung based on quantitative parameters of artificial intelligence CT [D]. Kunming Medical University, 2021. DOI: 10.27202/d.cnki.gkmyc.2021.000654.

[3] Wang Chen-Wang. Study on the origin characteristics and firing process of Donggou kiln celadon with modern analytical techniques[D]. Zhengzhou University, 2021. DOI:10.27466/d.cnki.gzzdu.2021.004138.

[4] Xie Xilin. Predictive value of CT features and quantitative analysis of pure ground glass nodules for pathological classification of lung adenocarcinoma[J]. Journal of Shanxi Health and Health Professions College,2021,31(01):69-71.

[5] Chen Yehang. Research on the segmentation method of subsolid pulmonary nodules based on the activity contour model[D]. Guilin University of Electronic Science and Technology, 2020. DOI: 10.27049/d.cnki.ggldc.2020.000015.

[6] Yang Zongyi. Research on the detection method of insulator surface fouling composition and the influence of composition on AC fouling flash characteristics[D]. Chongqing University, 2019.DOI: 10.27670/d.cnki.gcqdu.2019.003459.

[7] Zhao YJ. Human vision-based target segmentation technology and its application[D]. Northeastern University, 2018.DOI: 10.27007/d.cnki.gdbeu.2018.000497.

[8] Jian Chuanxia. Research on machine vision detection and classification method for cell phone glass screen surface defects [D]. Guangdong University of Technology,2017.

[9] Huang Qingdan,Yang Zhongyi,Song Haoyong,Jiang Xingliang,Wu Peiwei,Zhang Zhijin. Natural fouling component measurement and ion pairing analysis on insulator surfaces of different materials[J]. High Voltage Technology,2016,42(12): 3917-3923.DOI: 10.13336/j.1003-6520.hve.20161128028.

[10] Xia B.H.,Liu Y.C.,Xu G.L.. Experimental study of compressive properties of waste glass concrete[J]. Concrete,2013(07):26-29+43.

[11] Wang Hu. Research on the formation process and properties of blast furnace slag [D]. Beijing Institute of Technology, 2013.