ISSN: 2688-8645

Composition Analysis and Identification of Ancient Glass Products

Weisen Cheng, Shuyu Wei*, Genrui Zhang

Zhantianyou Honors College, Beijing Jiaotong University, Beijing, China

*21281202@bjtu.edu.cn

Abstract

This paper mainly studies the statistical analysis of chemical composition of ancient glass products, classification and prediction of types, and other issues. It is necessary to complete the analysis of the surface weathering of glass relics, the physical characteristics and chemical composition of glass, as well as the analysis of the classification rules of high-potassium glass and lead-barium glass. Through the analysis of chemical composition, the types of cultural relics are predicted and the correlation between different types of glass chemical composition is analyzed, which is decomposed into seven small problems to be solved. In order to solve these problems, this paper adopts several models such as stochastic decision forest and principal component analysis.

Keywords

Principal Component Analysis; Random Forest Classification; Chi-Square; Cluster Analysis.

1. Introduction

The Silk Road was a channel for cultural exchanges between China and the West in ancient times, and glass is a valuable evidence of early trade. Our ancient glass and foreign glass product appearance is similar, but the chemical composition is not identical. In the process of weathering, the interior elements of the ancient glass pole exchanged a lot with the environmental elements, which resulted in the change of its composition proportion, thus affecting the correct judgment of its category. There are a number of relevant [1] data of ancient glass products in China. According to the chemical composition and other detection methods of these cultural relics samples, archaeologists have divided them into two types: high-potassium glass and lead-barium glass.

The classification of glass, the correlation analysis of chemical composition with glass type and surface weathering, these are of great value for archaeologists and researchers of cultural relics. Therefore, this paper based on the chemical composition of the mathematical model [2] to analyze and identify glass products.

2. Data Preprocessing

According to the conditions shown in the question, considering that only the relationship between a single variable and the surface weathering of cultural relics is considered in the significance analysis of data, the data of four samples lacking color are still statistically [3] significant, so they are included in the analysis of the relationship between surface weathering of cultural relics and ornamentations and colors. The data that did not meet the accumulative proportion of components and the data between 85% and 105% were regarded as invalid data, so the two groups of data that did not meet the conditions were excluded. In order to facilitate the calculation, the position of the component vacancy is made up zero processing.

3. Based on Multivariate Linear Regression of Weathering Before Chemical Composition Prediction Model

All the data given are categorical variables, and the known information does not provide the criterion for transforming them into quantitative variables. And the internal relationship between the three is unknown, so chi-square analysis and the assumption of no difference are used to analyze the relationship between the surface weathering of glass relics and their glass types, patterns and colors

3.1. Chi-square

$$X^2 = \sum \frac{(A-T)^2}{T} \tag{1}$$

Where, A is the actual value, T is the theoretical value, and the significance of X^2 value is to measure the degree of difference between theory and practice.

Table 1. Chi square test table for surface weathering

		Surface weathering (%)		Total	χ2
		Non-weathering	weathering		
Color	Light green	2(8.33)	1(3.33)	3(5.56)	6.287
	Light blue	8(33.33)	12(40.00)	20(37.04)	
	Dark green	3(12.50)	4(13.33)	7(12.96)	
	Dark Blue	2(8.33)	0(0.00)	2(3.70)	
	purple	2(8.33)	2(6.67)	4(7.41)	
	green	1(4.17)	0(0.00)	1(1.85)	
	Blue-green	6(25.00)	9(30.00)	15(27.78)	
	black	0(0.00)	2(6.67)	2(3.70)	
Total		24	30	54	
		Surface weathering (%)		Total	χ2
		Non-weathering	weathering		
Type	Lead barium			40(68.97)	6.88
	hyperkalemia	12(50.00)	6(17.65)	18(31.03)	
Total		24	34	58	
		Surface weathering (%)		Total	χ2
		Non-weathering	weathering		4.957

The chi-square test results show in table1. The hypothesis of no significant difference between surface weathering and color and ornamentation is valid, while the hypothesis of no significant difference between surface weathering and type is not valid, so there is a significant difference between them. The available surface weathering has little correlation with color and ornamentation, but significant correlation with type.

3.2. Independent-samples T test

In the analysis of the statistical law of chemical content of cultural relics with or without weathering, because of the variety of chemical content, the data samples are few. Therefore, in order to obtain accurate and effective statistical rules, this paper adopts the method of combining correlation variables when dealing with such a complex problem of small sample and multi-dimension [4]. The principal component analysis method reduces the dimension and abstracts multiple complex variables into a few linearly independent variables. Firstly, KMO test and Bartlett test were used to confirm the validity of PCA.

Table 2. KMO test and Bartlett's test

KMO test and Bartlett's test		
KMO value		0.771
Bartlett's sphericity test	Approximate chi-square	64.882
	df	6.000
	P	0.000

As shown in table 2,KMO test (KMO>0.6) indicates that there is correlation between item variables, which meets the requirements of principal component analysis.

Bartlett test: P<0.05 is significant, then principal component analysis can be performed. Then the number of factors of principal components was determined by variance interpretation table and scree map.

For lead-barium glass, the content of target compound before weathering is Y, and the content of phosphorus pentoxide, barium oxide, lead oxide, copper oxide, iron oxide, alumina, calcium oxide and silicon dioxide in the weathered sample is$X_i(i = 1,2 \ldots 8)$, the coefficient is$k_i(i = 1,2 \ldots 8)$. The intercept of the multiple linear regression equation is b,therefore $Y = k_1 * X_1 + k_2 * X_2 + \cdots + k_8 * X_8 + b$. The high-potassium glasses correspond to different target compounds, and the rest are treated in the same way. The corresponding values of $X_i$and b of each target compound are shown in the following figure 1,2 and 3.

	P_2O_5	CuO	Fe_2O_3	Al_2O_3	MgO	CaO	K_2O	SiO_2	Intercept
SiO_2	0	0.09263214	0	0	0	0	-0.8465594	3.19416272	-234.59542
K_2O	0	-1.2923651	0	0	0	0	1.28457341	-0.1891194	26.8953947
CaO	0	-1.9398141	0	0	0	0	3.91101732	-2.6730879	256.70602
MgO	0	0.77037988	0	0	0	0	1.43987234	0.23532105	-23.18382
Al_2O_3	0	0.53357368	0	0	0	0	0.11986458	-0.2004103	24.3352259
Fe_2O_3	0	-1.2358009	0	0	0	0	1.16707399	-0.454017	46.0595862
CuO	0	0.90044159	0	0	0	0	-0.5472718	0.27995042	-25.163683
P_2O_5	0	-0.0251864	0	0	0	0	0.27497984	-0.0312984	3.6407716

Figure 1. Estimated Values of Hyperkalemi compounds

	P_2O_5	CuO	Fe_2O_3	Al_2O_3	MgO	CaO	K_2O	SiO_2	Intercept
SiO_2	0	0.09263214	0	0	0	0	-0.8465594	3.19416272	-234.59542
K_2O	0	-1.2923651	0	0	0	0	1.28457341	-0.1891194	26.8953947
CaO	0	-1.9398141	0	0	0	0	3.91101732	-2.6730879	256.70602
MgO	0	0.77037988	0	0	0	0	1.43987234	0.23532105	-23.18382
Al_2O_3	0	0.53357368	0	0	0	0	0.11986458	-0.2004103	24.3352259
Fe_2O_3	0	-1.2358009	0	0	0	0	1.16707399	-0.454017	46.0595862
CuO	0	0.90044159	0	0	0	0	-0.5472718	0.27995042	-25.163683
P_2O_5	0	-0.0251864	0	0	0	0	0.27497984	-0.0312984	3.6407716

Figure 2. Estimated Values of lead barium compounds

NaCl	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
0	0	4.58	1.47	5.38	2.74	0.7	34.18	6.1	11.1	0.46	0	0
0	0	3.19	0.47	1.87	0.33	1.13	44	14.2	6.34	0.66	0	0
0	0	1.21	0	1.85	0	0.79	41.25	15.45	2.54	0	0	0
0	0	1.31	0	2.18	0	1.16	45.1	17.3	0	0	0	0
0	1.05	2.34	1.18	5.73	1.86	0.26	47.43	0	3.57	0.19	0	0
0	0.74	1.66	0.64	3.5	0.35	0.55	0	0	0.21	0	0	0
0	0.25	2.32	1.47	7.25	1.68	0.5	13.74	4.68	6.42	0.31	0.05	0.17
0	0.29	1.46	0.94	1.87	0	0.95	22.11	10.26	1.34	0.36	0.06	0.2
0	0.29	0.56	0.81	1.59	0	0.78	18.98	13.26	0.39	0.31	0.05	0.17
0	0.22	0	0.69	3.35	0	0.82	21.87	13.26	0	0.26	0.04	0.14
0	0.19	0.65	0.6	11.86	1.05	0.27	25.32	1.68	0.34	0.23	0.04	0.13
0.73	10.13	12.23	0.04	6.5	4.52	0.82	0.43	0.63	0.98	0.04	0.21	0.11

Figure 3. Forecast Result Table

4. Glass Type Classification Model based on Random Forest Classification Algorithm

According to the chemical composition of the glass category of statistical law analysis. In view of the large number of chemical components, the final chemical components should be found to be significant. The specific problem can be abstracted as the data dimension is very high, and the analysis results need to get the mathematical model of different feature importance, so this paper uses the random forest classification algorithm to analyze it.

4.1. Subclass Division

According to the conclusions given by the stochastic decision forest, the subclasses can be divided into four categories: high potassium and high lead oxide, high potassium and low lead oxide, lead barium and high lead oxide.

After effective dimensionality reduction of data through random forest model, K-means algorithm was used to determine the specific percentage of each subclass in its corresponding glass type in order to give the specific partition results of subclass division. The results are shown in Table 3.

Table 3. Type of the corresponding specific percentage of the glass types

	Cluster category (mean ± standard deviation)		F	P
	Type 2(n=37)	Type 1(n=30)		
PbO	9.526±9.708	42.886±10.666	179.096	0.000

4.2. Sensitivity Analysis of Model Results

The random forest algorithm randomly extracts samples to form a decision tree, and finally outputs the high mean value of the prediction results of all randomly generated decision trees as the result, so the results are still random even in the case of the same data division, that is, the mean value of the prediction results may be different [5]. Therefore, the sensitivity of regression prediction algorithm should be evaluated by several experiments with the same parameters to check whether the answers given by regression prediction algorithm are the same.

After investigating the sensitivity of random forest algorithm under the same data partition, that is, the same proportion of training set and test set, sensitivity test of the algorithm under different data partition parameter values can be carried out. It should be noted that the sensitivity obtained by repeated experiments for each data partition is only for a single parameter value, that is, it cannot be shown that the sensitivity of random forest algorithm has the above property under all parameter values. Therefore, the above sensitivity analysis should be repeated after each change of data partition parameter value.

5. Analysis of Glass Samples based on Spearman Analysis Method

Table 4. The original data according to its average descending position in the overall data

Variable Xi	Descending position	Grade Xi
0.8	5	5
1.2	4	
1.2	3	
2.3	2	2
18	1	1

Spearman's correlation coefficient is defined as Pearson's correlation coefficient between rank variables. For samples of sample size n, the n original data are converted to rank number. The

ISSN: 2688-8645

original data are assigned a rank [6] according to their average position in descending order in the population. As shown in the table 4.

In practice, the linkage between variables is insignificant, so you can compute ρ in simple steps. The difference between the grades of the two variables being observed, then ρ is

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{2}$$

5.1. Comparative Element Analysis

Through the analysis of the elements of different glass products, it is found that the contents of different elements are related to decoration, weathering degree, color and type. So you can compare the differences between the chemical components of different classes by comparing the differences between the elements or the differences between the contents of the elements. At the same time, we can compare the different results caused by the content differences of the same elements in different glass products to explain the differences in glass structure and weathering degree. This article focuses on the differences between comparison elements for comparison purposes.

(1) Analysis of ornamentation

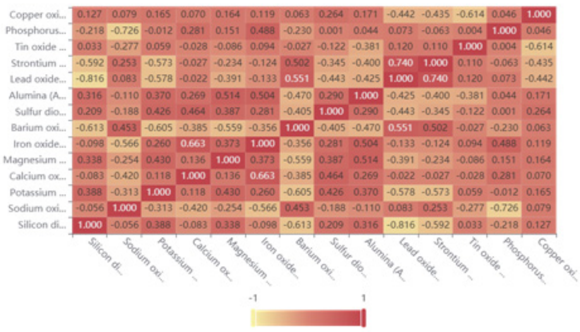

Figure 4. E Correlation coefficient thermogram analysis of ornamentation A

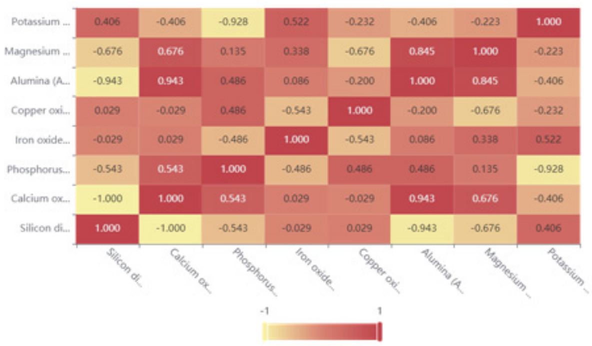

Figure 5. E Correlation coefficient thermogram analysis of ornamentation B

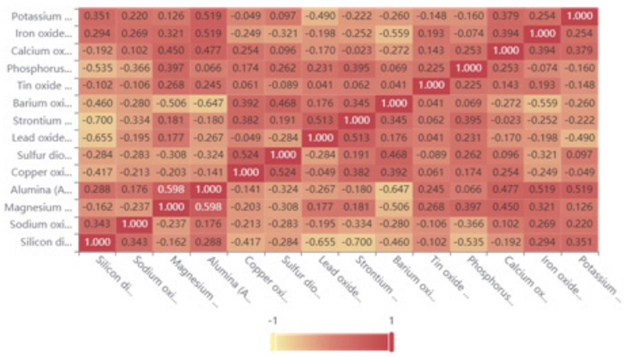

Figure 6. E Correlation coefficient thermogram analysis of ornamentation C

ISSN: 2688-8645

As shown in figure 4,5, and 6, according to the correlation coefficient thermogram, for decoration A, it is not difficult to find from the table that the correlation degree between lead oxide and strontium oxide is very high, reaching 0.740. Therefore, it is concluded that these two elements have A high correlation degree. For decoration B, it is not difficult to find from the table that the correlation degree between aluminum oxide and magnesium oxide is very high, reaching 0.845. Therefore, it is judged that these two elements have a high correlation degree. For decoration C, it is not difficult to find from the table that the correlation degree between aluminum oxide and magnesium oxide is relatively high, up to 0.598, so it is judged that these two elements have a high correlation degree.

Therefore, the related chemical elements of ornamentation A are significantly different from ornamentation B and C. Ornamentation A is A clear relationship between lead oxide and strontium oxide, while ornamentation B and C are A clear relationship between alumina and magnesium oxide.

(2) Analysis of types

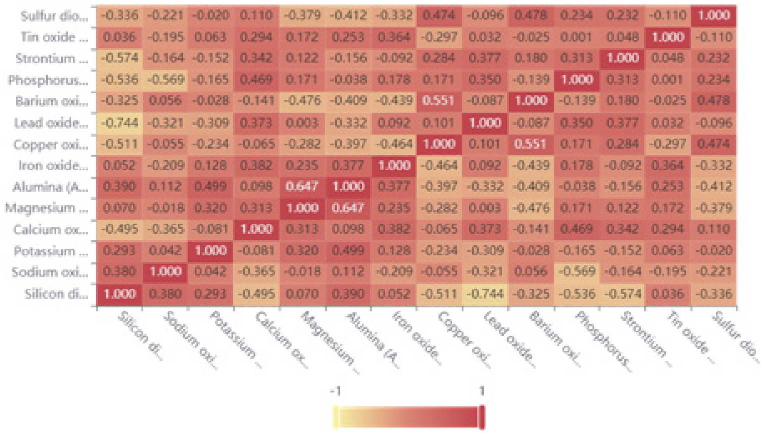

Figure 7. Correlation coefficient thermogram analysis of high potassium type

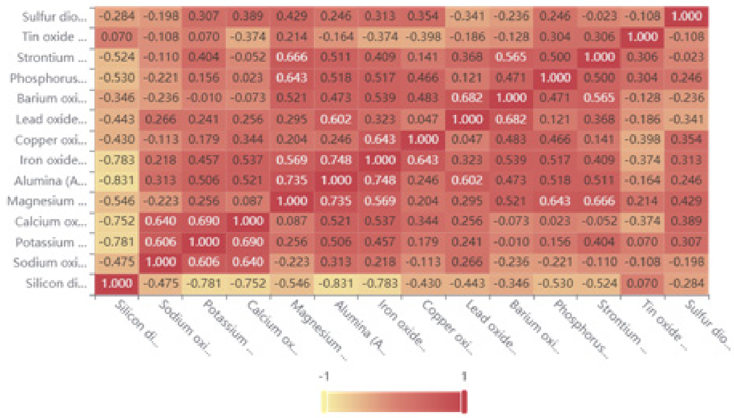

Figure 8. Correlation coefficient thermogram analysis of lead and barium type

As shown in figure 7and 8, according to the correlation coefficient thermogram, for the high potassium type, it is not difficult to find from the table that the correlation degree between alumina and iron oxide is the highest, reaching 0.748, so it is determined that these two elements have a high degree of correlation. For the lead-barium type, it is not difficult to find from the table that the correlation degree between alumina and iron oxide is the highest, up to 0.647, so it can be concluded that these two elements have a high degree of correlation.

Therefore, the high potassium type and lead-barium type are the highest correlation degree between alumina and iron oxide, so there is no strong correlation difference.

ISSN: 2688-8645

(3) Chemical differences between weathering of glass products

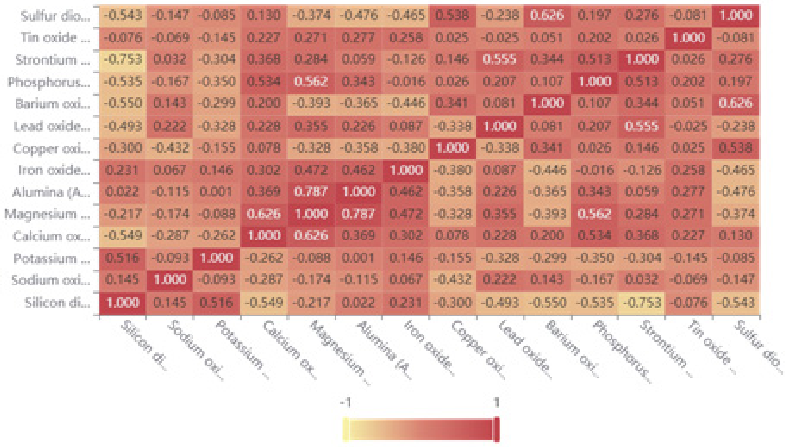

Figure 9. Correlation coefficient thermogram analysis of weathered glass products

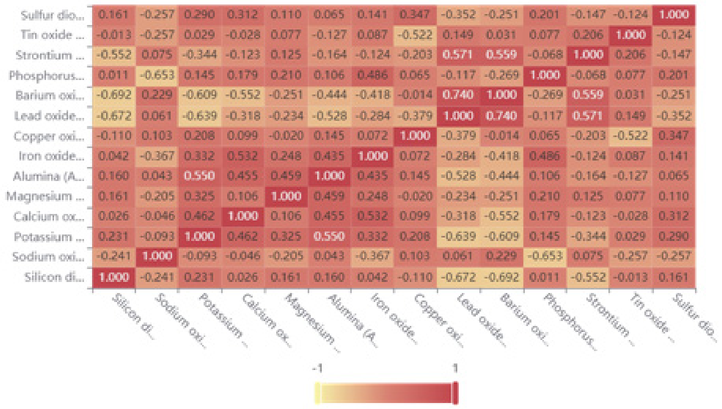

Figure 10. Correlation coefficient thermogram analysis of unweathered glass products

As shown in figure 9 and 10, according to the correlation coefficient thermogram, for weathered glass products, it is not difficult to find from the table that the correlation degree between alumina and magnesium oxide is the highest, reaching 0.626. Therefore, it is judged that these two elements have a high correlation degree. For unweathered glass products, it is not difficult to find from the table that the correlation degree between barium oxide and lead oxide is the highest, reaching 0.740, so it is concluded that these two elements have a high correlation degree.

Therefore, the correlation elements between weathered glass products and unweathered glass products are obviously different, for weathered glass products there is a clear correlation between alumina and magnesium oxide, while for unweathered glass products there is a clear correlation between barium oxide and lead oxide.

6. Conclusion

In this paper, the limited data research scope, this paper focuses on the analysis of the weathering elements of glass products content change and element species change. Based on this change, the functional relationship is constructed to describe and predict the types of some cultural relics, and then the previous state of different cultural relics and even the change of element content can be predicted.

For the study of cultural relics, the available cultural relics are extremely limited. The rough research in this paper is undoubtedly a deeper process of data mining and interpretation within a relatively small range of cultural relics data, which can greatly expand the data range and

ISSN: 2688-8645

increase the sample space for researchers. For front-line archaeologists, the correlation function analysis in this paper can be used for archaeologists to roughly check the general types and basic classification of unearthed glass relics, and lay a foundation for the next step of research.

References

[1] The SPSSAU project (2022). SPSSAU. (Version22.0) [Online Application Software]. Retrieved from https://www.spssau.com.

[2] Hosmane B S. Improved likelihood ratio tests and pearson chi-square tests for independence in two dimensional contingency tables[J]. Communications in Statistics, 1986, 15(6):1875-1888.

[3] MatthewGoldstein, EdwardWolf, WilliamDillon. On a test of independence for contingency tables[J]. Communications in Statistics, 1976, 5(2):159-169.

[4] Saroj, Kavita. Review: study on simple K means and modified K means clustering technique[J]. International Journal of Computer Science Engineering and Technology,2016,6(7): 279-281.

[5] Zhihua Zhou. Machine learning[M]. Tsinghua University Press, 2016.

[6] Xu Weichao. Review of correlation coefficient research[J]. Journal of Guangdong University of Technology, 2012,29(3):12-17.