ISSN 2616-5805 Vol. 3, Issue 2: 63-68, DOI: 10.25236/AJMS.2022.030210

A model for compositional analysis of ancient glass products based on correlation analysis

Sha Yuan

School of Mathematics and Computer Science, Yan'an University, Yan'an, 716099, China

Abstract: *In response to question 1, the key lies in the need to study the relationship between weathering and type, ornamentation and color on the surface of glass products, as well as the relationship and law between each glass type and chemical composition content. By establishing decision trees, chi-square tests, and multiple linear regression models, after eliminating two sets of invalid data from samples 15 and 17, it was concluded that the type and ornamentation of glass products have a greater influence on weathering, and the color of glass products has a smaller relationship with weathering The relationship between glass type and weathering is smaller, and the variation of chemical composition content with and without weathering is more obvious for different glass types, and the statistical pattern of these data is used to predict the chemical composition content of each glass type before weathering. For problem 2, different categories of glass need to be subclassified according to chemical composition. By establishing a decision tree model, the classification rules of glass types were found out, such as lead oxide content higher than 5.46 for lead-barium glass and lower than 5.46 for high-potassium glass, and the entropy weighting method was used to determine the weights of various chemical components for high-potassium glass with or without weathering and lead-barium glass with or without weathering, so that the chemical components with higher weights were selected to classify four types of glass: high-potassium weathered, high-potassium unweathered, lead-barium weathered, and lead-barium unweathered, respectively. The classification results were obtained by using the systematic hierarchical clustering algorithm, and the classification results of different types of glass differed, and then the results were analyzed for reasonableness and sensitivity.*

Keywords: *Correlation analysis; Decision tree model; Cardinality test; Entropy method; Spearman coefficient*

1. Introduction

Glass can be seen everywhere in our daily life, but in ancient times in China, glassware is relatively slow to develop compared to other industries, early for the production process of glass is not mature, and did not protect the glass in time, ancient glass is extremely vulnerable to environmental impact and weathering[1,2]. During the long storage process, the surface of glass products will come into contact with water and air, which will lead to a series of physical and chemical changes, and its chemical composition will also change, thus affecting the correct judgment of its category[3]. The research on the analysis and identification of the composition of ancient glass products is beneficial to the advancement of archaeological work[4]. The analysis and identification of the chemical composition of ancient products will be of great significance and value to the archaeological work in China, which attaches great importance to the protection of cultural heritage and strongly supports the development of archaeology[5].

This study intends to address the following questions.

Problem 1: By analyzing and studying the contents of the annexes, determine whether the surface weathering of glass artifacts is related to their glass type, decoration and color; combine the type of glass and the information given in Table 1 and Table 2 of the annexes to study where the pattern of the chemical composition content of glass artifact samples with or without weathering on their surface lies, and combine the weathering point detection data to predict the content of the chemical composition they contained before weathering.

Question 2: Based on the data given in the Annex, we analyze the classification pattern of high potassium glass and lead-barium glass; select the appropriate chemical composition of each category of high potassium glass and lead-barium glass for the Asia-Pacific classification, and give the specific classification method and results of the Asia-Pacific distribution, and then analyze the rationality and

ISSN 2616-5805 Vol. 3, Issue 2: 63-68, DOI: 10.25236/AJMS.2022.030210

sensitivity of the classification results.

2. Model building and solving

2.1 Modeling and solving of Problem 1

2.1.1 Data pre-processing

From the information known in the title, the data with the summation of component proportions between 85% and 105% were considered as valid data. We summed up the contents of 14 chemical components of each sample by combining the information in the attached Table I and Table II, and observed that the two groups of data numbered 15 and numbered 17 did not meet the requirements and were invalid data, and data preprocessing was taken to eliminate them.

2.1.2 Modeling and solving of the first subquestion

(1) Modeling and solution

For the analysis of the relationship between the surface weathering of glass artifacts and each attribute, the model assumptions are used, and the ornamentation ABC of glass artifacts are assumed to be A1, A2, A3, and their types and colors are also synthesized as such, and the results are obtained by constructing a decision tree model, as shown in Figure 1 below.

Each element has its corresponding regression coefficient and t-test value (in parentheses). A general t-test value of less than 0.05 indicates a significant difference, indicating that the independent variable has an effect on the dependent variable. Through observational analysis, it was found that the difference between the decoration and color of glass artifacts and their weathering was not significant, while the difference between the type and their weathering was significant.

Regression Table

	D2
A1	0.000
	(.)
A2	1.018***
	(4.862)
A3	0.257*
	(1.805)
B1	0.000
	(.)
B2	-0.797***
	(-5.424)
C1	-0.322
	(-0.848)
C2	0.232
	(0.801)
C3	0.030
	(0.088)
C4	0.000
	(.)
C5	-0.155
	(-0.427)
C6	-0.655
	(-1.347)
C7	0.381
	(1.289)
C8	0.601
	(1.585)
_cons	0.399
	(1.459)
N	54

Figure 1: Decision tree model results

(2) Model Analysis and Evaluation

The decision tree model visualizes the data, and it is easy to observe and analyze the results. When classifying the attributes, the different choices of judgment criteria also affect the classification results. There are 58 sets of data in Annex I. The number of data is small, so the decision tree model is easy to overfit and the results obtained are unstable, and the accuracy of the results obtained by this model is low, and the correlation of data attributes is easily ignored.

(3) Optimization of the model

The data in the appendix were sorted out statistically, and the missing color was recorded as white, and the discrete distribution tables of decoration, type, color and unweathered and weathered were used to determine whether the attributes were independent of each other or not.

Let the original hypothesis be that ornamentation is independent of weathering on the surface of glass artifacts, and adopt the discrete distribution fit test, and let p_i be the probability that the individuals in the total belong to A_i only, p_j be the probability that the individuals in the total belong to B_i only, and p_{ij} be the probability that the individuals in the total belong to A_i and B_i, where r denotes rows and c denotes columns In case the original hypothesis holds, these rc parameters are determined by the r+c parameters p_i and p_j, considering the two constraints $\sum_{i=1}^{r} p_i=1$ and $\sum_{j=1}^{c} p_j=1$, which are actually determined by the r+c-2 independent covariates.

$$X^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-np_{ij})^2}{np_{ij}} \tag{1}$$

When the original hypothesis holds, the above equation approximately obeys the x^2 distribution with degrees of freedom (r-1)(c-1), and at a given significance level α (0<α<1), the test rejection domain is W={ $x^2 \geq X_\alpha^2$ ((r-1)(c-1))},and by looking up the chi-square distribution table, X_α^2 ((r-1)(c-1)) is derived, which is compared with the test statistic x^2 to determine whether it is in the rejection domain, and if it is in the rejection domain, the original hypothesis is rejected and the ornamentation is considered not independent of the weathering of the glass surface, i.e., the two are correlated (Table 1).

Table 1: Result Analysis Table

Topic	Name	Unweathered	Wearthing	x^2	Revise x^2	Critic-al value
arnamentation	A	11	11	4.952	5.543	4.605 (α=0.1, n=2)
	B	0	6			
	C	13	17			
Add		24	34			
type	High K	12	6	6.875	5.997	5.024 (α=0.025, n=1)
	Lead barium	12	28			
		24	34			
color	Blue green	6	9	9.456	3.912	3,490 (α=0.9 n=)
	Wathet blue	8	12			
	Purple	2	2			
	Dark green	3	4			
	Dark blue	2	0			
	Light green	2	1			
	Green	1	0			
	Dark	0	2			
	White	0	4			
Add		24	34			

2.1.3 Modeling and solving of the second subquestion

Since there is surface weathering of lead barium but the test is for undifferentiated points, this situation needs to be taken into account. The new table combine.xlsx was finally obtained and differentially divided into five cell tables: high potassium weathered, high potassium undifferentiated, lead-barium undifferentiated, lead-barium weathered, and lead-barium weathered (not).

Visual analysis of the charts revealed that the high potassium types weathered were all B-textured and blue-green (Figure 2). From the overall data, it was found that the content of each chemical component of the high potassium glass weathered was significantly reduced, but the chemical component of the lead-barium glass weathered was significantly increased compared to that of the unweathered. A comparison of the chemical composition content of high potassium glass with and without weathering revealed that sulfur dioxide (SO2), tin oxide (SnO2), strontium oxide (SrO), barium oxide (BaO), lead oxide (PbO), and sodium oxide (Na2O), which had zero content after weathering. Comparing the

ISSN 2616-5805 Vol. 3, Issue 2: 63-68, DOI: 10.25236/AJMS.2022.030210

chemical composition content of lead-barium glass with and without weathering, it was found that the content of sulfur dioxide (SO2) and phosphorus pentoxide (P2O5) increased significantly and the content of sodium oxide (Na2O) decreased significantly before and after weathering.

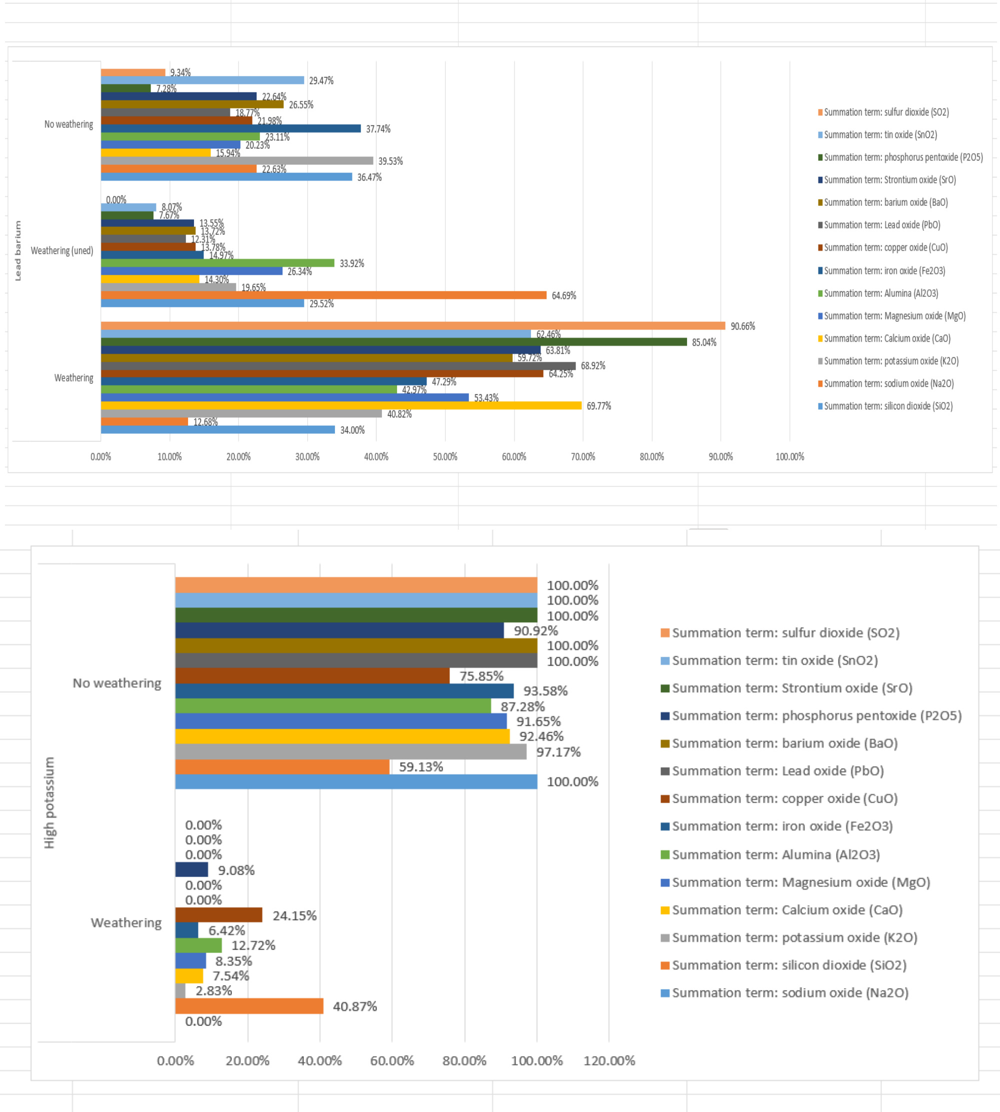

Figure 2: Visual analysis of the charts

2.1.4 Modeling and solving of the third subquestion

For the high potassium type, lead and barium type 14 kinds of components were found the maximum value, minimum value, mean value, standard deviation and variance of unweathered and weathered respectively (Table 2 and Table 3); the weathering of high potassium type, the color after weathering is all blue-green, while the content of unweathered color of blue-green oxide is 0, so the predicted weathering of its pre-weathering oxide content is 0. For the content of silica after differentiation is between 92.35 and 96.77, the content of un The differentiated content is distributed between 59.01 and 87.05, and the mean value is taken as 67.9842, potassium oxide as 9.3308, calcium oxide as 5.3325, magnesium oxide as 1.0792, aluminum oxide as 6.62, iron oxide as 1.9317, copper oxide (CuO): 2.4525, lead oxide (PbO): 0.4117, barium oxide (BaO). 0.5983, phosphorus pentoxide (P2O5): 1.4025, strontium oxide (SrO): 0.0417, tin oxide (SnO2): 0.1967, and sulfur dioxide (SO2): 0.1017.

Table 2: High potassium glass before weathering component content

Na2O	K2O	CaO	MgO	Al2O3	Fe2O3
0	9.3308	5.3325	1.0792	6.62	1.9317
PbO	BaO	P2O5	SrO	SnO2	SO2
0.4117	0.5983	1.4025	0.0417	0.1967	0.1017

ISSN 2616-5805 Vol. 3, Issue 2: 63-68, DOI: 10.25236/AJMS.2022.030210

Table 3: Composition of lead-barium glass before weathering

SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3
54.6596	1.6826	0.2187	1.3204	0.6404	4.4561	0.7365
CuO	PbO	BaO	P2O5	SrO	SnO2	SO2
1.4317	22.0848	9.0017	1.0491	0.2683	0.0465	0.1591

2.2 Modeling and solving Problem 2

2.2.1 Data pre-processing

From the known information of the topic, the data with the cumulative sum of component proportions between 85% and 105% are considered as valid data. We summed up the contents of 14 chemical components of each sample by combining the information in the attached Table I and Table II, and observed that the two groups of data numbered 15 and numbered 17 do not meet the requirements and are invalid data, so we took data preprocessing to eliminate them.

2.2.2 Establishment and solution of the first sub-question model

By fitting the data of high potassium and lead-barium before weathering with matlab, and then building a decision tree model, 1 indicates no weathering and 2 indicates weathering, to find out their laws X3 for potassium oxide (K2O) and x9 for lead oxide (PbO). The decision tree shows that before weathering, the potassium oxide of high potassium glass is higher than 3.3, the lead barium glass is lower than 3.3, the lead oxide of high potassium glass is lower than 5.46, and the lead barium glass is higher than 5.46. Using the entropy weight method, the decision matrix was constructed by taking high potassium class unweathered, high potassium class weathered, lead-barium class unweathered and lead-barium class weathered as the evaluation objects and 14 chemical components as the evaluation indexes.

$$X_{ij} = \begin{bmatrix} x11 & \cdots & x1,14 \\ \vdots & \ddots & \vdots \\ x41 & \cdots & x4,14 \end{bmatrix} \quad (2)$$

This decision matrix is normalized and denoted as W_{ij}, $W_{ij}=\frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$.

Further, the above matrix is processed to construct its non-negative matrix $W_{ij}=\begin{bmatrix} w11 & \cdots & w1,14 \\ \vdots & \ddots & \vdots \\ w41 & \cdots & w4,14 \end{bmatrix}$.

Calculate the probability, $P_{ij} = \frac{w_{ij}}{\sum_{i=1}^{n} w_{ij}}$ known by the statistical law of probability, $\sum_{i=1}^{n} P_{ij} = 1$, normalized to obtain the entropy weight of each chemical component of the glass artifact, and determine the weights of various chemical components according to the entropy. The weights of various chemical components in high potassium weathering, high potassium unweathering, lead-barium weathering, and lead-barium unweathering were calculated separately, and the top three with the largest weights were selected as the criteria for subclassification (Table 4).

Table 4: Calculation results

	High potassium weathering	High potassium is not weathered	Lead barium weathering	Lead barium is not weathered
Si02	0.0001	0.0007	0.0082	0.0032
Na20	0	0.1429	0.1678	0.0911
K20	0.1908	0.0123	0.1065	0.0649
Ca0	0.062	0.0239	0.0176	0.0325
Mg0	0.494	0.0259	0.0566	0.0451
Al203	0.0457	0.0066	0.0245	0.0176
Fe203	0.0133	0.0382	0.0649	0.0873
Cu0	0.0635	0.0258	0.0494	0.0499
Pb0	0	0.0901	0.0035	0.0068
Ba0	0	0.1197	0.0288	0.0136
P205	0.1306	0.0423	0.0309	0.0889
Sr0	0	0.0769	0.0199	0.0419
Sn02	0	0.2529	0.2322	0.184
S02	0	0.1417	0.1892	0.2732

ISSN 2616-5805 Vol. 3, Issue 2: 63-68, DOI: 10.25236/AJMS.2022.030210

2.2.3 Establishment and solution of the second sub-question model

Based on the results of solving the first sub-question, the analysis shows that the three chemical components of potassium oxide (K2O), magnesium oxide (MgO), and phosphorus pentoxide (P2O5) have a greater weight on high potassium weathering, so they are used as the subclassification criteria for high potassium weathering; similarly, the three chemical components of sodium oxide (Na2O), tin oxide (SnO2), and sulfur dioxide (SO2) are used as the subclassification criteria for high potassium no weathering, lead barium weathering, and lead-barium unweathered as subclassification criteria. A systematic hierarchical clustering algorithm was used to subclassify these four classes according to the selected chemical compositions.

3. Conclusion

This paper utilizes models such as correlation analysis, chi-square test, multiple linear regression, and decision tree to mainly investigate the variation patterns of chemical composition of ancient glass products with and without weathering, and to address the correlation relationships between surface weathering and ornamentation, type, and color of glass products, as well as the variability of the correlation relationships of chemical composition among different categories of glass. The principle of the systematic hierarchical clustering algorithm is relatively simple, the distance is easily defined, there are few restrictions, the results obtained by hierarchical clustering are more intuitive, the hierarchical relationship between different classes can be easily observed, and its algorithm is highly operable. The algorithm uses an iterative method, the optimal solution found is only a local optimal solution, and the model is unstable, which will have some influence on its results.

References

[1] Zhang Liyan, Li Hong, Chen Shubin, Li Zhongdi, Ruan Dangzhi, Xue Tianfeng, Qian Min, Fan Sijun. Simulation method of glass composition and properties [J]. Journal of Silicate, 2022, 50 (08): 2338-2350. DOI: 10.14062/j.issn.0454-5648.20220115.

[2] Ye Dingquan. High efficiency flux for making glass lithium mineral and lithium carbonate [J]. Glass fiber, 2000 (02): 42-44. DOI: 10.13354/j.cnki.cn32-1129/tq. 2000.02.013.

[3] Wang Chengyu, Tao Ying. Weathering of silicate glass [J]. Journal of Silicate, 2003 (01): 78-85.

[4] Chen Haiyan, Yu Ying, Jia Yizhen, Gu Bin. Incremental learning for transductive support vector machine [J]. Pattern Recognition, 2023,133.

[5] Xu Zhijun, Liu Tingting, Li Jianping, Yuan Fang. Prediction model of dynamic lateral pressure of grain storage silo wall based on support vector machine [J]. Research on Agricultural Mechanization, 2022, 44 (5): 9-16.