Research on the Composition and Identification of Ancient Glass Products

Xuan Lu[1,*,&] and Haoran Xiong[2,&]

[1] School of Mathematics and Statistics, Guangdong University of Technology, Guangzhou, Guangdong, 510520, China
[2] School of Automation, Guangdong University of Technology, Guangzhou, Guangdong, 510006, China

*Corresponding author's e-mail: 3221006905@mail2.gdut.edu.cn

Abstract. Because ancient glass is easily weathered by the burial environment, judgments about its type are easily affected. In addition, due to the influence of the surrounding environment, the internal elements of ancient glass will undergo a large number of exchanges with environmental elements, thus causing changes in the proportions of various components inside ancient glass. However, the research that has appeared so far has not conducted a detailed study on the connection between changes in the chemical composition of glass and glass-related properties. Consequently, in order to better conduct relevant analysis of ancient glass products, this paper uses Spearman's rank correlation coefficient and independent sample t test to analyze the weathering of glass. The results of the problem solved in this paper demonstrates that the effect of the modeling method in this paper is outstanding and significant.

Keywords: Spearman rank correlation coefficient, Pearson chi-square test, independent samples t-test, random forest model, hierarchical clustering analysis , Pearson correlation coefficient, Kruskal-Wallis test

1 Introduction

In related research on glass, it can be found that due to factors such as production processes, there are a variety of chemical substances in various types of glass. Under the influence of weathering, the chemical components inside the glass will exchange with environmental elements, so that the internal chemical elements get changed. Therefore, how to analyze the complicated chemical composition of ancient glass has encountered difficulties. However, the current emerging research focuses on the origin of glass, manufacturing technology, etc. And these studies rarely conduct an exhaustive study of glass chemical composition. For that reason, new relevant research needs to be proposed urgently.

Studying the chemical composition is very important to the study of glass. Numerous researchers have scrutinized related areas in this domain. For example, [1]

K. Subramaniyam et al. (eds.), *Proceedings of the 3rd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2024)*, Atlantis Highlights in Computer Sciences 21,
https://doi.org/10.2991/978-94-6463-419-8_14

designs a chemical composition analysis system based on the random forest model to analyze glass composition. On the basis of weathering change rates, [2] summarized and generalized relevant solution models for chemical content. In [3], the author of this paper utilizes principal component analysis (PCA) to extract the main influencing factors for classifying high potassium and lead-barium glass. Subsequently, it applies spectral clustering to categorize samples into sub[1]classes, enhancing the identification of ancient glass products. In [4], the research utilizes logistic regression and Bagged Trees for classifying ancient glass into high potassium and lead-barium types. Cross-validation and parameter optimization enhance model robustness, with barium oxide being a key predictor. Grey correlation analysis highlights silica's influence on high potassium glass composition. In [5], the paper employs logistic regression, decision trees, and random forests for classifying ancient Chinese glass artifacts. It uses grid search for hyperparameter optimization and Spearman correlation analysis to study the chemical composition correlations[6]. These papers provide assistance in studying the chemical composition of glass [7].

Although there are many studies on the analysis of glass components and classification of glass types, these studies are still lacking, and many issues remain unresolved. Many ancient glass products have weathered under the influence of the outside world. Therefore, studying the laws of weathering of glass products is also an important means of studying glass. In order to better study the related conditions of glass weathering, the relationship between the degree of glass weathering and its type, decoration, and color also needs to be investigated. In addition, in order to better analyze the chemical composition of glass, the classification rules of glass need to be fully analyzed and unknown glass products classified according to appropriate methods. Furthermore, to better study the properties of glass, the correlational relationships among its chemical components need to be analyzed, and the differences in the correlation of chemical components between different categories should be compared[8]. To address the aforementioned issues, this paper employs the following methods to strive for better research outcomes:

1. To gain a clearer understanding of the weathering condition of glass., this paper conducts a correlation analysis of the relationships between glass types, decorations, and colors. In this context, the paper employs the Spearman rank correlation coefficient[9] to assess the relationships between the features and uses the Pearson chi-square test to analyze their differences. Simultaneously, to explore the statistical regularities of weathering chemical composition content on the surface of cultural relics, this paper applies the independent samples t-test[10] to investigate the statistical regularities of the presence of weathering chemical composition in the analyzed cultural relics. By examining the significance and differences in chemical composition between samples with and without weathering across different types, statistical patterns are derived.
2. Potassium-rich glass and lead-barium glass are two major types of glass. To explore the classification patterns of these glasses, this paper employs a random forest model for determination. Moreover, there are subcategory divisions

[&] Xuan Lu and Haoran Xiong are co-first authors of the article.

within each type of glass. The paper uses hierarchical clustering analysis to delineate specific subcategories. In the subcategory division, the paper also applies the elbow rule to analyze the rationality and sensitivity of the subcategory divisions for glass artifacts.

2 Methology

Firstly, this paper uses the Spearman rank correlation coefficient to assess the relationships between various features of glass artifacts and employs the Pearson chi-square test for analysis. It also investigates the statistical regularities of weathering chemical composition on the surface of cultural relics using the independent samples t-test. Secondly, the paper employs a random forest model and hierarchical analysis for detailed classification of glass models.

2.1 Analyzing the weathering condition of glass.

The Spearman rank correlation coefficient [11] is an important method for correlation analysis. The Spearman rank correlation coefficient, also known as Spearman's rho, is a non-parametric measure of the strength and direction of the association between two ranked variables. It is particularly useful when the data is not normally distributed or when the relationship between variables is not linear. The method assigns ranks to the data points and then calculates the correlation coefficient based on these ranks.

When performing the Spearman rank correlation coefficient, each data point needs to be assigned a rank based on its position in the combined data set. If there are ties, assign the average rank to the tied values. After squaring the rank differences and the sum of ranks, calculate the total number of paired data points and apply the formula:

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

Where X and Y are two sets of data, and di is the rank difference between Xi and Yi.

To delve deeper into the properties of glass and investigate the phenomena related to glass weathering, this paper also employs an independent samples t-test[10] to analyze the statistical regularities of the presence of weathering chemical composition content on the surface of cultural relics.

2.2 Categorical operation

2.2.1Random Forest Models.

In this model, the robustness of the model can be improved by constructing the tree by randomly selecting training samples in the same decision tree, so as to suppress overfitting to the training samples. In this paper, the prediction of the classification problem is determined by a voting mechanism, where each decision tree votes for a category, and the final prediction is the category with the most votes. Also because of the randomness that exists in the random forest model, it increases the diversity of the model, its immunity to perturbation and generalisation ability, making the predictions more convincing. The random forest model i.e. in (1) $\hat{y}$ is the result of chemical composition categorisation, where $f_j(x)$ is the first j classification result of the decision tree, and m is the number of decision trees, and I is the indicator function

$$\hat{y} = \text{argmax}i\sum j = 1^m I(f_j(x) == i) \tag{1}$$

2.2.2systematic clustering.

This method creates a tree structure (dendrogram or dendrogram) in which each data point is a separate cluster, and then gradually merges these clusters to form larger clusters until all data points are contained in one cluster. This hierarchy can be seen as a kind of hierarchical division of data similarity.

For this clustering method firstly each data point needs to be considered as a separate cluster and the distance between each pair of clusters is calculated. For cluster distance imputation different distance metrics can be used such as Euclidean distance, Manhattan distance, correlation coefficient etc. For the above process one can assume n a number of data points, using a $n \times n$ similarity matrix of S .Among them, s_{ij} represents the similarity between the i and j data points, forming matrix (2). Find the smallest similarity value in S_k, that is, the most similar value s_{ij} of the i and j clusters s_{ij} , merge these two clusters into a new one. Finally iteratively generate hierarchical clusters and then analyse the data[12].

$$S_0 = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix} \tag{2}$$

In this paper, we use the Euclidean distance (3) to solve for the distance between two two samples.

$$\text{Euclidean distance: } d(\vec{x_i}, \vec{x_j}) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \tag{3}$$

3 Experimental Studies

3.1 Dataset and Comparison Algorithm

- Data Availability Statements: In this paper, the datasets are obtained from Problem C of the 11th China Undergraduate Mathematical Contest in Modeling (CUMCM) availably.
- Dataset Detail: The dataset consists of three attachments, with detailed information for each attachment as shown in Table 1:

Table 1. Dataset detail

Appendix Name	Feature
Appendix 1	Cultural relic number, decoration, type, color, surface weathering
Appendix 2	Cultural relic sampling location, content of various chemical components
Appendix 3	Weathered or not, content of various chemical components

3.2 Results and Analysis

3.2.1 Analyzing the weathering condition of glass.

The research results are shown in Table 2:

Table 2. Results of Spearman's correlation coefficient

	weathering	Type	Decoration	Color
weathering	1.000(0.000***)	0.344(0.008***)	0.116(0.384)	-0.116(0.385)
Type	0.344(0.008***)	1.000(0.000***)	-0.357(0.006***)	0.529(0.000***)
Decoration	0.116(0.384)	-0.357(0.006***)	1.000(0.000***)	-0.473(0.000***)
Color	-0.116(0.385)	0.529(0.000***)	-0.473(0.000***)	1.000(0.000***)

Analysis of the obtained correlation coefficient table and heatmap indicates that color has a strong correlation with type; the correlation of surface weathering with color and decoration is the weakest, while it has the strongest correlation with type.

The results of the independent samples t-test are in Table 3:

Table 3. Results of independent samples t-test

Name	Variable Value	Mean	Standard	t	P	Mean Difference	Cohen's d
PbO	Weathered	0.412	0.589	1.686	0.111	0.412	0.843
	Unweathered	0	0.000				
	Total	0.274	0.514				
SiO2	Weathered	67.984	8.755	-7.095	0.000***	25.979	3.547
	Unweathered	93.963	1.734				
	Total	76.644	14.467				
P2O5	Weathered	1.402	1.434	1.879	0.079*	1.122	0.940
	Unweathered	0.28	0.210				
	Total	1.028	1.281				
SO2	Weathered	0.102	0.186	1.322	0.205	0.102	0.661
	Unweathered	0	0.000				
	Total	0.068	0.157				
SnO2	Weathered	0.197	0.681	0.696	0.496	0.197	0.348
	Unweathered	0	0.000				
	Total	0.131	0.556				
K2O	Weathered	9.331	3.920	5.391	0.000***	8.788	2.696
	Unweathered	0.543	0.445				
	Total	6.402	5.308				
MgO	Weathered	1.079	0.676	3.011	0.008***	0.882	1.505

	Unweathered	0.197	0.306				
	Total	0.785	0.712				
CaO	Weathered	5.333	3.092	3.461	0.003***	4.463	1.731
	Unweathered	0.87	0.488				
	Total	3.845	3.308				
SrO	Weathered	0.042	0.048	2.077	0.054*	0.042	1.038
	Unweathered	0	0.000				
	Total	0.028	0.044				
Na2O	Weathered	0.695	1.287	1.303	0.211	0.695	0.651
	Unweathered	0	0.000				
	Total	0.463	1.089				
Fe2O3	Weathered	1.932	1.667	2.411	0.028**	1.667	1.206
	Unweathered	0.265	0.069				
	Total	1.376	1.566				
CuO	Weathered	2.452	1.660	1.210	0.244	0.890	0.605
	Unweathered	1.562	0.935				
	Total	2.156	1.492				
BaO	Weathered	0.598	0.982	1.470	0.161	0.598	0.735
	Unweathered	0	0.000				
	Total	0.399	0.842				
Al2O3	Weathered	6.62	2.492	4.393	0.000***	4.690	2.197
	Unweathered	1.93	0.964				
	Total	5.057	3.077				

3.2.2Categorical operation.

This paper trains the model using the existing data through a random forest model, allowing the determination of the feature importance for each chemical component. Visualizing the feature importance of each chemical component results in Table 4.

Table 4. the feature importance of each chemical component results

Chemical composition	Feature importance	Chemical composition	Feature importance
SiO2	0.13	Cu2O	0.02
Na2O	0	PbO	0.23
K2O	0.21	BaO	0.15
CaO	0.03	P2O5	0.02
MgO	0.01	SrO	0.13
Al2O	0.04	SnO	0
Fe2O	0.02	S2O	0

Observation reveals that lead oxide has the greatest impact on the determination of glass type, followed by potassium oxide; the least influential are sodium oxide, tin oxide, and sulfur dioxide.

4 Conclusion

This research paper investigates the composition and identification of ancient glass products, focusing on the weathering effects on the classification and properties of glass artifacts. The study employs Spearman's rank correlation coefficient and independent sample t-tests to analyze the weathering of glass, and uses random forest models and cluster analysis to classify glass types and analyze their chemical composition. The results demonstrate the effectiveness of the modeling methods used in this paper, providing valuable insights into the chemical composition and weathering patterns of ancient glass.

References

1. Atul Adya, Paramvir Bahl, Jitendra Padhye, Alec Wolman, and Lidong Zhou. 2004. A multi-radio unification protocol for IEEE 802.11 wireless networks. In Proceedings of the IEEE 1st International Conference on Broadnets Networks (BroadNets'04) . IEEE, Los Alamitos, CA, 210–217.
2. X. Li, J. Li, X. Yu and R. Zhang, "Analysis and Prediction of Glass Product Composition by Using Control Variable Method," 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), Xiangtan, China, 2023, pp. 1640-1645.
3. Y. Chen, J. Chu, T. Wang and S. Yu, "Composition Analysis and Identification of Ancient Glass Products Based on Spectral Clustering Algorithm," 2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA), Changchun, China, 2023, pp. 302-307.
4. J. Li and Y. Zhu, "Composition analysis and identification of ancient glass objects based on neural network models," 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2023, pp. 43-46.
5. J. Huang, X. Zhang, J. Zhao and Z. Cui, "Research on the Classification of Glass Products Based on Machine Learning Algorithm," 2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA), Changchun, China, 2023, pp. 11-16.
6. Yohei MISHINA, Ryuei MURATA, Yuji YAMAUCHI, Takayoshi YAMASHITA, Hironobu FUJIYOSHI, Boosted Random Forest, IEICE Transactions on Information and Systems, 2015, Volume E98.D, Issue 9, Pages 1630-1636
7. Fang Tao, Yao Yingsheng. A Test of Normality for Observation Series with Inhomogeneous Precision Using the Shapiro-Wilk Method [J]. Mining Surveying, 1990(02): 17-19.
8. Zhang Linquan. The Principle of Kruskal-Wallis Test for Multiple Independent Samples and Its Empirical Analysis [J]. Journal of Suzhou University of Science and Technology (Natural Science Edition), 2014, 31(01): 14-16+38.
9. E. Szmidt and J. Kacprzyk, "The Spearman rank correlation coefficient between intuitionistic fuzzy sets," 2010 5th IEEE International Conference Intelligent Systems, London, UK, 2010, pp. 276-280,
10. Kim HY. Statistical notes for clinical researchers: the independent samples t-test. Restor Dent Endod. 2019;44(3):e26. Published 2019 Jul 17.
11. Ostertagova, Eva & Ostertag, Oskar & Kováč, Jozef. (2014). Methodology and Application of the Kruskal-Wallis Test. Applied Mechanics and Materials. 611. 115-120.
12. SHAIK, Anjaneyulu Babu; SRINIVASAN, Sujatha. A brief survey on random forest ensembles in classification model. En International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2. Springer Singapore, 2019. p. 253-260.