Composition analysis and identification of ancient glass products

Minghao Wang [#, *], Ruochen Cui [#], Yulong Xue [#]

School of National University of Defense Technology, Changsha, China, 410073

* Corresponding author: 1649177887@qq.com

Abstract. Ancient glass chemical composition is diverse, and because of the environmental impact of weathering, the proportion of its composition will change, affecting the judgment of the glass category. Therefore, according to the relevant data, this paper queries a large number of literature, through reasonable statistical analysis to find the classification basis, and studies all kinds of glass relics before and after weathering, in order to play an important role in the composition analysis and identification of ancient glass products.To define the standard of correlation, so as to better explain the correlation degree of the respective chemical composition of the two types of glass; then, to compare the difference between the two types of glass in the above correlation relationship. The comparison model of individual differences was then established, and the gap between the two individual correlation coefficient was calculated from the absolute value distance, and compared with the defined evaluation criteria, which concluded that the glass of the two categories had significant individual differences in each pair of content of chemical composition. For example, both potassium oxide and calcium oxide are chemical components, with the absolute value distance of 1.319, and the difference level is very significant.

Keywords: Multiple linear regression, System clustering, Logit model, SVM, Spearman's correlation coefficient.

1. Introduction

The main chemical composition of ancient glass products is silica. But other chemicals are diverse. When refining, add different flux, the refined glass contains different main components is also, mainly divided into: lead-barium glass and potassium glass.

Ancient glass is susceptible to environmental influence and weathered, with a large number of internal and external elements exchanged, leading to the change of the proportion of its composition and affecting the judgment of the glass category. Therefore, it is of great significance for the ancient glass composition obtained through chemical composition and other detection means to judge the category of glass.

Through the classification information of cultural relics given in the attached form 1, analyze the relationship between the glass surface, the glass type, pattern and color, according to the form 2, and predict the content of weathering according to the weathering point monitoring data.

Analysis the classification basis of high potassium glass and lead-barium glass according to the decoration, color and chemical composition content of Annex Table 1 and 2, find out the appropriate chemical composition to further complete the subclassification, give specific division methods and results, and conduct rationality and sensitivity analysis.

Based on the study of high price and lead-barium glass classification law, the classification method is further explored. According to the chemical composition content of the known category glass data information, analyze the chemical composition of the unknown category glass relics in form 3, determine the type of glass, and analyze the sensitivity of the classification results.

Firstly, for the two types of glass relics samples, the correlation between the internal chemical components of the two types of glass is studied to determine whether they, and then compare the differences between the two types of glass chemical composition.

2. Problem analysis

For question 1, it can be divided into three small questions: First, the relationship between the surface weathering of glass cultural relics and its decoration, color and type is analyzed through the data of form 1. Consider the qualitative relationship between surface weathering and the three variables through visual processing, frequency analysis and square inspection. Second, the statistical analysis of the surface weathering and chemical composition content in the basis of the mean and variance of the weathered and weathered glass.

For question two, it can be divided into two small questions: first, according to the attached data, including the content of a chemical composition, analyze the classification rules of high potassium glass and lead-barium glass. The data of weathered and unweathered glass are considered separately, considering the qualitative description of the classification rules by drawing the chemical content of high potassium and lead-barium glass and establishing evaluation indicators; second, the subclassification of glass in each category. Since it is not known that it should be divided into several subclasses, the systematic clustering using Euclidean distance should be divided based on the actual classification situation in ancient China. Finally, the sensitivity analysis is conducted by changing the distance calculation formula in the clustering algorithm, and the rationality analysis is compared with the actual professional division.

For problem 3, analyze the chemical composition of the unknown category of glass relics in the attached form 3 and judge their glass type. The prediction requires the regression or learning of the known data first, so the Logit regression and SVM classification models are selected. After comparing the regression accuracy, the best unknown glass is predicted. Finally, perturbations to the input parameters can be added using Matlab's rand () function for sensitivity analysis.

According to question 4, the relationship between the respective chemical composition content of high potassium and lead-barium glass was analyzed, and the difference between the two types of glass pairs was compared. Consider using the correlation coefficient to describe the relationship between the chemical components, and customize the correlation criteria to describe the strength of the component correlation; the overall difference can be analyzed by the paired sample t-test or the individual difference by the absolute value distance of the correlation coefficient.

3. Model hypothesis

Hypothesis 1: Suppose that the null hypothesis can be rejected when the statistical significance is less than 0.05.

Rausibility: When the P value is less than 0.05, it can be considered that the probability of the difference caused by chance is less than 5%. The difference is statistically significant, and the difference cannot be ignored, so the null hypothesis [8] cannot be accepted.

Hypothesis 2: Suppose that the subclassification can be divided on the basis of the content of its main components.

Rationality: The main classification criteria of glass in ancient China were different in different dynasties, but they still focused on the content of the main chemical components such as silica, lead oxide and barium oxide, such as: high silica glass, low lead glass and other classification methods.

Hypothesis 3: The chemical composition content data of a pair of weathered and undifferentiated glass with the same color, decoration, and category can be used as the basis for the regression of the prediction model.

Rationality: According to the actual experience and the observation of the attachment data, the chemical composition content of the glass with the same color, pattern and type is relatively similar, so it can be roughly regarded as the chemical composition content before and after the weathering of the same piece of glass.

4. Model establishment and solution

4.1. The establishment and solution of the Problem-1 model

The figure can qualitatively analyze that the distribution of the surface weathering of cultural relics is related to the decoration, type and color. By observing the weathering frequency map of decoration, the weathering frequency of decoration B is significantly higher than that of decoration A and decoration C, while the frequency of decoration A and decoration C is similar, so the decoration should have a certain impact on weathering. By observing the weathering frequency map of the type, the weathering frequency of lead and barium is significantly higher than that of high potassium, and the type should have an obvious impact on the weathering system. By observing the weathering frequency map of the color, the frequency of green, dark blue and black is obviously too low or too high, and the number of samples is 1,2 and 2, respectively, and the number of samples is too small to have no reference value. Except for the three colors of green, dark blue and black, the weathering frequency of the other colors is relatively similar, and the color should have less impact on the weathering.

Obult: For glass type, P=0.020 <0.050, so the null hypothesis is rejected, surface weathering and type data are correlated with 98% probability. As for decoration and color, it is not significant, and there is no significant correlation between surface weathering and color data of decoration and color.

To better compare the differences between the chemical content of weathered and unweathered glass, Min-Max standardizes each chemical content data in Form 2.

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Through form 2 observed that even if the weathered glass sample, the composition content fluctuates in a wide range, considering the possible detection means, weathering time and other factors, should find the larger data points before the analysis of statistical rules, in order to better find the statistical rules, so the standardized data is divided into four categories according to glass type and weathering, draw the corresponding box diagram figure, find the outlier and replace the average.

Considering that the types of glass are divided into high potassium glass and lead-barium glass, the element content of lead and barium glass is higher, while potassium glass has higher potassium content, so the type of glass will seriously affect the proportion of chemical composition of cultural relics. According to the combination of weathering and unweathering, type of high potassium and lead and barium, cultural relics are divided into four categories: high potassium weathering, high potassium unweathering, lead and barium weathering, lead and barium unweathering. Draw a scatter plot of their chemical composition content ratio, as shown in Figure 1, with high-potassium glass on the left and lead-barium glass on the right.

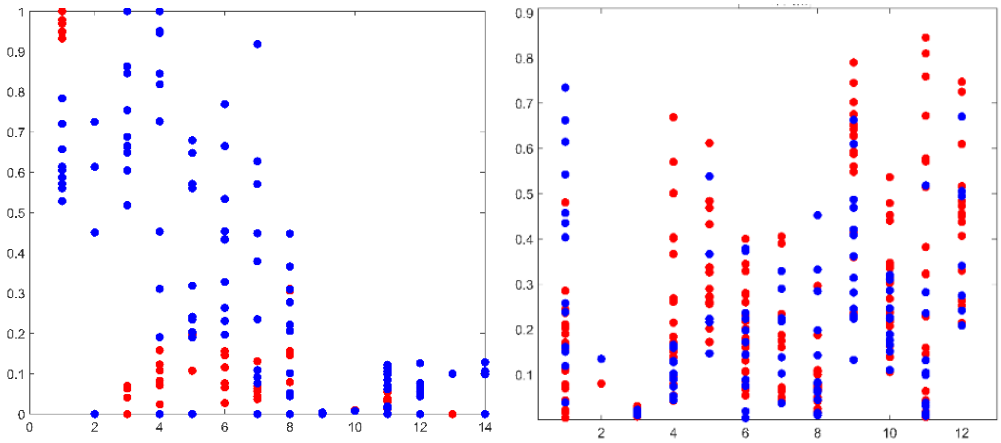


Figure 1. Sporter plot of the proportional chemical content of high potassium and lead-barium glass

As can be seen from the figure, when considering the high potassium and lead-barium glass separately into consideration, that is, after removing the influence of glass type on the cultural relic surface composition content, the composition content of glass before and after weathering has changed more significantly.

In general, the mean content and variance of each chemical composition changed before and after glass weathering. In addition, it was observed that for different types of glass, the chemical composition content on the surface of cultural relic samples before and after weathering was different, so they were analyzed separately.

For lead and barium glass, after weathering, the relative content of the components are: calcium oxide, iron oxide, copper oxide, strontium oxide; the relative content of the components are: silica, magnesium oxide, lead oxide, tin oxide, sulfur dioxide; the other components are not obvious changes.

For high potassium glass, after weathering, the relative content of significantly increased components are: silica, lead oxide; the relative content of significantly reduced components are: sodium oxide, potassium oxide, calcium oxide, magnesium oxide, alumina, iron oxide, copper oxide, phosphorus pentoxide, tin oxide, sulfur dioxide; the other components are not obvious changes.

For lead-barium glass, the chemical components obviously related to its weathering process are: silica, potassium oxide, calcium oxide, magnesium oxide, iron oxide, alumina, lead oxide, copper oxide, barium oxide, phosphorus pentoxide, strontium oxide.

For high-potassium glass, the chemical components obviously related to its weathering process are:

Therefore, a multiple linear statistical regression model was established:

$$y_i = \alpha_i + \sum_{j=1}^{N} \beta_j x_{ij} + \varepsilon_i, \varepsilon_i \sim N\left(0, \sigma^2\right) \tag{2}$$

Its loss function is:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} \left(y_i - \sum_{j} \beta_j x_{ij} \right)^2 \right\} \tag{3}$$

$$L^{OLS}(\beta) = \left\{ \sum_{i=1}^{N} \left(y_i - \sum_{j} \beta_j x_{ij} \right)^2 \right\} = \|Y - X\beta\|^2 \tag{4}$$

The least squares method has certain limitations. When there are many independent variables in the regression model, it will lead to unrelated redundant variables in the model that are found to have significant predictive effects. The results obtained by the model may only apply to the current sample and cannot be generalized to the overall [4], thus failing to make accurate predictions.

Therefore, it is more practical to obtain the regression coefficients at the cost of a small reduction of accuracy by abandoning the unbiased nature of the least squares method, so consider the Lasso method of introducing the penalty function.

4.2. The establishment and solution of the Problem-2 model

According to the decoration and color of a glass can be roughly judged according to the type of glass. This paper first analyzes the proportion of the decoration and color of high-potassium glass and lead-barium glass.

From the figure, it can be seen that the colors and patterns of different glasses are quite different. Although, through the patterns and colors, the glass classification can be preliminarily speculated. For example, type B decoration can speculate that it belongs to high potassium glass; and light blue can speculate that its large probability belongs to lead barium glass, etc.

From the scatter plot of the chemical composition, lead-barium glass is significantly higher than high-potassium glass for lead oxide, barium oxide and strontium oxide, while high-potassium glass is significantly higher than lead-barium glass, especially for weathered glass. Therefore, if the glass contains more lead barium, it can be speculated that it belongs to lead barium glass. However, for other chemical components, the difference between high-potassium glass and lead-barium glass content is not very obvious, and it cannot be used as the classification rule of glass.

Based on the analysis of chemical composition law, it is found that silica, lead oxide, barium oxide and strontium oxide have the greatest influence on the classification rule. Therefore, the self-built evaluation indicators of these four variables are used to more directly classify the two glass categories of high potassium and lead-barium.

Define the evaluation indicators:

$$Q = \frac{\sqrt[3]{x_{Pb} + x_{Ba} + x_{Sr}}}{x_{Si}} \tag{5}$$

Through the literature search, the ancient Chinese glass is roughly divided into: sodium and calcium glass, aluminate glass, high silica oxygen glass, lead-silicate glass, phosphate glass, lead-free glass, high silica oxygen glass, lead-barium silicate glass [1]. In this paper, lead-barium glass is divided into four subcategories: high silicon oxygen, high lead, high barium and low lead; high potassium glass is divided into potassium-calcium, high potassium and high silicon oxygen, which is included in ancient Chinese glass types, which proves the rationality of the classification results.

4.3. The establishment and solution of the Problem-3 model

Considering that only two high potassium and lead barium glass relics are classified, and the classification result can be regarded as a 0-1 binary variable as the explained variable, it was decided to use Logit regression to judge the category of unknown glass relics. Logit regression is a very important dichotomy algorithm that can effectively solve the difficulties in linear regression fitting.

Given the probability that the explanatory variable takes 1:

$$p(y=1|\ x;\theta) = \sigma\left(\theta^T x\right) = \frac{1}{1+e^{-\theta^T x}} \tag{6}$$

Where $\theta^T x = \theta_0 + \sum_{i=1}^{m} \theta_i x_i$, θ is the parameter to be estimated, and the parameter is usually estimated by the maximum likelihood method to write the overall probability density function:

$$p(y|\ x;\theta) = \sigma\left(\theta^T x\right)^y \left(1-\sigma\left(\theta^T x\right)\right)^{1-y} \tag{7}$$

The Logit regression achieved 96.9% accuracy and high prediction accuracy, proving that the model can be applied to judge the category of unknown categories of glass relics.

The basic model of SVM (Support Vector Machine) defines the linear classifier with the largest interval on the feature space, which finds a "maximum interval" in the desired sample space based on the training set, separating the samples of different categories from [7]. It is usually used to conduct pattern recognition as well as classification, so consider using the machine learning of SVM to solve the binary classification learning of glass types.

Also using the Matlab substitution into Form 2 data, the calculated support vector results are shown in the attachment: Support Vectors. xlsx. The accuracy of its regression is 100%.

Logic regression is relatively simple and easy to understand for models, especially for large-scale linear classification. The understanding and optimization of SVM is relatively complicated; SVM

learns the classifier by only considering the support vector, or the few points most relevant to the classification. However, the logistic regression greatly reduces the weight of the points far away from the classification plane through the non-linear mapping, and relatively improves the weight [9] of the data points most related to the classification.

Overall, when the perturbation is within the normal range, the developed Logit regression model presented here is more stable than the SVM model model, and is more insensitive to the perturbations in the input parameters. Therefore, this paper chose to use the Logit regression model.

4.4. The establishment and solution of the Problem-4 model

From the calculation result of the correlation coefficient, the chemical composition of lead-barium and high-potassium glass can be analyzed. For example, silicon dioxide and copper oxide are strongly correlated; while strontium oxide and alumina have little correlation.

By observing the correlation coefficient matrix of the two types of glass, we can preliminarily observe that the chemical composition correlation between the two types of glass has some difference, but further difference test is still needed. Because it is a small sample, consider the t-test for a significant difference in the mean of the two independent samples of the variables (chemical composition) columns corresponding to the two types of glass.

The Wilker statistics for each correlation coefficient column were calculated using SPSS and the corresponding P values are detailed in Appendix 4, and are compared with 0.05. If greater than 0.05, the null hypothesis cannot be rejected, meaning that the variable is considered to obey a normal distribution.

It is found that the data in both glass correlation coefficient columns are normal distributed.

However, the lack of obvious difference in the column mean of the correlation coefficient can only prove that the overall average level of the correlation degree between each chemical composition of the two glasses is relatively similar, and can not explain the correlation coefficient between the univariates, that is, the difference of the correlation relationship of each chemical composition. Therefore, consider further, with each correlation coefficient as the main body, to analyze the difference of each chemical composition association relationship between the two categories of glass.

Considering the difference in the correlation coefficients of each pair of chemical components, the correlation coefficient matrix of the two types of glasses was subtracted from the corresponding positions to the absolute value, i. e:

$$\rho^{(kj)} = \left|\rho_{Lead\ barium}{}^{(kj)} - \rho_{hyperkalemia}{}^{(kj)}\right| \tag{8}$$

It can be clearly seen from the table that when the correlation coefficient of the content of each pair of chemical components is considered separately, the difference between the two types of glass is more significant, especially the difference between potassium oxide and silica, and potassium oxide and calcium oxide.

5. Conclusion

Before the problem analysis, the data were standardized, eliminated the outliers, and classified the analysis, so that the subsequent model solution process was simplified and the results were more accurate. When solving the multiple linear regression model, the least squares and Lasso algorithms are used to solve the problem of overfitting in the regression process. The stability of the proposed model is verified by introducing the interference of the input parameters and changing the clustering distance. In this paper, qualitative analysis and quantitative analysis. This paper innovatively establishes the evaluation index and the individual difference comparison model, which has the originality. Due to the few data given by the topic, the accuracy of the procedure of performing parameter testing and Lasso regression cannot be well guaranteed. In the clustering process, the choice of the classification number is somewhat subjective, which increases the error caused by the

model in the classification. In order to simplify the model, the effects of fewer compounds (such as sulfur dioxide, etc.) were ignored in the analysis, resulting in the problem failed to achieve a comprehensive analysis.

References

[1] Gan Fuxi et al. The Development History of Glass Technology in Ancient China [M]. Shanghai: Shanghai Science and Technology Press, 2016.

[2] Wang Chengyu, Tao Ying. Weathing of the silicate glass [J]. Journal of Silicate, 2003, (1): 78 - 86.

[3] Written by Gao Mingqin. Glass weathering and its prevention method [M]. Wuhan: Wuhan University of Technology Press, 1995.

[4] Zhang Li today, Wei Xiyan, Lu Jiaqi, Pan Junhao. Lasso regression: from interpretation to prediction [J]. Progress in Psychological Science, 2020, 28 (10): 1777 - 1791.

[5] Jiang Qiyuan, Xie Jinxing, Ye Jun. Mathematical Model (3rd Edition) [M]. Beijing. Higher Education Press, 2003; 274 - 324.

[6] Zhang Yonglai, Zhou Yaojian. Summary of clustering algorithms [J]. Computer Applications, 2019, 39 (7): 1869 - 1882.

[7] Jike, Zhao Jianhui, Ren Xinhui, Bai Yimiao. A Support Vector Machine Review [J]. CD-ROM Technology, 2007, (2): 4 - 5.

[8] Liu Xibo. Probability Theory and Mathematical Statistics [M]. Beijing: Beijing University of Posts and Telecommunications Press, 2020.

[9] By Xue Yunfeng. Deep learning practices are based on Caffe's resolution [M]. Beijing: China National Machinery Industry Press, 2018.

[10] Zhang Liyan, Li Hong, Chen Shubin, Li Zhongdi, Ruan Chi, Xue Tianfeng, Qian Min, Fan Si Shijun. Method of composition and properties of glass [J]. Journal of Portland, 2022, Volume 50 (8): 2338 - 2350.