

Analysis and identification of glass components based on independent sample t-test and systematic clustering

Shuaiteng Liu*

Shijiazhuang Tiedao University, Shijiazhuang China

*Corresponding author: 934871477@qq.com

Abstract: With the development of archaeology, research on glass weathering has also received much attention. The glass composition was found to be an important factor influencing weathering through a large number of experiments. In this paper, glass composition and glass weathering are analyzed and studied by statistical analysis and integrated algorithm processing of data. The relationship between surface weathering and glass type, decoration, and color was analyzed and processed. Independent sample t-tests using SPSS were used to obtain the distribution patterns of weathering and chemical composition. The mean ratio analysis and BP neural network were then used to predict the pretreated data, respectively. The two results were obtained and statistically analyzed respectively. The classification patterns of high potassium and lead-barium glass were analyzed by statistical analysis and integrated algorithm Bagging model. The chemical composition data of high potassium and lead-barium glass were classified into subclasses and a systematic clustering model was used.

Keywords: Standing sample t-test, BP neural network, Integrated algorithm Bagging model, Systematic clustering.

1. Introduction

In the context of the great cultural exchange, glass was an important result of the early Sino-foreign trade exchanges. Studies related to ancient glass have received much attention. The earliest research on ancient Chinese glass can be traced back to the 1930s, with the study of Seligman[1] as the pioneer.

The main chemical component of glass is silicate, and its main chemical composition is silica. burnett and Douglas[2] conducted an analytical study on the content of the main chemical components of glass in the ternary system of $\text{Na}_2\text{O}-\text{CaO}-\text{SiO}_2$. In ancient times the process of refining glass often produced calcium oxide, lead oxide, barium oxide and other chemical components. We often classify glass into high potassium glass, lead-barium glass, etc. according to the main chemical composition in glass. The research on glass weathering has been paid much attention. In 1984, Liangzhi Zhou[3] found that the weathering of glass is mainly related to the composition of glass through a large number of experimental studies. In the process of weathering, the proportion of glass chemical composition will change, thus affecting the analysis of glass categories.

A batch of data related to ancient glass artifacts in China is known, and the relevant archaeologists have classified them into two types of glass artifacts, high potassium and lead-barium, based on the chemical composition of these artifact data and other testing methods. This paper analyzes the relationship between the surface weathering of this batch of glass relics and the type, decoration and color of the glass; from the type of glass, it analyzes the statistical pattern of the chemical composition content of the surface of glass relics with or without differentiation, and predicts the chemical composition content of the corresponding weathering points before weathering based on the data obtained from the testing of the weathering points. Based on the analysis of the basic data, the classification rules of high potassium glass and lead-barium glass were obtained; for each category, the appropriate chemical composition was selected to classify them into subcategories, and the specific classification scheme and results were given, and the rationality of the classification results was analyzed.

2. Model assumptions and notation

2.1. Assumptions[4]

1. Since some of the data for the chemical components are vacant, i.e., there are cases where they are not detected, the content of such data is assumed to be zero.

2. It is assumed that the chemical reactions occurring between the chemical components of the glass artifacts in the batch have no effect on the analysis of this question.

3. Assume that the chemical components are all determined within reasonable limits and ignore the effect of chemical components not given.

4. Assume that there is no instrument malfunction, operating error, etc. in the process of testing the glass composition that causes the measured data to be the data measured are assumed to be abnormal.

2.2. Notations

Important notations used in this paper are listed in Table 1.

Table 1. Notations

Symbols	Meaning
$ratio_c(j)$	Conversion rate of weathered to unweathered content of chemical composition in column j
$\overline{raw}_c(j)$	Average value of unweathered content of chemical components in column j
$\overline{pro}_c(j)$	Average weathering content of chemical components in column j
$raw_c(i,j)$	Pre-weathering content of column j of artifact i
$pro_c(i,j)$	The jth chemical composition of the i-th artifact after weathering content
x_i	The i-th position
u_k	kth center of gravity position
J	Total distortion degree
x_{ki}	The ith sample content of the kth chemical composition
β_k	Regression coefficient of the kth chemical composition
μ_i	Perturbation term of the i-th sample
y_i	Predicted value of the i-th sample
$F(x, \beta)$	Connection function
$Cov(X, Y)$	Covariance of sequence X sequence Y
σ_X	Standard deviation of sequence X
ρ_{XY}	Pearson's correlation coefficient

3. Model construction and solving

3.1. Analysis of surface weathering of glass artifacts in relation to glass type, decoration and color

During weathering, the composition ratio of the glass changes due to a large exchange of internal elements with the environmental elements of the location. The relationship between surface weathering on glass type, type of decoration, and type of color was analyzed separately. The data on glass type, ornament type, and color type without weathering and weathering were statistically analyzed. It was found that 45.45% of the weathered glass was high potassium glass and 54.55% was lead-barium glass, which was a balanced ratio. 17.65% of the weathered glass was high potassium glass and 82.35% was lead-barium glass, which indicated that the weathered glass had a larger proportion of lead-barium

glass. The analysis found that the types of ornamentation were A and C when there was no weathering, and the corresponding percentages were 50%; the types of ornamentation of weathering were A, C and B, and the corresponding percentages were 32.35%, 50.00% and 17.65%, and the C ornamentation of weathering accounted for more and the B ornamentation accounted for less, and the B ornamentation appeared under weathering. Analysis found that the weathered colors: light green, green, purple, dark blue accounted for a decrease in the ratio, light blue, blue-green, black accounted for a significant increase.

Independent sample t-tests were performed for each category of glass sample data using SPSS. The difference between the mean of the two groups of sample data for each category weathered and unweathered and the overall group it represents was tested for significance. Some groups of statistical results data are shown in Table 2, Table 3.

Table 2. High Potassium Glass Group Statistics

	With or without weathering	Number of cases	Average value	Standard deviation	Standard error mean
Silicon dioxide	0	12	67.9842	8.7551	2.5274
	1	6	93.9633	1.7336	0.7078
Potassium oxide	0	12	9.3308	3.9203	1.1317
	1	6	0.5433	0.4452	0.1817
Aluminum oxide	0	12	6.6200	2.4915	0.7192

The analysis showed that there were significant differences in the contents of silica, potassium oxide, calcium oxide, and alumina in the chemical composition of the high potassium glass before and after weathering. After further analysis of the data, it was found that for the high potassium glass the silica content was significantly higher than the unweathered amount after weathering, while the potassium oxide, calcium oxide and alumina contents were significantly lower than the unweathered amount.

Table 3. Lead barium glass group statistics

	With or without weathering	Number of cases	Average value	Standard deviation	Standard error mean
Silicon dioxide	0	23	54.6596	11.8286	2.4664
	1	23	27.0561	9.0051	1.8777
Potassium oxide	0	23	22.0848	8.2152	1.7130
	1	23	43.7100	12.0775	2.5183
Aluminum oxide	0	23	1.0491	1.8471	0.3851
	1	23	4.7600	3.9886	0.8317

The analysis showed that there were significant differences in the contents of silica, lead oxide and phosphorus pentoxide in each chemical composition of lead-barium glass before and after weathering. After further analysis of the data, it was found that the content of silica in the chemical composition of lead-barium glass after weathering was significantly lower than the amount of unweathered, while the content of lead oxide and phosphorus pentoxide was significantly higher than the amount of unweathered.

There are certain chemical compositions in each category that have changed significantly. For the chemical composition data with significant differences in content before and after weathering, the

corresponding mean values were obtained. The ratio of the mean values of the corresponding chemical compositions was obtained[5,6]:

$$ratio_c(j) = \frac{\overline{raw_c(j)}}{\overline{pro_c(j)}} \quad (1)$$

The content of the component before weathering can be predicted by multiplying each component with the corresponding ratio for the corresponding weathering point.

$$raw_c(i, j) = pro_c(i, j) \times ratio_c(j) \quad (2)$$

For the chemical components such as tin oxide, sulfur oxide and other data are sparse or have no obvious change pattern before and after weathering, the content is temporarily used to keep constant.

The final data of the preliminary prediction results partly calculated by MATLAB are shown in Table 4.

Table 4. Partial forecast results

Category	Artifacts	SiO ₂	Na ₂ O	K ₂ O	CaO	MgO	Al ₂ O ₃	Fe ₂ O ₃
High Potassium	07	67.02			7.87		6.79	1.49
	09	68.75		7.37	4.56		4.53	2.80
	10	70.01		11.49	1.54		2.78	2.27
Lead Barium	54	45.01		0.28	1.74	1.11	5.97	
	43 ₁	25.07			2.86	0.77	3.24	1.23
	43 ₂	43.84			3.50	0.83	4.90	2.25
Category	Cultural relic	CuO	PbO	BaO	P ₂ O ₅	SrO	SnO ₂	SO ₂
High potassium	07	5.55			2.78			
	09	2.66			1.59			
	10	1.44						
Lead barium	54	0.53	28.02	5.52	1.15	0.74		
	43 ₁	3.45	30.24	5.72		0.54		
	43 ₂	0.97	22.61	2.56	3.49	0.40		

Observation of the results reveals the presence of component proportion accumulation and the absence of data in the range of 85% to 105%. It means that the prediction results are better.

BP neural network model was used. It was observed that many groups of the same artifacts in the base data had different degrees of weathering in different parts. For example, the data of No. 08, No. 54 and No. 26 have weathering points and severe weathering points, which can be extracted from the weathering points to the process of severe weathering points, the transformation pattern of the content of different chemical components[7]. For example, No. 50 has unweathered and weathered points, so we can extract the change pattern of different chemical contents during the change from unweathered to weathered points[8].

Through such transformation relations, the weathering points required in the topic are predicted by the neural network toolbox to derive their pre-weathering chemical composition content.

The BP neural network model is capable of learning and storing a large number of input-output pattern mapping relations without revealing beforehand the mathematical equations describing such mapping relations.

The rule is to use the most rapid descent method to continuously adjust the weights and thresholds of the network by back propagation to minimize the sum of squared errors of the network. By designing the architecture and algorithms based on BP neural network, we are able to build a model more applicable to chemical composition.

A typical three-layer forward neural network model is constructed here, and this network topology includes an input layer, a hidden layer, and an output layer.

By setting the number of neurons in the hidden layer, the default is 10, with some adjustment after later observation to achieve the best prediction effect The process of BP neural network is mainly

divided into two stages, the first stage is the forward propagation of the signal, from the input layer through the hidden layer, and finally to the output layer; the second stage is the backward propagation of the error, from the output layer to the hidden layer, and finally to the input layer, in order to adjust the weights and biases from the hidden layer to the output layer, and the weights and biases from the input layer to the hidden layer.

For its direction propagation process, the basic idea is to adjust the network parameters by calculating the error between the output layer and the expected value, thus making the error smaller. Figure 1 show the regression fitting analysis performed by the neural network toolbox for one of them, and the rest of the predictions are similar with a better overall situation.

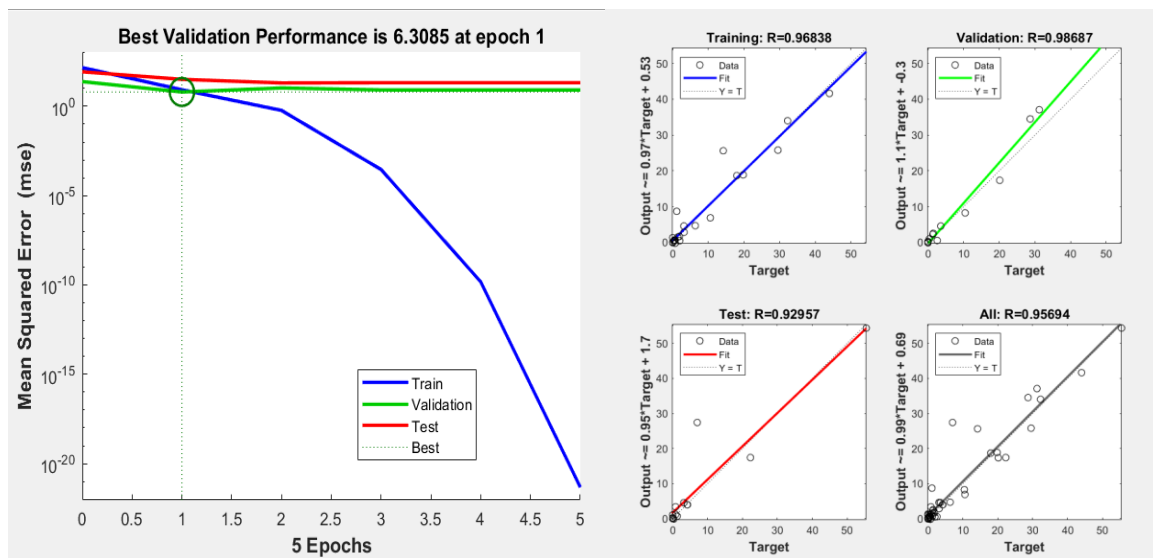


Figure 1. Mean square error and regression fitted curve

The fitted values were regressed on the true values, and the higher the goodness of fit, the better the results were indicated.

A total of three methods, the Levenberg-Marquardt method, the Bayesian regularization method, and the quantized conjugate gradient method, were used to judge the prediction results based on the resulting mean squared error MSE difference, and the methods used were continuously adjusted to achieve the best prediction results.

3.2. Subclasses of chemical composition

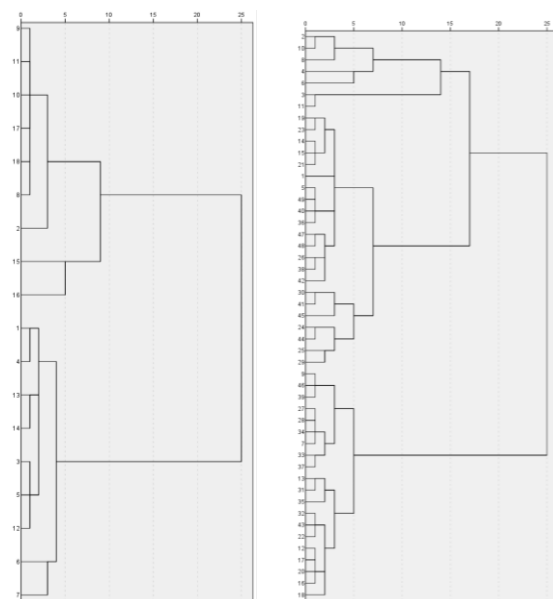


Figure 2. Lineage diagram of high potassium and lead-barium glass

The glasses of high potassium and lead-barium types were subclassed and systematically classified by SPSS to obtain the corresponding coefficients as well as the spectral diagram in Figure 2[9,10].

After the initial division, it is necessary to determine the optimal number of its categories, for which we analyze and determine them by the elbow method.

For this method, the squared sum of the distance between the center of gravity of the class and the position of its internal members can be used as the degree of distortion of each category. Considering $K \leq n - 1$, which is interpreted as having at least one class containing two elements, suppose that a total of n samples are divided into K classes, i.e., the k th class ($k = 1, 2, \dots, K$) is denoted by and the position of the center of gravity of that class is noted as u_k , then the degree of distortion of the k th class can be determined as

The degree of distortion of the K th class can be determined as

$$\sum_{i \in C_k} |x_i - u_k|^2 \tag{3}$$

Define the total degree of distortion for all categories as

$$J = \sum_{k=1}^K \sum_{i \in C_k} |x_i - u_k|^2 \tag{4}$$

The corresponding aggregation coefficients were derived by SPSS and then sorted by Excel to produce line graphs, see Figure 3.

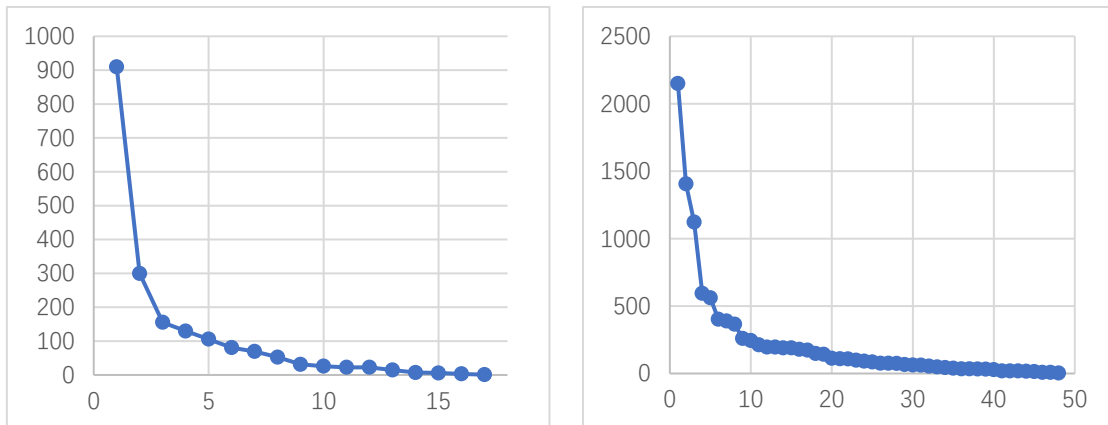


Figure 3. High potassium and lead-barium polymerization coefficient folding line graph

When the value of K changed from 1 to 3, the slope of the degree of distortion was larger, and when $K=3$, the decreasing trend of the folding graph slowed down significantly, thus dividing the category into 3 and determining its optimal subcategory division. Statistical analysis reveals that there is a significant difference in the percentage content of phosphorus pentoxide in these three categories.

High potassium glass and lead-barium glass are subclassified separately according to their chemical composition. According to the percentage content of phosphorus pentoxide, high potassium glass is specifically classified into high phosphorus and high potassium category, medium phosphorus and high potassium category, and low phosphorus and high potassium category. According to the percentage content of silica, copper oxide and phosphorus pentoxide, lead-barium glass is specifically classified into low silica lead-barium class, high copper lead-barium class, high silica lead-barium class, and high lead-lead-barium class.

For the subclassification of high potassium glasses, the decreasing curve of the fold is found to begin to moderate when the category number is set at 3 by the polymerization coefficient fold diagram. The polymerization coefficient changes faster when the category number is at 1 to 3 and starts to change slowly when the polymerization coefficient is at 3 or higher. When the number of categories is set at 1~2, the aggregation coefficient is too high, which means that the choice of categories is not

appropriate. When the number of categories is chosen above 3, there is no obvious change in the aggregation coefficient, but the process of trial classification becomes more complicated, and the sub-category number of aggregation coefficient is more appropriate to take 3. The analysis of the reasonableness of subclassification for lead barium glass is the same as that for high potassium glass, and the above can show that the aggregation classification model is more reasonable.

If the proportion of phosphorus pentoxide in high potassium glass is increased by 5%, the subclassification is again classified by cluster analysis, and the optimal subclassification is 3, which corresponds to an aggregation coefficient of 634.597, an increase of 0.004%. Similarly, if the proportion of silica, copper oxide and lead oxide in lead-barium glass is increased by 5%, the coefficient of aggregation is 593.901 when the optimal subclass is classified as 4 again by cluster analysis, which is an increase of 6.85%. Since the change is relatively small, it can indicate that the sensitivity is relatively low, which in turn indicates that the aggregation classification model is relatively stable.

4. Conclusion

This paper analyzes the relationship between surface weathering of glass artifacts and their glass type, ornamentation, and color requires the use of valid data, which were screened to eliminate invalid data. Then the relationship between surface weathering and glass type, decoration, and color was analyzed separately. The 14 chemical compositions were analyzed two by two to find the chemical composition that very easily distinguishes high potassium glass and lead-barium glass clearly as the classification law of high potassium glass and lead-barium glass. Then the attached data were then operated by random forest algorithm again, and the analysis of glass type classification was realized by MATLAB random forest function. And by analyzing the results of the test group to get a high accuracy rate, the importance of its random forest obtained for 14 chemical composition was observed and analyzed, so that the classification law of high potassium glass and lead-barium glass about chemical composition was derived, and the distribution laws obtained were compared with each other and verified with each other.

References

- [1] Zhou, Liang-Chi, The main factors affecting the weathering of silicate glasses, 1984.01.005
- [2] Burneet.D.G. ; Douglas.R.W. , Phys.Chem.Glass, 1970.11.125~135
- [3] Smith, A. Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm [J]. J Spat Sci, 2010.
- [4] Fu, Xiu-Feng, Yu, F. Xi. A study on the composition of a group of ancient glass from southern and southwestern China based on multivariate statistical analysis[J]. Conservation and Archaeological Science, 2006, 18(4):8.
- [5] Shi Yujiang, Sun Xiaoping. Stress sensitivity analysis of dense clastic reservoirs in Changqing[J]. Petroleum Exploration and Development, 2001, 28(5):3.
- [6] Zhang M.W., Liu Y., Zhang B., et al. A concept-based data clustering model[J]. Journal of Software, 2009(9):10.
- [7] Seligman C G, Ritchie P D, Beck H C. Early Chinese Glass from Pre - Han to Tang's Times [J]. Nature, 1936(138) : 721~724.
- [8] Wang J, Li Mo, Ma Qinglin, et al. Weathering study of an octagonal lead-barium glass vessel from the Warring States period[J]. Glass and enamel, 2014, 42(2):8.
- [9] Huang Qishan. The discovery of ancient glass objects in Guangxi and its research[J]. Archaeology, 1988(3):14.
- [10] Gao Hongshen, Tao Youde. Improvement of BP neural network model [J]. System Engineering Theory and Practice, 1996, 16 (1): 67-71.