

Composition Analysis and Identification of Ancient Glass Products

Kaijie Zhang^{1, †, *}, Ziwen Wang^{2, †}, Shaochen Yan^{3, †}

¹ Department of Intelligent Manufacturing, South China University of Technology, Guangzhou, China

² Department of Big Data, South China University of Technology, Guangzhou, China

³ Department of Robotic, South China University of Technology, Guangzhou, China

* Corresponding author email: zhangkaijie0621@163.com

†These authors contributed equally

Abstract. During the weathering process of ancient glass, due to the large amount of exchange between the internal elements and the environmental elements, the composition ratio of the glass changes, which affects the correct judgment of its category. In this paper, statistical analysis and classification of known data are carried out, and K-means algorithm of random forest and unsupervised learning is adopted to analyze and identify the components of glass products, and reasonable prediction is made for unknown types of glass products.

Keywords: Glass Composition Identification; Neural Network Algorithm; Random Forest Algorithm; K-Means Clustering Algorithm.

1. Introduction

As an important trade route between China and the West since the Han Dynasty, the Silk Road witnessed the economic and cultural exchanges between ancient China and countries in West Asia, South Asia, Europe and North Africa. Glass products originating in Mesopotamia are not only an ancient man-made material, but also an exquisite luxury. Like gold and jewelry, it is an important trade product and one of the main commodities imported from Western countries to China, witnessing the exchange of culture and technology between China and the West.

Glass was also an important item for exchange and communication along the Silk Road. However, due to the climate and burial environment, ancient glass is easily affected by burial environment and weathering. The main raw material of glass is quartz sand, and the main chemical composition is silica. When the glass is calcined with different melting AIDS, its main chemical composition is different. For example, lead ore is added as a melting aid in the paper burning process of barium glass, which has a high content of lead oxide and barium oxide. It is usually regarded as the glass variety invented by our country. The glass of Chu culture is mainly barium lead glass. Potassium glass is made of materials with high potassium content, such as plant ash, which is fired as flux. It is mainly popular in Lingnan, Southeast Asia and India. In this paper, two types of lead-barium glass and high-potassium glass need to be studied, and different classification standards should be established based on weathering, color, pattern and other characteristics for further solution. The specific process is shown in the following figure

2. Neural Network Predictive Analysis Model

Neural network is the basis of deep learning. It can make use of the weight relationship between the existing data input and output, and then use such weight relationship to conduct simulation. In the process of using neural network prediction, the data was firstly treated with dummy variables. It was noted that the weathering status was classified into three states: unweathered[1], weathered and severely weathered. Therefore, the weathering status was divided into three dummy variables. Due to the excessive variety of chemical components, the fitting effect was not good through linear

regression fitting, indicating that it is difficult to use simple linear regression method to predict the chemical components before weathering[2].Therefore, this paper adopts the most rapid descent method for neural network regression, which is suitable for more complex regression models. Through back propagation, the weight and threshold of the network are constantly adjusted to minimize the sum of error squares of the network[3].

Parameters of the Bp neural network fitting model are shown in Table 1.

Table 1. Bp neural network fitting model parameters

Parameter name	Parameter value
Data segmentation	1
Data shuffle	No
Cross validation	No
Activation function	identity
Solver	lbfgs
Learning rate	0.1
L2 regular term	1
Iterations	1000
Hide the number of neurons in the first layer	100

After the experiment, it can be seen that the neural network regression fitting results are good, and the total percentage content of each glass prediction substance is within [85-105%], which is an acceptable range.This model is used to fit weathered and severely weathered glass, and the results are shown in the appendix[4].

3. Glass Type Division Model

In this question, it is necessary to classify and subdivide high-potassium glass and lead-barium glass, and analyze the rationality and sensitivity of the classification model.First of all, data preprocessing is needed to screen the types of high-potassium glass and lead-barium glass, conduct elbow analysis, and determine the k value of the cluster.Screen for weathering, ornamentation, color of each category and total content range of chemical composition.Then this paper adopts the random forest algorithm, and chooses the genetic algorithm as the optimization algorithm, sets the parameters and solves to get the classification law.Then clustering was used to select appropriate chemical components for each category and subclassification was carried out.Finally, this paper classifies the clustering results by decision tree and obtains the specific classification method[5].

3.1 Random Forest Analysis

In order to seek the classification standard of high potassium, lead barium glass, it is necessary to find a suitable classification method[6].Considering that the chemical composition of glass is relatively complex and some errors are inevitable due to detection techniques and other reasons, the traditional decision tree classification is not effective in prediction. Therefore, random forest algorithm is chosen in this paper for classification.Random forest is a supervised learning algorithm, which is easy to implement and has low computation cost.When random forest is used, the initial population number is selected as 50, the maximum number of iterations is 150, the mutation probability is 0.01, and the crossover probability is 0.5.The model parameters are set according to these data as shown in Table 2.

The content of each component and the characteristic importance of its attributes can then be obtained, as shown in Figure 1.

Table 2. Data setting model parameter table

Parameter name	Parameter value
Training time	0.063 s
Data segmentation	0.7
Data shuffle	No
Cross validation	No
Evaluation criteria for node splitting	Gini Impurity
Number of decision trees	100
With put back sampling	Yes
Out of bag data test	No
Maximum feature scale considered in partition	Automatic
Minimum number of samples split by internal nodes	2
Minimum sample number of leaf node	1
Minimum weight of samples in leaf node	0
Maximum depth of tree	10

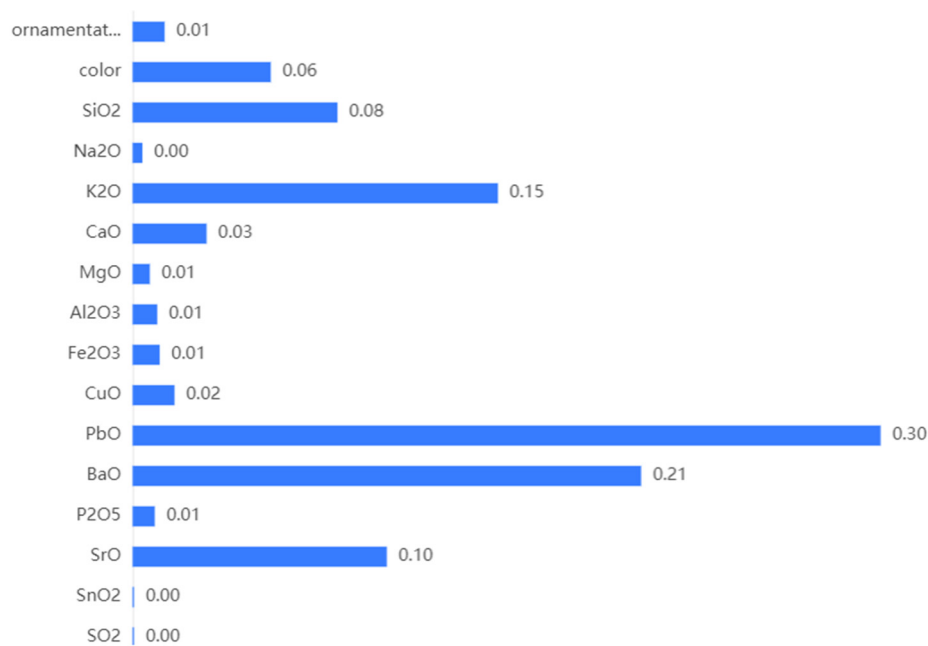


Figure 1. The content of each component and the characteristic importance of its attributes

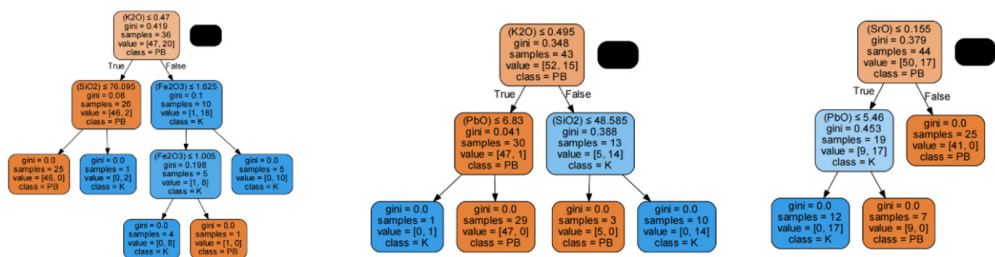


Figure 2. Part of the standard diagram of random forest classification

According to the chart and analysis, lead oxide, barium oxide and potassium oxide play a leading role in the classification of high-potassium glass and lead-barium glass. Color and other factors will

also have a certain degree of influence on the classification, but at the same time, decorative factors and other factors have little influence on the classification [7].

After testing, part of the random forest classification standards is shown in Figure 2.

3.2 Subclass Division Model with Clustering K-means Algorithm

Before the clustering division, this paper first needs to determine the specific elements through which to cluster. Since different substances have different contents in different samples, this paper decides to select the first several (first seven) elements with large fluctuations for clustering, so this paper solves the standard deviation of different substances.

Using standard deviation as the criterion, this paper can also get the dispersion degree of different chemical components.

The formula for calculating standard deviation is as follows:

$$\text{Overall standard deviation: } \sigma_0 = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

$$\text{Sample standard deviation: } \sigma_1 = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

$$\text{Standard error: } \sigma_n = \frac{\sigma}{\sqrt{n}}$$

$$\text{Sample mean: } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

The standard deviation of the content of each substance in the high-potassium and lead-barium glass was finally calculated as shown in the table3 below[8].

Table 3. Standard deviation of the content of each substance in high potassium and lead barium glass

Type	SiO ₂	Na ₂ O	K ₂ O	CaO	MgO	Al ₂ O ₃	Fe ₂ O ₃	CuO	PbO	BaO	P ₂ O ₅	SrO	SnO ₂	SO ₂
Hyperkalemia	14.46	1.08	5.37	3.31	0.71	3.07	1.56	1.49	0.51	0.84	1.28	0.04	0.55	0.15
Lead-barium	18.64	1.81	0.27	1.63	0.63	3.01	0.94	2.47	14.94	8.33	3.91	3.12	0.21	3.13

To sum up: when clustering high-potassium glass, the selected substances are: silica, potassium oxide, calcium oxide, alumina, iron oxide, copper oxide and phosphorus pentoxide; For the clustering of lead-barium glass, the materials selected in this paper are: silica, lead oxide, barium oxide, phosphorus pentoxide, sulfur dioxide, strontium oxide and alumina. Then this paper uses these substances to subdivide high potassium, lead barium glass.

Cluster analysis is used in this paper when the high potassium glass and lead barium glass are divided into subclasses. First of all, this paper determines K value according to the elbow method, that is, it is specifically divided into several categories, but the specific content of each category needs to be analyzed and judged by decision tree. First of all, for lead-barium glass, this paper draws its image and obtains the best K value when it is at the inflection point, as shown in Figure 3[9].

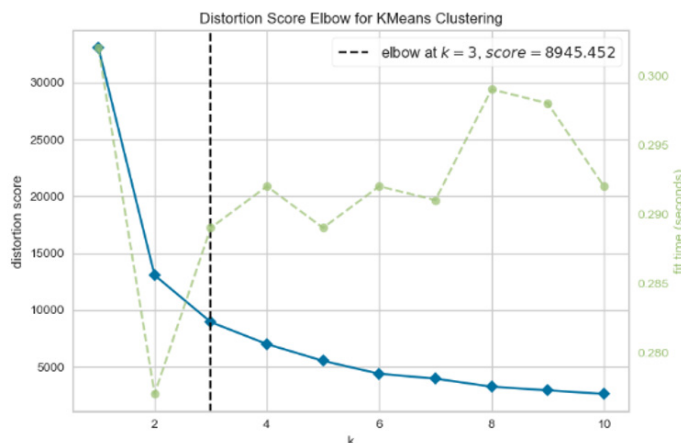


Figure 3. Elbow analysis of lead-barium glass

As can be seen from the chart, when the type of glass is lead-barium glass, the optimal K value is 3[10].

Similarly, elbow analysis was performed on the high potassium glass, as shown in Figure 4.

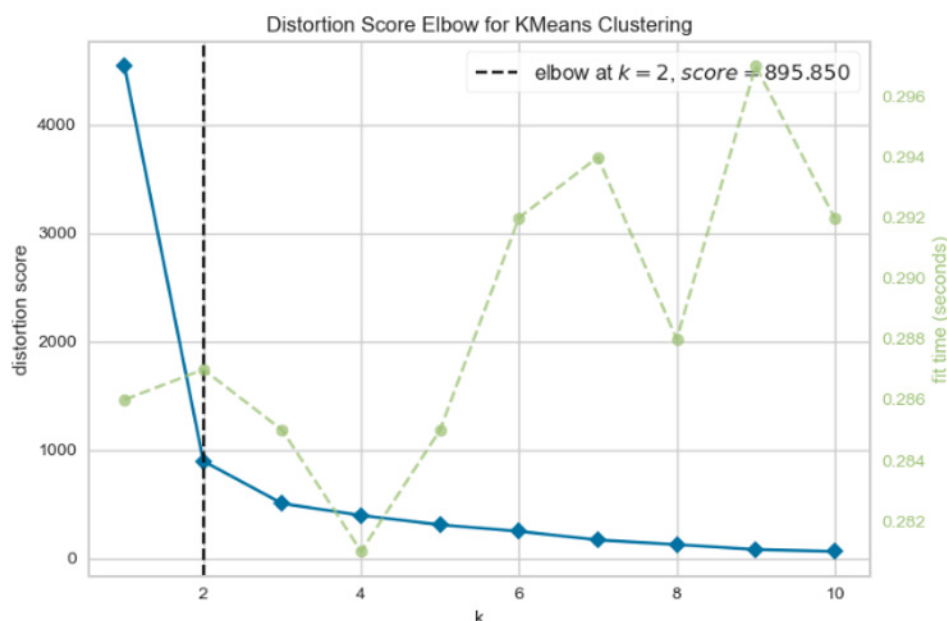


Figure 4. Elbow analysis of high potassium glass

As can be seen from the chart, when K=2, the image enters the inflection point, but the distortion score still drops significantly from K=2 to K=3. Therefore, clustering is carried out according to K=3 in this paper. (Note: The clustering division results of high-potassium glass are in the attachment).

Subsequently, this paper uses the decision tree algorithm to draw the decision trees of lead-barium glass and high-potassium glass in turn. From the decision tree, this paper can obtain numerous information such as Gini index and sample size, and also know the specific contents contained in each category.

The decision trees of lead-barium glass and high-potassium glass are respectively shown in Figure5 (left and right) :

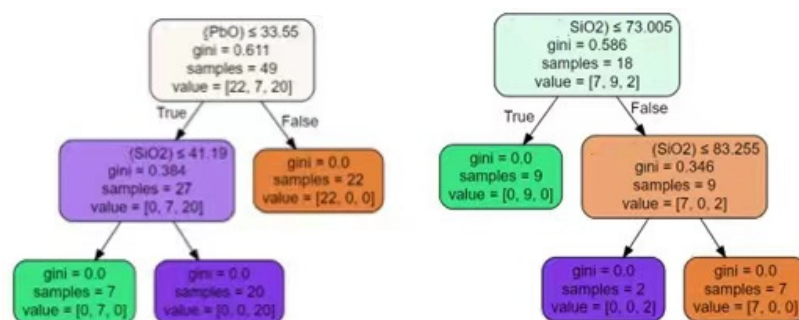


Figure 5. Decision trees of lead-barium glass and high-potassium glass
(Note: The above subclasses are denoted by 0, 1 and 2 in the model and below.)

4. Analyze the Rationality and Sensitivity of the Classification Results

4.1 Model Accuracy Analysis

After several subclasses are sorted out by clustering, this paper needs to analyze the rationality of the results, set a random range, bring the results back to the K-means algorithm to calculate the Euclidean distance from the coordinates of the original clustering center point, make a judgment again, and compare with the real value to calculate the accuracy of the model. In this paper, the sensitivity

changes were set as 1%, 2%, 5%, 10%, 20% and 30% respectively. The specific accuracy of the model is shown in Table 4.

Table 4. Specific accuracy of the model

Type	1%	2%	5%	10%	20%	30%
High potassium weathering 2	100%	100%	100%	100%	100%	99%
High potassium type unweathered 2	100%	100%	100%	100%	100%	98%
Lead barium type weathered 2	100%	100%	100%	100%	99%	99%
Lead barium type unweathered 2	100%	100%	100%	99%	98%	90%

As can be seen from the above table, when the high-potassium glass increases random disturbance before and after weathering, the accuracy of the model is still very high when the disturbance value is less than 30%, indicating good clustering effect and strong robustness of the model. For the unweathered lead-barium glass, only when the disturbance value is 30%, the accuracy of the model is reduced to 90%, and the remaining cases are still close to 100%. Therefore, the accuracy and stability of the model are reflected to a certain extent.

4.2 Composition and Sensitivity Analysis of Unknown Glass Relics

In the chemical composition analysis of glass cultural relics of unknown category, data preprocessing is carried out first. All the values with Nan content are treated as 0, and the weathering is treated as dummy variable. Meanwhile, the model is classified and the classification results are shown in Table 5.

Table 5. Classification results

Cultural relic number	Surface weathering	Inferential classification	Inferential subclass
A1	No	Hyperkalemia	1
A2	Yes	Lead-barium	0
A3	No	Lead-barium	2
A4	No	Lead-barium	0
A5	Yes	Lead-barium	1
A6	Yes	Hyperkalemia	2
A7	Yes	Hyperkalemia	2
A8	No	Lead-barium	1

On this basis, this paper carries out data disturbance processing:

Set the parameter k from 0.1 to 0.5, make the content of each component increase k times successively, and then make the prediction with the same model. To judge whether it is the same as the original prediction result, so as to judge the sensitivity of the model. (Note: Results are in the Appendix)

It can be seen from the results that all the predicted results are the same as the previous ones except for the range of copper oxide increase of 30% to 50%. Therefore, the paper concludes that the sensitivity of this model is low.

5. Model Evaluation

Through the modeling method of splitting problems and solving them in turn and getting the final answer by correlation, this paper can carry out rationality analysis and unreasonable data elimination in the process, so that the overall fit degree of the model can be optimized, and the results are more reasonable and authentic.

Through data preprocessing, the impact of wrong data and invalid data on the model is better solved, which makes the results more reasonable and avoids the generation of errors.

The weight and threshold of the predictive analysis model of the established neural network are adjusted by the fastest descent method and back propagation, and the sum of squares of the errors are reduced, so that the model has a better fitting effect and stronger authenticity.

Through random forest analysis, clustering K-means algorithm and decision tree algorithm, the model can better solve the classification and subclass of a given sample, and show high accuracy when processing the data of unknown samples, which reflects the high universality of the model.

Through the introduction of standard deviation judgment criterion, the glass type and its subclass classification model have a unique way of thinking and innovative characteristics.

References

- [1] Chen Shuyu. The Origin and Development of ancient glass in China [J]. Cultural Relics Identification and Appreciation,2019(04):44-45.
- [2] Gan Fuxi. Some Views on the study of ancient glass in China [J]. Journal of the Chinese Ceramics, 2004 (02): 182-188.
- [3] Fuxiufeng, dry happiness. Study on the composition of ancient glass from South China and Southwest China based on multivariate statistical analysis [J]. Protection of cultural relics and archaeological science, 2006 (04) : 6-13. DOI: 10.16334 / j.carol carroll nki cn31-1652 / k. 2006.04.002.
- [4] Li Qing-hui, Dong Jun-Qing, GAN Fu-xi. China's early glaze chemical composition and the technological characteristics of sand and glass study [J]. Journal of guangxi university for nationalities (natural science edition), 2009 (04) : 31-41. DOI: 10.16177 / j.carol carroll nki GXMZZK. 2009.04.014.
- [5] Wang Chengyu, TAO Ying. Weathering of silicate glass [J]. Journal of the Chinese Ceramic Society, 2003 (01):78-85.
- [6] Wang Chengyu, Tao Ying, Chen Min, Huang Ming. Sodium calcium magnesium aluminium silicate glass and alkali of lead silicate glass weathering [J]. Journal of silicate bulletin, 1989 (6) : 1-9. DOI: 10.16552 / j. carol carroll cnki issn1001-1625.1989.06.001.
- [7] Zhou C J. Main factors affecting the weathering of silicate glass [J]. Journal of Dalian Institute of Light Industry, 1984(01):34-44. (In Chinese).
- [8] Li Qinghui, Gan Fuxi, Gu Donghong. Some problems on the study of ancient Chinese glass [J]. Studies in the History of Natural Science,2007(02):234-247.
- [9] Wen Rui, ZHAO Zhiqiang, Ma Jian, Wang Jianxin. Composition analysis of glass beads unearthed from Shirenzigou Site in Balikun, Xinjiang [J]. Spectroscopy and Spectral Analysis,2016,36(09):2961-2965.
- [10] Li Mengmeng, Niu Ran, Song Qizhe, Su Chang, Wang Huijun, Zhang Bing, Wu Zhenyu. Study on the relationship between human agrochemical contaminant residues and hypercholesterolemia based on Principal component analysis and Logistic regression method [J/OL]. Chinese Journal of Food Hygiene:1-13 [2022-11-04].