Composition analysis and classification prediction of ancient glass

Jieyin Chen*

School of Management, Shenzhen University, Shenzhen, China

*Corresponding author: 2020044016@email.szu.edu.cn

Abstract. Ancient glass is easily affected by environmental factors, which cause chemical changes between glass composition and environmental substances and leads to weathering. Analysis of the chemical composition of the ancient glass plays an important role in the study of the Silk Road. To study the composition of glass relics, the Kruskal-Wallis test was used to determine the chemical composition of different glass types; A decision tree was established to explore and visualize the classification rules of glass types, and the sensitivity analysis of subclass classification was carried out by setting the value of chemical composition content of cultural relics sampling points from -10% to 10%. It was found that high potash glass was most sensitive to SiO_2, and lead-barium glass was most sensitive to PbO. Finally, based on the random forest algorithm, the sensitivity of the model is analyzed. According to the experimental hypothesis and data validation, it is proved that the comprehensive model has good robustness, and can accurately and quickly analyze and identify the chemical composition of ancient glass.

Keywords: Ancient Glass, Kruskal-Wallis, Decision Tree, Hierarchical Clustering, Random Forest.

1. Introduction

The land Silk Road originated from the land passage opened by Zhang Qian sent by Emperor Wudi of the Western Han Dynasty to the Western Regions, starting from the capital Chang'an (now Xi'an), through Gansu and Xinjiang, to Central Asia and West Asia, and connecting the Mediterranean countries. The ancient glass is the witness of the Silk Road[1], but because the ancient glass is vulnerable to temperature, humidity, environmental atmosphere and other factors, the chemical changes between the glass composition and the surrounding environmental substances lead to weathering[2] which will cause different degrees of weathering. Therefore, it is necessary to identify and analyze the chemical composition of ancient glass, which plays an important role in the study of ancient culture.

In this paper, a comprehensive analysis model for the chemical composition of ancient glass was established. Firstly, the Kruskal-Wallis test was used to determine the chemical composition of different glass types with significant differences; A decision tree was established to explore and visualize the classification law of glass types, and the decision criteria were summarized as the classification law, and finally, the conclusion that the classification of glass types was mainly based on the composition of flux was drawn. Inspired by the classification rules of high potassium-lead-barium, two methods of glass subclass classification are proposed in this paper. One was classified directly based on the content of calcium and aluminium[3], and the other was classified according to the distance of the content of each chemical component at different sampling points, and the decision tree was used to fit the classification results to visualize the rules, to construct a hierarchical clustering classification-decision tree visualization model. Then the random forest is used to build multiple decision trees to improve the robustness of the classification model. The results show that the comprehensive model can be used to identify and analyze the chemical composition of ancient glass.

2. Model establishment and solution

2.1. Classification model based on decision tree

2.1.1 Construction of decision tree model

Before studying the classification rules of glass types, the Kruskal-Wallis test was carried out for each chemical composition for different groups to confirm the chemical composition with significant differences in different glass types[4]. The test results are shown in Table 1.

Table 1. Results of the Kruskal-Wallis test

Analysis item	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
P	0.000 ***	0.341 **	0.000 ***	0.116 ***	0.441 **	0.050 ***	0.019 **	0.065 *	0.000 ***	0.000 ***	0.202 ***	0.000 ***	0.607	0.598
Cohen's f Value	0.151	0.026	0.156	0.074	0.021	0.043	0.059	0.012	0.165	0.116	0.061	0.116	0.021	0.026

Note : * * *, * *, * represents the significance level of 1 %, 5 %, 10 % respectively

According to the results of the table, it can be seen that SiO_2, K_2O, Fe_2O_3, PbO, BaO and SrO have significant differences in different types of samples.

The basic idea of the decision tree is to construct a tree with the fastest decrease of entropy based on information entropy. In the context of this problem, the tree chooses the feature with the largest information gain, that is, the specific chemical composition as the current split feature, until the entropy value at the leaf node is zero, which also means the end of glass type division. The formula for information entropy and information gain is as shown in formula (1) and (2).

$$H(X) = -\sum_{X} p(x)\log p(x) \tag{1}$$

$$g(X,A) = H(x) - H(X|A) \tag{2}$$

Formula 1 represents a calculation formula of information entropy, x represents a random variable, and $p(x)$ represents an output probability function. The greater the uncertainty of a variable, the greater the entropy. Formula 2 represents the calculation formula of information gain, where A is a feature, x is the original data set, $(X|A)$ represents the data set under classification A, and the subtraction of the information entropy of the two data sets is the information gain.

2.1.2Analysis of model results

When the decision tree[5] is established and pruned according to the information gained, the result is shown in Figure 1.

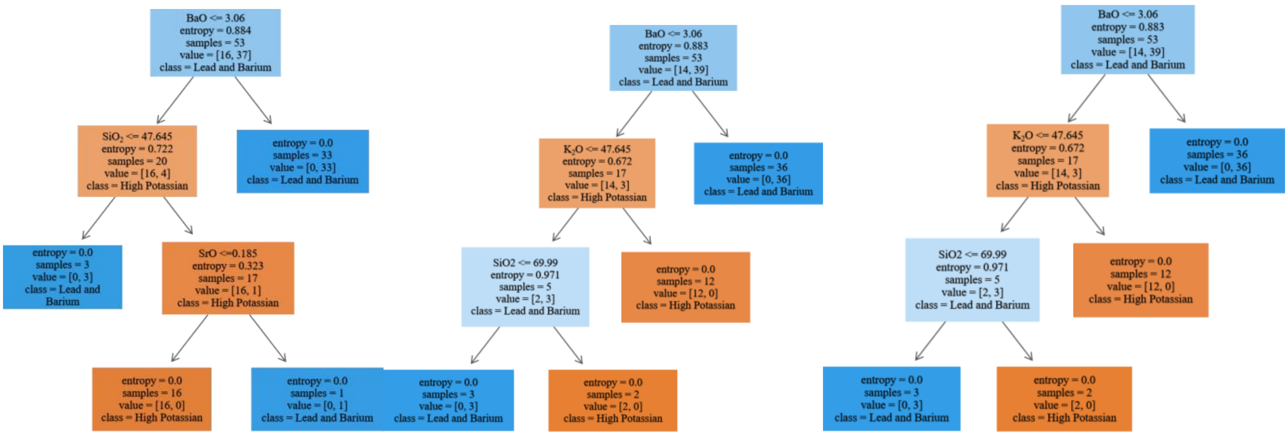

Figure 1. Decision tree based on the classification of different oxides

Fig.1 shows, from left to right, a decision tree based on the classification of barium silicon strontium oxide, barium phosphorus strontium oxide, and potassium silicon barium oxide, respectively, and the results of the model, which leads to the following conclusions.

From the four decision trees, it can be seen that the classification of high-potassium glass and lead-barium glass is mainly based on the content of its flux, namely K_2O, PbO, BaO and the content of its main component, SiO_2.

2.2. Subclass division based on hierarchical clustering

Because the number of subclasses can not be known in advance, we adopt the hierarchical clustering model, by calculating the Euclidean distance between samples, the nearest samples are clustered into one class first, and the Calinski-Harabasz index after each layer of classification is calculated, and the number of subclasses is judged according to the minimum covariance. It is noted that the samples need to be normalized before hierarchical clustering because of the large difference in the dimensions of different chemical components[6].

2.2.1 Classification of high potash glass

In the high-potash glass, we perform hierarchical clustering on all samples, and the results are shown in the figure 2. We calculate the Calinski-Harabasz index at different class numbers. When the Calinski-Harabasz index is minimum, the cluster class is 4 and the value is 1.62. The division results[7] are shown in Table 2, in which different sampling points of the same cultural relic are divided into one subclass.

Table 2. Classification results of high-potash glass subclasses

Subclass 1	Subclass 2	Subclass 3	Subclass 4
06	13、14、16	01、04、05、22、27	03、18、07、09、10、12、21

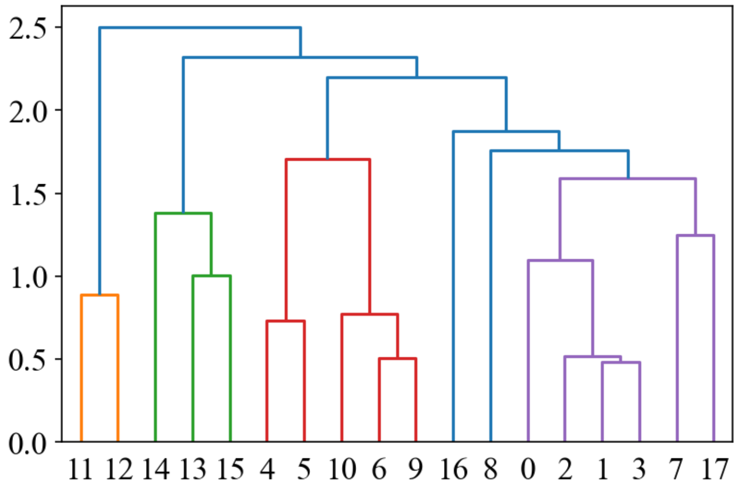

Figure 2. Hierarchical clustering of high-potash glasses

The oxides of calcium, strontium and sodium are selected as the classification characteristics of high potash glass, which can be divided into four subtypes: low calcium and high strontium type, high calcium and high sodium type, high calcium and low sodium type and low calcium and low strontium type, as shown in Figure 3.

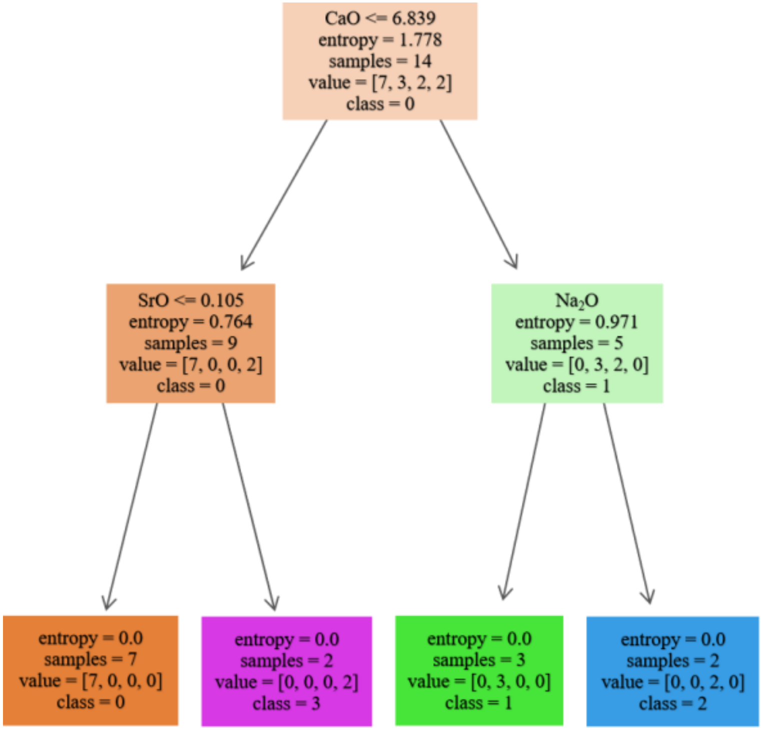

Figure 3. high-potash Glass Decision Tree Visualization

2.2.2 Classification of subclasses of lead-barium glass

In the lead-barium glass, we still use hierarchical analysis[8] to cluster the samples. Because the sample variance of lead-barium glass is large, the Calinski-Harabasz index is about 5 when the cluster class is small, and there is no significant reduction. We use the contour coefficient to determine the optimal number of clusters, and the number of clusters with the largest contour coefficient is 3, and the value is 0.10. The result of subclass division is as follows, in which different sampling sites with only cultural relics numbered 50 are divided into different subclasses.As shown in the Table 3 and figure 4.

Table 3. Results of Subclassification of Lead-Barium Glass

Subclass 1	Subclass 2	Subclass 3
24、37、20、08、26、11	52、42、23、46、45、53、47、39、40、56、57、34、36、32、35、25、55、28、33、54、19、50Unweathered	30、54、50、43、49、58、51、02、41、31、46、48、44

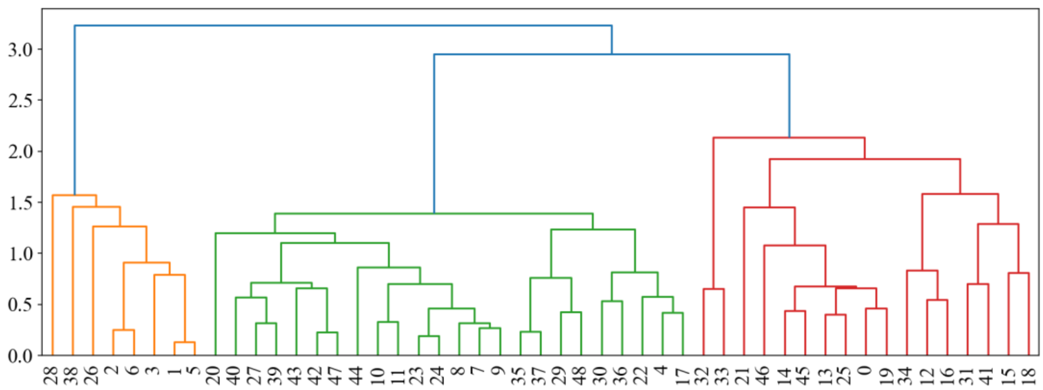

Figure 4. Lead-barium glass hierarchical clustering

The oxides of calcium and barium are selected as the classification characteristics of lead-barium glass, and finally, three subclasses of high calcium type, low calcium and high barium type and low calcium and low barium type are obtained.As shown in the figure 5 .

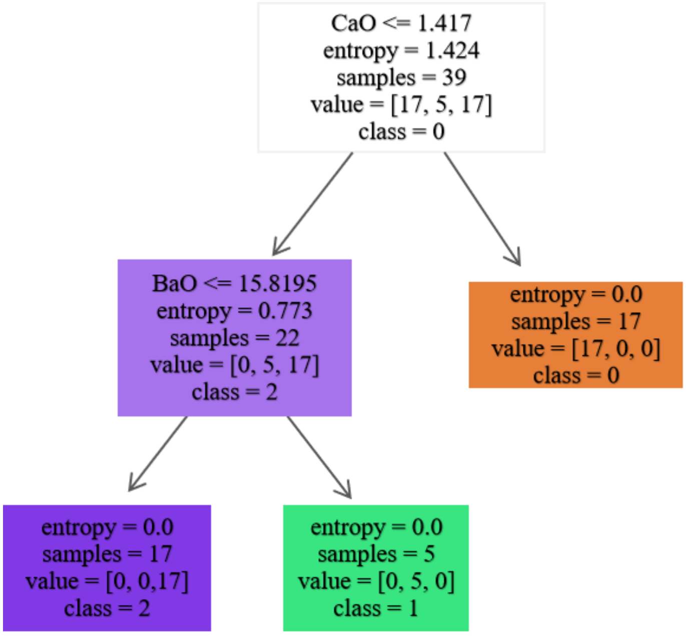

Figure 5. Lead-barium glass decision tree visualization

2.3. Model Rationality and Sensitivity Analysis

The division method based on the content of aluminium and calcium is based on the physical properties of chemical substances[9], and then the subclasses are divided, such as CaO can increase the chemical stability and mechanical strength of glass, Na_2O can increase the thermal expansion coefficient of glass, BaO can reduce the melting temperature and chemical stability, Sr can improve the insulation effect, increase high whiteness and transparency, etc. It is supported by literature and has certain rationality.

In the model based on the content of aluminium and calcium, we divide the subclasses by fixing the content range of Al_2O_3 and CaO, and when these two components exceed the critical value of classification, the class will change. In the hierarchical clustering model, because the samples are classified by distance, we are not clear about the sensitivity of the content of each element in the classification, and the classification criteria of subclasses are only an explanation of the results. Therefore, in this section, we will analyze the sensitivity of the classification results to the chemical content. The sensitivity of the hierarchical clustering model to each chemical element was explored by setting a numerical fluctuation of -10% ~ and 10% for different chemical compositions. Because the value of each sample is different, if we reduce or increase the value of a chemical substance by the same proportion at the same time, it will not affect the classification after normalization. Therefore, we adopt the method of changing the value of a certain chemical component content of only one sample at a time, and after traversing all samples, to classify the mean value of the number of samples with label changes, to measure the impact of chemical component changes on the subclass classification results, as shown in Figure 6, which shows the sensitivity analysis of chemical components of different glass types, and only six of the components are shown here.

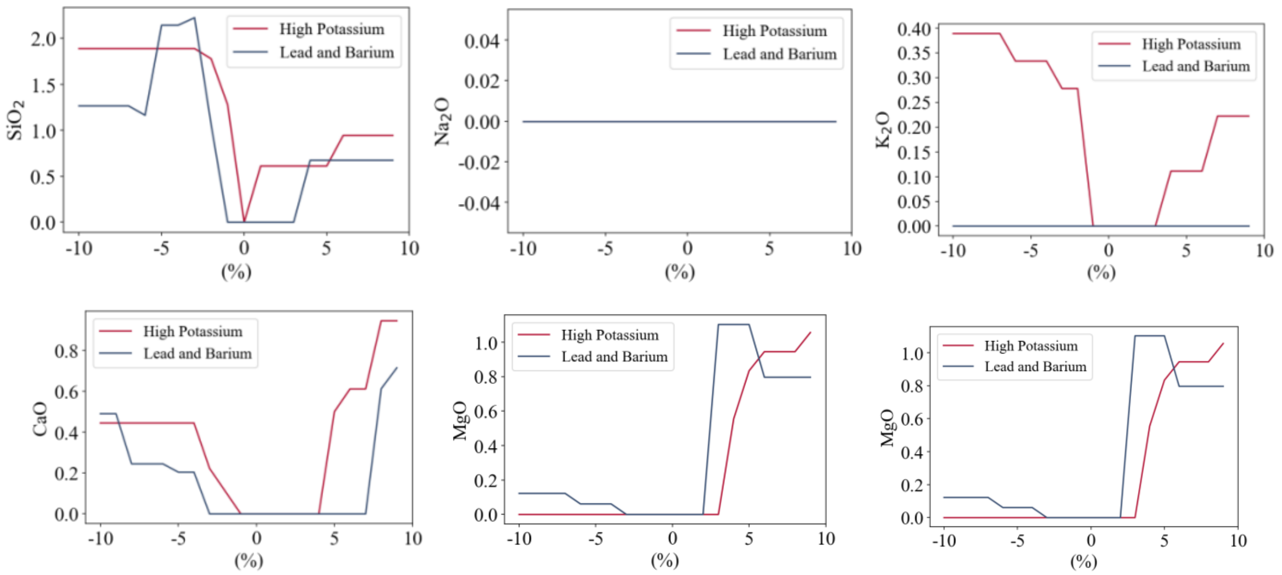

Figure 6. Chemical composition sensitivity analysis for different glass types

In the above figure, the red curve is the high potash glass, and the blue line is the lead-barium glass. Calculate the change number of each chemical element in the subclass classification results under different change degrees and take the average value. The sensitivity of the high potassium subclass model to each chemical component is as follows: SiO_2 > CuO > Al_2O_3 > PbO > SrO > BaO > CaO > MgO > SnO_2 > Fe_2O_3 > K_2O > P_2O_5 > Na_2O = SO_2

The sensitivity of the Pb-Ba subclass model to each chemical composition is: PbO > SiO_2 > SnO_2 > Fe_2O_3 > BaO > SrO > MgO > Al_2O_3 > CuO > P_2O_5 > SO_2 = Na_2O = K_2O

2.4. Glass type prediction based on random forest

From the previous article[10], we know that when using decision trees to classify high-potassium and lead-barium glasses, there are many ways to accurately classify existing samples. In this paper, the idea of ensemble learning is used to fit the decision tree which can distinguish glass types according to different criteria by using the characteristics of a random forest which randomly selects samples and features, so as to improve the robustness of the classification model.

The prediction results of cultural relic types in Table 4 are as follows:

Table 4. Prediction Results of Cultural Relic Types

Cultural relic number	A1	A2	A3	A4	A5	A6	A7	A8
Type	High potassium	Lead and barium	Lead and barium	Lead and barium	Lead and barium	High potassium	High potassium	Lead and barium

The important features of the random forest are extracted and sorted, as shown in Figure7.

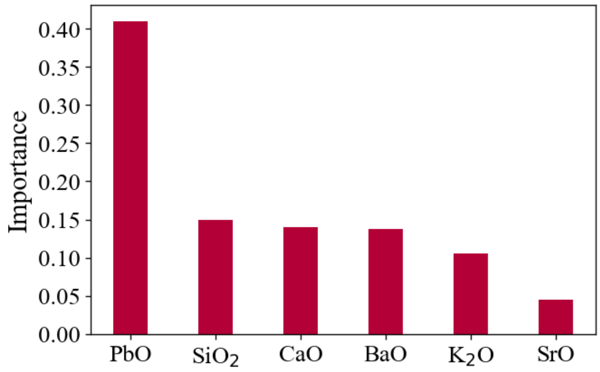

Figure 7. Random Forest Feature Importance Feature Distribution

The feature importance of the model corresponds to the sensitivity of the model to the feature. It can be seen from the figure that the sensitivity of the established model to PbO is far higher than that of other features. The model takes PbO as an important basis for glass classification. It can be seen from the literature that leads is an important part of the lead-barium glass and an important chemical substance to distinguish glass types. Therefore, the importance of the characteristics obtained from the analysis is in line with the assumption of this paper, and it also verifies the rationality of the classification rule established by model 2.1 based on lead, barium and potassium.

3. Conclusion

In this paper, a comprehensive analysis model for the chemical composition of ancient glass was established. Firstly, the chemical composition of different glass types was judged by inspection; then a decision tree was established to explore and visualize the classification rules of glass types, and the classification rules were summarized, and finally, the conclusion that the classification of glass types was mainly based on the composition of flux was drawn. Then the hierarchical clustering classification-decision tree visualization model is constructed, and the decision tree is established by using the random forest to improve the robustness of the classification model. The results show that the comprehensive model can be used to identify and analyze the chemical composition of ancient glass.

In this paper, the sampling points are innovatively grouped according to the physical properties of glass, which improves the accuracy of chemical composition prediction before weathering. The classification model of ancient glass types uses random forest, which makes the model sensitive to the change of different chemical compositions and reduces the possibility of overfitting the data. In the follow-up work, the hierarchical clustering-decision tree model established in this paper can be extended to more classes (subclasses) of material or classification law problems. The classification model provides a reference for the analysis and identification of the chemical composition of ancient glass.

References

[1] Han Hongchang. Alumina is used as a raw material in the glass industry. Light Metals, 6, 1981.

[2] Conscience Zhou. The main factors affecting the weathering of silicate glass. Journal of Dalian Institute of Light Industry, 1, 1984.

[3] An Jiayao. Early Chinese glassware. Acta Archaeologica, (4): 413 – 448, 1984. 4.1[4] Shi Meiguang, He Ouli, Zhou Fuzheng. Study on Potash Glass Unearthed from a Number of Han Tombs in China. Acta Silicate Sinica, 14 (03): 307 – 313, 1986.

[4] Cui Jianfeng, Wu Xiaohong, Tan yuanhui, Wang Yongbiao. Composition Analysis of Ancient Glassware Unearthed from Chu Tombs of the Warring States Period in the yuanshui Valley of Hunan Province. Acta Silicata Sinica, 37 (11): 1909 – 1913, 2009.

[5] Shao Hui, Zhou Hongqing, Liu Min, Zhu Haikui, Shen Xiaodong. Effect of alumina addition on barium silicate composite glass ceramics. Chinese Journal of Ceramics, 32 (1): 17 – 20, 2011.

[6] Wang Shengtie, Liu Yutang. Summary of Chu Culture. BEIJING BOOK CO. INC., 2013.

[7] Liu Song, Lv Liangbo, Li Qinghui, Xiong Zhaoming. Glass Beads Unearthed from Han Tombs in the South of the Five Ridges and the Sino-Foreign Exchange on the Maritime Silk Road of the Han Dynasty. Conservation and Archaeological Science, 31 (4): 18 – 29, 2019.

[8] Geological Dictionary Office, Ministry of Geology and Mineral Resources. Geological Dictionary: Geological Publishing House, 2005

[9] Wei Jie. Research on improved K-means clustering algorithm [J]. Information Communication, 2018 (5): 14-15.

[10] Chen Min. Factors affecting the weathering of lead glass [D]. Dalian University of Technology, 1987.