Glass component analysis based on Fisher's linear discriminant model and random forest classification model

Baoyin Liu*, Yapeng Liu

School of Traffic and Transportation, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, 050043 China

*Corresponding author: 2926386039@qq.com

Abstract: The production of ancient glass is the crystallization of the wisdom of the majority of ancient working people in China. The differences in fluxes and the differences in the burial environment during weathering have a great impact on its chemical composition content. In this paper, a relevant classification model is established based on the data of various properties of glass. In this paper, Fisher linear discriminant model and random forest classification model were established according to the content of different chemical components, and Fisher linear discriminant model, which is easier to test for sensitivity analysis, was selected. Next, the clustering results were obtained by systematically clustering high potassium glass and lead-barium glass, and the classification basis was determined by drawing a line graph and then establishing a subclass classification model to classify high potassium glass into low potassium subclasses and high calcium subclasses. Sensitivity analysis was performed by changing the values of key parameters to check the reasonableness and stability of the subclass classification model. A discriminant model was established to initially determine the categories of eight unknown classified glass artifacts. Sensitivity analysis was performed by changing the key variables through the gradient descent method, and the model was obtained to have good sensitivity.

Keywords: Glass artifacts; correlation analysis; variance analysis; Fisher's linear discriminant model; sensitivity analysis.

1. Introduction

Glass is an important physical evidence of cultural exchanges between China and the West along the Silk Road[1]. Early glass was introduced to China from Egypt and West Asia and other regions, and our ancient working people absorbed the technology and made glass products with similar appearance but different chemical composition[2].

Quartz sand as the main raw material for glass manufacturing, due to its high melting point, the ancient refining of glass often need to add natural soda, lead ore, grass ash and other fluxes to reduce the refining temperature. Lead ore as a flux made of glass with a high content of lead and barium called lead-barium glass[3]. Grasswood ash as a flux made of glass with a high content of potassium called potassium glass. During the weathering process of ancient glass, internal elements are exchanged with environmental elements to change its composition, so the weathering of glass is highly susceptible to the influence of the burial environment, and the difference in the burial environment affects the correct judgment of the type of glass[4].

In this paper, we explore the basis of classification of lead-barium glass and high-potassium glass based on the data, and derive the methods and results of appropriate subclass classification for each category, and test them. On this basis, we identify the types to which unknown categories of glass artifacts belong through chemical composition analysis, and analyze the sensitivity of the classification results.

2. Model assumptions and notation

2.1 Assumptions

1. When investors consider each investment choice, it is based on the probability distribution of sec 1. The sampling of artifacts is random.
2. the data with the sum of the proportion of components between 85% and 105% are considered as valid data.
3. do not consider the artifacts buried in high temperature, high humidity and other special environments.
4. do not consider the glass artifact samples in the sampling process weathering.

2.2 Notations

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

Table 1 Notations

Symbols	Description
C	Coefficient matrix
A	Post-weathering chemical content matrix
B	Pre-weathering chemical content matrix
U	Error range matrix
K	Elbow-law graphical distortion points, number of subcategories
W^T	Transpose of the optimal projection direction matrix W
X	Sample data matrix

3. 3. Model construction and solving

3.1 Establishment and solution of classification model for high potassium glass and lead-barium glass

Step1:Random forest classification model building and solving

The random forest decision tree training process is shown in Figure 1.

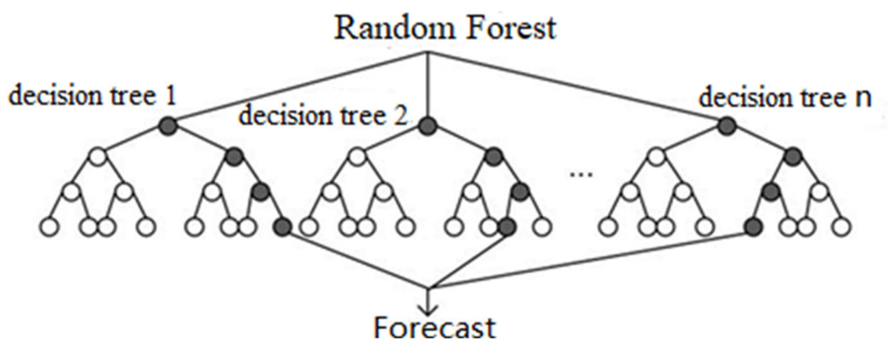

Figure 1 Random Forest Decision Tree Training Process

Importing the data into SPSS to solve the model yields. Heat map of training data is shown in Table 2

Table 2 Heat map of training data

Type	High Potassium	Lead Barium
High Potassium	18	0
Lead Barium	0	28

The good fit of the model can be seen from the heat map of the training set.

By testing the test set, the model has a good prediction effect and the prediction accuracy is 100% with high accuracy.

Step2:Fisher linear discriminant model building and solving

Fisher linear discriminant can be regarded as projecting all samples to a direction, and then determining a classification threshold in this one-dimensional space so that the two classes are as far apart as possible after projection, while the samples within each class are clustered as much as possible[5]. Accordingly we solve for the optimal projection direction and then we can obtain the discriminant function as follows.

$$g(X) = W^T X + b \tag{1}$$

Importing the data into SPSS, Table 3 and 4 can be obtained

Table 3 Training set prediction accuracy

Prediction Category	**Sample size**	**Precision**	**Recall**	**F1-score**
Lead Barium	42	100.00%	100.00%	100.00%
High Potassium	18	100.00%	100.00%	100.00%
Aggregate	60	100.00%	100.00%	100.00%

Table 4 Test set prediction accuracy

Prediction Category	**Sample size**	**Precision**	**Recall**	**F1-score**
Lead Barium	7	100.00%	100.00%	100.00%
High Potassium	0	0.00%	0.00%	0.00%
Aggregate	7	100.00%	100.00%	100.00%

The ratio of training set and test set is 9:1, i.e., 90% of the data is used to train the fitted model, and the remaining 10% is used to validate the model. From the data in the table, we can get that the correct rate and recall rate of each category of the training set and test set as well as the comprehensive weighted index of both are 100%, and the model fitting effect of the data is very good, the fitting quality is high and the model is reasonable and accurate[6].

The discriminant function is also obtained as follows.

$$g(X) = W^T X + b \tag{2}$$

Bringing in the sample values, g(x) > 0 is high potassium glass and g(x) < 0 is lead-barium glass, from which the two types of data can be discriminated.

Step3:Comprehensive comparison of models

By solving the above two models, we can see that both the Fisher linear discriminant model and the random forest classification model fit well and predict more accurately. However, the Fisher linear discriminant model has a discriminant function, which is more intuitive and simpler to predict which category the unknown sample data is, and facilitates the solution of subsequent problems. At the same time, because of the discriminant function, it is easier to analyze and test the sensitivity, so the Fisher linear discriminant model is finally chosen in this paper.

3.2 Establishment and solution of subclass classification model

In this paper, the subclasses were divided based on the similarities and differences of the contents of different chemical components within the same type of glass, and the data were imported into SPSS for systematic clustering by squared Euclidean distances to obtain aggregation coefficients and spectral plots. The aggregation coefficients were imported into Excel to make a line graph of aggregation coefficients, and the K values were determined by the elbow rule to obtain the following Figure 2 shown.

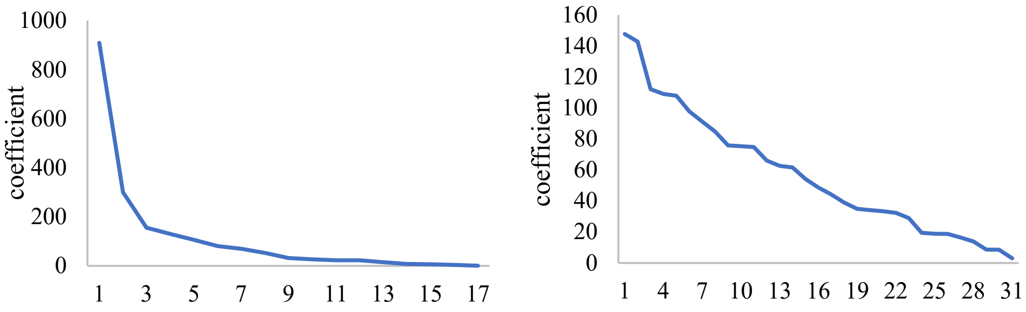

Figure 2 High potassium glass and lead-barium glass polymerization coefficient folding line graph

According to the elbow rule, observing the folding diagram of polymerization coefficient of high potassium glass, it can be seen that the degree of aberration is greatest when the K value is from 1 to 3. After exceeding 3, the change of aberration degree decreases significantly, so the elbow is K=3,that is, the high potassium glass is divided into 3 subclasses. Observing the folding graph of polymerization coefficient of lead-barium glass, we can see that the decreasing trend of the folding line slows down when K=3, that is, the lead-barium glass is divided into 3 subclasses. Therefore, the number of subclasses of both high potassium glass and lead-barium glass is determined to be 3.

When K=3, the spectral diagrams of high potassium glass and lead-barium glass are shown in Figure 3 below.

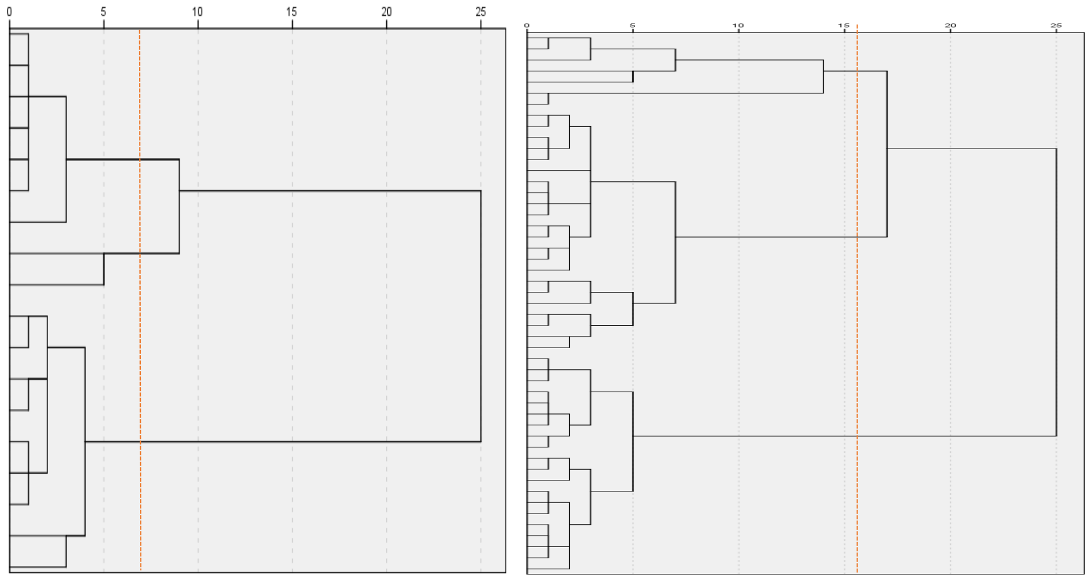

Figure 3 High potassium glass and lead-barium glass spectra

Figure 4 below shows a line graph of the chemical composition content of each glass artifact in the three subcategories of lead-barium glass.

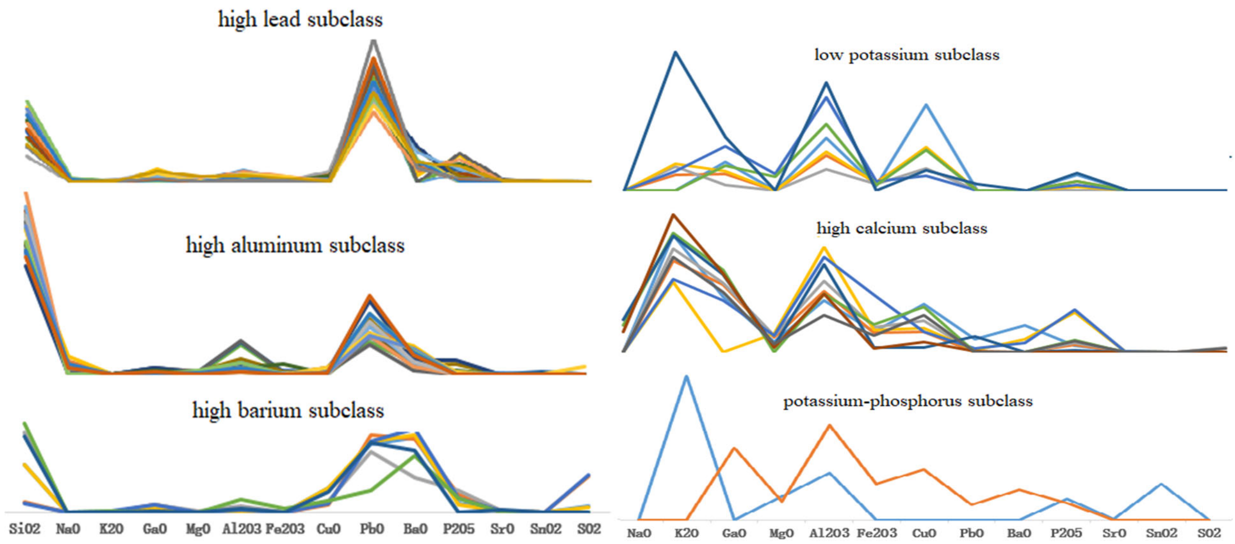

Figure 4 Chemical composition of lead barium glass and high potassium glass by subcategory

By analyzing the chemical composition content of lead barium glass, it is found that the lead content in the first subclass is higher compared with other chemical composition content, and its lead content is also higher than the other two subclasses, so the first subclass is named high lead subclass. The other subclasses are named in the same way, which are high aluminum subclass and high barium subclass.

Since the silica content is the main raw material, the content is relatively large, and in order to analyze the content of other chemical components more intuitively, the silica content is excluded from this figure. Through the analysis of the chemical composition content of high potassium glass artifacts, it is found that the potassium content of the first subclass (above) is less than other major chemical components and lower than the potassium content of the other two subclasses, so the first subclass is named low potassium subclass. The other subclasses were named in the same way, as high calcium subclass and potassium-phosphorus subclass, respectively.

3.3 Rationality and sensitivity analysis of subcategory classification

Step1:Rationalization analysis

In this paper, through the relevant literature, combined with the results of each subcategory division, to further explore the rationality of the results of the subcategory division method in this paper. Observe the histogram of the number of glass artifacts included in each subcategory[7-9].

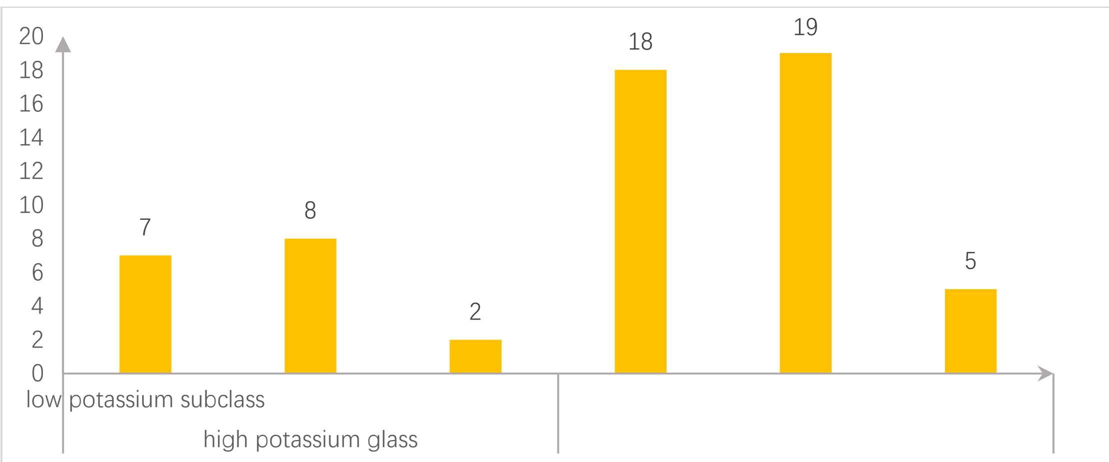

Figure 5 Bar chart of the number of glass artifacts included in each subcategory

From Figure 5, it can be seen that the highest content is in the high lead subclass and high aluminum subclass of lead-barium glass. It is clear from the literature[10] that lead-barium glass is an ancient glass with Chinese characteristics, and the content of aluminum trioxide is high from the spectral analysis of excavated ancient lead-barium glass, and the high-lead silicate glass excavated in China is much older than the surrounding countries, and most of them are high-lead and high-aluminum glass. Therefore, the classification of subclasses in this paper is reasonable in accordance with the actual situation.

Step2:Sensitivity analysis

In order to judge the sensitivity of this subclass classification model, this paper changes the content of silica, the main component of glass, by gradient, and calculates the K value corresponding to each gradient, and observes the change in K value to judge whether the model is sensitive.

In this paper, the gradient is 10%, and the following Table 5shows the K-value for each 10% decrease in silica.

Table 5 K value change table

High Potassium Glass SiO_2	**K**	**Lead barium glass** SiO_2	**K**
100%	3	100%	3
90%	3	90%	3
80%	3	80%	3
70%	3	70%	3*
60%	2*	60%	2
50%	2	50%	2

Note: * represents the start of a change in the category or number of clusters

Combining the K-value solving process and the results in Table, the K-value changes when the silica content decreases to 60% in high potassium glasses, and when the silica content decreases to 70% in lead-barium glasses, although the K-value does not change but their subclass types change. Since the general weathering process does not cause the silica content in glass to drop to 70% or below, the subclass classification model established in this paper is considered to be stable, accurate and reliable, and can be widely applied.

3.4 Testing the reasonable stability of the subclass classification model

According to the established discriminative model, the category to which the glass artifact belongs can be judged, and the chemical composition content of unclassified glass artifacts is imported into the model. When the judgment value is greater than 0, the category to which the glass artifact belongs is high potassium, and when the judgment value is less than 0, the category to which the glass artifact belongs is lead barium. Among them, artifacts A2, A3, A4, and A5 are lead-barium glass, and artifacts A1, A6, A7, and A8 are high-potassium glass.

Because the burial environment has a large impact on the chemical composition of glass artifacts, changes in the environment may change the content of key components and thus affect the judgment of the discriminatory model. In order to further determine the category to which the glass artifacts belong, a bar graph of the chemical composition of glass artifacts was drawn and compared with the bar graph of the mean value of the chemical composition of each of the two categories of glass in Problem 1 to determine whether the classification of the category is reasonable. After discriminating the unknown glass artifacts into broad categories, the data were added to the corresponding Excel sheets and imported into SPSS for cluster analysis.

The difference between high potassium glass and lead-barium glass is mainly the use of different fluxes when melting, so the classification is mainly based on the chemical composition content of the glass before the exchange of substances with the outside world, but in reality, the glass is bound to interact with the outside world, and the internal chemical composition is bound to change, so the higher the stability of the model the more accurate the judgment results. From the above table, the classification did not change even when the lead oxide content increased to 30%, so it can be considered that the Fisher linear discriminant model established in this paper is stable and has a good sensitivity analysis.

4. Conclusion

In this paper, the Fisher linear discriminant model is used to solve the problems of analyzing the classification laws of glass types and discriminating the types to which eight unknown types of glass artifacts belong, which simplifies the computational intensity and enhances the applicability of the model.

The composition analysis and identification of glass artifacts based on this paper can also be applied to the composition analysis and clustering identification of other artifacts, such as the change and prediction of the chemical composition of porcelain artifacts before and after weathering, because both glass and porcelain belong to silicate materials, and only some parameters need to be modified according to the data.

And the linear function for the chemical composition with small content changes before and after weathering, can not be well predicted before its weathering that content components.

References

[1] Dong Ximing. Analysis of the method of calculating correlation coefficients between two fixed categories of variables[J]. Jiangsu Statistics, 1997(05):15-16.

[2] Huang Zhaowang. Analysis of the distribution characteristics and influencing factors of differential weathering of granite in a wharf project[J]. China Water Transport(Second Half),2021,21(06):98-99.

[3] Li Mo. Preparation of Felons'Products and Research on Lead-Barium Glass during Warring States, Qin and Han Dynasty [D]. Beijing University of Chemical Industry, 2014

[4] Zhou C.-H. Machine Learning. Tsinghua University Press [M].2016

[5] Gan Fuxi. The ancient Silk Road and ancient Chinese glass [J]. Nature Journal,2006(05):253-260+246.

[6] Wang Chengxuan,Tao Ying. Weathering of silicate glass[J]. Journal of Silicates,2003(01):78-85.

[7] Liu Lu. Research on Location of Linear Demand Logistics Node Based on Euclidean Distance [D]. Beijing Jiaotong University, 2017.

[8] Li Qinghui, Gan Fuxi, Gu Donghong. Several Issues on the Study of Ancient Chinese Glass [J]. Research on the History of Natural Science, 2007 (02): 234-247

[9] Li Qinghui, Huang Jiaozhen, Li Fei, Gan Fuxi. Detection of chemical composition of a batch of ancient glass samples of the Warring States Period unearthed in China [J]. Cultural Relics Protection and Archaeological Science, 2006 (02): 8-13

[10] Wang Jie, Li Mo, Ma Qinglin, Zhang Zhiguo, Zhang Meifang, Wang Julin. Weathering research of an octagonal prism-shaped lead-barium glassware during the Warring States Period [J]. Glass and Enamel, 2014, 42 (02): 6-13.