Composition Analysis and Identification Study of Ancient Glass Products

Jingrui Yi[(✉)], Liya Song, and Jiapeng Meng

Shenyang University of Chemical Technology, 110142 Shenyang, China
`2115867373@qq.com`

Abstract. This paper addresses the problem of compositional analysis and classification identification of excavated ancient glass artifacts, using data visualization, chi-square test, support vector machine (SVM) algorithm, logistic regression algorithm and random forest algorithm to establish a mathematical model for artifact identification, as well as further refinement of the classification considering the expertise related to ancient glass artifacts, in order to provide a reference for the identification of ancient glass artifacts. The highlights of this paper are: firstly, this question uses a variety of models to train and predict the data, realizing the mutual test of the prediction results and satisfying the accuracy of the established model; secondly, a normal distribution test is conducted on the data chemical composition data, screening out a suitable test method for the subsequent data testing and analysis; finally, a clustering analysis model and the existing literature are combined to make the excavated ancient glass.

Keywords: Ancient glass analysis identification · BP neural network model · Cluster analysis model · Significance analysis · Correlation analysis · Mathematical model

1 Introduction

1.1 Background of the Problem

The main component of glass is silicon dioxide ($SiO2$), and because its raw material, quartz sand, has a high melting point, people usually add fluxes as auxiliary materials in the refining process. In our ancient glass is precisely the production of different fluxes used in the production of the main two kinds: the first is the use of lead ore flux process to Chu culture on behalf of our unique varieties of glass containing lead oxide (PbO), barium oxide (BaO) composition: lead barium glass. The second is the addition of grass ash as a flux composition in the high content of potassium (K) elements, widely distributed in China's Lingnan region, the South Central Peninsula and South Asia and other regions of the glass species: potassium glass. Ancient glass buried in the ground as cultural relics is susceptible to weathering, mainly due to chemical reactions of varying degrees between the metallic elements in the glass and the elements contained in the soil [1] The weathering is mainly manifested in two aspects: first, the glass objects are not severely weathered, and their color and decoration are clearly recognizable, but the presence of inconspicuous weathering cannot be excluded. Second, the surface of glass products weathering, divided into obvious weathering areas and general weathering areas.

1.2 Problem Description and Analysis

1.2.1 Problem Description

Based on the problem background and available data the following questions need to be addressed.

1, Analyze the relationship between surface weathering and glass type, decoration and color of glass artifacts; find the statistical pattern between these chemical elements and weathering through the distinction of glass elemental composition; predict their chemical composition content before weathering based on weathering point detection data.

2, Analyze the classification laws of two major categories of ancient glasses, high potassium glass and lead-barium glass, based on the attached data; select suitable chemical compositions for each major category for subcategory classification, provide specific classification methods and classification results, and analyze the reasonableness and sensitivity of the classification results.

1.2.2 Problem Analysis

(1) Analysis of Question 1

First, a visual analysis table of color, decoration, and surface weathering by glass type was made using frequency analysis. Secondly, a chi-square test was performed on the data in the table to determine whether there were significant differences in the types of artifacts (high potassium, high lead), color, and decoration for different degrees of surface weathering. Finally, a decision tree classification algorithm was used to further analyze the correlation of the above categorical variables. Next, the data in the annexes were processed for descriptive analysis and statistical analysis. Finally, a machine learning based neural network was established using the available data to effectively predict the chemical composition content of the sample test data before weathering.

(2) Analysis of Question 2

Firstly, unlike the analysis of the statistical law of chemical composition of samples focusing on the presence or absence of weathering in Problem 1, the classification work needs to be carried out by relevant classification laws such as: different ornamentation in the two glass classifications, different percentages of the two classifications, and the content of different chemical compositions and other indicators. Secondly, unsupervised, observational cluster analysis is used to learn the data and combine the ancient glass classification expertise to establish a mathematical model of cluster analysis + subclass division to determine the classes that cannot classify the samples. Finally, the sensitivity as well as the reasonableness of the obtained results are analyzed.

2 Mathematical Model Assumptions

(1) It is assumed that the errors arising from the processing of the data have a negligible impact on the results.

(2) It is assumed that the surface weathering of glass artifacts is only related to their glass type, decoration and color, without considering factors such as the time of burial of the artifacts.

3   Model Building and Solving

3.1   Model for Question 1

3.1.1   Data Preprocessing

In order to facilitate the subsequent modeling of this paper, the invalid data collected were eliminated, and frequency analysis and visualization of the data were performed to facilitate the subsequent testing and descriptive analysis of the data. The frequency, percentage, effective percentage and cumulative percentage of the valid data of lead barium glass and Korok glass by color, decoration and surface weathering were obtained from the visualization.

3.1.2   Model Ideas and Analysis

The visualized tabular data (Table 1) were first subjected to a chi-square test using the SPSS tool to test for significant differences between color, decoration and type of glass material and weathering type (weathered, unweathered) variables, respectively [2], with the following analysis steps.

(1) Analysis of whether the chi-square test showed significance ($p < 0.05$).
(2) If significance is presented, it is described specifically according to the percentage of differences in the categories.
(3) If significant, an in-depth quantitative analysis of the differences according to the effect indicators can follow

According to the inspection results (Table 2), the significant P value of surface weathering and color is 0.428, which is not significant horizontally, so there is no significant difference between surface weathering and color data; The significance P value of surface weathering and type is 0.049**, which is significant horizontally, so there are significant differences in surface weathering and type data; The significant P value of surface weathering and ornamentation is 0.057*, which is not significant at the level, so there is no significant difference in the data of surface weathering and ornamentation. (where: ***, ** and * represent the significance level of 1%, 5% and 10% respectively).

After chi-square test, the decision tree classification analysis model is established. The segmentation feature selected at each decision node of the decision tree determines the final classification result, so the model can be evaluated by testing the data classification effect. The model building steps are as follows:

1. determine variable X: {decoration, color, type}, variable Y: {surface weathering} and model parameters
2. Establish the decision number classification model by training set data, and get the decision tree structure.

Table 1. Visualization of effective data of barium glass decoration (From 2022 Mathematical Contest in Modeling)

Engraved pattern	Frequency	Percentage	Effective percentage	Cumulative percentage
A	14	38.9	38.9	38.9
C	22	61.1	61.1	100.0
Amount	36	100.0	100.0	

**Table 2.** Chi-square test analysis of color and surface weathering type (Self-painting)

Color	No weathering	Weather	Amount	X^2	CorrectionX^2	P
bluish-green	6	9	15	7.011	7.011	0.428
light blue	6	12	18			
purple	2	2	4			
dark green	3	4	7			
dark blue	2	0	2			
pale green	2	1	3			
black	0	2	2			
green	1	0	1			
bluish-green	22	30	52			
Amount	6	9	15			

The figure shows the decision tree structure, and the internal nodes give the specific segmentation of the branched features, that is, according to a certain tangent value of a feature:

① gini/ information entropy is used to determine which feature to segment.

② Sample type distribution is the number of samples belonging to each classification group in this node. For example, [10, 5, 5] indicates that there are 10, 5 and 5 samples in three classification groups.

③ Classification is the classification group to which the samples of this node are uniformly divided.

3. Calculate the feature importance through the established decision tree (Fig. 1).

The established decision tree classification model is applied to the training and testing data, and the partial probability assessment results of the model are obtained (Table 3).

Here, good data results are obtained through the training and testing of the decision tree model. By calculating the importance of features, the importance ratios of surface weathering, ornamentation and material types are 60% and 40% respectively, which are mutually verified with the significance level analysis of ornamentation and material types in chi-square test, thus improving the accuracy of the analysis conclusion. The data
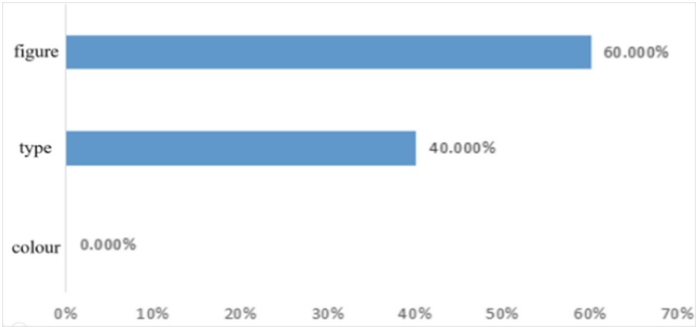
Fig. 1. Importance of color, ornamentation and type characteristics (Self-painting)

Table 3. Decision tree classification and probability prediction (Self-painting)

Prediction resultY	Surface weathering	Prediction Probability _ No Weathering	Prediction probability _ weathering	Engraved pattern	Color	Type
No Weathering	No Weathering	1	0	3	4	2
No Weathering	No Weathering	1	0	3	1	2
weathering	No Weathering	0	1	1	1	1
weathering	No Weathering	0	1	1	1	1
weathering	No Weathering	0	1	1	1	1

test results in Table 4 give the classification group with the largest prediction probability under the condition of the above data verification, and the probability of weathered and unweathered prediction results is represented by 0 and 1.

Secondly, it is necessary to statistically analyze and confirm the statistical characteristics of the chemical composition content on the surface of cultural relics, and use descriptive methods to summarize and describe the overall situation of the chemical composition data and the characteristics of the data. The basic data indicators described include 14 different chemical components: sample size, minimum value, maximum value, average value, standard deviation and median value from silica (SiO_2) to sulfur dioxide (SO_2) (Table 5). The analysis steps are described as follows:

1. 1, the overall description of the analysis of the average score;
2. Focus on the items with higher average value;
3. Summarize the analysis.

Median Comparison Chart
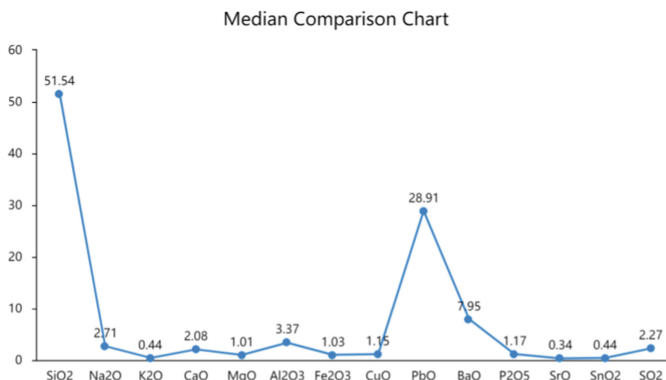
Fig. 2. Median comparison chart (self-painting)

In the actual analysis, it is found that the maximum values of four items, namely alumina (Al_2O_3), copper oxide (CuO), barium oxide (BaO) and phosphorus pentoxide (P_2O_5), exceed the average values by 3 standard deviations, indicating that the data fluctuates greatly, and the relative average value should be selected as the index to make a comparison chart of the average values of 14 chemical components (Fig. 2). It can be seen from the figure that silica (SiO_2) and lead oxide (PbO) as main components have the most prominent median contrast.

Further analysis of the basic indicators can obtain in-depth indicators, including the mean standard deviation, variance, 25th percentile, median, 75th percentile, standard error, mean 95% CI(LL), mean 95% CI(UL), IQR, kurtosis, skewness, coefficient of variation (CV) and P value of different samples.

Finally, the chemical composition of the samples before weathering is predicted based on the detection data of the weathering points. In this question, the matlab tool was used to build the prediction model of the artificial neural network based on the sample data.

In order to make the established BP neural network more accurately predict the chemical composition content before weathering, it is necessary to adjust the interlayer node threshold and weighting, so that the error between the actual output and the expected output after the sample data import is stable within a relatively ideal value. Training steps [3] as follows:

Initialization network and parameter learning were performed.

Design and provide data training methods and build training networks until the expected results are obtained.

Enter the data into the trained model and compare the output value to the expected value. If there is an error, enter the fourth step and return to the second step (Fig. 3).

Calculate the error of the same cells, correct weights and thresholds, return to the second step. This is a backpropagation process. In this question, the BP neural network model based on the Levenberg-Marquardt algorithm tool in matlab is used for data training, and the training results are obtained as follows (Fig. 4):

In this training, all weathered samples were selected as input data and the corresponding number of weathered samples as output data. Of these, 70% were training

Fig. 3. Neural network training process graph (Self-painting)
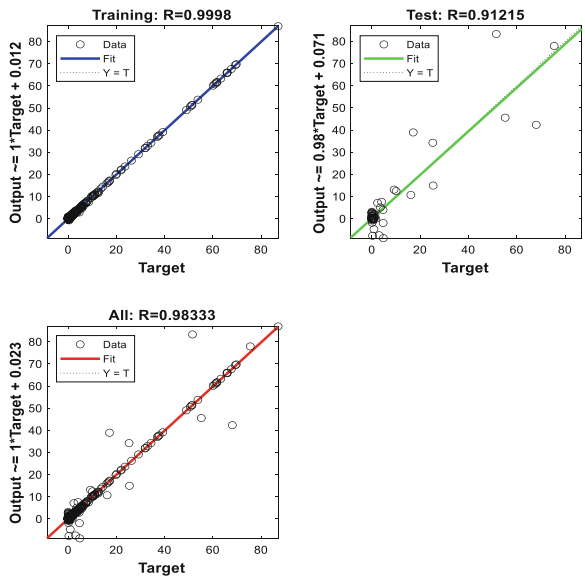

Fig. 4. Training set, test set, validation set fitting the effect map9 (Self-painting)

set, 15% were, test set, 15% were cross-validation set, with 10 hidden nodes, maximum iterations of 1000, and 3 retraining. And the data are not overfit, the accuracy of the model is high.

3.2   The Model for Problem 2

3.2.1   Description of the Indicator

According to the two types of glass indexes, high-potassium glass and lead-barium glass, each classification includes 14 chemical composition indexes: SiO_2, K_2O, Fe_2O_3, PbO, Na_2O, P_2O_5, BaO, CaO, MgO, Al_2O_3, CuO, SrO, SnO_2, SO_2.

3.2.2   Model Ideas and Analysis

Data from 14 indicators in both groups were first tested for normality to determine their significance. If the data does not show significance (p-value greater than 0.05 or 0.01, strictly 0.05, not strictly 0.01), indicating that the normal distribution, and otherwise does not conform to the normal distribution. Using the SPSS tool to test the data normality is as follows: The 14 indicators did not meet the normal distribution, so the t-test cannot be used to give the significance level to make decisions. Instead, the data were analysed using a non-parametric test. From the non-parametric inspection and analysis results (see supporting materials for details):

① same sample for Na_2O, CaO, MgO, Al_2O_3, P_2O_5, SnO_2, SO_2 A total of 7 items do not show significance ($p > 0.05$), meaning that different types of samples are for Na_2O, CaO, MgO, Al_2O_3, P_2O_5, SnO_2, SO_2 All showed consistency and no difference.

② type samples for SiO_2, K_2O, Fe_2O_3, CuO, PbO, BaO, SrO Total 7 items are significant ($p < 0.05$), meaning that different types of samples are for SiO_2, K_2O, Fe_2O_3, CuO, PbO, BaO, SrO are different.

Half of the indifference indicators were finally removed, leaving seven chemical indicators that were different by sample type.

A cluster analysis algorithm was then used in combination with reference references [4] The subclasses of the two types of glass samples were divided from a professional background. The cluster analysis steps are as follows:

The objective entropy weight method is used to screen the two types of indicators, and eliminate the part with the weight of less than 3% and the characteristic compounds of glass: such as high potassium glass, SiO_2 should be removed, K_2O ingredient. The remaining 18 data were analyzed and each sample was numbered (see the Supporting Material for the serial number). Chemical composition screening results are shown in Fig. 5:

Using the SPSS tool to import two types of data, using the four distance analysis methods of Q-type cluster analysis: square European distance, European distance, Cherbyshev distance, Chebyshev distance, and Pearson distance, respectively, we obtained the four cluster plots of the two types of glass (Fig. 6).

Classification method according to actual needs (see supporting materials for details): for example, the European distance of high potassium glass is divided into 2 and 4 categories in the Fig. 6 (the results are Table 4).

After consulting the professional data of ancient glass, the two types of glass were further classified into glass subcategories (take high potassium glass as an example):

① summarized the samples included in method 1 square Euro Type 2, Euro Type 2, Chebyshev Type 2, Pearson Type 2, Method 2 and Euro Type 4, and found P in samples 7,12,13,17,18$_2O_5$Content greater than 1 and no Na_2O, summarized as a high phosphorus
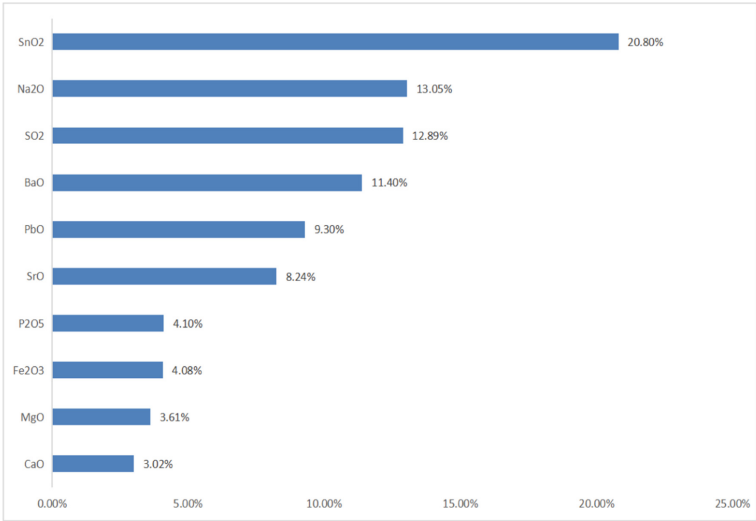
Fig. 5. Screening results of the chemical composition of high-potassium glass (Self-painting)
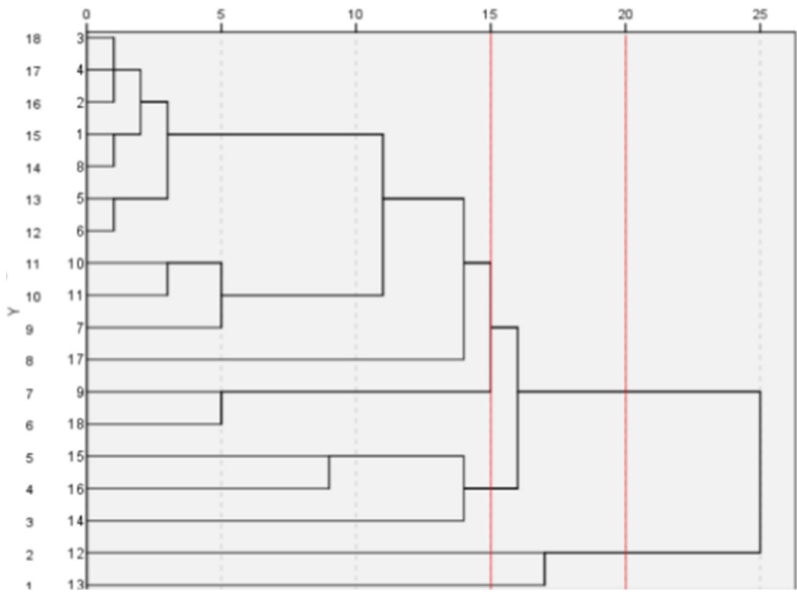
Fig. 6. Cluster diagram of the square distance of high potassium glass composition (Self-painting)

class.② comparison and summarizes the samples included by Pearson category 3 in Method 1, Method 2 Middle-European category 3 and Pearson category 314,15,16It contains Na2O, summarized as oxidative oxide.③ made a comparative summary of the samples included in Pearson category 1 in Method 1 and Method 2, and found that samples 9,10,11 samples Fe2O3The content is greater than 2Summarize it as the

Table 4. Two clustering methods of European distance of high potassium glass (Self-painting)

Select the location	20		15			
Individuals included in the category	1, 2, 3, 4, 5, 6, 7, 8 9, 10, 11, 14, 15, 16, 17, 18	12, 13	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 17, 18	9, 18	14, 15, 16	12, 13
Total number of individuals	16	2	11	2	3	2

Table 5. Results of high-potassium glass subclass (Self-painting)

type	condition	Includes samples	number of samples
High phosphorus class	P_2O_5Content > 1 Na_2O content $= 0$	7, 12, 13, 17, 18	5
Oxidation of the class	Na_2O content $= 0$	14, 15, 16	3
High-speed rail class	Fe_2O_3Content > 2	9, 10, 11	3
Low phosphorus and low iron	P_2O_5Content < 1 Fe_2O_3Content < 2 Na_2O content $= 0$	1, 2, 3, 4, 5, 6, 8	7

high-speed rail class.④ compared and summarized the samples included in Method 1 Middle-European category 1, Method 2 Middle-square Euro category 1, Euro category 1 Chebyshev category 1, and Pearson category 1, and found P in samples 1, 2, 3, 4, 5, 6, 82O5 The content of which is less than 1, Fe2O3. The content is less than 2 free from Na2O, summarized as low phosphorus and low iron class.

For the categories obtained by individual methods, there can not reflect obvious differences in the element content from other categories, and the reference value is not large, so it is abandoned. The final classification results are formed as follows (Table 5):

3.2.3 Analysis of Model Sensitivity and Rationality.

1. Sensibility analysis:
   Use the four distances in the model, and the results can be tested against each other.
2. Rationality analysis:
   This question is classified by Ancient Chinese Glass Technology [4]. The evolution of the glass composition in ancient China divided the glass into the original porcelain glaze (K_2O_2) Oil sands Glass sand (Na_2O-CaO-SiO_2) Potassium-silicate glass (K_2O-SiO_2), Lead-barium silicate glass (PbO-BaO-SiO_2), Potassium-lead-silicate glass (K2O-PbO-SiO2), Potassium, calcium, silicate glass (K_2O-CaO-SiO_2) According to the introduction of the chemical composition characteristics of the ancient glass

in the literature, it can be judged that the subclassification of glass in this paper has a certain rationality and accuracy.

4 Conclusions

4.1 Advantages and Disadvantages of the Model

4.1.1 Advantages of the Model

1. Use excel to organize the data and eliminate invalid values.
2. The decision tree model is established, which can be verified through numerical statistics and testing, which makes it possible to explain and verify the reliability of the conclusions drawn by the model.
3. A neural network model was established to predict the chemical composition, making the prediction results scientific and reasonable.
4. Using the four distance analysis methods of Q-type cluster analysis in SPSS, the results obtained can be tested against each other, thus reducing the impact of each method on the results and making the results more accurate and objective.
5. Support vector machine (SVM), logical regression and random forest three methods are used to identify the data and determine whether the data is overfit, which greatly improves the reliability of the retained data.

4.1.2 Disadvantages of the Model

1. The use of the decision tree model is easy to produce the overfitting phenomenon, which affects the accuracy of the results to some extent [5].
2. The subclassification of high-potassium glass and lead-barium glass has a certain subjectivity.
3. The reverse propagation of errors in the neural network model may have some impact on the accuracy of the results

4.2 Optimization and Generalization of the Model

4.2.1 Optimization of the Model

1. The decision tree model can be pruned to delete the cut branches with some influence on the prediction accuracy, and weaken the influence of the overfitting phenomenon on the rationality of the results.
2. In the subclassification of high-potassium glass and lead-barium glass, the data can be reduced in advance to simplify the classification process.

4.2.2 Extension of the Model

1. The Model establishes the classification model of different glass cultural relics, which has certain reference value for the classification of cultural relics in other fields.
2. The prediction model can be applied to predict the chemical composition of other materials.

References

1. Gao Mingqin. Glass weathering and its prevention method [M]. Wuhan: Wuhan University of Technology Press, 1995
2. MAO isong, Wang Jinglong, Pu Xiaolong, et al. Advanced Mathematical Statistics (second edition) [M]. Beijing: Higher Education Press, 2006.
3. Si Shoukui, Sun Zhaoliang, et al. Mathematical modeling algorithm and application (2nd edition) [M]. Beijing: National Defense Industry Press, 2015
4. Gan Fu Xi et al. Ancient Chinese Glass Technology [M]. Shanghai: Shanghai Science and Technology Press, 2016
5. Wang Peng et al. Research on Classification of Ancient Prescriptions Based on Decision Tree Algorithm [J].. Beijing: China Computer & Communication,2021,