Composition Analysis and Identification of ancient glass products

Hongpeng Wang[#], Xiang Li[#], Jiaju Kang[#,*], Yiming Qiao

College of Computer Science and Technology, Shandong Jianzhu University, Jinan, Shandong, 250101 China

[*]Corresponding author: kjj_python@163.com

[#]These authors contributed equally.

Abstract: This paper studies the chemical composition of two major types of glass products in ancient China, and the constituent structure of residual substances in a long-term weathering environment, and proposes a classification method based on composition analysis. When researchers provide unknown categories of ancient glass samples, this results can be used to make type judgment based on their chemical composition and assist generation and other work. Based on the quantitative analysis of the chemical composition of different sampling sites, the glass sample classification model is tested by using the bp neural network and the decision tree classification model and ID3 mechanism. When analyzing the problem, we found that some of the data in the attachment did not meet the requirements of the problem setting, and even had missing values. Therefore, we cleaned the data, completed the null value and eliminated the wrong data with 0, which was embodied in lines 15 and 17 of Table 2.Using Kappa consistency matrix analysis, it is an obvious connection between the surface weathering of glass cultural relics and their types, decoration, color and other characteristics. The discrete dot plot for pooling analysis was drawn using data characteristics of surface chemical composition changes before and after weathering. To predict the chemical composition after weathering, the prediction formula was established using vector iteration. Later, we used the vector iteration method to establish the weighted average ratio prediction formula to predict the chemical composition of the glass samples after weathering.

Keywords: Kappa Consistency Test, Adaboost Model, Bp Neural Network, Decision Tree

1. Restatement of the problem

Analyze the correlation between the surface weathering of cultural relics and the glass type, decoration and color in the question stem data. According to the types of cultural relics, whether the cultural relics are weathered or not is divided, and the change law of the surface composition of the cultural relics is explored. Chemical composition content before weathering is inferred from the weathering point detection results.

Distinguish the types of high potassium / lead-barium glass according to the different content of chemical elements. The distribution law of chemical element content for each category divides subclasses to each category, and its rationality and sensitivity are verified by testing.

Learn the quantitative relationship between glass types and chemical composition in Table 1 and Table 2, using the unclassified data in Table 3.

Explore the relationship between chemical components through the correlation analysis of chemical components, and reflect the difference of the association relationship of different categories of chemical components through the difference of feature importance.

2. Model hypothesis

(1) The error generated by the detection means makes the sum of the sample content acceptable in the range of 85% -105%, and the rest is identified as wrong data;
(2) the material described by the data is pure and no impurities not described in other accessories;
(3) the glass sample is relatively stable during the study, such as not react with the air;

(4) the data given in the accessories is basically accurate;
(5) the measurement error between the data and the real situation.

3. Model building and solution

Problem 1.1 It is required to study the surface weathering of glass relics, the relationship between types, decoration and color, and analyze it. Preliminary observation of the data we can clearly find that there are missing values appearing in the attachment data Table 1, and data completion is required. We observed the performance of other properties of the row data where the missing data was located, and proposed the criterion of "ensuring the consistency of similar data", so as to fit the actual situation as well as possible.(Here need to explain is that in the case of small data, we choose the number as the complete value strategy is relatively simple, very easy to lead to the data of the distribution deviation, late improvement can study the vacancy value in the data table in the possible value prior distribution is distributed in proportion, in order to ensure the rationality of the data distribution after adding the complete data.) After the data completion, we established the Kappa consistency matrix, calculated the correlation coefficients between the attributes, and intuitively expressed the degree of correlation by using the heat map. We used the chi-square test to study the effect of the model after the visualization of the results, which showed that the model can better describe the relationship of the data.

Problem 1.2 It is required to analyze the statistical law of the surface chemical composition content of some cultural relics samples under the condition of weathered environment. In looking at Table 2, we found outliers in lines 15,17 for the error that the sum of the sample content accepted was in the range of 85% to 105% because of the non-absolutely accurate test method. The empty value in Data Table 2 uses "content is 0" to complete the basic data processing. We will weathering before and after weathering as the binary state of a chemical element in the same glass samples, according to the basic classification of glass types, from the sampling objects of cultural relics under different degree of weathering, statistical quantitative analysis of chemical composition, study the frequency, variance, median, mean four indicators, in order to find the law of its composition distribution. At the same time, we studied the correlation between the various components for the glass classification model through the hierarchical clustering method.

Problem 1.3 Sample and detect according to the weathering point to predict the chemical composition content before weathering. We chose to use the vector iteration method to establish a weighted average value ratio prediction formula to predict the chemical composition of glass samples after weathering.

Question 1.1 We studied the data in Table 1. After data completion, we established the correlation coefficient matrix model, analyzed the correlation degree of the appearance properties of the glass cultural relics in the data table, and generated the correlation coefficient matrix and the lower triangle matrix in order to intuitively judge the correlation between the attributes.

Question 1.2 We studied the data in Table 2. After data cleaning, we established the statistical quantitative analysis of the 4 chemical components based on the four indicators, and compared the respective values using the graph array form in order to find the internal laws between the components. Later, we visualized the relationship between the components more intuitively through the hierarchical clustering method.

Question 1.3 We studied the data in Table 1 and Table 2. According to the weathering attribute of the glass samples identified in Table 1, we assigned a new attribute "degree of weathering" to each line in Table 1. Using the vector iteration method, the weighted average value ratio prediction formula was established to predict the chemical composition of the glass samples after weathering.

The kappa coefficient is a measure of the classification accuracy. It is obtained by multiplying the sum of the total images (N) in the confusion matrix diagonal (Xkk) by the sum of the sum of the sum of the total classified categories and by the sum of the total surface images.

Further analysis, due to the different chemical composition content of each type of glass, there may be some chemical content is not detected, so the data overall more empty value, we weighted average data, in the weight calculation part using the standard normal distribution function for weight allocation, the calculation process is as follows:

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{x^2}{2}} \tag{1}$$

Based on the variable type and color, we can see that the significance P-value is 0.000 * * *, showing significance at the level, and rejecting the null hypothesis, indicating that there is consistency between the two variables. Meanwhile, the value of the Kappa coefficient is-0.169, so the degree of correlation is extremely low for consistency. Based on the variable type and ornamentation, we can see that the significance P-value is 0.027 * *, showing significance at the level, and rejecting the null hypothesis, indicating that there is consistency between the two variables. Meanwhile, the value of the Kappa coefficient is-0.106, so the degree of correlation is extremely low for consistency. Based on the variable color and ornamentation, we can see that the significance P-value is 0.000 * * *, showing the significance at the level, and rejecting the null hypothesis, indicating the consistency between the two variables. Meanwhile, the value of the Kappa coefficient is 0.112, so the degree of correlation is extremely low for consistency.

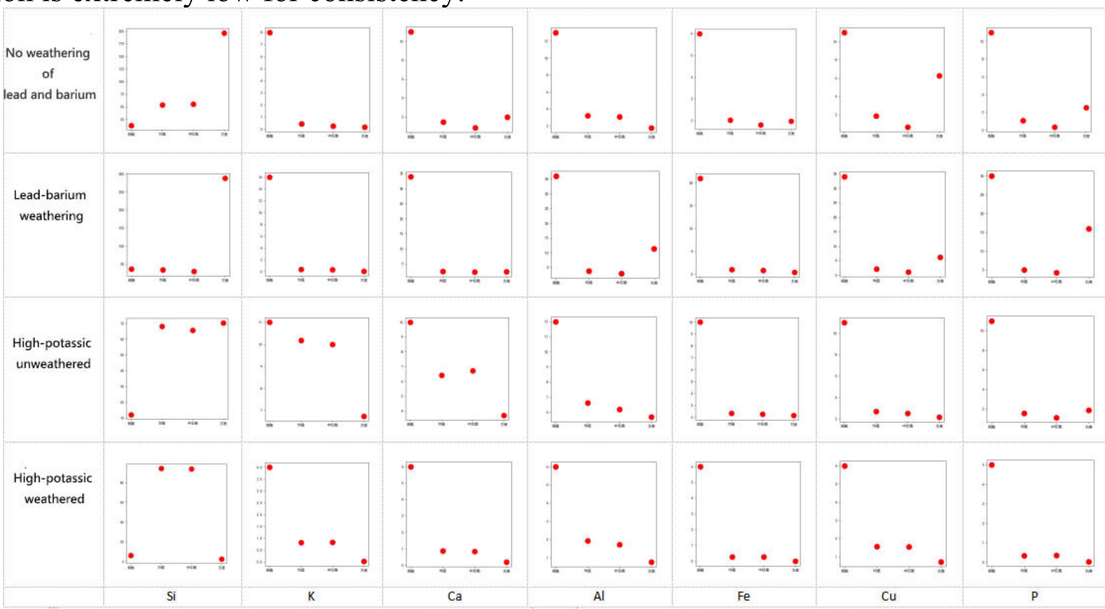

Figure 1. Element content index before and after lead-barium and high-potassium weathering

According to the categories of high potassium and lead-barium, the seven main elements of Si, K, Ca, Al, Fe, Cu and P were extracted, and the frequency, mean, median and variance of different chemical elements were extracted successively. With the difference of weathering, the scatter plot was drawn to show the difference of chemical elements before and after weathering. It can be intuitively seen from the figure array that the content of high potassium glass (silicon, potassium) decreases after weathering; the content of lead-barium glass (aluminum, copper) increases after weathering.

Starting from the chemical composition, to explore the classification of high potassium glass and lead barium glass, to weathering and after weathering, and distinguish, using principal component analysis, the type of chemical composition dimension, to rank higher factors as a representative characteristics of different kinds of glass relics, to k-means clustering algorithm for each classification of three typical categories, and use the multi-factor variance analysis to test the rationality and sensitivity of the results.

High-potassium glass and lead-barium glass before and after weathering were classified according to the adaboost model and the chemical composition.

Principal Components Analysis (PCA), also known as principal component analysis, is a method to simplify the data structure by reducing the dimension: how to turn multiple indexes into a few comprehensive indicators, which can reflect most of the information of multiple indicators. The purpose of the PCA is to take several linear combinations from the original multiple variables to retain as much information in the original data as possible. Here we still use the entropy weight method, and no longer explain it one by one.step 1 Since the index in the attachment is not only a benefit index, but also a cost index, here we first conduct regularization processing to standardize the data in order to eliminate the magnitude of magnitude and magnitude effect of the data.

The K-means algorithm, also known as the K-average or the K-mean, is a wide use of the most basic clustering algorithm, generally as the first algorithm to master the clustering algorithm. Suppose the input sample is; then the algorithm step is (using the Euclidean distance formula): select the k initialized category centers a_1 ,a_2, a_k. For each sample x;, mark it away from the class center a, and the nearest category j updates the center point a for each class; repeat the above two steps for the mean of all samples belonging to the class until an abort condition is reached.

On the basis of the deterministic analysis, the influence of the uncertain factors on the final economic effect index of the investment projects is further analyzed. Subjectively perturbated the data, and the perturbbed data were compared with the original data in QQ plots to analyze the rationality and sensitivity of the results. Before clustering, we first reduced the dimensionality of the number of elements in the data using the principal component analysis method to ensure that the research subjects were focused on significantly affecting the classification rule. But at the same time, the method is sensitive to the outliers in the data, and a large proportion of the principal components will occupy the effectiveness of the secondary components.

The attachment data was processed, the chemical composition was integrated according to the glass type and weathering type, and the data was classified using adaboost.

Due to the large number of vacancies in the original attachment and redundant chemical elements in the category, the results will lead to be inaccurate, so the dimension of chemical types is reduced, and the chemical elements accounting for the main components in each category are extracted by principal analysis, so as to explore the classification rules of different types. The results of the four categories have been included in the attachment, and the results of chemical element content in each category are shown in Figure 2 and 3.

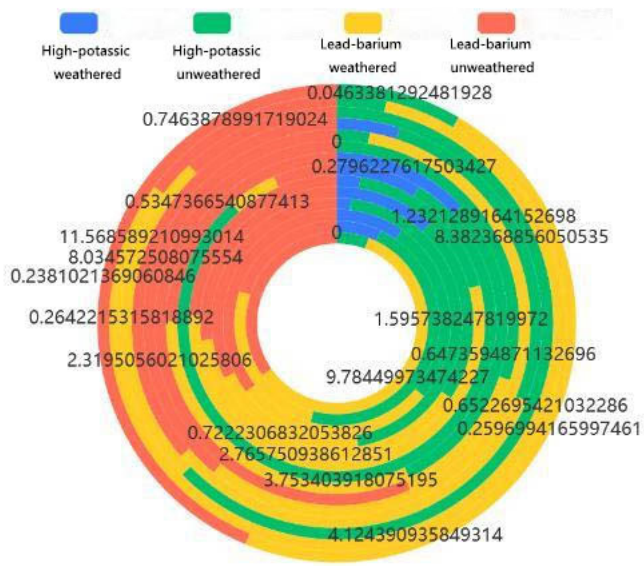

Figure 2. Results of the four-category principal component analysis

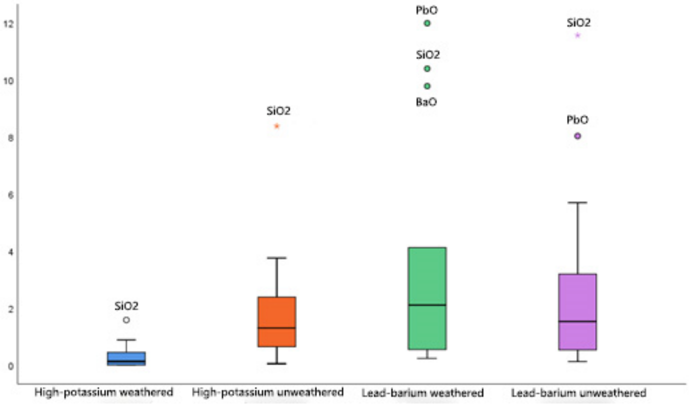

Figure 3. Four categories of display and the main representative elements

It can be seen that the main elements of high potassium weathering are magnesium oxide and alumina, while the representative elements of high potassium unweathering are potassium oxide, calcium oxide and phosphorus pentoxide; the main elements of lead and barium weathering are sulfur dioxide, barium oxide, lead oxide and copper oxide, and the main elements of lead and barium unweathering are strontium oxide and lead oxide. The difference of the content of the main elements indicates the classification law of high potassium glass and lead-barium glass.

The cluster category difference was analyzed according to the field and the frequency of each cluster category was summarized. The p-value was then judged based on the significance p-value of each analysis term, as the example.

Question 3 presents the unclassified glass relics with chemical composition data in order to classify them using the knowledge in the historical data. Here we observe the data. We have adopted the bp neural network and the decision tree classification model to learn the newly generated tables, respectively, in order to identify the classification guarantees in the existence model. After testing, the data in Table 3 meet the numerical requirements for the sum of the components, and can be used. We chose to take whether weathering, various chemical composition and other 15 data as the input, is high potassium / lead barium glass for the output to complete the classification process.

In order to preserve all the data representations reflected in the data in Tables 1 and 2 as much as possible, we chose the bp neural network and the decision tree model to learn about the past data. Relying on bp neural network based on neurons feedforward transmission mechanism, can be completed from the transmission of the input layer to the output layer forward, in the process of the weights of neurons change, the model learned the whole data change from input to output, and input new homogeneous data can be obtained according to the change of historical data classification results. Secondly, we use the decision tree classification model based on the ID3 mechanism to generate the tree structure with a top-down greedy search traversal method, outputting a tree attribute judgment mechanism that matches the training data. As shown in Figure 4.

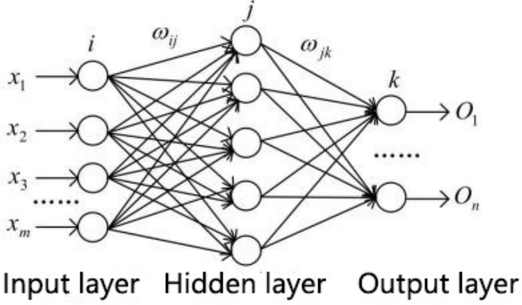

Figure 4. Schematic diagram of the bp neural network

The prediction effect of bp neural networks is measured by quantified metrics. Among them, the evaluation index of the cross-validation set can be constantly adjusted to obtain a reliable and stable model.

According to the analysis of the correlation relationship between the chemical components of different types of glass cultural relics samples, the correlation coefficient matrix is selected to realize it, and the correlation coefficient is also intuitively displayed through visualization. For the difference of chemical composition associations between different categories, the glass relics samples of different categories are distinguished by the different characteristic importance of the problem three different categories, so as to compare the differences of their associations.

Glass samples in high-potassium unweathered category are better characterized in the correlation of alumina-iron oxide and potassium oxide-calcium oxide; those in alumina-iron oxide and phosphorus pentoxide-copper oxide; glass samples in lead-barium unweathered category; and glass samples in lead-barium weathering category are better characterized in the correlation of barium oxide-copper oxide.

4. Scientific analysis of the model

In view of the different requirements of the various questions raised by the topic, The correlation analysis model based on the "the surface weathering of glass cultural relics, its type, decoration and color relationship" was established respectively, Used to evaluate the relationship between the appearance and the surface weathering of several glass cultural relics; A glass artifact classification model based on chemical composition was established, In order to automatically classify the glass samples with known chemical content; The bp neural network and decision tree classification models were developed, The processing discrimination of the classification method for the new data is retained, ; An analytical model of the correlation between the different chemical components was established, Ability to better response components between correlation and variability.

5. Conclusion

Model advantages :This paper aims at the different requirements of each problem raised by the topic, The correlation analysis model based on the "the surface weathering of glass cultural relics, its type, decoration and color relationship" was established respectively, To evaluate the relationship between the appearance and the surface weathering situation of several glass cultural relics; A glass artifact classification model based on chemical composition was established, In order to automatically classify the glass samples with known chemical content; The bp neural network and decision tree classification models were developed, The processing discrimination of the classification method for the new data is retained, ; An analytical model of the correlation between the different chemical components was established, Ability to better response components between correlation and variability.

Disadvantages of the model (1) In the data preprocessing, we used the masses of all values in the dimension of the missing value as the completion value, which is highly likely to affect the distribution of the data on the relatively small dataset of this problem and bring errors for subsequent studies. Therefore, we propose a complete value solution based on the same distribution of data to solve this problem, but we still use the crowd solution to ensure the continuity and logical rigor of the research process.(2) The problem using the data set is relatively small, to a certain extent, can not well represent the actual situation, which makes the model generalization ability is poor, and in the process of this study we found that often appear on the data training set, the test set success rate is 1, which increases the model effect and the actual data appear the possibility of underfitting phenomenon. One way to solve this phenomenon can be to find the relevant data augmentation data set with the same data distribution structure, or to use the existing data to generate the pseudo-data augmentation data set based on the data distribution.

Model promotion (1) Glass is an important material for the study of ore processing and smelting technology in ancient China, and it is also a common research object in the research of folk culture.(2) The method based on the chemical composition of ancient Chinese glass classification proposed in this paper can assist archaeologists to collect evidence at the physical level, and then judge the age of a certain cultural relic unearthed at the same time as the glass samples and the ore smelting level of the people in this age.

References

[1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11) [Online Application Software]. Retrieved from https://www.spsspro.com.

[2] Yang Xuebing, Zhang Jun. Decision tree algorithm and its core technology [J]. Computer Technology and Development, 2007 (01): 43-45.

[3] Zhou Zhihua. machine learning. Beijing: Tsinghua University Press, 2016:pp.121-139,298-300.

[4] Li Hang. Statistical learning methods. Beijing: Tsinghua University Press, 2012: Chapter 7, pp.95-135.

[5] Sinaga, Kristina P. Unsupervised K-Means Clustering Algorithm [J]. IEEE ACCESS,2020,8(2169-3536):80716-80727.