ISSN 2616-5775 Vol. 6, Issue 2: 94-99, DOI: 10.25236/AJCIS.2023.060212

Study of relic classification based on neural network and K mean classification

Yupeng Wang

School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, 400074, China

Abstract: *Over time, its own chemical elements will change in response to environmental changes, commonly known as weathering, which will lead to changes in the chemical composition ratios and affect the judgment of the class. This paper focuses on the chemical composition of different types of glass. By comparing the chemical indicators of each type, an identification model of the cultural relic category is established to predict the chemical content of the weathering process. To solve this, the neural network, chi-square test, and Spearman correlation analysis are used and the calculation results are given. According to the definition of sensitivity, after analysis, the accuracy was taken as the evaluation index, and the type of chemical composition was taken as the factor. K-value cluster analysis was performed for the type of chemical composition of the two glass categories in turn, and the accuracy was compared, and the accuracy of high potassium glass was close to 83%, and the accuracy of lead-barium glass was close to 69%, concluding that the type of chemical composition, which has little effect on the accuracy, has little sensitivity. In different categories, the variation of independent variables has a large impact on the accuracy and a large sensitivity. Finally, the advantages and disadvantages of the model are analyzed, and the model is extended to a larger domain.*

Keywords: *Chi-square test, BP neural network, K-value cluster analysis, Control variable, Spearman's correlation coefficient*

1. Introduction

With the excavation of the history of the Silk Road, glass cultural relics have gradually stepped into the public view [1]. The archaeology of ancient glass relics is moving towards modernization [2,3]. Archaecan measure the chemical composition and content of ancient glass products through instruments, have a more comprehensive understanding of the origin of cultural relics, and facilitate scientific investigation [4]. Over time, the surface chemicals are weathered and exchanged with other elements outside, changing the ratio of color to its components.

Therefore, it is the focus of cultural relics investigation to choose a reasonable analysis method to analyze the composition of cultural relics, and then calculate the type of cultural relics, whether they are imported, and the time period it is in the history of China [5].

Existing archaeologists have divided the cultural relic samples into high-potassium and lead-barium categories, and provided cultural relic data, such as form 1, form 2 and form 3. Please complete the following questions through mathematical modeling:

Problem 1: According to the data given in Form 1 and 2, study the correlation between the weathering and glass type, decorative type and color; count the chemical content of the weathered and unweathered surfaces; and compare the combination of the above results.

Problem 2: Study the classification rules of high potassium category and lead-barium category according to the form data; subclassify the two kinds of glass respectively, and provide reasonable classification basis; test the rationality and sensitivity of the classification results.

Problem 3: Analyze the category of cultural relics in form 3, point out the corresponding categories and verify the sensitivity of the results.

Problem 4: analyze the cultural relics of different categories of glass products, explore the associations of the different chemical components, and compare the differences between the categories.

2. Model hypothesis

In order to build more accurate mathematical models, this paper makes the following reasonable assumptions or conditions according to the actual situation:

When measuring the chemical composition and content, the chemical composition content of the external air will not affect the test results, except for the chemical composition of the given sampling point.

The chemical composition of the cultural relics is negligible after testing.

The testing equipment is in good condition and will not interfere with the chemical composition of the cultural surface during the testing process.

3. Model establishment and solution

3.1 Establishment and solution of Model 1

According to the conditions given in the topic, the proportional components accumulated and the data accumulated between 85% and 105% are included. Then fill the missing values in the valid data (assign 0).

This paper first analyzes the difference of surface weathering degree of cultural relics and glass type, decoration and color through chi-square test.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \tag{1}$$

χ^2 is used to measure the degree of difference between actual and theoretical values, divided by $n p_i$, to avoid the deviation between different observations and different expectations that varies too much by $n p_i$, so divided by $n p_i$ to eliminate this disadvantage.

The absolute magnitude of the deviation between the actual value and the theoretical value (the difference is magnified due to the presence of squares).

Since the larger the chi-square indicates, the greater the difference, the p-value at this time calculates the probability that the difference between the observed distribution and the expected distribution is greater than the chi-square value:

$$\text{Pvalue=p}(x \leq \chi^2)\text{=}\int_{\chi^2}^{\infty} f(x)dx \tag{2}$$

According to chi-square test, the significant p-value of glass type is less than 0.05 for surface weathering, so there is significant differences between surface weathering and type. Similarly, the p-value of color and decoration is greater than 0.05. There is no significant difference between surface weathering and decoration and color.

According to the title requirements, the type of glass is divided into high potassium and lead and barium. The combined table is screened in excel according to the type of glass, and then whether weathering as the condition for secondary screening, from which four categories can be obtained:

High potassium-weathered, high potassium-unweathered, lead barium-weathered, lead barium-unweathered.

The chemical composition of different types of glass relics is e_{ij}, i represents the i sampling point; j represents the j chemical composition, then:

$$E_j = \frac{1}{n} \sum_{i=1}^{i=n} e_{ij} \tag{3}$$

Similarly, the weathering chemical composition of different types of glass relics is f_{ij}, i represents the i sampling point; j represents the j chemical composition, then:

$$F_j = \frac{1}{n} \sum_{i=1}^{i=n} f_{ij} \tag{4}$$

Comparison of the two chemical components is available

$$G_j = \frac{E_j - F_j}{F_j} \times 100\% \tag{5}$$

Since the data has the presence of 0 without using the above formula, we used the difference between the chemical composition before and after weathering as the standard for change.

According to the above step, there is a relationship between the chemical composition content of the unweathered glass and the weathered glass. The simple prediction formula is as follows:

$$E_j = F_j\left(1 + G_j\right) \tag{6}$$

3.2 Establishment and solution of Model 2

Classify by glass type and analyze its distribution pattern. The average value of the chemical composition was calculated for both types of glass. According to the average chemical composition of the two types of glass, the chemical composition of the two types is quite different, which indicates that this index has certain classification significance. When the component difference is small, the reference significance of the classification results is small and cannot be used as the basis of classification. It is speculated that the difference of chemical components with large difference can be used as the basis for the classification of the two types of glass.

If the mean content of the fifth chemical composition of high potassium glass is m_i, and the fifth chemical composition content of lead-barium glass is n_i, the difference between the fifth chemical composition content of the two glass is o_i.

$$o_i = m_i - n_i \tag{7}$$

The chemical content of high-potassium glass and lead-barium glass varies greatly and slightly. There are 7 chemical components of the difference value is less than 1%, so 1% as the screening criterion, less than 1% is not significance. The chemical components with the difference value of o_i> 1%: silica, potassium oxide, calcium oxide, alumina, lead oxide, barium oxide and phosphorus pentoxide were selected as the reference indicators for the classification rules of high potassium glass and lead barium glass.

First for data preprocessing, because k in form 2 mean analysis need to set the cluster value in advance, we need to compare with real data to verify the accuracy of the classification results, , the high potassium glass cluster value in 2,3,4, cluster analysis, and the degree of weathering classification, grain classification, color classification data. Similarly, by analyzing the data of lead-barium glass in the form one, because the number of clusters should not be too much, so the cluster number of lead-barium glass was set to 2 to find the suitable subclassification respectively.

$$J = \sum_{j=1}^{k}\sum_{i=1}^{N_j}\left\|X_i - Z_j\right\|^2,\ X_i \in S_j \tag{8}$$

$$Z_j = \frac{1}{N_j}\sum_{i=1}^{N_j} X_i,\ X_i \in S_j \tag{9}$$

For high potassium glass: we used spss to set the number of clusters to 2,3,4, and conducted cluster analysis respectively. Comparing the data of the three with the data selected in form I, we found that the accuracy was 0.83 at n=2,0.55 for n=3 and n=4, and confirmed that the number of clusters of high potassium glass is 2.

According to the classification results of the above small questions, we can compare the data and get the accuracy of the classification results by calculating the same amount of data in the existing data. However, the accuracy of high-potassium class was 0.83, and that of lead-barium class was 0.71, which indicates that the clustering results have high accuracy and are more reasonable.

The classification result is the number of clusters of 2 and the number of chemical components of 7. It is speculated that the change of the number of chemical components will affect the accuracy of the classification results, so we reduce the number of chemical components, count the accuracy of the classification results in turn, and can judge the sensitivity of the classification results according to the size of the accuracy change.

We chose the number of clusters of 2, and the accuracy of decreasing chemical components did not changed, which indicates that the number of chemical components has little effect on cluster accuracy, and further shows that our classification results are less sensitive.

3.3 Establishment and solution of Model 3

According to question 2, the classification law of high-potassium glass and lead-barium glass is based on this theory by comparing the average differences between the chemical composition of the glass relics through the chemical composition and corresponding contents of the cultural relics of unknown glass types. We developed the neural network model based on the classification rules of SiO_2, K_2O, CaO, Al_2O_3, PbO, BaO, and P_2O_5 as the independent variable, type as the dependent variable, and the sample size of data 0.7 as the training set. We found that the accuracy of the training set was greater than 0.9, indicating that the model classifies it well in the training set, and that the model is highly practical.

The prediction results of high-level type glass are all high-potassium, and the prediction results of lead-barium glass are all lead-barium. Therefore, the model effect is good again, and the accuracy and reliability of the prediction results are high. To build the model, we show the process with a three-layer neural network model process.

Take the three-layer BP neural network for an example, as shown in Figure 1.

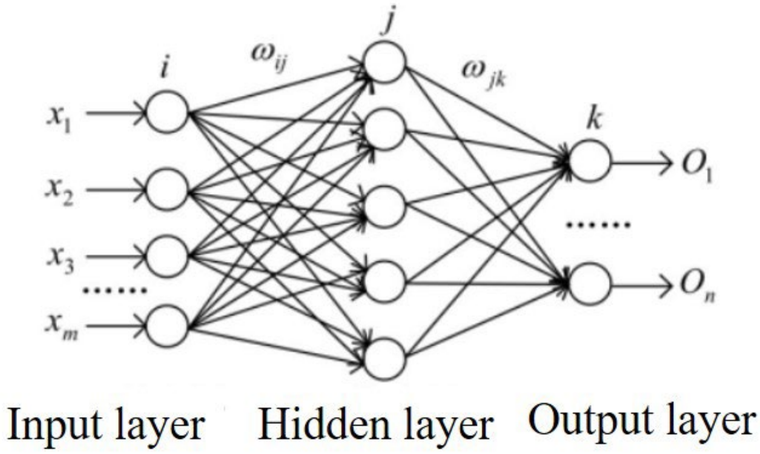

Figure 1: Schematic diagram of the three-layer neural network model

The desired output of the system is set to Tk, and the error E of the system can be expressed by the variance of the actual output value and the desired target value, specifically expressed as follows:

$$E = \frac{1}{2}\sum_{k=1}^{n}(T_k - O_k)^2 \tag{10}$$

$$e_k = T_k - O_k \tag{11}$$

Using the gradient descent principle, the system weights and bias are updated as follows:

$$\begin{cases} w_{ij} = w_{ij} + \beta F_j\left(1 - F_j\right)x_i \sum_{k=1}^{n} w_{ij}e_k \\ w_{jk} = w_{jk} + \beta F_j e_k \end{cases} \tag{12}$$

$$\begin{cases} a_j = a_j + \beta F_j\left(1 - F_j\right)x_i \sum_{k=1}^{n} w_{jk}e_k \\ b_k = b_k + \beta e_k \end{cases} \tag{13}$$

Based on question 2, we still use the neural network classification model for the sensitivity analysis. According to the definition of sensitivity and combined with the idea of machine learning, different chemical components are selected as the training set independent variables, and the training set accuracy is used as the evaluation index. We randomly select the number of types and chemical components to learn the model, and calculate the accuracy of the training results. The influence of different number of chemical composition types on the prediction accuracy of the training set.

According to the randomly selected chemical composition and the number of species, we compare the accuracy rate, and find that the selection of the chemical composition and the number of species has a great impact on the accuracy rate, and has a strong sensitivity (Figure 2).

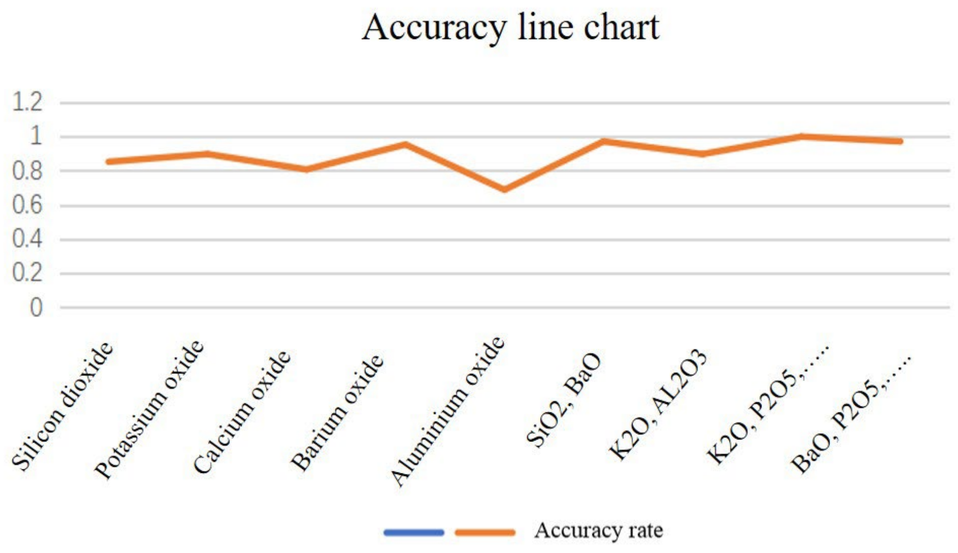

Figure 2: Accuracy line diagram

3.4 Establishment and solution of Model 4

Firstly, glass relics samples are divided into four categories according to glass type and weathering:

High potassium-weathering, high potassium-unweathered, lead barium-weathering, lead barium-unweathered, named as category one, two, three, four, respectively.

Taking high potassium and weathering as input variables, considering that the use conditions of Spearman correlation coefficient is wider than Pearson correlation coefficient, as long as the data meets the monotonous relationship can be used, so this formula is selected for data analysis. The correlation between glass type and weathering degree yielded four correlation coefficient heat maps. The form of the heat map shows the value of the correlation coefficient, which can show the correlation between the two through the depth of color.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{14}$$

High potassium, no weathering: as can be seen from the thermal coefficient diagram, the correlation between the chemical components under the conditions of high potassium and no weathering. If silica and silica 1 positive correlation, each chemical component is positively associated with k=1; calcium oxide, nano oxide and potassium oxide, forming a matrix, k value are positive, but not in the range of 0.661 to 0.827; such as copper oxide and calcium oxide k positive correlation; for example, silicon dioxide and an oxide small k value. The correlation coefficient thermal map shows the correlation between pairwise chemical components through the color depth, from depth to light.

High potassium, weathered: as can be seen from the thermal coefficient diagram, the correlation between the chemical components under the conditions of high potassium and each wind changes. For example, eight chemical components of silicon dioxide and phosphorus pentoxide are positively proportional to k=1; such as calcium oxide and magnesium oxide and alumina, k value is integer, in the range of 0.676 to 0.943; alumina, alumina, barium oxide, strontium oxide, tin oxide, sulfur dioxide and any other chemical components, k value is 0; such as negative correlation of k= -1.

Lead barium, unweathered: available from the thermal coefficient map, under the condition of lead barium and unweathered, each chemical composition is related to its own positive proportion of k value of 1. Among the many charts where chemicals are negatively correlated to each other, tin oxide and potassium oxide are the only two kinds of chemicals that have a positive correlation with their constituent k value greater than 0.5. The proportion of silica and the other thirteen categories of chemical components is mostly negatively correlated, mixed with four positive correlations, but the correlation coefficient k value does not exceed 0.2. Most of the correlation values of the rest and other chemicals are between ± 0.2, resulting in most of the lighter colors of the thermal coefficient map.

Lead barium, weathered: available by the thermal coefficient map, under the condition of lead barium, weathered, each chemical composition is related to its own positive proportion of k value of 1. Copper oxide, lead oxide and barium oxide are negatively correlated with each other or have a positive correlation with k values less than 0.1. Magnesium oxide, alumina and iron oxide are positively correlated with the k values of around 0.5. The distribution of similar colors of this thermodynamic coefficient map is a block, and several chemical classes synchronize into positive or negative correlations with similar k values.

Taking the matrix diagram with the positive correlation between lead barium without weathering and lead barium weathering as an example, the matrix blocks with a correlation coefficient greater than 0.5 do not coincide. Looking at the positive correlation of high price elegance and the positive correlation of high potassium and no weathering as an example, both matrix blocks with a correlation coefficient greater than 0.5 are the relative terms of k=1, but the difference between the two is not small, and the distribution of symmetric matrix blocks is completely different. Moreover, the negative correlation between lead barium weathering and the negative correlation between lead barium without weathering, and both matrix blocks are chaotic types, and the overlap part is less.

4. Conclusion

In this study, the chemical composition contents of high-potassium glass and lead-barium glass in weathered and unweathered conditions were analyzed respectively, and a chemical composition prediction model was established. Using chi-square test and other knowledge, it was analyzed that the type of cultural relics had a great correlation with weathering, which had certain reference value for the protection of cultural relics. Problem one uses chi-square test, chi-square test is helpful to compare the difference analysis between categorical variables and categorical variables. Therefore, by counting the degree of deviation between the observed value and the theoretical value of the type, color and decoration, the chi-square value can be obtained. The larger the chi-square value is, the greater the deviation between the two is. We can get the significant P value according to the chi-square test results. The relationship between surface differentiation and the three can be analyzed more intuitively by the size of the P value. In this paper, BP neural network classification is used in problem 3. BP network is a multi-layer feedforward network trained by error back propagation algorithm. BP neural network classification can explore the best weights and thresholds of types and chemical components, so as to minimize the error rate of network classification. Using the model to predict, make it get more accurate classification results.

References

[1] Chen Guifen, Zeng Guangwei, Chen Hang, Li Chunan. Study of RS image classification method based on texture features and neural network algorithm[J]. Journal of Chinese Agricultural Mechanization, 2014.

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

[3] Li Fei, Li Qinghui, Gan Fuxi, Zhang Bin, Cheng Huansheng. Proton excited X-ray fluorescence analysis of chemical composition of a batch of ancient Chinese glasses [J]. Journal of Silicate, 2005 (05): 581-586.

[4] Chen G, An K, Li X, et al. Identification and Classification of Adverse Geological Body Based on Convolution Neural Networks [J]. Geological Science and Technology Information, 2016.

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