ISSN 2616-5880 Vol. 4, Issue 1: 48-53, DOI: 10.25236/AJMC.2023.040109

Prediction of the chemical composition content of ancient glass artifacts before weathering

Haiyu Wang, Jiashu Zhai, Wujun Tao*

College of Mathematical Science and Engineering, Hebei University of Engineering, Handan, Hebei, 056038, China
**Corresponding author: m17839919209@163.com*

Abstract: *In this paper, the surface weathering, type, decoration, color and content of each chemical component of glass artifacts were studied and analyzed, and a K-means clustering model was established, using Spearman correlation analysis, chi-square test, and. It was solved to classify the glass types and analyze the change pattern of chemical composition of glass artifacts, and a better fitting effect was obtained. This paper characterized the problem as a prediction class, firstly, assigned values to four categorical variables: surface weathering, type, decoration and color of glass artifacts, and then used spss to perform Spearman correlation and chi-square test analysis to pre-process the data and eliminate invalid data. Then, it used descriptive statistical analysis to find that most of the chemical components of high potassium glass showed a decreasing trend after weathering, and most of the chemical components of lead-barium glass showed an increasing trend after weathering; finally, used Matlab matrix to derive a linear mapping relationship based on the changes of chemical components before and after weathering, and finally predicted the chemical components of glass artifacts before weathering.*

Keywords: *Antique glassware; K-means clustering model; Chi-square test; Spearman correlation*

1. Introduction

Glass as an important evidence of early trade in the Silk Road, in the cultural exchanges between China and the West played an important role. Early glass often came to China in the form of bead-shaped ornaments in West Asia and Egypt, and ancient glass in China was often made locally from local materials after absorbing their good techniques, so although the appearance of glass products was similar to that of the inflow, the composition of glass products was different [1].

Quartz sand, as the main raw material for glass, is composed mainly of Si02, and co-solvents are often added to lower the melting temperature to solve the problem of high melting point of pure quartz sand in the refining process. For example, potassium glass has a high potassium content due to the addition of grass ash as a flux during the refining process. Changes in the burial environment often lead to varying degrees of weathering of ancient glass products. In general, when weathering occurs, the environmental elements exchange with the internal elements to change their original composition, which affects the determination of their glass types. In this paper, glass samples with a cumulative composition of 85% to 105% are considered as valid data and the relationship between surface weathering and the other three elements is investigated; then the changes in the chemical composition of the surface of the samples are calculated according to the type of glass; finally, the pre-weathering chemical composition of the samples is analyzed based on the weathering data [2].

2. Assumptions and notations

2.1 Assumptions

Use the following assumptions.

1) Only the main conditions involved are considered, and no other cases are considered.

2) The basic information of the known glass artifact samples is made fidelity assumptions.

3) It is assumed that the data are processed in a way that reflects the actual weathering of the glass samples.

4) It is assumed that the data are representative and accurately reflect the relationship between surface weathering and weathering.

2.2 Notations

The primary notations used in this paper are listed as Table 1.

Table 1: Notations

Symbols	Description
t_i	Chemical composition of heritage samples
t_{ij}	Chemical composition of heritage samples before weathering
w_{ij}	Chemical composition of artifact samples after weathering
n_i	Chemical composition values
V_i	Variability of chemical composition
i	Number of clusters
j	Chemical composition of cluster center
m	Number of samples to be examined
n	Chemical composition of samples to be tested

3. Model construction and solving

3.1 Model Building

3.1.1 Model Preparation

Based on the data and requirements, Spielman's correlation analysis[3], chi-square test, descriptive statistics[4], and control variables were used in this question. In this paper, the type, decoration, and color of glass are categorical variables, so Spearman's correlation analysis[5] and chi-square test were used to analyze the relationship; when studying the type of glass, the variation pattern of other chemical components should be studied by using control variables and descriptive statistics. For the prediction of the chemical composition content before weathering, Matlab can be used to derive the corresponding expressions for the chemical composition before weathering based on the variables of the chemical composition, and then predict the chemical composition content of the samples before weathering.

3.1.2 Overall model algorithm flow

Model algorithm flow chart is shown in Figure 1.

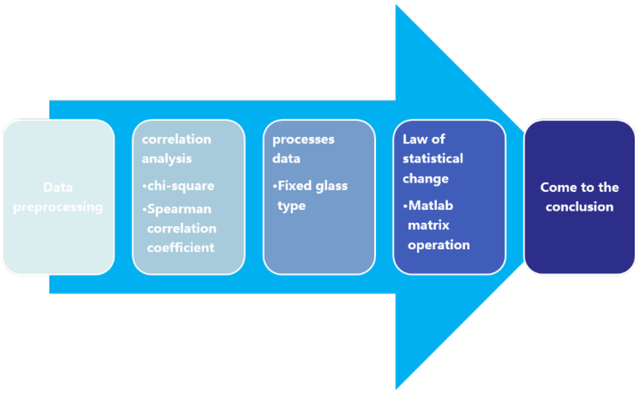

Figure 1: Model algorithm flow chart

3.1.3 Spearman Correlation Analysis

Spearman correlation analysis in short is solved by the order of the original data. Spearman's correlation coefficient is defined as follows.

Definition: A and B are two sets of data, then their spearman correlation coefficients are

ISSN 2616-5880 Vol. 4, Issue 1: 48-53, DOI: 10.25236/AJMC.2023.040109

$$r_s = 1 - \frac{6\sum_{i=0}^{n} d_i^2}{n\left(n^2 - 1\right)} \qquad (1)$$

3.1.4 Chi-square test

The chi-square test occupies an important place among the many hypothesis testing methods. It is a type of nonparametric test that is used to compare two and more sample rates as well as two categorical variables for correlation analysis. The key is to compare the goodness of fit or the degree of problem fit between the theoretical and actual frequencies. The principle is that the deviation between the theoretical inferred value and the actual observed value of the statistical sample, the deviation between the theoretical inferred value and the actual observed value determines the size of the chi-square value, if the chi-square value is smaller, the deviation is smaller; on the contrary, the deviation is larger; if the two are exactly equal, the chi-square value is 0, which means that it is consistent with the theoretical value.

The specific steps are shown below:

Step1:

Put forward the hypothesis: W_0: the distribution function of the overall T is $F(t)$, if the overall distribution is discrete, then the hypothesis is W0: the distribution law of the overall T is $P(T=t_i)=p_i$ (i=1, 2, ...)

Step2:

Divide the range of values of the overall T into k disjoint small intervals A_1, A_2, A_3, A_K (i=1, 2, 3, ...K), such as taking $A_1=(a_0,a_1)$, $A_2=(a_1,a_2)$, ...$A_k=(a_k -1,a_k)$. The division of intervals depends on the case, but make the number of sample values in each small interval not less than 5, and the number of intervals k should not be too large or too small.

Step3:

The number of sample values of Ai that fall into the i-th interval is recorded as f_i, which becomes the group frequency, and the sum of all group frequencies f_1+f_2+......+f_k is equal to the sample size n.

Step4:

When W_0 is true, according to the assumed overall theoretical distribution, the probability pi that the value of the overall T falls into the i-th interval can be calculated, so its product is the theoretical frequency of the sample value falling into the i-th interval.

Step5:

When W_0 is true, the frequency fin of sample values falling into the i-th smallest interval in n trials should be close to the probability, and when W_0 is not true, then fin is very different from W_0. Based on this idea, Pearson introduces the test statistic as shown below.

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

3.2 Analysis and solution of the model

In the process of solving this paper,it found that different models have their own advantages and disadvantages, therefore, the specific solution of this paper adopts spas correlation analysis and combines with excel to process the data, so as to use Matlab to run the program to get the prediction results.

3.2.1 Data pre-processing

First, according to the analysis of the basic data of this paper, the invalid data were eliminated, and the data blank indicates that the corresponding component is not detected, so that the chemical content of this component is 0, which can eliminate the detection data that do not satisfy the above equation.

3.2.2 Spss analysis of the relationship between sample parameters

For the study of the relationship between surface weathering and its sample parameters, it was first

ISSN 2616-5880 Vol. 4, Issue 1: 48-53, DOI: 10.25236/AJMC.2023.040109

assigned a quantified value, and then a Spearman correlation analysis and a chi-square test were performed to derive the relationship between surface weathering and its decoration, type and color. The relationship between surface weathering of glass artifacts and their ornamentation, type and color was determined by the Spearman correlation analysis and chi-square test.

3.2.3 Spss analysis of the surface of cultural relics samples with and without weathering chemical content of the statistical law

For the study of the type of glass and the change pattern of chemical content with or without weathering on the surface of the artifacts, we first used excel data processing to construct the content of each chemical component before and after glass weathering after fixing the glass type variable, and then import spss26.0 for descriptive statistical analysis to obtain basic statistics such as correlation trend, median, variance, etc., and then Then, the statistical patterns of the content of chemical components on the surface of artifact samples with and without weathering were obtained.

3.2.4 Matlab analysis to predict the chemical content of artifact samples before weathering

To predict the constituent content based on weathering point data, a matrix operation using Matlab was performed to derive the pre-weathering chemical content of each component content before weathering.

3.3 Results of the model

By passing the data into spss26.0 and transforming and recoding them into numerical variables and then using spearman correlation analysis, the results were obtained as shown in Figure 2 below.

Correlations

			ornamentation	type	color	surface. weathering
Spearman's rho	ornamentation	Correlation Coefficient	1.000	.119	-.035	-.037
		Sig. (2-tailed)	.	.374	.807	.781
		N	58	58	52	58
	type	Correlation Coefficient	.119	1.000	.392**	-.344**
		Sig. (2-tailed)	.374	.	.004	.008
		N	58	58	52	58
	color	Correlation Coefficient	-.035	.392**	1.000	-.125
		Sig. (2-tailed)	.807	.004	.	.379
		N	52	52	52	52
	surface.weathering	Correlation Coefficient	-.037	-.344**	-.125	1.000
		Sig. (2-tailed)	.781	.008	.379	.
		N	58	58	52	58

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 2: Correlation Analysis

According to Figure 3, it can be seen that the Spearman value of surface weathering and glass type of glass artifacts is 0.009 less than 0.05, which has a significant difference, so it can indicate that the surface weathering and glass type of glass artifacts are significantly related; however, the Spearman values of surface weathering and glass decoration and color of glass artifacts are 0.781 and 0.668, which are greater than 0.05 and do not have However, the Spearman values of surface weathering and glass decoration and color were 0.781 and 0.668, respectively, which were greater than 0.05 and not significantly different from each other. Next, the chi-square test was conducted on the surface weathering and glass type, ornamentation and color of glass artifacts, and the results were obtained as shown in Figure 3: according to Figure 2 above, it can be seen that the Pearson chi-square value of surface weathering and proportional type of glass artifacts is 0.009 less than 0.05, which is significantly different, while the Pearson chi-square values of surface weathering and glass ornamentation and color of glass artifacts are 0.084 and 0.507, both of which are greater than 0.05 and not significantly different, are consistent with the above findings and prove the consistency of the conclusions.

ISSN 2616-5880 Vol. 4, Issue 1: 48-53, DOI: 10.25236/AJMC.2023.040109

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6.880[a]	1	.009		
Continuity Correction[b]	5.452	1	.020		
Likelihood Ratio	6.889	1	.009		
Fisher's Exact Test				.011	.010
Linear-by-Linear Association	6.762	1	.009		
N of Valid Cases	58				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.45.

b. Computed only for a 2x2 table

Figure 3: Type and surface weathering chi-square test

3.4 K-means clustering algorithm for solving subclasses

The specific implementation steps are shown below:

Set X=x_1,x_2,......x_n as the input data set, use I and K to describe the maximum number of cycles and the number of clusters, respectively; use C_j,j=1,2......,K to describe the class clusters, and K class clusters to represent the output.

Step1:

To obtain the original cluster centers mj(i) of K class clusters, arbitrarily select data points and the number of these data points is K. Set I=1.

Step2:

Find d(x_i,m_j(i)),i=1,2......,n,j=1,2......,K,which denotes the distance between each data point xi and K cluster centers; in order to make $x_i \in C_j$,it is necessary to meet d(x_i,m_j(i)) as the minimum value.

Step3:

Reassign the K new cluster centers to obtain the formula:

$$m_j(I)=\frac{1}{N_j}\sum_{i=1}^{n}x_i \tag{3}$$

Step4:

Describe the objective function with the value of J(C) and solve the value of the formula as

$$J(C)=\sum_{j=1}^{K}\sum\nolimits_{x_i\in C_j}\left\|x_i-m_j(I)\right\|^2 \tag{4}$$

When I=I+1, return to step (2),re-solve the d(xi,mj(i)) value until J(C) obtains the minimum value, then return mj(i), k=1,2......, K, and the K-mean clustering algorithm terminates the calculation.

4. Conclusion

In this paper, the four categorical variables of surface weathering, type, decoration and color of glass artifacts were assigned and quantified, followed by Spielman correlation and chi-square test analysis using spss to eliminate invalid data. Finally, based on the changes of chemical composition before and after weathering, we used Matlab matrix operations to derive linear mapping relationships, and finally predicted the chemical composition of glass artifacts before weathering, and proposed a subclass classification method based on K-means.

ISSN 2616-5880 Vol. 4, Issue 1: 48-53, DOI: 10.25236/AJMC.2023.040109

References

[1] Bo YH, Yan ZQ, Wang ZP, Zheng CH. Study on the preparation of C4 olefins based on multiple regression model [J]. Science and Technology Innovation, 2022(11): 49-52.
[2] Li Chunsheng, Yu Hu. Construction of vehicle driving conditions based on improved K-means clustering algorithm [J]. Computer Technology and Development, 2022, 32(03): 169-174.
[3] Chen Guiru, Wang Bing, Cao Zhijie, Wang Shaoping. Research on wind power prediction based on clustering analysis and optimization neural network [J]. Electrical Automation, 2020, 42(03): 24-27.
[4] Liu Z , Liu Y , Xing H , et al. Determination of the Content of Pectin and Hemicellulose in Ramie Based on Near-infrared Technique[J]. Plant Fiber Sciences in China, 2018.
[5] Liang T, Zhou N, Lu TQ, Wu H, Ju P. Parameter filling method and typical parameter analysis of induction motor load transient model [J]. Power System Automation, 2020, 44(01): 74-82.