Study on Composition Analysis and Identification of Ancient Glass Products

Xin Sun
Jiangsu University of Science and Technology, Zhangjiagang, China
761689819@qq.com

Abstract. In this paper, an entropy-TOPSIS model is established to analyze and identify the composition of ancient glass products, which solves the relationship between the type, decoration, colour of glass and the surface weathering of glass cultural relics; The multivariate linear regression model is established, which solves the problem of statistical law of weathering chemical composition content on the surface of cultural relics samples. A random forest classification model based on an ant colony algorithm is established to solve the problem of determining the importance of chemical composition, and a K-means clustering model is established to solve the problem of sub-class division within the group; A grey correlation analysis model was established to solve the problem of the correlation between the chemical composition of the same type of glass.

Keywords: Entropy weight-TOPSIS method, Random forest, K-means, Grey Relational Analysis, Spearman's rank correlation coefficient.

1. Introduction

In this paper, an entropy-TOPSIS model is established to analyze and identify the composition of ancient glass products, which solves the relationship between the type, decoration, colour of glass and the surface weathering of glass cultural relics; The multivariate linear regression model is established, which solves the problem of statistical law of weathering chemical composition content on the surface of cultural relics samples. A random forest classification model based on an ant colony algorithm is established to solve the problem of determining the importance of chemical composition, and a K-means clustering model is established to solve the problem of sub-class division within the group; A grey correlation analysis model was established to solve the problem of the correlation between the chemical composition of the same type of glass.

The Silk Road is a channel for transportation, trade and cultural exchanges at all times and all over the world. Besides silk, the main commodities of its trade include gold and silver utensils, bronze and iron wares, glass and so on. As material evidence of past foreign trade, the analysis and identification of the composition of ancient glass have received great attention.

The composition analysis mainly includes three aspects: raw materials, cosolvents and stabilizers. In terms of raw materials, glass is mainly made of quartz sand, so the main component is silicon dioxide. However, because the melting point of pure quartz sand is too high, the melting point temperature is often reduced by adding cosolvent, and limestone is added as a stabilizer, so the composition is also affected by the addition of cosolvent and stabilizer. Generally, the main component formed after the refining of the stabilizer is calcium oxide, while the main chemical components formed by the addition of cosolvent are different due to the different types of addition.

In terms of identification, due to the weathering of ancient glass burial environment and time, the change of composition ratio inevitably affects the judgment of glass classification.

The absolute weathering rate of different glass types, different ornamentation types and different colours. Secondly, the entropy-TOPSIS model is established to analyze the relationship between the surface weathering of glass relics and their glass types, ornamentation and colours.

2. Model building

2.1. TOPSIS model based on the entropy weights method

According to the requirements of the first question, two methods are considered, the weight method, and the other is the fitting method. Considering the analysis of the topic, the weight method is selected. The first commonly used method is the analytic hierarchy process, but because the analytic hierarchy process is too subjective, it does not have the objectivity of data analysis. Therefore, we try to find a method that is both objective and in line with the requirements of the topic, that is, the TOPSIS model based on the entropy weight method [1].

2.1.1 Entropy weight method

Construction of evaluation matrix, as shown in formula (1).

$$\mathrm{X_{ij}} = \begin{bmatrix} \mathrm{X_{11}} & \cdots & \mathrm{X_{13}} \\ \vdots & \ddots & \vdots \\ \mathrm{X_{m1}} & \cdots & \mathrm{X_{m3}} \end{bmatrix},\ i = 1,2,\cdots,m,\ j = 1,2,3 \tag{1}$$

Among them, i is the number of cultural relics, j is the impact factor, that is, the evaluation index, respectively, glass type, decoration and color.

Calculate the information entropy of each index, as shown in formula (2), (3), (4).

$$e_j = \frac{1}{\ln m}\sum_{i=1}^{m} P_{ij}\cdot \ln P_{ij} \tag{2}$$

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{m} Y_{ij}} \tag{3}$$

Obtaining respective weight values according to the information entropy of each index ω_j.

$$\omega_j = \frac{1-\mathrm{e}_j}{n-\sum_{j=1}^{n} e_j} \tag{4}$$

2.1.2 TOPSIS method

The Euclidean distance formula is used to calculate the distance between each evaluation index value and the positive ideal solution d_{j}^{+} and the distance between each evaluation index value and the negative ideal solution d_{j}^{-}, as shown in formula (5), (6), (7).

$$d_{\mathrm{j}}^{+} = \sqrt{\sum_{j=1}^{n}(Z_{ij}-Z_j^{+})^2} \tag{5}$$

$$d_{\mathrm{j}}^{-} = \sqrt{\sum_{j=1}^{n}(Z_{ij}-Z_j^{-})^2} \tag{6}$$

Calculating the relative closeness between each evaluation index value and the ideal solution according to the distance between the positive value and the negative value T_i.

$$T_i = \frac{d_j^{-}}{d_{\mathrm{j}}^{+}+d_{\mathrm{j}}^{-}} \tag{7}$$

2.2. A random forest classification model based on ant colony algorithm

First of all, considering the basis of classification, it is necessary to select appropriate factors with certain importance that affect the classification of glass [2]. Here, the main factors are selected by the importance analysis of chemical composition characteristics obtained by random forest classification based on the ant colony algorithm. Although random forest classification itself can be used to select the main factors, However, considering that the ant colony algorithm (ACO) can accelerate the efficiency and accuracy of random forest decision tree classification, the ant colony algorithm is added to optimize random forest [3].

Firstly, when establishing the random forest classification model [4], it is assumed that the minimum sample number of internal node splitting is 2, the maximum sample number of the leaf node is 1, the minimum weight of sample in the leaf node is 0, the maximum depth of a tree is 10, and the maximum number of leaf nodes is 50.

The performance of the forest is improved by selecting the decision tree individuals of the integrated forest. On the one hand, the strength of the decision tree in the integrated forest is enhanced, and the original forest is recorded as [5]:

$$RF^{ALL} = \left\{T_1, T_2, \cdots, T_L\right\} \tag{8}$$

On the other hand, due to the decrease of forest scale [6] and the increase of output consistency of the selected trees with high prediction performance [7], the average similarity of sub-forests also increases, so it is necessary to reduce the average similarity between decision trees in the integrated forest.

The heuristic factor of the genetic algorithm [8] is constructed by the diversity matrix between the decision trees in the forest, namely, as shown in formula (9).

$$d\left(T_i, T_j\right) = \frac{1}{d_i v_{i,j}} \tag{9}$$

Using regression prediction, is the unobservable error vector, is the vector of regression coefficients, in order to minimize the sum of the squares of the errors, the least squares estimate of is:

$$\min_{\delta} \lambda^T \cdot \lambda = \min_{\delta}(Z - c\delta)^T (Z - c\delta)^{def} = Q(\hat{\delta}) \tag{10}$$

2.3. A random forest classification model based on genetic ant colony algorithm

first of all, we need to analyze the chemical composition to identify the type, which is the same as the problem to be solved in problem two, which is to classify the glass, so we use the idea of problem two to solve problem three. The second problem is to establish a random forest classification model based on the ant colony algorithm to solve the feature [9] importance and select the chemical elements whose importance value is greater than 10% as the independent variables affecting the classification. The third problem is to re-optimize the model through a genetic algorithm [10].

Choose a positive integer as the size of the population, and use the numerical solution to generate an approximate optimal point n :

$$p(0) = \left\{x_1^0, x_2^0, \cdots, x_n^0\right\} (i = 1, 2, \cdots n) \tag{11}$$

These points form the initial population, complete the initialization, which is taken here, and the upper limit of the number of iterations is set to 1000, given the fitness function:

$$Eval(x) \begin{cases} \dfrac{t_i * x_i^0}{\sum_{i=1}^{14} t_i} * 100\%, \dfrac{t_i}{\sum_{i=1}^{14} t_i} \geq 10\% \\ 0, \dfrac{t_i}{\sum_{i=1}^{14} t_i} \leq 10\% \end{cases} \quad i = 1, 2, \cdots, 14 \tag{12}$$

Assuming that whether the surface of lead oxide (PbO), potassium oxide (K_2O), barium oxide (BaO) and silicon dioxide (SiO_2) is weathered or not as the independent variables, and, the multiple linear regression model can be expressed as: $c_1 c_2 c_3 c_4 c_5$.

$$\begin{cases} z = \delta_0 + \delta_1 c_1 + \delta_2 c_2 + \delta_3 c_3 + \delta_4 c_4 + \delta_5 c_5 + \delta_6 c_6 + \lambda \\ \lambda \sim N(0, \ \sigma^2) \end{cases} \tag{13}$$

Where, $\delta_0\delta_1\delta_2\delta_3\delta_4\delta_5$ and $C_1C_2C_3C_4C_5$ are all unknown parameters unrelated to, $\delta_0\delta_1\delta_2\delta_3\delta_4\delta_5$ are called regression coefficients.

The specific data information of the four main factors can be obtained from Table 2. Suppose $c_{i1}c_{i2}c_{i3}c_{i4}c_{i5}c_1$ and $c_{i5}c_1c_2c_3c_4c_5$ are the observed values of, and, so the multiple linear regression model is:

$$\begin{cases} z_i = \delta_0 + \delta_1 c_{i1} + \delta_2 c_{i2} + \delta_3 c_{i3} + \delta_4 c_{i4} + \delta_5 c_{i5} + \lambda_i \\ \lambda_i \sim N(0, \ \sigma^2) \ , i = 1, 2, \cdots, n \end{cases} \tag{14}$$

3. Model solving

3.1. TOPSIS model based on the entropy weights method

3.1.1 Entropy weight method

According to the basic information of the glass relics in the attached table, the glass type, ornamentation and color are weighted by using the entropy weight method, and the results are shown in Table 1.

Table 1. Weighting of Glass Type, Texture and Color

	Glass type	Ornamentation	Color
Weight	50.3%	31.5%	18.2%

As can be seen from Table 1 of the weight result obtained from the entropy weight method analysis, the colour has the lar weight on the surface weathering, and the glass type has the smallest weight on the surface weathering.

3.1.2 Solution of TOPSIS method

Table 2. Ranking results of TOPSIS evaluation calculation

Item	Distance of positive ideal solution	Negative ideal solution distance	Relative proximity	Sort the results
1	0.503	3.089	0.86	1
2	3.089	0.503	0.14	54
3	0.807	3.023	0.789	16
4	0.807	3.023	0.789	15
5	0.807	3.023	0.789	14
⋮	⋮	⋮	⋮	⋮
54	0.535	3.083	0.852	2

The number of samples entering the analysis is 58. During the analysis, 4 missing sample data are filtered out first, so the remaining 54 complete samples enter the algorithm analysis. The results are shown in Table 2.

Due to a large amount of data, all the ranking results of the TOPSIS evaluation calculation are placed in the supporting material. Comparison of Weathering Rate Based on Entropy-TOPSIS.

3.2. A random forest classification model based on ant colony algorithm

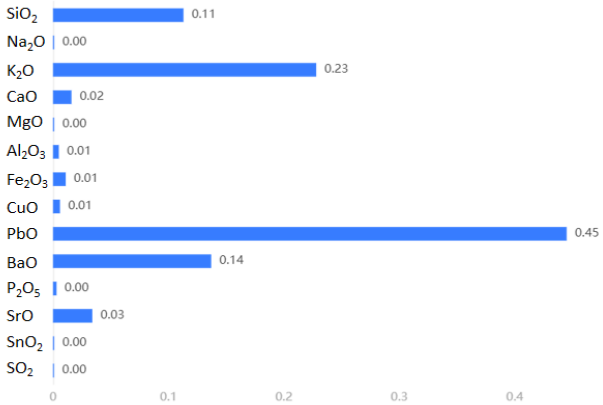

Figure 1. Analysis of the importance of chemical composition characteristics

It can be seen from Figure 1 that the characteristic importance of lead oxide (PbO) is the highest, with importance of 45%, followed by potassium oxide (K2O), barium oxide (BaO) and silicon dioxide (SiO2), with importance of 23%, 14% and 11% respectively. Here, only the factors whose importance accounts for more than 10% are considered.

Therefore, the above four chemical compositions were selected to analyze the classification rules of high potassium glass and lead-barium glass [11],which is shown in table 3.

Table 3. The significance level of the four main influencing factors on glass classification

	Non-standardized coefficients		standardized coefficients	t	p	VIF	R^2	F
	B	standard error	Beta					
C	1.95	0.352	-	5.533	0.000***	-	0.76	F=49.172 P=0.000***
SiO2	-0.006	0.004	-0.308	-1.468	0.147	11.414		
K2O	-0.054	0.012	-0.467	-4.487	0.000***	2.805		
PbO	0.005	0.005	0.211	1.062	0.292	10.181		
BaO	0.005	0.006	0.097	0.822	0.414	3.596		

3.3. A random forest classification model based on genetic ant colony algorithm

A classification model obtained by solving is as follows:

$$z = 2.068 - 0.006c_1 - 0.056c_2 + 0.008c_3 + 0.005c_4 - 0.31c_5 \tag{15}$$

The difference between the predicted value of the classification model obtained by fitting the curve and the true value is shown in Figure 2:

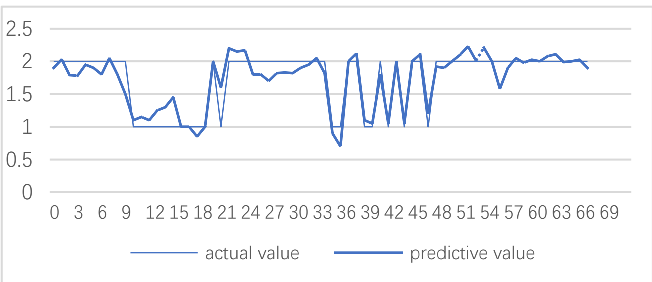

Figure 2. Comparison between the actual value and the predicted value

It can be seen from the figure that the difference between the actual value curve and the predicted value curve is not very large, so it can be concluded that the curve simulation result is better.

The glass of unknown type is analyzed and classified through the fitted relationship, and Table 4 is obtained:

Table 4. Classifies the unknown glass

Cultural relic number	Glass type	Cultural relic number	Glass type
A1	Lead and barium	A5	High potassium
A2	Lead and barium	A6	High potassium
A3	Lead and barium	A7	High potassium
A4	Lead and barium	A8	Lead and barium

4. Analysis of results

The weights of glass type, texture and colour on surface weathering were calculated by the entropy weight method, which was 50.3%, 31.5% and 18.2%, respectively. Then through the TOPSIS method score and rank, the higher the ranking, the easier the weathering. The Spellman correlation coefficient was used for correlation analysis, and the correlation coefficient values were 0.316, 0.128 and 0.112, which were consistent with the above data. Finally, the chi-square test was used for difference analysis, and the significant levels were 0.020, 0.056 and 0.507, respectively. Therefore, the glass type had the most obvious difference in surface weathering, which is shown in Figure 3 and 4.

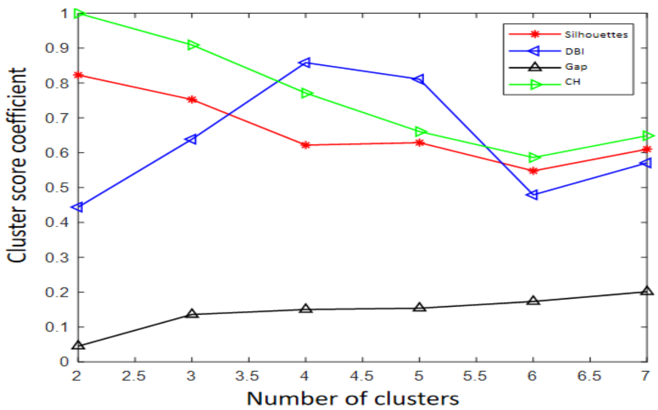

Figure 3. The optimal number of clustered points is determined using four coefficients

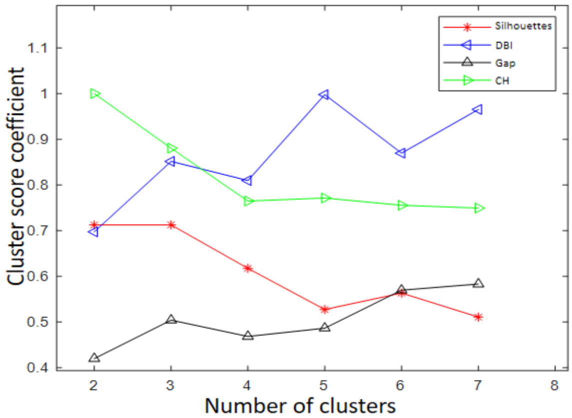

Figure 4. Four coefficients are used to find the optimal number of clustered points

As the contour coefficient is closer to 1 when the cluster score coefficient is smaller, the DB coefficient is smaller when the cluster score coefficient is larger, and the Gap coefficient and the CH coefficient are larger when the number of clusters is larger, it can be obtained from fig. 3 that when the number of clusters is 3, the characteristics of each coefficient at the same time are in line with the characteristics of each coefficient at the same time. Therefore, the optimal number of clustering points obtained from the four coefficients is also 3.

To sum up, the optimal number of clustering points is determined to be 3, that is, K = 3.

Ant colony algorithm is used to find the best split feature number of random forest classification, and the feature importance of lead oxide, potassium oxide, barium oxide and silicon dioxide is 45%, 23%, 14% and 11%. The four chemical compositions were selected as independent variables, the glass classification was used as the dependent variable to do linear regression analysis, and the glass classification rules were obtained. Finally, K-means clustering analysis was used to classify the subclasses, and the number of clustering points of the two types of glass was determined to be 3 by the elbow rule and contour coefficient.

According to the requirement of the third question, it is necessary to analyze the chemical composition to identify the type. Following the idea of the second sub-question of the second question, the main influencing factors are selected first, and the model is optimized by a genetic algorithm. The optimized random forest classification model based on the ant colony algorithm is used to solve the feature importance and select the factors whose importance is greater than 10%. According to the solution results, the main influencing factors are PbO, K_2O, BaO and SiO_2, which are the same as the results of the second sub-question of the second question.

A multiple linear regression model is established, whether the surface is weathered or not is added as a variable, the weathered surface is set as 1, the non-weathered surface is set as 0, and the model is obtained by solving.

$$z = 2.068 - 0.006c_1 - 0.056c_2 + 0.008c_3 + 0.005c_4 - 0.31c_5 \tag{16}$$

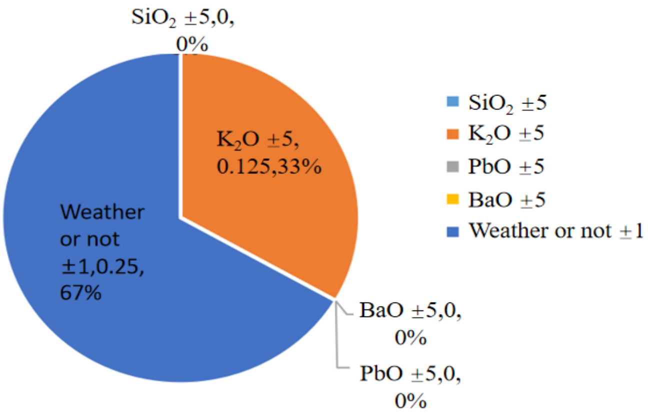

Figure 5. Rate of change of categories

The rate of change of categories is shown in Figure 5. Considering that the data with the cumulative proportion of each component between 85% and 105% are regarded as valid data, but the proportion of chemical components contained in unknown cultural relics is close to 100%, so only the chemical components with higher importance are selected for sensitivity analysis with ± 5 component changes. Because there are only "0" and "1" in weathering, the sensitivity analysis of weathering index is ±1. The results show that whether there is weathering on the surface of cultural relics has a great impact on classification, the content of K_2O has a great impact on classification, and the impact of PbO and BaO on the classification analysis of cultural relics is not particularly obvious.

5. Conclusion

In view of the composition analysis and identification of ancient glass products, this paper establishes the entropy right-TOPSIS model, which solves the relationship between glass type, ornamentation, colour and surface weathering of glass cultural relics. In this paper, the composition of ancient glass is determined, and the statistical law of whether there is weathered chemical composition content on the surface of cultural relics samples is solved by establishing a multiple linear regression model. Subsequently, a random forest classification model based on the ant colony algorithm is established, which solves the problem of determining the importance of chemical composition.

The random forest method of ant colony based on neural network optimization adopted in this paper can effectively solve the shortcomings of the ant colony algorithm such as slow convergence speed and long calculation time, greatly improve the solution efficiency of the model, etc., and achieve the expected effect through a variety of mathematical models, which has certain feasibility.

References

[1] Fu Weizhong, Chu Liuping. Research on the Evaluation of Manufacturing High Quality Development from the Perspective of Yangtze River Delta Integration -- TOPSIS Evaluation Model based on the improved CRITIC- entropy weight method combined weight [J]. Industrial technology and economics, 2020, 39(09): 145-152.

[2] L Syafaah, D Arisnanda, M Irfan, R H Hendaryati. Correlation of risk factors for coronary heart disease in type-2 diabetes mellitus using the chi-square method [J]. IOP Conference Series Materials Science and Engineering, 2020, 821(1).

[3] li xingqi, gao xiaohong. Dimensionless method of effectiveness evaluation [J]. Journal of statistics and decision, 2021 ((15): 24 to 28, DOI: 10.13546/j.carol carroll nki tjyjc. 2021.15.005.

[4] Shen Ying. T test and F Application of inspection in Trend Analysis of chemical test quality control data [J]. Chinese journal of inspection and quarantine, 2019, 29(02): 105-107.

[5] he xiuli. Multivariate Linear model and Ridge regression analysis [D]. Huazhong University of Science and Technology, 2005.

[6] Yan Chun, Li Yaqi, Sun Haitang. Research on Auto Insurance Fraud Identification based on Ant Colony Optimization Random Forest Model [J]. Insurance research, 2017(6): 114-127. The DOI: 10.13497/j.carol carroll nki is. 2017.06.011.

[7] Deng Yurui, xu-dong cheng, Tang Fang, zhou yong. Risk control of Rice mildew in storage based on multiple linear regression analysis and random forest algorithm [J]. Journal of University of Science and Technology of China, 202, 52(01): 47-54+72+55.

[8] Tian Bing, Huang Shan, Sun Ye, Yan Yujie, Jin Kangkang. Rockburst Prediction based on SOM neural network clustering and gray TOPSIS evaluation method [J]. China mining, 2021, 30(01): 188-192.

[9] guo wenjuan. K-means clustering algorithm based on Optimized initial clustering Center [J]. Wind technology, 2022(4): 63-65. The DOI: 10.19392/j.carol carroll nki. 1671-7341.202204021.

[10] Wu Huijie. Study on Application and Optimization Method of random machine forest Algorithm [D]. Jiangnan great learning, 2021. DOI: 10.27169/, dc nki. Gwqgu. 2021.001718.

[11] Zhang Weifeng. Statistical analysis of Spearman's simple correlation coefficient and Gini Gamma correlation coefficient [D]. wide East university of technology, 2020. DOI: 10.27029/, dc nki. Ggdgu. 2020.001333.