

Composition Analysis and Identification of Ancient Glass Products

Quanwen Tan

Department of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing, China
1940448289@qq.com

Abstract. The glass manufacturing technology of ancient China is well known in the world. The study of ancient glass relics is of great significance for today's archaeology and glass aesthetics. In this paper, the logarithmic index of principal component analysis was used to reduce the dimensionality of glass, and then the decision tree method was used to classify glass into two categories, with a prediction accuracy of up to 93%. On this basis, the glass was subclassified, and the decision tree model was used to classify lead-barium glass and high-potassium glass according to whether they were weathered or not. Secondly, K-means algorithm was used to classify weathered glass according to weathering degree. Finally, the sensitivity of the model is tested from the two aspects of parameter selection and data disturbance. The re-judgment results show that the model has high accuracy and strong practicability.

Keywords: Principal component analysis, decision tree algorithm, K-means, Kappa coefficient, sensitivity analysis.

1. Introduction

Ancient glass ornaments are valuable material evidence for modern archeologists to study the trade flows along the ancient Silk Road. Ancient glass is susceptible to weathering due to the influence of burial environment. Based on the basic data of the two types of glass to be studied, this paper classifies the chemical composition of each type of glass into sub-categories, obtains the specific classification methods and results, and conducts reasonable interpretation and sensitivity analysis on the classification results. The application of the model is helpful for relevant experts to scientifically and reasonably analyze the weathering degree of glass relics. At the same time, it also provides a certain solution for the classification of unknown types of ancient glass ornaments, and provides ideas and methods for the research and protection of ancient glass relics.

2. Establishment of the model

In this paper, based on the classification rules of high-potassium glass and lead-barium glass, suitable chemical composition is selected to carry out more detailed subclass division of glass. Firstly, the Z-Score data standardization method is used to unify the data standard [1]. Then, 14 components in glass were processed by principal component analysis (PCA) and the main components of the data stream were selected according to the results. On this basis, decision tree algorithm was introduced to obtain classification rules. Finally, K-means clustering method was used to classify the glass types into sub-classes.

Principal component analysis (PCA) is a comprehensive evaluation method, which has the advantages of dimensionality reduction analysis for a large number of variables Potential [2]. According to the large amount of data in the form, based on principal component analysis, 14 chemical content indexes were selected from the known data to screen out the most important component indexes, and the principal components corresponding to the eigenvalues whose cumulative contribution rate was more than 90% were screened out.

2.1. Decision tree classification law model

The nodes of the decision tree algorithm are used as the judgment conditions, and the specific judgment process is shown in Figure 1.

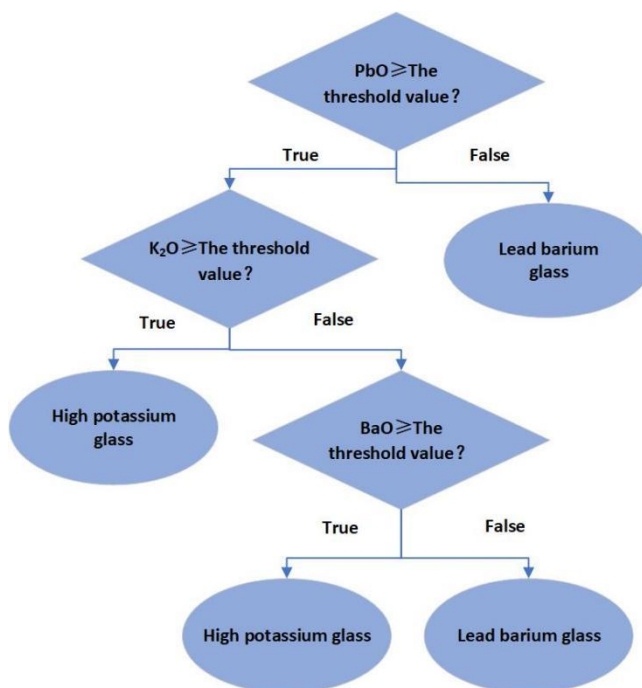


Figure 1. Flowchart of decision tree

2.2. K-means subclassification model

In decision tree model based on the results of classification using k means clustering algorithm to the classification, then two types of glass have samples according to the weathering degree of clustering, divided into two categories, serious and general weathering, requirements in the k cluster grouping, randomly selected initial point as the center of mass, selection of Euclidean distance and distance as a similarity criteria, the Euclidean distance between each sample and the centroid was calculated, and the sample points were clustered into the most similar category by comparing the distance between the samples and the centroid. The iteration is continued until the maximum number of iterations is reached, or the difference between the two iterations is less than a certain threshold, then the iteration is terminated, and the obtained cluster is the final clustering result. The flow chart of K-means algorithm is shown in Figure 2 [3].

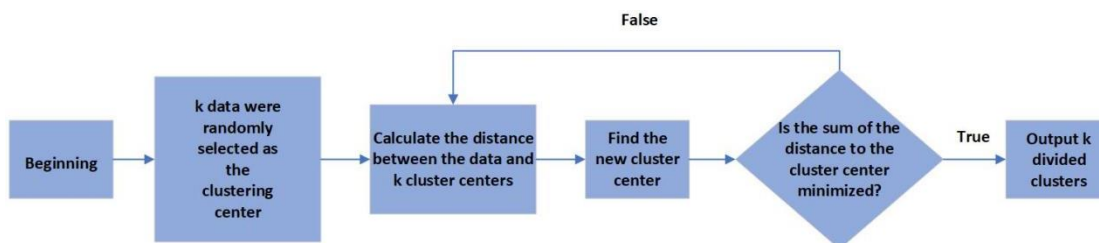


Figure 2. K-means algorithm flow chart

3. Solution and analysis of the model

3.1. Solution of multi-component screening dimensionality reduction model

Firstly, the Z-Score algorithm was used to standardize the chemical composition proportion data of known data. Then, the principal component analysis (PCA) algorithm was used to screen out the

chemical component proportion indexes with cumulative contribution greater than 90% in the known data, so as to achieve the purpose of dimension reduction.

The principal components with cumulative contribution rate of 90% were screened out from the 14 principal components. The ranking results according to the contribution degree were as follows: the first principal component, the second principal component, and the third principal component. According to the principal component analysis results, the proportion of the contribution degree of each principal component is shown in figure3 for the mother-child pie chart [4].

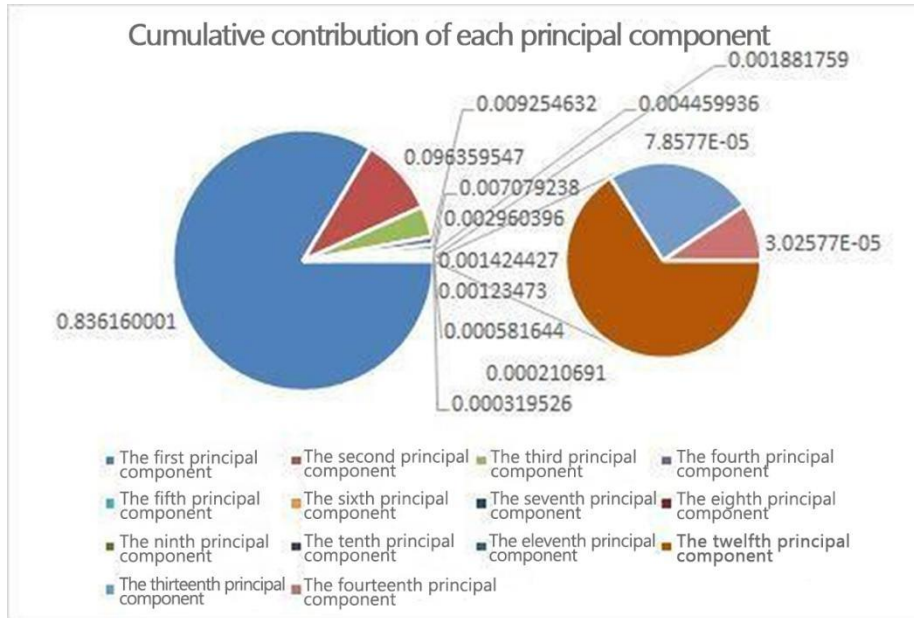


Figure 3. Parent-child pie chart of the contribution ratio of each chemical component

3.2. Solving the classification law model

After obtaining the principal components, the decision tree algorithm is introduced [5]. By judging the proportion of potassium, lead and barium in the glass in each node of the decision tree, the high-potassium glass and lead-barium glass are classified, and finally the classification rules of the two kinds of glass are extracted and summarized. The scatter plot of potassium oxide and sodium oxide content is shown in Figure 4.

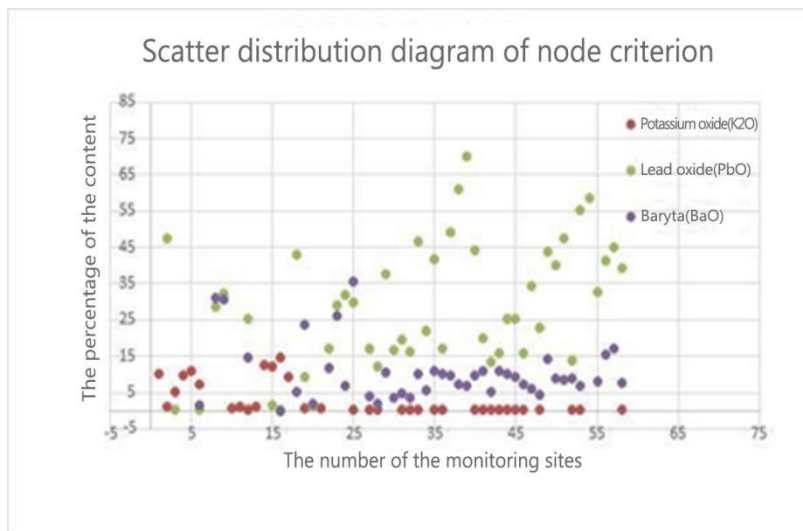


Figure 4. Scatter plot of potassium oxide and sodium oxide content

Based on the decision tree algorithm, the classification rules of glass are summarized and the types of glass are judged. Through python programming, the training samples are simulated and trained by

the algorithm for many times, and then the test samples are tested. Finally, the judgment accuracy reaches 90.47%. The error diagram of test sample results is shown in figure 5 [6].

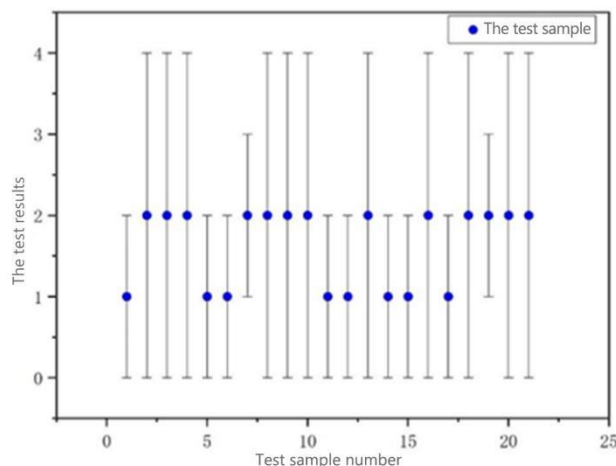


Figure 5. Error diagram of test sample results

Finally, the classification rules of high-potassium glass and lead-barium glass are summarized as follows:

Classification rule: By decision tree to its principal component potassium hydroxide (K₂O) of the first node, if potassium continues to the next node, if the potassium is no lead barium glass, the next node of sodium oxide (Na₂O) to identify whether the content reaches the preset threshold, if the threshold value, then it is k, but its content is low, so decide to lead barium glass, if the content is lower than the threshold, if the sodium content is lower than the threshold, it means that the potassium content reaches the standard and it is a high-potassium glass [7].

Table 1. High potassium glass weathering degree table

Cultural Relics Sampling Sites	The degree of weathering
02	General weathering
08	General weathering
11	Serious weathering
19	General weathering
25	General weathering
26	General weathering
28	Serious weathering
34	Serious weathering
36	General weathering
38	General weathering
39	General weathering
40	General weathering
41	Serious weathering
43	General weathering
48	General weathering
49	General weathering
50	General weathering
51	General weathering
52	General weathering
54	General weathering
56	General weathering
57	General weathering
58	General weathering

Solution and analysis of K-means algorithm subclassification model: In this paper, the known data according to the type of filter after the data to be carried out in accordance with the degree of weathering of classification, will be high in potassium and lead barium glass according to serious analysis and general weathering two, respectively for the weathering of high potassium and lead barium glass k means clustering, then according to the known data of known serious weathering point location category, type, python was used to solve the problem, and the results of subclass division were shown in Table 1, and the degree of weathering was shown in Table 2 [8].

Table 2. Weathering degree Table

Heritage Sampling Sites	The degree of weathering
07	General weathering
09	Serious weathering
10	General weathering
12	Serious weathering
22	General weathering
27	General weathering

3.3. Test and sensitivity analysis of the model

3.3.1 Test of decision tree model

Kappa coefficient can be used to evaluate the accuracy of supervised learning results. In this paper, Kappa coefficient is used to evaluate the classification results of decision tree models. The results of confusion matrix are shown in Figure 6 [9].

The Kappa coefficient was calculated as follows:

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{1}$$

The Kappa=0.8374>0.81 can prove that the classification of decision tree model is almost exactly the same, that is, the accuracy is high.

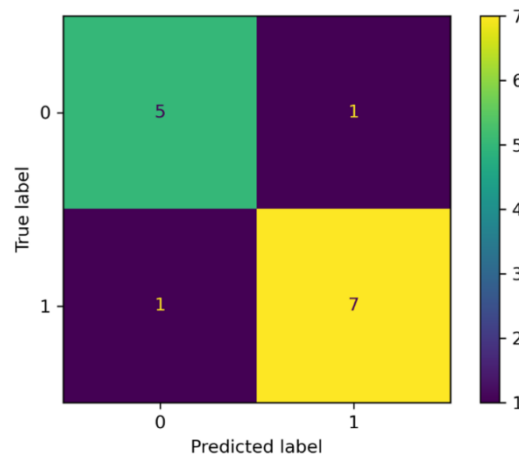


Figure 6. Confusion matrix results

3.3.2 Cluster model checking

The elbow was used to test the best K value of k-means clustering. With the increase of the number of categories selected, the trend of SSE would slow down after passing the inflection point, namely the best k value.

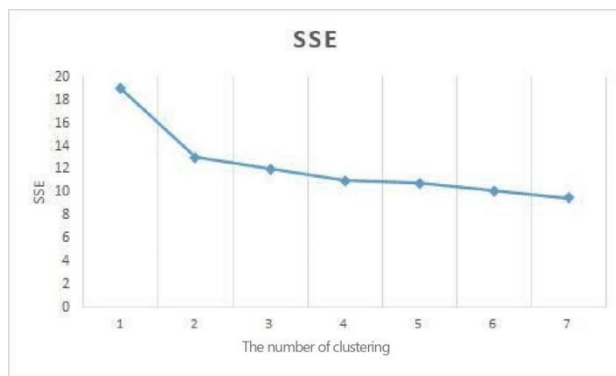


Figure 7. SSE value

As can be seen from figure 7, when K=2, the decrease trend of SSE gradually flattens out, so the optimal number of subcategories in this paper is 2, that is, the model sets the value of K reasonably (severe weathering and general weathering).

3.4. Sensitivity test of the model

3.4.1 Sensitivity test of decision tree model

Using the control variable method, we first control the feature selection criteria and feature partitioning criteria to study the maximum depth. We enumerate different depths to obtain the corresponding test set accuracy and training set accuracy, as shown in Figure 8 [10].

Then, the feature partitioning criteria and maximum depth are controlled to study the feature selection criteria, and 'gini' and 'entropy' are respectively selected to analyze the model accuracy, as shown in Table 3.

Finally, the feature selection criteria and the maximum depth are controlled to study the feature partitioning criteria. 'best' and 'random' are respectively selected to analyze the model accuracy, as shown in Table 4.

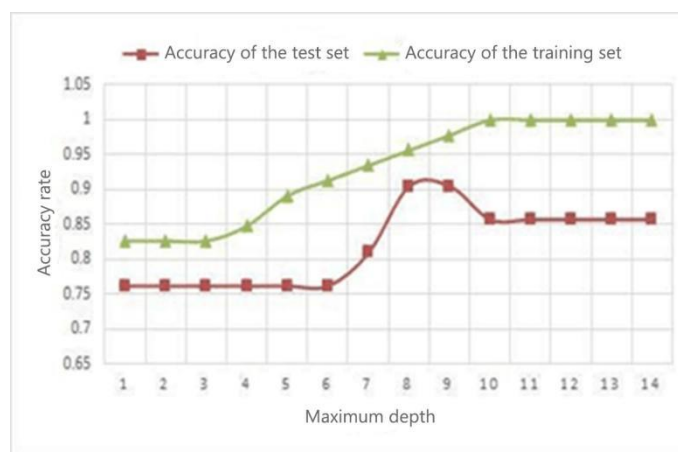


Figure 8. Test set accuracy and training set accuracy

Table 3. Research on Feature selection criteria

	The test set	The training set
gini	0.90	0.893
entropy	0.65	0.9782

Table 4. Research on feature partitioning Criteria

	The test set	The training set
best	0.90	0.903
random	0.857	0.905

From the above analysis results, it can be seen that the selection of the maximum depth parameter is sensitive to the accuracy of the model, while the selection of the feature partition standard parameter is highly anti-sensitive to the accuracy of the model.

3.4.2 K-means clustering model

The number of clusters was studied and analyzed, and the analysis results were shown in figure 9 as the accuracy curve.

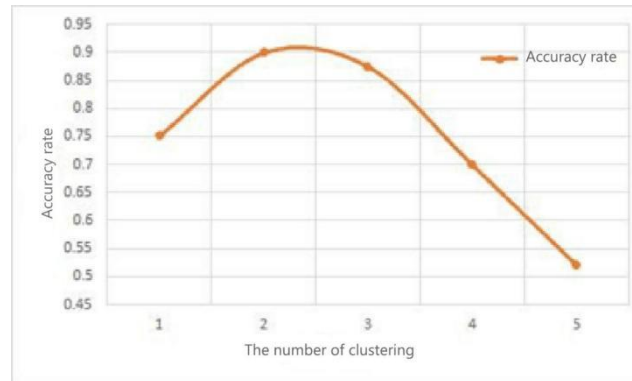


Figure 9. Accuracy curve

Finally, the number of clusters is controlled, and the analysis results are shown in Table 5.

Table 5. n_init study

	accuracy
k-means++	0.9123
random	0.8451

In conclusion, the selection of cluster number has a great impact on the accuracy of the model.

4. Conclusion

The main value of the decision tree algorithm model established in this paper is that it is insensitive and efficient to the absence of data when processing experimental data. The decision tree only needs to be constructed once and used repeatedly, and the maximum computation times of each prediction does not exceed the depth of the decision tree. K-means algorithm has simple principle and fast convergence speed. The clustering effect of K-means algorithm is better, and the specific process of K-means algorithm is more interpretable. It has a significant effect when nesting some classification objects with various types and complex data. The advantage of Pearson correlation coefficient is that it can avoid the problem of grade inflation, and reflect the correlation of the evaluation object objectively, which is scientific and convincing. In conclusion, the model established in this paper has strong practicability.

References

- [1] Kim Helena Kyunghye, Blumberger Daniel M, Karp Jordan F, Lenze Eric, Reynolds Charles F, Mulsant Benoit H. Venlafaxine XR treatment for older patients with major depressive disorder: decision trees for when to change treatment. [J]. Evidence-based mental health, 2022, 12, 09.
- [2] Yang Baojuan. Clothing Design Style Recommendation Using Decision Tree Algorithm Combined with Deep Learning [J]. Computational Intelligence and Neuroscience, 2022, 20, 22.
- [3] Sun Pingli. Mathematical Modeling and Simulation of Online Teaching Effect Evaluation Based on Decision Tree Algorithm [J]. Journal of Function Spaces, 2022, 12, 22.
- [4] Jin Chenxia, Li Fachao, Ma Shijie, Wang Ying. Sampling scheme-based classification rule mining method using decision tree in big data environment [J]. Knowledge-Based Systems, 2022, 2, 44.
- [5] Xie Shengkun. Feature extraction of auto insurance size of loss data using functional principal component analysis [J]. Expert Systems with Applications, 2022, 198, 198.

- [6] Bucherie Agathe, Hultquist Carolynne, Adamo Susana, Neely Colleen, Ayala Fernanda, Bazo Juan, Kruczkiewicz Andrew. A comparison of social vulnerability indices specific to flooding in Ecuador: principal component analysis (PCA) and expert knowledge [J]. *International Journal of Disaster Risk Reduction*, 2022, 73, 73.
- [7] Miao Rui, Guo Peng, Huang Wenjie, Li Qi, Zhang Bo. Profit model for electric vehicle rental service: Sensitive analysis and differential pricing strategy [J]. *Energy*, 2022, 249.
- [8] Xiaohui Lin, Jianmin Xu. Road network partitioning method based on Canopy-Kmeans clustering algorithm [J]. *Archives of Transport*, 2020, 54(2), 09.
- [9] Yu Ma, Wenjuan Cheng. Optimization and Parallelization of Fuzzy Clustering Algorithm Based on the Improved Kmeans++ Clustering [J]. *IOP Conference Series: Materials Science and Engineering*, 2020, 768(7), 768.
- [10] Performance Analysis of KMeans and KMediods Algorithms in Air Pollution Prediction [J]. *International Journal of Recent Technology and Engineering*, 2020, 8(5), 01.