ISSN 2616-5880 Vol. 5, Issue 3: 37-42, DOI: 10.25236/AJMC.2024.050306

Compositional Analysis and Identification of Ancient Glassware

Yanrui Wang

School of Mathematics and Statistics, Beijing Jiaotong University, Beijing, 100044, China

Abstract: *Ancient glass objects exhibit a wide range of compositions and styles, reflecting ancient societies' diverse production techniques and artistic traditions. The research of this paper focuses on the compositional analysis and identification of ancient glassware based on the K-means clustering algorithm, giving a specific classification of two types of glassware and verifying its rationality. Firstly, this paper constructs a decision tree classification model, and the importance of chemical composition characteristics of high-potassium glass, and lead-barium glass, both classified based on lead oxide. Using the K-means clustering algorithm to subclassify the two types of glass artifact samples, the elbow rule was used to determine the best subclasses of lead-barium glass into 5 subclasses and high-potassium glass into 3 subclasses, and the classification effect obtained is perfect, reaching 100%. This paper further refines the criteria and basis for subclass characteristics, conducive to simple and practical subdividing of types. The accuracy of the machine learning methods used in this paper is very high, which shows that the model chosen in this paper is very effective; secondly, this paper gives the specific classification standard for subclassification and reasonably verifies the accuracy of the standard. It is worth mentioning that the accuracy of the machine learning methods used in this paper is very high, which shows that the model this paper chose is very effective; secondly, this paper has given the specific classification standard for subclassification and reasonably verified the accuracy of the standard. This identification method not only ensures better accuracy but also helps to accelerate the identification speed, which is of guiding significance for the identification of ancient glass products in the future; this paper does not only stop at data analysis but also analyzes all aspects in combination with the chemical mechanism, which makes the model this paper built and the results this paper obtained more realistic value.*

Keywords: *Cluster analysis, Decision Tree, Random Forces*

1. Introduction

1.1 Background

The Silk Road was a channel for cultural exchange between China and the West in ancient times, in which glass was valuable material evidence of the early trade exchanges. Early glass in West Asia and Egypt was often made into bead-shaped jewelry imported into China, China's ancient glass absorbed its technology in the local material production, so the appearance of glass products with foreign similar, but the chemical composition is not the same. Analyzing the chemical composition and categorizing these artifacts based on their characteristics provide valuable insights into the materials used and the cultural contexts in which they were created. Recent studies have made significant contributions to the analysis and classification of ancient glass artifacts[1-2]. For instance, Brown et al.[3] utilized EPMA to identify the presence of trace elements in Hellenistic glass vessels, shedding light on regional production centers. Additionally, Smith and Jones[4] applied FTIR spectroscopy to categorize Roman glass beads based on their chemical composition and structural properties. Strugaj, Gentiana, Herrmann, Andreas, and Raedlein, Edda used AES and EDX surface analysis techniques to analyze float glasses exposed to different environmental conditions. The results showed that the surface element content and chemical state of glass changed after exposure to different environmental conditions. These changes may affect the optical properties and durability of the glass[5].

The main raw material of glass is quartz sand, and the main chemical composition is silicon dioxide(SiO_2). The main chemical composition differs for different fluxes added during the calcination process. For example, lead-barium glass in the firing process to add lead ore as a flux, its lead oxide(PbO), and barium oxide (BaO) content is higher, is usually regarded as China's invention of the glass varieties,

ISSN 2616-5880 Vol. 5, Issue 3: 37-42, DOI: 10.25236/AJMC.2024.050306

the glass of the Chu culture is based on lead-barium glass. Potassium glass is made by using substances with high potassium content, such as grass ash, as a flux, and is mainly popular in Lingnan, Southeast Asia, and India. The relationship between compositional analysis and identification of the glass products studied in this paper is shown in Figure 1 the following diagram. The data in this paper is from the 2022 Contemporary Undergraduate Mathematical Contest in Modeling (http://en.mcm.edu.cn/).

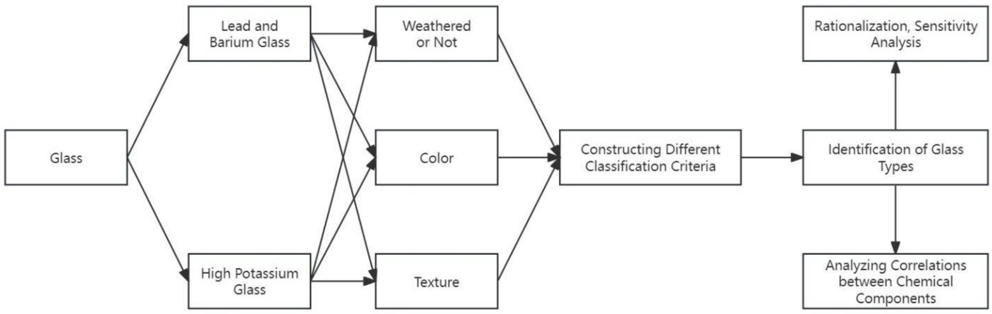

Figure 1: Relationship between compositional analysis and identification of the glass products

1.2 Formulation of the problem

This paper will address the following issues:

Analyze the classification rules of high-potassium glass and lead-barium glass; for each category, choose the appropriate chemical composition to classify them into subcategories, give the specific classification methods and results, and analyze the reasonableness and sensitivity of the classification results.

2. Analyze

In the solution to the problem, this paper should first analyze the statistics of high potassium glass and lead-barium glass and analyze their classification rules. After that, this paper should select appropriate indicators to classify each category into subcategories, and the indicators should be able to accurately classify the two types of glass into the categories with the highest degree of difference. Finally, the results of the classification should be clearly stated and analyzed for reasonableness and sensitivity.

3. Model Assumption

(1) Assuming that the weathering point is representative of the weathering area;

(2) Assume that the loss of color is completely random;

(3) A cumulative percentage of chemical constituents between 85% and 105% is assumed to be valid.

4. Description of Symbols

Descriptions of symbols are shown in Table 1.

Table 1: Description of symbols

Symbol	Clarification
Y_i	Weathered artifact sampling sites, i=1,2,…,58
X_i	Chemical composition in artifact samples(order from left to right), i=1,2,…,14
F_i	Predictable calculation factors, i=0, 1,…,15
μ	residual

where ai denotes the jth coefficient in the multiple linear regression equation for the ith chemical component, I from 1 to 14 denotes silicon dioxide, sodium oxide, potassium oxide, calcium oxide, magnesium oxide, aluminum oxide, iron oxide, copper oxide, lead oxide, barium oxide, diphosphorus pentoxide, strontium oxide, tin oxide, and sulfur dioxide, respectively; and j from 1 to 17 denotes the constant term, ornament_A, ornament_B, and ornament_C, respectively, Lead-barium, high-potassium,

ISSN 2616-5880 Vol. 5, Issue 3: 37-42, DOI: 10.25236/AJMC.2024.050306

light green, light blue, dark green, dark blue, purple, green, blue-green, black, unweathered, weathered.

5. Establishment and solution of K-means clustering algorithm

5.1 Analyze the classification law of high potassium glass and lead-barium glass

To analyze the classification law of high potassium glass and lead-barium glass, this paper takes the chemical compositions of the two as the object of analysis. Since the problem contains all the chemical compositions involved in the two types of glass, the problem is well suited for modeling and solving using a supervised learning model. And because the decision tree has the advantages of insensitivity to missing values and high efficiency, which makes it able to give more accurate results in this question, we use the decision tree for the analysis.

5.1.1 Introduction to the Algorithm

Decision Tree (Decision Tree), also known as decision tree, is an important classification and regression method in data mining technology, which is a predictive analytic model expressed in the form of a tree structure (including binary and multinomial trees), and is an algorithm for supervised learning. The purpose of decision tree learning is to produce a decision tree with strong generalization ability.

Information entropy is derived from Entropy in thermodynamics and is used to express the degree of disorganization of an individual. It represents a measure of the degree of uncertainty or randomness in a given data set. The mathematical formula for calculating information entropy is as follows:

$$Entropy(t) = \sum_{i=1}^{n} \left[P(x_i) \cdot (-\log_2 P(x_i)) \right] = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) \tag{1}$$

Gini is also a measure of the purity of a node and is very similar to the Information Entropy formula:

$$Gini(D) = 1 - \sum_{i=1}^{N} \left[P(x_i) \right]^2 \tag{2}$$

A common algorithm for measuring the purity of a node, i.e., its ability to purify, is to compare the increase or decrease in information entropy before and after the division of the sample, as shown in the formula below.

$$Gain(D,a) = Entropy(D) - \sum_{i=1}^{k} \left(\frac{|D_i|}{|D|} \right) Entropy(D_i) - Entropy(D_i) \tag{3}$$

5.1.2 Algorithm steps

The decision tree is a recursive process from root to leaf, looking for a "division" attribute at each intermediate node.

(1) Start: build the root node; all the samples are placed in the root node, select an optimal feature, split the samples set into subsets according to this feature, and enter the sub-nodes.

(2) All subsets are recursively partitioned according to the attributes of the internal nodes.

(3) If these subsets can be classified almost correctly, then build leaf nodes and assign these subsets to the corresponding leaf nodes.

(4) Each subset is assigned to a leaf node. That is to say, they all have a clear class, so a decision tree is generated.

5.1.3 Result Analysis

From the decision tree training, the feature importance of lead oxide PbO is 100%, and this paper can conclude that two classes of glass are classified according to the content of lead oxide PbO.

According to the trained decision tree, its performance in the training set is 100% in terms of accuracy, recall, and precision.

ISSN 2616-5880 Vol. 5, Issue 3: 37-42, DOI: 10.25236/AJMC.2024.050306

5.2 Subclassification

To find out the appropriate chemical compositions for the subclassification of the two categories, this paper first used the elbow rule to determine the number of species of the two types of glass and the specific classification method and then used K-means cluster analysis to obtain the classification results.

5.2.1 Introduction to the K-means clustering algorithm

K-means is our most commonly used clustering algorithm based on Euclidean distance, which considers that the closer the distance between two targets, the greater the similarity.

The steps of the K-means algorithm are:

(1) Select the initialized K samples as the initial clustering centers $a=a_1,a_2,...a_k$;

(2) For each sample xi in the data set, calculate its distance to the k clustering centers and assign it to the class corresponding to the clustering center with the smallest distance;

(3) For each class a_j, recalculate its cluster center (the center of mass of all samples belonging to that class);

(4) Repeat steps 2 and 3 above until a certain abort condition (number of iterations, change in minimum error, etc.) is reached.

The elbow rule which is shown in Figure 2 plays a guiding role in determining the value of k for the K-means algorithm, and the elbow algorithm can solve this problem effectively.

In the following figure, for example, the y-axis is the SSE (Sum of the Squared Errors) and the x-axis is the value of k. As x increases, the SSE decreases. The y-axis is the SSE (Sum of the Squared Errors-Sum of Errors) and the x-axis is the value of k. As x increases, the SSE decreases, and this value is taken to be the value of k when the decrease tends to be slow.

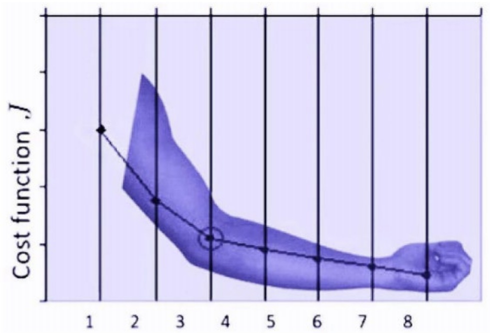

Figure 2: Diagram of the elbow rule

5.2.2 Introduction to the model

(1) Subclassification of high potassium types

A line graph based on the elbow rule was made using Python as shown in Figure 3.

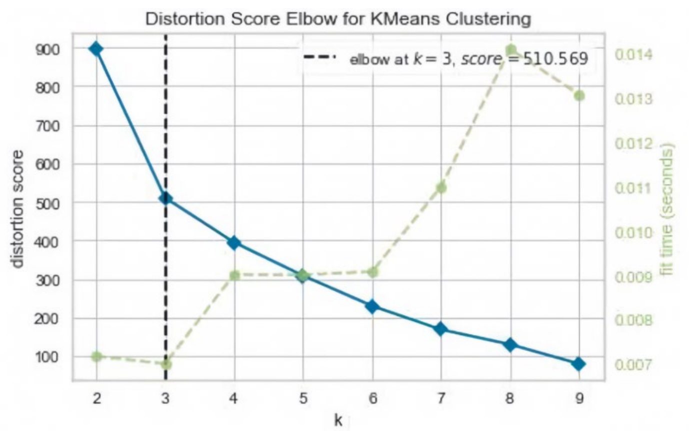

Figure 3: Line chart of the elbow rule (high potassium type subcategory)

ISSN 2616-5880 Vol. 5, Issue 3: 37-42, DOI: 10.25236/AJMC.2024.050306

It is easy to observe that k should be taken as 3, so this paper categorizes the artifacts of high potassium type into 3 subclasses.

Similar to the above still build the decision tree classification model. The established decision tree classification model is applied to the training and test data to obtain the classification evaluation results of the model. Its confusion matrix heat map is shown in Figure 4.

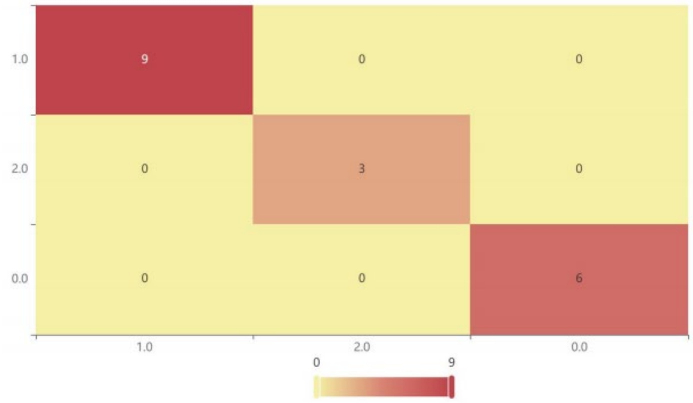

Figure 4: Confusion matrix heat map (high potassium type subcategory)

The accuracy, recall, and precision of the model in the training set are all 100%, which shows that the model achieves a good classification effect.

(2) Subclassification of Lead-Barium Types

The line graph based on the elbow rule using Python is shown in Figure 5.

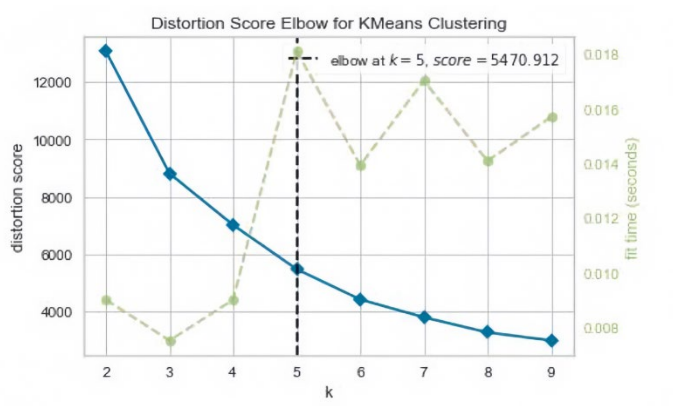

Figure 5: Line chart of the elbow rule (lead-barium type subcategory)

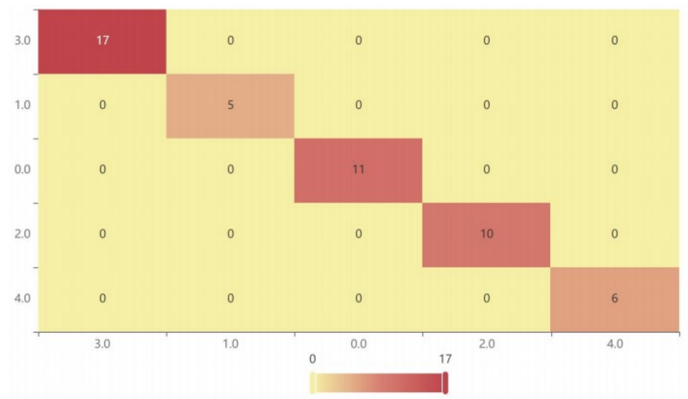

Figure 6: Confusion matrix heat map (lead-barium type subcategory)

It is easy to observe that k should be taken as 5, so this paper categorizes the artifacts of high potassium type into 5 subclasses.

ISSN 2616-5880 Vol. 5, Issue 3: 37-42, DOI: 10.25236/AJMC.2024.050306

Similar to the above still build the decision tree classification model.

Sensitivity analysis of classification results: the decision tree classification model is applied to the training and testing data to obtain the classification evaluation results of the model. The heat map of the confusion matrix is shown in Figure 6.

The accuracy, recall, and precision of the model in the training set are all 100%, which shows that the model achieves good classification results.

5.2.3 Analysis of Research Results

In this paper, it is hoped that the above clustering results can be used to provide a practical subclass classification method as succinctly and efficiently as possible, without relying on machine learning models. It is easy to observe that silica and lead oxide have a greater impact on classification. So in this paper, High Potassium and Barium Lead are subdivided according to the content of silica and lead oxide, and the detailed classification results are shown in Table 2.

Table 2: Subclass Classification Table

Class Name	Type Number	PbO Content	SiO_2 Content
High Potassium	1	0-1	76.68-79.46
	2	0-1.62	59.01-69.33
	3	0-0.25	87.05-96.77
Barium Lead	1	9.3-32.92	37.36-55.21
	2	28.68-32.45	3.72-31.94
	3	12.31-22.05	60.12-75.51
	4	25.39-49.31	17.98-39.57
	5	51.34-70.21	12.41-26.25

6. Conclusion

In this paper, the elbow rule is used to effectively and significantly optimize the clustering algorithm, which helps to find the best number of clusters quickly and accurately. At the same time, the specific classification standard is given for the subclass division, and the accuracy of the standard is reasonably verified, which has guiding significance.

In future research, more consideration can be given to the role of physical properties and other factors in the classification, which can also play a simplified and optimized role in the subclassification criteria. Meanwhile, more consideration can be given to the application of chemical mechanisms in the model to study the problem more clearly and comprehensively.

References

[1] White, E., et al. Typological Classification of Egyptian Glass Artifacts Based on Morphological Features. [J]. Glass Studies, 2021(17):421-435.
[2] Lee, H., et al. Comparative Study of Chinese and Persian Glassware Through FTIR Spectroscopy. [J]. Glass Research, 2023(12):55-68.
[3] Brown, A., et al. Trace Element Analysis of Hellenistic Glass Vessels Using Electron Probe Microanalysis. [J]. Archaeological Science, 2022(48):315-328.
[4] Smith, C., & Jones, D. Characterization of Roman Glass Beads by Fourier-Transform Infrared Spectroscopy. [J]. Archaeometry, 2020(35):82-95.
[5] Strugaj Gentiana, Herrmann Andreas, Rädlein Edda. AES and EDX surface analysis of weathered float glass exposed in different environmental conditions [J]. Journal of Non-Crystalline Solids, 2021, 572.