Chemical subclasses of ancient glass based on K-means

Han Liu [1*], Danbo He [2],Shunxin Li [2]

[1]Computer College of China West Normal University , Nanchong , China

[2]School of Mathematics and Information, China West Normal University , Nanchong , China

*Corresponding author :975226807@qq .com

Abstract. After the opening of the Maritime Silk Road, glass products flooded into China, which is an epitome of ancient China's foreign exchanges. Due to its exposure to the environment, the action of air, water, acid, and alkali substances, the chemical composition of the glass surface changes, thus affecting the judgment of the category. In this study, the subclassification of ancient cultural relics realized the chemical composition analysis of ancient glass and its subclass classification. First of all, the classification rules of high potassium glass and lead barium glass are obtained through analysis, which is divided into four categories: high potassium weathering, high potassium without weathering, lead barium weathering, and lead barium without weathering. Second, the cluster copula map was established, and the subclassification results were obtained combined with the K-means algorithm. Finally, the content of the chemical composition is different, so the division method is reasonable. After changing the original data, that is, changing the content of a chemical composition (-10%, -5%, 5%, 5%, 10%) into the original model, the experimental results can conclude that the model has high sensitivity. Finally, the construction of the subclassification model in this study can better provide the technical support for the analysis and identification of ancient chemical components, which is of great significance and value in the fields of archaeology, cultural heritage protection and, materials science.

Keywords: Ancient glass, chemical composition, subclass division, K-means.

1. Introduction

Glass in west Asia and Egypt to make bead jewelry through the silk road into China [1], ancient glass was easily affected by the buried environment and weathering, weathering process, affected the judgment of class[2], an existing batch of ancient glass products, archaeologists according to the chemical composition and other detection methods are divided into high potassium glass and lead barium glass two types, but because of detection means may lead to chemical composition proportion changes, so we should according to the attachment of cultural relics related data to solve the following problems:

Therefore, this paper aims to study and analyze the classification rules of high-potassium glass and lead-barium glass. We classify each category by selecting the appropriate chemical composition and its results.

2. Materials and Methods

2.1 Data acquisition and preprocessing

The relevant data of silicate glass before and after ancient weathering were sorted out by consulting the data, and the data were preprocessed. Due to missing data, we chose to replace it with 0 to detect the component. At the same time, in this study, the proportion of chemical composition and the data between 85% and 105% were used as valid data [3] to eliminate the remaining data, and then initially realize the dimension reduction of characteristic variables.

2.2 Method Introduction

2.2.1 Lineage map + K-means algorithm

According to the data[4], The K-means algorithm is a classical partitioning and clustering algorithm, Belongs to the distance-based clustering algorithm, With the advantages of simple and fast algorithms and is suitable for processing large data sets, however, because the K value in the k-means clustering algorithm is artificially set, More subjective, So, to optimize the algorithm, This paper will first establish the pedigree map, Lineage maps are drawn according to the interrelationships between categories in the coupling table, By establishing an analytical linkage table, Then according to the analysis of the connection table to establish the pedigree diagram, Finally, according to the lineage map, We can roughly divide it into several categories, That is, the K-value, The chemical composition was then analyzed by the K-means algorithm.

2.2.2 Specific steps

Step1. Establish the analytical association table, build the pedigree map, and determine the K value.

1. Establish four categories of analysis and connection table respectively, that is, the connection table is formed by the distance coefficient matrix. The specific process is to first calculate the distance coefficient between two pairs, obtain the distance coefficient matrix, then select the minimum value in the distance coefficient matrix, and write the minimum value and the class number of the corresponding two categories into the connection table.

2. According to the principle of "keep the small, cross out the large", recalculate the distance values of the rows and columns of the smaller classes in the two categories, cross out the rows and columns of the numbers of the larger classes in the two categories, to form a new distance coefficient matrix.

3. Repeat steps 1 and 2 until the last two categories remain in the matrix.

Finally, getting the spectrum map. The initial K values can then be obtained according to the pedigree plot.

Step2. Cluster analysis of chemical composition using K-means to obtain subclass division of each category.

1. has the first step to obtain the initial K value.

2. Calculate the distance between each chemical composition and the center point of the k cluster class, assign it to the cluster of the nearest cluster center, and then update the center point of the cluster.

3. The second step of the cycle is repeated until the center point no longer changes.

Step3. Finally, select the appropriate chemical components for each category for subclass division.

3. Model building and solution

3.1 The lineage map was constructed and the K values were determined

For each type of data, by using SPSS software, using the chemical composition as a variable, using the clustering method of the link between groups, according to the method of Pearson correlation measurement interval, the cluster association table and pedigree map of each category are obtained. Since the more the number of clusters, the closer the class is, the characteristics of the class will become more and more blurred. Therefore, for good clustering results, the distance between classes should be as large as possible [6]. According to the analysis of the lineage diagram of lead-barium weathering, k is taken as 3, as shown in Figure 1. According to the analysis of the unweathered lead barium, k is 4, as shown in Figure 2. According to the analysis of the cluster association table and lineage diagram of high potassium weathering, k is taken as 2, as shown in Figure 3. According to the analysis of the high potassium unweathered lineage map, k takes 3, as shown in Figure 4.

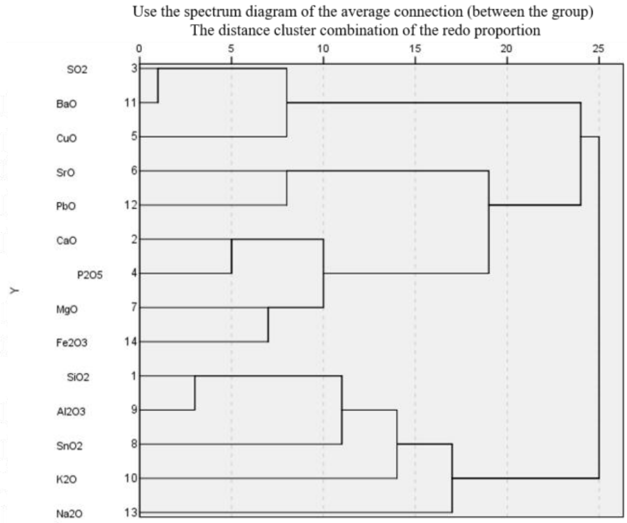

Figure 1. The pedigree diagram of lead-barium weathering

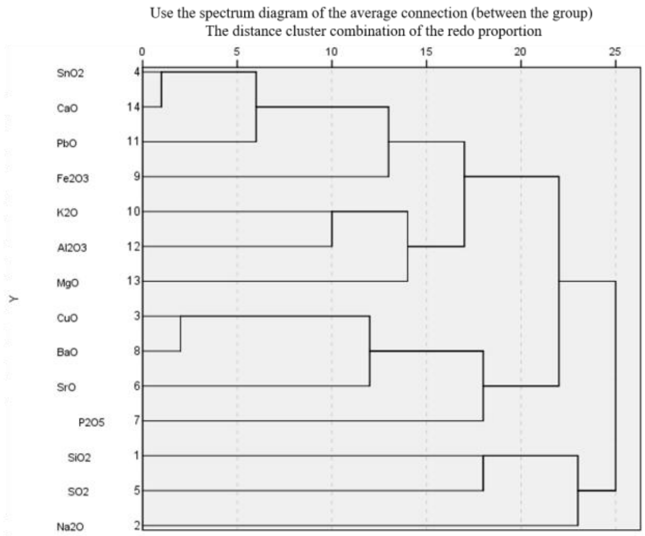

Figure 2.Unweathered pedigree diagram of lead and barium

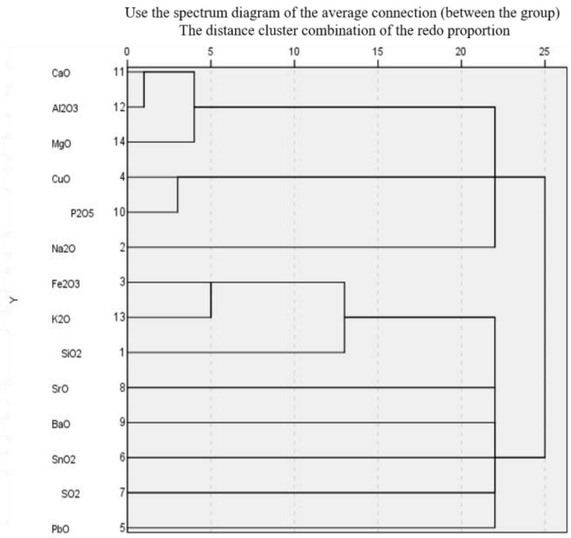

Figure 3. The pedigree diagram of high potassium weathering

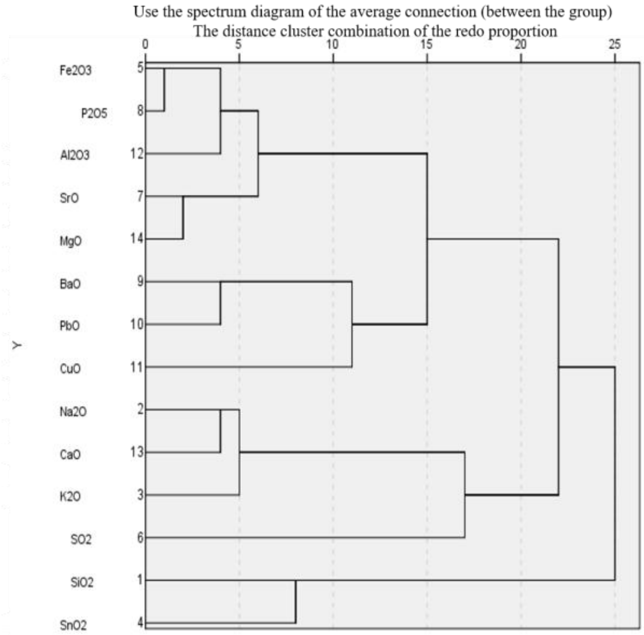

Figure 4. High K unweathered lineage

3.2K-means cluster analysis

Based on the obtained K values, K-means clustering analysis [7] was performed.

Step1 The obtained K value is taken as the initial K value.

Step2 Euclidean distance calculation formula:

$$d = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \tag{1}$$

Calculate the distance from each chemical composition to the cluster center, add the nearest one to a cluster, and then recalculate the cluster center point. The center point of each cluster species can be obtained by using the SPSS run.

(1) The central point of each cluster species, as shown in Table 1: (K=3)

Table.1. High potassium undefensed cluster center

Cluster species	Central value _ silica (SiO2)	Central value _ Sodium oxide (Na2O)	Center value _ potassium oxide (K2O)	Central value _ Calcium oxide (CaO)	Center value _ magnesium oxide (MgO)	Center value _ alumina (Al2O3)	Center value _ Iron oxide (Fe2O3)	Center value _ Copper oxide (CuO)	Central value _ lead oxide (PbO)
1	63.5942	1.1914	11.759	7.4085	0.9271	6.42	1.7685	2.9542	0.44857
2	81.0633	0	4.8700	2.24	0.9166	4.4333	0.79	1.3533	0.4166
3	63.7300	0	7.525	2.705	1.855	10.6000	4.215	2.3449	0.275

(2) Central point of each cluster, as shown in Table 2: (K=2)

Table. 2. Center coordinates of high K weathering clusters

Cluster species	Central value _ silica (SiO2)	Central value _ Sodium oxide (Na2O)	Center value _ potassium oxide (K2O)	Central value _ Calcium oxide (CaO)	Center value _ magnesium oxide (MgO)	Center value _ alumina (Al2O3)	Center value _ Iron oxide (Fe2O3)	Center value _ Copper oxide (CuO)	Central value _ lead oxide (PbO)
1	92.566	0	0.2466	1.22333	0.39333	2.6633	0.24	1.7766	0
2	95.36	0	0.84	0.5166	0	1.1966	0.29	1.3466	0

(3) The central point of each cluster species, as shown in Table 3: (K=4)

Table .3. Center coordinates of the lead-barium unweathered cluster

Cluster species	Central value _ silica (SiO2)	Central value _ Sodium oxide (Na2O)	Center value _ potassium oxide (K2O)	Central value _ Calcium oxide (CaO)	Center value _ magnesium oxide (MgO)	Center value _ alumina (Al2O3)	Center value _ Iron oxide (Fe2O3)	Center value _ Copper oxide (CuO)	Central value _ lead oxide (PbO)
1	34.65	0	0.355	0.235	0	3.52	0.755	6.6200	19.22
2	66.4066	0.4433	0.1166	0.8016	0.4383	2.83	0.96	0.7866	17.958336
3	35.6350	0	0.7049	4.365	0.745	4.1049	2.43	0	38.4800
4	51.92	2.4566	0.18	0.6667	0.7599	3.1	0	0.76	27.8566

(4) Central points of each cluster of PK weathering, as shown in Table 4 (K=2):

Table .4. Center coordinates of lead and barium weathering cluster

Cluster species	Central value _ silica SiO2	Center value _ Sodium oxide Na2O	Center value _ potassium oxide K2O	Center value _ calcium oxide CaO	Center value _ magnesium oxide MgO	Central value _ alumina Al2O3	Center value _ iron oxide Fe2O3	Center value _ CCr 2 CuO	Center value _ PO PbO
1	55.9754	2.6790	0.1809	1.5618	0.8972	6.7827	0.5309	1.1536	19.7218
2	23.7760	0.1927	0.126	2.6904	0.6144	2.5427	0.5668	2.3668	44.418

Finally, specific subclasses can be obtained, as shown in Table 5:

Table. 5. Results of subclass classification

type	Surface weathering	chemical composition	class
Lead barium	morals and manners	Sulfur dioxide, barium oxide, copper oxide	Lead-barium weathering · barium sulfur copper
		Strontium oxide, lead oxide, calcium oxide, phosphorus pentoxide, magnesium oxide, iron oxide	Lead barium weathering · strontium lead calcium phosphoramafic iron
		Silica dioxide, alumina, tin oxide, potassium oxide, sodium oxide	Lead barium weathering · silicon aluminum tin potassium sodium
	No weathering	Tin oxide, calcium oxide, lead oxide, iron oxide, potassium oxide, aluminum oxide, magnesium oxide	Lead and barium have no weathering · tin, calcium, lead, iron, potassium, aluminum and magnesium
		Copper oxide, barium oxide, strontium oxide, phosphorus pentoxide	Lead barium no weathering copper barium strontium and phosphorus
		Silica dioxide, sulfur dioxide	Lead and barium without weathering · silica-sulfur
		sodium oxide	Lead and barium without weathering · sodium
High potassium	morals and manners	Calcium oxide, aluminum oxide, magnesium oxide, copper oxide, phosphorus pentoxide, sodium oxide	High in potassium, weathering, calcium, aluminum, magnesium, copper, phosphorus, and sodium
		Iron oxide, potassium oxide, silicon dioxide, strontium oxide, barium oxide, tin oxide, sulfur dioxide, lead oxide	High potassium weathering · iron potassium silicon strontium barium tin sulfur lead
	No weathering	Iron oxide, phosphorus pentoxide, aluminum oxide, strontium oxide, magnesium oxide, barium oxide, lead oxide, copper oxide	High potassium, without weathering, iron, phosphorus, aluminum, strontium, magnesium, barium, aluminum, and copper
		Sodium oxide, calcium oxide, potassium oxide, sulfur dioxide	High in potassium, without weathering, sodium, calcium, potassium and sulfur
		Silica dioxide, tin oxide	High potassium without weathering · silicon tin

3.3 Rationality and sensitivity analysis

3.3.1Rationality analysis

That is, to analyze the difference in the chemical composition between various categories, if the difference is large, it shows that the division standard is reasonable [8]. Take three subcategories of lead and barium weathering as an example:

Firstly, all the data of lead and barium weathering were selected for a total of 36 sets of data, and then the data were expressed by a two-dimensional bar graph in Excel, as shown in Figure 5:

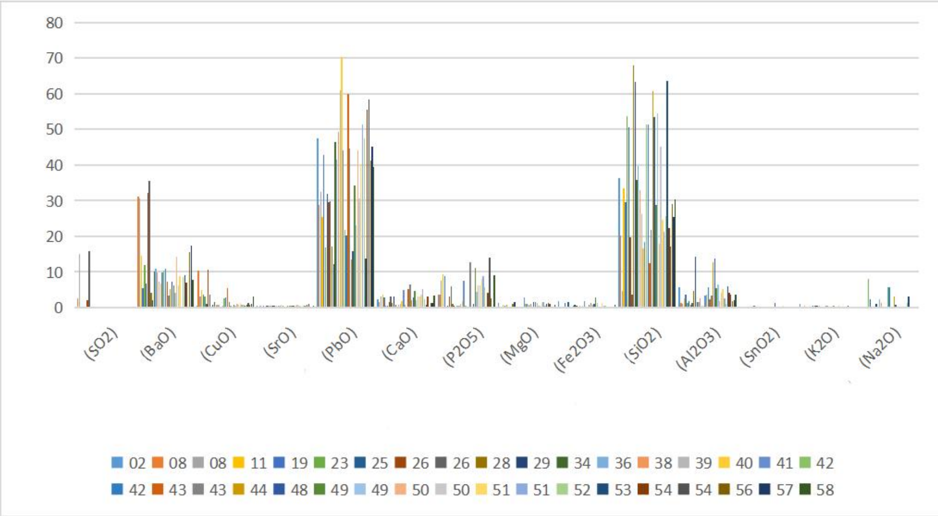

Figure 5 High K unweathered lineage

Through observation analysis can be seen, sulfur dioxide, barium oxide, copper oxide as a cluster, strontium oxide, lead oxide, calcium oxide, phosphorus pent, magnesium oxide, iron oxide as a cluster, silica, alumina, tin oxide, potassium oxide, sodium oxide gathered for a cluster, and the content difference between three clusters is larger, so the division way is reasonable.

3.3.2 Sensitivity analysis

To verify whether the classification results are relatively accurate, the sensitivity analysis of the classification results is conducted by adding a perturbation to the data of a certain variable [9]. The representative element silica of the chemical composition was selected, and the silica content data were increased by-10%, -5%, ,5% and 10%, and then the changed data were brought into the original model to observe whether the results will change [10], as shown in Figure 6. Therefore, the sensitivity is high, the model is good, and it is worth popularizing.

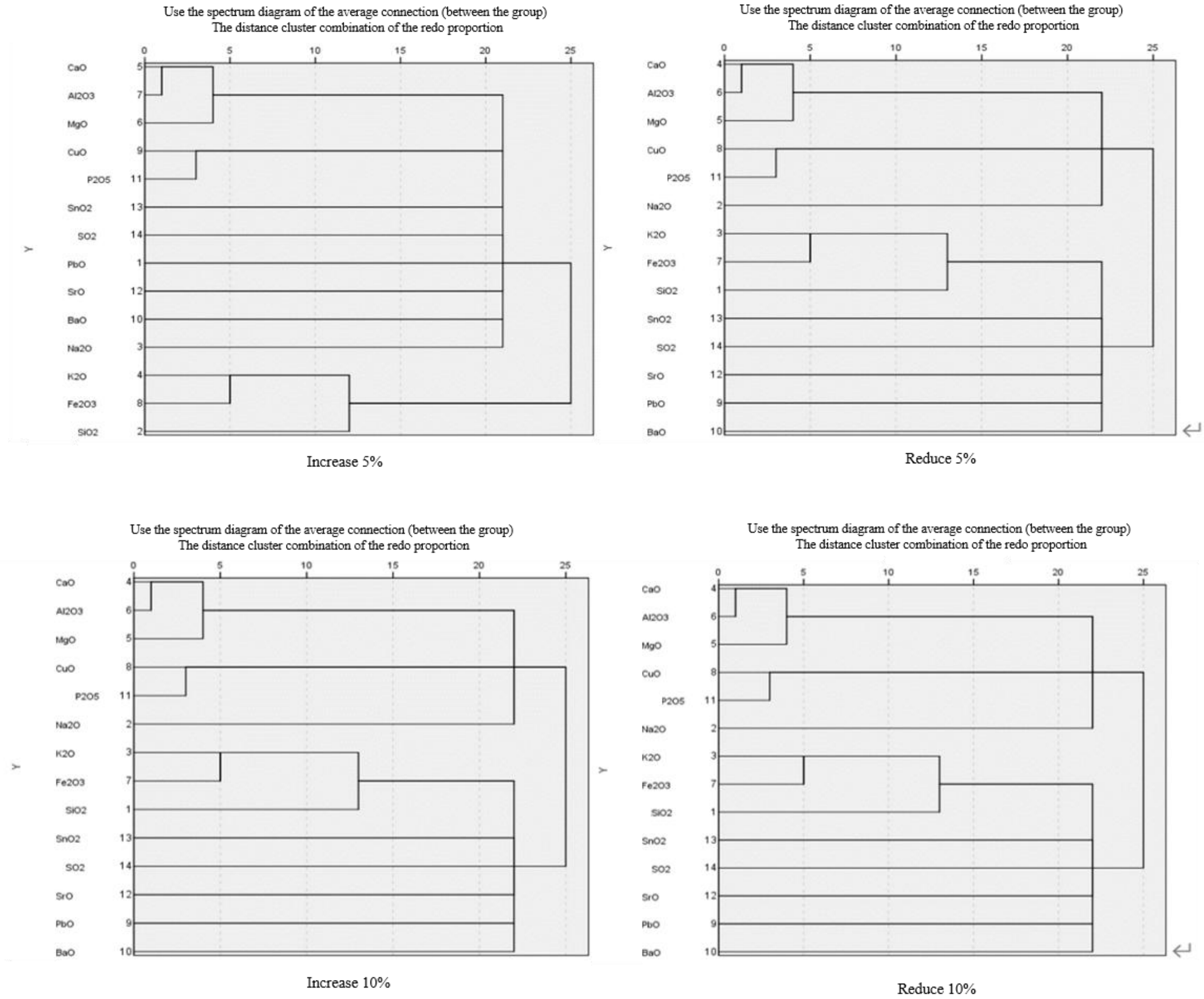

Figure 6. The pedigree plot after data variation

4. Conclusions

The establishment of the research model of the subclass division of ancient glass can better understand the production process, source, and development process of ancient glass, and provide basic data and a scientific basis for the identification, protection, restoration and research of ancient glass. Through the study of the chemical composition subclass division of ancient glass samples, the production period, origin and technical characteristics of glass can be determined, and the development trajectory and communication of the glass process of ancient glass can be inferred. In addition, the study of the chemical composition of ancient glass can also provide a reference for the modern glass manufacturing industry and promote the innovation and development of glass manufacturing technology. Therefore, the study of chemical subdivision of ancient glass has great significance and value in the fields of archaeology, cultural heritage protection and material science.

References

[1] Dussubieux Laure,Fenn Thomas R.,Abraham Shinu Anna,Kanungo Alok Kumar.Tracking ancient glass production in India: elemental and isotopic analysis of raw materials[J].Archaeological and Anthropological Sciences,2022,14(12).

[2] Zou Ying.Molecular-Composition Analysis of Glass Chemical Composition Based on Time-Series and Clustering Methods[J].Molecules,2023,28(2).

[3] Zlámalová Cílová Zuzana,Blažková Gabriela.Chemical composition of early modern and modern window glass from archaeological excavations in Prague-Hradčany, Czech Republic[J].Archaeometry,2022,64(6).

[4] Ikotun Abiodun M.,Ezugwu Absalom E.,Abualigah Laith,Abuhaija Belal,Heming Jia.K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data[J].Information Sciences,2023,622.Fangfang.Research on power load forecasting based on Improved BP neural network [D].Harbin Institute of Technology, 2011.

[5] Mussabayev Rustam,Mladenovic Nenad,Jarboui Bassem,Mussabayev Ravil.How to Use K-means for Big Data Clustering?[J].Pattern Recognition,2023,137.Ma Kunlong.Short term distributed load forecasting method based on big data [D].Changsha: Hunan University, 2014.

[6] Shengqiang Yao,Yang Nie,Wei Zhu,Xiaojing Yang.Study on Composition Analysis and Identification of Ancient Glass Products[J].Analytical Chemis try A Journal,2022,1(1).

[7] Ikotun Abiodun M.,Ezugwu Absalom E.,Abualigah Laith,Abuhaija Belal,Heming Jia.K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data[J].Information Sciences,2023,622.

[8] Qing Li,Chuang Wang,Jianwei Zhai.Research on Composition Analysis and Identification Based on Ancient Glass Products[J].Academic Journal of Materials & Chemistry,2022,3(2).

[9] Devika T.J.,Ravichandran J..A clustering method combining multiple range tests and K-means[J].Communications in Statistics - Theory and Methods,2022,51(21).

[10] Yang Jiannan.A general framework for probabilistic sensitivity analysis with respect to distribution parameters[J].Probabilistic Engineering Mechanics,2023,72.