

# 基于多元回归和聚类分析的古代玻璃制品成分分析与鉴别

包令言

江苏大学财经学院, 江苏 镇江

收稿日期: 2023年6月20日; 录用日期: 2023年7月24日; 发布日期: 2023年7月31日

## 摘要

本文主要研究古代玻璃制品在长期埋藏环境影响下不同类型玻璃文物表面风化前后化学成分比例之间的关系并进行相应的预测与鉴别。通过建立相关性分析和偏最小二乘回归模型进行各化学成分间关系分析, 结合持时放大比例进行风化前化学成分比例预测; 同时建立聚类分析模型进行文物分类并鉴别。此研究在现实物理化学意义的角度分析问题, 有助于古代玻璃制品出土后的成分分析与鉴别。

## 关键词

相关性分析, 偏最小二乘回归, 聚类分析, 轮廓系数, 物理化学变化

# Composition Analysis and Identification of Ancient Glass Products Based on Multiple Regression and Cluster Analysis

Lingyan Bao

School of Finance and Economics, Jiangsu University, Zhenjiang Jiangsu

Received: Jun. 20<sup>th</sup>, 2023; accepted: Jul. 24<sup>th</sup>, 2023; published: Jul. 31<sup>st</sup>, 2023

## Abstract

This paper mainly studies the relationship between the chemical composition ratio of different types of glass cultural relics before and after surface weathering of ancient glass products under the influence of long-term burial environment, and makes corresponding prediction and identification. Through the establishment of correlation analysis and partial least squares regression model to analyze the relationship between various chemical components, combined with the time-sustained

amplification ratio to predict the proportion of chemical components before weathering; at the same time, a cluster analysis model was established to classify and identify cultural relics. This study analyzes the problem from the perspective of realistic physical and chemical significance, which is helpful for the composition analysis and identification of ancient glass products after unearthed.

## Keywords

Correlation Analysis, Partial Least Squares Regression, Cluster Analysis, Silhouette Coefficient, Physical and Chemical Changes

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



## 1. 问题重述

### 1.1. 问题背景

世界最早的玻璃来源于古埃及，从发明至今已经有 4000 多年历史。作为中西方文化交流的通道，丝绸之路在中西方贸易往来中也发挥了巨大的作用，而其中的玻璃饰品则成为中西方文化交流的纽带之一。

玻璃的主要原料是石英砂，主要化学成分是二氧化硅( $\text{SiO}_2$ )。在炼制纯石英砂时需要添加助熔剂，古代常用的助熔剂有草木灰、天然泡碱、硝石和铅矿石等，并添加石灰石作为稳定剂，石灰石煅烧以后转化为氧化钙( $\text{CaO}$ )。添加的助熔剂不同，其主要化学成分也不同。

古代玻璃极易受埋藏环境(土壤湿度、温度等条件)的影响而风化。由于发生元素的大量交换，导致其成分比例发生变化，从而影响对其类别的正确判断。表面无风化的玻璃表面能明显看出文物的颜色、纹饰，但不排除局部有较浅的风化；表面风化的玻璃表面大面积灰黄色区域为风化层，是明显风化区域，紫色部分是一般风化表面。在部分风化的文物中，其表面也有未风化的区域。

现有一批我国古代玻璃制品的相关数据，考古工作者依据这些文物样品的化学成分和其他检测手段已将其分为高钾玻璃和铅钡玻璃两种类型，本文旨在对于所给数据建立预测与统计数学模型对问题进行分析并进行相应的合理性和敏感性检验。

### 1.2. 已知条件

- 1) 玻璃文物的基本信息；
- 2) 已分类玻璃文物的化学成分比例；
- 3) 未分类玻璃文物的化学成分比例。

### 1.3. 需解决的问题

1) 探究玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系并进行合理分析；结合玻璃的类型，配合分析文物样品表面有无风化化学成分比例的统计规律，并根据风化点检测数据，对其风化前的化学成分比例进行预测。

2) 依据附件数据分析高钾玻璃、铅钡玻璃的分类规律；对于每个类别选择合适的化学成分对其进行亚类划分，给出具体的划分方法及划分结果，并对分类结果的合理性和敏感性进行分析。

- 3) 对未知类别玻璃文物的化学成分进行分析, 鉴别其所属类型, 并对分类结果的敏感性进行分析。
- 4) 针对不同类别的玻璃文物样品, 分析其化学成分之间的关联关系, 并比较不同类别之间的化学成分关联关系的差异性。

## 2. 模型假设

- 1) 假设玻璃文物中存在颜色空缺, 假定无色进行处理;
- 2) 假设出现统一文物样品的不同部位(同为风化或无风化)时, 其化学成分比例按照两者平均值进行处理;
- 3) 假设假定同个文物风化前后种类保持不变;
- 4) 假设题中所给的数据均真实可靠。

## 3. 模型符号及说明

模型的符号及说明如表 1。

**Table 1.** Model symbols and descriptions

**表 1.** 模型符号及说明

符号	符号含义	单位
$\chi^2$	卡方检验卡方值	/
$O_i$	实际结果频数	/
$E_i$	期望值频数	/
$P$	显著性 P 值	/
$\rho$	Spearman 相关系数	/
$FJ_{1j}$	风化高钾玻璃化学成分均值	%
$WJ_{1j}$	无风化高钾玻璃化学成分均值	%
$FJ_{2j}$	风化铅钡玻璃化学成分均值	%
$WJ_{2j}$	无风化铅钡玻璃化学成分均值	%
$\mu$	样本均值	/
$s$	样本方差	/
$n$	聚类类的个数	/
$s(i)$	样本轮廓系数	/
$lm(i)$	变化幅度	/

## 4. 模型的建立与求解

### 4.1. 问题一模型建立与求解

#### 4.1.1. 数据预处理

针对玻璃文物的表面风化与其玻璃类型、纹饰和颜色数据进行观察后, 发现对于铅钡类型的玻璃存

在四个玻璃文物没有颜色数据,可能存在如题目所说玻璃文物表面风化较严重,无法观察出其原本颜色,故对这四个颜色数据采取赋予透明色处理。

在填补空白数据后,对数据进行分类整理,分为表面风化与无风化两类,便于后续检验。

#### 4.1.2. 对玻璃文物表面风化与其类型、纹饰、颜色进行分析

##### 1) 卡方检验

玻璃文物的基本信息所给数据为定类数据,问题一第一小问主要探究分类变量与分类变量之间的关系,故采用针对定类数据的卡方检验进行相关性分析。通过统计软件 SPSS 对玻璃文物表面是否风化与其类型、纹饰以及颜色进行卡方检验。

卡方检验作为检验两个定类变量是否相关的一种统计方法,卡方值的大小表示实际结果与期望值相关性大小。卡方检验的公式为:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

其中,  $\chi^2$  为卡方值,  $O_i$  为实际结果频数,  $E_i$  为期望值频数。

在此问中,通过卡方检验结果 P 值判断两定类变量间的相关性,运用 SPSS 得出卡方检验结果如表 2:

Table 2. Chi-square test results analysis results

表 2. 卡方检验结果分析结果

定类变量	卡方值	校正卡方值	P 值	变量与表面风化相关
颜色	6.287	6.287	0.507	无
类型	5.400	4.134	0.020	有
纹饰	5.747	5.747	0.056	无

因卡方检验要求样本最好是大样本且每个个案最好出现一次,1/4 个案需至少出现五次,但玻璃文物的基本信息中数据并不完全符合要求,故需在得出结果时校正卡方值,校正卡方值结果也列于表 2 中。

对于上述卡方检验结果进行分析,针对表面风化与玻璃文物颜色,其显著性 P 值为 0.507,接受原假设,在水平上并不呈现显著性,说明玻璃文物的颜色不同,其表面风化并不存在显著性差异;针对表面风化与玻璃文物类型,其显著性 P 值为 0.020,拒绝原假设,在水平上并呈现显著性,说明玻璃文物的类型不同,其表面风化存在显著性差异;针对表面风化与玻璃文物纹饰,其显著性 P 值为 0.056,接受原假设,在水平上并不呈现显著性,说明玻璃文物的纹饰不同,其表面风化同样不存在显著性差异。

总结上述分析得到玻璃文物表面风化与玻璃文物的颜色和纹饰关系不显著,与玻璃文物的类型有显著关系。对于玻璃文物类型与表面风化的关系,表 1 结果显示,铅钡玻璃中有 66.67% 表面风化,而高钾玻璃中仅有 33.33% 表面风化,证明玻璃文物的类型与表面是否风化存在相关性。

##### 2) Spearman 相关性分析

因玻璃文物数据的关系并不符合线性相关关系,且不服从绝对正态分布,故可以采用 Spearman 相关性分析。针对 Spearman 相关性分析需采用定量数据,故将玻璃文物的基本信息中数据赋值 0, 1, 2, 3, 以此表示定量数据且不会对结果造成影响。

对于 Spearman 相关系数计算公式如下:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2)$$

通过 SPSS 做出 Spearman 相关性分析结果, 如表 3:

**Table 3.** Spearman correlation coefficient results  
**表 3.** Spearman 相关性系数结果

定量变量	表面风化 Spearman 相关系数	显著性 P 值
颜色	-0.071	0.595
类型	0.344	0.008
纹饰	0.037	0.781

由表 3 可知, 玻璃文物类型的显著性 P 值 < 0.05, 呈现显著性, 颜色与纹饰显著性 P 值均>0.1, 不满足 10% 的显著性水平, 即只有玻璃文物类型与表面风化之间存在相关性, 与卡方检验结果相同。

综上所述方法, 查阅文献的得知: 高钾玻璃中 SiO<sub>2</sub> 化学成分比例 > 铅钡玻璃中 SiO<sub>2</sub> 化学成分比例, 对于 SiO<sub>2</sub> 化学成分比例较高的玻璃抗风化性能较好。对于高钾玻璃, 风华过程中表层 K<sub>2</sub>O 大量流失, 但 SiO<sub>2</sub> 化学成分比例会显著提高, 而 SiO<sub>2</sub> 化学成分比例的增加从一定程度上保护玻璃文物不被风化[1]。对于铅钡玻璃, SiO<sub>2</sub> 化学成分比例较高钾玻璃少, 故稳定性不高, 易被风化。

#### 4.1.3. 针对文物样品表面有无风化化学成分比例的统计规律

因问题一第一小问结论为玻璃文物表面风华与玻璃文物的类型有显著关系, 故第二小问延续第一小问结论, 通过文物样品的类型进行划分而后讨论其统计规律, 分别讨论高钾玻璃在有无风化条件下的化学成分比例与铅钡玻璃在有无风化条件下的化学成分比例。查阅文献得知, 可认为同类型玻璃在风化前化学成分比例大致相同[2]。

##### 1) 数据预处理

根据题目所述, 表单 2 中数据具有成分性特点, 即各成分比例累加和应为 100%, 但由于检测手段技术问题可能造成各成分累加和不为 100%, 此时需要将成分比例累加和范围进行界定, 即 85%~100% 间均为有效数据, 故需对附件表单 2 中数据进行数据清洗。文物编号为 15 和 17 的采样点化学成分采样累加和分别为 79.47% 和 71.89%, 不满足 85%~100% 区间, 故将文物编号为 15 和 17 的无效采样样本剔除后进行后续分析。

##### 2) 高钾玻璃在有无风化条件下化学成分比例统计规律

对处理后的数据进行简单处理, 即对应文物编号出现取样为两个部位情况时(如 06 编号文物, 取样部位 1 和部位 2), 取两部位化学成分比例均值替代整体化学成分比例。将表单 2 中数据进行分类, 具体分为有风化条件下的高钾玻璃、无风化条件下的高钾玻璃、有风化条件下的铅钡玻璃和无风化条件下的铅钡玻璃四类。在 2 中对高钾玻璃有无风化条件下化学成分比例(因玻璃主要成分为 SiO<sub>2</sub>, 其占比较高, 对结果影响不大, 故进行剔除)进行统计分析。

分类后对上述两种条件下文物采样样品化学成分比例进行简单可视化呈现, 如图 1。

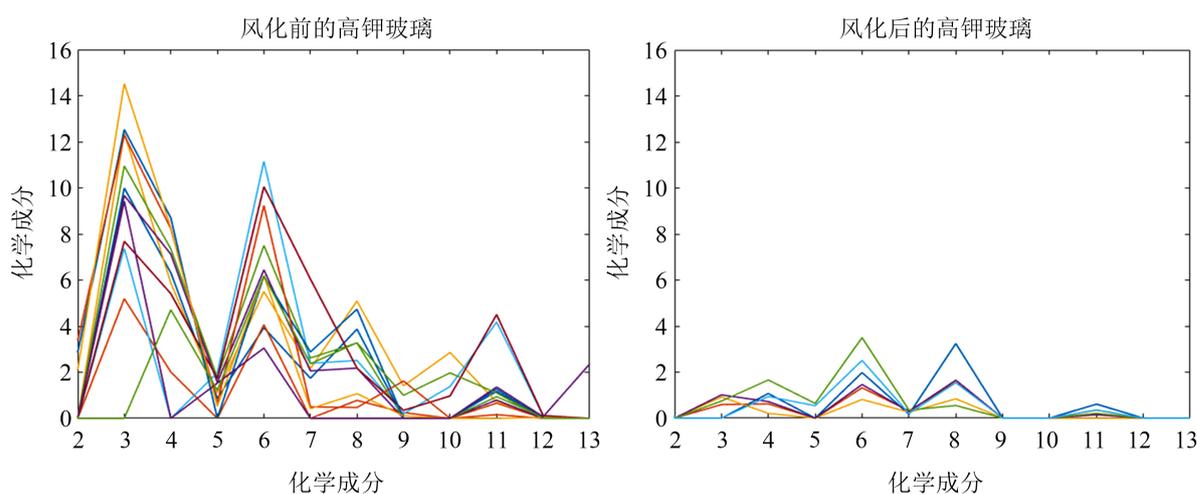
下述 0-13 分别代表附件中对应的化学成分, 折线图横坐标按 SiO<sub>2</sub>、Na<sub>2</sub>O、K<sub>2</sub>O、CaO、MgO、CuO、Al<sub>2</sub>O<sub>3</sub>、Fe<sub>2</sub>O<sub>3</sub> 等顺序依次延续; 纵坐标为百分数%。可得统计规律如下:

a) 不同文物样本有无风化情况下整体化学物质分布大致相同。

### b) 风化玻璃样品 $\text{SiO}_2$ 检测化学成分比例增加, $\text{K}_2\text{O}$ 化学成分比例降低。

查阅文献, 对于高钾玻璃文物风化过程, 玻璃表面中氧化钾大部分析出流失, 虽然玻璃风化比未风化样本有更富集的 Si、Al、Ca 和 Fe 元素, 但正如上述所说, K 元素流失严重; 与此相反, 风化样品较未风化样品  $\text{SiO}_2$  化学成分比例显著增加, 经化学分析, 可能是因为环境中存在  $\text{Al}_2\text{O}_3$  和  $\text{Fe}_2\text{O}_3$  等黏土类物质, 也对玻璃进行了一定程度上的保护[3]。

观察数据, 发现 3、6 对应的化学成分  $\text{K}_2\text{O}$  和  $\text{Al}_2\text{O}_3$  在风化前化学成分比例高, 而风华后化学成分比例变低。同时还发现未风化高钾玻璃样本  $\text{K}_2\text{O}$  化学成分比例约占 9.62%, 风化高钾玻璃样本  $\text{K}_2\text{O}$  化学成分比例约占 0.82%, 从图 1 也可直观地看出,  $\text{K}_2\text{O}$  化学成分比例明显减少; 而对于玻璃主要成分  $\text{SiO}_2$  而言, 无风化文物样本中  $\text{SiO}_2$  化学成分比例约占 67.03%, 均位于 60%~80% 区间段; 风化文物样本中  $\text{SiO}_2$  化学成分比例约占 93.96%, 均位于 90%~100% 区间段, 得出  $\text{SiO}_2$  化学成分比例也显著增加。上述主要物质均符合查阅文献所得的化学规律。



**Figure 1.** The proportion of chemical composition of some sampling points of high-potassium glass with or without weathering  
**图 1.** 高钾玻璃部分采样点有无风化条件下的化学成分比例

### 3) 铅钡玻璃在有无风化条件下化学成分比例统计规律

在 2 中对高钾玻璃有无风化条件下化学成分比例进行统计分析, 对于铅钡玻璃则同样绘制两种条件下铅钡玻璃文物采样样品化学成分比例进行简单可视化呈现, 如图 2。

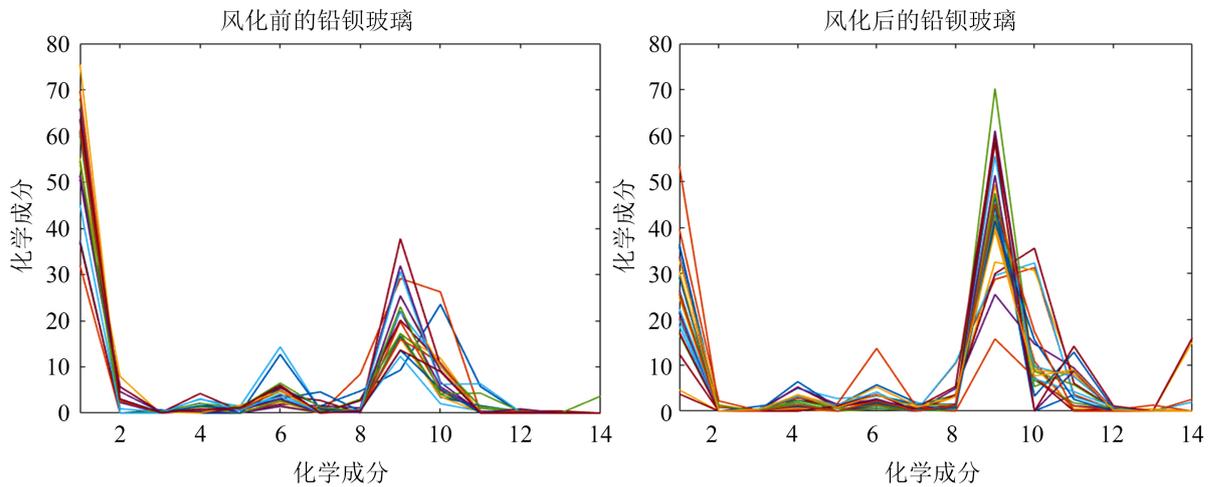
下述 0-14 分别代表附件中对应的化学成分, 如下述 2, 可得统计规律如下:

a) 不同文物样本有无风化情况下整体化学物质分布大致相同, 有部分样本成单独分布趋势。

b) 风化玻璃样品  $\text{BaO}$  检测化学成分比例增加,  $\text{SiO}_2$  化学成分比例降低。

同样查阅文献可知, 根据文献中不同区域分析的 EDX 结果, 铅钡玻璃在风化过程中 Si 元素由里至外顺着风化层的走向流失, 在最外层结壳中化学成分比例已经非常低; Ba 元素在最外层结壳中积聚形成夹心结构, 推测在最外层形成了大量难溶性的钡盐[4]。

观察数据, 发现序号 9、10 对应的化学成分  $\text{PbO}$  和  $\text{BaO}$  化学成分比例在风化后明显变多; 同时发现未风化铅钡玻璃样本  $\text{BaO}$  化学成分比例约占 10.50%, 风化铅钡玻璃样本  $\text{K}_2\text{O}$  化学成分比例约占 11.44%, 图 2 也可看出,  $\text{BaO}$  化学成分比例增加; 对于  $\text{SiO}_2$  而言, 无风化文物样本中  $\text{SiO}_2$  化学成分比例约占 53.44%, 主要位于 40%~60% 区间段; 风化文物样本中  $\text{SiO}_2$  化学成分比例约占 33.61%, 主要位于 30%~50% 区间段, 得出  $\text{SiO}_2$  化学成分比例也显著增加。上述主要物质均符合文献所述的化学规律。

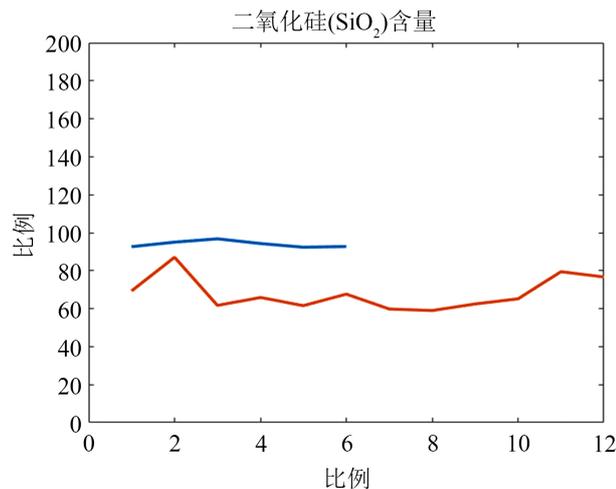


**Figure 2.** The chemical composition ratio of some sampling points of lead-barium glass with or without weathering  
**图 2.** 铅钡玻璃部分采样点有无风化条件下的化学成分比例

#### 4.1.4. 根据风化点检测数据预测风化前化学成分比例

##### 1) 利用持时放大比例建立模型

对风化前后的玻璃文物样本各种化学物质成分进行绘图,如图 3 可发现化学成分比例变化较为平稳,如图 3,可发现高钾玻璃风化前后  $\text{SiO}_2$  化学成分比例十分平稳,故使用持时放大比例[5]来分析各成分风化前后的关系。



**Figure 3.** The chemical composition ratio of  $\text{SiO}_2$  before and after weathering of high potassium glass

**图 3.** 高钾玻璃风化前后  $\text{SiO}_2$  的化学成分比例

采用持时放大比例,即对各化学成分取均值后求商的方法分别对高钾玻璃和铅钡玻璃有无风化条件下的玻璃样本取均值,分别如下:

风化高钾玻璃的各化学成分均值:

$$FJ_{1j} = \sum_{i=1}^6 \frac{gif(i,j)}{6}, j=1,2,\dots,14 \quad (3)$$

无风化高钾玻璃的各化学成分均值:

$$WJ_{1j} = \sum_{i=1}^{12} \frac{gfw(i,j)}{12}, j=1,2,\dots,14 \quad (4)$$

风化铅钡玻璃的各化学成分均值:

$$FJ_{2j} = \sum_{i=1}^{24} \frac{qbf(i,j)}{24}, j=1,2,\dots,14 \quad (5)$$

无风化铅钡玻璃的各化学成分均值:

$$WJ_{2j} = \sum_{i=1}^{21} \frac{qbw(i,j)}{21}, j=1,2,\dots,14 \quad (6)$$

其中,  $gfw(i,j)$  为风化高钾玻璃的各化学成分比例,  $qbf(i,j)$  为风化铅钡玻璃的各化学成分比例;  $gfw(i,j)$  为未风化高钾玻璃的各化学成分比例,  $qbw(i,j)$  为未风化铅钡玻璃的各化学成分比例。

根据上述求得风化前后的各化学成分的均值, 将无风化的化学成分均值与风化的均值相除, 分别得出高钾玻璃与铅钡玻璃在有无风化条件下的比值  $BI_1$  和  $BI_2$ ,  $BI_1$  和  $BI_2$  公式如下:

$$BI_1 = \frac{WJ_{1j}}{FJ_{1j}}, j=1,2,\dots,14 \quad (7)$$

$$BI_2 = \frac{WJ_{2j}}{FJ_{2j}}, j=1,2,\dots,14 \quad (8)$$

此过程中需要剔除  $x_{1j}$  与  $x_{2j}$  为零的情况, 单独将其比值赋值为 0, 避免出现比值不存在的情况。

根据风化点监测数据  $gfw(i,j)(i=1,2,\dots,6)$  和  $qbf(i,j)(i=1,2,\dots,24)$  分别将其乘上  $BI_1(j)$  和  $BI_2(j)$  预测出该点风化前的化学成分比例  $gfw'(i,j)$  和  $qbw'(i,j)$ , 最终可得公式为:

$$gfw'(i,j) = BI_1(j) \times gfw(i,j), i=1,2,\dots,6; j=1,2,\dots,14 \quad (9)$$

$$qbw'(i,j) = BI_2(j) \times qbf(i,j), i=1,2,\dots,24; j=1,2,\dots,14 \quad (10)$$

## 2) 模型的求解与检验

以高钾玻璃风化前化学成分比例为例, 模型求解结果部分如表 4 (详细结果见支撑材料):

**Table 4.** Solution results of chemical composition ratio of high-potassium glass samples before weathering  
**表 4.** 高钾玻璃样本风化前化学成分比例求解结果

文物采样点	SiO <sub>2</sub> (%)	Na <sub>2</sub> O (%)	K <sub>2</sub> O (%)	CaO (%)	MgO (%)
01	69.33	0	9.99	6.32	0.87
03	74.38	0	8.78	3.94	0.56
04	65.88	0	9.67	7.12	1.56
...	...	...	...	...	...
21	76.68	0	0	4.71	1.22
22	66.82	0	12.71	10.17	3.51
27	67.08	0	0	5.76	2.96

根据模型求解结果对该模型进行检验, 可知样本 49 本身作为分化后的铅钡玻璃却具有未风化部分, 故采用此样本数据对模型进行检验。经验证, 发现模型误差率约为 0.02%, 故  $BI_2$  精准可信, 使用相同方法得出的  $BI_1$  也具备较高的可信度, 从而印证了模型的合理性。

## 4.2. 问题二模型建立与求解

### 4.2.1. 通过偏最小二乘回归模型的建立分析分类规律

因题目所述, 古代玻璃极易受埋藏环境的影响而风化。在风化过程中, 内部元素与环境元素进行大量交换, 导致其成分比例发生变化, 从而影响对其类别的正确判断, 故需通过问题一的预测结论, 将所有风化后玻璃样本所含化学成分比例转化成风化前的化学成分比例, 统一研究玻璃样本风化前的化学成分与分类的关系从而对类别进行分类。

通过分析各项指标, 可知各化学成分之间由于化学转化的关系存在着极强的线性关系且高钾玻璃的样本数仅有 16 个, 铅钡玻璃的样本数仅有 40 个, 远远小于使用机器学习模型的样本条件, 故选用近年来兴起的偏最小二乘法回归模型对其进行回归分析。在分析前, 将高钾玻璃类型赋值 1, 铅钡玻璃类型赋值 10, 将二者分离开, 并利用  $\text{rand}(n, m)$  函数给予该赋值 1% 的震荡, 以避免后续回归分析无意义。

#### 1) 建立观测数据矩阵

以无风化铅钡玻璃为例,  $x_1, x_2, \dots, x_{14}$  分别表示无风化高钾玻璃的  $\text{SiO}_2$ 、 $\text{Na}_2\text{O}$ 、 $\text{K}_2\text{O}$  等 14 个化学成分比例指标;  $y$  表示无风化高钾玻璃的分类赋值指标。  $x$  对应  $a$ ,  $y$  对应  $b$ , 进而建立自变量和因变量的观测数据矩阵, 分别为:

$$A = (a_{ij})_{40 \times 14} \quad (11)$$

$$B = (b_{ij})_{40 \times 1} \quad (12)$$

其中  $i$  表示第  $i$  个玻璃样本,  $j$  表示第  $j$  个指标值。

#### 2) 数据标准化

在研究风化前铅钡玻璃化学成分与风化前玻璃类别赋值之间的联系的过程中, 每个指标的单位、最值、最优区间均不同, 要进行后续分析, 必须对上述理化指标进行数据标准化。

采用 Z-Score 标准化, 将不同量级的数据转化为统一量度的 Z-Score 分值进行比较, 提高数据可比性, 削弱数据解释性。

将各指标值  $a_{ij}$  转化成标准值  $\tilde{a}_{ij}$ , 有

$$\tilde{a}_{ij} = \frac{a_{ij} - \mu_j^{(1)}}{s_j^{(1)}}, i = 1, 2, \dots, 40; j = 1, 2, \dots, 14 \quad (13)$$

在此式中,  $\mu$  样本均值和  $s$  样本标准差分别用下式表示:

$$\mu_j^{(1)} = \frac{1}{40} \sum_{i=1}^{40} a_{ij} \quad (14)$$

$$s_j^{(1)} = \sqrt{\frac{1}{40-1} \sum_{i=1}^{40} (a_{ij} - \mu_j^{(1)})^2} \quad (15)$$

对应地, 称为标准化指标变量。

$$\tilde{x}_j = \frac{x_j - \mu_j^{(1)}}{s_j^{(1)}}, j = 1, 2, \dots, 14 \quad (16)$$

类似地, 也将  $b_{ij}$  转换成标准化指标值  $\tilde{b}_{ij}$ , 有

$$\tilde{b}_{ij} = \frac{b_{ij} - \mu_j^{(2)}}{s_j^{(2)}}, i=1, 2, \dots, 40; j=1 \quad (17)$$

在此式中,  $\mu$  样本均值和  $s$  样本标准差分别用下式表示:

$$\mu_j^{(2)} = \frac{1}{40} \sum_{i=1}^{40} b_{ij} \quad (18)$$

$$s_j^{(2)} = \sqrt{\frac{1}{40-1} \sum_{i=1}^{40} (b_{ij} - \mu_j^{(2)})^2} \quad (19)$$

与  $a_{ij}$  处理过程相似, 称为标准化指标变量:

$$\tilde{y}_j = \frac{y_j - \mu_j^{(2)}}{s_j^{(2)}}, j=1 \quad (20)$$

#### 4.2.2. 偏最小二乘回归模型的求解

##### 1) 求解相关系数矩阵

通过 MATLAB 求解相关系数矩阵, 如表 5:

**Table 5.** The partial correlation coefficient matrix of lead-barium glass index

**表 5.** 铅钡玻璃指标的部分相关系数矩阵

指标	SiO <sub>2</sub>	Na <sub>2</sub> O	K <sub>2</sub> O	CaO	Fe <sub>2</sub> O <sub>3</sub>	CuO	分类
SiO <sub>2</sub>	1.00	0.37	0.15	-0.19	0.13	-0.38	-0.27
Na <sub>2</sub> O	0.37	1.00	-0.10	-0.26	-0.16	-0.17	-0.23
K <sub>2</sub> O	0.15	-0.10	1.00	-0.05	0.27	-0.07	-0.07
CaO	-0.19	-0.26	-0.05	1.00	0.44	-0.23	0.07
Fe <sub>2</sub> O <sub>3</sub>	0.13	-0.16	0.27	0.44	1.00	-0.25	0.08
CuO	-0.38	-0.17	-0.07	-0.23	-0.25	1.00	0.10
分类	-0.27	-0.23	-0.07	0.07	0.08	0.10	1.00

从表 5 中的相关系数矩阵可以看出, 分类赋值指标与 SiO<sub>2</sub>、Na<sub>2</sub>O、K<sub>2</sub>O 等化学成分指标呈负相关, 而与 CaO、Fe<sub>2</sub>O<sub>3</sub>、CuO 等呈正相关。

##### 2) 提出自变量组和因变量组成分

使用 MATLAB 软件求得各对成分部分组成分如下(完整数据见支撑材料):

$$\begin{cases} u_1 = -0.0587\tilde{x}_1 - 0.0504\tilde{x}_2 + \dots + 0.0224\tilde{x}_{14} \\ v_1 = 2.4287y \\ u_2 = -0.0300\tilde{x}_1 - 0.0479\tilde{x}_2 + \dots + 0.0022\tilde{x}_{14} \\ v_2 = 1.6489y \\ u_3 = -0.0193\tilde{x}_1 + 0.0351\tilde{x}_2 + \dots + 0.0004\tilde{x}_{14} \\ v_3 = 0.9059y \\ \dots \end{cases} \quad (21)$$

经过自变量组和因变量组的成分分析,取前 12 个成分解释自变量的比率为 93.85%,故取前 12 个成分进行后续分析。

### 3) 标准化指标变量与成分变量回归

通过 MATLAB 进行自变量组和因变量组  $u_1, u_2, \dots, u_{12}$  之间的回归方程分别为(完整数据见支撑材料):

$$\begin{cases} \tilde{x}_1 = -4.8475u_1 + 0.8310u_2 + \dots + 0.2985u_{12} \\ \tilde{x}_2 = -3.3208u_1 - 2.2820u_2 + \dots + 0.5871u_{12} \\ \tilde{x}_3 = -1.3306u_1 + 1.1140u_2 + \dots - 1.5675u_{12} \\ \dots \\ y = 2.4287u_1 + 1.9519u_2 + \dots + 0.0190u_{12} \end{cases} \quad (22)$$

### 4) 因变量组与自变量组回归

将 2 中的成分  $u_i$  代入 3 中的  $y$  回归方程,得到标准化指标变量之间的回归方程,如下,其中常数项为 0:

$$\tilde{y} = 0.1784\tilde{x}_1 - 0.1750\tilde{x}_2 + \dots + 0.1750\tilde{x}_4 \quad (23)$$

将标准化变量分别还原成原始变量,得到如下原始数据回归方程系数,从而得到回归方程,如下:

$$y = 0.0002x_1 - 0.0006x_2 + \dots + 0.0070x_{14} + 9.8720 \quad (24)$$

上述均为风化前铅钡玻璃化学成分与分类赋值的偏最小二乘回归分析,风化前高钾玻璃化学成分与分类赋值偏最小二乘回归分析过程同上,得出高钾玻璃对应的偏最小二乘回归方程,如下:

$$y = 0.0003x_1 - 0.0003x_2 + \dots - 0.0002x_{14} + 0.9678 \quad (25)$$

### 4.2.3. 模型的解释与检验

为更直观地观察自变量与解释变量的边际作用,采用针对标准化数据绘制回归系数的直方图的方法,绘制出如图 4,从左到右依次为高钾玻璃、铅钡玻璃化学成分与分类赋值标准化数据回归系数直方图,如图 4:

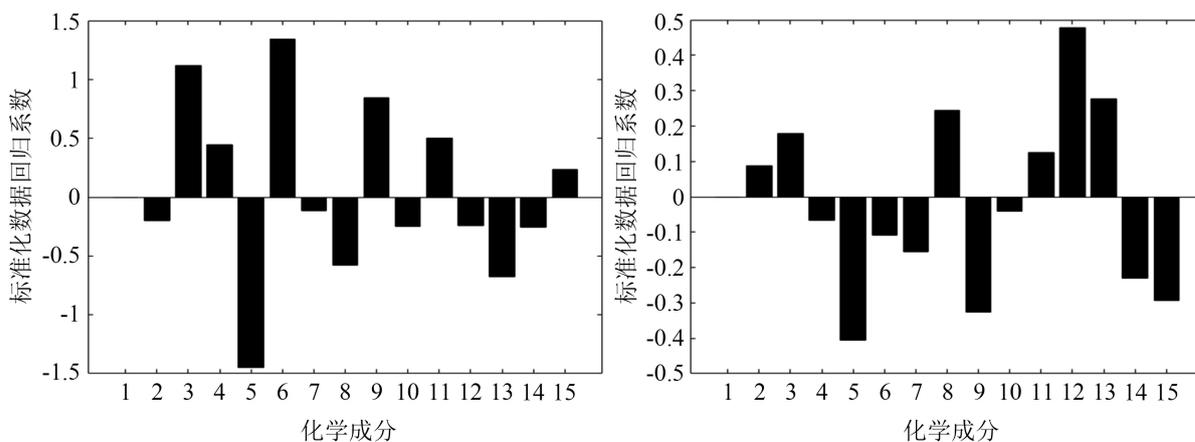


Figure 4. Histogram of regression coefficients of high potassium, lead-barium glass chemical composition and classification assignment standardized data

图 4. 高钾、铅钡玻璃化学成分与分类赋值标准化数据回归系数直方图

从标准化数据回归系数图中可以基本观察到,直方图对应的长度越长,标准化数据回归系数越大,则化学成分指标对分类赋值指标的影响越大。

为考察回归模型的精度，用 MATLAB 绘制预测图，在预测图上，若所有点都能在图的对角线两侧呈均匀分布，则方程的拟合值与原值差异很小，拟合效果令人满意，绘制高钾玻璃与铅钡玻璃化学成分与分类赋值指标间的预测图，如图 5：

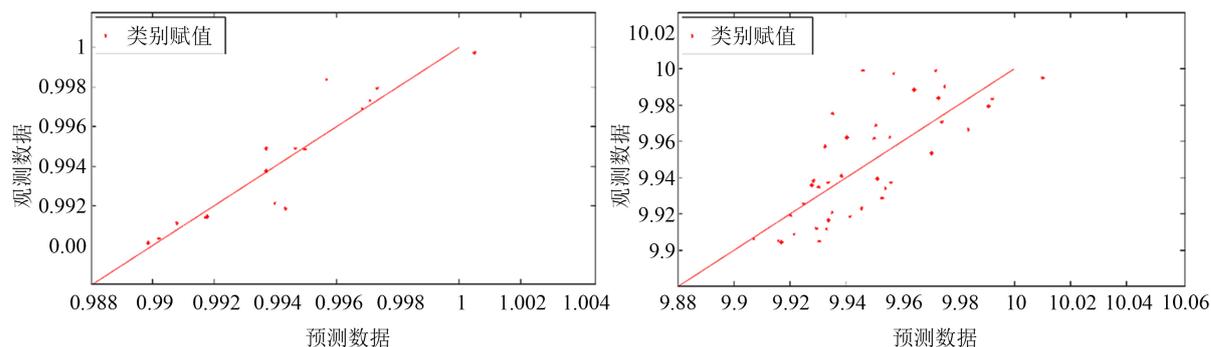


Figure 5. Prediction chart between chemical composition and classification index of high-potassium glass and lead-barium glass

图 5. 高钾玻璃与铅钡玻璃化学成分与分类赋值指标间预测图

通过高钾玻璃和铅钡玻璃的预测图均可以看出，所有点几乎都能在图的对角线两侧均匀分布，故拟合效果令人满意，则用偏最小二乘回归求得的原始变量回归方程可以定量解释化学成分与分类赋值间的联系。

#### 4.2.4. 利用 K-means 算法聚类模型的建立

对于每个类别选择合适的化学成分对其进行亚类划分，采用基于 K-means 算法的聚类分析，将高钾玻璃和铅钡玻璃样本分别划分为多个类。作为一种预测算法与模型，其具体建立过程，如图 6：

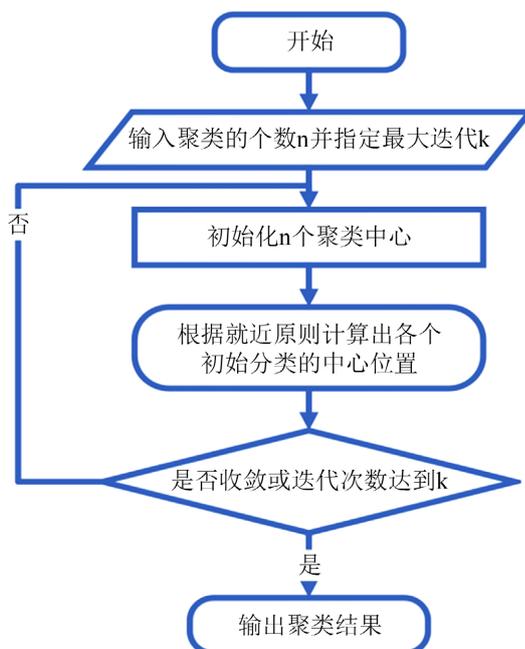


Figure 6. K-means cluster analysis flowchart

图 6. K-means 聚类分析流程图

经过计算机多次运算后，可得到所需分类结果，根据分类结果特征对亚类进行一定程度的描述。

#### 4.2.5. 利用 K-means 算法聚类模型的求解

##### 1) 对高钾玻璃进行模型求解

由于 K-means 聚类分析需人为确定类的个数，为排除人为因素的影响，建模求解过程中对数据进行多次尝试，分别尝试不同聚类数下指标是否符合均匀分布，最终确定类的个数为  $n = 4$ 。经聚类分析后，每类占比分别为 43.75%、18.75%、25.00%和 12.50%，较其他的类个数下分布更均匀，如图 7：

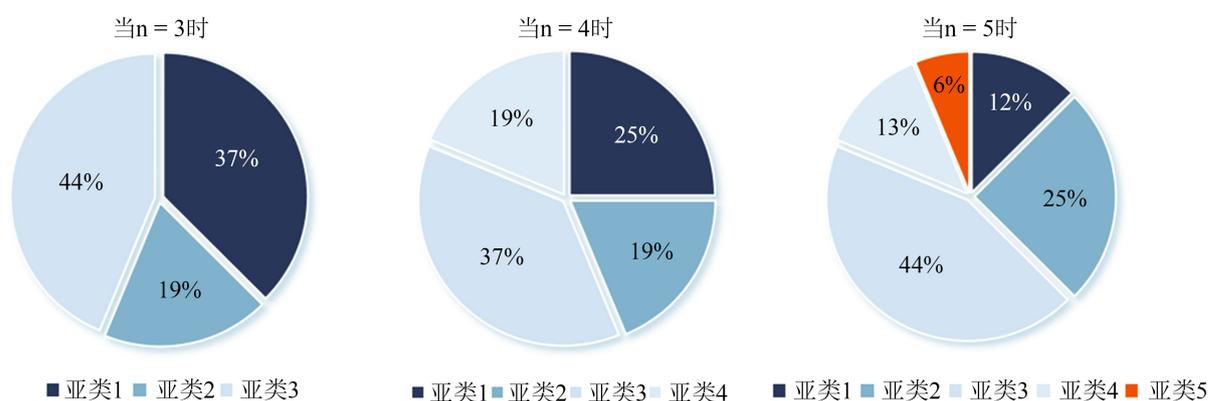


Figure 7. The proportion of cluster types when high-potassium glasses are different types

图 7. 高钾玻璃不同类时聚类种类占比

将聚类分析后结果如表 6 (每一聚类种类只列出两个文物样本的部分化学成分比例，所有结果详细数据见支撑材料)：

Table 6. Partial results of high potassium glass cluster analysis

表 6. 高钾玻璃聚类分析部分结果

文物样本编号	聚类种类	SiO <sub>2</sub> (%)	CaO (%)	MgO (%)	Al <sub>2</sub> O <sub>3</sub> (%)
4	1	65.88	9.67	7.12	1.56
5	1	61.58	10.95	7.35	1.77
7	2	67.02	6.56	0	6.79
21	2	76.68	4.71	1.22	6.19
1	3	68.75	3.80	0	4.53
9	3	70.01	1.29	0	2.78
3	4	74.38	3.94	0.555	4.78
18	4	79.46	0	1.53	3.05

经过相关文献查阅，共分析了 16 条玻璃文物样本，由于玻璃主要成分为 SiO<sub>2</sub>，每类样本二氧化硅化学成分比例均较大，故进行亚类划分时不讨论 SiO<sub>2</sub>。高钾玻璃样本可分为四个亚类，每个亚类根据其主化学成分比例进行相应命名[6]：

**亚类 1:** 由于其  $\text{Al}_2\text{O}_3$  和  $\text{CaO}$  高,  $\text{K}_2\text{O}$  化学成分比例较高,  $\text{MgO}$  化学成分比例高; 为便于辨别各亚类, 命名为 Al-Ca-K-Mg 亚类; 该亚类包含玻璃文物样本 4、5、6、13、14、16、22;

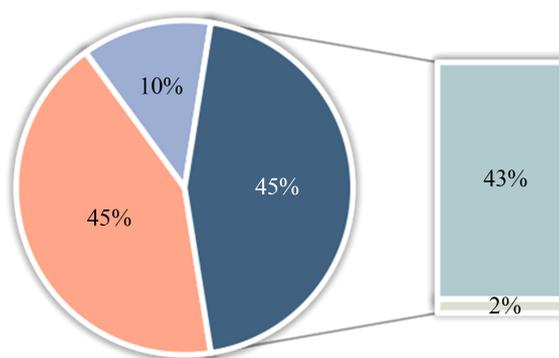
**亚类 2:** 由于其  $\text{Al}_2\text{O}_3$  和  $\text{CaO}$  化学成分比例较高,  $\text{K}_2\text{O}$  化学成分比例为 0,  $\text{CuO}$  化学成分比例较高; 为便于辨别各亚类, 命名为 Al-Ca-[z-K]亚类, 其中[z-K]代表  $\text{K}_2\text{O}$  化学成分比例为 0 (zero); 该亚类包括玻璃文物样本 7、21、27;

**亚类 3:** 由于其  $\text{Al}_2\text{O}_3$  和  $\text{CaO}$  化学成分比例中等,  $\text{K}_2\text{O}$  化学成分比例高,  $\text{MgO}$  化学成分比例较低, 且近似为 0; 为便于辨别各亚类, 命名为 Al-Ca-K 亚类; 该亚类包含玻璃文物样本 1、9、10、12;

**亚类 4:** 由于其  $\text{Al}_2\text{O}_3$  和  $\text{CaO}$  化学成分比例较少,  $\text{K}_2\text{O}$  化学成分比例中等, 为便于辨别各亚类, 命名为 Al-Ca-[m-K]亚类, 其中[m-K]代表  $\text{K}_2\text{O}$  化学成分比例中等(medium); 该亚类包含玻璃文物样本 3、18。

## 2) 对铅钡玻璃进行模型求解

同上述高钾玻璃求解过程, 人为设定类的个数为  $n = 3$ 。如图 8, 经聚类分析后, 每类占比分别为 45%、45%和 10%, 虽然  $n = 2$  时分布更均匀, 但其类个数与实际情况不符, 同时  $n = 3$  时较其他  $n > 3$  的类个数下分布更均匀, 比较如下复合条饼状图 12, 当  $n = 4$  时较  $n = 3$  时在 45%基础上继续进行划分, 故印证上述  $n = 3$  的合理性:



**Figure 8.** The proportion of cluster types when lead-barium glass is different  
**图 8.** 铅钡玻璃不同类时聚类种类占比

将聚类分析后结果如表 7 (每一聚类种类只罗列两个文物样本的部分化学成分比例, 所有结果详细数据见支撑材料):

**Table 7.** Some results of cluster analysis of lead-barium glass

**表 7.** 铅钡玻璃聚类分析部分结果

文物样本编号	聚类种类	$\text{SiO}_2$ (%)	$\text{PbO}$ (%)	$\text{BaO}$ (%)	$\text{CuO}$ (%)
23	1	53.79	16.98	11.86	2.99
25	1	50.61	31.90	6.65	1.12
11	2	73.34	12.66	11.11	3.31
19	2	64.72	21.36	4.07	2.36
8	3	43.97	14.31	23.76	6.99
20	3	37.36	9.30	23.55	4.78

对 40 条玻璃文物样本进行分析, 由于玻璃主要成分为  $\text{SiO}_2$ , 故同上述 2 不对  $\text{SiO}_2$  化学成分比例进行讨论。铅钡玻璃样本可分为三个亚类, 每个亚类仍根据其化学成分比例进行相应命名:

**亚类 1:** 由于其  $\text{PbO}$  化学成分比例高,  $\text{BaO}$  化学成分比例较低; 为便于辨别各亚类, 命名为  $\text{Pb-[l-Ba]}$  亚类, 其中  $[\text{l-Ba}]$  代表  $\text{BaO}$  化学成分比例较低(low); 该亚类包含玻璃文物样本 **23、25、30、39、40、41、42、43、45、46、47、49、50、51、52、54、55、57;**

**亚类 2:** 由于其  $\text{PbO}$  化学成分比例中等,  $\text{BaO}$  化学成分比例较低; 为便于辨别各亚类, 命名为  $[\text{m-Pb}]-[\text{l-Ba}]$  亚类, 其中  $[\text{m-Pb}]$  代表  $\text{PbO}$  化学成分比例中等,  $[\text{l-Ba}]$  代表  $\text{BaO}$  化学成分比例较低; 该亚类包含玻璃文物样本 **2、11、19、28、29、31、32、33、34、35、36、37、38、44、48、53、56、58;**

**亚类 3:** 由于其  $\text{PbO}$  化学成分比例中等,  $\text{BaO}$  化学成分比例高,  $\text{CuO}$  化学成分比例高; 为便于辨别各亚类, 命名为  $[\text{m-Pb}]-\text{Ba-Cu}$  亚类, 其中  $[\text{m-Pb}]$  代表  $\text{PbO}$  化学成分比例中等; 该亚类包含玻璃文物样本 **8、20、24、26。**

#### 4.2.6. 对分类结果进行合理性分析

对分类结果进行合理性分析时, 由于考虑到玻璃风化时一系列的物理化学变化, 故从数学角度与化学角度分别进行合理性分析, 分别采用**轮廓系数**和**物理化学变化**分析分类结果合理性。

##### 1) 理论意义分析——轮廓系数

在第二问第二小问中采用 K-means 聚类分析将高钾玻璃、铅钡玻璃进行了亚类划分, 亚类划分是否精确还需讨论分类结果的合理性, 故而引入**轮廓系数**这一概念, 轮廓系数是聚类分析性能的评估指标, 轮廓系数的取值范围为  $[-1, 1]$ , 取值越靠近 1 表明聚类效果越好, 与此同时, 取值越靠近 -1 表明聚类性能越差。

$a(i)$ : 样本与其所在簇内其他样本的平均距离;  $b(i)$ : 某个样本与其他簇样本的平均距离。则样本的轮廓系数  $s(i)$  为:

$$s(i) = \begin{cases} 1 - a(i)/b(i); & a(i) < b(i) \\ 0; & a(i) = b(i) \\ b(i)/a(i) - 1; & a(i) > b(i) \end{cases} \quad (26)$$

由上式, 通过计算可得出将高钾玻璃进行亚类划分的聚类分析轮廓系数  $s(i) = 0.83$ , 而对铅钡玻璃进行亚类划分的聚类分析轮廓系数  $s(i) = 0.75$ , 均在可接受范围内, 说明聚类分析的性能较好, 结果比较精确。

##### 2) 现实意义分析——物理化学变化

上述采用轮廓系数从理论角度对聚类分析的合理性进行分析, 与此同时, 还可以通过现实的化学意义, 即玻璃风化中进行的物理化学变化进行分析。

对于高钾玻璃进行聚类分析时分为四类, 很大程度上参考  $\text{Al}_2\text{O}_3$ 、 $\text{CaO}$ 、 $\text{K}_2\text{O}$  等物质的化学成分比例的高低, 该亚类划分具有现实意义。硅氧四面体  $[\text{SiO}_4]$  互相连接程度大, 化学稳定性好; 同时  $\text{Al}_2\text{O}_3$  所形成的  $[\text{AlO}_4]$  四面体结果对硅氧网络起到加固作用, 起到同样效果。而在  $\text{K}_2\text{O}$  析出后, 残留在表面的富硅薄膜, 形成了稳定的保护壳, 极大程度上提高了玻璃抗风化能力。上述均较好地印证了高钾玻璃四种亚类划分所具有的现实意义。

对于铅钡玻璃则是采用聚类分析分成三类, 在分类的过程中, 很大程度上参考  $\text{PbO}$  和  $\text{BaO}$  的化学成分比例, 该亚类划分同样具有现实意义,  $\text{BaO}$  所形成的  $\text{Ba}$  盐往往在最外层积聚形成夹心结构,  $\text{Pb}$  元素则是由外向内呈逐步减少的趋势, 通过这一规律便可判断铅钡玻璃的风化程度[7]。综合上述对高钾玻璃与铅钡玻璃分类现实意义的讨论, 进一步论证了聚类分析结构的合理性。

### 4.3. 问题三模型建立与求解

#### 4.3.1. 确定未知文物的类型

##### 1) 合理假设

经过观察表单 3 中可发现附件中的未知类别玻璃文物有风化和无风化之分, 故在分析问题三前需根据问题二结论将风化玻璃文物变成无风化玻璃文物统一求解。但问题在于第一问所求比值需事先清楚文物的类型才能预测无风化的指标, 而本问题文物类型变为所求量, 故针对问题四需要进行合理假设。

经过合理分析, 假设未知文物的类型均为高钾玻璃, 将风化文物指标  $gjf(i, j)$  转化成为  $gfw(i, j)$ 。根据问题一第三小问预测模型可得:

$$gfw(i, j) = BI_1(j) \times gjf(i, j), i = 1, 2, \dots, 8; j = 1, 2, \dots, 14 \quad (27)$$

##### 2) 判断是否符合假设

由问题二第一小问可得高钾玻璃化学成分与分类赋值的回归方程为:

$$y = 0.0003x_1 - 0.0003x_2 + \dots - 0.0002x_{14} + 0.9678 \quad (28)$$

将预测后的所有无风化文物的化学成分  $gfw'(i, j)$  分别代入该方程可得 8 个未知类别文物求得的分类赋值如表 8:

**Table 8.** Classification assignment of cultural relics of unknown category

**表 8.** 未知类别文物的分类赋值

未知类别文物编号	分类赋值	未知类别文物编号	分类赋值
A1	0.9952	A5	1.0072
A2	1.0000	A6	0.9979
A3	1.1731	A7	0.9967
A4	1.0843	A8	1.0696

若该文物为高钾玻璃类型, 则所求得的分类赋值应该在 **0.99 到 1** 之间, 因此可以判断出文物编号 A1、A6、A7 为高钾玻璃类型文物, 从而可以判断文物编号 A2、A3、A4、A5、A8 为铅钡玻璃类型文物。

##### 3) 准确性检验

在 2 中仅通过排除法判断出铅钡玻璃类型文物, 但不排除该方法存在一定误差, 故对于由排除法确定的铅钡玻璃类型文物再次运用上述方法, 依次进行假设、判断, 对结果进行准确性检验。检验结果如表 9:

**Table 9.** Classification assignments for cultural relics of unknown categories

**表 9.** 未知类别文物的分类赋值

未知类别文物编号	分类赋值	未知类别文物编号	分类赋值
A1	10.0218	A5	9.9735
A2	9.9731	A6	10.0389
A3	9.9750	A7	1.0035
A4	9.9808	A8	9.9461

如表 9 可以总结得出文物编号 A2、A3、A4、A5、A8 为铅钡玻璃类型文物，检验结果与合理假设所得结果完全一致，故采用此方法所得结果通过结果准确性检验，综上所述，文物编号 A1、A6、A7 为高钾玻璃类型文物；文物编号 A2、A3、A4、A5、A8 为铅钡玻璃类型文物。

#### 4.3.2. 按照亚类细分已知类型文物

为表述简便性，将高钾玻璃亚类 1、2、3、4 分别表示为高钾 A、B、C、D；将铅钡玻璃亚类 1、2、3 分别表示为铅钡 A、B、C。根据问题二第二小问的 K-means 聚类分析可知对于无风化高钾玻璃类，若进行亚类划分， $K_2O$ 、 $CaO$ 、 $Al_2O_3$  化学成分比例高，影响因素大故较为重要。在后文分析中，对于高钾玻璃只讨论  $K_2O$ 、 $CaO$ 、 $Al_2O_3$  这三个指标；同理，对于铅钡玻璃也只对  $PbO$  和  $BaO$  两个指标进行讨论。

##### 1) 数据震荡处理

以无风化高钾玻璃为例，由问题二第二小问可知，无风化高钾玻璃分为四种亚类，考虑到问题分析的简便性，对该四种亚类分为 A、B、C、D 类，分别对四种亚类赋值为 1、2、3、4，而后赋予数据 1% 的震荡保证回归过程可靠性。同理，对于无风化铅钡玻璃则采取对三种亚类 A、B、C 赋值为 1、2、3。

##### 2) 建立偏最小二乘回归方程

由于该问题本质与问题二第一小问相同，故仍沿用其偏最小二乘回归分析并利用第一问求解的化学指标进行分析。

对于高钾玻璃，由于仅考虑化学成分比例较高的  $K_2O$ 、 $CaO$ 、 $Al_2O_3$ ，故将化学成分指标改为 3 个，而后对所给分类赋值 1、2、3、4 分别进行偏最小二乘回归分析。由于问题二亚类划分中设定  $K_2O$  化学成分比例为 0 时自动将其划分为 B 类，故对于此类数据可直接筛选求解，故只需对分类赋值为 1、3、4 的亚类进行回归分析。

可建立如下三个方程：

$$\begin{aligned} y_1 &= 0.9718 + 0.0028x_1 - 0.0024x_2 - 0.0061x_3 \\ y_3 &= 3.0071 - 0.0106x_1 - 0.0076x_2 + 0.0274x_3 \\ y_4 &= 3.8277 + 0.0188x_1 - 0.0031x_2 + 0.0070x_3 \end{aligned} \quad (29)$$

同理，对于铅钡玻璃的三个亚类 A、B、C 也进行偏最小二乘回归分析，将化学成分指标改为 2 个，对分类赋值 1、2、3 分别进行回归分析，利用偏最小二乘回归分析建立如下三个方程：

$$\begin{aligned} y_5 &= 0.9494 - 0.0002x_1 + 0.0020x_2 \\ y_6 &= 1.9865 - 0.0011x_1 - 0.0032x_2 \\ y_7 &= 2.6514 - 0.0025x_1 + 0.0140x_2 \end{aligned} \quad (30)$$

##### 3) 判断所属亚类

对高钾玻璃类型文物，进行亚类划分后，观察数据可发现文物编号 A1 的  $K_2O$  化学成分比例为 0，故可直接将 A1 划分为 B 类，因此只需讨论 A6、A7 样本，将二者的三个指标带入公式(29)，求出其回归后得到的分类赋值，如表 10：

**Table 10.** Classification assignment of samples of high potassium glass A6 and A7 after regression

**表 10.** 高钾玻璃 A6、A7 样本回归后的分类赋值

文物样本编号	A	C	D
A6	0.9952	2.8741	4.2147
A7	0.9271	3.2515	4.0019

由表 10 数据可得, A1 划分为 B 类, 而 A6 和 A7 划分为 A 类, 未出现 C 类。

同理, 对铅钡玻璃类型文物, 进行亚类划分后, 也可使用该方法, 将两种指标带入公式(30), 如表 11:

**Table 11.** Classification assignment of lead-barium glass samples after regression

**表 11.** 铅钡玻璃样本回归后的分类赋值

文物样本编号	A	B	C
A2	0.9451	1.9671	2.6076
A3	0.9488	1.9271	2.6160
A4	0.9596	1.9328	2.7054
A5	1.0001	1.9743	2.6586
A8	0.8873	1.9266	2.9555

由表 11 可知, 铅钡玻璃的 A 类有 A2、A3、A4, B 类有 A5, C 类有 A8。综上所述, 分类如表 12:

**Table 12.** Summary of types and subcategories of unknown glass cultural relics

**表 12.** 未知类别玻璃文物所属类型与亚类汇总

文物样本	编号 A1	编号 A2	编号 A3	编号 A4
分类	高钾玻璃 B	铅钡玻璃 A	铅钡玻璃 A	铅钡玻璃 A
文物样本	编号 A5	编号 A6	编号 A7	编号 A8
分类	铅钡玻璃 B	高钾玻璃 A	高钾玻璃 A	铅钡玻璃 C

## 6. 模型的评价与推广

### 6.1. 模型的评价

#### 6.1.1. 模型优点分析

1) 整篇文章针对不同的问题采用多种模型进行求解, 并不同程度上检验改进了这些模型, 大大增强了结果的准确性。

2) 每一个问题的问题分析都有流程图配合阐述, 条理清晰, 逻辑鲜明。

3) 问题一中, 采用卡方检验与 Spearman 相关系数两种方法对玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系进行分析, 最后通过两种方法对比择优得出最终的结论。

4) 问题二中, 采用理论意义与现实物理化学意义相结合的方式对两种玻璃的亚类划分结果的合理性与敏感性。

5) 问题三中, 延用了第一问与第二问的结论, 使得文章连续性较好。利用偏最小二乘回归分析避免了使用机器学习需要大量数据样本的局限性。

#### 6.1.2. 模型缺点分析

在问题二、三中的分类赋值仍为人工赋值, 具有一定的主观性, 仅仅对本题的结论无影响, 若样本量增加或减少, 准确率将受到影响。

## 6.2. 模型的推广

1) 问题二对玻璃的化学成分与玻璃分类规律进行分析时构建了偏最小二乘回归模型, 该模型具有准确率高, 运行速度快等优势, 而当给定数据量增大时, 最后得到的结果将更加精确, 对于问题三这种对未知类别文物进行分类的问题时也会更加精确与高效, 可运用到日后现实中的实际问题里。

2) 在对两种玻璃进行亚类分析时结合理论意义与现实物理化学意义, 在给定更多样本后可对该模型进一步推广, 得到更多的亚类划分, 并保证每一种亚类有更加精确的现实物理化学意义。

## 参考文献

- [1] 史美光, 何欧里, 周福征. 一批中国汉墓出土钾玻璃的研究[J]. 硅酸盐学报, 1986(3): 307-313.
- [2] 张丽艳, 李洪, 陈树彬, 李忠镛, 阮莴秩, 薛天锋, 钱敏, 凡思军. 玻璃的成分和性质的模拟方法[J]. 硅酸盐学报, 2022, 50(8): 2338-2350.
- [3] 斯琴毕力格, 李青会, 干福熹. 激光剥蚀-电感耦合等离子体-原子发射光谱/质谱法分析中国古代钾玻璃组分[J]. 分析化学, 2013, 41(9): 1328-1333.
- [4] 马清林, 张治国, 大卫·斯科特, 海因兹·贝克, 戴特莱夫·昆特. 中国战国时期八棱柱状费昂斯制品成分及结构研究[J]. 中国国家博物馆馆刊, 2012(12): 112-132.
- [5] 许成顺, 豆鹏飞, 高岫成, 陈苏, 杜修力. 地震动持时压缩比对可液化地基地震反应影响的振动台试验[J]. 岩土力学, 2019, 40(1): 147-155.
- [6] Zhang, B., Li, Y.H., Li, Q.H., Ma, B. and Yang, F.J. (2004) Non-Destructive Analysis of Early Glass Unearthed in South China by External-Beam Pixe. *Journal of Radioanalytical & Nuclear Chemistry*, **261**, 387-392. <https://doi.org/10.1023/B:JRNC.0000034875.65356.72>
- [7] 王婕, 李沫, 马清林, 张治国, 章梅芳, 王菊琳. 一件战国时期八棱柱状铅钡玻璃器的风化研究[J]. 玻璃与搪瓷, 2014, 42(2): 6-13.