

# 基于成分分析的古代玻璃制品分析与鉴别模型

## 摘 要

一带一路建设推动了丝绸之路的历史文化探索，在古代中西方文化交流中，玻璃制品是一个重要的文化标志，具有深刻的研究意义。本文通过建立预测，分类，分析三个模型对玻璃制品的种类与化学成分深度分析。

对于问题一，使用数据挖掘分析玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系。利用 **K-means 聚类分析** 方法，将玻璃文物分为高钾 2 类和铅钡 3 类，且分类结果与风化情况对应关系良好。根据聚类中心以及化学成分统计规律利用**集合映射**将风化点映射至无风化点。

对于问题二，需调研各项影响指标并结合无机化学相关知识初步确定亚类划分影响因素，在此基础上采用 K-means 聚类方法进行定性分类分析。根据分类结果进行判别分析分别得到高钾和铅钡的定量分类判别函数。综合判别函数和亚类特征将高钾类型分成三个亚类  $A_1A_2A_3$ ，铅钡类型分成三个亚类  $B_2B_3B_4$ ，对  $B_4$  亚类进一步划分，提取出  $B_1$ ， $B_5$  两个小亚类，完成亚类的划分与定义。基于 F1-score 得分和判别函数微扰状态下的相对变化量对整体模型进行**合理性和敏感性分析**，F1-score 得分均高于 90，且判别函数的相对变化量位于区间  $[0.0014, 0.004]$  内，证明模型具有良好的稳定性。

对于问题三，利用问题一分类和预测模型将风化点转化为无风化点，并利用问题二判别函数进行亚类划分，分类结果均契合所定义亚类。基于判别函数最大变化量与原始量的比值完成敏感性分析，比值位于区间  $[0.0012, 0.0079]$  内，证明此分类过程具有强稳定性。

对于问题四，采用**相关性分析**得到 Pearson 相关可视化图，并分别分析得到高钾类型和铅钡类型各化学成分之间的关系。在此基础上采用**灰色关联法**，通过变量间关联度反映不同类别间化学成分相关性差异，其中二氧化硅，氧化钾，氧化铅，氧化钡指标存在明显相关性差异。

本文建立一个完整的分类，预测，分析体系，能很好地处理风化状态并且根据化学成分进行分类和相关性研究。

**关键词：**古代玻璃制品，K-means 聚类，判别分析，灰色关联法，相关性分析

## 一、问题重述

### 1.1 问题背景

丝绸之路是古代中西方文化交流的通道，其中玻璃是早期贸易往来的宝贵物证。早期的玻璃在西亚和埃及地区常被制作成珠形饰品传入我国，我国古代玻璃吸收其技术后在本地就地取材制作，因此与外来的玻璃制品外观相似，但化学成分却不相同。玻璃的主要原料是石英砂，主要化学成分是二氧化硅（ $\text{SiO}_2$ ）。由于纯石英砂的熔点较高，为了降低熔化温度，在炼制时需要添加助熔剂。由于助熔剂的种类不同，烧制出的玻璃类型也存在差异，主要常用的有高钾玻璃和铅钡玻璃。此外，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。因此，建立古代玻璃制品的成分分析与鉴别模型，排除表面风化带来的不利影响，正确完成玻璃文物成分分析、类型划分、鉴别保护工作显得尤为关键且必要。

### 1.2 问题重述

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。附件表单 1 给出了这些文物的分类信息，附件表单 2 给出了相应的主要成分所占比例（空白处表示未检测到该成分）。这些数据的特点是成分性，即各成分比例的累加和应为 100%，但因检测手段等原因可能导致其成分比例的累加和非 100% 的情况。本题中将成分比例累加和介于 85%~105% 之间的数据视为有效数据。

请你们团队依据附件中的相关数据进行分析建模，解决以下问题：

（1）对这些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；结合玻璃的类型，分析文物样品表面有无风化化学成分含量的统计规律，并根据风化点检测数据，预测其风化前的化学成分含量。

（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）对附件表单 3 中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

## 二、问题分析

### 2.1 问题一分析

问题一本质上是挖掘不同数据间所隐含的内部联系的数据分析问题，在此问题的基础上，进一步延伸了与前面数据问题高度相关的利用风化点的数据预测风化前的数据的预测问题。数据分析过程中需先分别对高钾玻璃和铅钡玻璃两种不同的玻璃类型其是否“表面风化”进行统计，再在每一类型中纹饰和颜色与文物“表面风化”与否进行分析，利用数据图可视化数据，观察其背后所隐藏的逻辑，做出判断。此外还需要分别去分析不同种类的玻璃其表面风化与化学成分含量的统计规律，首先需要从数据本身出发，进行数据可视化，通过观察图表对数据有一个认识，并根据图表进行合理的变化规律的归纳。在此基础上，利用隐去类型的数据进行 K-means 聚类分析，得到不同化学成分之间的统计规律。最后利用得到的统计规律，预测其风化前的化学成分含量。

2.2 问题二分析

问题二依据题意可知，本问题属于分类问题，分类依据主要聚焦于附件数据所给出的各种化学成分含量。故本文先根据第一问最后所得到的无风化后数据分析高钾玻璃和铅钡玻璃的分类规律，再通过聚类分析、鉴别分析和查阅文献等方式进行进一步的亚类划分，并最终对划分结果进行合理性以及敏感性分析。

2.3 问题三分析

问题三本质上是问题二结论的应用，利用问题二的亚类划分结果，鉴别附件三未知文物的种类，并最终对结果进行敏感性分析。利用问题一中聚类模型对判断未知文物属于高钾和铅钡有无风化类型进行区分，然后利用问题一的统计规律将风化类型的文物的化学成分含量预测出无风化类型的文物化学成分含量，最后利用问题二的亚类划分规律，完成种类鉴定。

2.4 问题四分析

依据题意可知，本问题属于相关性问題，分析同类型下不同化学成分之间的关联性以及不同类型间化学成分关联关系的差异性。先确定类型种类，选择分类粒度最大的类别划分，即分为高钾类型和铅钡类型，再利用附件数据分别对两种类型计算相同类型下不同化学成分之间的相关性，分析不同化学成分之间的关联性，之后对两种类型间化学成分关联关系差异性进行分析。

三、模型假设

- 1. 假设实验数据准确无误。
- 2. 假设文物采样点的风化属性与相应文物整体风化属性一致。
- 3. 假设不同种类的玻璃文物在风化过程中所处环境的温、湿度等气候条件基本相同。
- 4. 假设任意已风化的文物（检测点）在风化前都有一未风化的状态与之对应。
- 5. 假设相同类型的玻璃文物具有相似的化学成分组成和风化属性。

四、 符号说明

符号	说明	单位
F	两个均方的比值	
P	用来判定假设检验结果的一个参数	
A <sub>1</sub>	含钠类型(Na <sub>2</sub> O-K <sub>2</sub> O)	
A <sub>2</sub>	无钾类型	
A <sub>3</sub>	正常类型	
B <sub>1</sub>	高铅类型(PbO-SiO <sub>2</sub> )	
B <sub>2</sub>	高钡类型(SiO <sub>2</sub> -BaO)	
B <sub>3</sub>	含钠类型(Na <sub>2</sub> O-PbO-BaO)	
B <sub>4</sub>	正常类型(PbO-SiO <sub>2</sub> -BaO)	
B <sub>5</sub>	含硫类型	
F1-score	F1-score 值是指正确率与召回率的加权综合指标	%

## 五、模型建立与求解

### 5.1 基于聚类分析的化学成分含量的预测模型

#### 5.1.1 问题分析

分析文物“表面风化”与玻璃类型、纹饰和颜色的关系：

首先对于高钾玻璃和铅钡玻璃两种不同的玻璃类型，分别对其是否“表面风化”进行统计，结果见下图 1-1。从中可以看出高钾玻璃不容易发生风化，但其风化后会产生较大变化；铅钡玻璃容易发生风化，风化给其带来的变化较小。

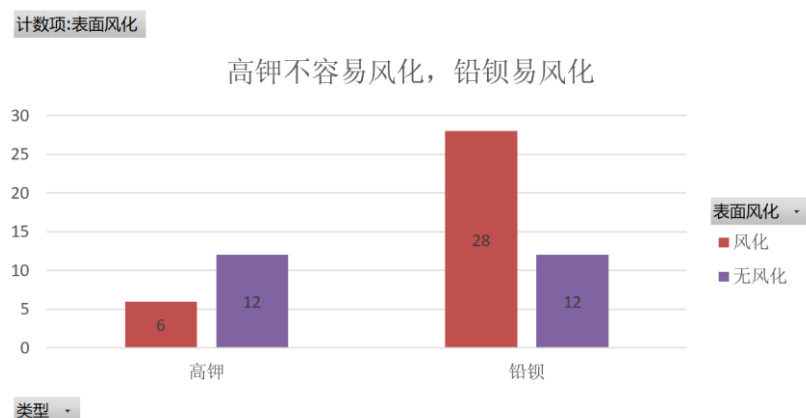


图 1-1 高钾、铅钡玻璃风化数量统计图

对于高钾玻璃，其整体颜色均分布于蓝绿附近，其中蓝绿色占大多数，浅蓝次之，深绿、深蓝较少，具体所占百分比见下图 1-2。其纹饰种类齐全，A、B、C 类纹均有分布，从图 1-3 中可以看出，具有 A、C 类纹的玻璃均不发生风化，其颜色分布较为多样；具有 B 类纹的玻璃全部发生风化且其颜色只有蓝绿一种，较为单一。

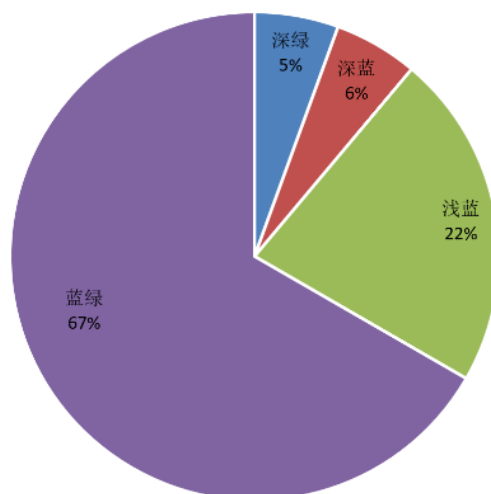


图 1-2 高钾玻璃不同颜色分布比例图

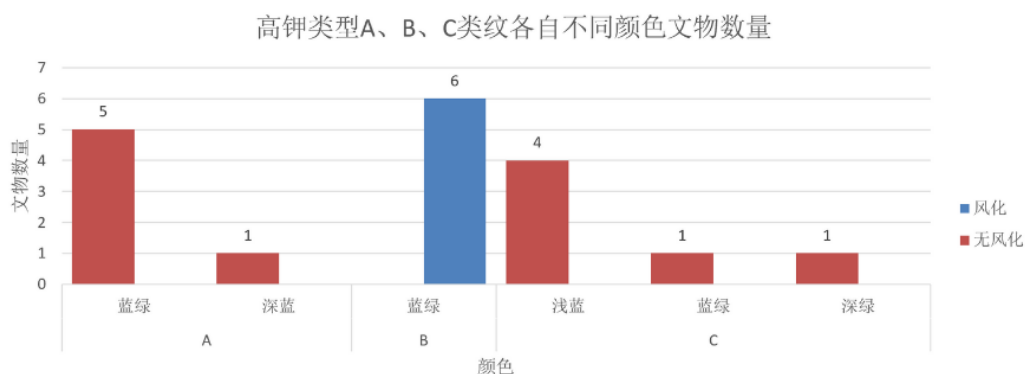


图 1-3 高钾玻璃 A、B、C 类纹各自不同颜色文物数量及风化情况统计图

对于铅钡玻璃，其颜色种类繁多，浅蓝色居多，深绿、紫次之，蓝绿、浅绿、黑、绿、深蓝较少，具体所占百分比见下图 1-4。关于不同颜色铅钡玻璃风化情况，黑色、蓝绿色玻璃全部风化，其余颜色风化规律不尽相同，具体风化情况见图 1-5，总体上其颜色与“表面风化”间无明显关系。在其纹饰中，只有 A、C 两类纹，两类纹中均存在风化的现象，两者无明显区别。

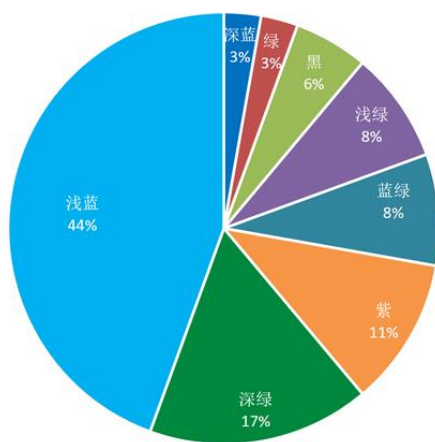


图 1-4 铅钡玻璃不同颜色分布比例图

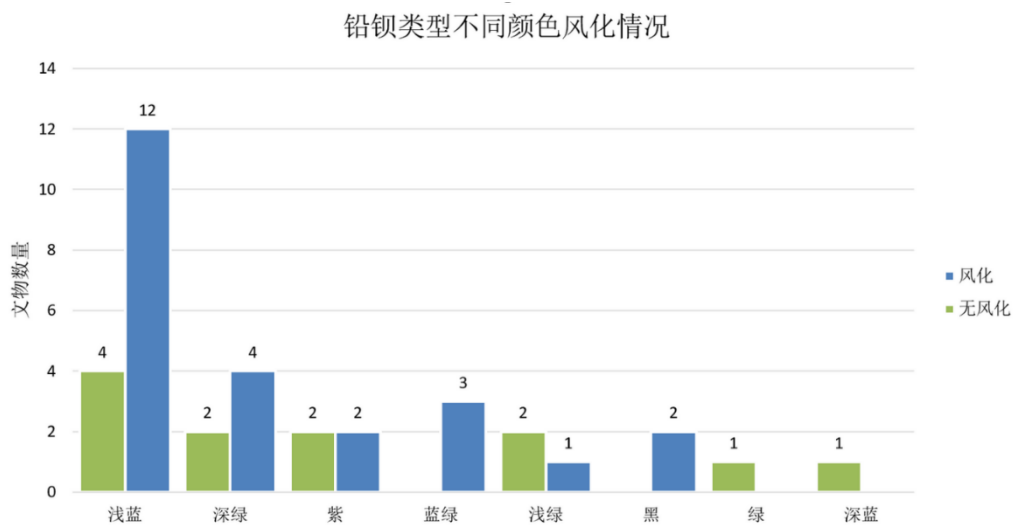


图 1-5 铅钡玻璃不同颜色风化情况统计图

5. 1. 2 模型建立

分析文物表面有无风化的化学成分含量的统计规律：

首先对于高钾玻璃，其是以含钾量高的物质如草木灰作为助燃剂烧制而成，故其氧化钾（ $K_2O$ ）含量较高。由之前统计结果可知发生风化的均是具有 B 类纹的玻璃，在从无风化到有风化的过程中，发生明显变化的化学成分有：氧化钾（ $K_2O$ ）含量减少、氧化钙（ $CaO$ ）含量减少、二氧化硅（ $SiO_2$ ）含量增加，具体含量变化情况见图 1-6；存在小幅变化的化学成分有：氧化镁（ $MgO$ ）含量减少、氧化铝（ $Al_2O_3$ ）含量减少、氧化铁（ $Fe_2O_3$ ）含量减少，具体含量变化情况见图 1-7；其余化学成分变化可忽略不计。

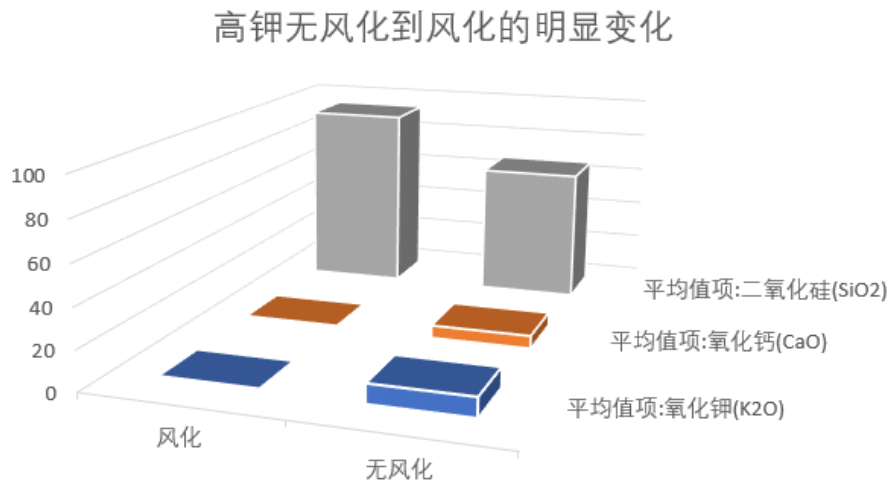


图 1-6 高钾玻璃化学成分含量变化情况图（明显变化）

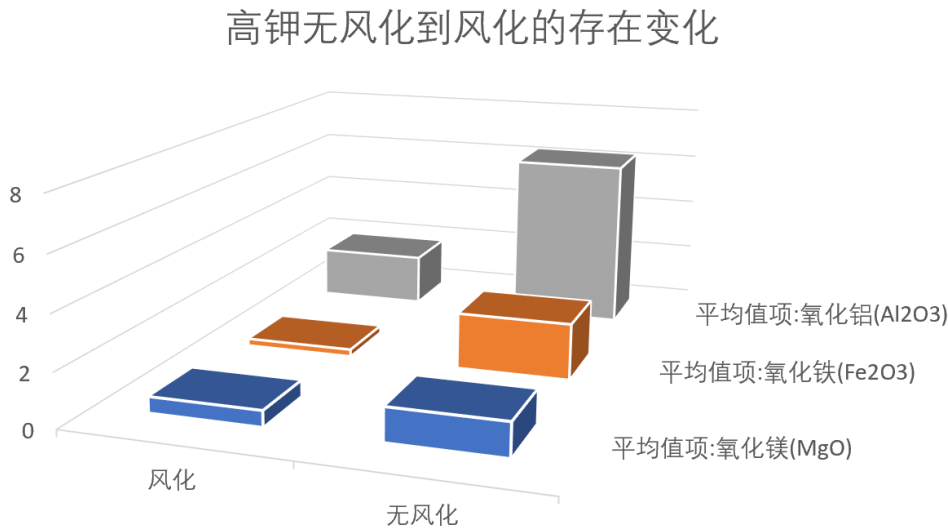


图 1-7 高钾玻璃化学成分含量变化情况图（小幅变化）

对于铅钡玻璃，其在烧制过程中加入铅矿石作为助熔剂，其氧化铅（ $PbO$ ）、氧化钡（ $BaO$ ）的含量较高，同时有稳定剂氧化钙（ $CaO$ ），此外可以较明显的看出紫色玻璃对应钡含量高于其他颜色的玻璃。在从无风化到有风化的过程中，发生明显变化的化学成分有：氧化铅（ $PbO$ ）含量增加、二氧化硅（ $SiO_2$ ）含量减少，具体含量变化情况见图 1-8；存在小幅变化的化学成分有：氧化钡

( $BaO$ ) 含量减少、氧化磷 ( $P_2O_5$ ) 含量增加、氧化铁 ( $Fe_2O_3$ ) 含量减少、氧化钙 ( $CaO$ ) 含量增加，具体含量变化情况见图 1-9；其余化学成分变化可忽略不计。

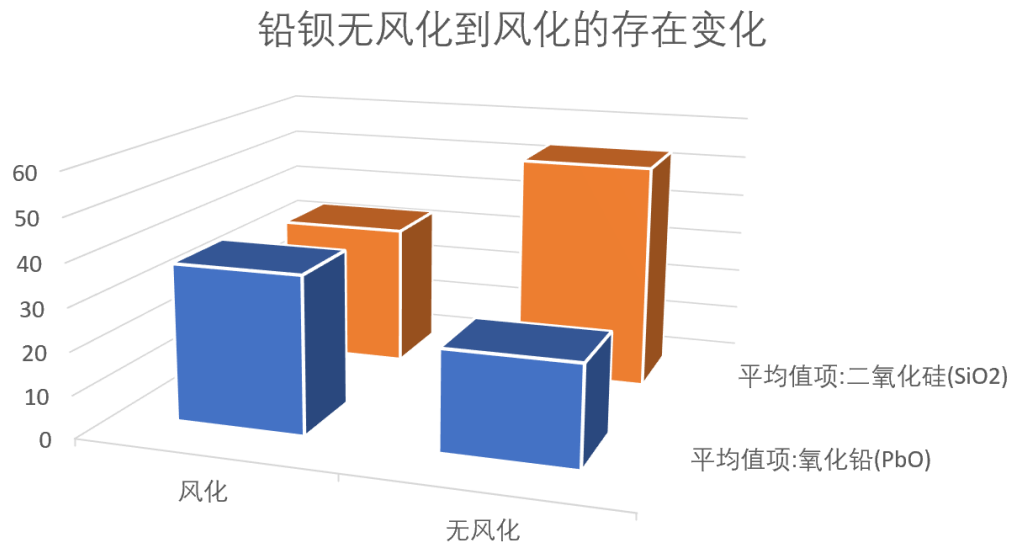


图 1-8 铅钡玻璃化学成分含量变化情况图（明显变化）

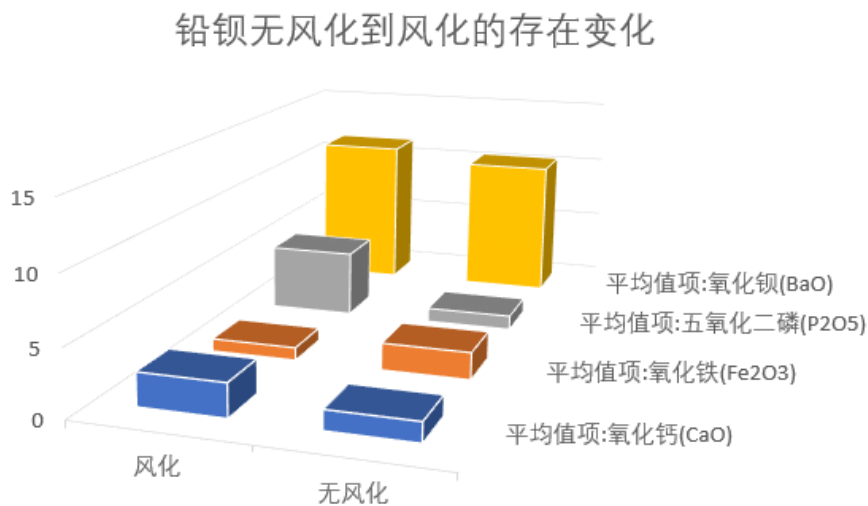


图 1-9 铅钡玻璃化学成分含量变化情况图（小幅变化）

分析其风化前后化学成分含量变化原因：对于高钾玻璃，在风化过程中氧化钾会与二氧化硅发生反应，导致高钾相对比例减少；对于铅钡玻璃，烧制过程中残余的铅元素进一步氧化，同时二氧化硅发生一系列化学反应，导致氧化铅相对比例增加。同时对于两种玻璃，在风化过程中，内部元素与环境元素进行大量交换、风化层与未风化区域之间发生物质交换，这些都会导致其化学成分含量发生变化。

在总结完风化前后各化学成分含量的变化规律后，我们利用聚类分析的方法研究其统计规律。由于表单中所给化学成分指标种类较多，且有些指标平均值近似相等且在风化过程中变化趋势相近，我们可以绘制出其各成分含量平均值变化曲线，根据其在坐标平面内的空间分布，选出几个最具代表性的指标（含量达、数据齐全）作为判断一文物是否（严重）风化的依据。随后根据所

选取的指标，将表单中所给各文物（采样点）进行聚类，划分为无风化、风化、严重风化等若干类，最后得到各类别各化学成分含量的统计规律。

对于高钾玻璃，绘制出其各化学成分含量平均值变化曲线如图 1-10 所示，从上往下依次选取氧化铜（ $CuO$ ）、氧化铝（ $Al_2O_3$ ）、氧化钙（ $CaO$ ）、氧化钾（ $K_2O$ ）、二氧化硅（ $SiO_2$ ）五个指标作为判断依据，对所给文物（采样点）进行聚类分析，利用 SPSSAU 软件进行计算，划分为风化、无风化两类，聚类类别汇总见图 1-11，聚类项（指标）重要性对比图见图 1-12，其统计规律整体上与变化规律相符，各指标对应取值范围见表 1。

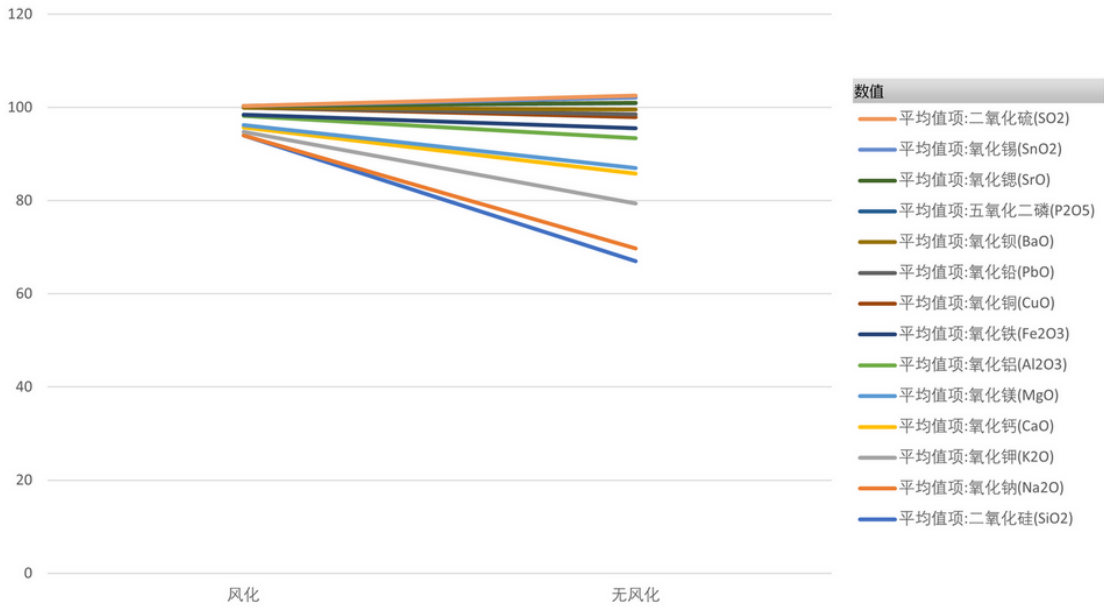


图 1-10 高钾玻璃各化学成分含量平均值变化曲线图

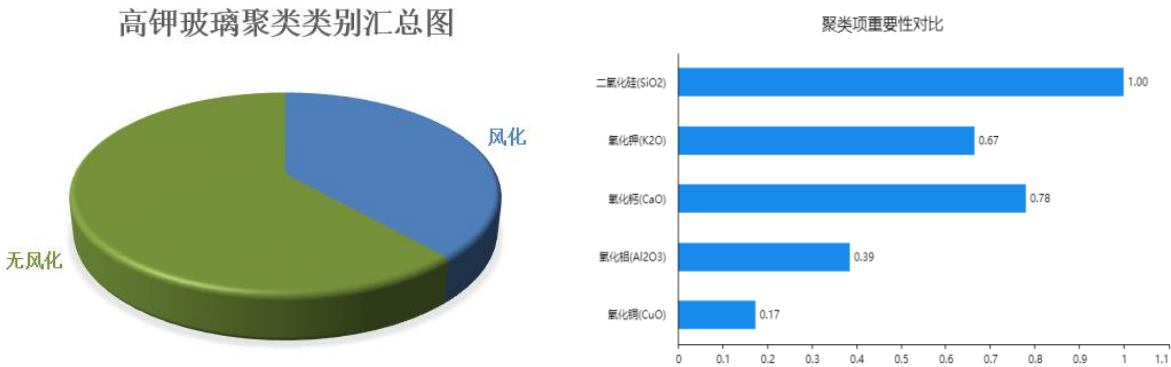


图 1-11 高钾玻璃聚类类别汇总图

图 1-12 高钾玻璃聚类项（指标）重要性对比图

表 1-1 高钾玻璃聚类各指标取值范围表

指标	聚类类别方差分析差异对比结果（平均值±标准差）	
	风化（n=5）	无风化（n=8）
二氧化硅（ $SiO_2$ ）	93.10±3.73	63.12±3.44
氧化钾（ $K_2O$ ）	1.69±1.96	11.25±2.12
氧化钙（ $CaO$ ）	1.04±0.76	7.16±1.21
氧化铝（ $Al_2O_3$ ）	2.23±1.45	6.87±1.99
氧化铜（ $CuO$ ）	1.07±0.49	2.86±1.67



对于铅钡玻璃，绘制出其各化学成分含量平均值变化曲线如图 1-13 所示，从上往下依次选取氧化锶（ $SrO$ ）、氧化钡（ $BaO$ ）、氧化铅（ $PbO$ ）、二氧化硅（ $SiO_2$ ）四个指标作为判断依据，对所给文物（采样点）进行聚类分析，利用 SPSSAU 软件进行计算，划分为风化、严重风化、无风化三类，聚类类别汇总见图 1-14，聚类项（指标）重要性对比图见图 1-15，其统计规律整体上与变化规律相符，各指标对应取值范围见表 2。

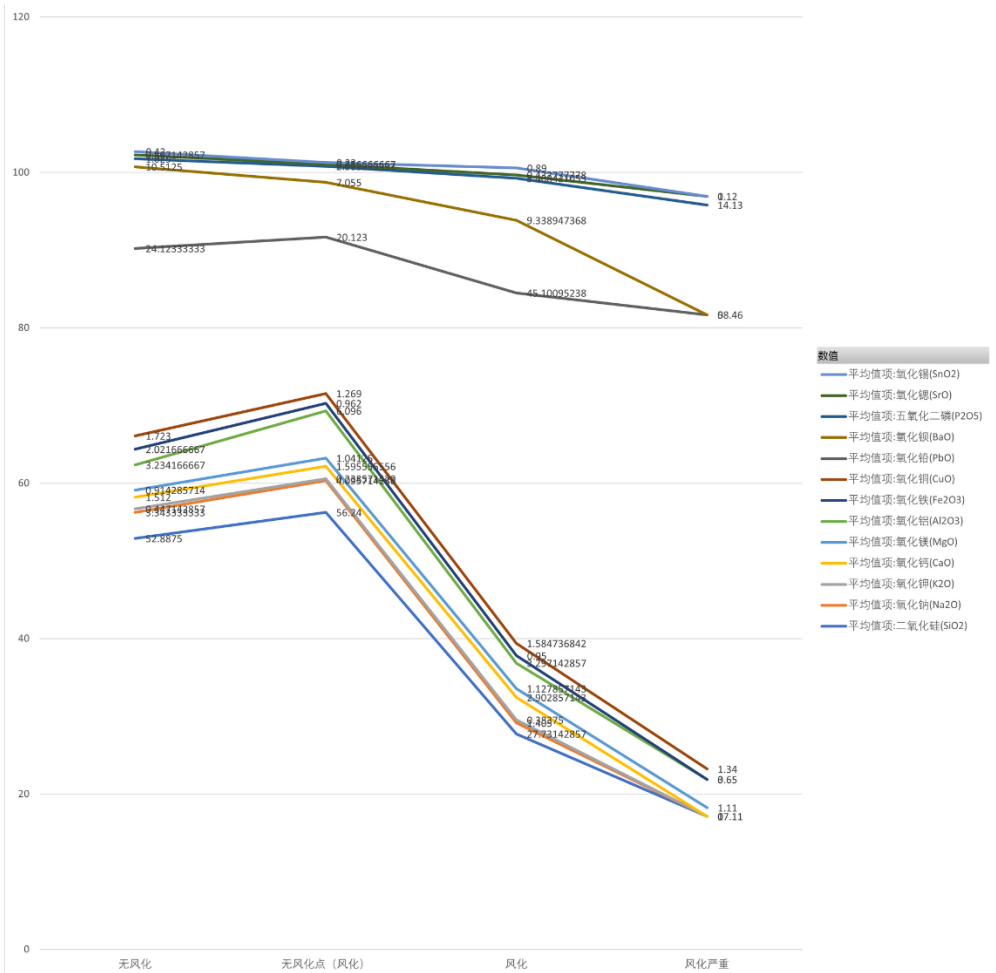


图 1-13 铅钡玻璃各化学成分含量平均值变化曲线图

铅钡玻璃聚类类别汇总图

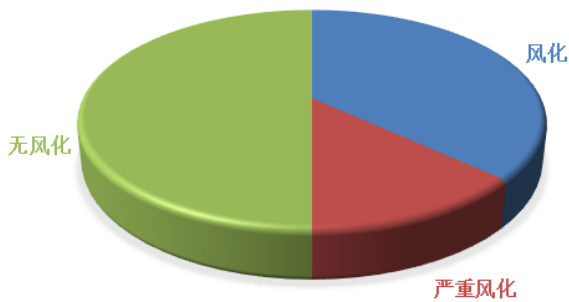


图 1-14 铅钡玻璃聚类类别汇总图

聚类项重要性对比

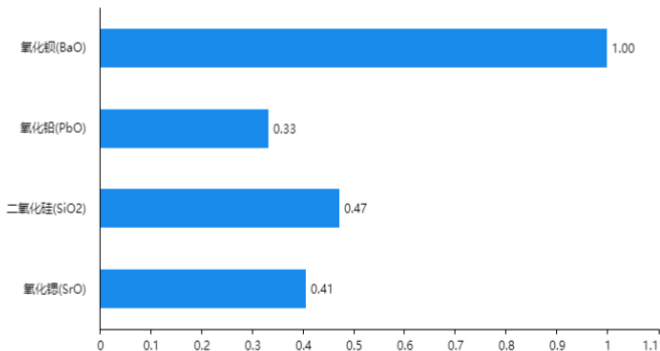


图 1-15 铅钡玻璃聚类项（指标）重要性对比图

表 1-2 铅钡玻璃聚类各指标取值范围表

指标	聚类类别方差分析差异对比结果（平均值±标准差）		
	风化（n=14）	严重风化（n=5）	无风化（n=19）
氧化钡（ $BaO$ ）	8.49±2.55	31.16±3.32	7.91±3.36
氧化铅（ $PbO$ ）	46.64±11.84	29.94±1.47	24.61±11.05
二氧化硅（ $SiO_2$ ）	26.48±10.08	16.04±11.90	51.36±12.32
氧化锶（ $SrO$ ）	0.56±0.17	0.58±0.21	0.27±0.07

### 5.1.3 模型求解

根据风化点检测数据，预测其风化前的化学成分含量：

由之前求得的变化规律和统计规律可知，各化学成分含量风化前后数值均可进行一定精度的双向推导。结合聚类分析的原理以及集合映射思想，可以利用公式（1）对已知风化后具体数值及聚类标准差的指标（聚类分析过程中选取的主指标）进行风化前含量预测。其中 $x_1$ 表示所求风化前化学成分含量预测值， $x_2$ 表示附件表单中所给各采样点对应化学成分含量检测值， $x_{C_1}$ 表示聚类分析中

求得的该指标风化前（无风化）聚类中心值， $x_{C_2}$ 表示聚类分析中求得的该指标风化后（风化）聚类中心值， $\sigma_1$ 表示聚类分析中求得的该指标风化前（无风化）对应的标准差， $\sigma_2$ 表示聚类分析中求得的该指标风化后（风化）对应的标准差。

$$x_1 = \frac{(x_2 - x_{C_2}) \sigma_1}{\sigma_2} + x_{C_1} \quad (1)$$

对于非主指标，由于其变化趋势与相近的主指标基本相同，可以利用等比例放缩的方法将其对应主指标的聚类中心值、标准差一一映射到非主指标上，所用公式（2）如下， $x_{C_1}'$ 表示所求非指标风化前（无风化）聚类中心值， $x_{C_2}'$ 表示所求非指标风化后（风化）聚类中心值， $\sigma_1'$ 表示所求非主指标风化前（无风化）对应的标准差， $\sigma_2'$ 表示所求非主指标风化后（风化）对应的标准差， $\bar{x}_1$ 、 $\bar{x}_2$ 分别表示对应主指标风化前后含量平均值， $\bar{x}_1'$ 、 $\bar{x}_2'$ 分别表示所求非主指标风化前后含量平均值。在映射完成后，便可利用公式（1）进行计算，得到对应的风化前成分含量预测值。

$$\begin{cases} \begin{pmatrix} x_{C_1}' \\ \sigma_1' \end{pmatrix} = \frac{\bar{x}_1'}{\bar{x}_1} \begin{pmatrix} x_{C_1} \\ \sigma_1 \end{pmatrix} \\ \begin{pmatrix} x_{C_2}' \\ \sigma_2' \end{pmatrix} = \frac{\bar{x}_2'}{\bar{x}_2} \begin{pmatrix} x_{C_2} \\ \sigma_2 \end{pmatrix} \end{cases} \quad (2)$$

经过计算，得到全部化学成分对应风化前预测值，部分结果见图 1-16，全部具体结果见附录 1。通过观察结果，可以发现一些普遍性规律：对于高钾玻璃，从风化到无风化，二氧化硅（ $SiO_2$ ）占比减少到 60%左右，氧化钾（ $K_2O$ ）、

氧化铝 ( $Al_2O_3$ ) 占比减少到 2%左右, 氧化钙 ( $CaO$ )、氧化铜 ( $CuO$ ) 占比减少到 1%左右; 对于铅钡玻璃, 从严重风化到无风化, 二氧化硅 ( $SiO_2$ ) 占比明显增加到 50%左右, 氧化钡 ( $BaO$ ) 占比从 31%下降到 8%左右, 氧化铅 ( $PbO$ )、氧化锶 ( $SrO$ ) 均有部分减少但不明显; 从风化到无风化, 二氧化硅 ( $SiO_2$ ) 占比同样增加到 50%左右, 氧化钡 ( $BaO$ )、氧化锶 ( $SrO$ ) 占比无明显变化, 氧化铅 ( $PbO$ ) 占比减半。

文物采样点	纹饰	类型	颜色	二氧化硅 ( $SiO_2$ )	氧化钠 ( $Na_2O$ )	氧化钾 ( $K_2O$ )	氧化钙 ( $CaO$ )	氧化镁 ( $MgO$ )	氧化铝 ( $Al_2O_3$ )	氧化铁 ( $Fe_2O_3$ )	氧化铜 ( $CuO$ )	氧化铅 ( $PbO$ )	氧化钡 ( $BaO$ )	五氧化二磷 ( $P_2O_5$ )	氧化锶 ( $SrO$ )	氧化锡 ( $SnO_2$ )	二氧化硫 ( $SO_2$ )
22	B	高钾	蓝绿	62.43	0	10.22	8.15	1.35	8.61	2.42	1.09	0	0	1.38	0	0	0
05	A	高钾	蓝绿	61.58	0	10.95	7.35	1.77	7.5	2.62	3.27	0	0	0.94	0.06	0	0.47
16	C	高钾	浅蓝	65.18	2.1	14.52	8.27	0.52	6.18	0.42	1.07	0.11	0	0	0.04	0	0
37	C	铅钡	深绿	60.12	0	0.23	0.89	0	2.72	0	3.01	17.24	10.34	1.46	0.31	0	3.66
45	A	铅钡	浅蓝	61.28	2.66	0.11	0.84	0.74	5	0	0.53	15.99	10.96	0	0.23	0	0
52	C	铅钡	浅蓝	50.46	1.34	0	2.64	1.07	2.95	0.9	2.31	25.34	8.11	0.47	0.22	0	0
58	C	铅钡		56.14	0	0.43	3.12	1.27	3.48	1.07	2.73	17.81	6.82	0.74	0.14	0	0
51部位1	C	铅钡	浅蓝	49.07	0	0	2.53	1.03	2.82	0.86	2.21	18.64	8.5	0.67	0.2	0.23	0

图 1-16 风化前各化学成分含量预测值（部分）

5.2 基于玻璃文物化学成分含量的聚类判别分类模型

5.2.1 问题分析

依据题意可知，本问题属于分类问题，分类依据主要聚焦于附件数据所给出的各种化学成分含量。故本文先根据第一问最后所得到的无风化后数据分析高钾玻璃和铅钡玻璃的分类规律，再通过聚类分析、鉴别分析和查阅文献等方式进行进一步的亚类划分，并最终对划分结果进行合理性以及敏感性分析。

从高钾类型与铅钡类型的命名可以推测，这两种不同类型的划分应与他们命名的化学成分具有相关关系。通过观察附件数据，将高钾类型与铅钡类型的氧化钾的化学成分含量进行均值比较，比较结果如图 2-1 所示，发现高钾类型的氧化钾含量远远高于铅钡类型，故可以初步认为，氧化钾是高钾类型与铅钡类型分类的主要分类依据之一。

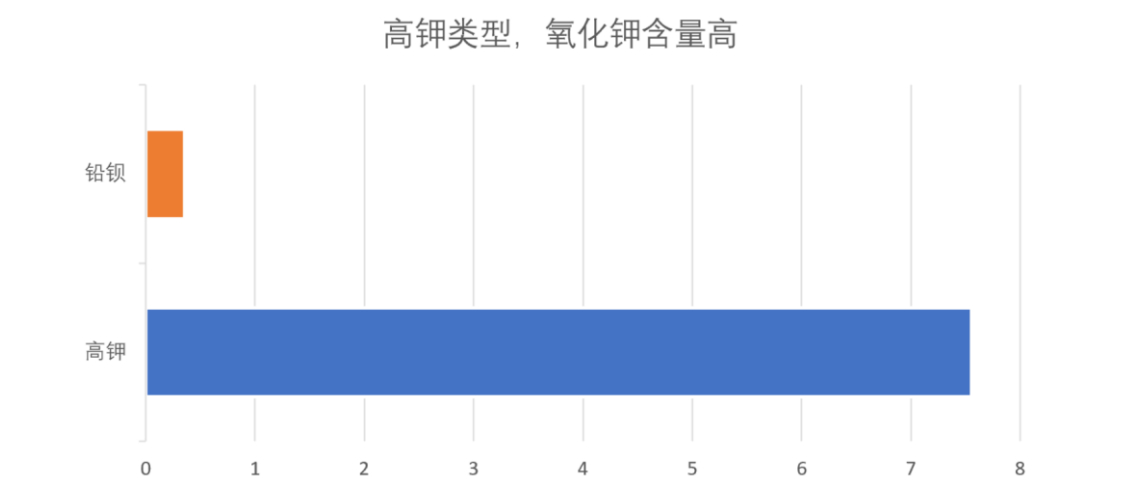


图 2-1 高钾类型与铅钡类型的氧化钾的化学成分含量比较

再将高钾类型与铅钡类型的氧化铅与氧化钡的化学成分含量进行均值比较，比较结果如图 2-2 所示，有图可以发现，氧化铅与氧化钡的化学成分含量，在两种不同类型之间差异明显，故可以初步认为，氧化铅与氧化钡是这两种类型的分类的主要依据之一。

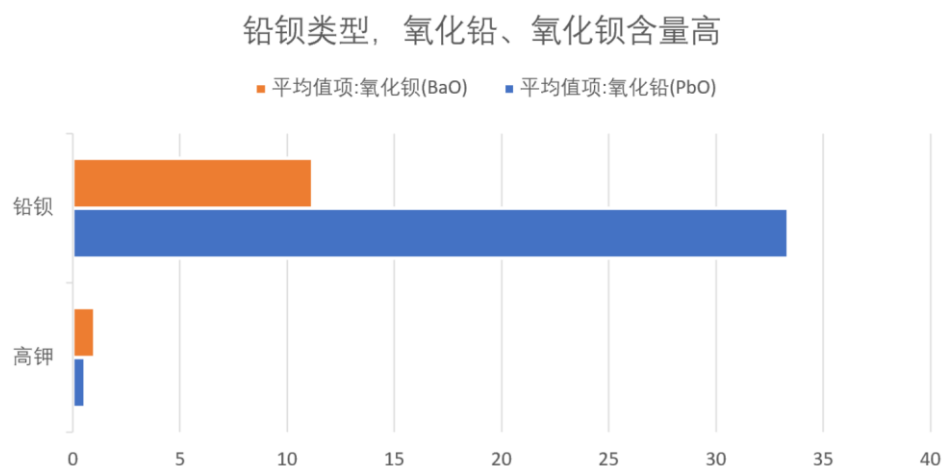


图 2-2 高钾类型与铅钡类型的氧化铅与氧化钡化学成分含量比较

### 5.2.2 模型建立

根据问题分析，本文决定采用聚类分析、判别分析和相关背景调研的方法，建立“基于玻璃文物化学成分含量的聚类判别分类模型”。

首先隐去高钾类型和铅钡类型的分类结果，对问题一中所得到的无风化数据进行聚类分析，对样本进行分类，使用 K-means 聚类分析方法，最终生成聚类类别汇总图如图 2-3 所示，观察分类结果与原始数据，发现 K-means 聚类分析完全把高钾类型（cluster 1）和铅钡类型（cluster 2）区分开，符合客观事实，说明本次聚类分析结果具有可靠性。

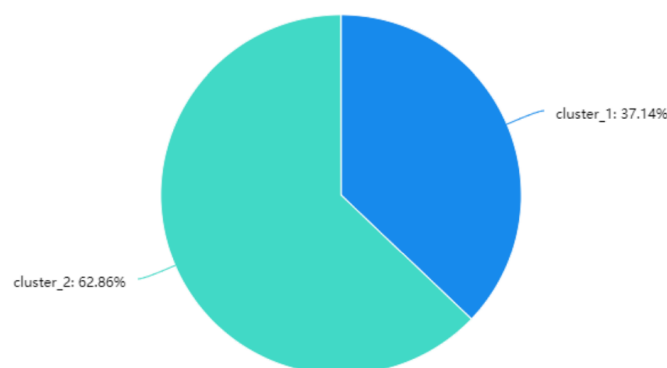


图 2-3 聚类类别汇总图

为了探索各个类别的具体特征，使用方差分析方法研究各个类别群体的差异性如下表 2-1 所示，可知聚类类别群体对所有研究项均呈现出显著性，意味着聚类分析得到的两类群体在研究项的特征上具有明显的差异性。

表 2-1 聚类类别方差差异分析对比结果

	cluster_1	cluster_2	F	P
二氧化硅(SiO <sub>2</sub> )	67.62±8.48	54.27±11.95	12.452	0.001**
氧化钾(K <sub>2</sub> O)	8.64±4.51	0.22±0.32	77.595	0.000**
氧化钙(CaO)	5.15±3.03	1.24±1.26	28.630	0.000**
氧化镁(MgO)	1.11±0.66	0.60±0.53	6.346	0.017*

	cluster_1	cluster_2	F	P
氧化铝( $\text{Al}_2\text{O}_3$ )	7.21±3.21	4.01±2.51	10.862	0.002**
氧化铁( $\text{Fe}_2\text{O}_3$ )	1.85±1.63	0.73±1.18	5.464	0.026*
氧化铅( $\text{PbO}$ )	1.33±3.35	22.53±8.12	79.783	0.000**
氧化钡( $\text{BaO}$ )	0.71±1.02	9.32±5.76	28.227	0.000**
氧化锶( $\text{SrO}$ )	0.06±0.07	0.27±0.25	8.800	0.006**

\*  $p < 0.05$  \*\*  $p < 0.01$

为探究高钾类型和铅钡类型的分类规律，结合聚类分析结果的聚类项贡献程度进行分析，聚类项贡献程度对比如图 2-4 所示。可以观察得到氧化钾、氧化铅和氧化钡指标重要性较高，与问题分析中对数据的观察结果相吻合，符合事实情况，进一步证明了使用聚类分析的合理性，以及氧化钾、氧化铅和氧化钡在高钾类型和铅钡类型分类过程中的重要性。

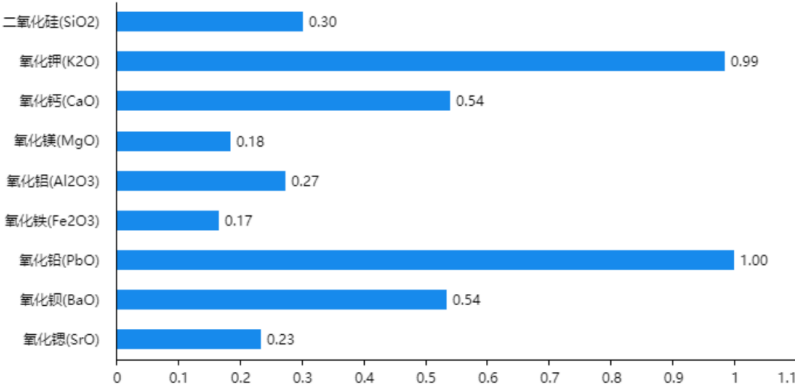


图 2-4 聚类项贡献程度对比图

故根据聚类分析结果，可以得到高钾类型和铅钡类型的分类规律如下：综合考虑二氧化硅、氧化钾、氧化钙、氧化镁、氧化铝、氧化铁、氧化铅、氧化钡和氧化锶共九种化学成分含量。

将这九种化学成分含量依次与表 2-1 的聚类类别结果（平均值±标准差）相比较，观察他们分别落入 cluster 1 或 cluster 2 类别，对落入 cluster 1 和 cluster 2 的项分别赋值为 1 和-1。再对九种化学成分的值按照图 2-4 各项贡献程度进行加权平均，若加权平均值大于 1 则为 cluster 1 类（即高钾类型），反之则为 cluster 2 类（即铅钡类型）。若加权平均值为 0 理论上分为 cluster 1 和 cluster 2 类皆可，但在对所有数据处理的过程中均未碰到此情况，故在此简单将其放入 cluster 2 类。

5.2.3 模型求解

下面将分别针对两种类型进行亚类划分，首先给出本文的亚类划分命名即符号表示如下表 2-2:

表 2-2 亚类划分命名表

	符号表示	亚类名称
高钾类型	A1	含钠类型 ( $\text{Na}_2\text{O}$ – $\text{K}_2\text{O}$ )
	A2	无钾类型
	A3	正常类型

铅钡类型	B1	高铅类型 (PbO-SiO <sub>2</sub> )
	B2	高钡类型 (SiO <sub>2</sub> -BaO)
	B3	含钠类型 (Na <sub>2</sub> O-PbO-BaO)
	B4	正常类型 (PbO-SiO <sub>2</sub> -BaO)
	B5	含硫类

在上表所示的铅钡类型中 B1 和 B5 类型，是从 B4 正常类型中进一步划分的特殊子类，严格意义上来说 B1 和 B5 都可以归属与 B4 类型，但在后期定量分析后，观察到其含有特殊的化学成分含量，故将其从 B4 中脱离出来，以提高分类的准确性。

在上述高钾类型亚类划分的基础上，基于所有高钾类型无风化数据，先根据调研和理论依据确定基本含量指标，高钾类型亚类划分基本含量指标为：氧化钠、氧化钾和二氧化硅，铅钡类型亚类划分基本含量指标为：氧化铅、氧化钡、氧化钠和二氧化硅。

根据参考指标利用聚类方法结合人工标注对高钾类型亚类划分进行定性分类处理，使用 Kmeans 聚类分析方法，最终生成聚类类别汇总图如图 2-5 所示。

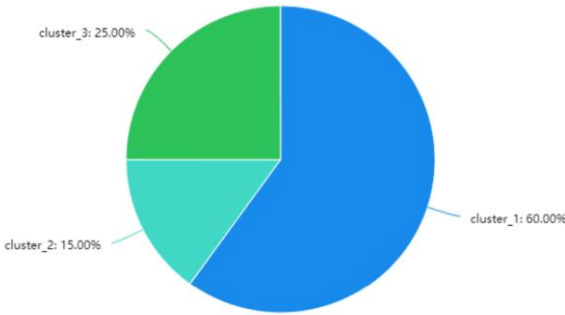


图 2-5 高钾类型聚类类别汇总图

使用方差分析方法研究各个类别群体的差异性如下表 2-3 所示，可知聚类类别群体中二氧化硅没有表现出显著性，氧化钠、氧化钾均呈现出显著性，意味着聚类分析得到的三类群体，在氧化钠、氧化钾特征上具有明显的差异性，而二氧化硅在此次划分中不起作用。

表 2-3 高钾类型亚类划分聚类类别方差差异分析对比结果

	cluster_1	cluster_2	cluster_3	F	P
二氧化硅(SiO <sub>2</sub> )	67.54±8.00	67.38±8.05	61.85±2.28	1.20	0.326
氧化钠(Na <sub>2</sub> O)	0.00±0.00	0.00±0.00	2.73±0.60	165.60	0.000**
氧化钾(K <sub>2</sub> O)	9.49±1.90	0.00±0.00	10.50±3.73	22.05	0.000**

\*  $p < 0.05$  \*\*  $p < 0.01$

01

为探究高钾类型亚类划分的分类规律，结合聚类分析结果的聚类项贡献程度进行分析，聚类项贡献程度对比如图 2-6 所示。可以观察得到氧化钾和氧化钠指标重要性较高，二氧化硅指标重要性低。结合三类指标的重要性以及对高钾类型玻璃的背景调研，分别对 cluster 1、cluster 2 和 cluster 3 进行命名，命名结果如下：cluster 1（A3 正常类型）、cluster 2（A2 无钾类型）和 cluster 3（A1

含钠类型)。

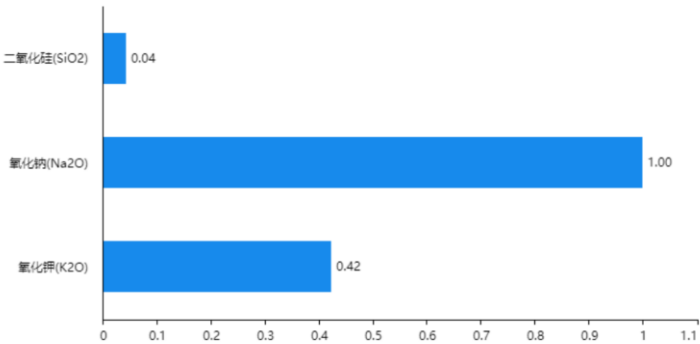


图 2-6 高钾类型亚类划分聚类项贡献程度对比图

根据参考指标利用聚类方法结合人工标注对铅钡类型亚类划分进行定性分类处理，使用 K-means 聚类分析方法，最终生成聚类类别汇总图如图 2-7 所示。

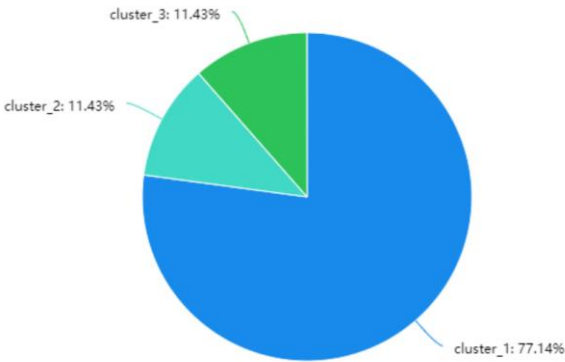


图 2-7 铅钡类型聚类类别汇总图

使用方差分析方法研究各个类别群体的差异性如下表 2-4 所示，可知聚类类别群体中二氧化硅没有表现出显著性，氧化钠、氧化钾均呈现出显著性，意味着聚类分析得到的三类群体，在氧化钠、氧化钾特征上具有明显的差异性，而二氧化硅在此次划分中不起作用。

表 2-4 铅钡类型亚类划分聚类类别方差差异分析对比结果

	cluster_1	cluster_2	cluster_3	F	P
氧化铅(PbO)	24.17±9.08	23.51±4.05	7.64±2.00	6.960	0.003**
氧化钠(Na2O)	1.48±2.25	0.00±0.00	0.00±0.00	1.654	0.207
氧化钡(BaO)	7.11±2.78	19.57±4.75	29.15±11.29	49.628	0.000**

\*  $p<0.05$  \*\*  $p<0.01$

为探究铅钡类型亚类划分的分类规律，结合聚类分析结果的聚类项贡献程度进行分析，聚类项贡献程度对比如图 2-8 所示。可以观察得到氧化钡和氧化铅指标重要性较高，氧化钠指标重要性低。结合三类指标的重要性以及对铅钡类型玻璃的背景调研，分别对 cluster 1、cluster 2 和 cluster 3 进行命名，命名结果如下：cluster 1（B3 含钠类型）、cluster 2（B4 正常类型）和 cluster 3（B2 高钡类型）。



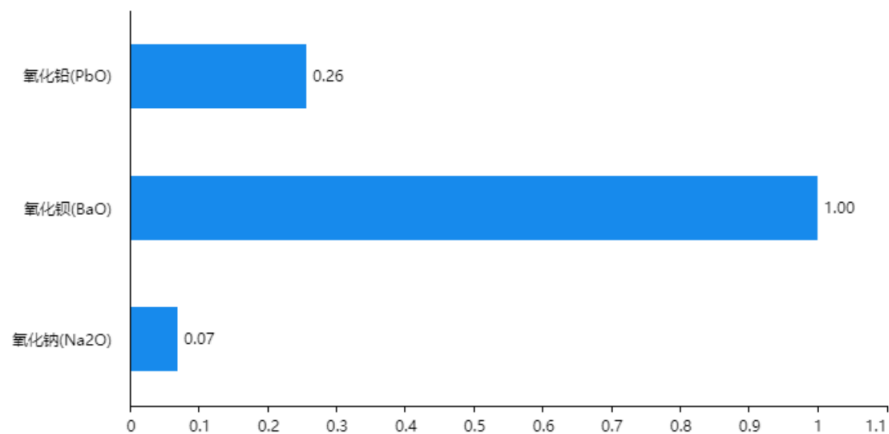


图 2-8 铅钡类型亚类划分聚类项贡献程度对比图

定性分类处理部分结果如下表 2-5:

表 2-5 定性分类处理部分结果展示表

文物采样点	类型	亚类	二氧化硅(SiO <sub>2</sub> )	氧化钠(Na <sub>2</sub> O)	氧化钾(K <sub>2</sub> O)	氧化铅(PbO)	氧化钡(BaO)
16	高钾	A1	65.18	2.10	14.52	0.11	0.00
27	高钾	A2	62.77	0.00	0.00	0.00	0.00
09	高钾	A3	64.89	0.00	10.06	0.00	0.00
02	铅钡	B1	63.34	0.00	0.51	25.35	0.00
11	铅钡	B2	60.05	0.00	0.48	4.78	15.97
56	铅钡	B2	54.62	0.00	0.00	19.58	17.08
55	铅钡	B3	49.01	2.71	0.00	32.92	7.95
52	铅钡	B3	50.46	1.34	0.00	25.34	8.11
31	铅钡	B4	65.91	0.00	0.00	16.55	3.42
33	铅钡	B4	75.51	0.00	0.15	16.16	3.55
39	铅钡	B5	51.08	0.00	0.00	38.04	6.24

最后利用判别分析进行亚类划分的定量分类处理，选用线性判别分析（LDA）方法来对每一项数据进行类别划分，在建立线性判别模型之前，先将数据划分为训练数据和测试数据两类，划分比例选取 8：2，利用 20%的测试数据来验证线性判别模型的有效性。

对高钾类型进行亚类划分的定量分类处理，首先依照上文所选择的高钾类型亚类划分基本含量指标氧化钠、氧化钾和二氧化硅建立判别函数。由于高钾类型的亚类划分的类别为 3 类，故应建立三个判别函数，分别对应 A1、A2 和 A3 三种类型，各分类判别函数如下表 2-6:

表 2-6 高钾类型亚类划分判别函数表

	类别 1(A1)	类别 2(A2)	类别 3(A3)
截距	-389.593	-274.052	-422.576
二氧化硅(SiO <sub>2</sub> )	10.022	8.671	10.370
氧化钠(Na <sub>2</sub> O)	12.279	13.058	8.594
氧化钾(K <sub>2</sub> O)	11.761	5.400	13.338



$$\begin{cases} \text{判别函数1} = -389.593 + 10.022 \cdot n_{Si} + 12.279 \cdot n_{Na} + 11.761 \cdot n_K \\ \text{判别函数2} = -274.052 + 8.671 \cdot n_{Si} + 13.058 \cdot n_{Na} + 5.400 \cdot n_K \\ \text{判别函数3} = -422.576 + 10.370 \cdot n_{Si} + 8.594 \cdot n_{Na} + 13.338 \cdot n_K \end{cases} \quad (3)$$

分别使用这三种判别函数 1, 2, 3, 根据数据的氧化钠、氧化钾和二氧化硅进行求值, 得到类别 1、2 和 3 的判别函数值, 选取判别函数值的最大值所对应的类别作为该数据的类别。在训练集上进行预测得到训练集预测准确率 (表 2-7), 在测试集上进行预测得到测试集预测准确率 (表 2-8)。

表 2-7 高钾类型亚类划分训练集预测准确率

预测类别	正确率 Precision	召回率 Recall	F1-score
类别 1 (A1)	100.00%	80.00%	88.89%
类别 2 (A2)	100.00%	100.00%	100.00%
类别 3 (A3)	90.00%	100.00%	94.74%
汇总	94.38%	93.75%	94.06%

表 2-8 高钾类型亚类划分测试集预测准确率

预测类别	正确率 Precision	召回率 Recall	F1-score
类别 1 (A1)	100.00%	100.00%	100.00%
类别 2 (A2)	100.00%	100.00%	100.00%
类别 3 (A3)	100.00%	86.67%	80.00%
汇总	100.00%	95.00%	90.71%

由上表可以看出, 正确率、召回率和 F1 指标值均在 90%, 说明线性判别分类模型在高钾类型亚类划分上表现较好, 使用该线性判别分类模型在高钾类型亚类划分的部分结果如下表 2-9 所示:

表 2-9 高钾类型亚类划分测试集预测准确率

文物标号	大类	亚类	类别 1 (A1)	类别 2 (A2)	类别 3 (A3)
07	高钾	A2	238.6862	269.533	227.5193
13	高钾	A1	384.2885	342.6316	381.0617
15	高钾	A1	357.3856	344.515	345.8374
27	高钾	A2	239.4879	270.2267	228.3489
10	高钾	A3	399.4196	358.8375	406.011
05	高钾	A3	356.3447	319.0382	362.0597

对铅钡类型进行亚类划分的定量分类处理, 首先依照上文所选择的铅钡类型亚类划分基本含量指标氧化铅、氧化钡、氧化钠和二氧化硅建立判别函数。根据上文所分析得到 B1、B5 在定性分析时皆为 B4 的子类, 故铅钡类型的亚类划分的类别为 3 类分别是 B2、B3 和 B4 类型, 应建立三个判别函数, 各分类判别函数如下表 2-10:

表 2-10 铅钡类型亚类划分判别函数表

	类别 1 (B2)	类别 2 (B3)	类别 3 (B4)
截距	-1042.381	-1063.244	-1079.412
二氧化硅 (SiO <sub>2</sub> )	23.034	23.468	23.760

	类别 1 (B2)	类别 2 (B3)	类别 3 (B4)
氧化钠 (Na <sub>2</sub> O)	15.255	23.809	16.868
氧化铅 (PbO)	24.787	25.396	25.864
氧化钡 (BaO)	26.861	25.193	25.424

$$\begin{cases} \text{判别函数1} = -389.593 + 10.022 \cdot n_{Si} + 12.279 \cdot n_{Na} + 11.761 \cdot n_K \\ \text{判别函数2} = -274.052 + 8.671 \cdot n_{Si} + 13.058 \cdot n_{Na} + 5.400 \cdot n_K \\ \text{判别函数3} = -422.576 + 10.370 \cdot n_{Si} + 8.594 \cdot n_{Na} + 13.338 \cdot n_K \end{cases} \quad (4)$$

分别使用这三种判别函数 1, 2, 3, 根据数据的氧化铅、氧化钡、氧化钠和二氧化硅进行求值, 得到类别 1、2 和 3 的判别函数值, 选取判别函数值的最大值所对应的类别作为该数据的类别。在训练集上进行预测得到训练集预测准确率 (表 2-11), 在测试集上进行预测得到测试集预测准确率 (表 2-12)。

表 2-11 铅钡类型亚类划分训练集预测准确率

预测类别	正确率 Precision	召回率 Recall	F1-score
类别 1 (B2)	100.00%	100.00%	100.00%
类别 2 (B3)	100.00%	100.00%	100.00%
类别 3 (B4)	100.00%	100.00%	100.00%
汇总	100.00%	100.00%	100.00%

表 2-12 铅钡类型亚类划分测试集预测准确率

预测类别	正确率 Precision	召回率 Recall	F1-score
类别 1 (B2)	100.00%	100.00%	100.00%
类别 2 (B3)	100.00%	81.82%	90.00%
类别 3 (B4)	85.71%	100.00%	92.31%
汇总	94.47%	93.55%	94.01%

由上表可以看出, 正确率、召回率和 F1 指标值均在 90%, 说明线性判别分类模型在铅钡类型亚类划分上表现较好。使用该线性判别分类模型在铅钡类型亚类划分的部分结果如下表 2-13 所示:

表 2-13 铅钡类型亚类划分测试集预测准确率

文物标号	大类	亚类	类别 1 (B2)	类别 2 (B3)	类别 3 (B4)
02	铅钡	B4	1044.943	1067.008	1081.199
46	铅钡	B4	1125.42	1127.115	1141.209
47	铅钡	B3	1093.396	1134.836	1115.392
45	铅钡	B3	1100.462	1120.404	1113.682
20	铅钡	B2	681.265	642.998	647.532
08	铅钡	B2	1173.936	1113.613	1122.601

在铅钡类型亚类划分的过程中, 注意到在类别 3 (B4) 的分类效果并不理想, 同时观察到效果较差的文物, 例如文物标号 02 氧化铅含量异常高, 故将这类氧化铅含量高的文物从 B4 中单独抽出, 结合铅钡玻璃的背景调研, 单独分类为 B1, 并命名为高铅类型。此外注意到 B4 中有部分文物含有氧化硫, 结合铅钡玻

璃的背景调研，将含有氧化硫的文物，单独分类为 B5 类，并命名为含硫类。

合理性分析：

为了评价模型的准确程度我们给予以下三种评价指标的定义：  
正确率是指预测为某类别时，实际情况下属于该类别的样本比例；  
召回率是指实际为某类别时，被正确预测到该类别的样本比例；  
F1-score 值是指正确率与召回率的加权综合指标，其计算公式为：

$$F1 - score = 2 \cdot \text{正确率} \cdot \frac{\text{召回率}}{\text{正确率} + \text{召回率}} \tag{5}$$

首先在训练判别分析模型时，我们将高钾类型训练集和测试集的比例调成 8：2，发现训练集预测准确率和测试集预测准确率均能够达到 100%，见下图 2-9、2-10，这足够说明我们对高钾类型的亚类进行划分时具有合理性。

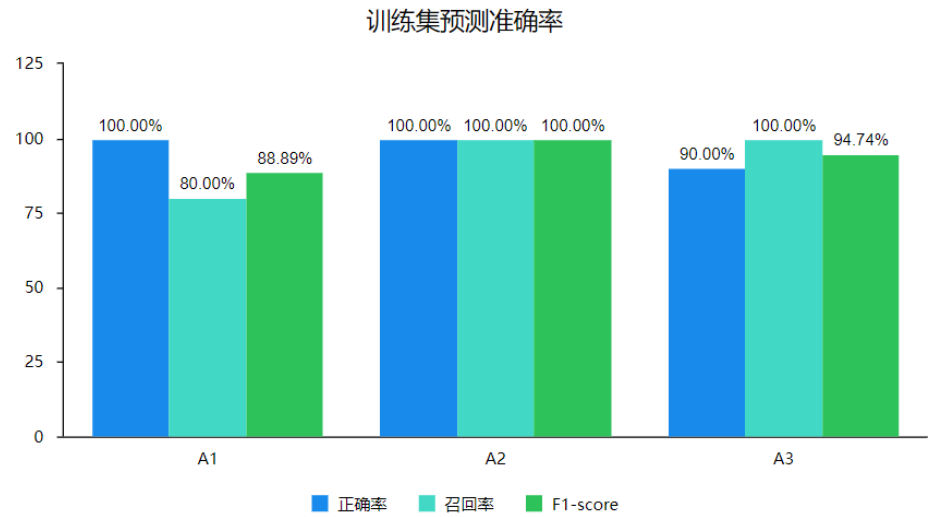


图 2-9 8：2 训练集预测准确率

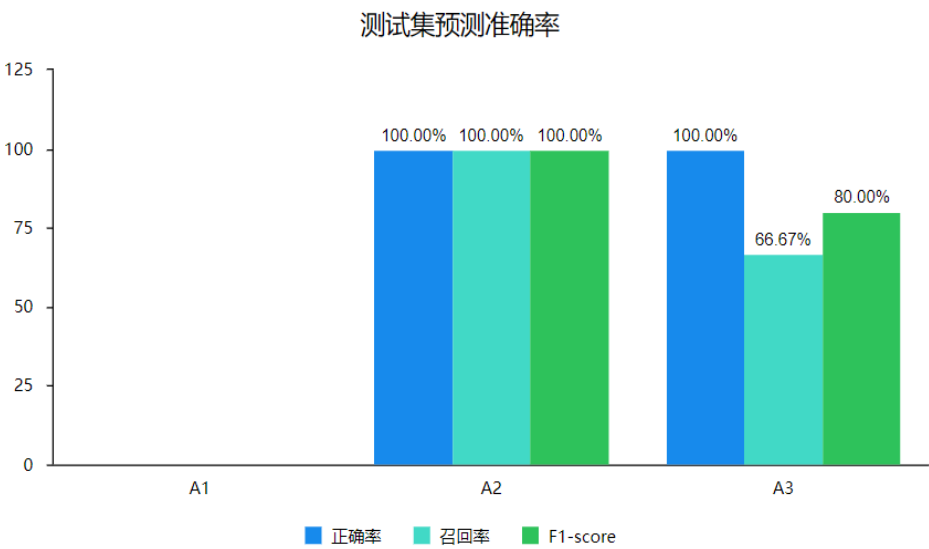


图 2-10 8：2 测试集预测准确率

同时，我们将铅钡类型的训练集和测试集的比例调整成 9：1，发现不仅测

试集的准确率，召回率和 F1-score 得分都能够达到 100%，并且也不难发现训练集的准确率，召回率和 F1-score 得分均能够达到较高的水平，见图 2-11、2-12，这无疑说明出模型的划分函数有着很好地划分效果，进一步说明出模型的合理性。

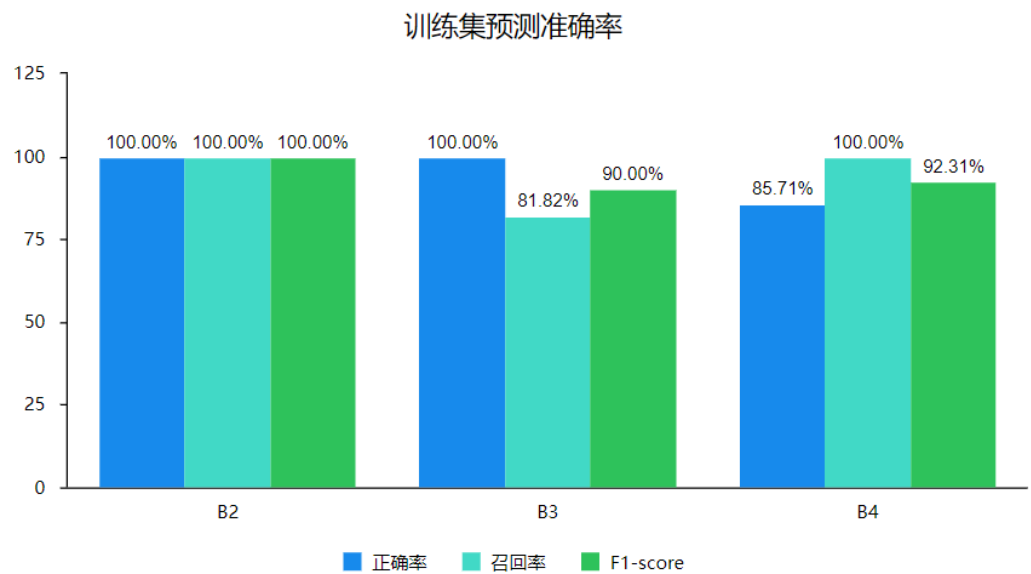


图 2- 11 9： 1 训练集预测准确率

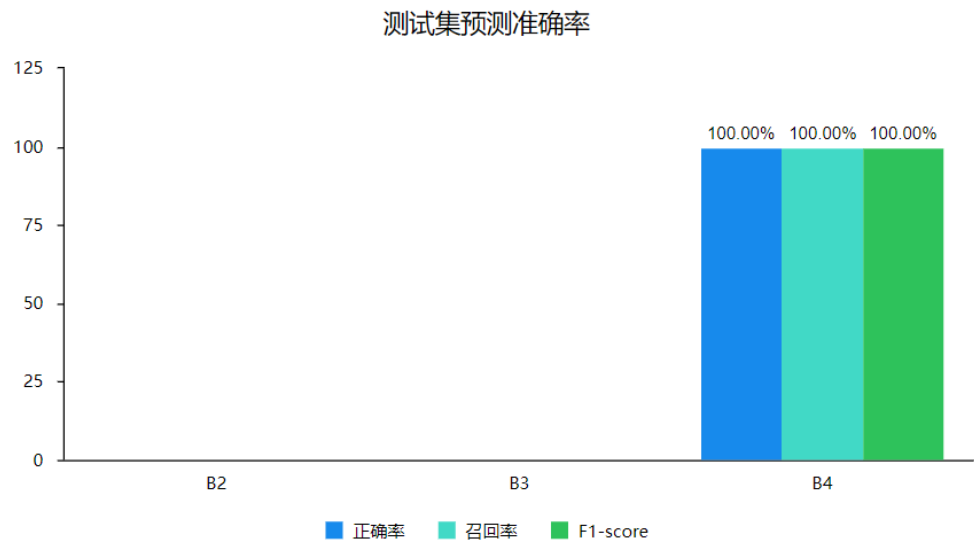


图 2- 12 9： 1 测试集预测准确率

**敏感性分析：**

首先根据百分量成分的基本守恒可以得到变量的变化关系，即一个百分比变量所占比的增加势必会引起其他变量的减少，将微小的变量的变化量之和近似忽略，得到变化量相对较大的变化关系如下面的公式（6）（7）所示，其中公式（6）为铅钡类型的近似守恒关系。而公式（7）则表示高钾类型的近似守恒关系。

$$\Delta_{Si} + \Delta_{Na} + \Delta_{Pb} + \Delta_{Ba} = 0 \quad (6)$$

$$\Delta_{Si} + \Delta_{Na} + \Delta_K = 0 \quad (7)$$

然后根据前面得到的判别公式得到判别公式 1，判别公式 2，判别公式 3 推导出变化量如下面的公式（8）（9）所示。其中公式（8）代表铅钡类型的判别公式变化量，公式（9）代表高钾类型的判别公式变化量。

$$\begin{cases} 23.034\Delta_{Si} + 15.255\Delta_{Na} + 24.787\Delta_{Pb} + 26.861\Delta_{Ba} \\ 23.468\Delta_{Si} + 23.809\Delta_{Na} + 25.396\Delta_{Pb} + 25.193\Delta_{Ba} \\ 23.760\Delta_{Si} + 16.868\Delta_{Na} + 25.864\Delta_{Pb} + 25.424\Delta_{Ba} \end{cases} \quad (8)$$

$$\begin{cases} 10.022\Delta_{Si} + 12.279\Delta_{Na} + 11.761\Delta_K \\ 8.671\Delta_{Si} + 13.058\Delta_{Na} + 5.400\Delta_K \\ 10.370\Delta_{Si} + 8.594\Delta_{Na} + 13.338\Delta_K \end{cases} \quad (9)$$

为了更好地得到近似，也是去除变化量较大的二氧化硅对于整体变化量的影响，我们将公式（6）（7）与公式（8）（9）联立从而消去变量二氧化硅得到公式（10）（11），其中公式（10）代表铅钡类型中判别函数变化量的简化结果，公式（11）代表高钾类型中判别函数变化量的简化结果。

$$\begin{cases} 8.120\Delta_{Na} + 0.175\Delta_{Pb} - 2.102\Delta_{Ba} \\ 0.887\Delta_{Na} + 0.351\Delta_{Pb} - 2.163\Delta_{Ba} \\ 7.333\Delta_{Na} + 0.196\Delta_{Pb} - 0.061\Delta_{Ba} \end{cases} \quad (10)$$

$$\begin{cases} 2.257\Delta_{Na} + 1.739\Delta_K \\ 4.387\Delta_{Na} - 3.271\Delta_K \\ -1.776\Delta_{Na} + 2.968\Delta_K \end{cases} \quad (11)$$

通过得到的最终近似公式，我们不妨令他们二组公式的 $\Delta_{Na}$ 等于 0.1，然后求解方程（10）（11）所能达到的最大值。解得最后的划分函数变化的最大值如下表 2-14 所示。

表 2-14 划分函数变化的最大值

类型	划分函数 1	划分函数 2	划分函数 3
高钾类型	0.518	1.116	1.5208
铅钡类型	3.952	1.587	4.788

为了更好地表示它的相对变化量，我们取到高钾类型和铅钡类型划分函数的均值如下表 2-15 所示。

表 2-15 划分函数的均值

类型	划分函数 1	划分函数 2	划分函数 3
高钾类型	378.88	354.15	379.30
铅钡类型	1031.37	1038.17	1042.33

将划分函数的最大值除以划分函数的均值，不难发现变化之均为 0.001 数量级，这无疑说明了划分函数具有良好的稳定性，即在变量发生微小变化时，划分函数的值也仅仅会发生微小的变化。进一步说明模型对于微小变化的敏感度不高，在分类时会有更好的稳定性。

5. 2. 4 问题三求解

首先利用模型一中的聚类模型首先对 8 个值进行分析，首先根据模型一中 K-means 聚类所分的五个类来对 8 个值进行区分并判断风化情况，在铅钡中发现 $A_5$ 虽然是表面风化，但它的的数据属于聚类中的第一类，所以初步判断其为表面风化的无风化点，所以无需进行数据转化直接进行判断。然后 $A_2$ 属于第二类铅钡普通分化类， $A_3$ ， $A_4$ ， $A_8$ 则是第一类铅钡无风化类。同样判断得到 $A_1$ 是高钾无风化类， $A_6$ ， $A_7$ 是高钾风化类，然后再次利用模型来预测其风化前的化学成分含量并进行替换结果见表 2-16、2-17。

表 2-16 原始数据表格

文物编号	表面风化	二氧化硅 (SiO <sub>2</sub> )	氧化钠 (Na <sub>2</sub> O)	氧化钾 (K <sub>2</sub> O)	氧化钙 (CaO)	氧化镁 (MgO)	二氧化硫 (SO <sub>2</sub> )
$A_1$	无风化	78.45			6.08	1.86	0.51
$A_2$	风化	37.75			7.63		
$A_3$	无风化	31.95		1.36	7.19	0.81	
$A_4$	无风化	35.47		0.79	2.89	1.05	
$A_5$	风化	64.29	1.2	0.37	1.64	2.34	
$A_6$	风化	93.17		1.35	0.64	0.21	
$A_7$	风化	90.83		0.98	1.12		0.11
$A_8$	无风化	51.12	0.00	0.23	0.89	0.00	2.26
氧化铝(Al <sub>2</sub> O <sub>3</sub> )	氧化铁 (Fe <sub>2</sub> O <sub>3</sub> )	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P <sub>2</sub> O <sub>5</sub> )	氧化锶 (SrO)	氧化锡 (SnO <sub>2</sub> )
7.23	2.15	2.11			1.06	0.03	
2.33			34.3		14.27		
2.93	7.06	0.21	39.58	4.69	2.68	0.52	
7.07	6.45	0.96	24.28	8.31	8.45	0.28	
12.75	0.81	0.94	12.23	2.16	0.19	0.21	0.49
1.52	0.27	1.73			0.21		
5.06	0.24	1.17			0.13		
2.12	0.00	9.01	21.24	11.34	1.46	0.31	

表 2-17 转化后的数据表格

文物编号	表面风化	二氧化硅 (SiO <sub>2</sub> )	氧化钠 (Na <sub>2</sub> O)	氧化钾 (K <sub>2</sub> O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al <sub>2</sub> O <sub>3</sub> )
$A_1$	无风化	78.45			6.08	1.86	7.23
$A_2$	无风化	65.13			3.875		4.32
$A_3$	无风化	31.95		1.36	7.19	0.81	2.93
$A_4$	无风化	35.47		0.79	2.89	1.05	7.07
$A_5$	风化（未风化点）	64.29	1.2	0.37	1.64	2.34	12.75
$A_6$	无风化	63.18		10.88	6.52	1.16	5.90
$A_7$	无风化	61.02		10.48	7.28	1.06	10.75
$A_8$	无风化	51.12		0.23	0.89		2.12
氧化铁(Fe <sub>2</sub> O <sub>3</sub> )	氧化铜(CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P <sub>2</sub> O <sub>5</sub> )	氧化锶 (SrO)	氧化锡 (SnO <sub>2</sub> )	二氧化硫 (SO <sub>2</sub> )
2.15	2.11			1.06	0.03		0.51
		13.09		1.176			
7.06	0.21	39.58	4.69	2.68	0.52		

6.45	0.96	24.28	8.31	8.45	0.28	
0.81	0.94	12.23	2.16	0.19	0.21	0.49
2.15	5.10			1.38		
2.05	3.20			0.69		0.11
	9.01	21.24	11.34	1.46	0.31	2.26

具体转化方式见问题一。

然后用其风化前的化学成分含量，利用模型二中的划分函数进行数值的计算与比较，得到划分函数中最小的划分函数所对应的类型，然后我们再将根据二氧化硫等特征性因素将元素分配到 $B_4$ 类对应的子类中，从而将 8 个值分到了不同的亚类中。划分结果如下表 2-18 所示。

表 2-18 亚类划分结果表

文物编号	判别函数 1	判别函数 2	判别函数 3	得到结果
$A_1$	396.63	406.18	390.95	$A_2$
$A_2$	782.47	797.85	806.83	$B_1$
$A_3$	800.60	809.88	822.65	$B_4$
$A_4$	599.67	595.13	602.61	$B_4$
$A_5$	817.95	839.09	839.59	$B_4$
$A_6$	371.63	332.59	377.79	$A_3$
$A_7$	345.29	311.71	350.08	$A_3$
$A_8$	966.20	961.54	972.86	$B_5$

**敏感性分析：**

为了分析分类的敏感性，我们将对模型设置一个微小的扰动，将表 2-14 中所得到的划分函数变化的最大值带入。从而得到在该种微扰状态下，铅钡类型和高钾类型的划分函数变化，然后再除上划分函数的原始值从而得到在这种微扰状态下划分函数的变化占比情况，从而可以观察出模型在微扰状态下的敏感性和稳定性。将表 2-14 中所得到的划分函数变化的最大值带入，计算出结果如表 2-19。

表 2-19 变化量占比原始值的情况

需判断的文物编号	占比划分函数 1	占比划分函数 2	占比划分函数 3
$A_1$	0.13%	0.27%	0.39%
$A_2$	0.50%	0.20%	0.49%
$A_3$	0.49%	0.19%	0.58%
$A_4$	0.66%	0.27%	0.79%
$A_5$	0.48%	0.19%	0.57%
$A_6$	0.14%	0.12%	0.16%
$A_7$	0.15%	0.14%	0.19%
$A_8$	0.41%	0.17%	0.49%

通过表 2-19 的数据不难发现，划分函数的变化量占原始量的比值在 $10^{-4}$ 数量级即所表示的变化对于原始值非常微小甚至可以忽略，这也说明模型在受到变量微扰时很难改变分类结果，即分类结果具有很强的稳定性。

5.3 基于相关性的灰色关联分析模型

5.3.1 问题分析

依据题意可知，本问题属于相关性问題，分析同类型下不同化学成分之间的关联性以及不同类型间化学成分关联关系的差异性。先确定类型种类，选择分类粒度最大的类别划分，即分为高钾类型和铅钡类型，再利用附件数据分别对两种类型计算相同类型下不同化学成分之间的相关性，分析不同化学成分之间的关联性，之后对两种类型间化学成分关联关系差异性进行分析。

5.3.2 模型建立

根据问题分析，采用相关分析和灰色关联的方法，建立“基于相关性的灰色关联分析模型”。

利用相关分析，分别研究高钾类型和铅钡类型不同化学成分之间的关系情况，是否有关系与关系紧密程度情况。再利用灰色关联，分别针对不同的化学成分，比较高钾类型和铅钡类型的不同化学成分之间的关系的差异性。

5.3.3 模型求解

首先对高钾类型的不同化学成分进行相关分析，利用相关分析去研究氧化钠、氧化钾、氧化钙、氧化镁、氧化铝、氧化铁、氧化铜、二氧化硅、氧化铅、氧化钡、五氧化二磷、氧化锶、氧化锡和二氧化硫共 14 项之间的相关关系，使用 Pearson 相关系数去表示相关关系的强弱情况，Pearson 相关结果如图 3-1 所示。

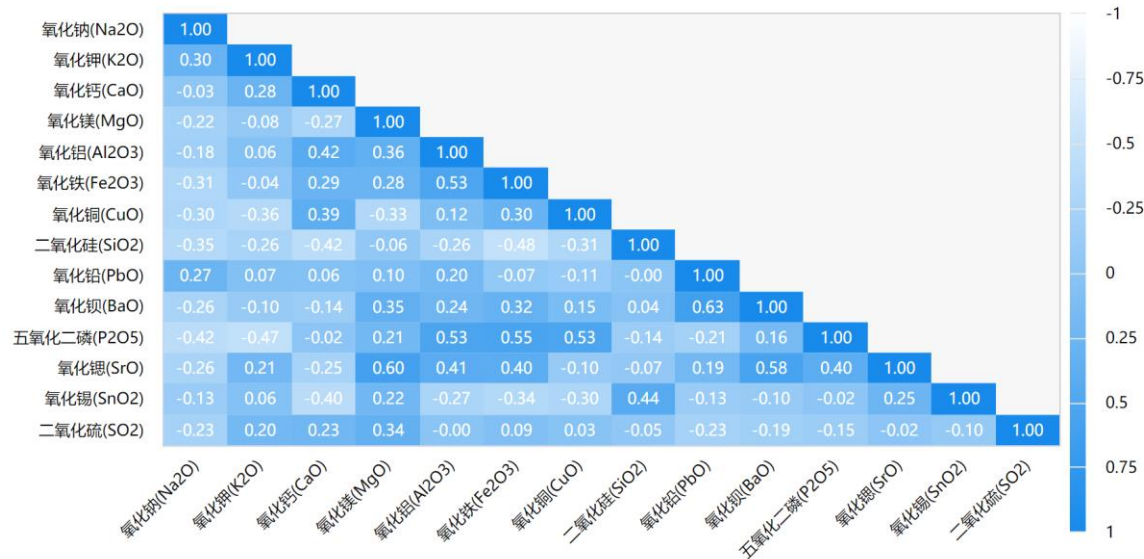


图 3-1 高钾类型 Pearson 相关可视化图

具体分析可知：

- 1. 氧化钠与其他化学成分之间无显著相关。
- 2. 氧化钾与五氧化二磷呈现出显著性，相关系数值是-0.467。
- 3. 氧化钙与其他化学成分之间无显著相关。
- 4. 氧化镁与氧化锶呈现出显著性，相关系数值是 0.601。



- 氧化铝与氧化铁、五氧化二磷呈现出显著性，相关系数值分别是 0.527，0.531。
- 氧化铁与氧化铝、二氧化硅、五氧化二磷呈现出显著性，相关系数值分别是 0.527，-0.479，0.551。
- 氧化铜与五氧化二磷呈现出显著性，相关系数值是 0.529。
- 二氧化硅与氧化铁呈现出显著性，相关系数值是-0.479。
- 氧化铅与氧化钡呈现出显著性，相关系数值是 0.632。
- 氧化钡与氧化铅、氧化锶呈现出显著性，相关系数值分别是 0.632，0.578。
- 五氧化二磷与氧化钾、氧化铝、氧化铁、氧化铜呈现出显著性，相关系数值分别是-0.467，0.531，0.551，0.529。
- 氧化锶与氧化镁、氧化钡呈现出显著性，相关系数值分别是 0.601，0.578。
- 氧化锡与其他化学成分之间无显著相关。
- 二氧化硫与其他化学成分之间无显著相关。

再对铅钡类型的不同化学成分进行相关分析，利用相关分析去研究氧化钠、氧化钾、氧化钙、氧化镁、氧化铝、氧化铁、氧化铜、二氧化硅、氧化铅、氧化钡、五氧化二磷、氧化锶、氧化锡和二氧化硫共 14 项之间的相关关系，使用 Pearson 相关系数去表示相关关系的强弱情况，Pearson 相关结果如图 3-2 所示。

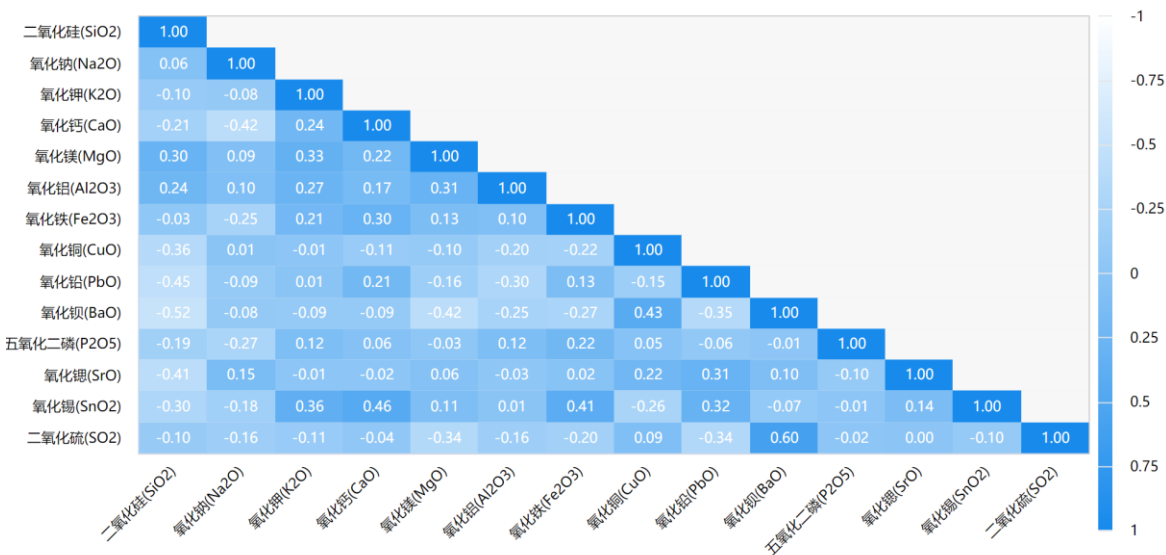


图 3-2 Pearson 相关可视化图

具体分析可知：

- 二氧化硅与氧化铜, 氧化铅, 氧化钡, 氧化锶共 4 项之间全部均呈现出显著性，相关系数值分别是-0.357, -0.449, -0.520, -0.413。
- 氧化钠与氧化钙呈现出显著性，相关系数值分别是-0.421。

3. 氧化钾与氧化镁、氧化锡呈现出显著性，相关系数值分别是 0.333，0.362。
4. 氧化钙与氧化钠、氧化锡呈现出显著性，相关系数值分别是-0.421，0.457。
5. 氧化镁与氧化钾、氧化钡、二氧化硫呈现出显著性，相关系数值分别是 0.333，-0.423，-0.342。
6. 氧化铝与其他化学成分之间无显著相关。
7. 氧化铁与氧化锡呈现显著性，相关系数 0.412。
8. 氧化铜与二氧化硅、氧化钡呈现显著性，相关系数值分别是-0.357，0.428。
9. 氧化铅与二氧化硅、氧化钡、二氧化硫呈现显著性，相关系数值分别是 -0.449，-0.350，-0.341。
10. 氧化钡与二氧化硅、氧化镁、氧化铜、氧化铅、二氧化硫呈现显著性，相关系数值分别是 -0.520，-0.423，0.428，-0.350，0.601。
11. 五氧化二磷与其他化学成分之间无显著相关。
12. 氧化锶与二氧化硅呈现显著性，相关系数值分别是 -0.413。
13. 氧化锡与氧化钾、氧化钙、氧化铁，相关系数值分别是 0.362，0.457，0.412。
14. 二氧化硫与氧化镁、氧化铅、氧化钡呈现显著性，相关系数值分别是-0.342，-0.341，0.601。

针对不同类别之间的化学成分关联关系的差异性：利用灰色关联法，以二氧化硅作为参考值（母序列），其余所有项作为特征序列（评价值）。从而判定评分值与“参考值”（母序列）的相似程度，进而针对评价项进行评价。主要分为以下三个步骤：

第一步：针对数据进行无量纲化处理；

第二步：计算关联系数，关联系数表示某项与“参考值”（母序列）的相关程度；

第三步：结合关联系数值计算关联度。

关联系数的全部计算结果见附录 2、3，此处展示二氧化硅的关联系数计算结果，其余变量的计算过程与其相同。最后通过关联系数值所计算出来的关联度如下图 3-3、3-4 所示，关联度很好地表达出二氧化硅与其他参考变量之间的关系，一般来说关联度越高，变量之间的相关性越强。同时分别做出两种不同类型的参考变量关系，不难发现其明显的差异性。比如在铅钡类型中二氧

化硅与其他变量的关联度要明显的平均低于其在高钾类型中与其他变量的关联度，这无疑说明了高钾类型中二氧化硅与其他变量的相关性要强于铅钡类型。

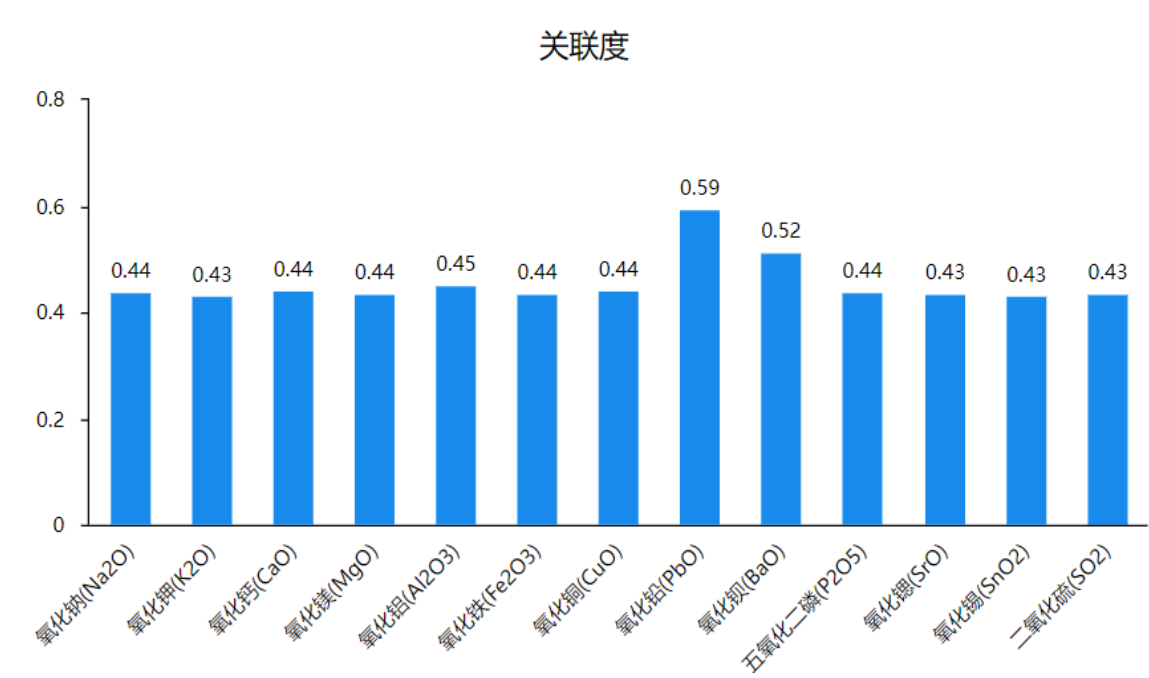


图 3-3 铅钡类型中二氧化硅与其他变量关联度

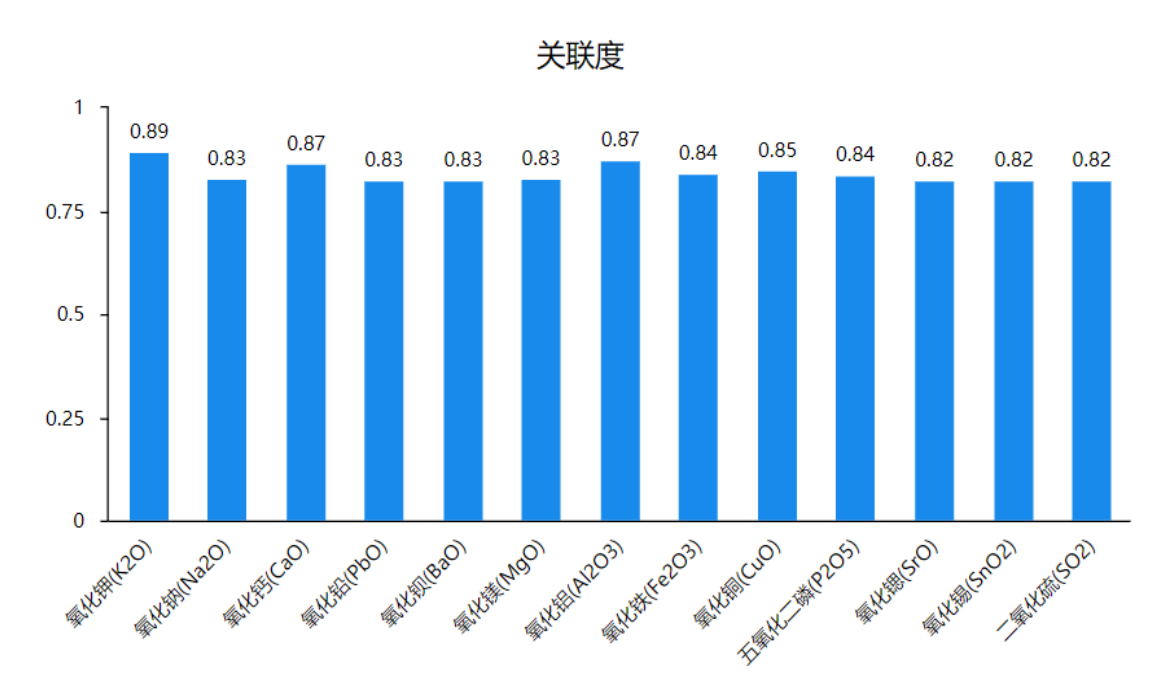


图 3-4 高钾类型中二氧化硅与其他变量关联度

同理可以得到其他的一系列关系。图 3-5、3-6 说明了铅钡类型中氧化钾与其他变量关联度整体略低于高钾类型中氧化钾与其他变量关联度。但铅钡类型

中氧化钾与氧化铅和氧化钡的关联度却高于高钾类型，这说明在铅钡类型中虽然氧化钾与其他变量相比相关性不及高钾类型，但是却和氧化铅和氧化钡呈强相关。

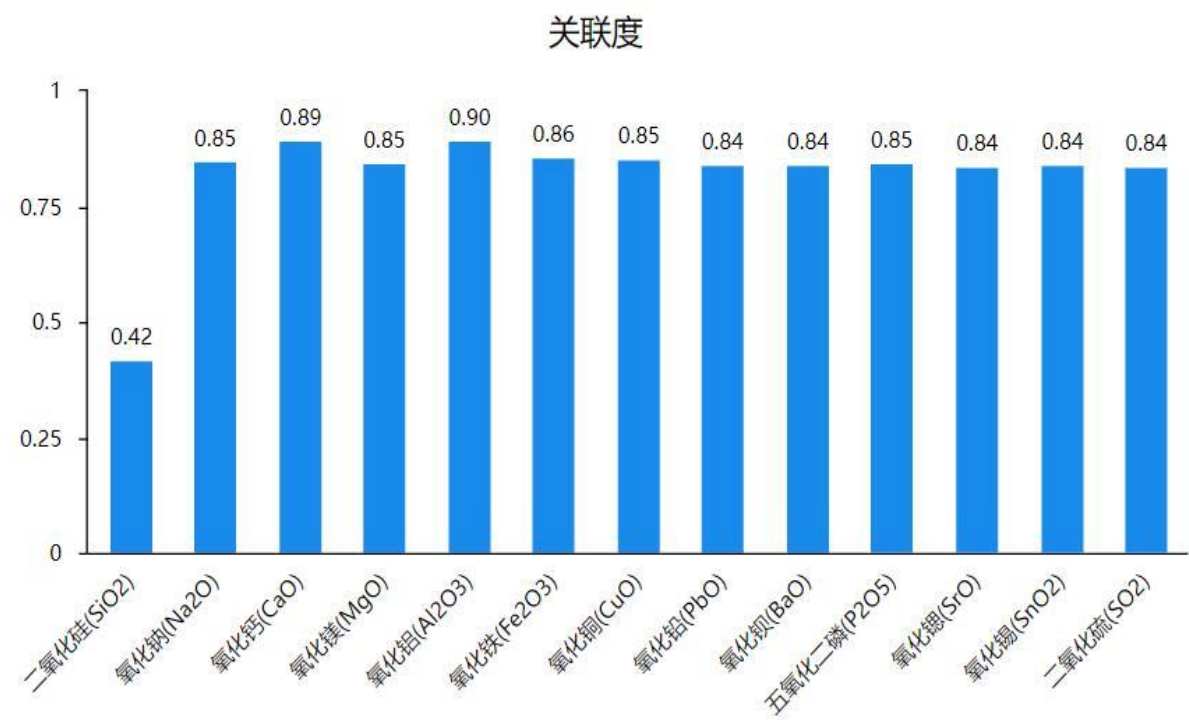


图 3-5 铅钡类型中氧化钾与其他变量关联度

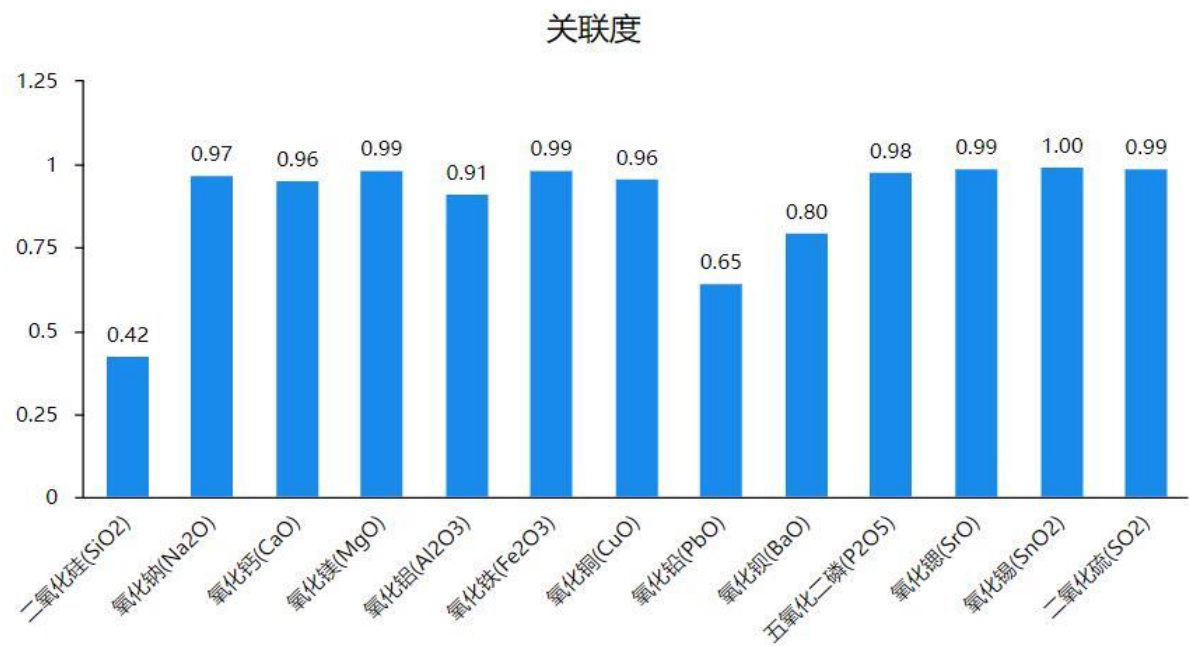


图 3-6 高钾类型中氧化钾与其他变量关联度

图 3-7、3-8 以及图 3-9、3-10 说明了在铅钡类型中，氧化铅和氧化钡与其他变量的关联度高于高钾类型。当然与二氧化硅这一变量的关联度却没有铅钡类型高。这也充分说明了铅钡类型中，氧化铅和氧化钡与其他变量的相关性要强于氧化钾（二氧化硅除外）。进一步说明出铅钡类型与氧化铅，氧化钡的强相关性和高钾类型与氧化钾的强相关性。

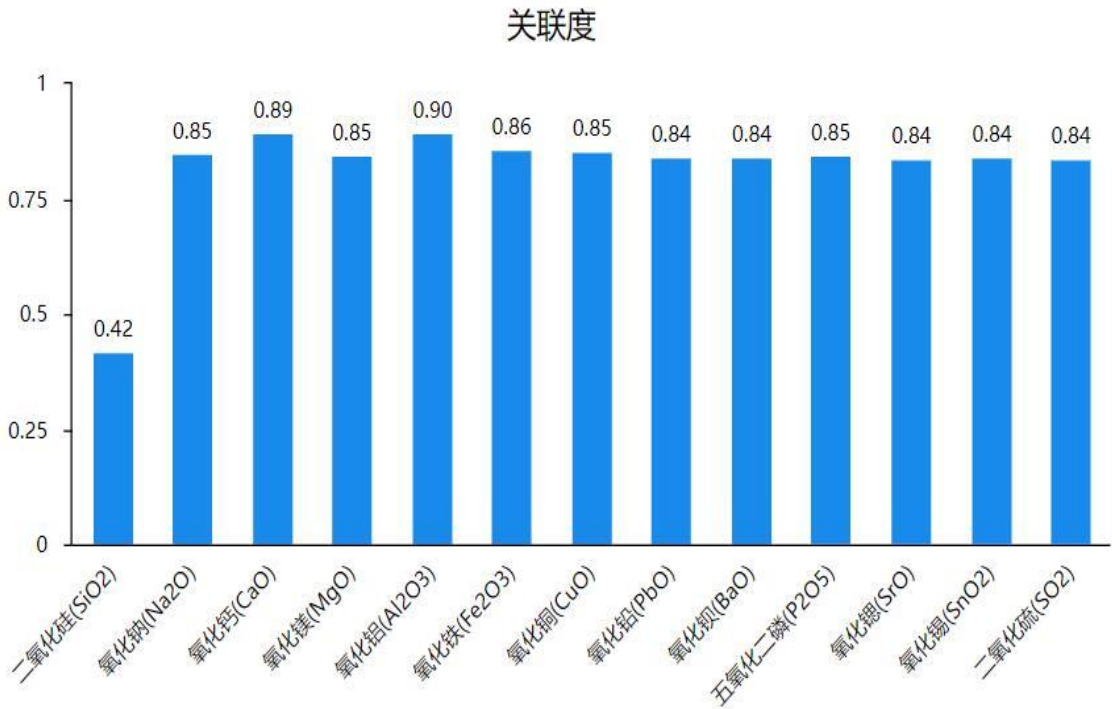


图 3-7 铅钡类型中氧化铅与其他变量关联度

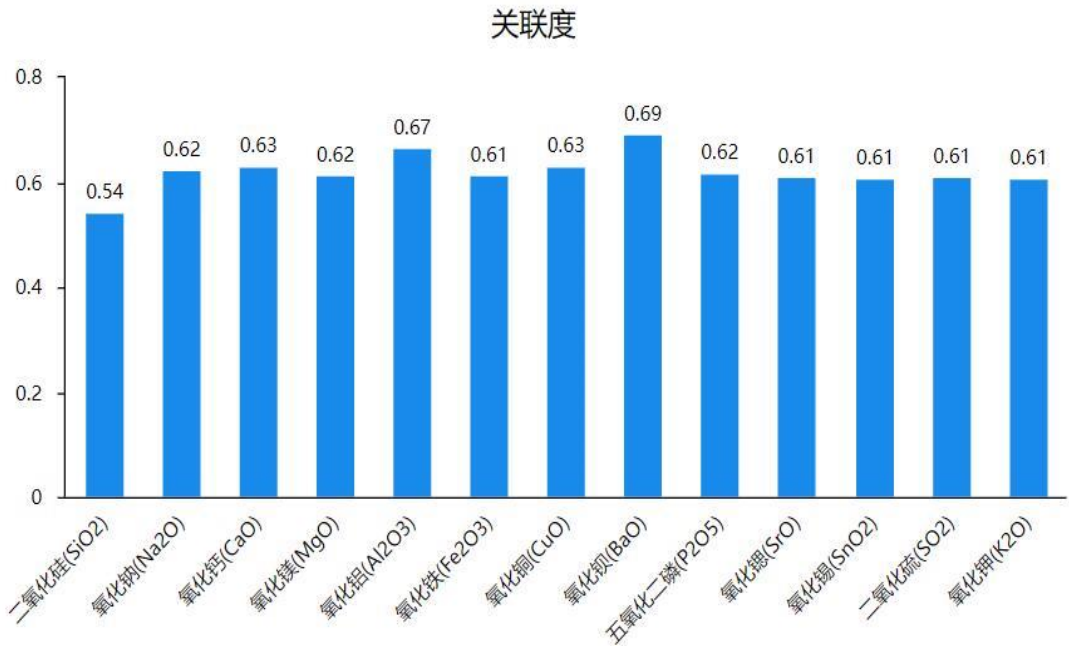


图 3-8 高钾类型中氧化铅与其他变量关联度

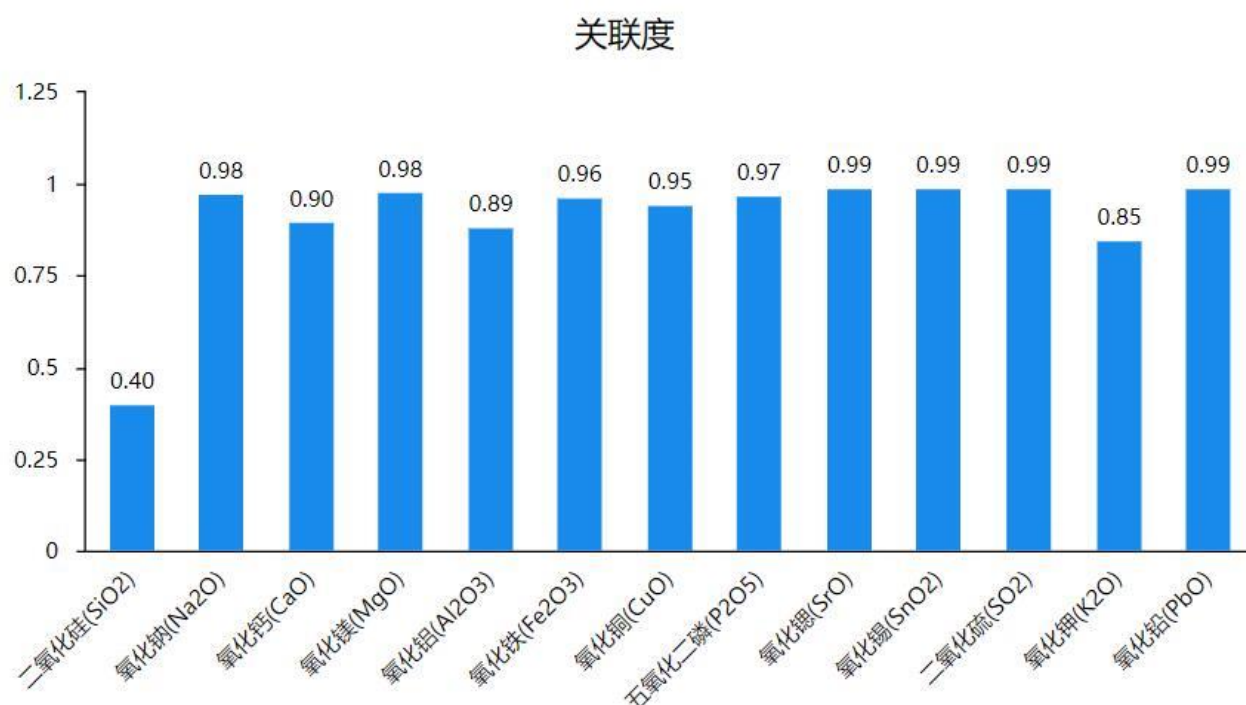


图 3-9 铅钡类型中氧化钡与其他变量关联度

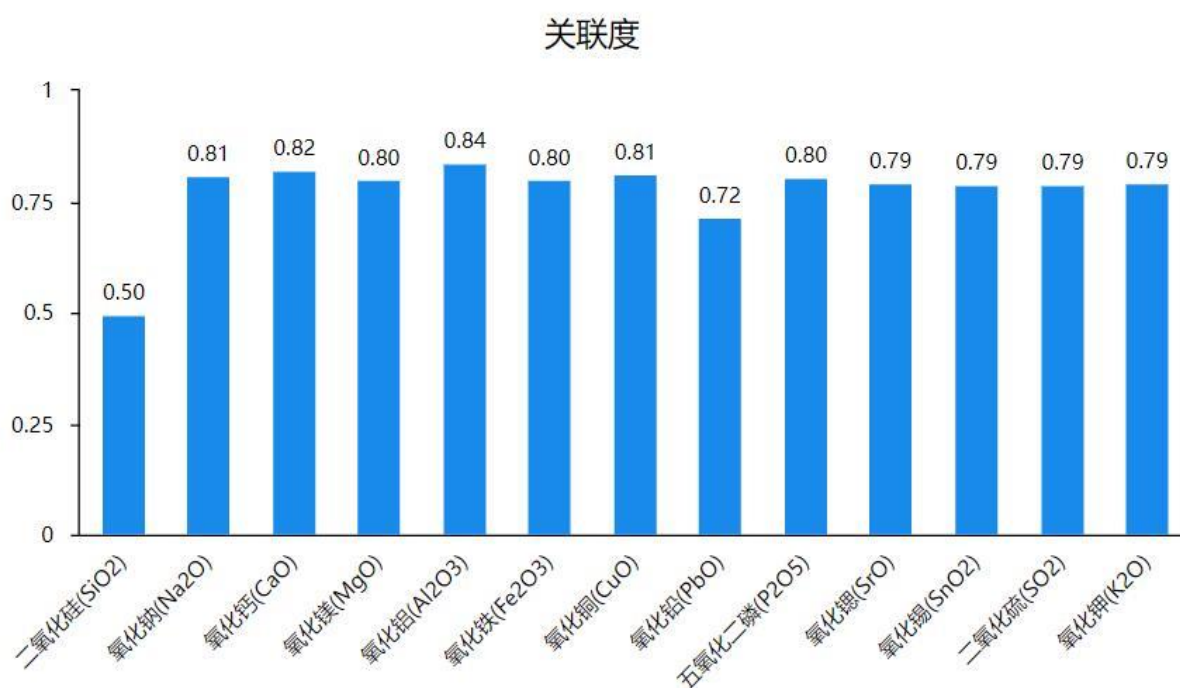


图 3-10 高钾类型中氧化钡与其他变量的关联度

## 六、模型评价

### 6.1 模型优点

1. 基于玻璃文物的化学成分含量的聚类判别模型训练集和测试集均具有很高的正确率和 F1-score 得分, 说明模型具有良好的合理性。

2. 基于玻璃文物的化学成分含量的聚类判别模型在敏感性分析中, 模型的敏感性和分类结果的敏感性在微扰状态下的相对变化量为  $10^{-3}$  数量级, 说明微小的成分变化很难影响到分类结果, 说明模型具有良好的稳定性。

3. 基于聚类分析的化学成分含量预测模型很好的遵循了统计规律以及集合映射的特点, 良好的数学理论支持说明模型具有很强的鲁棒性。

4. 基于相关性的灰色关联分析模型, 在不同种类相关性比较时很好地将定性的结论用定量的方式来加以表达, 数据和结论更加的直观。

### 6.2 模型缺点

1. 虽然采用了亚类的子类来尽量遍历特殊情况, 但仍然存在一部分特殊的个例没有很好地进行表达。

2. 虽然引入了化学原理来进行辅助数据分析但仍然对考古层面的知识涉及不够。

## 七、参考文献

[1] 骆占斌, 樊军, 邵明安. 地球关键带基岩风化层生态水文研究进展[J/OL]. 科学通报: 1-13[2022-09-18].

[2] 罗健. 两个化学风化反应方程式的讨论[J]. 地质论评, 1987(03): 291-296. DOI:10.16509/j.georeview.1987.03.012.

[3] 刘壮飞. 玻璃制品的风化及其预防措施[J]. 商业科技, 1987(05): 30-31.

[4] 李青会, 干福熹, 顾冬红. 关于中国古代玻璃研究的几个问题[J]. 自然科学史研究, 2007(02): 234-247.

[5] 干福熹. 中国古代玻璃的起源和发展[J]. 自然杂志, 2006(04): 187-193+184.

## 八、附录清单

- 1、风化前各化学成分含量预测结果表
- 2、铅钡类型关联系数结果表
- 3、高钾类型关联系数结果
- 4、K-means 聚类代码
- 5、灰色关联分析算法代码
- 6、相关系数代码
- 7、鉴别分析代码

# 附录

附录 1								
风化前各化学成分含量预测结果表								
文物采样点	纹饰	类型	颜色	二氧化硅 (SiO2)	氧化钠 (Na2O)	氧化钾 (K2O)	氧化钙 (CaO)	氧化镁 (MgO)
20	A	铅钡	浅蓝	37.36	0	0.71	0	0
24	C	铅钡	紫	31.94	0	0	0.47	0
31	C	铅钡	紫	65.91	0	0	1.6	0.89
32	C	铅钡	浅绿	69.71	0	0.21	0.46	0
33	C	铅钡	深绿	75.51	0	0.15	0.64	1
35	C	铅钡	浅绿	65.91	0	0	0.38	0
37	C	铅钡	深绿	60.12	0	0.23	0.89	0
45	A	铅钡	浅蓝	61.28	2.66	0.11	0.84	0.74
46	A	铅钡	浅蓝	55.21	0	0.25	0	1.67
47	A	铅钡	浅蓝	51.54	4.66	0.29	0.87	0.61
55	C	铅钡	绿	49.01	2.71	0	1.13	0
30 部位 1	A	铅钡	深蓝	34.34	0	1.41	4.49	0.98
30 部位 2	A	铅钡	深蓝	36.93	0	0	4.24	0.51
23 未风化点	A	铅钡	蓝绿	53.79	7.92	0	0.5	0.71
25 未风化点	C	铅钡	浅蓝	50.61	2.31	0	0.63	0
28 未风化点	A	铅钡	浅蓝	68.08	0	0.26	1.34	1
29 未风化点	A	铅钡	浅蓝	63.3	0.92	0.3	2.98	1.49
42 未风化点 1	A	铅钡	浅蓝	51.26	5.74	0.15	0.79	1.09
42 未风化点 2	A	铅钡	浅蓝	51.33	5.68	0.35	0	1.16
44 未风化点	A	铅钡	浅蓝	60.74	3.06	0.2	2.14	0
49 未风化点	A	铅钡	黑	54.61	0	0.3	2.08	1.2
50 未风化点	A	铅钡	黑	45.02	0	0	3.12	0.54
53 未风化点	A	铅钡	浅蓝	63.66	3.04	0.11	0.78	1.14
03 部位 1	A	高钾	蓝绿	87.05	0	5.19	2.01	0
03 部位 2	A	高钾	蓝绿	61.71	0	12.37	5.87	1.11
04	A	高钾	蓝绿	65.88	0	9.67	7.12	1.56
05	A	高钾	蓝绿	61.58	0	10.95	7.35	1.77
06 部位 1	A	高钾	蓝绿	67.65	0	7.37	0	1.98
06 部位 2	A	高钾	蓝绿	59.81	0	7.68	5.41	1.73
18	A	高钾	深蓝	79.46	0	9.42	0	1.53
21	A	高钾	蓝绿	76.68	0	0	4.71	1.22
01	C	高钾	蓝绿	69.33	0	9.99	6.32	0.87
13	C	高钾	浅蓝	59.01	2.86	12.53	8.7	0
14	C	高钾	深绿	62.47	3.38	12.28	8.23	0.66
16	C	高钾	浅蓝	65.18	2.1	14.52	8.27	0.52
氧化铝 (Al2O3)	氧化铁 (Fe2O3)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P2O5)	氧化锶 (SrO)	氧化锡 (SnO2)	二氧化硫 (SO2)
5.45	1.51	4.78	9.3	23.55	5.75	0	0	0
1.59	0	8.46	29.14	26.23	0.14	0.91	0	0
3.11	4.59	0.44	16.55	3.42	1.62	0.3	0	0
2.36	1	0.11	19.76	4.88	0.17	0	0	0
2.35	0	0.47	16.16	3.55	0.13	0	0	0



1.44	0.17	0.16	22.05	5.68	0.42	0	0	0
2.72	0	3.01	17.24	10.34	1.46	0.31	0	3.66
5	0	0.53	15.99	10.96	0	0.23	0	0
4.79	0	0.77	25.25	10.06	0.2	0.43	0	0
3.06	0	0.65	25.4	9.23	0.1	0.85	0	0
1.45	0	0.86	32.92	7.95	0.35	0	0	0
4.35	2.12	0	39.22	10.29	0	0.35	0.4	0
3.86	2.74	0	37.74	10.35	1.41	0.48	0.44	0
1.42	0	2.99	16.98	11.86	0	0.33	0	0
1.9	1.55	1.12	31.9	6.65	0.19	0.2	0	0
4.7	0.41	0.33	17.14	4.04	1.04	0.12	0.23	0
14.34	0.81	0.74	12.31	2.03	0.41	0.25	0	0
3.53	0	2.67	21.88	10.47	0.08	0.35	0	0
5.66	0	2.72	20.12	10.88	0	0	0	0
12.69	0.77	0.43	13.61	5.22	0	0.26	0	0
6.5	1.27	0.45	23.02	4.19	4.32	0.3	0	0
4.16	0	0.7	30.61	6.22	6.34	0.23	0	0
6.06	0	0.54	13.66	8.99	0	0.27	0	0
4.06	0	0.78	0.25	0	0.66	0	0	0
5.5	2.16	5.09	1.41	2.86	0.7	0.1	0	0
6.44	2.06	2.18	0	0	0.79	0	0	0.36
7.5	2.62	3.27	0	0	0.94	0.06	0	0.47
11.15	2.39	2.51	0.2	1.38	4.18	0.11	0	0
10.05	6.04	2.18	0.35	0.97	4.5	0.12	0	0
3.05	0	0	0	0	1.36	0.07	2.36	0
6.19	2.37	3.28	1	1.97	1.1	0	0	0
3.93	1.74	3.87	0	0	1.17	0	0	0.39
6.16	2.88	4.73	0	0	1.27	0	0	0
9.23	0.5	0.47	1.62	0	0.16	0	0	0
6.18	0.42	1.07	0.11	0	0	0.04	0	0

附录 2				
铅钡类型关联系数结果表				
铅钡类型关联系数结果表				
项	氧化钠(Na <sub>2</sub> O)	氧化钾(K <sub>2</sub> O)	氧化钙(CaO)	氧化镁(MgO)
第 1 项	0.490	0.490	0.502	0.495
第 2 项	0.394	0.396	0.409	0.400
第 3 项	0.417	0.417	0.431	0.417
第 4 项	0.439	0.439	0.453	0.439
第 5 项	0.513	0.518	0.513	0.513
第 6 项	0.553	0.553	0.557	0.553
第 7 项	0.474	0.474	0.486	0.474
第 8 项	0.476	0.476	0.489	0.476
第 9 项	0.385	0.383	0.393	0.387
第 10 项	0.404	0.392	0.400	0.392
第 11 项	0.448	0.436	0.440	0.436
第 12 项	0.459	0.444	0.450	0.444
第 13 项	0.444	0.437	0.451	0.443
第 14 项	0.392	0.381	0.383	0.385
第 15 项	0.456	0.433	0.436	0.435
第 16 项	0.463	0.434	0.437	0.439
第 17 项	0.462	0.435	0.433	0.439
第 18 项	0.400	0.390	0.393	0.392
第 19 项	0.461	0.421	0.424	0.425
第 20 项	0.415	0.416	0.415	0.422
第 21 项	0.535	0.546	0.570	0.542
第 22 项	0.516	0.516	0.547	0.520
第 23 项	0.372	0.372	0.378	0.375
第 24 项	0.340	0.341	0.342	0.344
第 25 项	0.415	0.415	0.429	0.420
第 26 项	0.364	0.365	0.369	0.368
第 27 项	0.418	0.419	0.427	0.423
第 28 项	0.359	0.360	0.360	0.359
第 29 项	0.500	0.500	0.511	0.500
第 30 项	0.372	0.372	0.373	0.372
第 31 项	0.459	0.461	0.472	0.464
第 32 项	0.466	0.466	0.484	0.469
第 33 项	0.434	0.434	0.448	0.434
第 34 项	0.411	0.413	0.425	0.416
第 35 项	0.444	0.444	0.458	0.450
项	氧化铝(Al <sub>2</sub> O <sub>3</sub> )	氧化铁(Fe <sub>2</sub> O <sub>3</sub> )	氧化铜(CuO)	氧化铅(PbO)
第 1 项	0.503	0.494	0.500	0.682
第 2 项	0.410	0.394	0.407	0.415
第 3 项	0.433	0.417	0.430	0.530
第 4 项	0.454	0.439	0.451	0.597
第 5 项	0.554	0.524	0.548	0.586
第 6 项	0.566	0.553	0.630	0.951
第 7 项	0.488	0.474	0.485	0.525
第 8 项	0.490	0.476	0.487	0.533
第 9 项	0.445	0.385	0.384	0.435

第 10 项	0.449	0.395	0.393	0.454
第 11 项	0.446	0.444	0.442	0.683
第 12 项	0.452	0.444	0.449	0.716
第 13 项	0.452	0.442	0.449	0.613
第 14 项	0.404	0.380	0.382	0.439
第 15 项	0.447	0.432	0.435	0.604
第 16 项	0.451	0.433	0.447	0.574
第 17 项	0.462	0.433	0.447	0.559
第 18 项	0.410	0.389	0.392	0.464
第 19 项	0.428	0.421	0.435	0.517
第 20 项	0.437	0.415	0.418	0.570
第 21 项	0.569	0.551	0.535	0.905
第 22 项	0.545	0.536	0.516	1.000
第 23 项	0.384	0.389	0.374	0.443
第 24 项	0.348	0.340	0.342	0.397
第 25 项	0.431	0.419	0.427	0.536
第 26 项	0.381	0.366	0.366	0.435
第 27 项	0.449	0.423	0.420	0.556
第 28 项	0.367	0.362	0.359	0.440
第 29 项	0.512	0.504	0.500	0.858
第 30 项	0.377	0.373	0.373	0.473
第 31 项	0.474	0.459	0.470	0.754
第 32 项	0.491	0.466	0.470	0.739
第 33 项	0.449	0.434	0.446	0.759
第 34 项	0.427	0.415	0.423	0.507
第 35 项	0.459	0.449	0.456	0.566

氧化钡(BaO)	五氧化二磷(P2O5)	氧化锶(SrO)	氧化锡(SnO2)	二氧化硫(SO2)
0.609	0.493	0.492	0.490	0.490
0.471	0.397	0.395	0.394	0.394
0.512	0.418	0.417	0.417	0.417
0.565	0.439	0.439	0.439	0.439
0.748	0.556	0.513	0.513	0.513
0.887	0.554	0.561	0.553	0.553
0.887	0.476	0.475	0.474	0.496
0.924	0.478	0.478	0.476	0.495
0.389	0.383	0.383	0.382	0.382
0.413	0.392	0.393	0.392	0.392
0.472	0.437	0.437	0.436	0.436
0.489	0.446	0.444	0.444	0.444
0.481	0.440	0.438	0.437	0.437
0.417	0.380	0.381	0.380	0.380
0.482	0.432	0.436	0.432	0.432
0.491	0.434	0.435	0.433	0.433
0.493	0.433	0.433	0.433	0.433
0.438	0.389	0.390	0.389	0.389
0.484	0.421	0.423	0.421	0.421
0.465	0.416	0.417	0.415	0.415
0.624	0.535	0.538	0.538	0.535
0.599	0.526	0.520	0.519	0.516

0.385	0.378	0.373	0.372	0.372
0.352	0.341	0.340	0.340	0.340
0.432	0.418	0.415	0.415	0.415
0.379	0.368	0.365	0.365	0.364
0.437	0.438	0.419	0.418	0.418
0.376	0.359	0.359	0.359	0.359
0.538	0.501	0.502	0.500	0.500
0.394	0.374	0.372	0.372	0.372
0.495	0.461	0.461	0.459	0.459
0.504	0.505	0.467	0.466	0.466
0.467	0.435	0.436	0.434	0.434
0.443	0.414	0.411	0.411	0.411
0.492	0.448	0.445	0.445	0.444

附录 3				
高钾类型关联系数结果				
高钾类型关联系数结果				
项	氧化钾(K <sub>2</sub> O)	氧化钠(Na <sub>2</sub> O)	氧化钙(CaO)	氧化铅(PbO)
第 1 项	0.847	0.847	0.909	0.847
第 2 项	0.915	0.830	0.883	0.830
第 3 项	0.904	0.818	0.864	0.818
第 4 项	0.926	0.835	0.890	0.835
第 5 项	0.940	0.849	0.920	0.849
第 6 项	0.847	0.847	0.906	0.847
第 7 项	0.902	0.823	0.880	0.823
第 8 项	0.956	0.856	0.921	0.856
第 9 项	0.793	0.732	0.732	0.732
第 10 项	0.718	0.689	0.700	0.691
第 11 项	0.875	0.798	0.845	0.798
第 12 项	0.956	0.844	0.896	0.829
第 13 项	0.914	0.881	0.863	0.865
第 14 项	1.000	0.903	0.959	0.878
第 15 项	0.919	0.881	0.854	0.856
第 16 项	0.960	0.877	0.921	0.862
第 17 项	0.941	0.871	0.919	0.874
第 18 项	0.867	0.810	0.810	0.811
第 19 项	0.749	0.749	0.779	0.755
第 20 项	0.969	0.855	0.906	0.867
项	氧化钡(BaO)	氧化镁(MgO)	氧化铝(Al <sub>2</sub> O <sub>3</sub> )	氧化铁(Fe <sub>2</sub> O <sub>3</sub> )
第 1 项	0.847	0.847	0.903	0.862
第 2 项	0.830	0.830	0.876	0.848
第 3 项	0.818	0.818	0.856	0.834
第 4 项	0.835	0.835	0.883	0.853
第 5 项	0.849	0.860	0.925	0.869
第 6 项	0.847	0.857	0.909	0.862
第 7 项	0.823	0.835	0.874	0.838
第 8 项	0.856	0.871	0.922	0.878
第 9 项	0.732	0.741	0.750	0.732
第 10 项	0.689	0.689	0.711	0.689
第 11 项	0.798	0.804	0.826	0.810
第 12 项	0.828	0.832	0.878	0.831
第 13 项	0.863	0.871	0.863	0.872
第 14 项	0.878	0.878	0.934	0.903
第 15 项	0.854	0.862	0.880	0.862
第 16 项	0.849	0.854	0.930	0.853
第 17 项	0.879	0.886	0.965	0.925
第 18 项	0.820	0.824	0.900	0.827
第 19 项	0.761	0.756	0.789	0.764
第 20 项	0.879	0.864	0.902	0.873
氧化铜(CuO)	五氧化二磷(P <sub>2</sub> O <sub>5</sub> )	氧化锶(SrO)	氧化锡(SnO <sub>2</sub> )	二氧化硫(SO <sub>2</sub> )
0.938	0.888	0.847	0.847	0.847
0.866	0.851	0.830	0.830	0.830

0.834	0.818	0.818	0.818	0.818
0.875	0.842	0.835	0.835	0.835
0.858	0.861	0.849	0.849	0.849
0.884	0.869	0.847	0.847	0.847
0.839	0.829	0.823	0.823	0.825
0.884	0.864	0.857	0.856	0.860
0.732	0.740	0.732	0.746	0.732
0.693	0.693	0.689	0.689	0.689
0.826	0.806	0.798	0.798	0.800
0.836	0.828	0.828	0.828	0.828
0.873	0.865	0.863	0.863	0.863
0.920	0.889	0.878	0.878	0.878
0.865	0.856	0.854	0.854	0.854
0.853	0.850	0.849	0.849	0.849
0.890	0.911	0.872	0.871	0.871
0.828	0.841	0.810	0.810	0.810
0.770	0.756	0.749	0.749	0.749
0.899	0.861	0.856	0.855	0.855

## 附录 4

### K-means 聚类代码

```
import numpy as np
import matplotlib.pyplot as plt

# 两点距离
def distance(e1, e2):
    return np.sqrt((e1[0]-e2[0])**2+(e1[1]-e2[1])**2)

# 集合中心
def means(arr):
    return np.array([np.mean([e[0] for e in arr]), np.mean([e[1] for e in arr])])

# arr 中距离 a 最远的元素，用于初始化聚类中心
def farthest(k_arr, arr):
    f = [0, 0]
    max_d = 0
    for e in arr:
        d = 0
        for i in range(k_arr.__len__()):
            d = d + np.sqrt(distance(k_arr[i], e))
        if d > max_d:
            max_d = d
            f = e
    return f

# arr 中距离 a 最近的元素，用于聚类
def closest(a, arr):
    c = arr[1]
    min_d = distance(a, arr[1])
    arr = arr[1:]
    for e in arr:
        d = distance(a, e)
        if d < min_d:
            min_d = d
            c = e
    return c

if __name__=="__main__":
    df = pd.read_csv('全部转换完的未风化高钾加铅钼.xlsx')
    arr = df[11:14]

    ## 初始化聚类中心和聚类容器
    m = 5
    r = np.random.randint(arr.__len__() - 1)
    k_arr = np.array([arr[r]])
    cla_arr = [[]]
    for i in range(m-1):
        k = farthest(k_arr, arr)
        k_arr = np.concatenate([k_arr, np.array([k])])
        cla_arr.append([])

    ## 迭代聚类
    n = 20
    cla_temp = cla_arr
```

```
for i in range(n): # 迭代 n 次
    for e in arr: # 把集合里每一个元素聚到最近的类
        ki = 0 # 假定距离第一个中心最近
        min_d = distance(e, k_arr[ki])
        for j in range(1, k_arr.__len__()):
            if distance(e, k_arr[j]) < min_d: # 找到更近的聚类中心
                min_d = distance(e, k_arr[j])
                ki = j
        cla_temp[ki].append(e)
# 迭代更新聚类中心
for k in range(k_arr.__len__()):
    if n - 1 == i:
        break
    k_arr[k] = means(cla_temp[k])
    cla_temp[k] = []

## 可视化展示
col = ['HotPink', 'Aqua', 'Chartreuse', 'yellow', 'LightSalmon']
for i in range(m):
    plt.scatter(k_arr[i][0], k_arr[i][1], linewidth=10, color=col[i])
    plt.scatter([e[0] for e in cla_temp[i]], [e[1] for e in cla_temp[i]], color=col[i])
plt.show()
```



## 附录 5

### 灰色关联分析算法代码

```
#导入相关库
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
# 解决图标题中文乱码问题
import matplotlib as mpl
mpl.rcParams['font.sans-serif'] = ['SimHei'] # 指定默认字体
mpl.rcParams['axes.unicode_minus'] = False # 解决保存图像是负号'-'显示为方块的问题
#导入数据
data=pd.read_excel(r'D:\问题 C\标准 高钾 训练集.xlsx')
# print(data)
#提取变量名 x1 -- x7
label_need=data.keys()[1:]
# print(label_need)
#提取上面变量名下的数据
data1=data[label_need].values
print(data1)
#0.002~1 区间归一化
[m,n]=data1.shape #得到行数和列数
data2=data1.astype('float')
data3=data2
ymin=0.002
ymax=1
for j in range(0,n):
    d_max=max(data2[:,j])
    d_min=min(data2[:,j])
    data3[:,j]=(ymax-ymin)*(data2[:,j]-d_min)/(d_max-d_min)+ymin
print(data3)
# 绘制 x1,x4,x5,x6,x7 的折线图
t=range(2007,2014)
plt.plot(t,data3[:,0], '*-',c='red')
for i in range(4):
    plt.plot(t,data3[:,2+i], '-.')
plt.xlabel('year')
plt.legend(['x1','x4','x5','x6','x7'])
plt.title('灰色关联分析')
# 得到其他列和参考列相等的绝对值
for i in range(3,7):
    data3[:,i]=np.abs(data3[:,i]-data3[:,0])
#得到绝对值矩阵的全局最大值和最小值
data4=data3[:,3:7]
d_max=np.max(data4)
d_min=np.min(data4)
a=0.5 #定义分辨系数
# 计算灰色关联矩阵
data4=(d_min+a*d_max)/(data4+a*d_max)
xishu=np.mean(data4, axis=0)
print(' x4,x5,x6,x7 与 x1 之间的灰色关联度分别为: ')
print(xishu)
```

## 附录 6

### 相关系数代码

```
#导入依赖库
import numpy as np
import pandas as pd

data = pd.read_excel(r'D:\问题 C\标准 高钾 训练集.xlsx')

data.corr('spearman') # spearman 秩相关
```

## 附录 7

### 鉴别分析代码

```
import xlrd as xr
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis as LDA
import numpy as np
from sklearn.neighbors import KNeighborsClassifier

##导入数据
# 类别数据
file_location = "/Users/lifangjian/Desktop/本科课程资料/数学建模/判别分析案例.xls"
data = xr.open_workbook(file_location)
sheet = data.sheet_by_index(0)
lbstats = [[sheet.cell_value(r, c) for c in range(2, sheet.ncols)] for r in range(1, sheet.nrows)]
lbstats_Frame = pd.DataFrame(lbstats)
print(lbstats_Frame)
# 样本数据
sheet1 = data.sheet_by_index(1)
ybstats = [[sheet1.cell_value(r, c) for c in range(1, sheet1.ncols)] for r in range(1, sheet1.nrows)]
ybstats_Frame = pd.DataFrame(ybstats)
print(ybstats_Frame)

# 数据堆砌
X0 = np.array(lbstats_Frame)
x = np.array(ybstats_Frame)
g = np.hstack([np.ones(5), 2 * np.ones(5)])

##计算类间协方差及其转置矩阵
```

```

l1 = np.vstack(X0[0:5])
l2 = np.vstack(X0[5:])
v11 = np.cov(l1.T)
v12 = np.cov(l2.T)

##计算总体协方差矩阵以及转置矩阵
v1 = np.cov(X0)
v = np.cov(X0.T)

# 计算类内均值
l1_mean = l1.mean(axis=0)
l2_mean = l2.mean(axis=0)

# 判断行数(一类中的个数)
n1 = l1.shape[0]
n2 = l2.shape[0]
# 计算类内协方差,rowvar=False 代表计算方式为列, 每一列称之为一个变量
s1 = np.cov(l1, rowvar=False)
s2 = np.cov(l2, rowvar=False)
# 计算联合协方差
s = ((n1 - 1) * s1 + (n2 - 1) * s2) / (n1 + n2 - 2)

##计算系数和常数项
# 矩阵求逆 np.linalg.inv(s)
# np.dot 代表矩阵相乘
B1 = np.array(l1_mean).reshape(1, 3).dot(np.linalg.inv(s))[0]
print(B1)
# 就是  $\hat{y}=0$ ,  $\hat{x}=\text{li.mean}$ ,  $d=(\hat{y}-\hat{x}*B)/2$ 
d1 = 0
for i in range(len(B1)):
    d1 -= l1_mean[i] * B1[i]
d1 = d1 / 2
print("y1 的 Baye 判别函数为:  $y1=\{ \}+\{ \}x1+\{ \}x2+\{ \}x3$ ".format(d1, B1[0], B1[1], B1[2]))

B2 = np.array(l2_mean).reshape(1, 3).dot(np.linalg.inv(s))[0]
print(B2)
d2 = 0
for i in range(len(B2)):
    d2 -= l2_mean[i] * B2[i]
d2 = d2 / 2
print("y2 的 Baye 判别函数为:  $y2=\{ \}+\{ \}x1+\{ \}x2+\{ \}x3$ ".format(d2, B2[0], B2[1], B2[2]))

```