

# 古代玻璃制品的成分分析与鉴别

## 摘 要

古代玻璃极易受埋藏环境的影响而风化。本文通过定性定量分析而确定了玻璃的种类及其化学成分；分析了因风化所导致的成分变化；初步探讨了风化机理。

对于问题一，首先由于玻璃类型、纹饰和颜色都是离散值，故对其进行量化处理，将它们对玻璃文物表面风化情况分别进行差异性分析，通过卡方检验得到风化情况与类型和颜色均显著相关，风化情况与纹饰无显著相关性。其次通过对不同类型玻璃的物理指标结合表单 1 和表单 2 进行分组列出表格，比较每组的化学成分含量得到统计规律：铅钡玻璃中化学成分的变化范围较大，相比于高钾类玻璃铅钡类玻璃化学成分流失率更大。最后将两类玻璃风化前后的化学成分含量按照其在标准正态曲线上的位置，分配权重，就得到加权平均值，作为风化前的值。

对于问题二，通过观察附件中表单 2 的数据可以得到两类玻璃是根据氧化铅和氧化钡的含量进行分类的；为了对两类玻璃进行亚类划分，建立随机森林算法模型，使用 SPSSPRO 求解绘制出柱形图，将各化学成分的重要性可视化，筛选 5 种指标，加和为 90% 以上，根据两种玻璃中 5 种指标的含量，使用个案排秩法进行综合排名，将每类玻璃都分成 5 组。最后通过绘制混淆矩阵进行合理性分析，得到训练集与测试集都被完美预测，效果极好；并将表单 2 中的数据赋予上下 10% 的扰动，重新对数据进行预测分类并和实际分类进行比较，发现给予扰动后的分类准确度在 98.3%。

对于问题三，要预测表单 3 中未知类别玻璃文物的所属类别，根据其化学成分和表单 2 中两种类别玻璃的化学成分建立 logistic 回归预测模型，初步计算得出模型的预测结果并不理想，在 logistic 模型中加入交互平方项，再次对数据进行预测，出现过拟合现象，为了消除过拟合现象，采用 Fisher 线性判别法重新对数据进行预测。分组结果为：高钾类玻璃有 A1, A6, A7；铅钡类玻璃有 A2, A3, A4, A5, A8；最后将表单 3 中的数据赋予上下 10% 的扰动进行敏感度分析，重新对数据进行预测分类并和实际分类进行比较，发现给予扰动后的分类准确度在 98.5%。

对于问题四，分别将高钾类玻璃样品中的氧化钾和铅钡类玻璃样品中的氧化铅和氧化钡作为母序列，其他化学成分作为子序列进行三次灰色关联度分析；通过观察比较三次灰色关联度分析的关联程度表，得到各类化学成分对于铅钡类玻璃的关联程度更高，对于高钾类玻璃的关联程度较低；其中氧化钙和氧化铜对两类玻璃都有明显的关联性。

**关键词：** 随机森林   logistic 回归   Fisher 线性判别法   灰色关联度分析

## 一、问题重述

丝绸之路是我国古代中西方文化交流的方式，这其中玻璃是早期的时候经济文化交流的具有代表性证物。玻璃的主要原材料是石英砂，主要化学成分是二氧化硅。由于纯石英砂的熔点过高，为了降低其熔化温度，在炼制时一般要添加使用助熔剂。其中草木灰、天然泡碱、硝石和铅矿石等是古时最常见的助熔剂，并添加石灰石作为稳定剂，石灰石经过煅烧可以转化为氧化钙。增加的助熔剂不同，其主要化学成分也不同。例如，铅钡玻璃在烧制过程中加入铅矿石作为一个助熔剂，其氧化铅、氧化钡的含量相对较高，基本上被认为是我们国家自行研制出的玻璃品种，楚文化的玻璃正是以铅钡玻璃为主体。钾玻璃是以含钾量高的物质作为一个助熔剂烧制而成的如草木灰，主要流行于我国岭南乃至东南亚和印度地区。古代玻璃极易受埋藏环境的影响而风化。在风化过程中，玻璃内部的元素与环境的元素通过大量交换，导致了其成分比例出现了变化，进而影响对其种类的准确判断。有些文物标记为表面无风化，但不排除个别有较浅的风化；有些文物标记为表面风化，在部分风化的文物中，其表面也有未风化的区域。

附件中给出一批我国古代玻璃制品的化学成分，考古工作者凭借这些文物样品的化学成分和除此以外检测手段将其主要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。现依据附件中的相关数据进行分析建模，解决以下问题：

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。

问题 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。

问题 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。

问题 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 问题一的分析

针对问题一，需要对玻璃的表面风化与玻璃类型、纹饰和颜色这三个物理指标的关系进行分析，考虑对其进行差异性分析，通过观察表单一可以发现这三个指标均为离散值，因此应将其进行量化，根据卡方检验，得出它们的显著性关系；再通过将各物理指标排列组合，并结合表单 2 中的化学成分含量，列出表格，根据风化的化学反应机理分析出玻璃的化学成分含量统计规律；最后将两类玻璃风化前后的化学成分含量进行正态分布概率密度函数的计算，将“中心点”作为原点，其他点按照其在正态曲线上的位置，分配权重，就得到加权平均值，作为风

化前的各化学成分的值。

## 2.2 问题二的分析

针对问题二，依据附件数据分析高钾玻璃、铅钡玻璃的分类规律，通过观察附件中表单 2 的数据，对比两类玻璃在哪些化学成分含量有明显的分层变化，并将其作为两类玻璃的分类依据；由于要根据化学成分进行亚类划分，考虑使用机器学习中的随机森林模型，得到各特征（自变量）的重要性，筛选出对化学成分含量影响最大的指标，通过两种玻璃的这些指标含量的多少，使用个案排秩进行综合排名，根据分数高低，将每类玻璃进行合适的分组。最后通过绘制观察混淆矩阵作为模型评价进行合理性分析；将表单 2 中的数据赋予上下 10%的扰动进行敏感性分析，重新对数据进行预测分类，对比预测分类和实际分类的差异。

## 2.3 问题三的分析

针对问题三，要预测表单 3 中未知类别玻璃文物的所属类别，考虑建立 logistic 回归预测模型，根据表单 2 中两种类别玻璃的化学成分来预测未知类别的 8 组数据，如果求解过程中发现模型的预测结果并不理想，还应考虑在 logistic 模型中加入交互平方项，再次对数据进行预测，但一般会出现过拟合现象，为了消除过拟合现象采用 Fisher 线性判别法重新对数据进行预测。最后将表单 3 中的数据赋予上下 10%的扰动进行敏感性分析，重新对数据进行预测分类，对比预测分类和实际分类的差异。

## 2.4 问题四的分析

针对问题四，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，考虑使用灰色关联度分析，将高钾类玻璃样品中的氧化钾作为母序列，其他化学成分作为子序列进行灰色关联度分析；同理将铅钡类玻璃样品中的氧化铅和氧化钡作为母序列，其他化学成分作为子序列分别进行灰色关联度分析。通过将高钾类玻璃和铅钡类玻璃的各化学成分与母序列的关联程度进行比较，对不同类别之间的化学成分关联关系的差异性进行分析。

# 三、模型假设

- (1) 假设所有未检测到成分极少，可近似看为 0；
- (2) 假设附件所给数据均真实有效；
- (3) 假设化学成分在测量过程中不会再发生变化；
- (4) 假设颜色空白处为白色。

# 四、符号说明

符号	符号描述
$A$	实际值
$T$	理论值
$X^2$	理论与实际的差异程度
$H_0$	玻璃是否风化分布律
$n$	试验次数
$N$	玻璃样本数
$M$	化学成分指标总数

$m_i$	化学成分指标数
$X^i$	玻璃样本中第 <i>i</i> 个样本
$\hat{y}$	预测分类指标
$\hat{\beta}_i$	逻辑回归叙述
$p$	分辨系数
$\xi_i$	灰色关联系数
$\eta$	灰色关联度
$X_i$	玻璃类别 <i>i</i> 的集合
$U_i$	均值向量
$\sum_i \omega$	协方差矩阵

## 五、模型建立与求解

### 5.1 问题一模型的建立与求解

对玻璃的表面风化与玻璃类型、纹饰和颜色这三个物理指标的关系进行分析，考虑对其进行差异性分析，通过观察表单一可以发现这三个指标均为离散值，因此应将其进行量化，根据卡方检验，得出它们的显著性关系；再通过将各物理指标排列组合，并结合表单 2 中的化学成分含量，列出表格，根据风化的化学反应机理分析出玻璃的化学成分含量统计规律；最后将两类玻璃风化前后的化学成分含量进行正态分布概率密度函数的计算，将“中心点”作为原点，其他点按照其在正态曲线上的位置，分配权重，就得到加权平均值，作为风化前的各成分值。

#### 5.1.1 问题一中数据的预处理

将附件表单 1 中的纹饰、类型颜色、表面风华分别量化，如表 1：

表 1 分类量化表

类型	量化
纹饰	A=1, B=2, C=3
类型	高钾=0, 铅钡=1
颜色	空白=0, 黑=1, 蓝绿=2, 绿=3 浅蓝=4, 浅绿=5, 深蓝=6, 深绿=7 紫=8
风化情况	风化=1, 无风化=0

根据合理假设将表单 2 中的空值均填为 0；且本题要求将成分比例累加和介于 85%~105%之间的数据视为有效数据，经过计算应剔除第 15 和 17 组样本。

#### 5.1.2 卡方检验进行差异性分析

卡方检验<sup>[1]</sup>，是使用范围十分广的一种假设检验方法，它在分类数据统计推断中的广泛应用，涵盖多个率或多个共同构成比对比的卡方检验以及分类数据的相应分析等。

$$CHI(x, y) = X^2(x, y) = \sum \frac{(A - T)^2}{T} \quad (1)$$

其中  $A$  是实际值,  $T$  是理论值  $X^2$  是衡量理论与实际的差异程度。

### 5.1.2 卡方检验的求解和结果分析

(1) 提出原假设:

$H_0$ : 总体  $X$  玻璃样本数的分布函数为  $F(x)$ 。

由于玻璃是否分化的值为离散型, 则假设总体  $X$  的分布律为

$$P\{X = x_i\} = p_i, i = 0, 1$$

(2) 将总体  $X$  玻璃样本数的取值范围分成  $k$  个互不相交的小区间  $A_1, A_2, A_3, \dots, A_k$ , 如可取  $A_1 = (a_0, a_1], A_2 = (a_1, a_2], \dots, A_k = (a_{k-1}, a_k)$  其中  $a_0$  可取  $-\infty$ ,  $a_k$  可取  $+\infty$ , 区间的分类视实际需求而定, 但要使各个小区间所含的样本值个数不低于 5, 而区间个数  $k$  不要太大也千万不要太小。

(3) 把落入第  $i$  个小区间的  $A_i$  的样本值的个数记作  $f_i$ , 成为组频数(真实值), 将所有组频数之和  $f_1 + f_2 + \dots + f_k$  等于样本容量  $n$ 。

(4) 当  $H_0$  为真时, 参照所假设的整体理论分布, 可计算出总体  $X$  的值落入第  $i$  个小区间  $A_i$  的概率  $p_i$ , 于是,  $np_i$  就是落入第  $i$  个小区间  $A_i$  的样本值的理论频数(理论值)。

(5) 当  $H_0$  为真时,  $n$  次试验中样本值落入第  $i$  个小区间  $A_i$  的频率  $f_i/n$  与概率  $p_i$  应很接近, 当  $H_0$  不真时, 则  $f_i/n$  与  $p_i$  相差很大。基于这种思想, 皮尔逊引进如下检验统计量

$$X^2 = \sum_{i=1}^k \frac{(f_i - np_i)^2}{np_i} \quad (2)$$

在  $H_0$  假定成立的情况下服从自由度为  $k-1$  的卡方分布。

使用 SPSS 得到如下卡方检验结果, 见表 2:

表 2 卡方检验

		皮尔逊卡方	线性关联
类型	值	6.880 <sup>a</sup>	6.762
	显著性	0.009	0.009
颜色	值	25.683 <sup>a</sup>	25.569
	显著性	0.000	0.000
纹饰	值	11.023 <sup>a</sup>	2.996
	显著性	0.200	0.083

通过上表对卡方检验的结论进行分析, 对于类型和颜色来说, 显著性数值分别为 0.009 和 0.000 均都小于 0.05, 因此有显著性差异, 不能接受无关假设, 即风化情况与类型和颜色均明显相关, 但纹饰的显著性数值为 0.200, 显然大于 0.05, 故风化情况与纹饰无明显相关性。

### 5.1.3 化学成分的统计规律

参照附件表单 2 中的化学成分含量, 按照三个物理指标进行排列组合进行统计与划分从而得到以下结果, 见表 3 和表 4:

表 3 高钾类玻璃风化情况分组

	纹饰	表面风化	组号	颜色	化学成分含量
高钾	A	风化	无	无	无
		无风化	3、4、5、6、18、21	蓝绿、深蓝	96.06%~100%
	B	风化	7、9、10、12、22、27	蓝绿	98.81%~99.81%
		无风化	无	无	无
	C	风化	无	无	无
		无风化	1、13、14、15、16、17	蓝绿、浅蓝、深绿	97.61%~99.00%

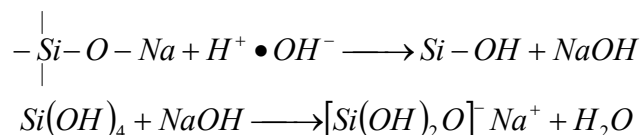
由上表可以看出在高钾类玻璃中纹饰为 A 时样本均无风化，说明 A 类纹饰为对于高钾玻璃的保护性良好，且化学成分的含量变化在 4%左右；纹饰为 B 时，样本全部风化说明 B 类纹饰对于高钾类玻璃的保护性极差，且其化学成分的含量变化在 1%左右；纹饰为 C 时样本均无风化，说明纹饰 C 对于高钾类玻璃的保护性良好，化学成分含量变化在 1.5%左右。通过对其化学成分含量变化可以发现高钾类玻璃化学成分的流失情况并不明显。

表 4 铅钡类玻璃风华情况分组

	纹饰	表面风化	组号	颜色	化学成分含量
铅钡	A	风化	2、19、23、28、29、42、44、48、49、50、53	浅蓝、空白、蓝绿、黑	90.17%~100.00%
		无风化	20、30、45、46、47	浅蓝、深蓝	88.41%~98.7%
	C	风化	8、11、25、26、34、36、38、39、40、41、43、51、52、54、56、57、58	紫、浅蓝、深绿、浅绿、空白、蓝绿	91.70%~99.89%
		无风化	31、32、33、35、37、55	紫、浅绿、绿、深绿	96.21%~99.98%

由上表可以看出在铅钡类玻璃中并无 B 类纹饰，说明该种类型不适合 B 类纹饰的制造；在 A 类纹饰中均存在风化和无风化两种样本，其中风化样本的数目明显大于无风化样本的数目，并且两种样本的化学成分的含量变化都在 10%左右，说明 A 类纹饰在铅钡玻璃中的保护效果一般，无法很好地控制化学成分的流失；纹饰为 C 类时风化玻璃的样本明显增多说明 C 类文是对于铅钡玻璃的表述性不好且其化学含量的变化在 9%左右而无风化玻璃的化学含量变化在 3%左右说明 C 类玻璃对于铅钡类玻璃化学成分的流失性控制良好。

按照玻璃水解原理<sup>[2]</sup>，风化显微形貌，化学成分的变化等，可假想以下风化模式，水解氢化反应，对风化有着至关重要的作用，钠、钾玻璃极为突出。可以使用：



这样的化学方程式来描述。

经水解氢化反应后的玻璃表面，其  $Si-O$  键、或其他键价结合之间的关系受到破坏，使玻璃表面上的结构出现了变化。因所处环境、水文地理条件的变化乃至玻璃化学成分的不相同，因此呈现各不相同的风化形貌。

由于铅玻璃在  $Si-O$  键断裂后，可能有  $Pb-O$  或  $Pb^+$  出现。因地层中存在着多种酸根，它们长期与  $Pb-O$  或  $Pb^+$  作用，一部分酸根逐渐富集，一部分酸根则渐渐流失，这样就会导致  $PbCO_3$  或  $Pb_3(CO_3)_2(OH)_2$  等结晶物析出或铅的含量减少。这很好地解释了上述分析中化学成分的流失情况。

#### 5.1.4 风化前的化学含量预测

观察附件表单 2 中数据，发现数据量小并且数据量为 0 较多，所以为了提高预测准确度，本题使用正态分布加权平均数来处理风化前和风化后的化学成分。正态分布显然是一种可取的权重分配模式。

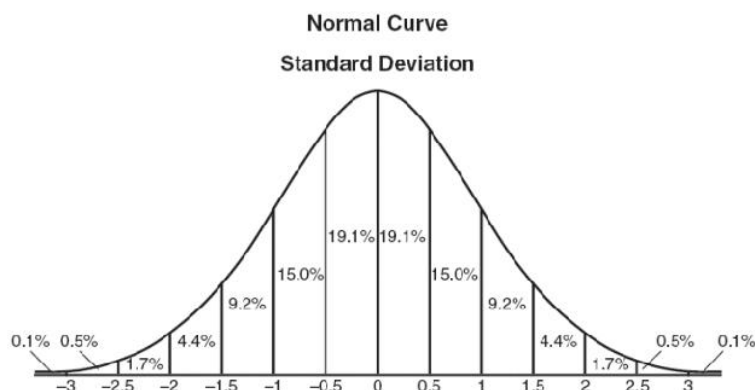


图 1 权重分配模式

在图形上，正态分布是一种钟形曲线，越接近中心，取值越大，越远离中心，取值越小。将附件表单 2 中数据的初步处理：将玻璃类别和风化表面分为高钾风化、高钾无风化、铅钡风化、铅钡无分化四组；将“中心点”作为原点，其他点按照其在正态曲线上的位置，分配权重，就可以得到一个加权平均值，即为风化前的值。通过 matlab（代码详见附录一）将四组数据进行正态分布概率密度函数的计算；计算结果如表 5：

表 5 化学成分预测表

	高钾风化	高钾无风化	铅钡风化	铅钡无风化
二氧化硅( $SiO_2$ )	93.9214	66.578	33.5345	54.6061
氧化钠( $Na_2O$ )	0	0.4247	0.5612	0.5039
氧化钾( $K_2O$ )	0.4217	9.3299	0.1197	0.2089
氧化钙( $CaO$ )	0.6709	5.3478	2.1678	1.0389
氧化镁( $MgO$ )	0.0491	1.0547	0.6373	0.4639
氧化铝( $Al_2O_3$ )	1.4937	6.3477	3.3613	3.2031
氧化铁( $Fe_2O_3$ )	0.2382	1.7726	0.4383	0.7174
氧化铜( $CuO$ )	1.587	2.3159	1.7223	0.7201
氧化铅( $PbO$ )	0	0.2962	35.4302	21.9275
氧化钡( $BaO$ )	0	0.3923	9.5842	8.6457
五氧化二磷( $P_2O_5$ )	0.2566	1.1741	3.4867	0.6719
氧化锶( $SrO$ )	0	0.034	0.3524	0.2094
氧化锡( $SnO_2$ )	0	0.0884	0.0251	0.0376
二氧化硫( $SO_2$ )	0	0.0634	0.4576	0.1371
总含量	99.522	98.1636	97.3764	97.581

## 5.2 问题二的模型的建立与求解

依据附件数据分析高钾玻璃、铅钡玻璃的分类规律，通过观察附件中表单 2 的数据，以及第一问两类玻璃成分含量柱状图，对比两类玻璃在哪些化学成分含量有明显的分层变化，并将其作为两类玻璃的分类依据；由于要根据化学成分进行亚类划分，考虑使用机器学习中的随机森林模型，得到各化学成分的重要性，筛选出对化学成分含量影响最大的指标，通过两种玻璃的这些指标含量的多少，使用个案排秩进行综合排名，根据分数高低，将每类玻璃进行合适的分组。最后通过绘制观察混淆矩阵进行合理性分析；最后将表单 2 中的数据赋予上下 10% 的扰动进行敏感性分析，重新对数据进行预测分类，对比预测分类和实际分类的差异。

### 5.2.1 两类玻璃的分类规律分析

通过观察表单 2 中的数据可以发现高钾类玻璃的氧化铅和氧化钡两个化学指标数据缺失较多并且有数据的样本含量也很少，而铅钡类玻璃这两种成分的含量很高，并且缺失值很少，高钾类玻璃出现了明显的断层；

将两类玻璃风化前和风化后各化学成分含量求平均值绘制出柱状图如图 2，图 3：



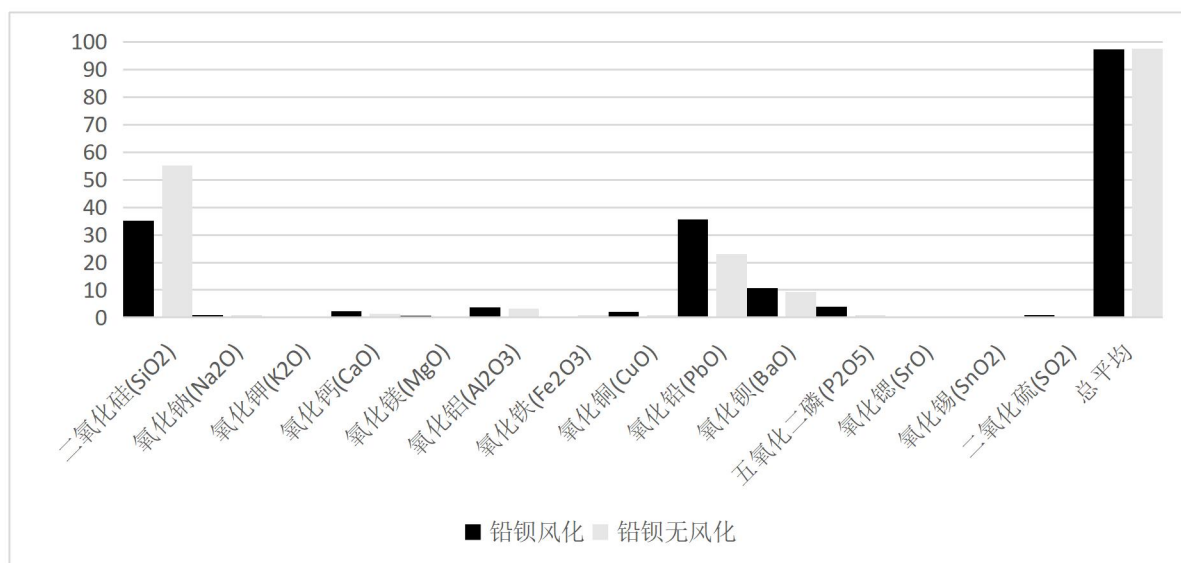


图 2 铅钡风化前成分对比

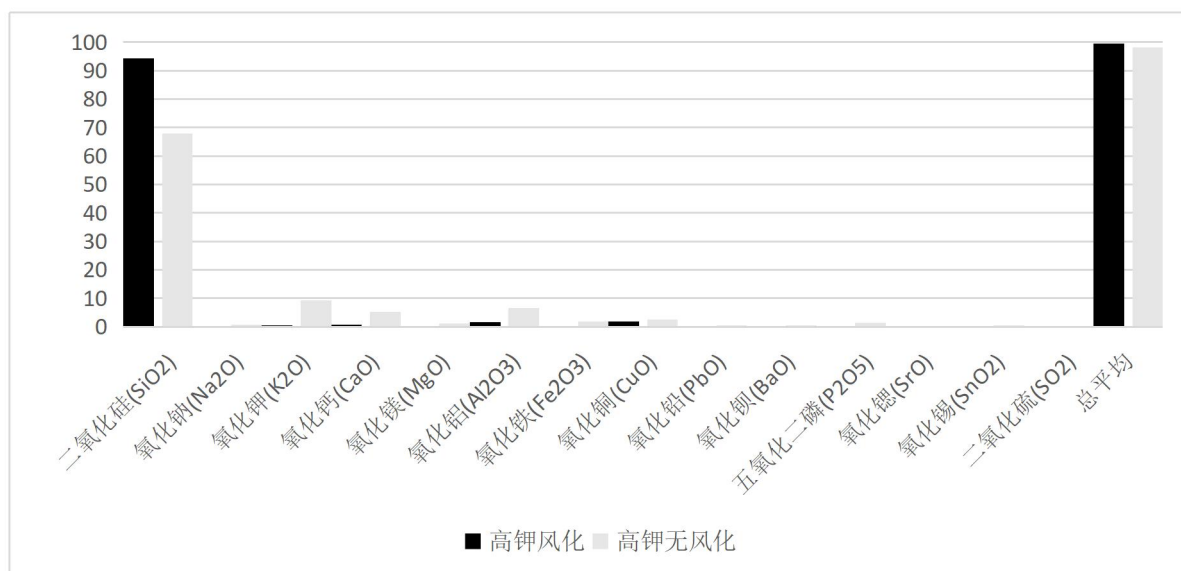


图 3 高钾风化前成分对比

通过观察图 2，图 3 可以发现铅钡类玻璃氧化铅和氧化钡的含量明显高于高钾类玻璃，而且高钾类玻璃中的二氧化硅的含量明显高于铅钡类玻璃，结合上述分析，可以得出，两类玻璃的划分依据是根据氧化铅、氧化钡以及氧化硅的含量进行划分：氧化硅的含量相对较少氧化铅氧化钡含量相对较高的为铅钡类玻璃反之则为高钾类玻璃。

### 5.2.2 随机森林模型的建立

1. 假如有  $N$  个玻璃样本，则有放回的随机选择  $N$  个玻璃样本(每次随机选择一个玻璃样本，然后返回继续选择)。这选择好了的  $N$  个玻璃样本用来训练一个决策树，作为决策树根节点处的样本。

2. 当每个玻璃样本有  $M$  个化学成分时，在决策树的每个节点需要分裂时，随机从这  $M$  个属性中选取  $m$  个化学成分，满足条件  $m \ll M$ 。然后从这  $m$  个属性中采用某种策略(比如说信息增益)来选择 1 个化学成分作为该节点的分裂属性。

3. 决策树形成过程中每个节点都要按照步骤 2 来分裂（很容易理解，如果下一次该节点选出来的那一个化学成分是刚刚其父节点分裂时用过的化学成分，则该节点已经达到了叶子节点，无须继续分裂了）一直到不能够再分裂为止。注意整个决策树形成过程中没有进行剪枝。
4. 按照步骤 1~3 建立大量的决策树，这样就构成了随机森林。
- 在建立每一棵决策树的过程中，有两点需要注意采样与完全分裂。

### 5.2.3 模型的求解和结果分析

具体操作过程如图 4：

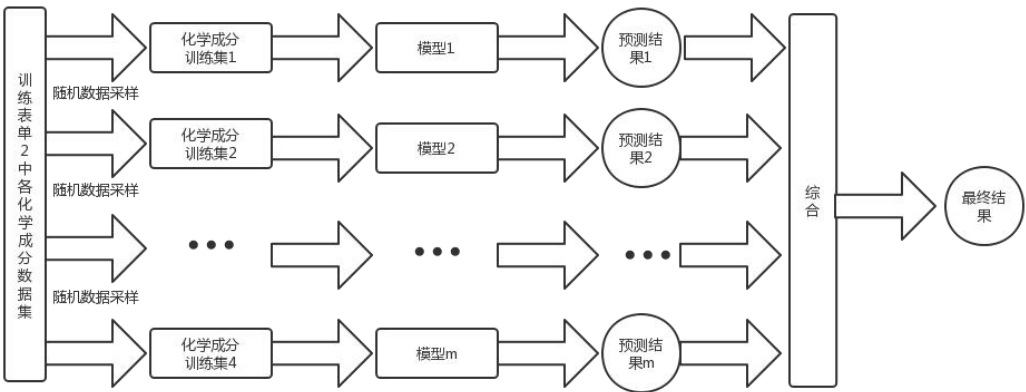


图 4 随机森林预测的构建

使用 SPSSPRO 对随机森林模型进行求解，得到各特征（自变量）的重要性比例柱形图，见图 5：

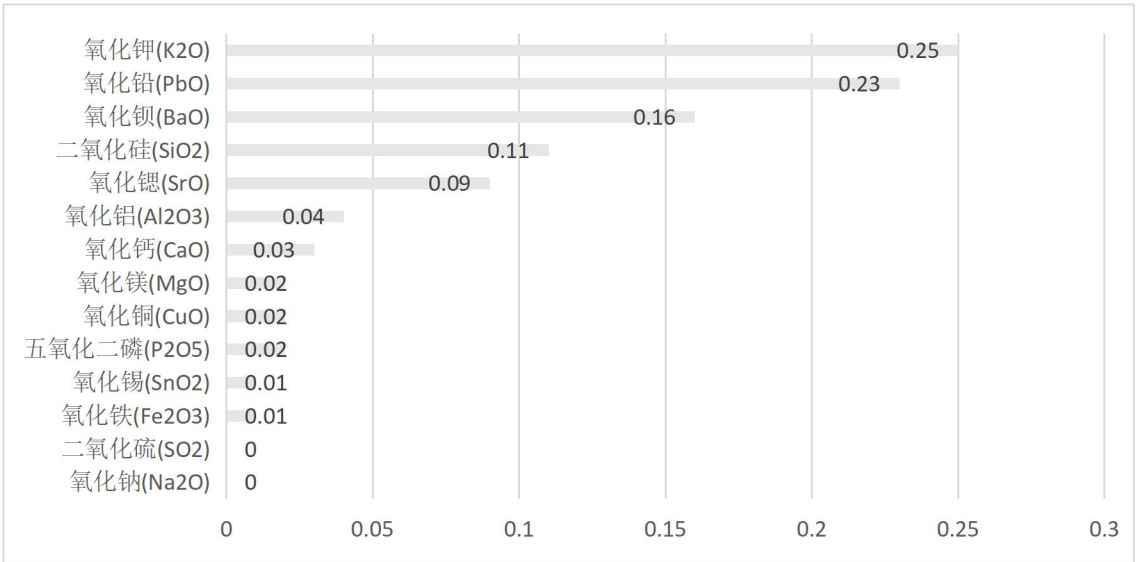


图 5 化学成分的重要性比例柱形图

这里选取二氧化硅、氧化铅、氧化钾、氧化锶和氧化钡这 5 种比较重要的化学成分作为分类指标，这 5 种化学成分的重要性相加大于 90%，可以近似的代表每个样本的化学成分总量。

### 5.2.4 基于个案排秩法的亚类划分结果

由于要根据合适的化学成分进行亚分类，这里采用上述选取的 5 种化学成分

的相对含量进行排序分组，但由于每组样本中的含量有所不同，无法根据单个化学成分的含量进行排序，因此采取个案排秩法进行综合排序，下表 6 仅展示一部分样品的得分情况详见附录二。

表 6 个案排秩得分表

高钾类玻璃		铅钡类玻璃					
组号	分数	组号	分数	组号	分数	组号	分数
3	24	47	87	45	123	44	145
6	32	54	88.5	40	124	32	145
6	36	30	88.5	8	125	43	147.5
16	40	26	92	52	125	22	148
14	43	41	96.5	26	126	25	153
21	44	46	96.5	43	126	50	153.5

在 SPSS 中“个案排秩”的对话框里勾选将秩 1 指定给最大值。如 100 分，那么它的秩就是 1，那么 99 就是 2.000，因此秩值越小排名越靠前。对两类玻璃进行以下亚分类，见表 7，表 8：

表 7 铅钡类玻璃亚分类

划分依据	铅钡类玻璃：组号（按排名顺序）
80~100	47、54、30、26、41、46
101~120	24、34、11、37、30、36、38、42、50、39、48
121~140	2、45、40、8、52、26、43、49、20、23、48、42、50、39、48
141~160	56、29、49、44、32、43、22、25、50、55、33、31
>160	35、19、51

表 8 高钾类玻璃亚分类

划分依据	高钾类玻璃：组号（按排名顺序）
20~30	3
30~40	32、36、40
41~50	14、21、18、3
51~60	5、1、10、15、12、4、9
≥ 60	13、17、27、7

### 5.2.5 合理性和敏感性进行分析

混淆矩阵（Confusion Matrix）也称为可能性矩阵或错误矩阵。混淆矩阵是一种可视化工具，一般是用来监督学习的，在没有监督的学习中，称为匹配矩阵。图像精度评估主要用于比较分类结果和实际测量值，分类结果的精度可以显示在混淆矩阵中。

混淆矩阵的结构一般如下表 9 表示的方法。

表 9 混淆矩阵的结构

	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

混淆矩阵要表达的含义：

混淆矩阵的第一列代表了预测的玻璃类别，每一列的数据总数表示预测为该类别的玻璃样本的数目；

第一行代表了数据的真实归属类别，每一行的数据总数表示该玻璃类别样本实际的数目；

**True Positive (TP)**：真正类。玻璃样本的真实类别是高钾类，并且模型识别的结果也是高钾类。

**False Negative (FN)**：假负类。玻璃样本的真实类别是高钾类，但是模型将其识别为铅钡类。

**False Positive (FP)**：假正类。玻璃样本的真实类别是铅钡类，但是模型将其识别为高钾类。

**True Negative (TN)**：真负类。玻璃样本的真实类别是铅钡类，并且模型将其识别为铅钡类。

该矩阵可用于易于理解的二类分类问题，但通过向混淆矩阵添加更多行和列，可轻松应用于具有 3 个或更多类值的问题。

借助 SPSSPRO 可以得到混淆矩阵如下结果：

表 10 混淆矩阵

测试集	高钾	铅钡	训练集	高钾	铅钡
高钾	10	0	高钾	9	0
铅钡	0	37	铅钡	0	12

上表中展示了训练集和测试集的分类评价指标，通过量化指标来衡量决策树对训练、测试数据的分类效果。由混淆矩阵得到，训练集与测试集都被完美预测，效果极好。

最后将表单 3 中的数据赋予上下 10%的扰动，进行敏感性分析，重新对数据进行预测分类，对比预测分类和实际分类的差异。得到如下结果：

表 11 分类表

分类结果				
		预测组成员信息		总计
		高钾	铅钡	
计数	高钾	19	0	19
	铅钡	1	48	49
	未分组个案	0	0	0
准确率	高钾	100.0	0	100.0
	铅钡	2.0	98.0	100.0
	未分组个案	37.5	62.5	100.0
正确地对 98.3% 个原始已分组个案进行了分类				

### 5.3 问题三模型的建立与求解

要预测表单 3 中未知类别玻璃文物的所属类别，考虑建立 logistic 回归预测模型<sup>[4]</sup>，根据表单 2 中两种类别玻璃的化学成分来预测未知类别的 8 组数据，如果求解过程中发现模型的预测结果并不理想，还应考虑在 logistic 模型中加入交互平方项，再次对数据进行预测，但一般会出现过拟合现象，为了消除过拟合现象采用 Fisher 线性判别法重新对数据进行预测。最后将表单 3 中的数据赋予上下 10% 的扰动进行敏感性分析，重新对数据进行预测分类，对比预测分类和实际分类的差异。

#### 5.3.1 数据预处理

由于玻璃的类型是离散值，将表单 2 和表单 3 中的玻璃类型进行量化，方便后续的预测。高钾类玻璃=1，铅钡类玻璃=0。

#### 5.3.2 逻辑回归模型建立

本问中，将已知玻璃样本的 14 个化学成分作为预测指标，判断另一组未知类别的玻璃样本是高钾类还是铅钡类。针对这个问题，可以先测量  $N$  个玻璃样本的各化学成分和相应高钾、铅钡的两种类型，并将高钾和铅钡分别标记为 1 和 0，来训练这个数据输入模型。训练后，将要分类的玻璃样本的各化学成分输入模型，观察这组玻璃样本是高钾类还是铅钡类。

如果数据是两个指标，则可以用平面上的点表示数据，其中一个为  $x$  轴，另一个为  $y$  轴；如果数据有三个指标，可以用空间中的点表示数据；如果是  $p$  维的话( $p > 3$ )，就是  $p$  维空间中的点。

从本质上来说，逻辑回归训练后的模型是平面的一条直线 ( $p = 2$ )，或是平面 ( $p = 3$ )，超平面 ( $p > 3$ )。并且这条线或平面把空间中的散点分成两半，属于同一类的数据大多数分布在曲线或平面的同一侧。如图 6：

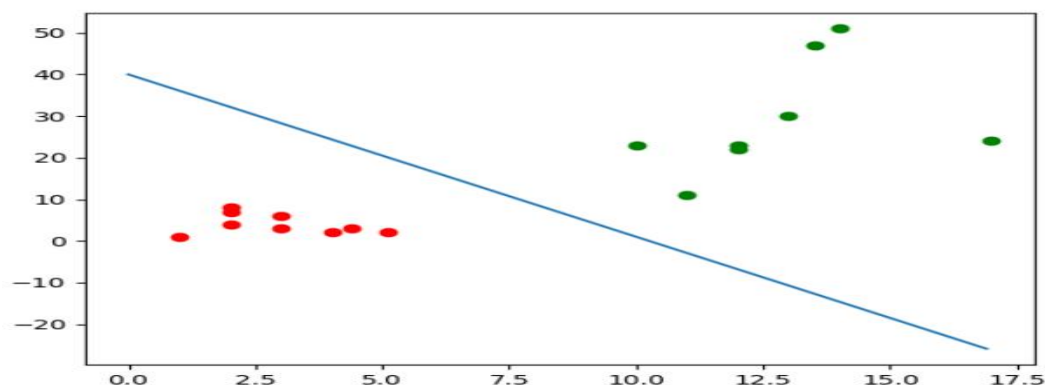


图 6 逻辑回归直线图

如上图 6 所示，点数是玻璃样本数，两种颜色表示两个玻璃类型。这条直线可以看作是分割经过这些样品的训练而获得的样品的直线。对于表单 3 的玻璃样品中各化学成分的值，可以根据这条线确定属于哪种类型。

#### 5.3.3 逻辑回归模型的求解

玻璃分为两类，用 1 表示高钾类，用 0 表示铅钡类。对于输入的每个数据集，需要能够映射到 0 到 1 之间的数字的函数。而且，如果函数值大于 0.5，则判定为 1，否则判定为 0。此外，函数需要待定参数。利用样本训练，可以准确预测培训集中的数据。

这个函数就是西蒙德函数，形式为  $\sigma(x) = \frac{1}{1 + e^{-x}}$ 。所以在这里可以设函数为：

$$h(x^i) = \frac{1}{1 + e^{-(W^T X^i + b)}}$$

这里  $X^i$  是表单 2 中所有玻璃样本中的第  $i$  个样本，是  $p$  维列向量  $(x_1^i, x_2^i, \dots, x_p^i)^T$ ； $W$  是  $p$  维列向量  $W = (w_1, w_2, \dots, w_p)^T$ ，为待求参数； $b$  是一个数，也是待求参数。

对于  $w^T x + b$ ，其结果是  $w_1 x_1 + w_2 x_2 + \dots + w_p x_p + b$ 。所以可以把  $W$  写成  $(w_1, w_2, \dots, w_p, b)^T$ ， $X^i$  写成  $(x_1^i, x_2^i, \dots, x_p^i, 1)^T$ 。 $w^T x + b$  就可以写成  $w^T x$ ，则：

$$h(x^i) = \frac{1}{1 + e^{-W^T X^i}}。$$

这样就可以把另一个参数  $b$  合并到  $W$  中。之后就是根据训练样本来求参数  $W$  了。

首先，借助 SPSS 实现逻辑回归，先将全部数据导入进 SPSS，然后给因变量种类创建虚拟变量（01 变量），将高钾类玻璃量化作为 1，铅钡类玻璃量化作为 0。删除所有不需要的数据，其中高钾和铅钡为分类变量，将该变量放入分类协变量里面，得到逻辑回归系数表详见附录 3，将回归系数和各化学成分含量带如下式：

$$\hat{y}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 m_i + \hat{\beta}_2 m_{2i} + \dots + \beta_k m_{ki}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 m_i + \hat{\beta}_2 m_{2i} + \dots + \beta_k m_{ki}}} \quad (3)$$

其中  $\hat{\beta}_i (i=1, \dots, k)$  为逻辑回归系数， $m_i$  为各化学成分。

如果  $\hat{y}_i \geq 0.5$  则预测的是高钾类玻璃；否则则认为其预测的是铅钡类玻璃。得到预测分类结果见表 12：

表 12 预测分类表

类型	组号
高钾	A6, A7, A8
铅钡	A1, A2, A3, A4, A5

将该模型进行灵敏度分析，将表单 2 中已有数据带入模型预测其类型，并与真实类型进行对比，得到如下结果：

表 13 预测准确率表

实测	0.00	1.00	正确百分比
铅钡	0.00	14	4
高钾	1.00	14	35
总体百分比			72.1

用 SPSS 得出来的结果发现，68 个样本中，预测出来铅钡有 35 个，预测出的正确率为 71%；预测出高钾有 14 个，预测正确率为 77%，对于整个样本，逻辑回归的预测成功率为 72.1%。

可以发现两类玻璃的预测率都在 80%以下，所以预测结果较差，因此考虑在逻辑回归模型中加入平方项与交互项，其余操作和上述类似，得到预测结果如表 14：

表 14 预测分类表

类型	组号
高钾	A6, A7, A8
铅钡	A1, A2, A3, A4, A5

并将加入交互项的模型进行灵敏度分析，将表单 2 中已有数据带入模型预测其类型，并与真实类型进行对比，得到如下结果：

表 15 预测准确率表

实测	0.00	1.00	正确百分比
铅钡	0.00	19	100.0
高钾	1.00	0	100.0
总体百分比			100.0

用 SPSS 得出来的结果发现，68 个样本中，预测出来铅钡有 49 个，预测出的正确率为 100%；预测出高钾有 19 个，预测正确率为 100%，对于整个样本，逻辑回归的预测成功率为 100%。所有百分比均为 100%，虽然预测能力提高了，但是容易发生过拟合的现象。对于样本数据的预测会非常好，但是对于样本外的数据的预测效果可能会很差。

因此为了消除过拟合现象，采取 Fisher 线性判别法来优化模型。

### 5.3.3 Fisher 线性判别法进行模型优化

Fisher 判别<sup>[5]</sup>思想用投影的方式将多维化问题降为一维问题来处理。选择适当的投影轴，以便所有取样点都可以投影到此轴上，从而获得投影值。对该投影轴方向的具体要求是：确保由每个类中的投影值形成的类内偏差尽可能小，由类间投影值形成的类间偏差尽可能大。

对于附件中给定玻璃样本化学成分数据， $D$ (已经设置好分类标签)， $X_i, U_i, \sum_i \omega$  分别表示玻璃类别  $i$  的集合，均值向量和协方差矩阵。现将数据投影到直线  $x = 0$  上，则样本中心的投影为  $0 = \omega_1 u_1 + \omega_2 u_2 + \dots + \omega_n u_n$ 。（ $n$  为玻璃样本维度，接下来的讨论中将统一设置为 2），写成向量的形式  $\omega^T u = 0$  如果将所有的玻璃样本都投影到直线上，则两类玻璃样本的协方差分别为  $\omega^T \sum_0 \omega$  和  $\omega^T \sum_1 \omega$ 。想要达到较好的分类效果，应该使同类玻璃样本的投影点尽可能的接近，也就是让同类玻璃样本投影点的协方差尽可能的小。即  $(\omega^T \sum_0 \omega + \omega^T \sum_1 \omega)$  尽可能小的分别表示给定类别的集合、平均向量和协方差矩阵。

现在，将所有数据投影到直线上时，如果所有样例都投影到直线上，则两种玻璃样本的协方差以矢量格式创建。以获得更好的分类效果作为结果，相似玻璃

样品的投影点必须尽可能的接近。也就是说，相似玻璃样本投影点的协方差必须尽可能的小。另外，要确保不同玻璃种类的采样投影点尽可能的彼此远离，也就是说，尽可能的远离对方。考虑到两者之间的关系，则可以获得以下需要最大化的目标：

$$J = \frac{\|\omega^T u_0 - \omega^T u_1\|}{\omega^T \sum_0 \omega}$$

这里定义类内散度矩阵为：

$$S_{\omega} = \sum_0 + \sum_1 = \sum_{x \in X_0} (x - u_0)(x - u_0)^T + \sum_{x \in X_1} (x - u_1)(x - u_1)^T$$

以及类间离散度矩阵为：

$$S_b = (u_0 - u_1)(u_0 - u_1)^T$$

则  $J$  可重写为：

$$J = \frac{\omega^T S_b \omega}{\omega^T S_{\omega} \omega} \tag{4}$$

### 5.3.4 优化模型的求解和结果的分析

操作和逻辑回归做法一致，首先为分类指标创建虚拟变量，将高钾类玻璃量化作为 1，铅钡类玻璃量化作为 0。

接着借助 SPSS 实现 Fisher 线性判别分类，其中，定义范围为分类指标的种类，本问中，有两个种类，使用 SPSS 计算，得到表 16：

表 16 贝叶斯判别函数系数表					
分类函数系数					
类型	铅钡	高钾	类型	铅钡	高钾
表面风化	-3.143	-6.195	氧化铝(Al2O3)	3.162	4.033
二氧化硅(SiO2)	7.187	7.470	氧化铁(Fe2O3)	10.739	10.713
氧化钠(Na2O)	11.780	13.330	氧化铜(CuO)	3.017	2.151
氧化钾(K2O)	10.912	9.726	氧化铅(PbO)	7.584	8.197
氧化钙(CaO)	2.974	3.381	氧化钡(BaO)	9.342	10.431
氧化镁(MgO)	20.178	21.585	五氧化二磷(P2O5)	10.335	10.951

将样本的每个参数导入到两个贝叶斯判别函数中。比较得出的函数值，哪个样本的函数值较大，就将该样品归于哪一类。根据 Fisher 预测结果（详见附录七）得出分组表格：

表 17 Fisher 线性判别分类结果	
类型	组号
高钾	A1, A6, A7
铅钡	A2, A3, A4, A5, A8



最后将表单 3 中的数据赋予上下 10%的扰动进行敏感性分析，重新对数据进行预测分类，对比预测分类和实际分类的差异。得到如下结果：

表 18 分类表

分类结果				
		预测组成员信息		总计
		高钾	铅钡	
计数	高钾	19	0	19
	铅钡	1	48	49
	未分组个案	3	5	8
准确率	高钾	100.0	0	100.0
	铅钡	2.0	98.0	100.0
	未分组个案	37.5	62.5	100.0
正确地对 98.5% 个原始已分组个案进行了分类				

#### 5.4 问题四模型的建立与求解

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，考虑使用灰色关联度分析，将高钾类玻璃样品中的氧化钾作为母序列，其他化学成分作为子序列进行灰色关联度分析；同理将铅钡类玻璃样品中的氧化铅和氧化钡作为母序列，其他化学成分作为子序列分别进行灰色关联度分析。通过将高钾类玻璃和铅钡类玻璃的各化学成分与母序列的关联程度进行比较，对不同类别之间的化学成分关联关系的差异性进行分析。

##### 5.4.1 灰色关联度模型建立

针对两个系统之间的综合因素，测量随时间或根据其他因素变化的关联性大小称为关联度。在系统开发体系中，两个因素之间变化地趋势是一致的，即同步变化程度较高，可以说关联程度较高；反之，则较低。灰色关联度分析<sup>[6]</sup>就是对系统发展状况的定量描述和比较方法。核心逻辑为确定参考数据列和多个比较数据列的几何相似性，以判断关系是否紧密相连。反映曲线之间的关联性程度。

(1) 确定反映系统行为特征的参考序列  $X^{(0)}$  和影响系统行为的比较序列  $X^{(m)}$ ：

其中反映系统行为特征的数据序列为参考序列为：

$$X_0 = \{X_0(1), X_0(2), \dots, X_0(n)\}$$

影响系统行为的因素组成的数据序列为比较序列为：

$$X_1 = \{X_1(1), X_1(2), \dots, X_1(n)\}$$

$$X_2 = \{X_2(1), X_2(2), \dots, X_2(n)\}$$

⋮

$$X_m = \{X_m(1), X_m(2), \dots, X_m(n)\}$$

(2) 求出各序列的初值像(无量纲化处理)。令

$$X' = X_i / X_i(1) = \{X'_i(1), X'_i(2), \dots, X'_i(n)\}$$

其中  $i = 0, 1, 2, 3, \dots, m$  得到：

$$X_0', X_1', X_2', \dots, X_m'$$

(3) 求得出  $X_0$  与  $X_i$  的初值像对应分量之差的绝对值序列。

$$\Delta_i(k) = |X_0'(k) - X_i'(k)|$$

$$\Delta_i = (\Delta_i(1), \Delta_i(2), \dots, \Delta_i(n))$$

$$i = 1, 2, \dots, m$$

$$k = 1, 2, \dots, n$$

(4) 解出  $\Delta_i(k) = |X_0'(k) - X_i'(k)|$  的最小值与最大值。分别记作

$$\Delta_{\min} = \min_i \min_k \Delta_i(k)$$

$$\Delta_{\max} = \min_i \min_k \Delta_i(k)$$

(5) 则关联系数  $\xi_i(k)$

$$\xi_{0i}(k) = \frac{\Delta_{\min} + p\Delta_{\max}}{\Delta_i(k) + p\Delta_{\max}}$$

其中,  $p$  为分辨系数,  $0 < p < 1$ , 一般取  $p = 0.5$

(6) 求关联度

$$\gamma_{0i} = \frac{1}{n} \sum_{k=1}^n \xi_{0i}(k) \quad (5)$$

(7) 按  $\gamma_{0i}$  的排序, 区分其关联程度的大小, 若  $\gamma_i$  值越大, 说明其关联程度越大; 反之若  $\gamma_i$  值越小, 则关联程度越小

#### 5.4.2 模型的求解和结果分析

在本问中, 先分别先将高钾类玻璃中的氧化钾作为母序列, 将其余化学成分作为子序列进行灰色关联度分析; 再将铅钡中的氧化铅和氧化钡分别作为母序列, 将其余化学成分作为子序列进行灰色关联度分析。

关联系数代表着该子序列二氧化硅(SiO<sub>2</sub>)、五氧化二磷(P<sub>2</sub>O<sub>5</sub>)、氧化锶(SrO)、氧化钡(BaO)、二氧化硫(SO<sub>2</sub>)、氧化铁(Fe<sub>2</sub>O<sub>3</sub>)、氧化锡(SnO<sub>2</sub>)、氧化钠(Na<sub>2</sub>O)、氧化钙(CaO)、氧化铅(PbO)、氧化铝(Al<sub>2</sub>O<sub>3</sub>)、氧化铜(CuO)、氧化镁(MgO)对与母序列对应维度上的关联程度值(数字越大, 代表关联性越强)。关联系数表详见附录 4, 5, 6。

(1) 当氧化钾为母序列时得到关联程度如表 19:

表 19 高钾关联程度

关联度结果					
评价项	关联度	排名	评价项	关联度	排名
氧化钙(CaO)	0.946	1	二氧化硅(SiO <sub>2</sub> )	0.92	6
氧化铝(Al <sub>2</sub> O <sub>3</sub> )	0.945	2	五氧化二磷(P <sub>2</sub> O <sub>5</sub> )	0.911	7
氧化铁(Fe <sub>2</sub> O <sub>3</sub> )	0.934	3	氧化锶(SrO)	0.898	8
氧化镁(MgO)	0.932	4	氧化铅(PbO)	0.897	9
氧化铜(CuO)	0.93	5			

从上表可以看出：针对本次 13 个评价项，氧化钙(CaO)评价最高(关联度为：0.946)，其次是氧化铝(Al<sub>2</sub>O<sub>3</sub>)(关联度为：0.945)。

(3) 当氧化铅为母序列时得到关联程度如表 20：

表 20 氧化铅关联程度

关联度结果					
评价项	关联度	排名	评价项	关联度	排名
氧化锶(SrO)	0.956	1	氧化铜(CuO)	0.922	8
氧化钙(CaO)	0.948	2	氧化铁(Fe <sub>2</sub> O <sub>3</sub> )	0.911	9
氧化钡(BaO)	0.945	3	氧化钾(K <sub>2</sub> O)	0.906	10
氧化铝(Al <sub>2</sub> O <sub>3</sub> )	0.939	4	二氧化硫(SO <sub>2</sub> )	0.89	11
二氧化硅(SiO <sub>2</sub> )	0.935	5	氧化钠(Na <sub>2</sub> O)	0.888	12
氧化镁(MgO)	0.928	6	氧化锡(SnO <sub>2</sub> )	0.887	13
五氧化二磷(P <sub>2</sub> O <sub>5</sub> )	0.927	7			

从上表可以看出：针对本次 13 个评价项，氧化锶(SrO)评价最高(关联度为：0.956)，其次是氧化钙(CaO)(关联度为：0.948)。

(3) 当氧化钡为母序列时得到关联程度如表 21：

表 21 氧化钡关联程度

关联度结果					
评价项	关联度	排名	评价项	关联度	排名
氧化锶(SrO)	0.944	1	二氧化硫(SO <sub>2</sub> )	0.91	8
氧化铅(PbO)	0.944	2	氧化钾(K <sub>2</sub> O)	0.91	9
氧化铜(CuO)	0.944	3	氧化镁(MgO)	0.907	10
二氧化硅(SiO <sub>2</sub> )	0.937	4	氧化铁(Fe <sub>2</sub> O <sub>3</sub> )	0.898	11
氧化铝(Al <sub>2</sub> O <sub>3</sub> )	0.929	5	氧化钠(Na <sub>2</sub> O)	0.891	12
氧化钙(CaO)	0.924	6	氧化锡(SnO <sub>2</sub> )	0.887	13
五氧化二磷(P <sub>2</sub> O <sub>5</sub> )	0.913	7			

从上表可以看出：针对本次 13 个评价项，氧化锶(SrO)评价最高(关联度为：0.944)，其次是氧化铅(PbO)(关联度为：0.944)。

关联度表示各评价项与母序列之间的相似关联程度，其是由关联系数进行计算平均值得出，关联度值介于 0~1 之间，该值越大表示评价项与母序列相关性越强，关联度越高，意味着评价项与母序列之间关系越紧密，因而其评价越高。结合关联度值，针对所有评价项进行排序，得到各评价项排名。

#### 5.4.3 化学成分的关联关系的差异性

根据以上三个表格可以看出，氧化钙对高钾的关联程度最为重要，其次为氧化铝、氧化铁等；对于铅钡类玻璃来说，两次灰色关联分析与之关联程度最高的皆为氧化锶，并且可以发现氧化钡关联度结果中氧化锶氧化铅和氧化铜的关联度相同，说明对于氧化钡同等重要；

虽然钡类玻璃关联程度最高和铅类玻璃关联程度最高的化学成分皆为氧化锶，但两者关联度大小有所不同，氧化锶对于铅钡类玻璃中氧化铅的关联程度为 0.956，但氧化锶关联程度为 0.944，可以明显看出铅类玻璃中的氧化锶高于钡类玻璃中氧化锶含量。

并且可以发现在高钾类玻璃与氧化钾关联度较高的成分只有 9 项,而铅钡类玻璃中有 13 项,说明各化学成分对于铅钡类玻璃的关联度更高,对于高钾类玻璃的关联度较低。

综上,各类化学成分对于铅钡类玻璃的关联程度更高,其中氧化钙和氧化铜对两类玻璃都有明显的关联性。

## 六、模型的评价与推广

### 6.1 模型的优点

1.本文主要是对于数据的分析和处理,采用功能强大的 SPSS 和 SPSSPRO 软件进行求解,可以避免调试代码,使问题变得简单易处理;

2.第三问采用随机森林和 Fisher 线性判别两种方法对于样本进行分类,使结论更加准确可信,更有说服力;

3.第四问采用灰色关联度分析,计算量小,对样本的量多少和有无规律都同样适用,不会出现量化结果与定性分析结果不符的情况。

### 6.2 模型的缺点

1.第二问随机森林中,可能有很多相似的决策树,掩盖了真实的结果,导致各特征的重要性发生变化,对于分类结果产生影响;

2.第三问中,对于线性不可分的情况, Fisher 方法无法确定分类;

3.灰色关联发要求需要对各项指标的最优值进行现行确定,主观性过强。

### 6.3 模型的推广

本文的预测模型还可应用到实际生活当中,例如全面二孩政策对我国人口结构的影响某地水资源短缺风险综合评价等。

## 七、参考文献

- [1]本刊编辑部.卡方检验的应用条件[J].现代电生理学杂志, 2021, 28(03): 192.
- [2]王玥,陈之源,蒋远东,腾子豪,胡鹏远,常玉雪,向阳.基于随机森林和 logistic 回归的结核病危险因素研究[J].医学动物防制, 2022, 38(10): 924-928.
- [3]吴宗道,周福征,史美光.几个古玻璃的显微形貌、成分及其风化的初步研究[J].电子显微学报, 1986, (04): 65-71.
- [4]王凌妍,张鑫雨,许胜楠,王禹力,甄志龙.逻辑回归的敏感性分析及在特征选择中的应用[J].信息记录材料, 2022, 23(07): 30-33.DOI: 10.16009/j.cnki.cn 13-1295/tq.2022.07.051.
- [5]汪家清,张鑫,曹彤,王能才,张海英.基于 Fisher 线性判别分析对乳腺微钙化性质的预测研究[J].中国医学装备, 2022, 19(02): 5-9.
- [6]司守奎.数学建模算法与程序[M].北京:国防工业出版社, 2013.

## 附录

### 支撑材料列表:

#### 所用数据:

风化情况.xlsx

附件.xlsx

附件 1.xlsx

#### 第一问:

第一问卡方检验.xlsx

卡方检验.sav

类型.spv

纹饰卡方.spv

颜色.spv

颜色卡方.spv

含量预测.rar

#### 第二问:

第二问.xlsx

第二问随机森林.xlsx

随机森林分类.docx

高钾个案排秩.sav

铅钡个案排秩.sav

2+90.xlsx

2+90 预测.xlsx

2+110.xlsx

2+110 预测.xlsx

#### 第三问:

第三问灵敏度分析.xlsx

第三问逻辑回归.xlsx

第三问数据.sav

Fisher 判别.spv

问题三预测结果.spv

逻辑回归灵敏度 3.docx

3+90.xlsx

3+90 预测.xlsx

3+110.xlsx

3+110 预测.xlsx

#### 第四问:

第四问关联分析 1.xlsx

第四问关联分析 2.xlsx

灰色关联分析钡.docx

灰色关联分析钾.docx

灰色关联分析铅.docx

灰色关联系数钡.csv

灰色关联系数钾.csv

## 附录 1 问题一：分配权重，得到加权平均值

```
clc
clear
close all
fName=["gjfh.xlsx", "gjwfh.xlsx", "qbfh.xlsx", "qbwfh.xlsx"];
Pingjaunzhi=zeros(4, 15);%行为 4，列为 15 的矩阵
for t=1: 4
    A=xlsread(fName(t));
    [n, m]=size(A);
    for i=1: m
        temp=sort(A(:, i));%排序
        ave=0;
        sum=0;
        for j=1: n
            alpha=normpdf((j/n-0.5)*3);%正态分布概率密度函数
            ave=ave+temp(j)*alpha;
            sum=sum+alpha;
        end
        ave=ave/sum;
        Pingjaunzhi(t, i)=ave;%输出结果
    end
end
Pingjaunzhi
```

## 附录 2 问题二：两类玻璃得分表，其中浅色为铅钡类玻璃，深色为高钾类玻璃

组号	分数	组号	分数	组号	分数	组号	分数
3	24	47	87	45	123	44	145
6	32	54	88.5	40	124	32	145
6	36	30	88.5	8	125	43	147.5
16	40	26	92	52	125	22	148
14	43	41	96.5	26	126	25	153
21	44	46	96.5	43	126	50	153.5
18	45.5	24	104	49	126	55	153.5
3	49	34	107	20	127.5	33	155
5	52.5	11	108	23	128	31	156
1	54.5	37	109	48	129	35	161
10	55.5	30	110	53	129.5	19	161.5
12	56.5	36	110.5	8	133.5	51	179

15	56.5	38	113	54	137.5
9	57.5	42	113	51	138
4	57.5	50	113	57	138.5
13	60.5	39	115	28	140
27	61.5	58	118	56	141.5
17	61.5	42	119.5	29	144
7	62.5	2	122	49	144

附录 3 问题三:逻辑回归系数表  
方程中的变量

步 骤		B	标准误差	瓦尔德	自由度	显著性	Exp(B)
1 <sup>a</sup>	表面风化(1)	-12.448	22385.822	.000	1	1.000	.000
	二氧化硅(SiO <sub>2</sub> )	7.224	2323.454	.000	1	.998	1371.365
	氧化钠(Na <sub>2</sub> O)	.956	2211.381	.000	1	1.000	2.601
	氧化钾(K <sub>2</sub> O)	5.528	3423.909	.000	1	.999	251.625
	氧化钙(CaO)	11.710	3927.539	.000	1	.998	121823.112
	氧化镁(MgO)	42.397	14092.792	.000	1	.998	2587827471599864800.000
	氧化铝(Al <sub>2</sub> O <sub>3</sub> )	6.330	3191.082	.000	1	.998	561.007
	氧化铁(Fe <sub>2</sub> O <sub>3</sub> )	2.410	4583.112	.000	1	1.000	11.131
	氧化铜(CuO)	-11.897	4114.801	.000	1	.998	.000
	氧化铅(PbO)	7.806	2612.501	.000	1	.998	2455.629
	氧化钡(BaO)	16.426	3414.035	.000	1	.996	13601105.878
	五 氧 化 二 磷 (P <sub>2</sub> O <sub>5</sub> )	6.258	2668.181	.000	1	.998	522.178
	氧化锶(SrO)	15.296	29757.691	.000	1	1.000	4395048.209
	氧化锡(SnO <sub>2</sub> )	-65.205	86393.167	.000	1	.999	.000
	二氧化硫(SO <sub>2</sub> )	1.233	6015.820	.000	1	1.000	3.432
	常量	-719.734	229797.289	.000	1	.998	.000

a. 在步骤 1 输入的变量: 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>), 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>)。

附录 4 问题四: 母序列为氧化钾时的灰色关联系数

关联系数结果													
	SiO <sub>2</sub>	P <sub>2</sub> O <sub>5</sub>	SrO	BaO	SO <sub>2</sub>	Fe <sub>2</sub> O <sub>3</sub>	SnO <sub>2</sub>	Na <sub>2</sub> O	CaO	PbO	Al <sub>2</sub> O <sub>3</sub>	CuO	MgO
7	0.88	0.93	1.00	1.00	1.00	0.99	1.00	1.00	0.97	1.00	0.95	0.85	1.00
9	0.88	0.97	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.98	0.93	0.99
10	0.88	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.99	0.98	1.00	0.97	0.98
12	0.89	1.00	0.98	0.98	0.98	0.99	0.98	0.98	0.99	0.98	0.98	0.93	0.98

27	0.88	0.96	1.00	1.00	1.00	0.98	1.00	1.00	0.97	1.00	0.94	0.92	0.93
1	0.94	0.97	0.86	0.86	0.66	0.97	0.86	0.86	0.97	0.86	0.93	0.96	0.96
3	0.96	0.99	0.92	0.92	0.92	0.92	0.92	0.92	0.98	0.99	0.99	0.96	0.92
3	0.90	0.89	0.82	0.61	0.83	0.97	0.83	0.83	0.98	0.73	0.93	0.94	0.95
4	0.94	0.93	0.86	0.86	0.68	1.00	0.86	0.86	0.94	0.86	0.99	0.95	0.95
5	0.92	0.93	0.93	0.84	0.61	0.97	0.84	0.84	0.95	0.84	0.99	0.99	0.94
6	0.98	0.74	0.74	0.77	0.89	0.93	0.89	0.89	0.89	0.96	0.88	0.99	0.87
6	0.96	0.72	0.72	0.86	0.89	0.73	0.89	0.89	0.96	0.99	0.90	0.99	0.90
13	0.89	0.94	0.83	0.83	0.83	0.98	0.83	0.81	0.94	0.83	0.94	0.96	0.83
14	0.90	0.84	0.83	0.83	0.83	0.86	0.83	0.75	0.95	0.69	0.99	0.85	0.90
15	0.97	0.91	0.89	0.89	0.89	0.96	0.89	0.72	0.89	0.95	0.95	0.95	0.98

## 附录 5 问题四：母序列为氧化铅时的灰色关联系数

关联系数结果													
	SiO2	Na2O	K2O	SnO2	SO2	CaO	SrO	CuO	Fe2O3	P2O5	BaO	Al2O3	MgO
2	0.95	0.89	0.73	0.89	0.89	0.97	0.93	0.90	0.89	0.97	0.89	0.99	0.97
8	0.97	0.93	0.93	0.93	0.82	0.99	0.98	0.70	0.93	0.98	0.84	0.96	0.93
8	0.93	0.92	0.92	0.92	0.38	0.95	0.95	0.94	0.92	0.89	0.85	0.94	0.92
11	0.99	0.94	0.97	0.94	0.94	0.92	0.97	0.86	0.94	0.84	0.95	1.00	0.97
19	0.95	0.90	0.90	0.90	0.90	0.99	0.94	0.95	0.94	0.89	0.93	0.97	0.97
22	0.83	1.00	0.74	1.00	1.00	0.93	1.00	0.97	0.95	0.99	1.00	0.92	0.92
23	0.93	0.57	0.96	0.96	0.96	0.98	0.96	0.91	0.96	0.96	0.95	0.99	0.95
24	0.99	0.93	0.93	0.93	0.93	0.94	0.86	0.75	0.93	0.93	0.87	0.96	0.93
25	0.97	0.87	0.92	0.92	0.92	0.94	0.97	0.97	0.89	0.92	0.97	0.96	0.92
26	0.96	0.93	0.93	0.93	0.88	0.98	0.96	0.70	0.93	0.99	0.83	0.94	0.93
26	0.93	0.92	0.90	0.92	0.37	0.95	0.93	0.92	0.92	0.92	0.82	0.95	0.92
28	0.91	0.96	0.93	0.76	0.96	0.99	0.98	0.97	0.99	0.98	0.99	0.94	0.92
29	0.90	0.94	0.90	0.97	0.97	0.91	0.97	1.00	0.93	0.98	0.98	0.76	0.85
34	0.96	0.89	0.99	0.89	0.89	0.92	0.94	0.95	0.94	0.90	0.96	0.92	0.89
36	0.98	0.90	0.96	0.90	0.90	0.91	0.95	0.93	0.94	0.90	0.98	0.93	0.90

## 附录 6 问题四：母序列为氧化钡时的灰色关联系数

关联系数结果													
	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	P2O5	PbO	SrO	SnO2	SO2
2	0.92	1.00	0.66	0.91	0.86	0.88	0.80	0.99	0.91	0.88	0.95	1.00	1.00
8	0.81	0.79	0.79	0.83	0.79	0.81	0.79	0.81	0.85	0.84	0.85	0.79	0.98
8	0.80	0.79	0.79	0.89	0.79	0.81	0.79	0.90	0.95	0.85	0.89	0.79	0.41
11	0.95	0.89	0.98	0.97	0.97	0.94	0.89	0.90	0.88	0.95	0.97	0.89	0.89
19	0.98	0.96	0.96	0.92	0.97	0.96	0.88	0.89	0.83	0.93	1.00	0.96	0.96
22	0.83	1.00	0.74	0.93	0.92	0.92	0.95	0.97	0.99	1.00	1.00	1.00	1.00



23	0.98	0.59	0.91	0.92	1.00	0.94	0.91	0.96	0.91	0.95	0.98	0.91	0.91
24	0.86	0.81	0.81	0.83	0.81	0.84	0.81	0.85	0.82	0.87	0.99	0.81	0.81
25	0.95	0.85	0.95	0.97	0.95	0.99	0.86	1.00	0.95	0.97	0.99	0.95	0.95
26	0.81	0.78	0.78	0.82	0.78	0.79	0.78	0.81	0.84	0.83	0.86	0.78	0.95
26	0.77	0.76	0.90	0.85	0.76	0.78	0.76	0.88	0.88	0.81	0.87	0.76	0.40
28	0.89	0.97	0.92	0.98	0.91	0.93	0.98	0.98	0.99	0.99	1.00	0.75	0.97
29	0.89	0.93	0.89	0.90	0.84	0.75	0.91	0.98	0.99	0.98	0.95	0.98	0.98
34	0.99	0.92	0.97	0.95	0.92	0.95	0.98	0.99	0.93	0.96	0.97	0.92	0.92
36	0.99	0.88	0.97	0.93	0.91	0.95	0.95	0.94	0.92	0.98	0.96	0.91	0.91

## 附录七：fisher 判别预测结果

组号	0 概率（铅钡）	1 概率（高钾）
A1	0.57789	0.42211
A2	0.00004	0.99996
A3	0.00002	0.99998
A4	0.00000	1.00000
A5	0.00048	0.99952
A6	0.99996	0.00004
A7	0.99907	0.00093
A8	0.37557	0.62443