古代玻璃制品的成分分析与鉴别

摘要

文物的考古研究在现代社会具有重要意义，研究分析文物特点是重要课题。其中，玻璃文物极易受环境影响而风化，如何探究玻璃文物表面风化与自身特性关系、对其化学成分展开鉴别分析、找寻不同文物类别的关联性与差异性，成为了亟待解决的问题。

问题一中，我们首先建立**R-Q 型因子分析模型**，对玻璃文物与其类型、纹饰、颜色的关系展开研究分析，并对初步分析的结果进行卡方检验，进一步确定风化与各因素的关系，得到**玻璃文物风化与其类别密切相关**的结论。在研究化学成分含量规律时，我们首先进行了数据清洗，剔除了总含量在 85%-105% 外的无效数据，而后对各类别文物的化学成分含量进行了**分类汇总**，深入研究有无风化化学成分含量的统计规律。结合研究得出的规律，我们采用平均值比值乘积的方法，对各类别已风化的文物进行了未风化前化学成分含量的预测，得出的结果在正文和附录中展示。

问题二中，在研究高钾玻璃与铅钡玻璃的分类规律时，我们首先建立 **CART 决策树模型**，将数据清洗后的样本集导入作为训练集，得到该玻璃类型的分类以 PbO **百分比含量与 5.46% 的大小关系作为分类依据**，完成二分。在进行亚分类工作时，我们首先建立随机森林模型确定各化学成分含量的分类权重，决定选取权重较高的 BaO、PbO、SrO、SiO_2 与 K_2O 当做亚分类指标，建立 **K-means 聚类**模型，并予以优化，改进衍生出 **K-means++** 模型，完成类型的进一步亚分，并带入足够样本集进行合理性与敏感性检验。

问题三中，对于未知文物的类型鉴别，我们构建了**逻辑回归模型**，将已知文物化学成分数据作为训练集，完成模型构建并通过交叉验证验证了模型的准确性。而后将未知文物的化学含量数据导入该模型，完成鉴别工作，最终判定文物 **A1**、**A2**、**A6**、**A7 属于高钾玻璃，A3**、**A4**、**A5**、**A8 属于铅钡玻璃**。在此基础上，我们利用逻辑回归判别函数特性设计敏感性特征值计算公式，将其与理论特征值作比，完成敏感度检验。

问题四中，为研究同类别文物化学含量的关系，我们对两类文物进行**相关性分析**，计算得相关系数，并以相关系数热力图形式表示。为检验结果准确性，我们建立了**灰色关联分析模型**，选特定数据进行相关系数计算，与前述相关性分析结果对比检验，发现计算十分准确。进行差异性分析时，我们利用 R-Q 型因子分析中的显著值计算方法，将不同类别同种元素两两配对计算显著值和效应值，根据范围划分完成差异性比较。

最后，结合实际情况对模型的优劣进行讨论及推广，综合判定模型具有较强适用性与泛用性。

关键字：R-Q 型因子分析　决策树　K-means 聚类　灰色关联分析　逻辑回归

一、 问题重述

1.1 问题的背景

在百年前的康熙年间，玻璃器是一种珍贵的奢侈品，寥寥可数凤毛麟角。其中彩釉玻璃有许多不同的颜色和花纹，如条状、网状和电状图案等等。在自然状态下，玻璃成分失去了结晶水,玻璃变得脆弱,透光率降低,产生裂缝,和鳞片状剥落，我们称之为玻璃风化。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，玻璃文物风化研究也十分重要。

玻璃文物表面风化与文物的各种特性密切相关，其中便包括玻璃类型、颜色与纹饰。例如，在玻璃生产中，其配料中金属氧化物越多，钠离子富集区越大，非桥阳离子就越多，形成“Si-O-Na”型结构的可能性越大，风化可能性越大。所以，通常情况下茶色、绿色、蓝宝石色等各种颜色玻璃比普通平板玻璃更容易风化。因此探索玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系对文物保护工作有重大意义。

1.2 问题的提出

为了进一步研究玻璃文物风化的相关问题，题目要求结合所提供的文物数据，建立数学模型研究以下问题:

1. 分析玻璃文物表面风化与其类型、纹饰、颜色的关系，结合玻璃类型及是否风化，分析化学成分含量的统计规律，并根据该规律，预测风化玻璃未风化前的各化学成分含量。
2. 分析高钾玻璃、铅钡玻璃的分类依据，选择合适化学成分对其进行亚分类，并检验分类结果合理性及敏感性。
3. 分析未知玻璃文物化学成分，鉴别其类型，并进行敏感性分析。
4. 分析不同类别玻璃文物中化学成分之间的联系，比较不同类别间化学成分关系的差异性。

二、 问题分析

2.1 问题一分析

问题一要求我们分析玻璃文物表面风化与三者的关系，并结合相关指标，分析化学成分含量的统计规律，并根据该规律完成预测工作。在分析玻璃风化与三者关系时，基

于多因素分析，我们建立了 R-Q 型因子分析模型，分别对玻璃类型、纹饰与颜色展开分析，并进行卡方检验，从而得到相关结论。

对于化学成分含量统计规律问题，首先基于题目所给有效值范围对数据进行筛选取舍，而后进行分类汇总工作，以更加直观研究出化学成分含量的规律。

针对风化玻璃化学成分含量在未风化前的预测问题，鉴于各类别多组数据线性拟合效果良好，本次预测采用比值预测，最终得到未风化前数据的预测值。

2.2 问题二分析

问题二要求研究高钾玻璃与铅钡玻璃的分类规律，在此基础上亚划分，并检验分类的合理性与敏感性。玻璃类型的分类为简单的二分，需要对各成分数据建立分类模型研究，我们构建了 CART 决策树模型，对相关指标进行了具体研究，得到了分类依据。

在进行亚划分时，首先要在各类型中选择合适的化学成分，这里采用随机森林模型，首先之于各化学成分对分类的影响建立权重，决定好各成分的取舍。接着建立 K-means 聚类模型，并予以优化改进，最终完成亚分类。针对敏感度分析，我们选取了一系列指标，代入样本进行分析，最终检验该模型的合理性与敏感度。

2.3 问题三分析

问题三中需要对附件中未知玻璃文物 A1-A8 进行类型的鉴别与划分，并探讨该类别划分的敏感性。在进行类型鉴别时，我们构建了逻辑回归模型，并进一步建立相关指标使其适用于分类与预测，而后代入 8 组数据进行类别的鉴定预测。

为了检验该类型鉴定的合理性，我们根据判别函数特性确立了敏感性特征值的计算方式，代入真实数据进行对比，以此完成敏感度的分析。

2.4 问题四分析

问题四主要在于探究玻璃文物同类中化学成分含量的关联性以及非同类间化学成分含量的差异性。对于关联性分析，我们采用相关性分析并建立灰色关联模型予以验证，二者相辅相成，以得到更加准确的结论。

对于差异性，则计算影响力较大的显著性值以及效应值，进一步完成不同类别间的元素差异性比较，完成相关分析。

三、 模型假设

1. 仅考虑文物表面风化与文物类型、纹饰、颜色的联系，忽略其他影响因素。
2. 忽略缺失值的影响，将化学成分含量缺失值统一视作 0。
3. 认为玻璃文物类型的分类仅以化学成分含量为准，忽略其他影响。

4. 认为所提供文物数值与实际情况一致，真实可靠。

四、 符号说明

符号	定义
P	显著值
x_{ij}	影响玻璃表面风化第 i 行第 j 列的因素
X^2	卡方检验统计量
$P(y_i)$	单个化学成分样本含量预测值
S	模型敏感性特征值
ρ	皮尔逊相关系数
x_n	第 n 个化学元素含量百分比

五、 问题一模型的建立、求解、分析

5.1 玻璃文物表面风化的因素分析

5.1.1 R-Q 型因子分析

初步分析了影响玻璃文物表面风化的影响因素后，利用 R-Q 型因子分析，分别对玻璃类型、纹饰、颜色进行了对应分析，以此更加直观地表示出因素间的相关性。

一、R-Q 型因子分析具体步骤

（1）R 型因子分析原理

共有 n 个玻璃文物，每个玻璃文物中有 m 个影响因素，则其矩阵为：

$$X=\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & & & \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \tag{1}$$

式中，$x_{ij} \geq 0(i=1,2,\dots,m;j=1,2,\dots,n)$，并且每一行和每一列至少有一个数据不为 0。根据式 (2) 对式 (1) 进行变换，即：

$$x_{ij}=\frac{x_{ij}-\frac{x_i x_j}{T}}{\sqrt{x_i x_j}}(i=1,2,\dots,m;j=1,2,\quad,n) \tag{2}$$

式中，x_i,x_j,T 分别满足以下条件：

$$x_i = \sum_{j=1}^{n} x_{ij} \quad (j = 1, 2, \ldots, m) \tag{3}$$

$$x_j = \sum_{i=1}^{n} x_{ij} \quad (j = 1, 2, \ldots, n) \tag{4}$$

$$T = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} = \sum_{i=1}^{m} x_i \tag{5}$$

根据影响因素协方差矩阵特征值 $\lambda_1 \geq \lambda_2 \geq \ldots, \geq \lambda_m$，取其累积特征值百分比≥85% 的前 p 个特征值 $\lambda_1, \lambda_2, \ldots, \lambda_p.$，计算与之相对应的单位特征向量 $\mathrm{u}_1, \mathrm{u}_2, \ldots, \mathrm{u}_\mathrm{p}$, 得到 R 型因子载荷矩阵：

$$U = \begin{bmatrix} u_{11}\sqrt{\lambda_1} & u_{12}\sqrt{\lambda_2} & \cdots & u_{1p}\sqrt{\lambda_p} \\ u_{21}\sqrt{\lambda_1} & u_{22}\sqrt{\lambda_2} & \cdots & u_{2p}\sqrt{\lambda_p} \\ \cdots & \cdots & \cdots & \cdots \\ u_{m1}\sqrt{\lambda_1} & u_{m2}\sqrt{\lambda_2} & \cdots & u_{mp}\sqrt{\lambda_p} \end{bmatrix} \tag{6}$$

以此为基础进行的散点数据分析为 R 型因子分析，表示了研究影响因素之间的相互关系 [1]。

（2）Q 型因子分析原理

由 $v_j = z' u_j$ 可得，Q 型因子载荷矩阵为：

$$V = \begin{bmatrix} v_{11}\sqrt{\lambda_1} & v_{12}\sqrt{\hat{i}_2} & \cdots & v_{1p}\sqrt{\lambda_p} \\ v_{21}\sqrt{\lambda_1} & v_{22}\sqrt{\lambda_2} & \cdots & v_{2p}\sqrt{\lambda_p} \\ \cdots & \cdots & \cdots & \cdots \\ v_{n1}\sqrt{i_1} & v_{n2}\sqrt{\hat{h}_2} & \cdots & v_{np}\sqrt{\lambda_p} \end{bmatrix} \tag{7}$$

根据 Q 型因子载荷矩阵在因子平面上作散点图并进行分析，这被称为 Q 型因子分析，旨在研究不同因素之间的相关性 [2]。

（3）R-Q 型因子分析原理

R-Q 因子分析是联合应用了影响因素相关分析的 R 型分析和风化相关分析的 Q 型分析方法，分别计算 R 型和 Q 型因子载荷矩阵，据此进行相关性研究。

二、R-Q 型因子分析结果

维度	奇异值	惯量	X_2	贡献率	累计贡献率
λ_1	34.442	11.863	688.039	100%	100%
λ_2	29.233	8.546	495.654	100%	100%
λ_3	34.122	11.643	628.714	100%	100%

表 5.1　因子分析表

根据上述因子分析表，可以分析字段提取的维度的贡献率。从表中贡献率计算结果可以得出，当选取玻璃类型、纹饰和颜色三种因素分别进行因子分析，其累计贡献率均大于 80%，模型表现较为优秀。因此，通过模型运算，我们可以得到以下三个因素维度分析表。

字段名	项	维度1	维度2
表面风化	无风化	0.41	-1
	风化	-0.289	-1
类型	高钾	0.513	1
	铅钡	-0.231	1

表 5.2　玻璃类型维度分析表

字段名	项	维度1	维度2
表面风化	无风化	-0.348	-0.927
	风化	0.246	-0.927
纹饰	C	-0.04	-1
	A	-0.175	-1
	B	0.84	-1

表 5.3　玻璃纹饰维度分析表

字段名	项	维度1	维度2
表面风化	无风化	0.381	-0.657
	风化	-0.305	-0.657
颜色	蓝绿	-0.089	1
	浅蓝	-0.089	1
	紫	0.112	1
	深绿	-0.032	1
	深蓝	1.118	1
	浅绿	0.447	1
	黑	0.894	1
	绿	1.118	1

表 5.4　玻璃颜色维度分析表

由维度分析表可以直观观察出，仅玻璃类型的两个因素——高钾与铅钡，与是否风化的条件维度相离较近，说明玻璃类型是影响文物表面是否风化的关键性因素。

5.1.2 卡方检验

由上述 R-Q 型因子分析可以初步得到：玻璃文物的表面风化与其玻璃类型关系紧密，而与纹饰、颜色等因素关联不大。为了进一步验证结果，需进一步采用卡方检验进一步得出观测值与理论值之间偏离程度的直观结果。

一、卡方检验具体步骤

(1) 提出原假设:

H_0: 总体 X 的分布函数为 $F(x)$。

如果总体分布为离散型，则假设具体为

H_0: 总体 X 的分布律为 $p\{X=x_i\}=p_i, i=1,2,\ldots$

(2) 将总体 X 的取值范围分成 k 个互不相交的小区间 A_1、A_2、$A_3 \ldots A_k$，如可取 $A_1=(a_0,a_1], A_2=(a_1,a_2], \ldots A_k=(a_{k-1},a_k)$, 其中 a_0 可取 $-\infty$，a_k 可取 $+\infty$，区间的划分视具体情况而定，但要使每个小区间所含的样本值个数不小于 5，而区间个数 k 不要太大也不要太小。

(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)，所有组频数之和 $f_1+f_2+\ldots+f_k$ 等于样本容量 n。

(4) 当 H_0 为真时，根据所假设的总体理论分布，可算出总体 X 的值落入第个小区间 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 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 $X^2=\sum_{i=1}^{k}\frac{(f_i-np_i)^2}{np_i}$，在 O 假设成立的情况下服从自由度为 $k-1$ 的卡方分布 [3]。

二、卡方检验结果

题目	名称	表面风化		总计	X^2	校正 X^2	P
		无风化	风化				
类型	高钾	12	6	18			
	铅钡	12	28	40	6.88	5.452	0.009***
合计		24	34	58			
纹饰	C	13	17	30			
	A	11	11	22	4.957	4.957	0.084*
	B	0	6	6			
合计		24	34	58			
颜色	蓝绿	6	9	15			
	浅蓝	8	12	20			
	紫	2	2	4			
	深绿	3	4	7			
	深蓝	2	0	2	6.287	6.287	0.507
	浅绿	2	1	3			
	黑	0	2	2			
	绿	1	0	1			
合计		24	30	54			

表 5.5　卡方检验结果

该表展示了模型检验的结果，包括数据的频数、频数百分比、卡方值、显著性 P 值。分析显著性 P 值不难发现，只有玻璃类型对应的显著性 P 值小于 0.05，不存在显著性差异，卡方检验通过。而纹饰与颜色的各样本之间存在显著性，不能满足假设前提，因此给予否定。

5.1.3 结果分析

根据 R-Q 型因子分析以及卡方检验可以判断：玻璃类型对文物表面风化影响较为显著，而纹饰与颜色则影响较小。

5.2 化学成分含量统计规律

5.2.1 数据预处理

附件表单二中给出了相应主要成分所占比例，由数据特点知，各成分比例的累加和应当为 100%，但因检测手段原因可能导致其成分累加和非 100%，于是选取累加和介于 85%-105%之间的数据视为有效数据。经过各成分加和处理后发现不满足条件的样本文物为 15 号以及 17 号，于是进行无效样本剔除处理。

5.2.2 分类汇总

根据 5.1 所得结论采取玻璃类型为分类依据，将有效数据分为高钾和铅钡两组，每一组中进一步以是否风化再次细分。考虑到样本数值的随机性与准确性，在进行各化学成分含量计算时采取均值处理法完成分类汇总操作。

表面风化	类型	氧化锡(Sn02)	二氧化硅(Si02)	二氧化硫(S02)	氧化钠(Na20)	五氧化二磷(P205)	氧化锶(Sr0)	氧化钡(Ba0)	氧化铅(Pb0)	氧化铝(Al203)	氧化铜(Cu0)	氧化铁(Fe203)	氧化镁(Mg0)	氧化钾(K20)	氧化钙(Ca0)
无风化	高钾	0.197	67.984	0.102	0.695	1.402	0.042	0.598	0.412	6.62	2.453	1.932	1.079	9.331	5.333
	铅钡	0.065	53.444	0.282	0.772	0.904	0.297	10.499	23.594	3.195	1.557	0.933	0.492	0.258	1.232
风化	高钾	0	93.963	0	0	0.28	0	0	0	1.93	1.562	0.265	0.197	0.543	0.87
	铅钡	0.056	33.615	0.987	0.953	4.155	0.366	10.487	36.872	3.838	1.996	0.556	0.701	0.143	2.346

表 5.6　化学成分分类汇总展示

5.2.3 结果分析

表 5.6 汇总整理出各化学成分含量的具体均值，并按照高钾无风化、铅钡无风化、高钾风化、铅钡风化依次排列，由此可以直观得到文物样品表面有无风化化学成分含量的统计规律。

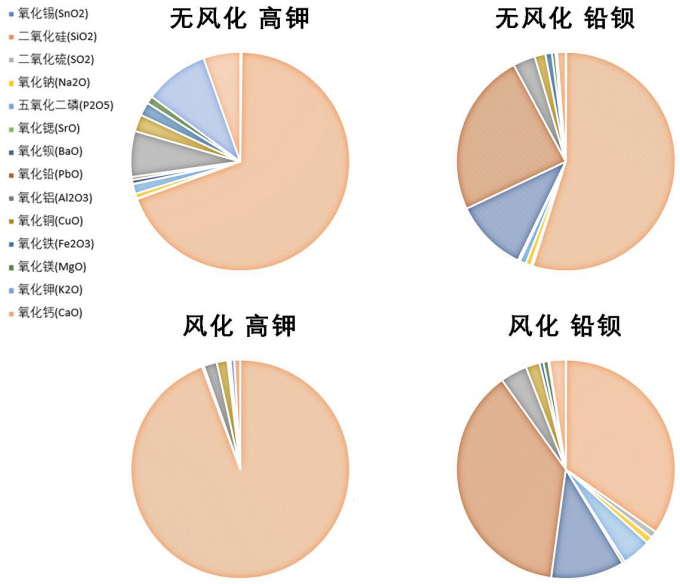
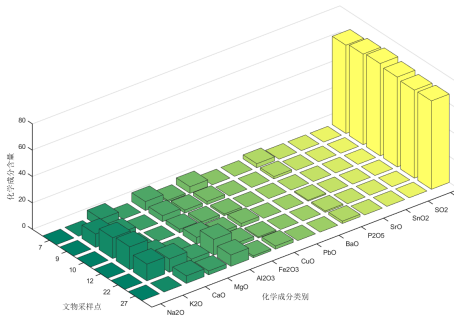


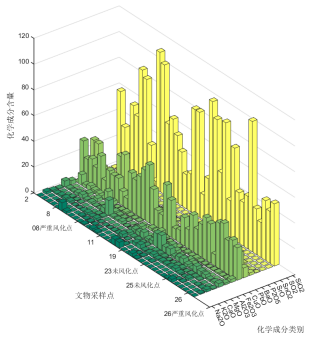
图 5.7　化学成分含量扇形统计图

5.3 风化前化学成分含量预测

根据前述化学成分含量统计规律，这里针对于未风化前成分含量的预测，选取四类（高钾未风化、高钾风化、铅钡未风化、铅钡风化）中的化学成分含量，分别计算上述分类汇总中的均值。利用各成分未风化前的均值与风化后的均值计算比值，与各组各成分数据对应相乘计算得未风化前的数值预测值。如下图所示：



高钾类预测值



铅钡类预测值

图 5.8　未风化前数值预测值

六、 问题二模型的建立、求解、分析

6.1 玻璃文物分类规律

6.1.1 分类概述

由附件可知，玻璃文物被分为高钾玻璃和铅钡玻璃两种类型，不同玻璃类型间化学成分的差异导致了这样的分类。为了进一步探究玻璃类型的分类依据，本文采取 CART 决策树模型对各化学成分含量数据进行进一步的研究。

6.1.2 CART 决策树模型

CART 算法是以最小分割 $Gini$ 系数的属性作为划分依据的二分递归算法，可以避免数据过分拟合，有效提高预测精度。其主要计算步骤如下:

(1) 计算初始 $Gini$ 系数值。针对样本训练集 Q，在此训练集中的属性为 A，根据每一属性 A，计算此时的初始 Gini 系数，计算公式如下:

$$Gini\,(Q) = \sum_{i=1}^{n} P_i\,(1 - P_i) = 1 - \sum_{i=1}^{n} {P_i}^2 \tag{1}$$

式中，n 为训练集中类别个数，P_i 为样本点属于第 n 类的概率。(2) 计算分割 Gini 系数值。针对训练集中的每一属性 A，以阈值 a 作为属性 A 的分割依据，将训练集 Q 分为 Q; 和 Qz 两个子集，分别计算两个子集的分割 Gini 系数值，数值越大，表示分割后集合出错的概率越大，计算公式如下:

$$Gini\,(Q, A) = \frac{|Q_1|}{|Q|} Gini\,(Q_1) + \frac{|Q_2|}{|Q|} Gini\,(Q_2) \tag{2}$$

式中，$Gini\,(Q, A)$ 为属性 A 的分割阈值为 a 时，训练集 Q 分割子集的错误概率。

(3) 确定最佳属性及分割阈值。对于每一属性 A，选择最小分割 $Gini$ 系数的属性及其阈值作为最佳划分依据，生成两个子节点，并进行样本划分 [4]。

依据上述步骤，分别对高钾玻璃以及铅钡玻璃进行画像构建。影响玻璃类型分类的因素共有 14 个，将 46 个样本作为训练集，21 个样本作为测试集，当交叉验证误差最小时，生成决策树分类图。

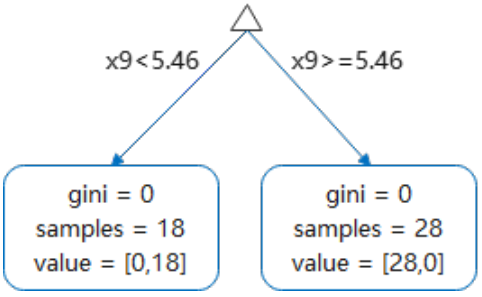


图 6.1　决策树分类图

依图可知，影响图像构建的拆分点为 x9 的含量百分比，即氧化铅（PbO）的含量百分比。以 PbO 含量 =5.46% 为分类依据，完成对玻璃类型的分类。

6.2 化学成分亚分类

6.2.1 随机森林

使用随机森林对客户流失进行预测，随机森林是集成学习方法 Bagging 的一个扩展变体 [5], 从名称上可以了解到该算法的两大特点: “随机” 和” 森林”。

(1)” 随机”。Bagging 算法基于自助采样法抽取 m 个训练样本，每 m 个训练样本用来训练一个基学习器，总共采样 T 个采样集，用来构建 T 个基学习器 [6]。自助采样法的采样” 随机” 是随机森林的” 随机” 之一; 另外，在基学习器训练的过程中，引入属性选择的” 随机”. 这也正是随机森林对 Bagging 的扩展之处。

(2)” 森林”。随机森林的基学习器为 CART 决策树，集成多棵决策树的学习结果确定模型最终结果，即为” 森林”。

由于每个基学习器训练样本和属性选择的” 随机性”, 各基学习器间差异性较大，进而提升了集成结果的泛化性 [7]。CART 决策树可以用于分类和回归，所以随机森林同样可以处理分类和回归问题，相较于单棵 CART 决策树，随机森林不需要进行剪枝，且不易产生过拟合现象.

具体步骤为:

(1)” 随机” 选择训练样本集。采用自助采样法从 N 个样本中采样 m 个训练样本，采样 T 轮得到 T 个样本集。

(2)” 随机” 生成 CART 决策树。每次训练决策树过程中，从 p 个属性中随机选择 k(k<p) 个属性用于构建 CART 决策树。

(3) 集成” 森林” 结果。T 棵决策树之间相互独立，重要性相等，因而在将它们进行组合时，认为它们具有相同的权值. 分类预测时，由所有的决策树投票确定最终分类结果; 回归预测时，所有决策树输出的均值作为最终的输出结果 [8]。

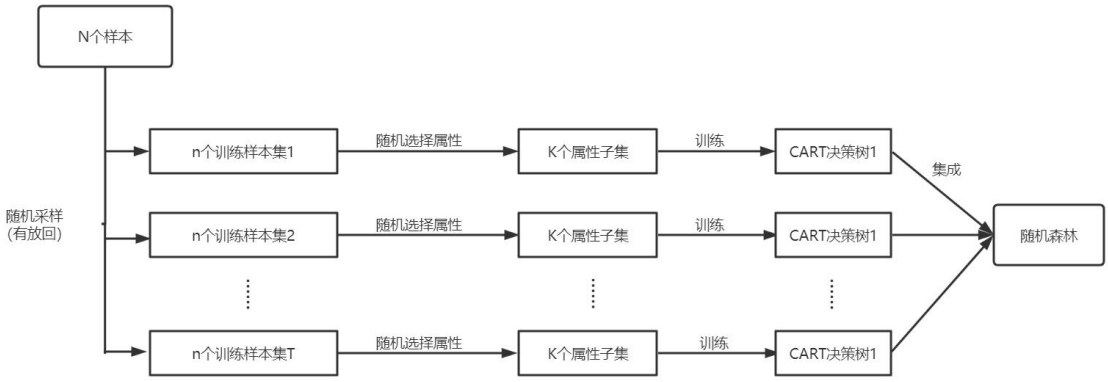


图 6.2　随机森林算法流程图

经由随机森林模型计算出的各化学成分在分类中的特征重要性比例如图:

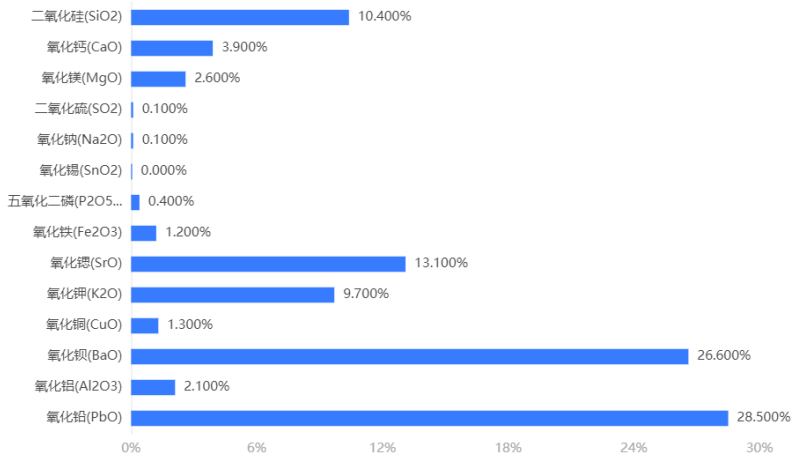


图 6.3　各成分特征重要性权重

6.2.2 K-means 聚类

一、数据标准化

数据标准化处理的目的是降低不用特征值的量纲和取值范围差异造成的影响，本文采用零——均值规范化方法将特征值映射到 [-1,1];

$$x^* = \frac{x - x_{mean}}{x_{std}}$$

其中，x^* 为归一化后的值，x_{mean} 为数据集中特征值的均值，x_{std} 为数据集中特征值的标准差。

二、基于 K-means 算法的优化

K-means 算法具有简单快速的处理流程，对于处理大数据集拥有较高效率。但其要求用户必须实现给出要生成的簇的数目 K，并且对初值以及孤立点数据较为敏感。基于以上 K-means 算法的缺点，我们对该算法进行了改进与优化，衍生出较为适用的 K-means++ 聚类算法。

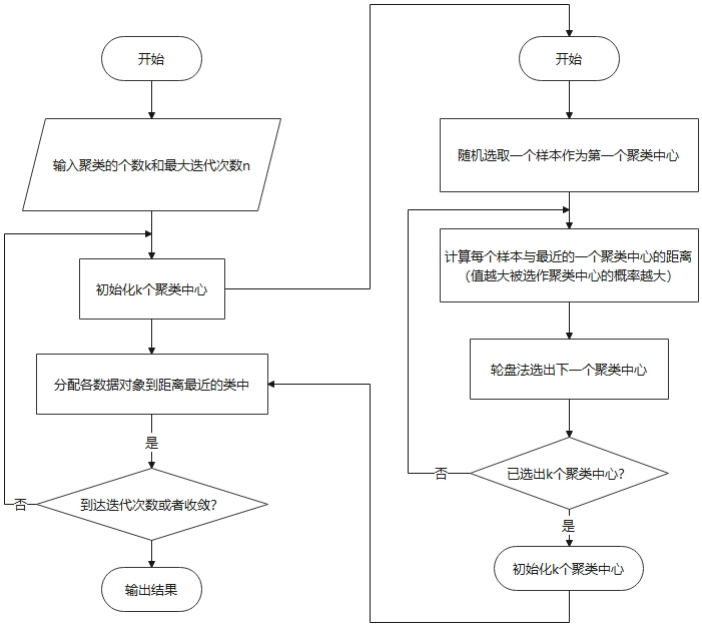


图 6.4　K-means 算法实现与优化

K-means++ 算法能够很好解决初值与孤立值敏感的问题。

6.2.3 结果分析

经由随机森林模型确定的分类权重以及 K-means 聚类算法确定的聚类中心和标准，选取权重最高的可以对每个玻璃类别根据各化学成分的比例做出以下亚分类：

1. 对于高钾玻璃类，氧化钡（BaO）、氧化铅（PbO）、氧化锶（SrO）含量百分比的影响不大，不作为分类标准考虑。而针对二氧化硅（SiO_2）以及氧化钾（K_2O）的含量，分类边界值为二氧化硅含量 63.62%、氧化钾含量 10.82%，并以二者的权重占比加和完成高钾玻璃类的进一步划分。

2. 对于铅钡玻璃类，氧化钾（K_2O）、氧化锶（SrO）含量百分比的影响不大，不作为分类标准考虑。而针对二氧化硅（SiO_2）、氧化钡（BaO）以及氧化铅（PbO）的含量，分类边界值为二氧化硅 57.49%、氧化钡 7.98%、氧化铅 19.88%，并以三者权重占比相加和完成铅钡玻璃类的进一步划分。

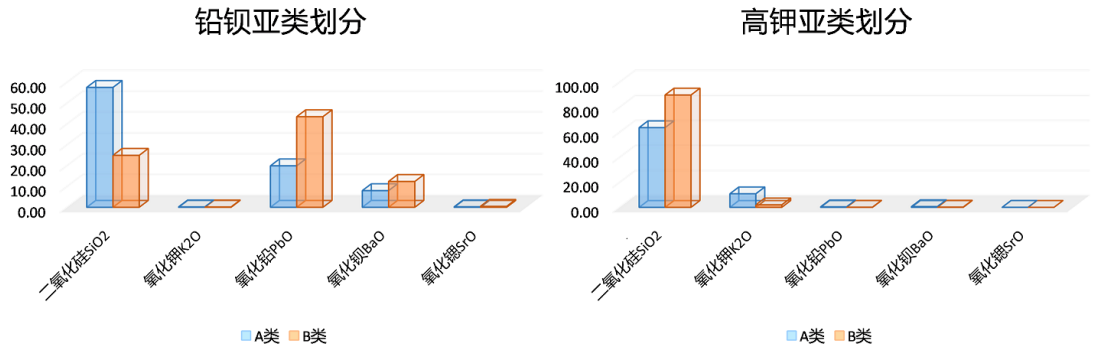


图 6.5　亚分类划分

6.2.4 合理性及敏感性分析

基于 K-means++ 聚类模型，我们选取正确率、误报率及漏报率作为评估指标。定义以下式子：

正确率：$CR = \frac{TP+TN}{TP+TN+FP+FN}$

误报率：$FR = \frac{FN}{TP+TN+FP+FN}$

漏报率：$MR = \frac{FP}{TP+TN+FP+FN}$

其中,TP 表示分类为正常样本的正常样本数量,TN 表示分类为异常样本的异常样本数量,FN 表示分类为异常样本的正常样本数量,FP 表示分类为正常样本的异常样本数量。

在代入样本集进行测试后，我们将其与标准 KNN 算法的运算指标进行对比，得到以下表格。

	准确率	误报率	漏报率
K-means++	99.97	0.01	0.01
标准 KNN	99.95	0.02	0.03

表 6.6　对比表

可以发现，改进优化后的 K-means++ 聚类模型在三方面表现均十分优秀，针对样本集的选取与分析能够满足要求。所以该模型的设计较为合理，能够完成稳定的亚分类工作。

七、 问题三模型的建立、求解、分析

7.1 鉴别未知文物的类别

7.1.1 逻辑回归模型

一、逻辑回归模型的构建过程

逻辑回归模型中的因变量为玻璃文物类型的二分类变量，自变量为玻璃文物的各种化学成分等变量。基于多因素建模的需要，本次模型选用了由线性回归模型发展而来的逻辑回归模型。逻辑回归模型能够在对自变量进行多维建模的同时纳入连续型及离散型自变量。

逻辑回归模型: 其因变量 Y 为二分类变量,即取值为 0 或 1,自变量为 X_1、$X_2 \ldots X_n$,P 表示在 n 维自变量的作用下 Y 发生的概率，当 P 大于等于某一值域 Q 时,Y 取 1; 当 P 小于某一值域 Q 时，Y 取 0。概率 P 的算法公式:

$$P = 1/\left(1 + e^{-z}\right) \tag{1}$$

公式中的 Z 为统计量，其算法如公式:

$$Z = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \ldots + \theta_n X_n \tag{2}$$

其中 θ_0 为常数项，表示所有自变量均取 0 时个体样本发生概率与不发生概率之比的自然对数的变化值，θ_1、θ_2 θ_n 为模型的回归系数，表示样本的某一特征 X 改变一个单位时，个体样本发生概率与不发生概率之比的自然对数的变化值，统计量 Z 的取值范围为 $(-\infty, +\infty)$，由逻辑回归函数将其映射到 P 上，P 的取值范围为 (0,1)。

本研究中的逻辑回归模型采用了最大似然估计求得模型的回归系数 θ_0、θ_1 θ_n。所有样本的预测值与其真实值一致的概率 $T(\theta)$ 最大时的回归系数 0 的取值即为所求，其算法公式:

$$T(\theta) = \prod P(y_i) \tag{3}$$

其中 $P(y_i)$ 为单个样本的预测值与其真实值一致的概率，其算法公式:

$$P(y_i) = X_i^{y_i}(1 - X_i)^{(1-y_i)} \tag{4}$$

对公式 (3）中等号两边进行取对数处理，然后对 0 求导数，当导数等于 0 时 0 的值即为所求。

二、将逻辑回归模型应用于分类与预测

在给定 x 的情况下，考虑 y 的两点分布概率

$$\text{s.t} \begin{cases} P(y = 1|x) = F(x, \beta) \\ P(y = 0|x) = 1 - F(x, \beta) \end{cases} \tag{5}$$

由于

$$E\left(y|x\right)=1\times P\left(y=1|x\right)+0\times P\left(y=0|x\right)=P\left(y=1|x\right) \tag{6}$$

将 $\hat{y}$ 表示为“y=1”的发生概率。

在这里 y=1 即表示玻璃文物的类型为高钾，y=0 表示玻璃文物的类型为铅钡。

$$\hat{y_i}=P\left(y_i=1|x\right)=S\left(x_i'\hat{\beta}\right)=\frac{exp\left(x_i'\hat{\beta}\right)}{1+exp\left(x_i'\hat{\beta}\right)}=\frac{e^{\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}+...+\hat{\beta}_kx_{ki}}}{1+e^{\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}+...+\hat{\beta}_kx_{ki}}} \tag{7}$$

如果 $\hat{y_i}\geqslant 0.5$, 则认为其预测的 y=1; 否则则认为其预测的 y=0。

三、模型结果

	铅钡	高钾	正确百分比 (%)
铅钡	49	0	100
高钾	0	18	100
总体百分比			100

表 7.1　回归模型预测正确率

根据表 7.1 可知此模型对于已知数据的预测效果非常好，为了避免出现过拟合现象，我们对已知数据分为训练组和测试组进行交叉验证，通过多次选取不同的训练组与测试组测试该模型的准确性，经验证模型比较准确。

同时据表 7.1，逻辑回归模型对于测试集预测准确。接下来将未知文物的化学成分数据（A1-A8）代入该逻辑回归模型，得到对于未知玻璃文物类型分类的预测值。

文物编号	类型
A1	1
A2	1
A3	0
A4	0
A5	0
A6	1
A7	1
A8	0

表 7.2　玻璃文物类型分类的预测

其中，1 代表高钾，0 代表铅钡。

7.1.2 结果分析

据表 7.2，可以初步判断出 A1、A2、A6、A7 属于高钾玻璃，A3、A4、A5、A8 属于铅钡玻璃。

7.2 敏感性分析

敏感性计算方式：

逻辑回归的判别函数为：

$$y = \frac{1}{1 + e^{-W^T x}} \tag{1}$$

其中

$$W^T x = (w_0 \cdot 1 + w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + \ldots + w_n \cdot x_n)$$

令

$$e^{-(w_0 \cdot 1 + w_2 \cdot x_2 + w_3 \cdot x_3 + \ldots + w_n \cdot x_n)} = m$$

即

$$m \cdot e^{-w_1 x_1} = e^{-W^T x} \tag{2}$$

于是对 x_1 求偏导得：

$$\frac{\partial y}{\partial x_1} = \frac{\partial \left(\frac{1}{1+e^{-W^T x}} \right)}{\partial x_1} = w_1 \cdot e^{-w_1 x_1} \cdot \left(1 + m \cdot e^{-W^T x} \right)^{-2}$$

将 m 代入后得：

$$\frac{\partial y}{\partial x_1} = w_1 \cdot e^{-W^T x} \cdot \left(1 + e^{-W^T x} \right)^{-2}$$

同理可得：

$$\frac{\partial y}{\partial x_2} = w_2 \cdot e^{-W^T x} \cdot \left(1 + e^{-W^T x} \right)^{-2}$$

$$\frac{\partial y}{\partial x_i} = w_i \cdot e^{-W^T x} \cdot \left(1 + e^{-W^T x} \right)^{-2} \tag{3}$$

最终得到每个特征的敏感性的计算公式为:

$$S = \frac{1}{n} \sum_{j=1}^{n} \left| \frac{\partial y}{\partial x_i^j} \right| \tag{4}$$

这里 i 表示第 i 个特征，j 表示第 j 个样本 [9]。

敏感性检验步骤：选用附件中给出的已知玻璃文物类型数据集中随机五组进行训练和预测，用逻辑回归和求偏导数两种方法计算出敏感性特征值，分别利用二者与真实数据的特征值的关系，计算逻辑回归模型和求偏导数方法各自的测试精确率。

最后画出二者数据折线图，更加直观地展示结果。

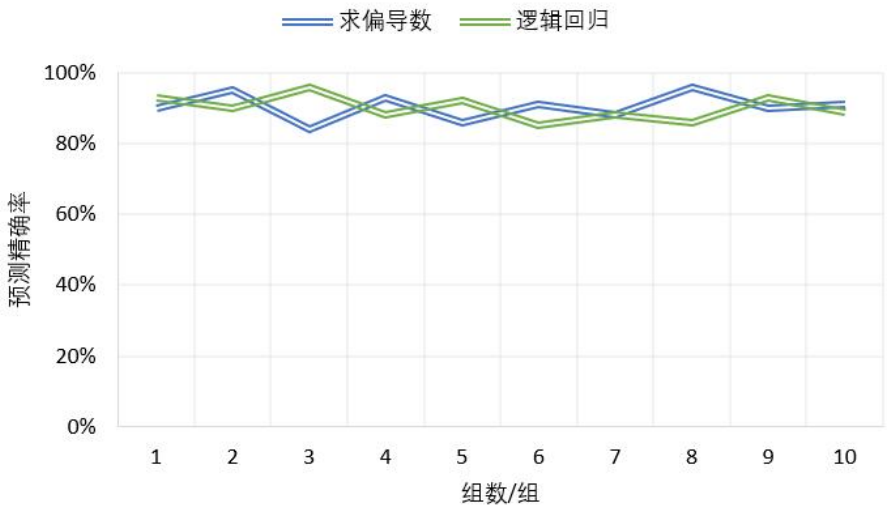


图 7.3　求偏导数与逻辑回归精确率对比

由图 7.3 可以看出，该逻辑回归模型的精确率与求偏导数得出的精确率较为吻合，且预测的精确率也较为稳定。

八、 问题四模型的建立、求解、分析

8.1 同类别成分关联性分析

8.1.1 相关性分析

针对于高钾玻璃与铅钡玻璃这两个类别中各自化学成分关联关系的研究，首先考虑到对各化学成分两两间的相关程度进行分析，这里采用皮尔逊相关系数法。

两个变量之间的皮尔逊相关系数定义为两个变量之间的协方差和标准差的商 [10]:

$$\rho_{X,Y} = \frac{cov\left(X,Y\right)}{\sigma_X \sigma_Y} = \frac{E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]}{\sigma_X \sigma_Y} \tag{1}$$

上式定义了总体相关系数，常用希腊小写字母 ρ 作为代表符号。估算样本的协方差和标准差，可得到皮尔逊相关系数，常用英文小写字母 r 代表:

$$r = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{X}\right)^2}} \tag{2}$$

r 亦可由 (X_i, Y_i) 样本点的标准分数均值估计，得到与上式等价的表达式:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma_X}\right)\left(\frac{Y_i-\bar{Y}}{\sigma_Y}\right) \tag{3}$$

其中 $\frac{X_i-\bar{X}}{\sigma_X}$、$\bar{X}$ 及 σ_X 分别是对 X_i 样本的标准分数、样本平均值和样本标准差。

由此计算出来两种类别各自的化学成分之间的相关系数，并用热力图的形式表示出来。

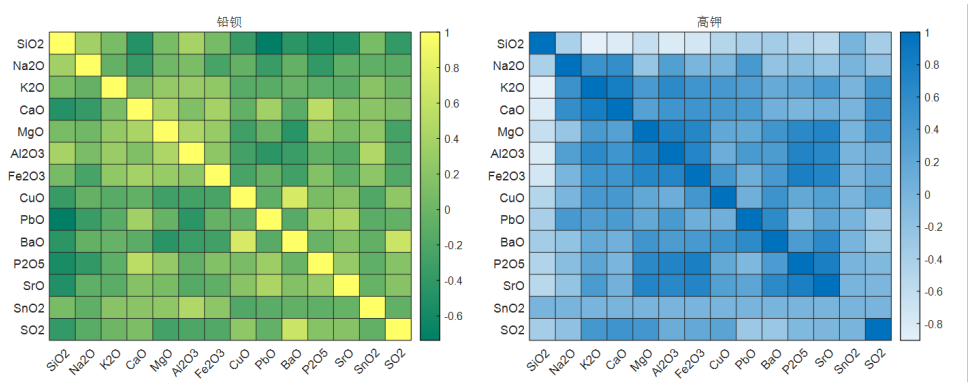


图 8.1　相关系数热力图

8.1.2 灰色关联分析

为了检验上述利用皮尔逊系数进行的相关性分析是否准确可靠，我们建立了灰色关联分析模型对单个指标进行准确性检验。

一、模型的构建：

1. 确定特征数列和母数列

比较序列为

$$\begin{bmatrix} X_1' & X_2' & \cdots X_n' \end{bmatrix} = \begin{bmatrix} x_1'(1) & x_2'(1) & \cdots & x_n'(1) \\ x_1'(2) & x_2'(2) & \cdots & x_n'(2) \\ \cdots & \cdots & & \cdots \\ x_1'(m) & x_2'(m) & \cdots & x_n'(m) \end{bmatrix}$$

母序列 (即评价标准) 为 [11]

$$X_0' = (x_0'(1), x_0'(2), \ldots, x_0'(m))^T$$

2. 对样本数据进行标准化处理。

3. 由下式计算关联系数

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta min + \rho \Delta max}{\Delta_{ik} + \rho \Delta max}$$

$$\Delta min = \min_i \min_k |x_0(k), x_i(k)|$$

$$\Delta max = \max_i \max_k |x_0(k), x_i(k)|$$

$$\Delta_{ik} = |x_0(k), x_i(k)|$$

其中 ρ 为分辨系数 [12]，这里取 0.5。

4. 计算关联序度

$$r_{0i} = \frac{1}{m} \sum_{k=1}^{m} W_k \zeta_i(k)$$

5. 得出计算结果

关联度越大，则计算所得结果越接近 1。

二、数据的选取与计算

根据相关性分析所得相关系数热力图可以直观看出一些化学成分含量的关联性。为了进一步检验，本次仅选取氧化钾（K_2O）、氧化铝（Al_2O_3）、氧化钡（BaO）、氧化锶（SrO）与二氧化硅（SiO_2）的相关性进行分析。最终得到以下结果：

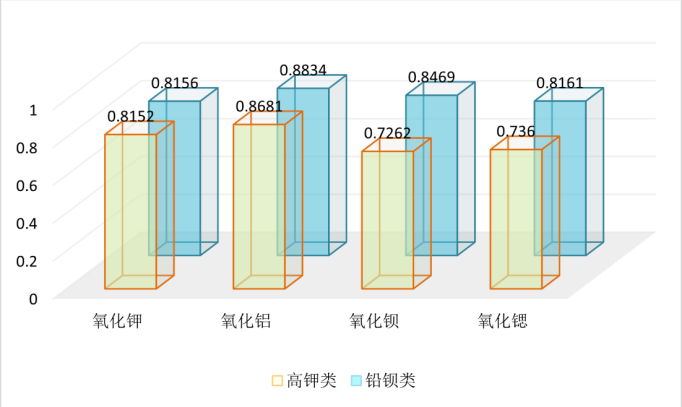


图 8.2　灰色关联分析图

8.1.3 结果分析

由灰色关联分析对单个数据完成的分析检验与相关性分析热力图中的值对比相似，可以确定经由相关性分析确定的相关系数准确可靠。结合两个相关系数热力图可以直观看出两种类型各化学成分之间的关联关系。

8.2 不同类别间成分差异性分析

8.2.1 差异性分析

上文中 R-Q 型因子分析模型可以有效计算样本数据间的显著性问题，针对此问题的差异性分析依旧采用计算 P 值表示显著性，以进一步研究判断。将数据按类型分开，两两一组进行配对，如：高钾中二氧化硅含量与铅钡中二氧化硅含量配对。如此进行，完成共计 14 组配对，代入模型，计算得各组 P 值及相关差异性效应值如下：

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3
P	0.077**	0.015**	0.028**	0.054*	0.010***	0.002***	0.061*
效应量	0.356	0.536	0.693	0.284	0.594	0.596	0.143

	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
P	0.000***	0.000***	0.000***	0.000***	0.000***	0.5	0.093**
效应量	0.544	3.145	1.752	1.027	1.802	0.035	0.349

***、**、* 分别代表 1%、5%、10% 的显著性水平

表 8.3　显著值与效应量表

根据 P 值大小以及显著值对应关系可以得到：当 P<0.05 时，存在显著性。根据效应值及其对应关系可以得到：0.20 以下表示效应过小，0.20-0.50 表示效应偏小，0.50-0.80 表示效应较大，0.80 以上表示大效应。如：配对样本氧化钾（K2O）检验的结果显示，其显著性 P 值为 0.028**，水平上呈现显著性，拒绝原假设，因此该氧化钾配对之间存在显著性差异。其差异幅度效应值为：0.693，差异幅度中等。

据此我们可以得到以下两类间各化学成分含量的差异关系：

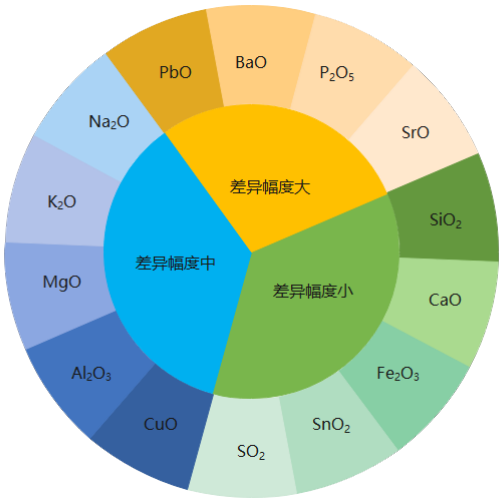


图 8.4　两类间各化学成分含量的差异关系

8.3 结论

据上述分类及显著性及效应值表能够初步得出高钾玻璃与铅钡玻璃在化学成分含量数值上的差异，根据差异性幅度能够进一步具体探究完成差异性关联度的分析工作。

九、 模型评价

9.1 模型与论文的优点

- 合理的假设：本文基于大量文献阅读，通过对问题的深入研究与分析建立了一系列科学的假设，将一些对结果影响甚微的因素予以忽略，简化了模型与算法，提高了模型的运算效率。
- 结果可视化：本文使用 MATLAB 软件对相关模型的运算及相关变化规律进行了可视化工作，使得模型运算结果更加直观。
- 模型创新型：在问题二中，我们改进了传统 K-means 聚类算法，优化了其运算过程，在其基础上衍生出 K-means++ 算法，使得模型能够更加切合题目。
- 模型泛用性：在解题过程中建立的分类模型、对应分析模型等，均可以更换不同样本集来适应不同实际情况，模型泛用性良好。

9.2 模型的缺点与不足

- 在进行未风化前数据的预测中，基于图像拟合效果直接采用了均值乘积方法，未进一步检验该不操作的合理性，缺乏严谨。
- 未对建立的模型进行仿真验证，具有一定的局限性。

参考文献

[1] 程浩, 王起琮, 景帅.R-Q 因子分析法在储层分类评价中的应用——以塔里木盆地奥陶系碳酸盐岩储层为例 [J]. 中国资源综合利用,2019,37(11):161-165.

[2] 孙宁. 对应分析法在储层定量分类评价中的应用——以子长油田安定区长 6 储层为例 [J]. 中国石油和化工标准与质量,2019,39(01): 107-108+110.

[3] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1 .0.11)[Online Application Software]. Retrieved from https://www.spsspro .com.

[4] 魏山森, 梁建芳, 雷钦渊. 基于 CART 决策树算法的服装可持续消费者画像构建 [J/OL]. 丝绸:1-10[2022-09-18] .http://kns.cnki.net/kcms/detail /33.1122.TS.20220822.1600.002. html

[5] 曹正凤, 谢邦昌, 纪宏. 一种随机森林的混合算法 [J]. 统计与决策,2014(04):7-9.DOI:10. 13546/j.cnki.tjyjc.000109.

[6] 刘余霞, 吕虹, 胡涛, 孙小虎. 基于 Bagging 集成学习的字符识别方法 [J]. 计算机工程与应用, 2012,48(33):194-196+211.

[7] 周志华. 机器学习 [M]. 北京：清华大学出版社.2016：178-180

[8] 龙冰婷. 基于随机森林和 K-means 算法的网络视频客户流失预测与分析 [J]. 湖北民族大学学报 (自然科学版),2022,40(02):202-207.DOI:10.13501/j.cnki.42-1908/n.2022.06.013.

[9] 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.cn13-1295/tq.2022.07.051.

[10] [4]Piantadosi, J.; Howlett, P.; Boland, J. (2007) ”Matching the grade correlation coefficient using a copula with maximum disorder”, Journal of Industrial and Management Optimization, 3 (2), 305–312

[11] ohamInadA，Daniel N，PeterI C. Fu2zy grey relational analvsis for software effort estimation[J]. Empircal Sotware Engineering，2010，15(1)：60—90.

[12] 付雅芳, 杨任农, 刘晓东, 等. 基于灰色关联分析的软件工作量估算方法 [J]. 系统工程与电子技术,2012,34(11):2384-2389.

附录 A　支撑材料文件列表

序号	支撑材料文件名
1	0.1.xlsx
2	0.2.xlsx
3	0.xlsx
4	1.3.xlsx
5	2.2.1.sav
6	2.2.2.sav
7	2.2.xlsx
8	3.1.sav
9	3.1.xlsx
10	code4.m
11	gray_anal.xls
12	HighK.mat
13	img.m
14	img.mat
15	k方.sav
16	map.m
17	map.mat
18	PbBa.mat
19	relate.xlsx
20	res3.1.xlsx
21	决策树.sav

附录 B　卡方检验 SPSS 指令

```
GET DATA
  /TYPE=XLSX
  /FILE='D:\OneDrive\桌面\国赛\附件.xlsx'
  /SHEET=name '表单1'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
CROSSTABS
  /TABLES=纹饰 类型 颜色 表面风化 BY 文物编号
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

附录 C　CART 决策树 SPSS 指令

```
DATASET ACTIVATE 数据集1.
DATASET CLOSE 数据集2.

GET DATA
  /TYPE=XLSX
  /FILE='D:\OneDrive\桌面\code\0.xlsx'
  /SHEET=name '铅钡'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集3 WINDOW=FRONT.
DATASET ACTIVATE 数据集1.
DATASET CLOSE 数据集3.

GET DATA
  /TYPE=XLSX
  /FILE='D:\OneDrive\桌面\code\0.2.xlsx'
  /SHEET=name '表单'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集4 WINDOW=FRONT.
* 决策树.
TREE 类型 [n] BY 二氧化硅SiO2 [s] 氧化钠Na2O [s] 氧化钾K2O [s] 氧化钙CaO [s] 氧化镁MgO [s]
    氧化铝Al2O3 [s] 氧化铁Fe2O3 [s]
    氧化铜CuO [s] 氧化铅PbO [s] 氧化钡BaO [s] 五氧化二磷P2O5 [s] 氧化锶SrO [s] 氧化锡SnO2 [s]
       二氧化硫SO2 [s]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO
  /DEPCATEGORIES USEVALUES=[VALID]
  /PRINT MODELSUMMARY CLASSIFICATION RISK
  /METHOD TYPE=CHAID
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50
  /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
  /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO CHISQUARE=PEARSON CONVERGE=0.001
    MAXITERATIONS=100 ADJUST=BONFERRONI INTERVALS=10.
```

附录 D　K-means++ 高钾类 SPSS 指令

```
GET DATA
  /TYPE=XLSX
  /FILE='D:\OneDrive\桌面\code\2.2.xlsx'
  /SHEET=name '高钾2'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化铅PbO 氧化钡BaO 氧化锶SrO
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(30) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER DISTANCE
  /PRINT ID(文物采样点) INITIAL.
```

附录 E　K-means++ 铅钡类 SPSS 指令

```
GET DATA
  /TYPE=XLSX
  /FILE='D:\OneDrive\桌面\code\2.2.xlsx'
  /SHEET=name '铅钡2'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集2 WINDOW=FRONT.
DATASET CLOSE 数据集1.

SAVE OUTFILE='D:\OneDrive\桌面\code\2.2.2.sav'
  /COMPRESSED.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化铅PbO 氧化钡BaO 氧化锶SrO
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(50) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER
  /PRINT ID(文物采样点) INITIAL.
```

附录 F　逻辑回归模型 SPSS 指令

```
SET TLook=None
FOOTNOTE=ON
Small=0.0001 SUMMARY=None THREADS=AUTO Printback=On SIGLESS=YES TFit=Both DIGITGROUPING=No
    LEADZERO=No TABLERENDER=light.
LOGISTIC REGRESSION VARIABLES
虚拟变量
  /METHOD=ENTER 二氧化硅SiO2 氧化钠Na2O 氧化钾K2O 氧化钙CaO 氧化镁MgO 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
  /SAVE=PRED PGROUP
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

附录 G　灰色关联分析 matlab 代码

```
%% 4.1灰色关联分析
clear;clc
%分别计算高钾类和铅钡类
%load HighK.mat % 导入高钾类数据
%disp("高钾类：")
load PbBa.mat %导入铅钡类数据
disp("铅钡类：")
Average = mean(data); % 每一列均值
data = data ./ repmat(Average,size(data,1),1); %进行数据预处理
  Y = data(:,1); % 因变量（母序列）
  X = data(:,2:end); %自变量（子序列）
  num_abs = abs(X - repmat(Y,1,size(X,2)));
  %计算两级最大差和最小差
  min_num = min(min(num_abs));
  max_num = max(max(num_abs));

  r = 0.5; %分辨系数
  gam = (min_num+r*max_num) ./ (num_abs + r*max_num); % 计算关联度
  disp('氧化钾，氧化铝，氧化钡，氧化锶与二氧化硅的灰色关联度分别为：')
  disp(mean(gam))
  xlswrite('gray_anal',mean(gam));
```

```
%对预测结果进行可视化
clear;clc
%读取数据
load img.mat
bar3(highK); %高钾类
%bar3(pbBa);  %铅钡类
title = title('高钾类预测值');
```

```
%title = title('铅钡类预测值');
x_values = {'Na2O','K2O','CaO','MgO','Al2O3','Fe2O3',...
    'CuO','PbO','BaO','P2O5','SrO','SnO2','SO2','SiO2'};
 y_label_highK = {7,9,10,12,22,27};
 y_label_pbba = {'2','8','08严重风化点',11,19,'23未风化点','25未风化点',26,...
'26严重风化点','28未风化点','29未风化点',34,36,38,39,40,41,'42未风化点1','42未风化点2',...
    '43部位1','43部位2','44未风化点',48,49,'49未风化点',50,'50未风化点',...
    '51部位1','51部位2',52,'53未风化点',54,'54严重风化点',56,57,58};
 %设置坐标轴刻度以及一些细节
 set(gca, 'Box', 'off', ...
        'LineWidth', 1, 'GridLineStyle', '-',...
        'XGrid', 'off', 'YGrid', 'off','ZGrid', 'on', ...
        'TickDir', 'out', 'TickLength', [.015 .015], ...
        'XMinorTick', 'off', 'YMinorTick', 'off', 'ZMinorTick', 'off',...
        'XColor', [.1 .1 .1], 'YColor', [.1 .1 .1], 'ZColor', [.1 .1 .1],...
        'Xticklabel',x_values,...
        'Yticklabel',y_label_highK)
 % set(gca, 'Box', 'off', ...
 %         'LineWidth', 1, 'GridLineStyle', '-',...
 %         'XGrid', 'off', 'YGrid', 'off','ZGrid', 'on', ...
 %         'TickDir', 'out', 'TickLength', [.015 .015], ...
 %         'XMinorTick', 'off', 'YMinorTick', 'off', 'ZMinorTick', 'off',...
 %         'XColor', [.1 .1 .1], 'YColor', [.1 .1 .1], 'ZColor', [.1 .1 .1],...
 %         'Xticklabel',x_values,...
 %         'Yticklabel',y_label_pbba)
 x_label = xlabel('化学成分类别');
 y_label = ylabel('文物采样点');
 z_label = zlabel('化学成分含量');
 %设置字号
 set(gca, 'FontSize', 10)
 set(gca, 'FontName', 'Helvetica')
 set(title, 'FontSize', 12, 'FontWeight' , 'bold')
 set([x_label, y_label,z_label], 'FontName', 'AvantGarde')
 set([x_label, y_label,z_label], 'FontSize', 12)
```

```
%热力图
clear;clc
%加载数据文件
load map.mat
%设置横纵坐标
x_values =
    {'SiO2','Na2O','K2O','CaO','MgO','Al2O3','Fe2O3','CuO','PbO','BaO','P2O5','SrO','SnO2','SO2'};
y_values =
    {'SiO2','Na2O','K2O','CaO','MgO','Al2O3','Fe2O3','CuO','PbO','BaO','P2O5','SrO','SnO2','SO2'};
pb = heatmap(x_values,y_values,pbba); %铅钡
```

```
9. pb.Title = '铅钡';
10.  hi = heatmap(x_values,y_values,highK); %高钾
11.  hi.Title = '高钾';
```

附录 H　5.3 高钾预测结果

文物采样点	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)	二氧化硅 (SiO_2)
7	0	0	6.558362	0	6.791503	1.239182	5.088218	0	0	3.055446	0	0	0	67.01948
9	0	10.13225	3.800172	0	4.527668	2.332579	2.434178	0	0	1.753125	0	0	0	68.74868
10	0	15.79945	1.287155	0	2.778342	1.89522	1.319168	0	0	0	0	0	0	70.01484
12	0	17.34505	4.413103	0	5.007876	2.113899	2.591222	0	0	0.751339	0	0	0	68.22052
22	0	12.70825	10.17466	3.511864	12.00518	2.551258	0.863741	0	0	1.051875	0	0	0	66.81689
27	0	0	5.761552	2.963136	8.60943	1.457862	2.418474	0	0	1.803214	0	0	0	67.08459

附录 I　5.3 铅钡预测结果

文物采样点	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)	二氧化硅 (SiO_2)
2	0	1.90075	1.22862	0.8289	4.76904	3.12238	0.20279	30.3498	0	0.77654	0.15398	0	0	57.6814
8	0	0	0.77708	0	1.11527	0	8.11957	18.3519	31.2658	0.78089	0.29985	0	0.73619	32.0205
08严重风化点	0	0	1.67492	0	0.92384	0	2.44913	20.7643	30.6551	1.64443	0.42951	0	4.28871	7.32941
11	0	0.38015	1.84293	0.49875	2.23887	0	3.84529	16.2467	14.6267	2.04032	0.29985	0	0	53.4045
19	0	0	1.5384	0.41445	2.97128	2.23267	2.73772	27.3999	5.35613	1.92068	0.15398	0	0	47.1245
23未风化点	6.41531	0	0.26253	0.49875	1.18186	0	2.33213	10.8653	11.8736	0	0.26743	0	0	85.5204
25未风化点	1.87113	0	0.33078	0	1.58136	2.60198	0.87358	20.4124	6.65761	0.04133	0.16208	0	0	80.4645
26	0	0	0.75607	0	0.5826	0	8.24437	18.8958	32.2869	0.68083	0.36468	0	0.55927	31.464
26严重风化点	0	0.72409	1.58041	0	0.9821	0	2.80792	19.1454	35.4906	1.31381	0.50245	0	4.55122	5.91441
28未风化点	0	0.47066	0.70357	0.70246	3.91177	0.68827	0.25739	10.9676	4.04463	0.22622	0.09725	0.26618	0	108.24
29未风化点	0.74521	0.54307	1.56465	1.04667	11.9351	1.35975	0.57718	7.877	2.03232	0.08918	0.2026	0	0	100.64
34	0	0.45256	0.40954	0	1.34831	0.78899	1.17777	29.7867	10.0115	0.07396	0.17829	0	0	56.8864
36	1.79823	0.25343	0.19427	0	1.33167	0.53718	0.53039	26.6257	10.8424	0.01523	0.17829	0	0	62.9121
38	1.11782	0	0.35704	0	2.13899	0.48682	0.56938	31.5528	9.80121	0.10441	0.33227	0	0	52.3552
39	0	0	0.58281	0	0.41615	0	0.68638	39.0523	7.22827	0.25232	0.49435	0	0	41.7347
40	0	0	0.98185	0	0.37453	0.31895	0	44.9264	6.69766	0.38501	0.55107	0	0	26.5671
41	0	0.7965	2.60426	1.91772	2.77153	3.00487	0.1482	28.2318	9.77118	1.62268	0.38089	0	0	29.3494
42未风化点1	4.64948	0.27154	0.41479	0.76568	2.93799	0	2.08254	14.0007	10.482	0.0174	0.28364	0	0	81.498
42未风化点2	4.60088	0.63358	0	0.81485	4.71077	0	2.12154	12.8745	10.8925	0	0	0	0	81.6093
43部位1	0	0	2.75127	0.62519	1.87266	1.27581	4.17288	38.2972	7.29835	0	0.51866	0	0	19.7306
43部位2	0	0	3.36033	0.66734	2.83812	2.33339	1.17777	28.6349	3.26373	2.79075	0.38089	0	0	34.5007
44未风化点	2.47864	0.36205	1.12361	0	10.5618	1.2926	0.33539	8.70885	5.22598	0	0.21071	0	0	96.5702
48	0.64801	0.57928	1.48065	1.08179	11.3608	1.72906	0	10.0526	7.31837	0.23927	0.2026	1.51605	0	84.789
49	0	0	2.40474	1.03262	4.47773	4.59964	0.54598	21.8713	6.10698	2.41445	0.37279	0	0	45.773
49未风化点	0	0.54307	1.09211	0.84295	5.4099	2.13195	0.35099	14.7302	4.1948	0.93968	0.24312	0	0	86.8241
50	0	0	1.67492	0.33016	1.55639	0.55397	0.88138	28.155	14.2163	1.37906	0.53487	0	0	28.5863
50未风化点	0	0	1.63816	0.37933	3.46234	0	0.54598	19.5869	6.22712	1.37906	0.18639	0	0	71.577
51部位1	0	0	1.87969	0.83593	4.36953	1.99765	1.06857	25.749	8.95024	1.76189	0.31606	0.54393	0	39.1273
51部位2	0	0	2.69352	1.01857	2.08905	0.70505	0.58498	32.8518	0	1.90328	0	0	0	33.9442
52	0.98822	0	1.19187	0.38635	0.96546	0.3861	0.54598	30.3434	8.64989	1.24203	0.35658	0	0	40.9239
53未风化点	2.46244	0.19913	0.40954	0.8008	5.04369	0	0.42119	8.74084	9.00029	0	0.21881	0	0	101.213
54	0	0.57928	1.67492	0.89915	3.45401	0	0.64738	35.4881	7.04806	0.92227	0.71316	0	0	35.4228
54严重风化点	0	0	0	0.77973	3.03787	0	1.04517	37.4077	0	3.07352	0.90765	0	0	27.2031
56	0	0	0.63531	0	1.53974	0	0.61618	26.3953	15.4677	0.55249	0	0	0	46.3454
57	0	0	0.68782	0	1.8144	0	0.90477	28.8589	17.3198	0	0	0	0	40.4151
58	0	0.61548	1.83243	0.55494	2.92967	1.44368	2.44133	25.1795	7.66877	1.95548	0.1945	0	0	48.3169

附录 J　8.1 高钾相关系数表

	二氧化硅	氧化钠	氧化钾	氧化钙	氧化镁	氧化铝	氧化铁	氧化铜	氧化铅	氧化钡	五氧化二磷	氧化锶	氧化锡	二氧化硫
二氧化硅	1	-0.414	-0.903	-0.827	-0.636	-0.815	-0.742	-0.499	-0.4	-0.355	-0.476	-0.538	0	-0.37
氧化钠	-0.414	1	0.5	0.567	-0.25	0.315	-0.004	-0.01	0.406	-0.214	-0.142	-0.227	0	-0.194
氧化钾	-0.903	0.5	1	0.811	0.393	0.621	0.466	0.42	0.343	0.149	0.228	0.383	0	0.423
氧化钙	-0.827	0.567	0.811	1	0.293	0.465	0.483	0.417	0.377	0.056	-0.009	0.048	0	0.475
氧化镁	-0.636	-0.25	0.393	0.293	1	0.769	0.674	0.187	0.172	0.458	0.637	0.699	0	0.435
氧化铝	-0.815	0.315	0.621	0.465	0.769	1	0.691	0.098	0.397	0.362	0.717	0.658	0	0.107
氧化铁	-0.742	-0.004	0.466	0.483	0.674	0.691	1	0.44	0.084	0.412	0.784	0.699	0	0.18
氧化铜	-0.499	-0.01	0.42	0.417	0.187	0.098	0.44	1	0.072	0.491	0.171	0.305	0	0.243
氧化铅	-0.4	0.406	0.343	0.377	0.172	0.397	0.084	0.072	1	0.614	-0.057	0.222	0	-0.299
氧化钡	-0.355	-0.214	0.149	0.056	0.458	0.362	0.412	0.491	0.614	1	0.357	0.627	0	-0.272
五氧化二磷	-0.476	-0.142	0.228	-0.009	0.637	0.717	0.784	0.171	-0.057	0.357	1	0.779	0	-0.066
氧化锶	-0.538	-0.227	0.383	0.048	0.699	0.658	0.699	0.305	0.222	0.627	0.779	1	0	-0.032
氧化锡	0	0	0	0	0	0	0	0	0	0	0	0	0	0
二氧化硫	-0.37	-0.194	0.423	0.475	0.435	0.107	0.18	0.243	-0.299	-0.272	-0.066	-0.032	0	1

附录 K　8.1 铅钡相关系数表

	二氧化硅	氧化钠	氧化钾	氧化钙	氧化镁	氧化铝	氧化铁	氧化铜	氧化铅	氧化钡	五氧化二磷	氧化锶	氧化锡	二氧化硫
二氧化硅	1	0.362	0.087	-0.488	0.088	0.401	0.082	-0.354	-0.738	-0.436	-0.565	-0.503	0.085	-0.386
氧化钠	0.362	1	-0.059	-0.373	0.026	0.103	-0.244	-0.058	-0.349	-0.056	-0.396	-0.093	-0.083	-0.13
氧化钾	0.087	-0.059	1	0.096	0.269	0.306	0.248	-0.146	-0.131	-0.041	-0.106	-0.121	0.195	0.01
氧化钙	-0.488	-0.373	0.096	1	0.417	0.135	0.387	-0.068	0.356	-0.124	0.535	0.171	0.21	0.104
氧化镁	0.088	0.026	0.269	0.417	1	0.45	0.293	-0.309	-0.036	-0.453	0.272	0.087	0.246	-0.266
氧化铝	0.401	0.103	0.306	0.135	0.45	1	0.23	-0.282	-0.441	-0.339	-0.074	-0.155	0.469	-0.206
氧化铁	0.082	-0.244	0.248	0.387	0.293	0.23	1	-0.247	-0.048	-0.295	0.153	-0.091	0.226	-0.18
氧化铜	-0.354	-0.058	-0.146	-0.068	-0.309	-0.282	-0.247	1	-0.085	0.719	0.096	0.196	-0.184	0.241
氧化铅	-0.738	-0.349	-0.131	0.356	-0.036	-0.441	-0.048	-0.085	1	-0.14	0.331	0.434	-0.126	-0.065
氧化钡	-0.436	-0.056	-0.041	-0.124	-0.453	-0.339	-0.295	0.719	-0.14	1	-0.02	0.164	-0.077	0.634
五氧化二磷	-0.565	-0.396	-0.106	0.535	0.272	-0.074	0.153	0.096	0.331	-0.02	1	0.28	-0.082	0.173
氧化锶	-0.503	-0.093	-0.121	0.171	0.087	-0.155	-0.091	0.196	0.434	0.164	0.28	1	-0.038	0.181
氧化锡	0.085	-0.083	0.195	0.21	0.246	0.469	0.226	-0.184	-0.126	-0.077	-0.082	-0.038	1	-0.071
二氧化硫	-0.386	-0.13	0.01	0.104	-0.266	-0.206	-0.18	0.241	-0.065	0.634	0.173	0.181	-0.071	1