

古代玻璃制品的成分分析与鉴别

摘要

古代玻璃制品是古代丝绸之路的重要历史见证，不仅具有极高的艺术价值和收藏价值，更揭示出我古代化学工艺制造水平，对其化学成分的研究对考古学者来说意义重大。但是其易风化的特性为考古工作者的研究带来了极大的不便。基于此，本文以玻璃化学成分含量为突破口，力求得出风化前后古代玻璃化学成分含量的统计规律，并进行合理的亚类划分，正确判别分析不同类别的玻璃文物样品化学成分之间的关联关系，以此求解得出符合题目要求的结论，更好地帮助考古工作者进行科考研究。

针对问题一，通过卡方检验来反映玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系，结果显示在 95% 的置信水平下，不同类型的玻璃文物样品之间表面风化有显著差异，而不同颜色和纹饰之间无显著差异。观测可知，铅钡玻璃比高钾玻璃更易风化。在根据玻璃类型和是否风化进行划分为四类后，分别对四类数据进行描述性统计和绘制相关图表进行可视化，可明显观测出结风化前后化学成分含量的特征及变化。对于风化前的化学成分估计，通过对样本进行预处理之后，将样本按是否风化分为两类，并计算出两类样本各种化学成分均值，其差值作为风化过程使得任一样本受风化后化学成分的变化量的近似估计，风化样本风化后的化学成分加上这个差值就可估计风化前的化学含量。

针对问题二，采用决策树模型从化学成分及纹饰类型等特征中找出氧化铅是最能解释类别划分的变量，并通过十折交叉检验验证决策树的分类准确性和可靠性较高，采取决策树对两者的分类标准作为参考分类规律。当玻璃样本的氧化铅含量 ≤ 5.46 时则可归为高钾类，否则归为铅钡类，该分类规律同时在问题四进行敏感性分析时得到再次验证。对于亚类划分，可对已根据类别划分的两类样本数据分别通过层次聚类算法可获取冰状图，从冰状图中可观测到适合进行聚类的聚类数目，得到适宜的聚类数后采用 K-means 算法到聚类中心和聚类结果，划分方法就是寻找距离样本最近的聚类中心，并将其归入该类。最后进行合理性和敏感性分析，对于合理性，数形结合观察出不同的类与类之间的差异性较大而同一类间样本的相似性较高，认为是有效且合理的划分；对于敏感性，通过蒙特卡洛方法对每类玻璃的各种化学成分含量随机增或减一个原始数据-5%~5% 大小的值模拟由于各种不确定因素导致成分检测结果不准确的干扰，检验其分类结果的稳健性，重复进行 100 次敏感性实验，计算所得铅钡玻璃和高钾玻璃的平均分类准确率约在 99.8% 和 89.4% 左右，表明分类结果较稳健可靠，敏感性较低。

针对问题三，通过随机森林分类模型学习其依据化学成分的分类规律，通过十折交叉检验确保分类结果较为准确后，用其来对未知类别玻璃文物的化学成分进行分析并鉴别其类别。对于敏感性检验同问题二的方式采用蒙特卡洛方法机对数据造成上下 5% 波动的干扰，发现其划分完全相同，分类结果具有较好的抗干扰和稳健性。

针对问题四，对每个类别的玻璃化学成分之间绘制散点图，在直观观察其是否存在相关性之后，计算斯皮尔曼相关系数并进行显著性检验，分析其是否显著相关及正负相关并将两个类别进行对比，绘制热力图直观观测两个类别相同的化学成分之间相关系数的大小关系。

关键词：随机森林模型 层次聚类算法 K-means 算法 决策树 蒙特卡洛方法

一、问题重述

2013年，我国首倡“一带一路”伟大构想，这不仅显著提升了我国的国际地位，还促进了沿线各国友好交流。毫无疑问“一带一路”建设将在后疫情时代在提振国民经济上发挥重要作用。而早在两千年前，丝绸之路就已经成为了古代东西方交流的重要纽带，玻璃就是其重要历史见证。但是古代玻璃受埋藏环境等因素的制约而极易风化，因而不利于对其类别的正确判断。该问题也成为考古工作者开展研究的重大阻力。

根据题目和附件所提供的相关数据，建立数学模型解决下列4个问题：

1.对玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系进行分析；结合玻璃类型，分析文物样品表面是否存在风化化学成分含量的统计规律，并依据风化点检测数据，预测玻璃文物风化前的化学成分含量。

2.分析两种不同类型玻璃的分类规律；根据化学成分对每个类别进行亚类划分，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.分析附件表单3中未知类别玻璃文物的化学成分，甄别其所属类型，并对分类结果的敏感性进行分析。

4.分析不同类别玻璃文物样品的化学成分之间的关联关系，并比较出这种关联关系的差异性。

二、问题分析

2.1 问题一的分析

问题一要在相关已述条件下分析玻璃文物的玻璃类型、纹饰和颜色与表面风化的关系，得出文物样品表面有无风化化学成分含量的统计规律，并预测其风化前的化学成分含量。本文认为解题思路如下：

1.对于玻璃的表面风化与相关变量之间的分析，将颜色空缺的样本剔除后，本文采用卡方检验的方式，观察在特定显著性水平下检验玻璃的纹饰、颜色和类型与表面风化之间是否具有显著差异，并分析其所产生的影响。

2.对于文物样品表面有无风化化学成分含量的统计规律的探究，在对所给数据样本进行预处理后，将高钾和铅钡两种玻璃类型按风化和未风化的特征划分为四类，对每种类别进行描述统计后绘制出图表，将每个类别风化前后的化学含量进行对比，总结风化前后化学成分含量的特征及变化。

3.将所给数据按高钾和铅钡两种类型划分后再进行数据处理，计算出样本风化和无风化的化学成分含量的均值，以风化前后均值差值来作为任一样本受风化后化学成分的变化量的近似估计，风化后的化学成分加上这个差值就可估计风化前的化学含量

2.2 问题二的分析

1.为找出高钾和铅钡玻璃的分类规律，本文从其化学成分含量及纹饰等特征入手，找出能较好解释其类别的变量，因此考虑使用决策树来找出最能代表其特征的一个或几个变量。在样本量不多的前提下，采用十折交叉检验法进行验证。若验证结果正确率高，可认为决策树分类结果较为准确，从而可采用决策树对两者的分类标准作为参考分类规律。

2.对于亚类划分的问题，本文认为可采用聚类算法对该问题进行解答。本文首先通过层次聚类模型获得冰状图，以此得到每种类型玻璃可划分的聚类数，再通过 K-means++聚类算法，求出具体分类方法和分类结果。

3.对于合理性，观察不同的类与类之间的差异大小与同一类间样本的相似性高低，若组间差异大，同组样本相似度高，则为有效且合理的划分；对于敏感性，通过蒙特卡洛方法对每类玻璃的各种化学成分含量随机增或减一个原始数据-5%~5%大小的值模拟由于各种不确定因素导致成分检测结果不准确的干扰，检验其分类结果的稳健性，重复进行 100 次敏感性实验，若分类平均准确率较高，表明分类结果较稳健，敏感度较低。

2. 3 问题三的分析

1.通过随机森林分类模型学习其依据化学成分的分类规律，通过十折交叉检验确保分类结果较为准确后，用其来对未知类别玻璃文物的化学成分进行分析并鉴别其类别。

2.对于分类结果敏感性分析，可以同样采取问题二的蒙特卡洛方法随机对数据造成上下 5%波动的干扰，观察其划分是否完全相同，若是则分类结果具有较好的抗干扰和稳健性。

2. 4 问题四的分析

对不同类别的文物的样品要分析其化学成分之间的关联性，可以先画出散点图进行直观观察，再对每个类别的玻璃样品化学成分进行斯皮尔曼相关分析，并进行显著性检验，得到各成分之间是否有显著相关性，再观察哪些成分是显著正相关的，哪些是显著负相关的。

对于不同类别之间化学成分关联关系的差异，可从同两种成分之间一方有相关一方无相关，一方正相关一方负相关及相关系数大小来分析，而相关系数大小可通过热力图直观测得。

三、模型假设

- 1.假设检测仪器足够精确，未检测到的成分即视为样品中没有该成分；
- 2.假设古代玻璃的风化不考虑时期等因素的影响，而仅考虑埋藏环境因素的影响，其成分比例的变化也仅与内部元素和外部元素的交换有关；
- 3.假设颜色空缺的样本不影响本文进行的相关数据分析，将其从样本中除去。

四、问题一的建立与求解

4.1 卡方检验影响表面风化的变量

根据附件所给出的纹饰、类型和颜色与表面风化的相关数据，结合题目信息，对于玻璃的表面风化与相关变量之间的分析，将颜色空缺的样本剔除后，本文采用卡方检验的方式，观察在特定显著性水平下检验玻璃的纹饰、颜色和类型与表面风化之间是否具有显著差异，并分析其所产生的影响。

表 4-1 卡方检验

名称	表面风化		总计	χ^2	校正 χ^2	P
	无风化	风化				
颜色	蓝绿	6	9	15	6.287	6.287
	浅蓝	8	12	20		
	紫	2	2	4		
	深绿	3	4	7		
	深蓝	2	0	2		
	浅绿	2	1	3		
	黑	0	2	2		
	绿	1	0	1		
	合计	24	30	54		
类型	高钾	12	6	18	5.400	4.134
	铅钡	12	24	36		
	合计	24	30	54		
纹饰	C	13	15	28	5.747	5.747
	A	11	9	20		
	B	0	6	6		
	合计	24	30	54		

注：***、**、*分别代表 1%、5%、10% 的显著性水平

经上表可知：

卡方检验分析的结果显示，在 95% 的置信水平下：

基于表面风化和颜色，显著性 P 值为 0.507，水平上不呈现显著性，接受原假设，因此对于表面风化和颜色数据不存在显著性差异。

基于表面风化和类型，显著性 P 值为 0.020**，水平上呈现显著性，拒绝原假设，因此对于表面风化和类型数据存在显著性差异。显然铅钡玻璃比高钾玻璃更易风化。

基于表面风化和纹饰，显著性 P 值为 0.056*，水平上不呈现显著性，接受原假设，因此对于表面风化和纹饰数据不存在显著性差异。

4.2 风化前后的统计规律

先对所给数据样本进行预处理，包括：

1. 将表单一二合并，缺失值填补为 0
2. 增添采样点是否风化列，对于无特殊说明的数据采样点是否风化默认与表单一的表面风化值相同（如表面无风化的样本 30 采样点 1 和 2），对于特殊注明的（如风化采样点上的无风化点）采样点是否风化与标注相同（无风化）
3. 剔除异常值（成分累加不在 85%~105% 之间的数据，包括文物 15 和 17）

由于是否风化会对化学含量产生较大影响，为确保结果的准确性，应按照采样点是

否风化来进行分类和分析(如表面风化玻璃上无风化采样点的取样值应列入无风化玻璃中分析),因此后面的分析均以采样点风化而非表面风化来分析。

注: 经过预处理后, 将样本数据按高钾和铅钡两种玻璃类型进行划分, 再分别按风化和未风化的特征总划分为四类, 对每种类别进行描述统计后绘制出图表, 并对每个类别风化前后的化学含量进行比对, 初步观察出每个类别的玻璃风化前后内部成分都具有较为显著的差异性。

表 4-2 高钾风化

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3
平均数	93.96	0.00	0.54	0.87	0.20	1.93	0.27
标准误	0.71	0.00	0.18	0.20	0.13	0.39	0.03
中位数	93.51	0.00	0.67	0.83	0.00	1.72	0.28
众数	/	0.00	0.00	/	0.00	/	/
标准差	1.73	0.00	0.45	0.49	0.31	0.96	0.07
方差	3.01	0.00	0.20	0.24	0.09	0.93	0.00
最小值	92.35	0.00	0.00	0.21	0.00	0.81	0.17
最大值	96.77	0.00	1.01	1.66	0.64	3.50	0.35
观测数	6.00	6.00	6.00	6.00	6.00	6.00	6.00

续表:

	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
平均数	1.56	0.00	0.00	0.28	0.00	0.00	0.00
标准误	0.38	0.00	0.00	0.09	0.00	0.00	0.00
中位数	1.55	0.00	0.00	0.28	0.00	0.00	0.00
众数	/	0.00	0.00	/	0.00	0.00	0.00
标准差	0.93	0.00	0.00	0.21	0.00	0.00	0.00
方差	0.87	0.00	0.00	0.04	0.00	0.00	0.00
最小值	0.55	0.00	0.00	0.00	0.00	0.00	0.00
最大值	3.24	0.00	0.00	0.61	0.00	0.00	0.00
观测数	6.00	6.00	6.00	6.00	6.00	6.00	6.00

注: 高钾未风化、铅钡风化、铅钡未风化的描述统计图体现在附录 1.

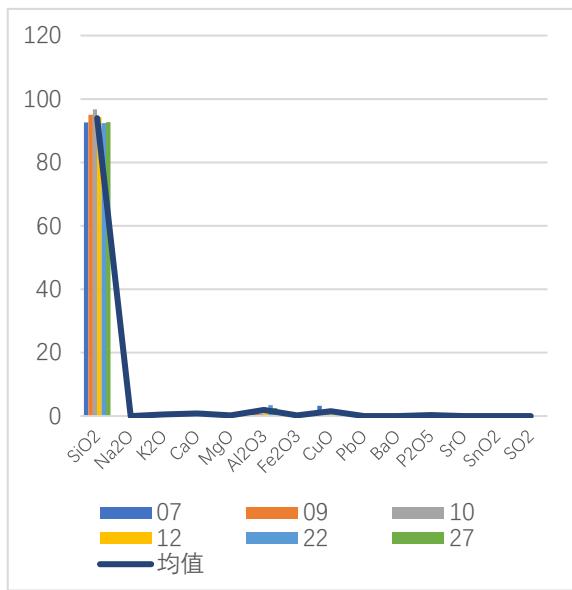


图 4-1 风化高钾玻璃化学含量示意图

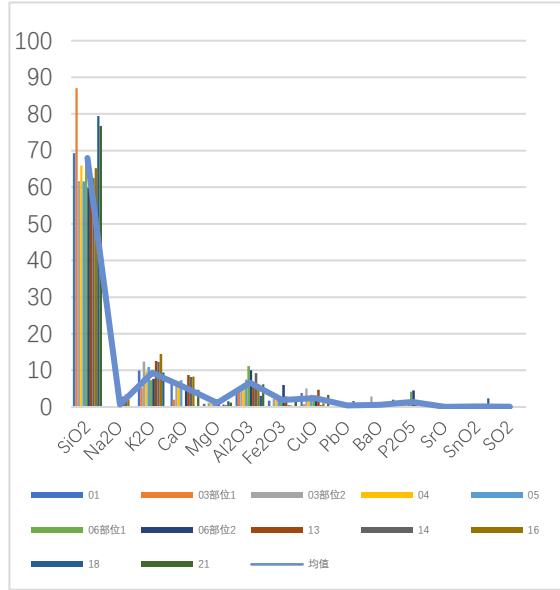


图 4-2 无风化高钾玻璃化学含量示意图

由图 5-1 和 5-2,可以明显观察出有风化和无风化的高钾玻璃都具有较高的 SiO₂ 含量，但风化前高钾玻璃还含有较高的 K₂O、CaO、Al₂O₃ 等化学成分，在风化后上述化学成分明显减少。

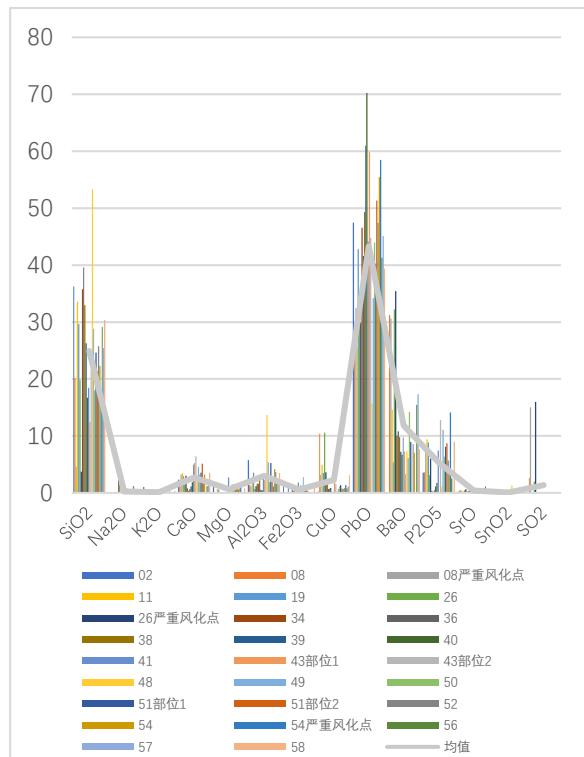


图 4-3 风化铅钡玻璃化学含量示意图

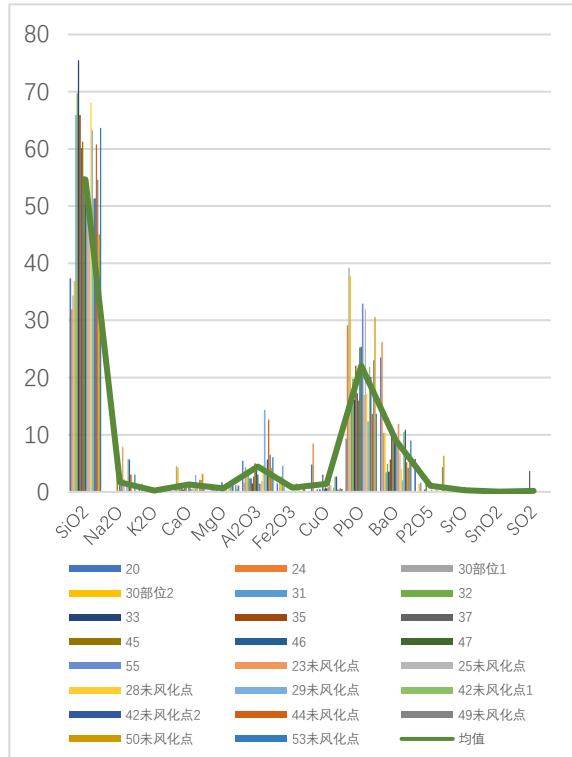


图 4-4 无风化铅钡玻璃化学含量示意图

由图 5-1 和 5-2,可以观察出有风化和无风化铅钡玻璃都具有较高的 SiO₂、PbO 和

BaO 含量，风化前铅钡玻璃的 SiO₂ 化学成分含量明显高于风化后上述化学成分含量，而风化后 PbO、BaO 化学成分明显增加。

4.3 玻璃样品风化前的的化学成分含量预测

计算出样本风化和无风化的化学成分含量的均值，以风化前后均值差值来作为任一样本受风化后化学成分的变化量的近似估计，风化后的化学成分加上这个差值就可估计风化前的化学含量。

数据处理：由于要求均值，为确保相对准确性，

- 1.以具体采样点的是否风化作为求均值的原始数据划分；
- 2.由于部分相同的文物（如 54、26、8）采样点以风化和严重风化数据成对出现，为避免极端值影响，删去严重风化数据；
- 3.对于部分风化文物（如 43）有两个采样部位都为风化的数据用其均值作为该样本风化后的值；
- 4.对于部分风化文物（如 49、50）同时有风化采样点和无风化采样点的数据直接以无风化采样点数据作为其风花前化学含量的估计，若有两个无风化采样点（如 42）数据取均值作为风花前的化学含量。

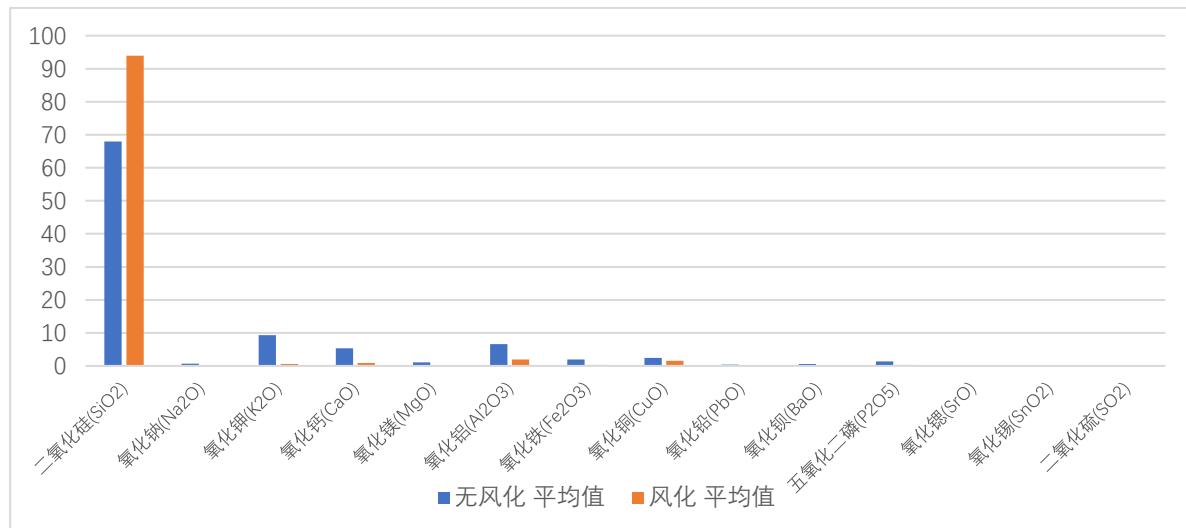


图 4-5 高钾玻璃风化前后平均化学含量对比图

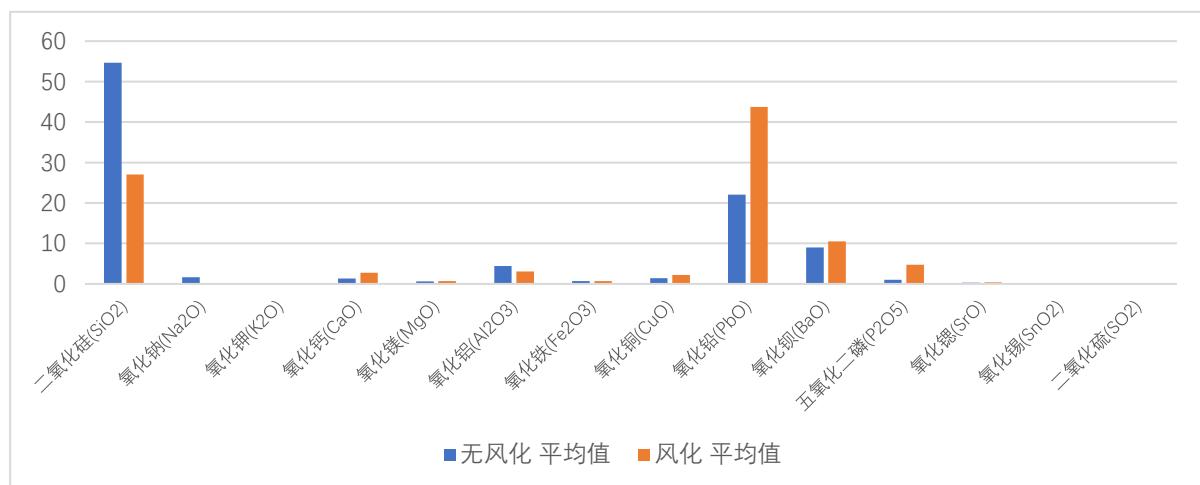


图 4-6 铅钡玻璃风化前后平均化学含量对比图

表 4-3 各化学成分风化前后含量差值

	二氧化 硅(SiO ₂)	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)
高钾	-25.98	0.70	8.79	4.46	0.88	4.69	1.67
铅钡	27.60	1.44	0.09	-1.46	-0.05	1.36	0.08
续表:							
	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化 二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化 硫(SO ₂)
高钾	0.89	0.41	0.60	1.12	0.04	0.20	0.10
铅钡	-0.79	-21.63	-1.47	-3.71	-0.11	-0.03	-0.04

注：差值表示高钾和铅钡两种玻璃类型平均风化前后的差值

各化学成分具体结果详见附录 2。

五、问题二的建立与求解

5.1 基于决策树模型的分类规律探究

对于高钾玻璃和铅钡玻璃的分类规律的探究，针对题目所给样本数据为小数据集的特点，本文主张采用决策树算法进行解决。首先要训练集数据设置出决策树分类模型，得到决策树结构；之后通过建立的决策树来计算特征重要性；再次将建立的决策树分类模型应用到训练、测试数据，得到模型的分类评估结果。

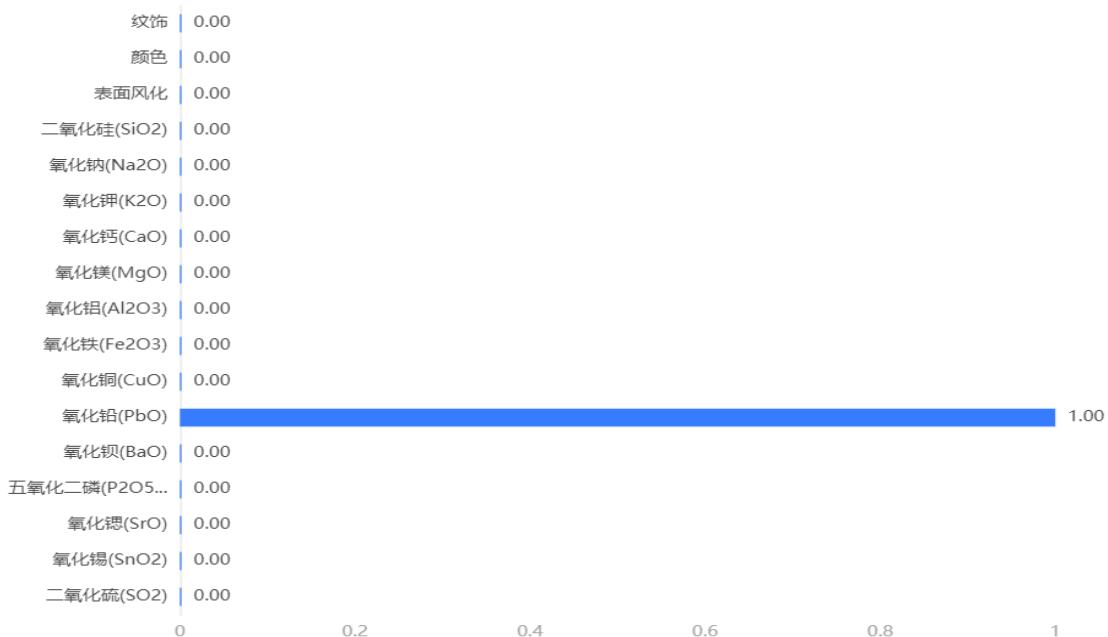


图 5-1 决策树特征重要性示意图

该示意图展示出各变量之间的重要程度的比例。其中 PbO 的特征重要性为 1，表明其可以完全解释玻璃类别的划分，可作为划分依据。

表 5-1 模型评估结果

	准确率	召回率	精确率	F1
训练集	1	1	1	1
交叉验证集	1	1	1	1
测试集	1	1	1	1

上示模型评估结果表中表示出训练集和测试集的分类评价指标，通过量化指标来评定决策树对训练、测试数据的分类效果。

- 准确率：所有预测样本/总样本；
- 召回率：将正类预测为正类/预测的正类；
- 精确率：将正类预测为正类/原本的正类；
- F1：精确率和召回率的调和平均值。

上述结果全部为 1 表明决策树分类的验证结果正确率高，可认为决策树分类结果较为准确，从而可采用决策树对两者的分类标准作为参考分类规律。

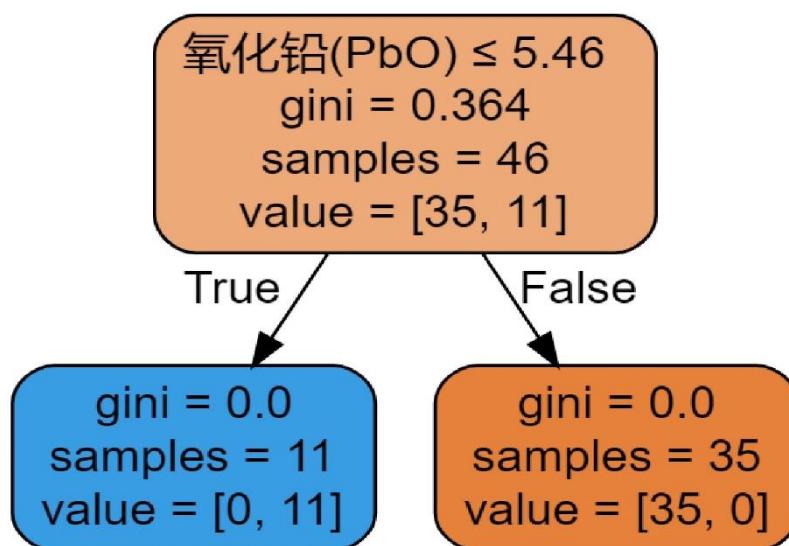


图 5-2 决策树结构分类图

由上图所得到的规律即：当 $\text{PbO} \leq 5.46$ 时，将玻璃类型归为高钾类；当 $\text{PbO} > 5.46$ 时，将玻璃归为铅钡类。

5.2 基于聚类算法对玻璃进行亚分类

聚类算法是一种区别于分类的无监督学习算法，即面对无标签的样本点，根据样本数据及数据间的信息关系，对数据对象进行分组。而 K-means++ 算法是基于划分的聚类算法，采用距离作为评价指标，即认为两个对象的距离越近，它们就越相似^[3]。

将附件数据根据类型分为高钾和铅钡两大类后，分别对每一类别的玻璃样本的数据进行亚分类。对每一类别的玻璃样本的数据的具体处理步骤如下：

1. 构建分层聚类模型，得到冰状图，从图中可观测处可聚类数目；
2. 选取可聚类数目 k，通过 K-means 算法进行聚类，得到聚类中心和聚类结果；
3. 通过聚类中心和聚类结果给出划分方法和划分结果；
4. 对分类结果合理性和敏感性进行分析。

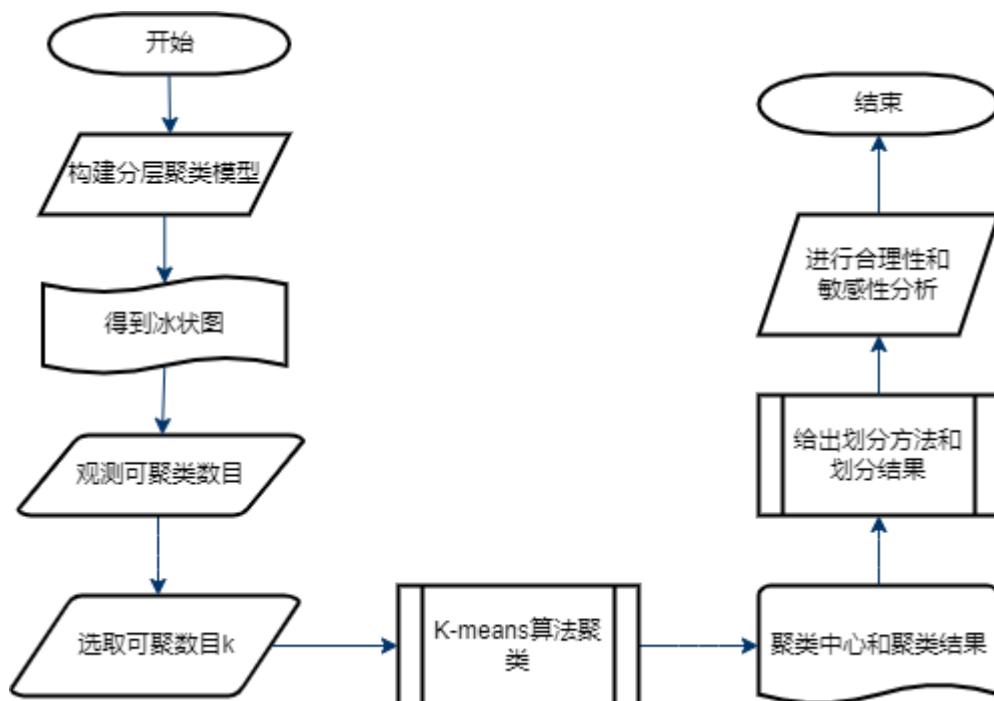


图 5-3 分析步骤思维导图

下文将分别分析研究铅钡玻璃与高钾玻璃的亚类划分的问题。

5.2.1 铅钡玻璃亚类划分问题的研究

构建分层聚类模型，得到冰状图：

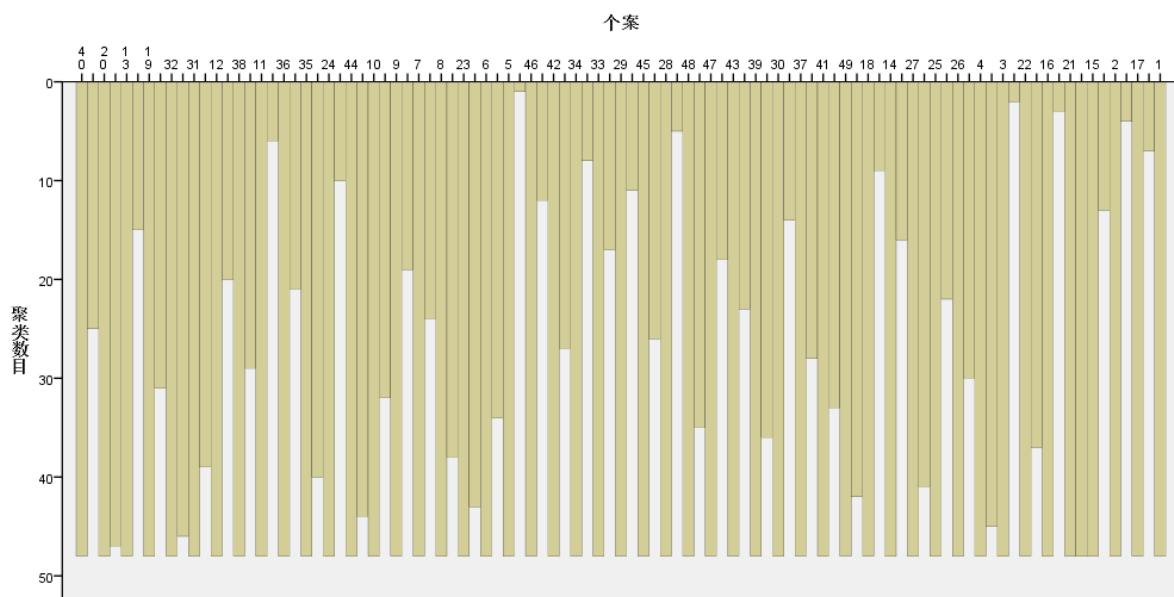


图 5-4 铅钡类玻璃样本冰状图

由上示冰状图可知适宜选取的聚类数目，我们选取的聚类数目为 4。

再利用 SPSS 软件运用 K-means++算法进行聚类，得到聚类中心和分类结果。聚类中心如下：

表 5-2 铅钡玻璃样本聚类中心

	聚类			
	1	2	3	4
二氧化硅(SiO ₂)	39.82	24.89	12.07	60.31
氧化钠(Na ₂ O)	.80	.14	.00	2.03
氧化钾(K ₂ O)	.27	.13	.10	.19
氧化钙(CaO)	2.00	2.86	2.28	1.12
氧化镁(MgO)	.30	.77	.00	.84
氧化铝(Al ₂ O ₃)	3.01	2.89	1.08	5.49
氧化铁(Fe ₂ O ₃)	.92	.73	.00	.59
氧化铜(CuO)	2.39	1.36	6.93	1.00
氧化铅(PbO)	30.87	48.57	30.15	18.40
氧化钡(BaO)	12.96	7.62	32.39	7.24
五氧化二磷(P ₂ O ₅)	2.63	5.60	5.08	.65
氧化锶(SrO)	.31	.42	.49	.25
氧化锡(SnO ₂)	.09	.02	.00	.09
二氧化硫(SO ₂)	.00	.00	8.88	.22

铅钡玻璃亚分类的依据为：将高钾玻璃样本的每个样本和聚类中心都视为对应包含 14 种化学成分的 14 维向量，对每个样本计算出其到四个聚类中心的距离，找出其中距离最小对应的聚类中心，则该样本划分入该类。

对于铅钡玻璃中的 18 个样本做如上划分，所得分类如下：

表 5-3 铅钡玻璃样本聚类成员

个案号	文物采样点	聚类	距离
1	20	1	24.702
2	24	1	17.041
3	30 部位 1	1	11.466
...
48	57	2	11.905
49	58	2	11.438

下面分析该亚类划分的合理性和敏感性：

1. 铅钡玻璃合理性分析：

合理的划分应使类与类之间的差异较大并使同一类内的样品相似度高。

由本文的划分方式可知，若不同聚类中心之间的距离较大而属于同一类的样本距离较小则该划分合理且有效。

计算得到不同聚类中心间的距离矩阵如下：

表 5-4 铅钡聚类中心之间的距离

聚类	1	2	3	4
1	0	24.003	35.482	24.957
2	24.003	0	35.110	46.934
3	35.482	35.110	0	57.042
4	24.957	46.934	57.042	0

将聚类成员到聚类中心的距离绘制示意图如下：

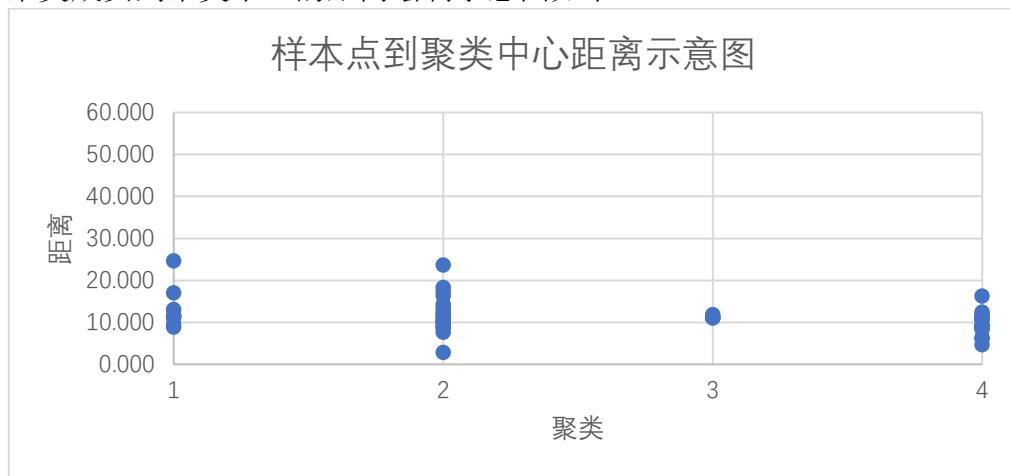


图 5-5 样本点到聚类中心距离示意图

不同聚类中心之间的距离显然较大而属于同一类的样本距离较小，因此可说明该划分是有效且合理的。

2. 铅钡玻璃亚类划分敏感性分析：

敏感性分析 (Sensitivity Analysis)，就是假设模型表示为 $y=F(x^1, x^2, \dots, x^n)$ (x^i 表示模型的第 i 个属性值)，控制其他属性不变的情况下，使某个属性在一定取值范围内变动，研究和预测这些属性的变动对模型输出值的影响程度^[1]。其含义和通用步骤可以简单理解为控制变量法，与经济学和管理学意义中的杠杆概念类似。

由于玻璃各种不确定因素会导致成分检测结果不准确的干扰，一个合理的分类应具有敏感性小，稳定性高的特性，使得对于小的误差不会对分类结果造成太大影响。

我们通过蒙特卡洛方法对每类玻璃的各种化学成分含量随机增或减一个相对原始数据-5%~5%大小的值，模拟各种因素对检测化学成分结果不准确的干扰，检验其分类

结果的稳健性，重复进行 100 次敏感性实验，计算平均分类准确率，若平均分类准确率较高，表明分类结果敏感性较低，也较稳健。

计算所得铅钡玻璃平均分类准确率约在 99.8% 左右，表明分类结果较稳健。

5.2.2 高钾玻璃亚类划分问题的研究

构建分层聚类模型，得到冰状图：

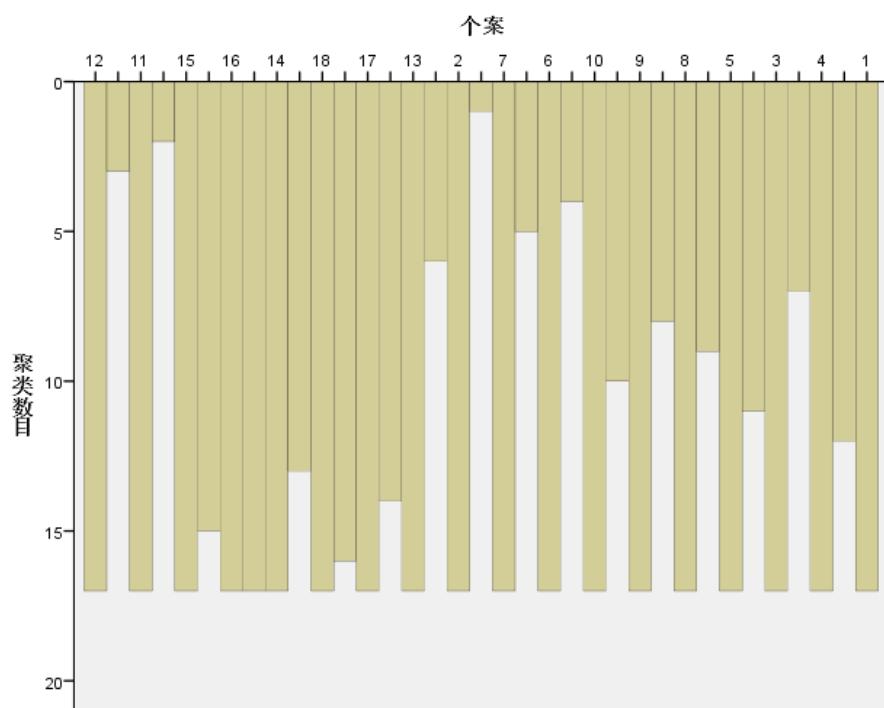


图 5-6 高钾类玻璃样本冰状图

由上示冰状图可知可聚类数目，我们选取的聚类数目为 4。

再利用 SPSS 软件运用 K-means++ 算法进行聚类，得到聚类中心和分类结果。聚类中心如下：

表 5-5 高钾玻璃样本聚类中心

	聚类			
	1	2	3	4
二氧化硅(SiO_2)	78.07	92.98	62.23	68.49
氧化钠(Na_2O)	.00	.00	1.19	.00
氧化钾(K_2O)	4.71	1.21	11.43	8.68
氧化钙(CaO)	2.36	1.03	7.28	3.16
氧化镁(MgO)	1.38	.17	1.05	1.43
氧化铝(Al_2O_3)	4.62	2.23	7.29	7.54

氧化铁(Fe ₂ O ₃)	1. 19	. 23	2. 38	2. 07
氧化铜(CuO)	1. 64	1. 45	2. 71	3. 19
氧化铅(PbO)	. 50	. 04	. 50	. 10
氧化钡(BaO)	. 99	. 00	. 55	. 69
五氧化二磷(P ₂ O ₅)	1. 23	. 33	1. 19	2. 68
氧化锶(SrO)	. 04	. 00	. 05	. 06
氧化锡(SnO ₂)	1. 18	. 00	. 00	. 00
二氧化硫(SO ₂)	. 00	. 00	. 12	. 20

高钾玻璃亚分类的依据为：将高钾玻璃样本的每个样本和聚类中心都视为对应包含14种化学成分的14维向量，对每个样本计算出其到四个聚类中心的距离，找出其中距离最小对应的聚类中心，则该样本划分入该类。

对于高钾玻璃中的18个样本做如上划分，所得分类如下：

表 5-6 高钾玻璃样本聚类成员

聚类成员			
个案号	文物采样点	聚类	距离
1	01	4	5. 395
2	03 部位 1	2	7. 480
3	03 部位 2	3	4. 464
4	04	3	4. 470
5	05	3	1. 935
6	06 部位 1	4	5. 395
7	06 部位 2	3	7. 596
8	13	3	4. 867
9	14	3	4. 647
10	16	3	5. 478
11	18	1	6. 235
12	21	1	6. 235
13	07	2	2. 226
14	09	2	2. 370
15	10	2	4. 207
16	12	2	1. 604
17	22	2	1. 916
18	27	2	1. 325

下面分析该亚类划分的合理性和敏感性：

1.高钾玻璃合理性分析

由本文的划分方式可知，若不同聚类中心之间的距离较大而属于同一类的样本距离较小则该划分合理且有效。

计算得到不同聚类中心间的距离矩阵如下：

表 5-7 高钾聚类中心之间的距离

聚类	1	2	3	4
1	0	15.738	18.249	11.120
2	15.738	0	33.524	26.496
3	18.249	33.524	0	8.245
4	11.120	26.496	8.245	0

将聚类成员到聚类中心的距离绘制示意图如下：

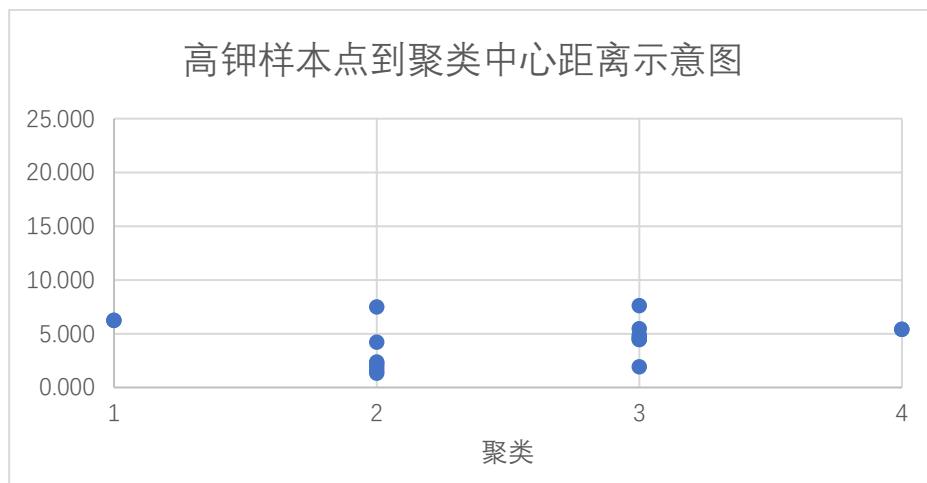


图 5-7 高钾样本点到聚类中心距离示意图

不同聚类中心之间的距离显然较大而属于同一类的样本距离较小，因此可说明该划分是有效且合理的。

2.高钾玻璃的敏感性分析

在计算高钾玻璃的敏感性分析时，我们同样通过蒙特卡洛方法对每类玻璃的各种化学成分含量进行计算，计算所得高钾玻璃平均分类准确率约在 89.4% 左右，表明分类结果较为稳健，敏感度较低，尚可接受。

六、问题三的建立与求解

6.1 未知玻璃文物类型分类及结果敏感性分析

随机森林模型（Random Forests, RF）是以决策树为基分类器，通过 Bagging 集成方法与随机子空间方法的结合而形成的。随机森林模型作为集成算法的典型代表之一，拥有预测准确度高、对数据中的异常值与噪声有较好的容忍度、且不易出现过拟合问题等诸多优点^[2]。

我们通过预处理过的数据来训练随机森林分类模型学习玻璃数据依据化学成分对类别进行分类的规律，并以十折交叉验证法以提高模型分类的可靠性。

表 6-1 随机森林分类模型评估结果

	准确率	召回率	精确率	F1
训练集	1	1	1	1
交叉验证集	1	1	1	1
测试集	1	1	1	1

评估结果显示经过训练的随机森林模型分类结果较可靠，反馈的预测结果如下：

表 6-2 未知类别的文物的预测

文物编号	预测结果	预测结果概率_铅钡	预测结果概率_高钾
A1	高钾	0.09	0.91
A1_u1	高钾	0.09	0.91
A1_u2	高钾	0.09	0.91
A1_u3	高钾	0.09	0.91
A2	铅钡	0.85	0.15
A2_u1	铅钡	0.85	0.15
A2_u2	铅钡	0.85	0.15
A2_u3	铅钡	0.85	0.15
A3	铅钡	0.88	0.12
...
A7	高钾	0.06	0.94
A7_u1	高钾	0.06	0.94
A7_u2	高钾	0.06	0.94
A7_u3	高钾	0.06	0.94
A8	铅钡	1	0
A8_u1	铅钡	1	0
A8_u2	铅钡	1	0
A8_u3	铅钡	1	0

详细预测见附录 5

上表中的 A_i_{ui} 同样是我们采取蒙特卡洛方法对每类玻璃的各种化学成分含量随机增或减一个相对原始数据-5%~5%大小的值所构造出来的带干扰的样本（具体数据查看支撑材料），分类结果全部一致，我们认为分类结果具有相当的稳健性，敏感性较低。

同时对上表文物的氧化铅含量和预测结果进行联系，可与问题二中得到的玻璃类型划分规律相互验证。

七、问题四的建立与求解

对不同类别的文物的样品要分析其化学成分之间的关联性，可以先画出每个类别化学成分之间的散点图进行直观观察，再对每个类别的玻璃样品化学成分进行斯皮尔曼相关分析，并进行显著性检验，得到各成分之间是否有显著相关性

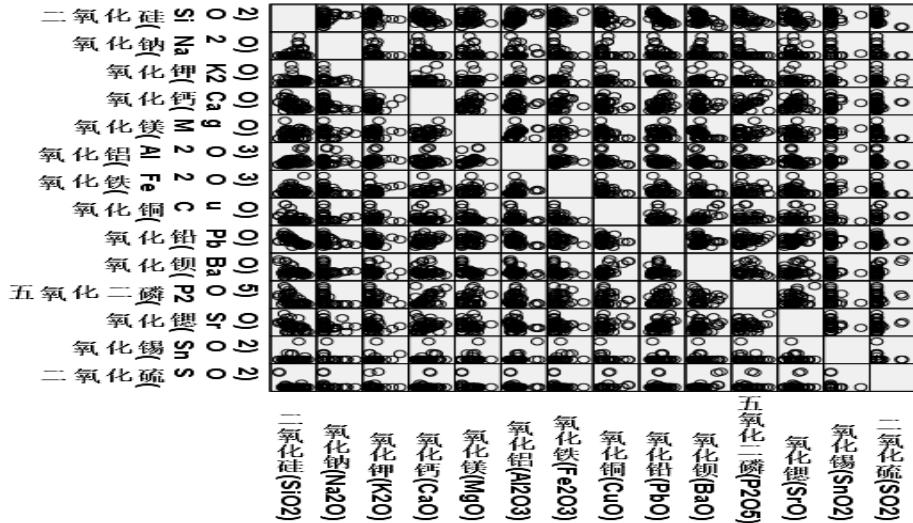


图 7-1 铅钡化学成分散点图

	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)	
二氧化硅(SiO2)	1.000(0.000**)	0.362(0.011*)	0.087(0.553)	0.488(0.000*)	0.088(0.549)	0.401(0.004**)	0.082(0.574)	0.354(0.012*)	0.738(0.000*)	0.436(0.002*)	0.565(0.000*)	0.503(0.000*)	0.085(0.560)	0.386(0.006*)	
氧化钠(Na2O)	0.362(0.011*)	1.000(0.000*)	-	0.059(0.689)	0.373(0.008**)	0.026(0.861)	0.103(0.483)	0.244(0.091*)	-	0.058(0.692)	0.349(0.014*)	-	0.056(0.700)	0.396(0.005*)	
氧化钾(K2O)	0.087(0.553)	-	1.000(0.000*)	0.096(0.510)	0.269(0.061*)	0.306(0.033*)	0.248(0.086*)	-	0.146(0.316)	0.131(0.370)	0.041(0.779)	0.106(0.470)	0.121(0.407)	0.195(0.180)	
氧化钙(CaO)	0.488(0.000*)	0.373(0.008*)	0.096(0.510)	1.000(0.000*)	0.417(0.003**)	0.135(0.354)	0.387(0.006**)	-	0.068(0.642)	0.356(0.012*)	0.124(0.395)	0.535(0.000*)	0.171(0.240)	0.210(0.148)	
氧化镁(MgO)	0.088(0.549)	0.026(0.861)	0.269(0.061*)	0.417(0.003**)	1.000(0.000*)	0.450(0.001**)	0.293(0.041*)	-	0.309(0.031*)	0.036(0.805)	0.453(0.001**)	0.272(0.058*)	0.087(0.554)	0.246(0.089*)	
氧化铝(Al2O3)	0.401(0.004**)	0.103(0.483)	0.306(0.033*)	0.135(0.354)	0.450(0.001**)	1.000(0.000*)	0.230(0.112)	-	0.282(0.050*)	0.441(0.002*)	0.339(0.017*)	-	0.074(0.612)	0.155(0.287)	
氧化铁(Fe2O3)	0.082(0.574)	0.244(0.091*)	0.248(0.086*)	0.387(0.006**)	0.293(0.041*)	0.230(0.112)	1.000(0.000*)	-	0.247(0.087*)	0.048(0.745)	0.295(0.040*)	0.153(0.295)	0.091(0.535)	0.226(0.119)	
氧化铜(CuO)	-	-	-	-	0.309(0.031*)	0.282(0.050*)	0.247(0.087*)	1.000(0.000*)	-	0.085(0.562)	0.719(0.000*)	0.096(0.510)	0.196(0.177)	-	
氧化铅(PbO)	0.738(0.000**)	0.349(0.014*)	0.131(0.370)	0.356(0.012*)	-	0.036(0.805)	0.441(0.002*)	0.048(0.745)	0.085(0.562)	-	1.000(0.000*)	0.140(0.337)	0.331(0.020*)	0.434(0.002*)	
氧化钡(BaO)	0.436(0.002*)	0.056(0.700)	0.041(0.779)	0.124(0.395)	0.453(0.001*)	0.339(0.017*)	0.295(0.040*)	0.719(0.000*)	-	0.140(0.337)	1.000(0.000*)	-	0.164(0.262)	0.077(0.601)	
五氧化二磷(P2O5)	0.565(0.000*)	0.396(0.005**)	0.106(0.470)	0.535(0.000*)	0.272(0.058*)	0.074(0.612)	0.153(0.295)	0.096(0.510)	0.331(0.020*)	-	1.000(0.000*)	0.280(0.052*)	-	0.082(0.576)	
氧化锶(SrO)	0.503(0.000**)	-	0.093(0.526)	0.121(0.407)	0.171(0.240)	0.087(0.554)	0.155(0.287)	0.091(0.535)	0.196(0.177)	0.434(0.002*)	0.164(0.262)	0.280(0.052*)	1.000(0.000*)	0.038(0.796)	
氧化锡(SnO2)	0.085(0.560)	-	0.083(0.573)	0.195(0.180)	0.210(0.148)	0.246(0.089*)	0.469(0.001*)	0.226(0.119)	0.184(0.207)	0.126(0.387)	0.077(0.601)	0.082(0.576)	0.038(0.796)	-	
二氧化硫(SO2)	0.386(0.006**)	-	0.130(0.374)	0.010(0.944)	0.104(0.478)	0.266(0.064*)	0.206(0.155)	0.180(0.216)	-	0.241(0.096*)	-	0.634(0.000*)	0.173(0.235)	0.181(0.213)	1.000(0.000*)

注：***、**、*分别代表1%、5%、10%的显著性水平

图 7-2 铅钡斯皮尔曼相关系数表

详见支撑材料。

从散点图和皮埃尔相关系数表中可看出，标色单元格对应行列之间都有显著相关性（绿色、蓝色、红色分别表示90%、95%、99%的置信水平）（由于是对称矩阵，为明显起见只标注出了矩阵主对角线之上的部分）

1. 铅钡玻璃中呈现显著正相关的成分有(二氧化硅, 氧化钠), (二氧化硅, 氧化铝), (氧化钾, 氧化镁), (氧化钾, 氧化铝), (氧化钾, 氧化铁), (氧化钙, 氧化镁), (氧化镁, 氧化铝), (氧化镁, 氧化铁), (氧化镁, 五氧化二磷), (氧化镁, 氧化锡), (氧化铝, 氧化锡), 氧化铜, 氧化钡), (氧化铜, 二氧化硫), (氧化铅, 五氧化二磷), (氧化铅, 氧化锶), (氧化钡, 二氧化硫), (五氧化二磷, 氧化锶)。

呈现显著负相关的成分为表中标色但不在上述显著正相关的成分中列出的成分。

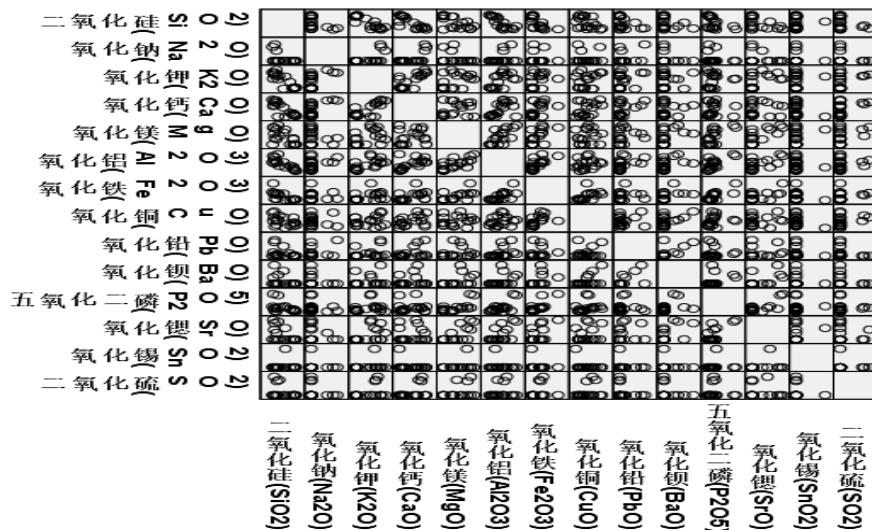


图 7-3 高钾化学成分散点图

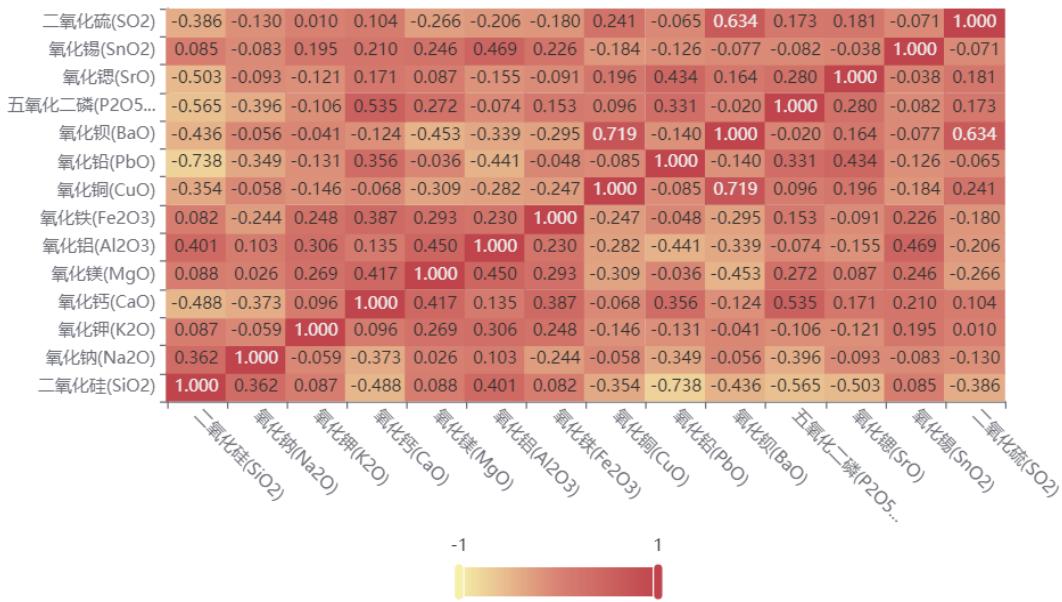
	二氧化硅(SiO ₂)	氯化钠(Na ₂ O)	氯化钾(K ₂ O)	氯化钙(CaO)	氯化镁(MgO)	氯化铝(Al ₂ O ₃)	氯化铁(Fe ₂ O ₃)	氯化铜(CuO)	氯化铅(PbO)	氯化钡(BaO)	五氧化二磷(P ₂ O ₅)	氯化锶(SrO)	氯化锡(SnO ₂)	二氧化硫(SO ₂)
二氧化硅(SiO ₂)	1.000(0.000*) **)	0.475(0.046*) **) 0.781(0.000*) **) 0.752(0.000*) **) 0.546(0.019*) **) 0.831(0.000*) **) 0.763(0.000*) **) 0.430(0.075*) **) 0.443(0.066*) **) 0.346(0.160) - 0.530(0.024*) **) 0.524(0.026*) **) 0.070(0.782) 0.284(0.253)	-	-	-	-	-	-	-	-	-	-	-	-
氯化钠(Na ₂ O)	-	1.000(0.000*) **) 0.606(0.008*) **) 0.640(0.004*) **) 0.223(0.373) 0.313(0.206) 0.218(0.385) 0.113(0.656) 0.266(0.287) 0.236(0.347) 0.221(0.378) 0.110(0.665) 0.108(0.670) 0.198(0.431)	-	-	-	-	-	-	-	-	-	-	-	-
氯化钾(K ₂ O)	0.781(0.000*) **) 0.606(0.008*) **) 1.000(0.000*) **) 0.690(0.002*) **) 0.256(0.305) 0.506(0.032*) **) 0.457(0.056*) **) 0.179(0.477) 0.241(0.335) 0.010(0.969) 0.156(0.536) 0.404(0.096*) **) 0.070(0.782) 0.307(0.215)	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化钙(CaO)	0.752(0.000*) **) 0.640(0.004*) **) 0.690(0.002*) **) 1.000(0.000*) **) 0.087(0.731) 0.521(0.026*) **) 0.537(0.022*) **) 0.344(0.162) 0.256(0.305) 0.073(0.773) 0.023(0.929) 0.052(0.839) 0.374(0.126) 0.389(0.110)	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化镁(MgO)	0.546(0.019*) 0.223(0.373) 0.256(0.305) 0.087(0.731) 1.000(0.000*) **) 0.735(0.001*) **) 0.569(0.014*) **) 0.204(0.417) 0.295(0.234) 0.521(0.027*) **) 0.643(0.004*) **) 0.668(0.003*) **) 0.214(0.393) 0.429(0.076*) -	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化铝(Al ₂ O ₃)	0.831(0.000*) **) 0.313(0.206) 0.506(0.032*) **) 0.521(0.026*) **) 0.735(0.001*) **) 1.000(0.000*) **) 0.748(0.000*) **) 0.246(0.326) 0.602(0.008*) **) 0.473(0.047*) **) 0.518(0.026*) **) 0.511(0.030*) **) 0.164(0.517) 0.246(0.325)	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化铁(Fe ₂ O ₃)	0.783(0.000*) **) 0.218(0.385) 0.457(0.056*) **) 0.537(0.022*) **) 0.569(0.014*) **) 0.748(0.000*) **) 1.000(0.000*) **) 0.643(0.004*) **) 0.323(0.191) 0.539(0.021*) **) 0.517(0.028*) **) 0.409(0.092*) **) 0.374(0.126) 0.313(0.206)	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化铜(CuO)	0.430(0.075*) 0.113(0.656) 0.179(0.477) 0.344(0.162) 0.204(0.417) 0.246(0.326) 0.643(0.004*) **) 1.000(0.000*) **) 0.047(0.853) 0.483(0.042*) **) 0.466(0.051*) **) 0.141(0.577) 0.398(0.102) 0.354(0.149)	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化铅(PbO)	0.443(0.066*) 0.266(0.287) 0.241(0.335) 0.256(0.305) 0.295(0.234) 0.602(0.008*) **) 0.323(0.191) 0.047(0.853) 1.000(0.000*) **) 0.682(0.002*) **) 0.121(0.633) 0.368(0.133) 0.186(0.460) 0.341(0.165)	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化钡(BaO)	0.346(0.160) 0.236(0.347) 0.010(0.969) 0.073(0.773) 0.521(0.027*) **) 0.473(0.047*) **) 0.539(0.021*) **) 0.483(0.042*) 0.682(0.002*) **) 1.000(0.000*) **) 0.471(0.049*) **) 0.565(0.015*) **) 0.128(0.612) 0.236(0.347)	-	-	-	-	-	-	-	-	-	-	-	-	-
五氧化二磷(P ₂ O ₅)	0.530(0.024*) 0.221(0.378) 0.156(0.536) 0.023(0.929) 0.643(0.004*) **) 0.518(0.028*) **) 0.517(0.028*) **) 0.466(0.051*) **) 0.121(0.633) 0.471(0.049*) **) 1.000(0.000*) **) 0.500(0.035*) **) 0.304(0.220) 0.306(0.216) 1.000(0.000*) **) 0.108(0.670)	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化锶(SrO)	0.524(0.026*) 0.110(0.665) 0.404(0.096*) **) 0.052(0.839) 0.666(0.003*) **) 0.511(0.030*) **) 0.409(0.092*) **) 0.141(0.577) 0.368(0.133) 0.565(0.015*) **) 0.500(0.035*) **) 1.000(0.000*) **) 0.306(0.216) 0.023(0.929) 0.108(0.670) 0.100(0.000*) **) 0.108(0.670)	-	-	-	-	-	-	-	-	-	-	-	-	-
氯化锡(SnO ₂)	0.070(0.782) 0.108(0.670) 0.070(0.782) 0.374(0.126) 0.214(0.393) 0.164(0.517) 0.374(0.126) 0.398(0.102) 0.186(0.460) 0.128(0.612) 0.304(0.220) 0.306(0.216) 1.000(0.000*) **) 0.108(0.670)	-	-	-	-	-	-	-	-	-	-	-	-	-
二氧化硫(SO ₂)	0.284(0.253) 0.198(0.431) 0.307(0.215) 0.389(0.110) 0.429(0.076*) **) 0.246(0.325) 0.313(0.206) 0.354(0.149) 0.341(0.165) 0.236(0.347) 0.246(0.324) 0.023(0.929) 0.108(0.670) 0.100(0.000*) **) 0.108(0.670)	-	-	-	-	-	-	-	-	-	-	-	-	-

注: **、* 分别代表1%、5%、10%的显著性水平
图 7-4 高钾斯皮尔曼相关系数表

2. 高钾玻璃中呈现显著正相关的成分有(氧化钠, 氧化钾), (氧化钠, 氧化钙), (氧化钾, 氧化钙), (氧化钾, 氧化铝), (氧化钾, 氧化铁), (氧化钾, 氧化锶), (氧化钙, 氧化铝), (氧化钙, 氧化铁), (氧化镁, 氧化铝), (氧化镁, 氧化铁), (氧化镁, 氧化锶), (氧化铝, 氧化锶), (氧化铜, 氧化钡), (氧化铜, 二氧化硫), (氧化铅, 五氧化二磷), (氧化铅, 氧化锶), (氧化钡, 二氧化硫), (五氧化二磷, 氧化锶)。

钡), (氧化镁, 五氧化二磷), (氧化镁, 氧化锶), (氧化镁, 二氧化硫), (氧化铝, 氧化铁), (氧化铝, 氧化铅), (氧化铝, 氧化钡), (氧化铝, 五氧化二磷), (氧化铝, 氧化锶), (氧化铁, 氧化铜), (氧化铁, 氧化钡), (氧化铁, 五氧化二磷), (氧化铁, 氧化锶), (氧化铜, 氧化钡), (氧化铜, 五氧化二磷), (氧化铅, 氧化钡), (氧化钡, 五氧化二磷), (氧化钡, 氧化锶), (五氧化二磷, 氧化锶)。

呈现显著负相关的成分有(二氧化硅, 氧化纳), (二氧化硅, 氧化钾), (二氧化硅, 氧化钙), (二氧化硅, 氧化镁), (二氧化硅, 氧化铝), (二氧化硅, 氧化铁), (二氧化硅, 氧化铜), (二氧化硅, 氧化铅), (二氧化硅, 五氧化二磷), (二氧化硅, 氧化锶)。



7-5 铅钡斯皮尔曼相关系数热力图

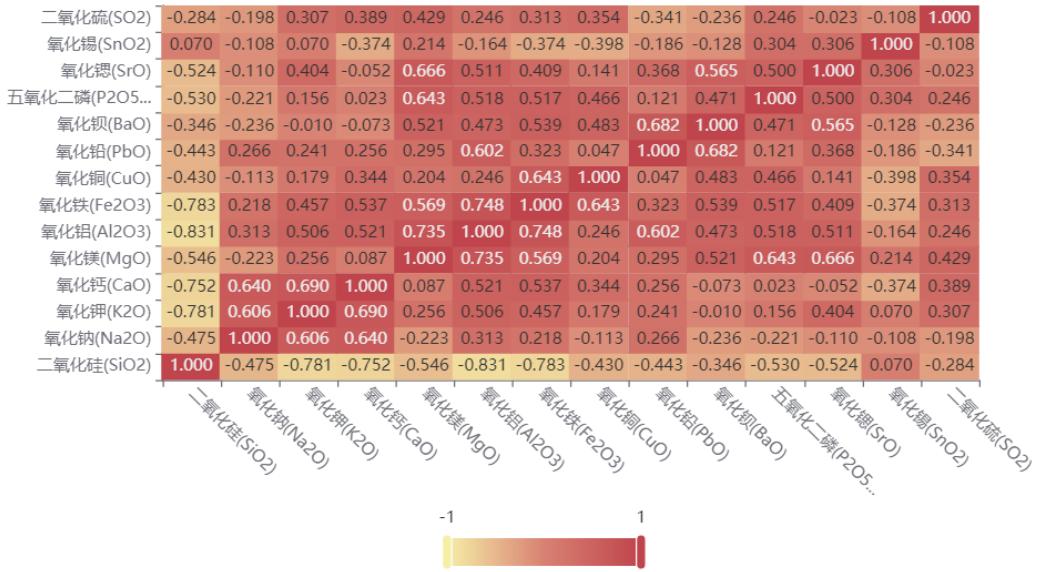


图 7-6 高钾斯皮尔曼相关系数热力图

铅钡玻璃和高价玻璃成分关联之间的差别由上述分析及两者相关系数热力图对比可得出:

1. 从相关系数表标色部分可看出，许多成分在二者之间一方显著相关而另一方并不显著相关，如二氧化硅和氧化钡。
2. 许多成分在二者之间一方显著正相关而另方显著负相关，如氧化镁和氧化钡。
3. 从热力图对比可看出，高钾玻璃大部分成分间的相关系数高于铅钡玻璃成分的相关系数。
4. 在显著相关的成分之间，高钾玻璃中二氧化硫和其他所有成分均呈现负相关关系，而铅钡玻璃中二氧化硅和氧化纳、氧化铝均呈现显著正相关关系。

八、模型的评价、改进与推广

8.1 模型的评价

8.1.1 模型的优点

1. 卡方检验的增添，在分析描述统计时增添了模型的合理性和可靠性
2. 问题一中风化前后图形的比对使得化学成分含量的比较上更加直观
3. 数形结合的工作思路和方法可以较好的指导实践工作
4. 采用了随机森林算法对未知类型玻璃进行判断，充分利用了模型优异的分类性能，充分发挥其对异常值和噪声数据不敏感的特点，保持了精确度；
5. 本文着重强调了敏感性分析在玻璃的亚类划分和未知玻璃的预测中的作用，在分析中去掉敏感性系数很小的属性，重点考虑敏感性系数较大的属性，大大降低模型的复杂度，减少数据分析处理的工作量，在很大程度上提高了模型的精度。

8.1.2 模型的缺点

1. 问题1中排除了无颜色标记样本的影响，增添了模型的理想性，从而可能会在一定程度上影响本文预测结果的准确性；

为了简化分析，在部分分析中没有了考虑严重风化样本的影响，异常值可能影响结果全面性。

8.2 模型的改进

时间条件允许，可采用如svm等程序编辑更复杂的模型进行数据处理，以此提升预测结果的精确度。

8.3 模型的推广

本文所构建的各类模型可以为众多成分含量和检测的问题进行分析，如食品成分含量的分析、某产品的配比等。通过对成分含量的全面了解，经过大量试验找出某事物的特征，便于生活中新事物的合理分类。

九、 参考文献

- [1] 蔡毅,邢岩,胡丹.敏感性分析综述[J].北京师范大学学报(自然科学版),2008,(01):9-16.
- [2] 武炜杰.随机森林算法的应用与优化方法研究[D].江南大学,2021.
- [3] 王森,刘琛,邢帅杰.K-means 聚类算法研究综述[J/OL].华东交通大学学报: 1-8[2022-09-15 10: 22].10220914.001.DOI: 10.16749/j.cnki.jecjtu.20220914.001.

十、附录

附录 1：问题一中玻璃的表面风化与其纹饰、颜色、玻璃类型的关系分析

表 5-2 高钾未风化

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3
平均数	67.98	0.70	9.33	5.33	1.08	6.62	1.93
标准误	2.53	0.37	1.13	0.89	0.20	0.72	0.48
中位数	65.53	0.00	9.83	6.10	1.17	6.19	2.11
众数	/	0.00	/	0.00	0.00	/	0.00
标准差	8.76	1.29	3.92	3.09	0.68	2.49	1.67
方差	76.65	1.66	15.37	9.56	0.46	6.21	2.78
最小值	59.01	0.00	0.00	0.00	0.00	0.00	0.00
最大值	87.05	3.38	14.52	8.70	1.98	11.15	6.04
观测数	12.00	12.00	12.00	12.00	12.00	12.00	12.00

续表：

	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
平均数	2.45	0.41	0.60	1.40	0.04	0.20	0.10
标准误	0.48	0.17	0.28	0.41	0.01	0.20	0.05
中位数	2.35	0.16	0.00	1.02	0.02	0.00	0.00
众数	2.18	0.00	0.00	/	0.00	0.00	0.00
标准差	1.66	0.59	0.98	1.43	0.05	0.68	0.19
方差	2.76	0.35	0.96	2.06	0.00	0.46	0.03
最小值	0.00	0.00	0.00	0.00	0.00	0.00	0.00
最大值	5.09	1.62	2.86	4.50	0.12	2.36	0.47
观测数	12.00	12.00	12.00	12.00	12.00	12.00	12.00

表 5-3 铅钡风化

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3
平均	24.91	0.22	0.13	2.70	0.65	2.97	0.58
标准误	2.08	0.11	0.05	0.33	0.14	0.52	0.14
中位数	25.02	0.00	0.00	2.88	0.57	2.38	0.31
众数	/	0.00	0.00	3.19	0.00	/	0.00
标准差	10.61	0.56	0.24	1.66	0.71	2.63	0.74
方差	112.48	0.31	0.06	2.75	0.50	6.94	0.54
最小值	3.72	0.00	0.00	0.00	0.00	0.45	0.00
最大值	53.33	2.22	1.05	6.40	2.73	13.65	2.74
观测数	26.00	26.00	26.00	26.00	26.00	26.00	26.00

续表：

	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
平均	2.28	43.31	11.81	5.28	0.42	0.07	1.37
标准误	0.55	2.40	1.96	0.82	0.05	0.05	0.82
中位数	1.15	44.06	8.79	4.98	0.43	0.00	0.00
众数	1.51	/	0.00	0.00	0.00	0.00	0.00

标准差	2.82	12.23	9.98	4.20	0.26	0.27	4.21
方差	7.96	149.58	99.57	17.61	0.07	0.07	17.69
最小值	0.00	15.71	0.00	0.00	0.00	0.00	0.00
最大值	10.57	70.21	35.45	14.13	1.12	1.31	15.95
观测数	26.00	26.00	26.00	26.00	26.00	26.00	26.00

表 5-4 铅钡未风化

	SiO ₂	Na ₂ O	K ₂ O	CaO	MgO	Al ₂ O ₃	Fe ₂ O ₃
平均数	67.98	0.70	9.33	5.33	1.08	6.62	1.93
标准误	2.53	0.37	1.13	0.89	0.20	0.72	0.48
中位数	65.53	0.00	9.83	6.10	1.17	6.19	2.11
众数	/	0.00	/	0.00	0.00	/	0.00
标准差	8.76	1.29	3.92	3.09	0.68	2.49	1.67
方差	76.65	1.66	15.37	9.56	0.46	6.21	2.78
最小值	59.01	0.00	0.00	0.00	0.00	0.00	0.00
最大值	87.05	3.38	14.52	8.70	1.98	11.15	6.04
观测数	12.00	12.00	12.00	12.00	12.00	12.00	12.00

续表：

	CuO	PbO	BaO	P ₂ O ₅	SrO	SnO ₂	SO ₂
平均数	2.45	0.41	0.60	1.40	0.04	0.20	0.10
标准误	0.48	0.17	0.28	0.41	0.01	0.20	0.05
中位数	2.35	0.16	0.00	1.02	0.02	0.00	0.00
众数	2.18	0.00	0.00	/	0.00	0.00	0.00
标准差	1.66	0.59	0.98	1.43	0.05	0.68	0.19
方差	2.76	0.35	0.96	2.06	0.00	0.46	0.03
最小值	0.00	0.00	0.00	0.00	0.00	0.00	0.00
最大值	5.09	1.62	2.86	4.50	0.12	2.36	0.47
观测数	12.00	12.00	12.00	12.00	12.00	12.00	12.00

附录 2：风化前的预测结果

编号	风化前后	二氧化硅 (SiO ₂)	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)
23	风化后 (已知)	53.79	7.92	0.00	0.50	0.71	1.42	0.00	2.99	16.98	11.86	0.00	0.33	0.00	0.00
	风化前 (预测)	81.39	9.36	0.09	-0.96	0.66	2.78	0.08	2.20	-4.65	10.39	-3.71	0.22	-0.03	-0.04
25	风化后 (已知)	50.61	2.31	0.00	0.63	0.00	1.90	1.55	1.12	31.90	6.65	0.19	0.20	0.00	0.00
	风化前 (预测)	78.21	3.75	0.09	-0.83	-0.05	3.26	1.63	0.33	10.27	5.18	-3.52	0.09	-0.03	-0.04
28	风化后 (已知)	68.08	0.00	0.26	1.34	1.00	4.70	0.41	0.33	17.14	4.04	1.04	0.12	0.23	0.00
	风化前 (预测)	95.68	1.44	0.35	-0.12	0.95	6.06	0.49	-0.46	-4.49	2.57	-2.67	0.01	0.20	-0.04
29	风化后 (已知)	63.30	0.92	0.30	2.98	1.49	14.34	0.81	0.74	12.31	2.03	0.41	0.25	0.00	0.00
	风化前 (预测)	90.90	2.36	0.39	1.52	1.44	15.70	0.89	-0.05	-9.32	0.56	-3.30	0.14	-0.03	-0.04
42	风化后 (已知)	51.26	5.74	0.15	0.79	1.09	3.53	0.00	2.67	21.88	10.47	0.08	0.35	0.00	0.00
	风化前 (预测)	78.86	7.18	0.24	-0.67	1.04	4.89	0.08	1.88	0.25	9.00	-3.63	0.24	-0.03	-0.04

	无风化部位 1	51.26	5.74	0.15	0.79	1.09	3.53	0.00	2.67	21.88	10.47	0.08	0.35	0.00	0.00
42	无风化部位 2	51.33	5.68	0.35	0.00	1.16	5.66	0.00	2.72	20.12	10.88	0.00	0.00	0.00	0.00
	风化前(平均)	51.30	5.71	0.25	0.40	1.13	4.60	0.00	2.70	21.00	10.68	0.04	0.18	0.00	0.00
	风化后(已知)	60.74	3.06	0.20	2.14	0.00	12.69	0.77	0.43	13.61	5.22	0.00	0.26	0.00	0.00
44	风化前(预测)	88.34	4.50	0.29	0.68	-0.05	14.05	0.85	-0.36	-8.02	3.75	-3.71	0.15	-0.03	-0.04
	无风化点(已知)	54.61	0.00	0.30	2.08	1.20	6.50	1.27	0.45	23.02	4.19	4.32	0.30	0.00	0.00
49	无风化点(已知)	45.02	0.00	0.00	3.12	0.54	4.16	0.00	0.70	30.61	6.22	6.34	0.23	0.00	0.00
53	风化后(已知)	63.66	3.04	0.11	0.78	1.14	6.06	0.00	0.54	13.66	8.99	0.00	0.27	0.00	0.00
	风化前(预测)	91.26	4.48	0.20	-0.68	1.09	7.42	0.08	-0.25	-7.97	7.52	-3.71	0.16	-0.03	-0.04
02	风化后(已知)	36.28	0.00	1.05	2.34	1.18	5.73	1.86	0.26	47.43	0.00	3.57	0.19	0.00	0.00
	风化前(预测)	63.88	1.44	1.14	0.88	1.13	7.09	1.94	-0.53	25.80	-1.47	-0.14	0.08	-0.03	-0.04
08	风化后(已知)	20.14	0.00	0.00	1.48	0.00	1.34	0.00	10.41	28.68	31.23	3.59	0.37	0.00	2.58
	风化前(预测)	47.74	1.44	0.09	0.02	-0.05	2.70	0.08	9.62	7.05	29.76	-0.12	0.26	-0.03	2.54
08	风化后(已知)	4.61	0.00	0.00	3.19	0.00	1.11	0.00	3.14	32.45	30.62	7.56	0.53	0.00	15.03
	风化前(预测)	32.21	1.44	0.09	1.73	-0.05	2.47	0.08	2.35	10.82	29.15	3.85	0.42	-0.03	14.99
11	风化后(已知)	33.59	0.00	0.21	3.51	0.71	2.69	0.00	4.93	25.39	14.61	9.38	0.37	0.00	0.00
	风化前(预测)	61.19	1.44	0.30	2.05	0.66	4.05	0.08	4.14	3.76	13.14	5.67	0.26	-0.03	-0.04
19	风化后(已知)	29.64	0.00	0.00	2.93	0.59	3.57	1.33	3.51	42.82	5.35	8.83	0.19	0.00	0.00
	风化前(预测)	57.24	1.44	0.09	1.47	0.54	4.93	1.41	2.72	21.19	3.88	5.12	0.08	-0.03	-0.04
26	风化后(已知)	19.79	0.00	0.00	1.44	0.00	0.70	0.00	10.57	29.53	32.25	3.13	0.45	0.00	1.96
	风化前(预测)	47.39	1.44	0.09	-0.02	-0.05	2.06	0.08	9.78	7.90	30.78	-0.58	0.34	-0.03	1.92
26	风化后(已知)	3.72	0.00	0.40	3.01	0.00	1.18	0.00	3.60	29.92	35.45	6.04	0.62	0.00	15.95
	风化前(预测)	31.32	1.44	0.49	1.55	-0.05	2.54	0.08	2.81	8.29	33.98	2.33	0.51	-0.03	15.91
34	风化后(已知)	35.78	0.00	0.25	0.78	0.00	1.62	0.47	1.51	46.55	10.00	0.34	0.22	0.00	0.00
	风化前(预测)	63.38	1.44	0.34	-0.68	-0.05	2.98	0.55	0.72	24.92	8.53	-3.37	0.11	-0.03	-0.04
36	风化后(已知)	39.57	2.22	0.14	0.37	0.00	1.60	0.32	0.68	41.61	10.83	0.07	0.22	0.00	0.00
	风化前(预测)	67.17	3.66	0.23	-1.09	-0.05	2.96	0.40	-0.11	19.98	9.36	-3.64	0.11	-0.03	-0.04
38	风化后(已知)	32.93	1.38	0.00	0.68	0.00	2.57	0.29	0.73	49.31	9.79	0.48	0.41	0.00	0.00
	风化前(预测)	60.53	2.82	0.09	-0.78	-0.05	3.93	0.37	-0.06	27.68	8.32	-3.23	0.30	-0.03	-0.04
39	风化后(已知)	26.25	0.00	0.00	1.11	0.00	0.50	0.00	0.88	61.03	7.22	1.16	0.61	0.00	0.00
	风化前(预测)	53.85	1.44	0.09	-0.35	-0.05	1.86	0.08	0.09	39.40	5.75	-2.55	0.50	-0.03	-0.04
40	风化后(已知)	16.71	0.00	0.00	1.87	0.00	0.45	0.19	0.00	70.21	6.69	1.77	0.68	0.00	0.00
	风化前(预测)	44.31	1.44	0.09	0.41	-0.05	1.81	0.27	-0.79	48.58	5.22	-1.94	0.57	-0.03	-0.04

41	风化后 (已知)	18.46	0.00	0.44	4.96	2.73	3.33	1.79	0.19	44.12	9.76	7.46	0.47	0.00	0.00
	风化前 (预测)	46.06	1.44	0.53	3.50	2.68	4.69	1.87	-0.60	22.49	8.29	3.75	0.36	-0.03	-0.04
43	风化部位 1 (已 知)	12.41	0.00	0.00	5.24	0.89	2.25	0.76	5.35	59.85	7.29	0.00	0.64	0.00	0.00
	风化部位 2 (预 测)	21.70	0.00	0.00	6.40	0.95	3.41	1.39	1.51	44.75	3.26	12.83	0.47	0.00	0.00
	风化后 (平均)	17.06	0.00	0.00	5.82	0.92	2.83	1.08	3.43	52.30	5.28	6.42	0.56	0.00	0.00
	风化前 (预测)	44.66	1.44	0.09	4.36	0.87	4.19	1.15	2.64	30.67	3.80	2.70	0.45	-0.03	-0.04
48	风化后 (已知)	53.33	0.80	0.32	2.82	1.54	13.65	1.03	0.00	15.71	7.31	1.10	0.25	1.31	0.00
	风化前 (预测)	80.93	2.24	0.41	1.36	1.49	15.01	1.11	-0.79	-5.92	5.84	-2.61	0.14	1.28	-0.04
49	风化后 (已知)	28.79	0.00	0.00	4.58	1.47	5.38	2.74	0.70	34.18	6.10	11.10	0.46	0.00	0.00
	风化前 (预测)	56.39	1.44	0.09	3.12	1.42	6.74	2.82	-0.09	12.55	4.63	7.39	0.35	-0.03	-0.04
50	风化后 (已知)	17.98	0.00	0.00	3.19	0.47	1.87	0.33	1.13	44.00	14.20	6.34	0.66	0.00	0.00
	风化前 (预测)	45.58	1.44	0.09	1.73	0.42	3.23	0.41	0.34	22.37	12.73	2.63	0.55	-0.03	-0.04
51	风化后 (已知)	24.61	0.00	0.00	3.58	1.19	5.25	1.19	1.37	40.24	8.94	8.10	0.39	0.47	0.00
	风化前 (预测)	52.21	1.44	0.09	2.12	1.14	6.61	1.27	0.58	18.61	7.47	4.39	0.28	0.44	-0.04
51	风化后 (已知)	21.35	0.00	0.00	5.13	1.45	2.51	0.42	0.75	51.34	0.00	8.75	0.00	0.00	0.00
	风化前 (预测)	48.95	1.44	0.09	3.67	1.40	3.87	0.50	-0.04	29.71	-1.47	5.04	-0.11	-0.03	-0.04
52	风化后 (已知)	25.74	1.22	0.00	2.27	0.55	1.16	0.23	0.70	47.42	8.64	5.71	0.44	0.00	0.00
	风化前 (预测)	53.34	2.66	0.09	0.81	0.50	2.52	0.31	-0.09	25.79	7.17	2.00	0.33	-0.03	-0.04
54	风化后 (已知)	22.28	0.00	0.32	3.19	1.28	4.15	0.00	0.83	55.46	7.04	4.24	0.88	0.00	0.00
	风化前 (预测)	49.88	1.44	0.41	1.73	1.23	5.51	0.08	0.04	33.83	5.57	0.53	0.77	-0.03	-0.04
56	风化后 (已知)	29.15	0.00	0.00	1.21	0.00	1.85	0.00	0.79	41.25	15.45	2.54	0.00	0.00	0.00
	风化前 (预测)	56.75	1.44	0.09	-0.25	-0.05	3.21	0.08	0.00	19.62	13.98	-1.17	-0.11	-0.03	-0.04
57	风化后 (已知)	25.42	0.00	0.00	1.31	0.00	2.18	0.00	1.16	45.10	17.30	0.00	0.00	0.00	0.00
	风化前 (预测)	53.02	1.44	0.09	-0.15	-0.05	3.54	0.08	0.37	23.47	15.83	-3.71	-0.11	-0.03	-0.04
58	风化后 (已知)	30.39	0.00	0.34	3.49	0.79	3.52	0.86	3.13	39.35	7.66	8.99	0.24	0.00	0.00
	风化前 (预测)	57.99	1.44	0.43	2.03	0.74	4.88	0.94	2.34	17.72	6.19	5.28	0.13	-0.03	-0.04

附录 3：铅钡类玻璃样品聚类成员表

个案号	文物采 样点	聚类	距离	25	34	2	12.752
1	20	1	24.702	26	36	1	11.685
2	24	1	17.041	27	38	2	10.18
3	30 部位 1	1	11.466	28	39	2	13.678
4	30 部位 2	1	8.921	29	40	2	23.674

5	31	4	8.76	30	41	2	8.964
6	32	4	10.59	31	42 未风化点1	4	11.215
7	33	4	16.277	32	42 未风化点2	4	10.745
8	35	4	8.365	33	43 部位1	2	18.346
9	37	4	6.361	34	43 部位2	2	10.462
10	45	4	4.729	35	44 未风化点	4	9.092
11	46	4	9.387	36	48	4	11.417
12	47	4	12.001	37	49	2	16.403
13	55	1	11.369	38	49 未风化点	4	9.144
14	2	2	14.316	39	50	2	10.685
15	8	3	11.111	40	50 未风化点	1	9.669
16	08 严重风化点	3	11.099	41	51 部位1	2	9.178
17	11	1	11.299	42	51 部位2	2	9.727
18	19	2	8.762	43	52	2	2.894
19	23 未风化点	4	11.064	44	53 未风化点	4	6.273
20	25 未风化点	1	13.072	45	54	2	7.712
21	26	3	11.219	46	54 严重风化点	2	17.293
22	26 严重风化点	3	11.904	47	56	2	12.162
23	28 未风化点	4	8.817	48	57	2	11.905
24	29 未风化点	4	12.523	49	58	2	11.438

附录 4:

%蒙特卡罗方法检验铅钡玻璃亚分类结果的敏感性 matlab 代码

```

clc;clear
load data_qianbei_middle.mat
t = 100; % 重复实验次数
middle=middle';
[n,m] = size(X);
u = zeros(t,1); % 扰动项 u
b = zeros(t,1); %每次试验的分类准确率
result=zeros(n,1);
mindis=ones(n,1)*+inf;

```

```

for j= 1: t
for i = 1: 4
    u = -1+rand(1,m)*2;    % u 在[-1,1]上随机分布，大小为 1*m
    x1=X+X.*repmat(0.05*u,n,1); %x1 为原数据数值随机增减 0~5%之后的带干扰
的数据
    zhongxin=repmat(middle(i,:),n,1);
    dis=sum((x1-zhongxin).^2,2).^0.5;
    if sum(dis<mindis) %寻找并归类距离最近的聚类中心点
        result(dis<mindis)=i;
        mindis(dis<mindis)=dis(dis<mindis);
    end
end
b(j)=sum(cluster==result)/n;
end
disp(['重复',num2str(t),'次实验平均预测准确度为:',num2str(sum(b)/t)])

```

附录 5

%蒙特卡罗方法检验高钾玻璃亚分类结果的敏感性 matlab 代码

```

clc;clear
load data_gaojia_middle.mat
t = 100; % 重复实验次数
middle=middle';
[n,m] = size(X);
u = zeros(t,1); % 扰动项 u
b = zeros(t,1); %每次试验的分类准确率
result=zeros(n,1);
mindis=ones(n,1)*+inf;
for j= 1: t
for i = 1: 4
    u = -1+rand(1,m)*2;    % u 在[-1,1]上随机分布，大小为 1*m
    x1=X+X.*repmat(0.05*u,n,1); %x1 为原数据数值随机增减 0~5%之后的带干扰
的数据
    zhongxin=repmat(middle(i,:),n,1);
    dis=sum((x1-zhongxin).^2,2).^0.5;
    if sum(dis<mindis) %寻找并归类距离最近的聚类中心点
        result(dis<mindis)=i;
        mindis(dis<mindis)=dis(dis<mindis);
    end
end
b(j)=sum(cluster==result)/n;
end
disp(['重复',num2str(t),'次实验平均预测准确度为:',num2str(sum(b)/t)])

```

附录 6:

文物编号	预测结果_Y	预测结果概率_铅钡	预测结果概率_高钾
------	--------	-----------	-----------

A1	高钾	0.09	0.91
A1_u1	高钾	0.09	0.91
A1_u2	高钾	0.09	0.91
A1_u3	高钾	0.09	0.91
A2	铅钡	0.85	0.15
A2_u1	铅钡	0.85	0.15
A2_u2	铅钡	0.85	0.15
A2_u3	铅钡	0.85	0.15
A3	铅钡	0.88	0.12
A3_u1	铅钡	0.88	0.12
A3_u2	铅钡	0.89	0.11
A3_u3	铅钡	0.88	0.12
A4	铅钡	0.87	0.13
A4_u1	铅钡	0.87	0.13
A4_u2	铅钡	0.87	0.13
A4_u3	铅钡	0.87	0.13
A5	铅钡	0.87	0.13
A5_u1	铅钡	0.87	0.13
A5_u2	铅钡	0.87	0.13
A5_u3	铅钡	0.87	0.13
A6	高钾	0.02	0.98
A6_u1	高钾	0.02	0.98
A6_u2	高钾	0.02	0.98
A6_u3	高钾	0.02	0.98
A7	高钾	0.06	0.94
A7_u1	高钾	0.06	0.94
A7_u2	高钾	0.06	0.94
A7_u3	高钾	0.06	0.94
A8	铅钡	1	0
A8_u1	铅钡	1	0
A8_u2	铅钡	1	0
A8_u3	铅钡	1	0