

基于多种模型对古代玻璃制品的成分分析与鉴别

摘要

本文主要研究了古代玻璃制品的成分分析与鉴别问题。通过对附件中的数据进行多角度分析，本文建立了卡方检验、肯德尔等级相关系数、Kmeans 聚类、随机森林等多种数据模型，并以此为基础，探究了这些玻璃文物的各项性质和化学成分与表面风化的关系，还原了风化点数据，同时得到了不同玻璃类型的分类规律并对不同类别进行了亚类划分，解决了未知古代玻璃文物的玻璃类型预测问题，并对不同化学成分之间的关系进行了研究。

对于问题一，本文建立了卡方检验模型，对古代玻璃的纹饰、类别、颜色和表面风化进行差异性检验，以此判断古代玻璃的表面风化与这三种性质的关系；同时，本文建立了肯德尔等级相关系数模型，探究了不同类型下古代玻璃表面风化与各种化学成分之间的关系；同时，本文根据相关性较为显著的多项化学成分，通过分别乘以对应的化学成分的无风化有风化平均值比值，将不同玻璃类型的风化后数据还原为风化前。结果显示：古代玻璃文物的纹饰、颜色与表面风化无关，类型与表面风化有关，铅钡玻璃有更大可能表面有风化，高钾玻璃有更大可能表面无风化。铅钡玻璃表面风化与二氧化硅、氧化钙、氧化铅、五氧化二磷显著相关，而高钾玻璃表面风化与二氧化硅、氧化钾、氧化铝显著相关。还原后的风化点数据见附录 1.3。

对于问题二，本文建立了随机森林模型和 Kmeans 聚类模型，训练了随机森林模型预测并分类玻璃类型的能力，并得到了高钾玻璃和铅钡玻璃的分类规律；同时，本文通过选择合适的化学成分含量作为维度，并选取合适的 K 值，以此建立了 Kmeans 聚类模型，对不同的玻璃类型、颜色、纹饰进行了亚类划分，并且进行了合理性和敏感度分析。结果显示：高钾玻璃和铅钡玻璃对分类主要取决于氧化铅、氧化钡、氧化钾、氧化钙、氧化锶、二氧化硅、五氧化二磷的数值大小，具体区分标准见表 7。

对于问题三，本文使用了问题二中已经训练好了的随机森林模型，对未知玻璃类型的数据进行玻璃类型分类预测。因为本题使用的是随机森林模型，参数调优的过程等同于敏感度检验。结果表明：预测结果为 A1、A6 和 A7 为高钾，A2、A3、A4、A5 和 A8 为铅钡。

对于问题四，本文建立了肯德尔等级相关系数模型，分别对不同玻璃类型、不同纹饰的化学成分两两计算肯德尔等级相关系数，并进行显著性检验，从而判断化学成分之间的相关性并且衡量不同玻璃类型、不同纹饰之间化学成分关联关系的差异性。结果表明：高钾玻璃与铅钡玻璃中的化学成分之间均存在多对强或中等强度相关性的化学元素；高钾玻璃和铅钡玻璃之间相差较大的化学成分关联关系为：二氧化硅与氧化钠、二氧化硅与氧化钾、二氧化硅与氧化铝、氧化钠与氧化钙、氧化镁与氧化钡、氧化铝与氧化铅、氧化铁与氧化铜、氧化铁与氧化钡、氧化铅与氧化钡。各个纹饰的玻璃其化学成分之间的关联关系均为弱或无相关。不同类型玻璃文物样品之间的相关系数无显著差异性。

关键词：古代玻璃，卡方检验，肯德尔等级相关系数，随机森林，Kmeans 聚类算法

一、问题重述

1.1 问题背景

丝绸之路是古代中西方文化交流的通道，其中玻璃是早期贸易往来的宝贵物证。早期的玻璃在西亚和埃及地区常被制作成珠形饰品传入我国，我国古代玻璃吸收其技术后在本土就地取材制作，因此与外来的玻璃制品外观相似，但化学成分却不相同。玻璃的主要原料是石英砂。由于纯石英砂的熔点较高，为了降低熔化温度，在炼制时需要添加助熔剂。古代常用的助熔剂有草木灰、天然泡碱、硝石和铅矿石等，并添加石灰石作为稳定剂。添加的助熔剂不同，其主要化学成分也不同。古代玻璃极易受埋藏环境的影响而风化。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。被标记为表面无风化的文物，仍可能在局部有较浅的风化，可能会影响到某些关键信息的辨认；标记为表面风化的文物，甚至可能纹饰都无法辨别，颜色的辨认也会受到影响。已经风化的古代玻璃文物可能难以从外观上判断其类别，轻微风化的玻璃文物也可能因风化区域影响类别鉴定。如果能够通过化学成分来判断、预测古代玻璃文物的类别，甚至通过分析划分出新的亚类，将会对一些脆弱或损坏严重的古代玻璃文物的鉴别、保护、修复和已有的古代玻璃文物的研究起到重要作用。

1.2 问题提出

为了更好地对古代玻璃文物进行成分分析和鉴别，根据题目所附带的附件数据，我们建立数学模型解决下列问题：

- (1) 对这些玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系进行分析；结合玻璃的类型，分析文物样品表面有无风化化学成分含量的统计规律，并根据风化点检测数据，预测其风化前的化学成分含量。
- (2) 依据附件数据分析高钾玻璃、铅钡玻璃的分类规律；对于每个类别选择合适的化学成分对其进行亚类划分，给出具体的划分方法及划分结果，并对分类结果的合理性和敏感性进行分析。
- (3) 对附件表单 3 中未知类别玻璃文物的化学成分进行分析，鉴别其所属类型，并对分类结果的敏感性进行分析。
- (4) 针对不同类别的玻璃文物样品，分析其化学成分之间的关联关系，并比较不同类别之间的化学成分关联关系的差异性。

二、问题分析

1. 该问题首先要求我们分析玻璃文物表面风化与其玻璃类型、纹饰和颜色之间的关系。这是一个对不同变量进行相关性判断的问题。考虑到附件中的数据类型均为离散型数据（定类数据），我们可以以卡方检验作为主要思路，将玻璃类型、纹饰和颜色数据分别与玻璃文物表面风化数据分别构成两组变量，利用附件中的表一数据分别对这三对变量做卡方检验，取某一数值作为卡方检验的 p 值临界值，若某一组变量的显著性检验的 p 值小于该临界值，则认为该种性质（玻璃类型、纹饰或颜色）的不同取值会导致玻璃文物表面风化数据出现显著差异，由此认为该种性质与玻璃文物表面风化相关，具体相关性通过卡方检验的结果表格进行具体分析。 p 值越小，不同性质（如高钾玻璃与铅钡玻璃）导致的表面风化的差异性越显著，即该种性质与表面风化的相关性越强，由此判断出三种性质与玻璃文物表面风化之间的关系。之后，问题要求我们结合玻璃类型，分析文物样品表面有无风化化学成分含量的统计规律。这是一个对不同玻璃类型的数据分别进行玻璃文物表面风化和不同化学成分含量之间的相关性判断的问题。考虑到样本数据较少，我们可以采用肯德尔等级相关系数作为主要工具，将不同玻璃类型的不同化学成分含量和表面风化作为两组肯德尔等级相关变量，利用附件中的表一和表二数据得到 28 个肯德尔等级相关系数，取某一组数值为肯德尔等级相关系数的临界值，若某两组数据的相关系数大于该值，则认为在某一种玻璃类型中某种化学成分含量与文物表面风化有直接关联，否则认为在某一种玻璃类型中某种化学成分含量与文物表面风化没有直接关联。为了检验该种方法的正确性，可以进行显著性检验，若某两组数据的显著性检验通过，则证明使用肯德尔等级相关系数分析该两组数据的正确性。为了分析不同玻璃类型中不同化学成分含量的重要性，可建立肯德尔等级相关模型，基于肯德尔等级相关系数的区间范围，将不同玻璃类型的不同化学成分含量与表面风化的关系进行分类，相关系数越大，关系越密切，从而判断出不同玻璃类型下不同化学成分含量与表面风化的对应关系。同时，问题还要求我们根据风化后的数据预测出风化前的数据。这就根据风化前后样本平均值形成比例进行优化计算。
2. 问题二要求我们给出古代玻璃文物玻璃类型的分类规律，并对每个类别（玻璃类型、纹饰、颜色）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分析分类结果的合理性和敏感性。前半题是分类问题，后半题是聚类问题。因此，我们以随机森林和 Kmeans 聚类模型作为主要思路，首先利用数据集和化学成分属性集训练随机森林分类玻璃类型，并且得到最优特征子集，结合随机森林中多个决策树的评判标准得到玻璃类型的分类规律；之后，我们继续利用数据集和化学成分属性集训练随机森林分类纹饰、颜色，最终得到划分玻璃类型、纹饰、颜色的最优特征子集。同时我们通过轮廓系数得到各自最优的 K 值。我们根据最优特征子集和最优 K 值构建 Kmeans 聚类模型，最终得到玻璃类型、颜色、纹饰下的多种亚类。敏感性可以通过分析 K 值在最优 K 值附近波动时轮廓系数的变化来得到结论，合理性可以通过分析亚类的性质统一性来得到结论。

- 问题三要求我们对附件表单 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。这是一个分类预测问题，预测所需要的模型用问题二中已经训练好的随机森林模型，将附录中表三数据输入到随机森林模型中得到预测结果。随机森林模型的参数调优过程等同于敏感度检验。
- 问题四要求我们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。这是一个对不同变量进行相关性判断，并对相关性数值进行差异性判断的问题。考虑到样本数较低，我们以肯德尔等级相关系数作为主要思路，对不同玻璃类型的数据，将不同化学成分含量两两组合构成两组变量，利用附件中的表二数据计算它们的肯德尔等级相关系数，从而得到不同玻璃类型下不同化学成分含量之间的相关性，同时进行显著性检验来检验相关系数的可靠性。之后，对通过了显著性检验的相关系数，

三、基本假设

- 假设题目附件所给的古代玻璃文物鉴定和成分分析数据真实有效；
- 假设已鉴定和未鉴定的古代玻璃文物在进行化学成分检验时均没有出现明显误差；

四、符号说明

符号	说明
χ^2	卡方检验卡方值
τ_B	肯德尔相关系数
$Gini_index$	基尼指数
T_{rate}	预测准确率
\hat{f}	随机森林预测结果
$dis(X'_i, C_j)$	第 i 个数据点到第 j 个类簇中心点的欧氏距离
C_y	第 y 个类簇中心点
X	输入到 Kmeans 中的训练数据集
X'	标准化后的 X
D	输入到随机森林中的训练数据集
D_t	输入到随机森林中的检测数据集
A	属性集

五、数据预处理

数据清洗：对于含有缺失值的数据，我们采取了删除处理，如表一中不含颜色数据的19、40、48、58号文物数据。同时，对附录中表二数据，我们计算了各化学成分占比的总和，依照题干所给的成分比例累加有效区间，排除了15、17号文物的数据，剩余有效数据67个。

六、模型的建立与求解

6.1 问题一模型的建立与求解

6.1.1 问题一模型的建立

问题一可以分为两个小题。小题一要求我们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；小题二要求我们结合玻璃的类型，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。对小题一，我们建立卡方检验模型，分别以纹饰、玻璃类型、颜色作为X值，表面风化作为Y值，进行卡方检验并进行显著性检验。若通过显著性检验，则说明不同的某种表面性质（纹饰、玻璃类型、颜色等）会使表面风化产生显著差异，该种表面性质与表面风化显著相关；否则则说明不同的纹饰或玻璃类型或颜色下表面风化具有一致性，该种表面性质与表面风化不相关。

问题一小题一模型的准备

卡方检验模型。

查找相关资料^[1]可得，卡方检验模型是用来统计样本的实际观测值和理论推断值之间的偏离程度，实际观测值与理论推断值之间的偏离程度决定了卡方值的大小。卡方值越大，偏离程度越大，理论值和实际值之间的差异也越大。因此，对于两个变量X、Y，我们可以设定 H_0 为“X与Y有关联”，并利用独立性检验来考察X、Y之间是否有关系。

因本小题数据为离散型，则设定假设为：

H_0 ：总体X的分布率为 $P\{X = x_i\} = p_i, i = 1, 2, 3, \dots$

把等于 x_i 的样本值个数记为 f_i ，成为组频数（真实值），样本容量 n 为所有组频数之和。

当 H_0 为真时，根据假设的总体理论分布，可算出总体 X 的值等于 x_i 的概率 p_i ，于是， np_i 就是等于 x_i 的样本值的理论频数（理论值）

若 H_0 为真，则 n 次实验中等于 x_i 值的频率 f_i/n_i 应该接近概率 p_i ，则引入检验统计量：

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

r 行 c 列表数据的卡方检验的卡方值为： $n[(\frac{A_{11}}{n_1 n_1} + \frac{A_{12}}{n_1 n_2} + \dots + \frac{A_{rc}}{n_r n_c}) - 1]$

问题一小题二模型的准备

肯德尔等级相关系数 (Kendall' s tau correlation coefficient)

查找相关资料可得^[2]，肯德尔等级相关系数通常被用来测量两个被测量量之间的序数联系。

$(x_1, y_1), \dots, (x_n, y_n)$ 是对一组联合随机变量的观察，任取 i ， (x_i) 和 (y_i) 都是独一无二的。

如果任取 $i < j$ ， $x_i > x_j$ ， $y_i > y_j$ 或者 $x_i < x_j$ ， $y_i < y_j$ ，则认为 X 和 Y 是一致的

(*concordant*)；如果任取 $i < j$ ， $x_i < x_j$ ， $y_i > y_j$ 或者 $x_i < x_j$ ， $y_i > y_j$ ，则认为 X 和 Y 是不一致的 (*discordant*)；如果存在 $i < j$ ， $x_i = x_j$ ， $y_i = y_j$ ，则称 $\{(x_i, y_i), (x_j, y_j)\}$ 为绑定对。则肯德尔系数为：

$$\begin{aligned} \tau &= \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{(\text{number of pairs})} \\ &= 1 - \frac{2(\text{number of discordant pairs})}{\binom{n}{2}} \end{aligned}$$

τ 取值区间为 $[-1, 1]$ ，若 X 与 Y 之间的一致性完美的，则 τ 等于1；若 X 与 Y 之间的分歧是完美的，则 τ 等于-1；若 X 与 Y 独立且不恒定，则 τ 的期望为0。

Tau-b 统计量：Tau-b 的值范围从 -1（100% 负关联，或完全反转）到 +1（100% 正关联，或完全一致）。零值表示不存在关联。

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \quad (1-2-1)$$

$$\begin{aligned} n_0 &= n(n-1)/2; \\ n_1 &= \sum_i t_i(t_i-1)/2; \end{aligned}$$

$$n_2 = \sum_j u_j(u_j - 1)/2;$$

n_c = number of concordant pairs;

n_d = number of discordant pairs ;

t^i = number of tied values in the i^{th} group of ties for the first quantity ;

u^j = number of tied values in the j^{th} group of ties for the second quantity ;

τ 的显著性检验:

$$z_B = \frac{n_c - n_d}{\sqrt{v}} \quad (1-2-2)$$

$$v = (v_0 - v_t - v_u)/18 + v_1 + v_2;$$

$$v_0 = n(n-1)(2n+5);$$

$$v_t = \sum_i t_i(t_i-1)(2t_i+5);$$

$$v_u = \sum_j u_j(u_j-1)(2u_j+5);$$

$$v_1 = \sum_i t_i(t_i-1) \sum_j u_j(u_j-1)/(2n(n-1));$$

$$v_2 = \sum_i t_i(t_i-1)(t_i-2) \sum_j u_j(u_j-1)(u_j-2)/(9n(n-1)(n-2));$$

模型的建立

一、 卡方检验

将 C 题附件中表一的处理后的数据作为数据集，删去其中缺失颜色数据的四个文物数据（19、40、48、58 号文物）。之后，分别对表面风化与纹饰、表面风化与类型、表面风化与颜色进行卡方检验并进行显著性检验，画出模型结构如下：

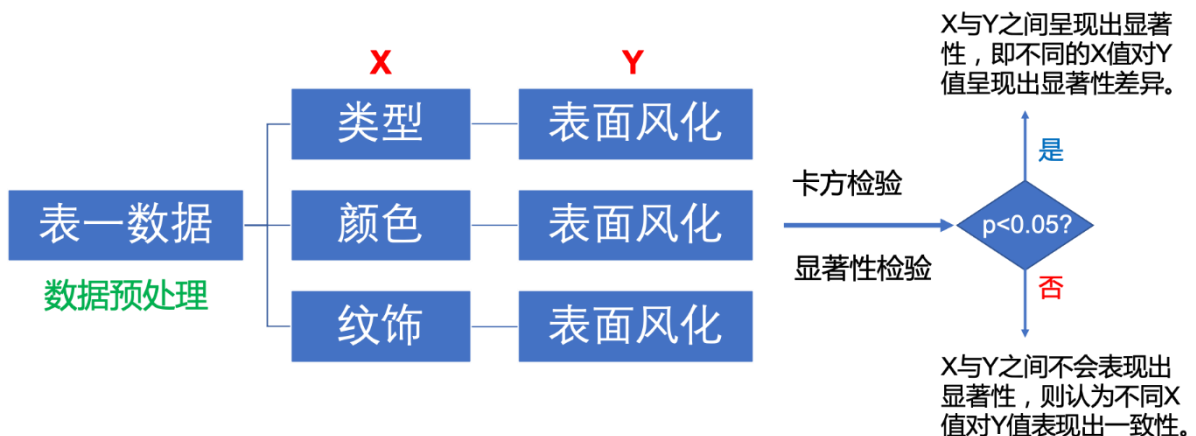


图 1：卡方检验模型结构图

由上图可知，表面风化与纹饰、类型、颜色的关系可以参考卡方检验的显著性和表格数据来判断。至此，我们建立卡方检验相关模型如下：

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - np_i)^2}{np_i} \quad (1-1-3)$$

f_i 为等于 x_i 的样本值个数，样本容量 n 为所有组频数之和。 H_0 为理论假设：总体 X 的分布率为 $P\{X = x_i\} = p_i, i = 1, 2, 3, \dots$ 。 np_i 为等于 x_i 的样本值的理论频数。

如果某两项的卡方检验的显著性检验的 p 值小于0.05，则认为这两项数据之间呈现出显著性，则该项性质（如类型、颜色、纹饰等）的不同数值会导致文物表面风化呈现出显著性差异，该项性质与文物表面风化相关；否则该项性质的不同数值下文物表面风化表现出一致性，该项性质与文物表面风化无关。

二、肯德尔等级相关系数

1、数据预处理

将C题附件中表二的处理后的数据作为数据集，删去其中不符合比例累加总数要求的两个数据（15、17号文物），并将颜色、有无风化、纹饰这三列数据添加到对应编号文物的数据中。将所有空值赋值对表面风化的文物的未风化点数据，将其“有无风化”的数据改为“无风化”。之后，将“有无风化”中的“无风化”替换为0、“有风化”替换为1。之后，将“类别”为“高钾玻璃”和“铅钡玻璃”的数据分别提取出来，得到表2钾（高钾玻璃）和表2铅（铅钡玻璃）。

2、计算肯德尔等级相关系数并进行显著性检验

在表2钾和表2铅中，分别将每个化学成分和有无风化作为 X 、 Y 值，计算相对应的肯德尔等级相关系数，从而得到不同类别下，古代玻璃的不同化学成分与有无风化的相关性；同时我们进行显著性检验，以确定相关性的可信度。则画出肯德尔等级相关系数模型结构如下：

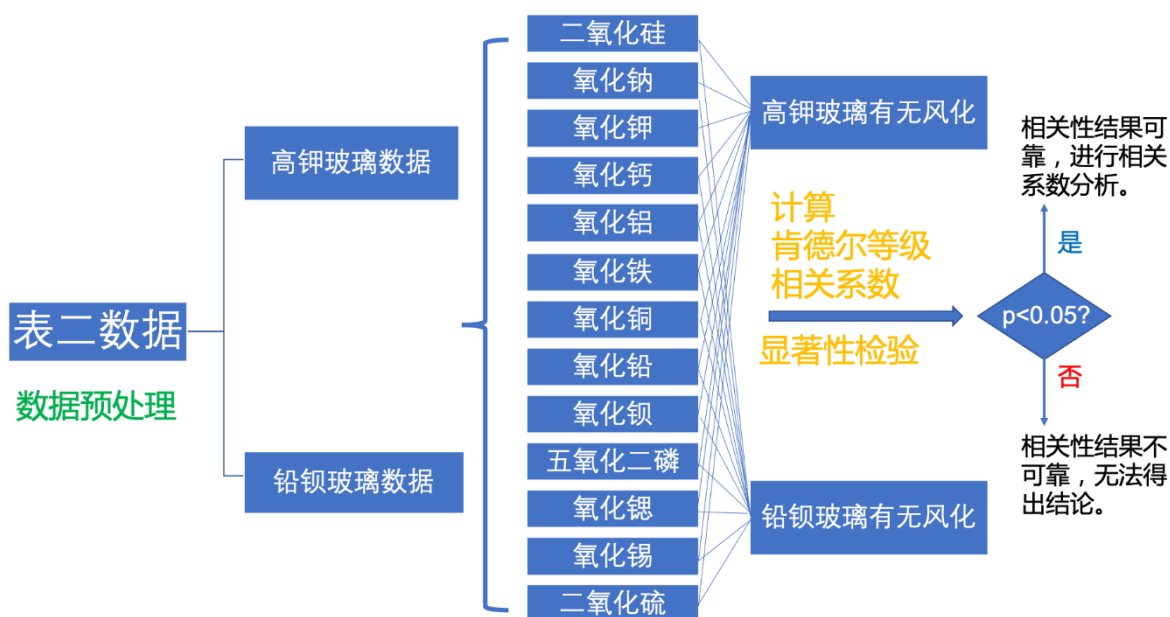


图 2：肯德尔等级相关系数模型结构图

从上图可知，不同类型的古代玻璃文物表面有无风化的化学成分含量统计规律可以通过分别计算不同类型下文物表面风化数据和不同化学成分之间的肯德尔等级相关系数来得到。由此构建肯德尔等级相关模型如下：

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \quad (1-2-3)$$

若最终计算得出 p 值小于 0.05，则认为该组数据相关，之后根据相关系数来判断不同类型玻璃表面有无风化与不同化学成分之间的相关性强弱；否则则认为该组数据不相关，无法通过相关系数得出任何统计规律。相关系数的意义如表 1 所示

表 1：相关系数与相关性强度对应值

相关系数绝对值	相关性强弱
0.8-1.0	强相关性
0.6-0.8	较强相关
0.4-0.6	中等相关
0.2-0.4	弱相关性
0-0.2	无相关或相关程度极弱

6.1.2 问题一模型的求解

一、卡方检验

根据题目所给的附件中表一的数据，在经过数据预处理之后，分别以纹饰、类型、颜色作为变量 X，以表面风化作为变量 Y，通过 SPSSAU 系统分别进行三次卡方检验，并得到表 2、表 3、表 4。

表 2：纹饰与表面风化的卡方检验分析结果

卡方检验分析结果 数字							
题目	名称	纹饰			总计	χ^2	p
		A	B	C			
表面风化	无风化	11	0	13	24	5.747	0.056
	风化	9	6	15	30		
	总计	20	6	28	54		

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

由表二可知， p 值大于 0.05，则文物不同纹饰的对表面风化表现出一致性，即认为纹饰对表面风化并无影响。

表 3：类型与表面风化的卡方检验分析结果

卡方检验分析结果 数字						
题目	名称	类型		总计	χ^2	p
		铅钡	高钾			
表面风化	无风化	12	12	24	5.400	0.020*
	风化	24	6	30		
	总计	36	18	54		

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

由表三可知， p 值小于 0.05 并大于 0.01，则类型与表面风化之间具有统计学差异。观察表三可知，铅钡类型的玻璃无风化与风化占比分别为 33.3%和 66.7%，而高钾类型的玻璃无风化与风化的占比分别为 66.7%和 33.3%。因此，我们认为古代玻璃文物的不同类型对表面是否风化具有显著影响，铅钡玻璃表面风化的可能性较大，高钾玻璃表面无风化的可能性较大。

表 4：颜色与表面风化的卡方检验分析结果

题目	名称	颜色								总计	χ^2	p
		浅绿	浅蓝	深绿	深蓝	紫	绿	蓝绿	黑			
表面	无风化	2	8	3	2	2	1	6	0	24		
风化	风化	1	12	4	0	2	0	9	2	30	6.287	0.507
	总计	3	20	7	2	4	1	15	2	54		

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

由表四可知， p 值大于 0.05，则文物的不同颜色对表面风化表现出一致性，即认为颜色对表面是否风化并无影响。

综上所述，我们认为古代玻璃文物的纹饰和颜色对文物表面风化情况并无影响，文物的玻璃类型对文物表面风化情况有显著影响。铅钡玻璃有更大的可能性表面风化，高钾玻璃有更大的可能性表面无风化。

二、肯德尔等级相关系数

根据题目所给的附件中表二的数据，在经过数据预处理之后，对高钾玻璃数据和铅钡玻璃数据，分别以不同的化学成分占比和表面风化数据作为 X 和 Y 值，计算肯德尔等级相关系数并进行显著性检验，最终得到数据如下：

表 5：铅钡玻璃和高钾玻璃的不同化学成分含量和表面风化的肯德尔等级相关系数

变量名	kendall 相关系数		p 值	
	铅钡	高钾	铅钡	高钾
二氧化硅(SiO ₂)	-0.666	0.685994	0.000	0.001
氧化钠(Na ₂ O)	-0.343	-0.306186	0.011	0.194
氧化钾(K ₂ O)	-0.189	-0.596595	0.143	0.004
氧化钙(CaO)	0.369	-0.458831	0.002	0.025
氧化镁(MgO)	-0.023	-0.521672	0.853	0.013
氧化铝(Al ₂ O ₃)	-0.269	-0.666939	0.024	0.001
氧化铁(Fe ₂ O ₃)	0.041	-0.458831	0.745	0.025
氧化铜(CuO)	0.232	-0.229416	0.052	0.261

氧化铅(PbO)	0.601	-0.5	0.000	0.025
氧化钡(BaO)	0.076	-0.359211	0.521	0.123
五氧化二磷(P2O5)	0.462	-0.506626	0.000	0.013
氧化锶(SrO)	0.261	-0.454859	0.031	0.045
氧化锡(SnO2)	-0.070	-0.171499	0.620	0.480
二氧化硫(SO2)	0.178	-0.306186	0.208	0.194

由上表可知，其中标黄和标橙的 p 值通过了显著性检验，说明对应类型下的文物的表面是否风化与对应化学成分的占比有关。标黄的 p 值大于 0.01 并小于 0.05，说明对应 X 与 Y 相关；而标橙的 p 值小于 0.01，说明对应 X 与 Y 显著相关。

在铅钡玻璃中，二氧化硅与表面风化呈显著较强负相关，二氧化硅含量提高会导致铅钡玻璃表面风化的可能性降低；氧化钠与表面风化呈弱负相关，氧化钠含量提高会导致铅钡玻璃表面风化的可能性降低；氧化钙与表面风化呈显著弱负相关，氧化钙含量提高会导致铅钡玻璃表面风化可能性升高；氧化铝与表面风化呈弱负相关，氧化铝含量提高会导致铅钡玻璃表面风化可能性升高；氧化铅与表面风化呈显著较强正相关，氧化铅含量提高会导致铅钡玻璃表面风化可能性升高；五氧化二磷与表面风化呈显著中等正相关，五氧化二磷含量提高会导致铅钡玻璃表面风化可能性升高；氧化锶与表面风化呈弱正相关，氧化锶含量提高会导致铅钡玻璃表面风化可能性升高。

在高钾玻璃中，二氧化硅与表面风化呈显著较强正相关，二氧化硅含量提高会导致高钾玻璃表面风化可能性升高；氧化钾与表面风化呈显著中等负相关，氧化钾含量提高会导致高钾玻璃表面风化可能性降低；氧化钙与表面风化呈中等负相关，氧化钙含量提高会导致高钾玻璃表面风化可能性降低；氧化镁与表面风化呈中等负相关，氧化镁含量提高会导致高钾玻璃表面风化可能性降低；氧化铝与表面风化呈显著较强负相关，氧化铝含量提高会导致高钾玻璃表面风化可能性降低；氧化铁与表面风化呈中等负相关，氧化铁含量提高会导致高钾玻璃表面风化可能性降低；氧化铅与表面风化呈中等负相关，氧化铅含量提高会导致高钾玻璃表面风化可能性降低；五氧化二磷与表面风化呈中等负相关，五氧化二磷含量提高会导致高钾玻璃表面风化可能性降低；氧化锶与表面风化呈中等负相关，氧化锶含量提高会导致高钾玻璃表面风化可能性降低。

三、 预测无风化值

首先，我们将表二中严重风化点的数据暂时剔除。重新取 p 的临界值为 0.02，由肯德尔相关系数以及显著性检验结果可知，铅钡玻璃表面风化与二氧化硅、氧化钠、氧化钙、氧化铅、五氧化二磷显著相关，而高钾玻璃表面风化与二氧化硅、氧化钾、氧化镁、氧化铝、五氧化二磷显著相关。因此，对铅钡玻璃的风化点数据还原，我们

将二氧化硅、氧化钠、氧化钙、氧化铅、五氧化二磷的含量分别乘以对应化学成分无风化与风化的平均值的比值；对高钾玻璃的风化点数据还原，我们将其二氧化硅、氧化钾、氧化镁、氧化铝、五氧化二磷的含量分别乘以对应化学成分无风化与风化的平均值的比值。

然后，将处理后的所有化学成分含量相加，得到累加的含量。将所有化学成分含量除以累加含量，使所有预测数据的化学成分含量累加总和为 100%。由此，我们得到了高钾玻璃和铅钡玻璃的风化点数据的未风化前预测值。下表为部分预测值，全部预测值见附录 1.3。

6.2.1 问题二模型的建立与求解

问题二模型的建立

问题二要求我们分析得出区分高钾玻璃和铅钡玻璃的主要化学成分及分类规律，并从每个类别中寻找合适的化学成分以此划分亚类，得出划分的方法及划分结果，并对模型进行合理性及敏感性分析。这就要求我们通过机器学习的模型进行训练，在根据所得的各个化学成分的权值系数从大到小选择特征，以此获得主要化学成分。再依据所得的化学成分进行聚类，得出亚类。合理性可由可视化证明，敏感性可由轮廓系数证明。

模型的准备

随机森林

随机森林 (*Random Forest*) 是包含多个决策树的分类器，并且其输出的类别是由个别树输出的类别的众数决定的。随机森林在一个数据集上的不同部分上训练着不同的决策树，平均了多个深决策树以降低变异数。

一、CART 分类决策树

CART 分类决策树用来处理分类问题，利用基尼指数进行多分，从而生成决策树并得到对应的预测模型和准确率。

因为题目中数据不存在连续值和缺失值，因此不进行连续值和缺失值处理

导入数据集和属性集

导入数据集并从中随机抽出一定数量的数据作为检测集 D_t ，其余作为训练数据集 D ，同时导入属性集 A

分类讨论情况

创建根节点 node:

- 1) 若 D 中样本均属于同一类别 C ，则将 node 标记为 C 类叶节点，返回结果。
- 2) 若 A 为空集或 D 中样本在 A 中取值相同，则将 node 标记为叶节点，叶节点的类标记为 D 中样本数最多的类的值，返回结果。
- 3) 其他情况下：计算基尼指数，并根据基尼指数得到最优划分属性 a_* ：

$$Gini_index(D, a) = \sum_{v=1}^V \frac{|D^v|}{|D|} Gini(D^v) \quad (2-1-1)$$

取使划分后基尼指数最小的属性为最优划分属性，即：

$$a_* = \underset{a \in A}{\operatorname{argmin}} Gini_index(D, a) \quad (2-1-2)$$

对 a_* 的每个值 a_*^v ，生成 node 分支， D^v 为 D 在 a_* 上取值为 a_*^v 的样本子集。

(i) 若 D^v 为空集，则将这一分支节点标记为叶节点，此叶节点的类标记为 D 中样本最多的类的值。

(ii) 其他情况下，以这一分支节点为根节点，继续进行判断和分类。

当最新一轮的分支节点都成为叶节点并返回结果之后，决策树生成完成。

最终，导入检测集 D_t 到此决策树中进行结果预测并计算对应准确率，并输出结果。

准确率的计算方法如下：

$$T_{rate}(a_1^{v_1}, a_2^{v_2}, \dots, a_n^{v_n}) = \frac{N_{correctly-predicted}}{N}, \quad (2-1-3)$$

$$N = \text{total number of test data with feature } (a_1^{v_1}, a_2^{v_2}, \dots, a_n^{v_n})$$

二、Bagging 算法

给定一个大小为 n 的训练集 D ，Bagging 算法从中均匀、有放回地选出 m 个大小为 n' 的子集 D_i ，作为新的训练集。随机森林模型对输入的训练集 X 和目标 Y ，运用 Bagging 算法得到多个训练集(取 B 次)，由此得到多个决策树。训练结束之后，随机森林模型可以通过对 X 上所有单个决策树的预测结果取平均值来得到最终预测结果：

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (2-1-4)$$

\hat{f} 为最终预测结果， $f_b(x')$ 为第 b 个决策树根据数据集 x' 得到的预测结果。

或者取所有决策树预测结果的众数作为最终预测结果。

Kmeans 聚类算法

Kmeans 算法是一种常用的聚类算法，基本思想为：在给定 K 值和 K 个初始类簇中心点的情况下，把每个数据点分到离其最近的类簇中心点代表的类簇中。所有点完成分配后，根据一个类簇中的所有点通过取平均值的方式重新计算类簇的中心点，并不断迭代地进行分配点和更新类簇中心点的步骤，直到中心点的变化很小或达到一定的迭代数。

基本原理:

给定数据集 X 和属性集 A , X 包含 n 个数据: $\{X_1, X_2, \dots, X_n\}$, A 包含 m 个属性 $\{A_1, A_2, \dots, A_m\}$ 。首先对 X 进行数据标准化处理:

$$X'_{A_i} = \frac{X_{A_i} - \mu}{\sigma} \quad (2-2-1)$$

其中 X'_{A_i} 为 X 的第 i 个属性值的数据标准化后的数据, X_{A_i} 为 X 的第 t 个属性值的数据标准化前的数据, μ 是 X_{A_i} 的平均值, σ 是 X_{A_i} 的标准差。 X' 是所有属性值数据都经过标准化之后的数据集, 取 X' 作为新的数据集。

然后, 初始化 K 个类簇中心点 $\{C_1, C_2, \dots, C_k\}$, $1 < k \leq n$, 然后通过计算每一个对象到每一个类簇中心的欧氏距离:

$$dis(X'_i, C_j) = \sqrt{\sum_{t=1}^m (X'_{it} - C_{jt})^2} \quad (2-2-2)$$

X'_i 表示第 i 个数据, C_j 表示第 j 个类簇中心点, X'_{it} 表示第 i 个数据的第 t 个属性值, C_{jt} 表示第 j 个类簇中心点的第 t 个属性值。

依次比较数据到每个中心点的距离, 将数据分到距离最近的聚类中心的类簇中, 得到 k 个类簇。然后, 通过计算平均值得到新的类簇中心:

$$C_y = \frac{\sum_{X'_i \in S_y} X'_i}{|S_y|} \quad (2-2-3)$$

C_y 表示第 y 个聚类中心, X'_i 表示属于 S_y 的第 i 个数据, $|S_y|$ 表示第 y 个聚类中数据的总数。

之后重复上述的分配数据到类簇和计算类簇中心的过程, 直到达到预先确定好的最大迭代次数 max_iter , 或者在计算新的聚类中心时, 所有聚类中心的变化量都小于预先确定好的阈值 tol 。然后, 停止迭代, 输出最终聚类中心结果, 得到训练完毕的 Kmeans 模型。

在未知最优 K 值的情况下, 可以通过轮廓系数来获得最优的 K 值选择。轮廓系数的计算公式如下:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2-2-4)$$

$$S = \frac{1}{|X'|} \sum_i S(i)$$

$a(i)$ 为簇内不相似度, 是 i 向量到同簇内其他点不相似程度的平均值, 体现凝聚度;

$b(i)$ 为簇间不相似度, 是 i 向量到其他簇的平均不相似程度的最小值, 体现分离度;

$S(j, i)$ 是第 i 个对象的轮廓系数, S 是所有样本的 $S(i)$ 的均值。

S 取值范围为 $[-1, 1]$, 值越大, 说明聚类效果越好。

模型的建立

一、随机森林

根据题目所给的表二的数据，我们先删去了不符合比例累加总数的 15、17 号文物数据，并暂时剔除了三个严重分化点数据。然后利用 Bagging 算法，从中抽取一定固定量的数据集和属性集，共抽取 B 次，由此训练出 B 个决策树。我们预先赋予数据类，并训练决策树判断玻璃类型、颜色和纹饰的能力。

依靠随机森林模型，我们还可以得到具有特征重要性的属性集，并且对他们的特征重要性进行降序排列，从而得到其中较为重要的几种化学成分，从而分别得到划分玻璃类型、颜色和纹饰的最优特征子集，用来建立后续的 Kmeans 聚类模型。

我们画出随机森林模型结构图如下

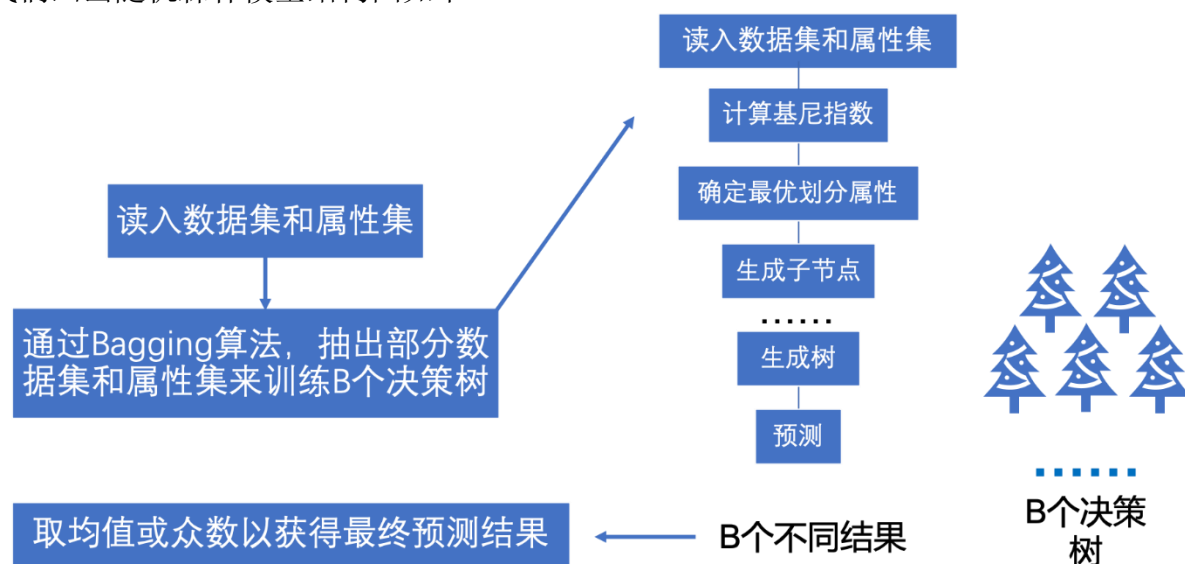


图 3：随机森林模型结构图

二、Kmeans 聚类模型

在通过随机森林得到划分玻璃类型、颜色、纹饰的最优特征子集之后，以最优特征子集中的多种属性作为维度，在这一新的空间中构建 Kmeans 聚类模型。我们画出 Kmeans 聚类模型结构图如下：

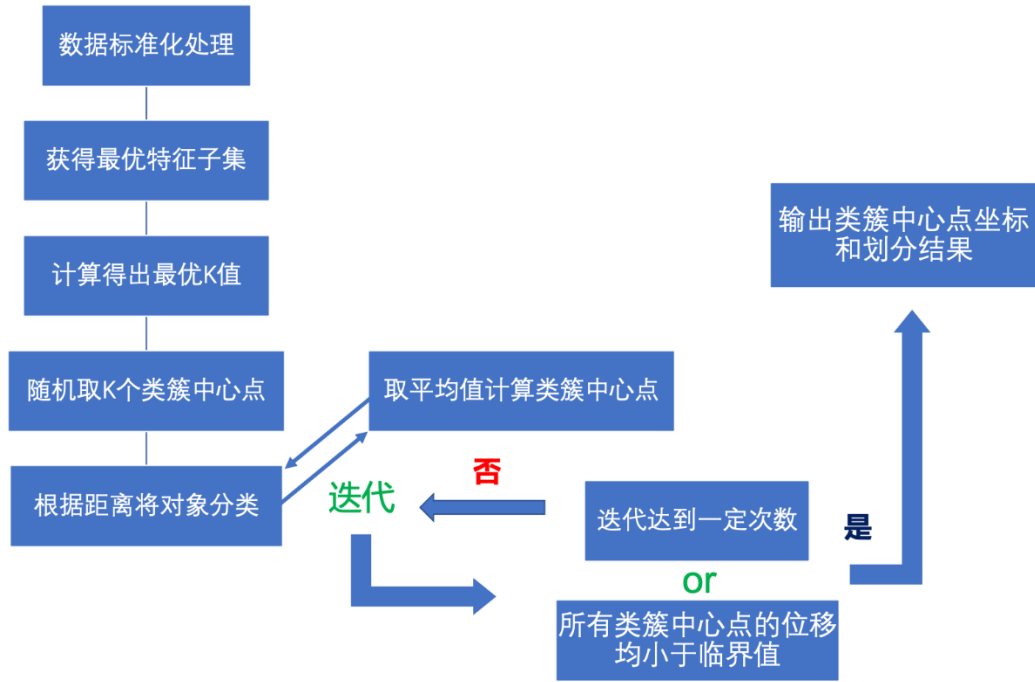


图 4: Kmeans 聚类模型结构图

由上图可知最终亚类的中心点是由 Kmeans 聚类模型产生的，同时所有的数据也可以通过计算到各个中心点的距离来放到各个亚类之中。求解过程以及结果在下面模型求解中给出。至此，我们建立出 Keams 聚类模型如下：

$$dis(X'_i, C_j) = \sqrt{\sum_{t=1}^m (X'_{it} - C_{jt})^2} \quad (2-2-5)$$

X'_i 表示第*i*个数据， C_j 表示第*j*个类簇中心点， X'_{it} 表示第*i*个数据的第*t*个属性值， C_{jt} 表示第*j*个类簇中心点的第*t*个属性值。

$$C_y = \frac{\sum_{X'_i \in S_y} X'_i}{|S_y|} \quad (2-2-6)$$

C_y 表示第*y*个聚类中心， X'_i 表示属于 S_y 的第*i*个数据， $|S_y|$ 表示第*y*个聚类中数据的总数。

最终结束迭代得到聚类中心之后，我们可以继续根据公式（）计算点到中心点距离，并完成分类，由此得到各个亚类。

6.2.2 问题二模型求解

一、玻璃类型分类-随机森林模型

首先为了分析高钾玻璃和铅钡玻璃的分类规律，我们构建用于预测玻璃类型分类的随机森林模型。

根据题目所附带的附件中表一和表二的数据，在经过数据预处理之后，我们通过 Python 调取 GridSearchCV 和 RandomForestClassifier 包，将一组可选择参数给到随机森林，随机森林取最佳参数为：

表 6：玻璃类型分类-随机森林模型的参数选择

参数	值
criterion	entropy
max depth	5
n estimators	11
max features	0.3
min sample split	4

之后利用数据训练出了一个随机森林模型，并得到了具有特征重要性的属性子集和对应的特征重要性，取 0.1 作为特征重要性的临界值，我们取其中特征重要性高于 0.1 的属性为最优特征子集。

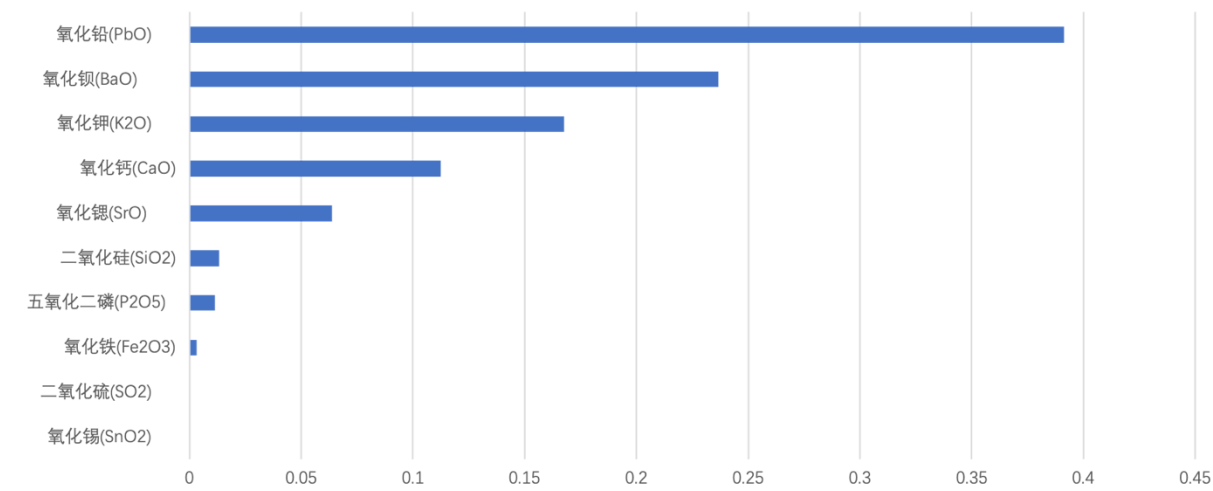


图 5：玻璃类型分类-随机森林特征重要性

如上图，我们取 $A_1 = \{\text{氧化铅(PbO)}, \text{氧化钡(BaO)}, \text{氧化钾(K2O)}, \text{氧化钙(CaO)}\}$ 作为最优特征子集，后续用于高钾玻璃和铅钡玻璃的亚类划分。

通过对随机森林中 11 棵决策树的分类临界值取平均值，我们得到了表 7 作为铅钡玻璃和高钾玻璃的分类规律。经过测试集的检验，本随机森林预测模型的准确率高达 100%。

图 6 图 7 为随机森林中的两棵决策树，玻璃类型分类-随机森林模型中的全部 11 棵决策树可见于附录中。

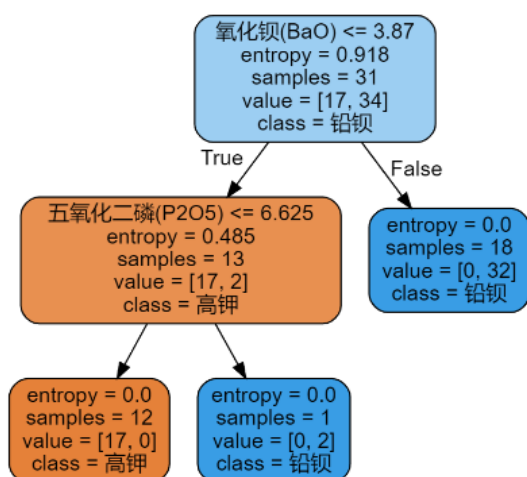


图 6：玻璃类型分类-随机森林 决策树 1

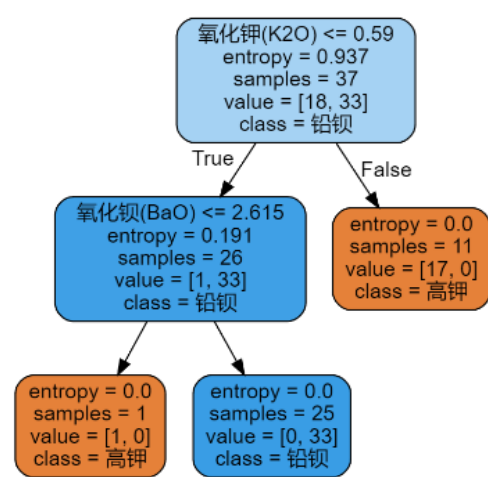


图 7：玻璃类型分类-随机森林 决策树 2

表 7：玻璃类型分类临界值

化学成分	临界值(铅钡)
氧化铅(PbO)	≤ 5.4425
氧化钡(BaO)	≤ 2.83375
氧化钾(K ₂ O)	≤ 0.725
氧化钙(CaO)	> 2.3275
氧化锶(SrO)	≤ 0.1525
二氧化硅(SiO ₂)	≤ 55.57
五氧化二磷(P ₂ O ₅)	> 6.625

由表 7 所示，若氧化铅 ≤ 5.4425 且氧化钡 ≤ 2.83375 且氧化钾 ≤ 0.725 且氧化钙 > 2.3275 且氧化锶 ≤ 0.1525 且二氧化硅 ≤ 55.57 且五氧化二磷 > 6.625 时，认为该类玻璃为铅钡玻璃，否则认为该类玻璃为高钾玻璃。

二、Kmeans

根据随机森林得到的最优特征子集，我们构建以氧化铅、氧化钡、氧化钾、氧化钙的含量为四个维度的空间。

对高钾玻璃数据，我们以轮廓系数为依据，得到最优 K 值为 7。高钾 K 值-轮廓系数折线图如下：

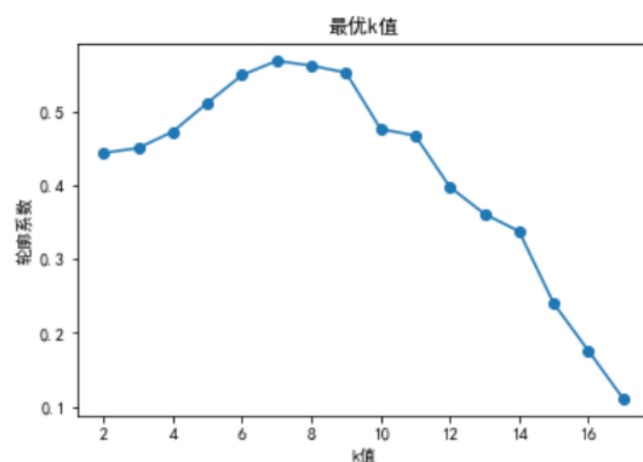


图 8：高钾玻璃亚类划分不同 k 值寻优

由上图可得知，K=7 时，轮廓系数取最大值，7 为最优 K 值。

根据最优 K 值，利用附件中表二的玻璃类型为高钾玻璃的数据，我们在以氧化铅、氧化钡、氧化钾、氧化钙的含量为四个维度的空间上构建 Kmeans 聚类模型。并最终得到 7 个亚类，7 个类簇中心点如下表：

表 8：高钾亚类不同类别中心值

类别	中心点
0	[0.0357, 0.0000, 0.0000, 1.2071]
1	[0.0367, 0.0000, 0.0567, 11.6300]
2	[1.4100, 2.8600, 0.1000, 12.3700]
3	[1.6200, 0.0000, 0.0000, 12.2800]
4	[0.2750, 1.1750, 0.1150, 7.5250]
5	[1.0000, 1.9700, 0.0000, 0.0000]
6	[-0.5493, -0.4877, -0.6516, -0.8391]

并且得到 7 个亚类划分结果如下表：

表 9：高钾亚类划分结果

类别	划分结果
0	03 部位 1, 10, 27, 22, 12, 07, 09
1	05, 16, 18
2	03 部位 2
3	14
4	06 部位 1, 06 部位 2
5	21

将所有高钾亚类数据绘制为安德鲁斯曲线如下图：

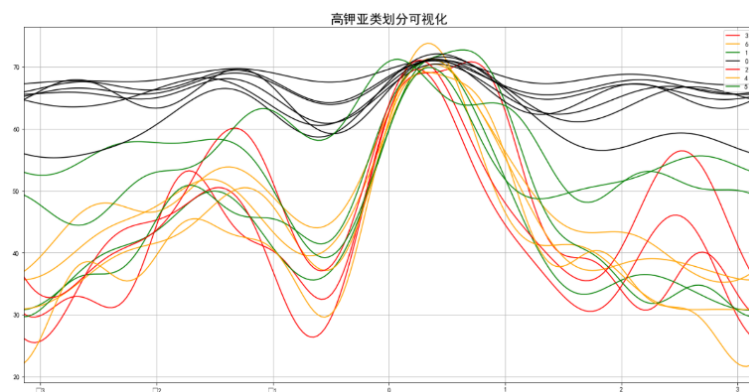


图 9：高钾亚类划分合理性

由上图可以看出，相同颜色曲线之间的聚集程度较高，分类效果较好，此方法得出的高钾玻璃亚类合理性较强

同时，由高钾 K 值-轮廓系数折线图可以看出，在 K 小于 7 时，曲线较为平缓，轮廓系数变化较小，敏感度低；当 K 大于 7 时，轮廓系数在 K 达到 9 之后开始迅速下跌，聚类效果快速劣化。总体而言，在我们所取的最优 K 值附近，K 的敏感度较低。

对铅钡玻璃数据，我们以轮廓系数为依据，得到最优 K 值为 4。铅钡 K 值-轮廓系数折线图如下：

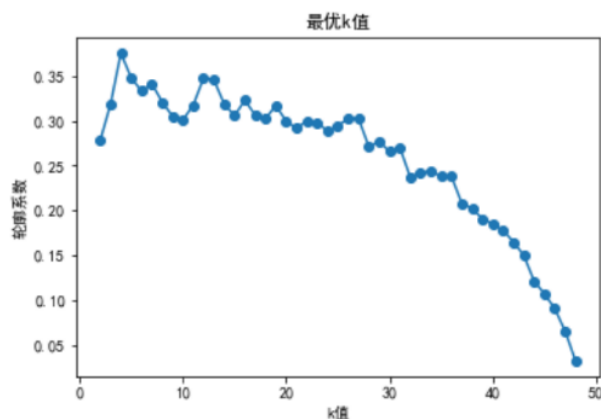


图 10：铅钡亚类划分不同 k 值寻优

由上图可得知，K=4 时，轮廓系数取最大值，4 为最优 K 值。

根据最优 K 值，利用附件中表二的玻璃类型为铅钡玻璃的数据，我们在以氧化铅、氧化钡、氧化钾、氧化钙的含量为四个维度的空间上构建 Kmeans 聚类模型。并最终得到 4 个亚类，4 个类簇中心点如下表：

表 10：高钾亚类不同类别中心值

类	中心点
0	[26.5285, 7.9963, 0.1878, 0.1437]
1	[26.5033, 29.8883, 0.4800, 0.1850]
2	[48.0121, 7.7507, 0.6114, 0.0750]
3	[43.3250, 5.1450, 0.2700, 1.2300]

并且得到 4 个亚类划分结果如下表

表 11：铅钡亚类划分结果

	划分结果
0	11, 19, 23 未风化点, 25 未风化点, 26, 28 未风化点, 29 未风化点, 31, 32, 33, 34, 35, 36, 37, 42 未风化点 1, 42 未风化点 2, 44 未风化点, 45, 46, 48, 49 未风化点, 50 未风化点, 51 部位 2, 53 未风化点, 55, 56, 57, 58
1	08, 08 严重风化点, 20, 24, 26, 26 严重风化点
2	30 部位 2, 38, 39, 40, 41, 43 部位 1, 43 部位 2, 47, 49, 50, 51 部位 1, 52, 54, 54 严重风化点
3	02, 30 部位 1

将所有铅钡亚类数据绘制为安德鲁斯曲线如下图：

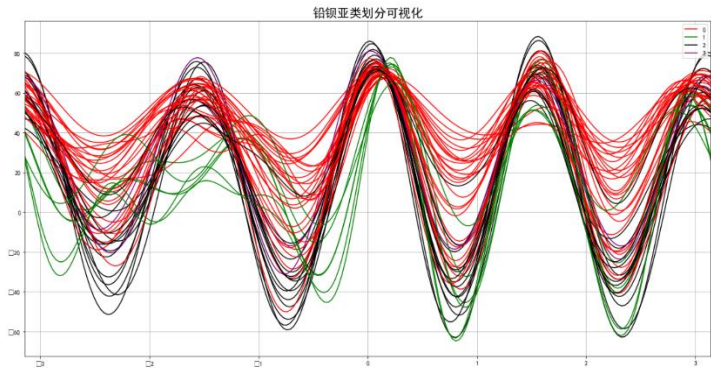


图 11：铅钡亚类划分合理性

由上图可以看出，黑色、紫色曲线之间的聚集程度较高，分类效果较好；而红色和绿色曲线较为分散，分类效果较弱。总体而言，此方法得出的铅钡玻璃亚类合理性一般。同时，由铅钡 K 值-轮廓系数折线图可以看出，在 K 小于 4 时，曲线较为陡峭，轮廓系数变化较快，敏感度高；当 K 大于 7 时，轮廓系数先迅速下跌，然后快速抖动，敏感度高。总体而言，在我们所取的最优 K 值附近，K 的敏感度较高。

综上所述，根据 Kmeans 模型得到的高钾玻璃亚类合理性较高，敏感度较低；而铅钡玻璃亚类合理性一般，敏感度较高。

6.3.1 问题三模型的建立和求解

问题三要求我们预测表三数据的类型。我们利用问题二中已经训练好的随机森林模型，对表三的数据进行分类预测。

模型准备

随机森林，见问题二的模型准备

问题三的模型建立

问题三的分类预测使用问题二训练完毕的玻璃类型分类-随机森林模型，见问题二的模型建立。

6.3.2 问题三的模型求解

根据附件中表三数据，将化学成分含量数据输入到问题二训练完毕的玻璃类型分类-随机森林模型中，得到相应的分类预测结果如下表：

表 12：表三预测结果

文物编号：	A1	A2	A3	A4	A5	A6	A7	A8
预 测 玻 璃 类 型：	高钾	铅钡	铅钡	铅钡	铅钡	高钾	高钾	铅钡

在使用 GridSearchCV 调参的过程中我们通过表 13 的参数几何进行模型比对，并寻找最优的模型，以此实现了敏感度分析。

6.4.1 问题四模型的建立与求解

问题四要求我们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。我们可以利用“Kendall's tau-b”计算不同类别内，其化学成分之间的相关系数。若 p 值大于 0.05，且相关系数绝对值大于 0.8 为极强相关，位于 0.6-0.8 为强相关性，位于 0.4-0.6 之间为中等强度相关性，0.4 以下为弱或无相关性。接着比较不同类别之间的相关系数，以 0.7 为界限值，判断其相关性。若类别间相关系数差值的绝对值大于 0.7，则认为不同类别间该化学成分有较大差异性；若 0 位于 0.4-0.7，则认为不同类别间该化学成分存在差异性；否则认为不同类别间该化学成分无差异性。

模型准备

肯德尔相关系数，见问题一的模型准备

模型建立

利用题目所给的表二的数据，首先根据玻璃类型将其分为高钾玻璃数据和铅钡玻璃数据。然后，对分别对不同玻璃类型的数据中的化学成分进行两两组合得到相关变量，计算肯德尔相关系数并进行显著性检验。画出肯德尔相关系数模型结构如下

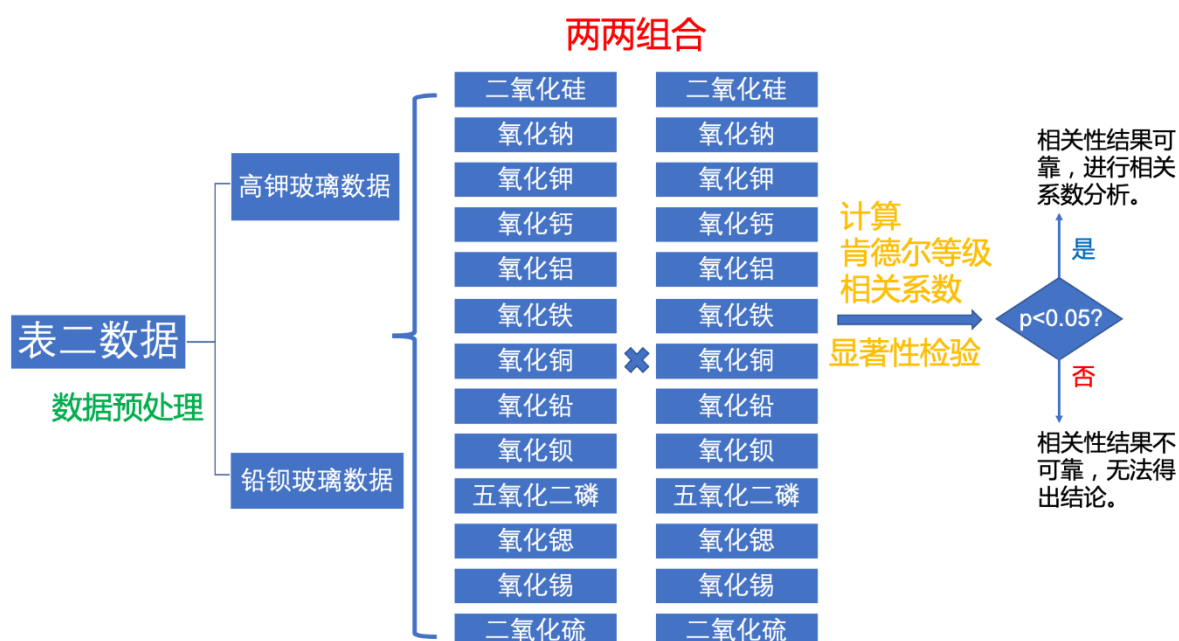


图 12：不同玻璃类型不同化学成分之间的肯德尔相关系数

同时，根据纹饰将表二数据分成纹饰 A 数据、纹饰 B 数据和纹饰 C 数据。然后，分别对不同纹饰的数据中的化学成分进行两两组合得到相关变量，计算肯德尔相关系数并进行显著性检验

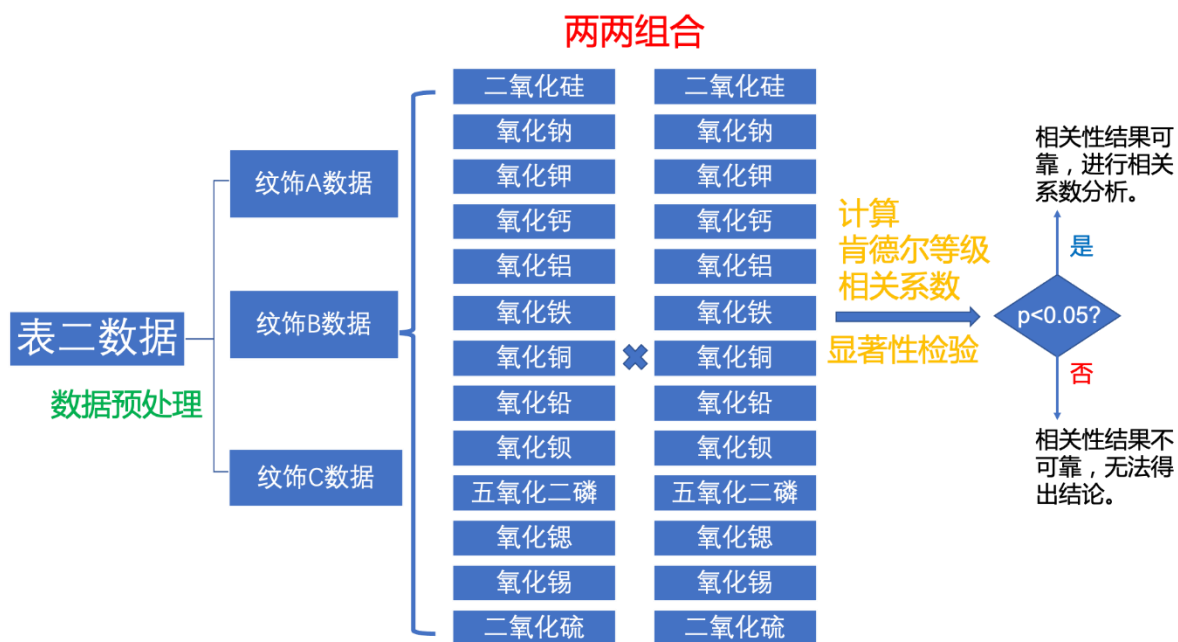


图 13：不同纹饰类型下化学成分之间的肯德尔等级相关系数模型

从上两图可知，不同类型的古代玻璃文物表面有无风化的化学成分含量统计规律可以通过分别计算不同类型下文物表面风化数据和不同化学成分之间的肯德尔等级相关系数来得到。由此构建肯德尔等级相关模型如下：

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \quad (4-1-1)$$

若最终计算得出 p 值小于 0.05，则认为该组数据相关，之后根据相关系数来判断不同类型玻璃表面有无风化与不同化学成分之间的相关性强弱；否则则认为该组数据不相关，无法通过相关系数得出任何统计规律。相关系数的意义如表 1 所示。

同时，对于在不同玻璃类型之中，若相同组合的某两种化学成分之间的肯德尔相关系数均通过了显著性检验，且相关系数差值大于 0.7，则认为不同玻璃类型间，这两种化学成分的关联关系具有差异性。对于不同纹饰亦然。

6.4.2 问题四的模型求解

根据附录中的表二数据，我们根据玻璃类型将其分为高钾玻璃和铅钡玻璃，并且分别对不同玻璃类型中的化学成分两两求肯德尔相关系数并进行显著性检验。高钾玻璃的相应数据可见于下图，带显著性 p 值的数据见附录 4.1.1。

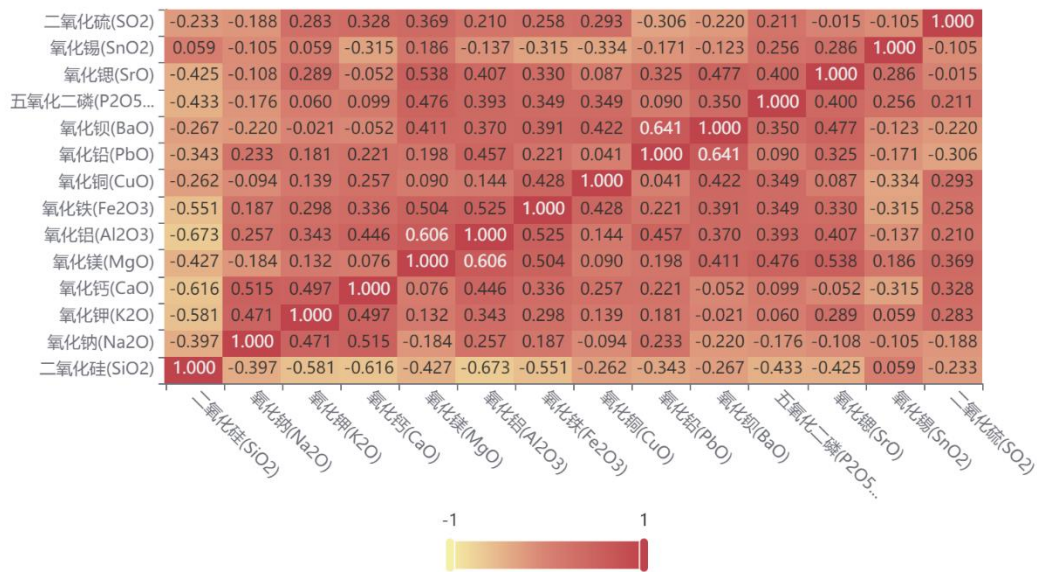


图 14：高钾玻璃不同化学成分的肯德尔相关系数热力图

结合上图和附录 4.1.1，我们可以看到通过了显著性检验的不同化学成分的相关系数，从而分析出对应的关联关系。如：氧化铝氧化镁，强正相关；氧化钙二氧化硅，强负相关；氧化钡氧化铅，强正相关；氧化钾二氧化硅，中等负相关；氧化锶氧化镁；中等正相关；氧化铁氧化铝，中等正相关；等等。以上列举的相关性均通过显著性检验。

铅钡玻璃的相应数据可见于下图，带显著性 p 值的数据见附录 4.1.2。

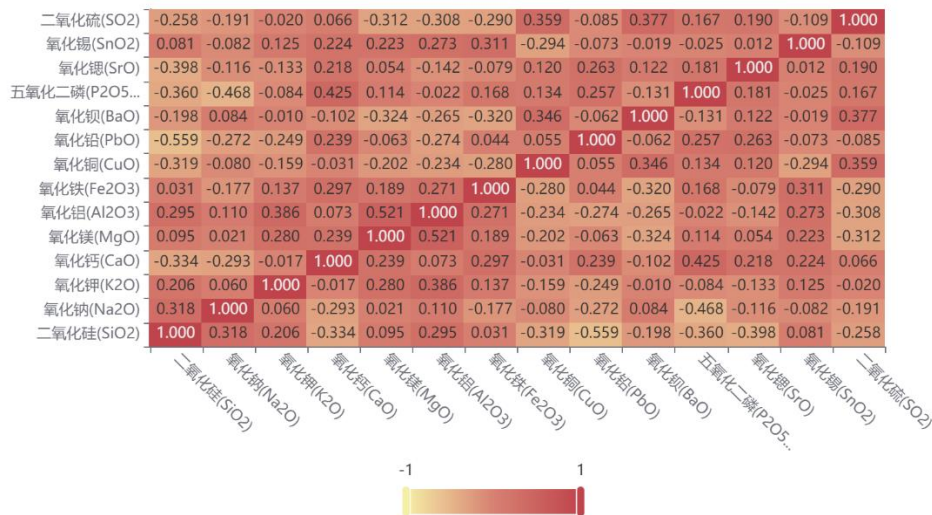


图 15：铅钡玻璃不同化学成分的肯德尔相关系数热力图

结合上图和附录 4.1.1，我们可以看到通过了显著性检验的不同化学成分的相关系数，从而分析出对应的关联关系。如：氧化铝氧化镁，强正相关；氧化钙二氧化硅，强负相关；氧化钡氧化铅，强正相关；氧化钾二氧化硅，中等负相关；氧化锶氧化

镁；中等正相关；氧化铁氧化铝，中等正相关；等等。以上列举的相关性均通过显著性检验。

仅比较图 14 于图 15，我们发现，高钾玻璃中化学成分之间的相关性多为中等相关和强相关，铅钡玻璃中化学成分之间的相关性多为弱相关和中等相关。在不同玻璃类型之中，若相同组合的某两种化学成分的肯德尔相关系数均通过了显著性检验，且相关系数差值大于 0.7，则认为不同玻璃类型间，这两种化学成分的关联关系具有差异性。根据计算得出的数据，依照这一标准进行筛选，我们得到具有差异性的化学成分关联关系如下表和下图

表 14：不同玻璃类型中具有差异性的化学成分关联关系的数值和差值

AB	铅钡	高钾	差值
硅钠	0.318	-0.397	0.715
硅钾	0.206	-0.581	0.787
硅铝	0.295	-0.673	0.968
钠钙	-0.293	0.515	0.808
镁钡	-0.324	0.411	0.735
铝铅	-0.274	0.457	0.731
铁铜	-0.28	0.428	0.708
铁钡	-0.32	0.391	0.711
铅钡	-0.062	0.641	0.703

上表中，A、B 均为化学成分，AB 指这两种化学成分之间的关联关系，“硅”代表二氧化硅，“钠”代表氧化钠，“钾”代表氧化钾，“铝”代表氧化铝，“镁”代表氧化镁，“铅”代表氧化铅，“铁”代表氧化铁，“铜”代表氧化铜。

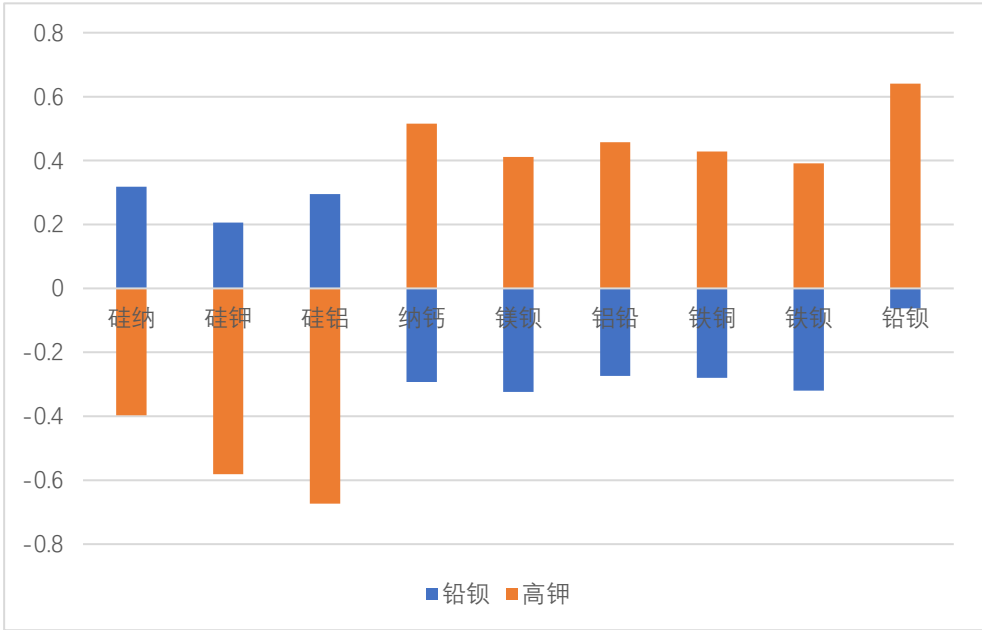


图 16：不同玻璃类型中差异性大的不同化学成分关联关系的肯德尔相关系数

如上图所示，上述相关系数均通过显著性检验，且相关系数差值大，正负值均不同，因此判定以上化学成分关联关系在不同玻璃类型之间差异性大。

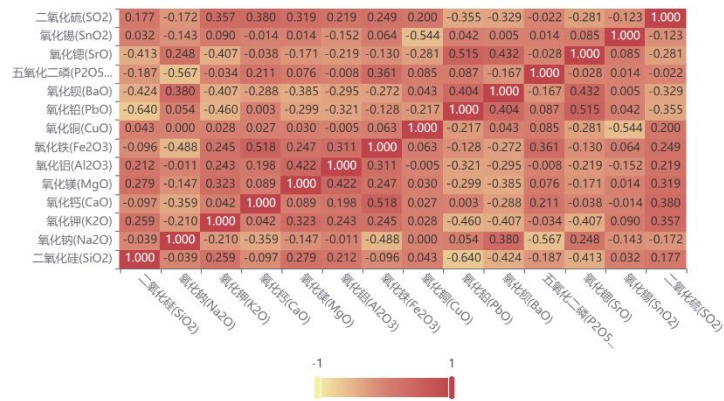


图 17: A 纹饰不同化学成分的肯德尔相关系数热力图

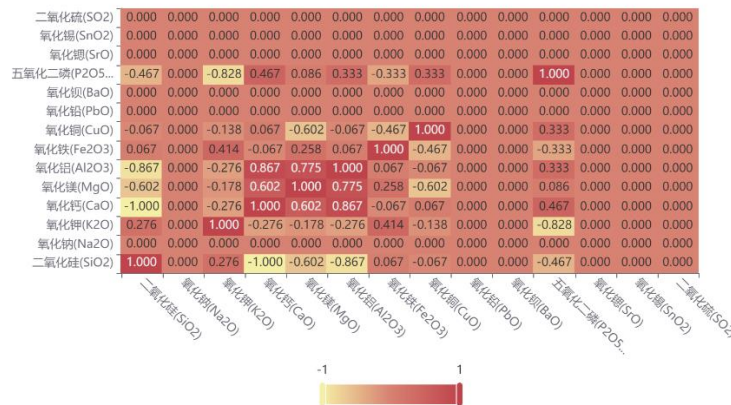


图 18: B 纹饰不同化学成分的肯德尔相关系数热力图

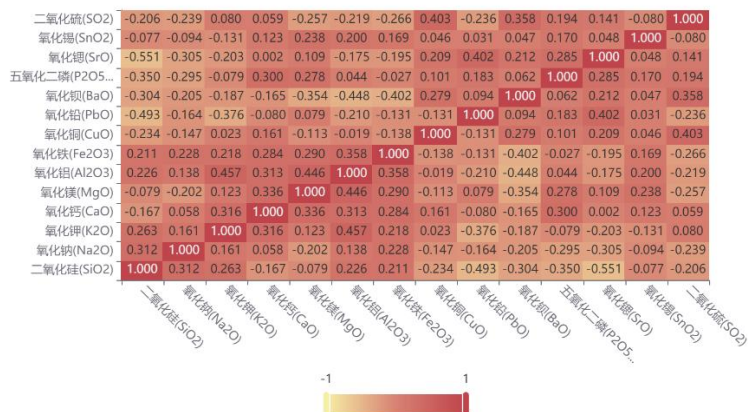


图 19: C 纹饰不同化学成分的肯德尔相关系数热力图

由图 17、18、19 可知，纹饰类别 A、B、C 相关系数多为弱相关及无相关，相关性弱。由附录 4.2.1 可得，通过显著性检验的相关系数均小于 0.1，为弱或无相关，因此判断类型 A、B、C 纹饰化学成分无显著相关性，且 A、B、C 类别之间并无显著差异性。

7. 模型的评价与推广

7.1 模型的优点

1. 我们对于化学成分与类型所搭建的随机森林模型准确率达到了 100%，能够完美预测测试集和训练集，并且能够提供各个化学成分的判定阈值。
2. 安德鲁斯曲线反映出了亚类划分模型具备相当的合理性。

7.2 模型的缺点

1. 搭建的亚类划分模型只考虑了模型 k 值的最优，而未考虑其他因素，例如样本大小的影响。
2. 未针对样本大小进行扰动处理，样本过小导致部分预测不准确。
3. 在预测风化前的化学成分时使用了平均值进行按比例预测，但平均值只能体现部分样本的分布趋势，而并非能使得所有样本被准确预测。

7.3 模型的推广

1. 对不同类型的亚类划分可以推广到文物种类研究的领域，从亚类中推测导致风化的因素。
2. 通过搭建化学成分与类别的模型可以实现对风化文物类型的预测。

8. 结论

针对古代玻璃的成分分析与鉴定问题，本文建立了卡方检验、肯德尔等级相关系数等模型，探究了不同外表性质之间的关联以及内在化学成分与外表性质的关联，并且通过训练随机森林模型和 Kmeans 聚类模型，实现了仅频化学成分分析对玻璃类型的预测，以及对不同类别根据化学成分之间的差异进行的亚类划分。最终，本文分析了不同类别（纹饰、玻璃类型）之下，化学成分之间的关联，以及不同类别（如不同纹饰、不同玻璃类型）之间化学成分的关联的差异。本文中得到的不同内外性质之间的关联与差别有助于未来在古代玻璃方面的研究；同时，本文中训练完成的随机森林模

型有助于特殊情况下对古代玻璃文物的鉴定、保护与修复；本文中训练完成的 Kmeans 聚类模型也有助于未来对古代玻璃文物的亚类细分以及更加专业细化的研究。

9. 参考文献

- [1] 肖志勇, 卡方检验, https://baike.baidu.com/item/卡方检验/2591853?fromModule=lemma_inlink, 2022/9/16
- [2] Wikipedia, 肯德尔等级相关系数, https://en-m-wikipedia-org.translate.goog/wiki/Kendall_rank_correlation_coefficient?_x_tr_sl=en&_x_tr_tl=zh-CN&_x_tr_hl=zh-CN&_x_tr_pto=sc, 2022/9/16
- [3] Wikipedia, 随机森林, <https://zh.wikipedia.org/wiki/随机森林>, 2022/9/16
- [4] 打不死的小强, 随机森林-Random Forest, <https://easyai.tech/ai-definition/random-forest/>, 2022/9/16
- [5] 千鹤, 规则分类——决策树算法, <https://aistudio.baidu.com/aistudio/projectdetail/619590?channelType=0&channel=0&hmsr=aladdin>, 2022/9/16
- [6] 思绪无限, Kmeans 算法详解, https://blog.csdn.net/qg_32892383/article/details/80107795, 2022/9/16

10. 附录

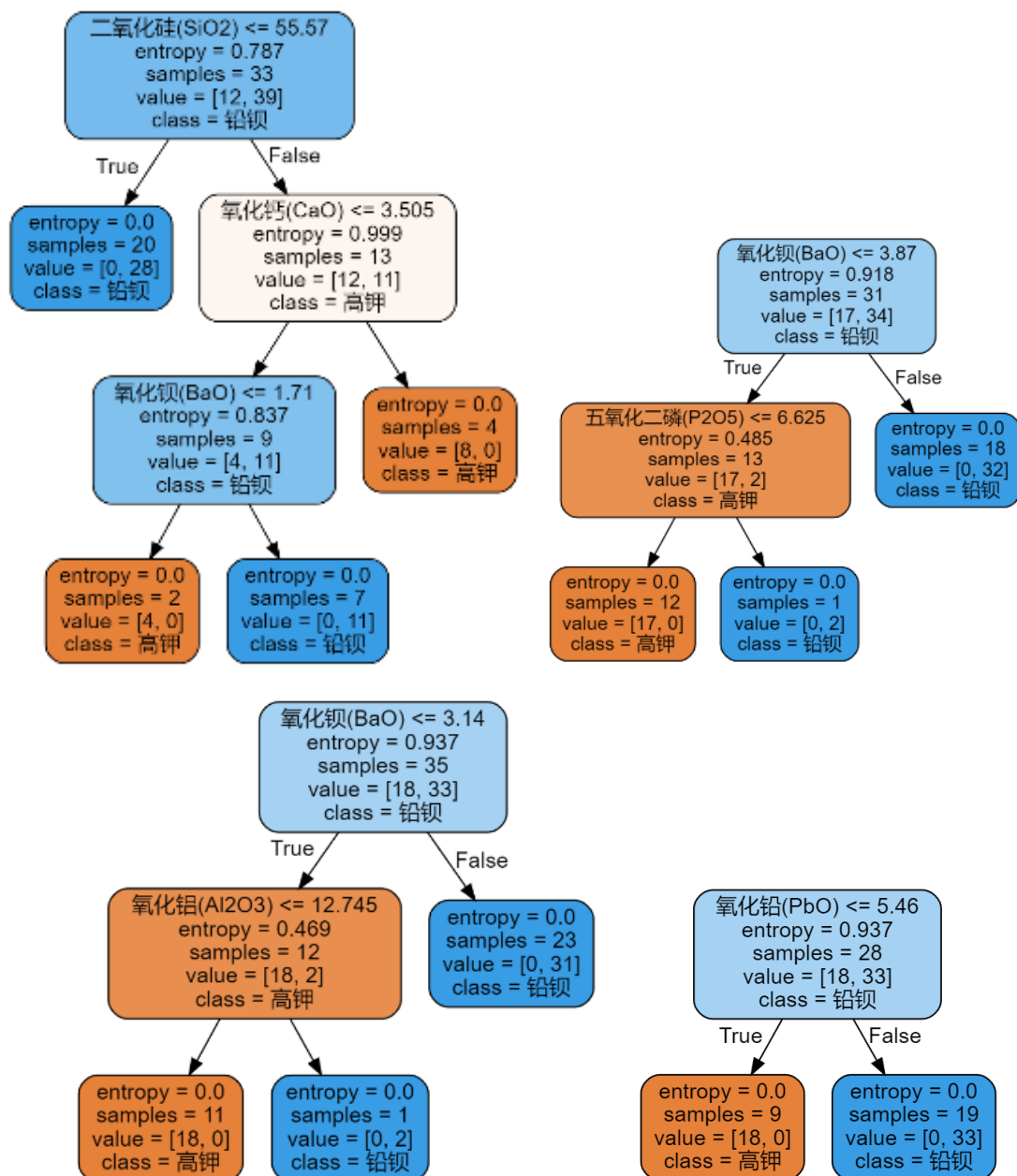
附录 1.1 使用 SPSSAU 进行卡方检验的步骤为：将数据预处理后的表一上传至 SPSSAU，点击左侧“实验/医学研究”下属第一个方法“卡方检验”，然后分三次将“纹饰”、“类型”、“颜色”放置到“X（定类）”中，将“表面风化”放置到“Y（定类）”中，先后生成三个卡方检验分析结果。

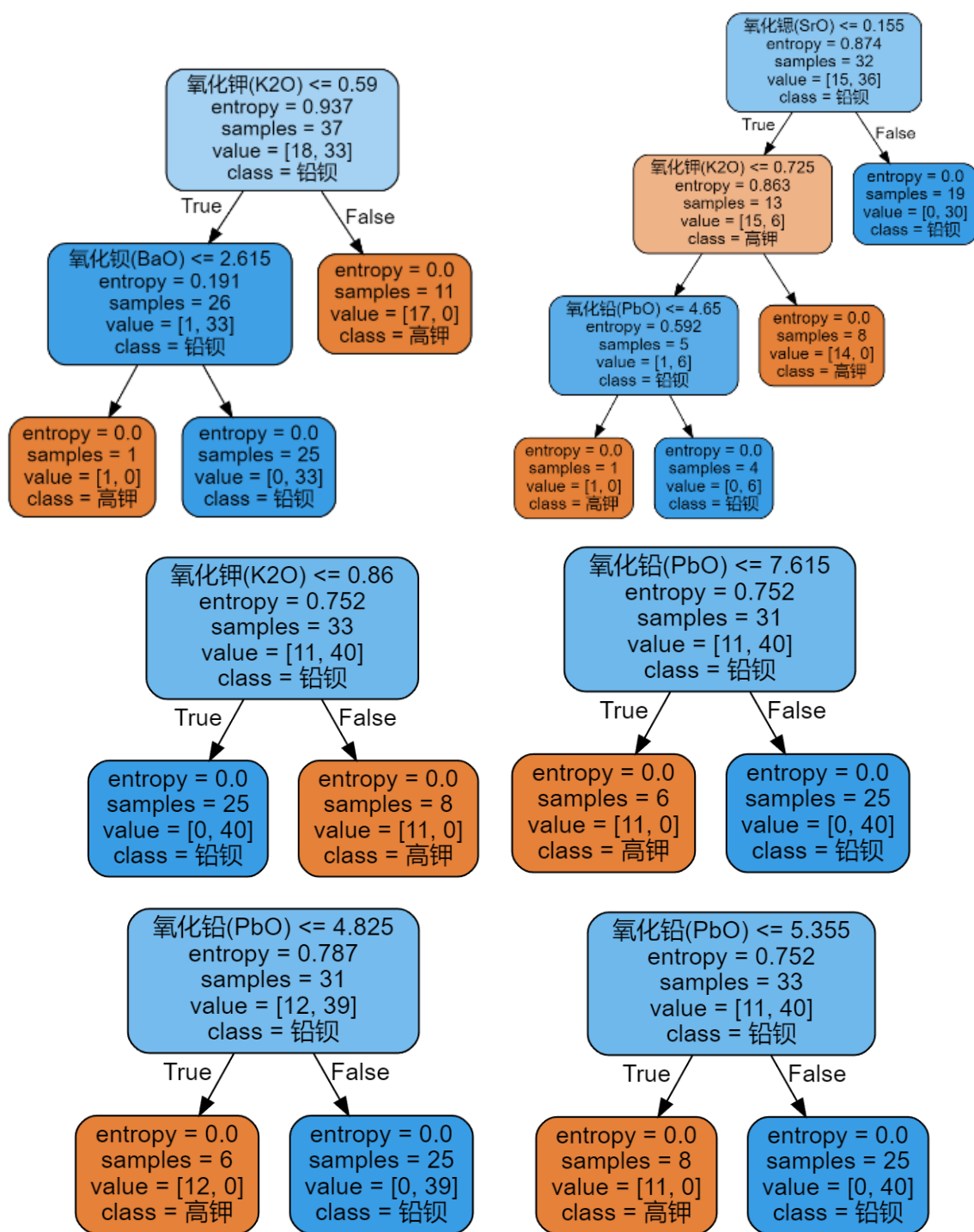
附录 1.2 肯德尔等级相关系数的 python 代码见于支撑材料中“Kendall’s tau-b 等级相关系数”

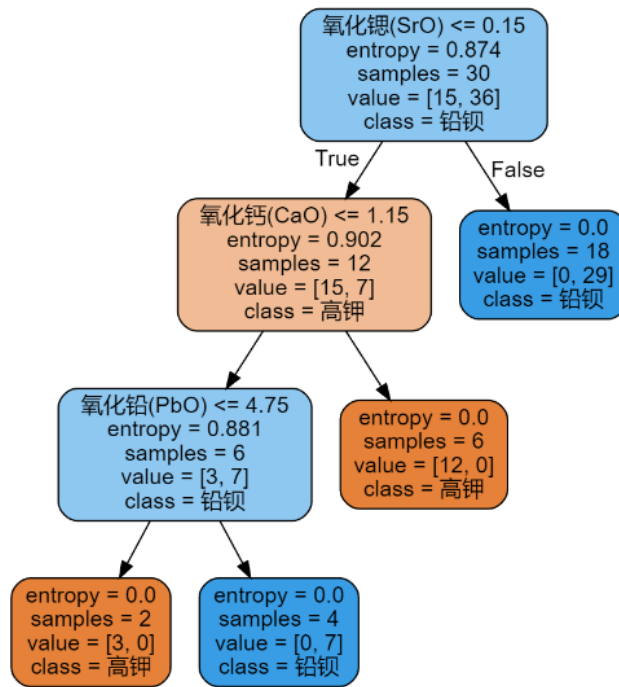
附录 1.3

检测点	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)
2	65.10	0.00	1.05	0.99	1.18	7.32	1.86	0.26	21.28	0.00	0.70	0.19	0.00	0.00
8	43.06	0.00	0.00	0.74	0.00	2.04	0.00	10.41	15.34	31.23	0.84	0.37	0.00	2.58
08严重风化点	44.02	0.00	0.00	3.02	0.00	3.70	0.00	3.14	26.36	30.62	1.27	0.53	0.00	15.03
11	64.45	0.00	0.21	1.58	0.71	3.67	0.00	4.93	12.18	14.61	1.96	0.37	0.00	0.00
19	60.47	0.00	0.00	1.41	0.59	5.18	1.33	3.51	21.85	5.35	1.97	0.19	0.00	0.00
7	74.67	0.00	0.00	1.07	0.00	7.57	0.17	3.24	0.00	0.00	3.40	0.00	0.00	0.00
9	73.35	0.00	10.81	0.62	0.00	4.83	0.32	1.55	0.00	0.00	1.87	0.00	0.00	0.00
26	42.66	0.00	0.00	0.73	0.00	1.07	0.00	10.57	15.92	32.25	0.74	0.45	0.00	1.96
10	75.21	0.00	16.97	0.21	0.00	2.98	0.26	0.84	0.00	0.00	0.00	0.00	0.00	0.00
26严重风化点	37.96	0.00	0.40	3.04	0.00	4.20	0.00	3.60	25.97	35.45	1.09	0.62	0.00	15.95
12	67.92	0.00	17.27	0.72	0.00	4.99	0.29	1.65	0.00	0.00	0.75	0.00	0.00	0.00
22	60.92	0.00	11.59	1.66	0.64	10.95	0.35	0.55	0.00	0.00	0.96	0.00	0.00	0.00
27	74.46	0.00	0.00	0.94	0.54	9.56	0.20	1.54	0.00	0.00	2.00	0.00	0.00	0.00
34	66.17	0.00	0.25	0.34	0.00	2.13	0.47	1.51	21.53	10.00	0.07	0.22	0.00	0.00
36	61.86	11.83	0.14	0.14	0.00	1.78	0.32	0.68	16.27	10.83	0.01	0.22	0.00	0.00
38	58.07	8.29	0.00	0.28	0.00	3.23	0.29	0.73	21.75	9.79	0.09	0.41	0.00	0.00
39	57.28	0.00	0.00	0.57	0.00	0.78	0.00	0.88	33.31	7.22	0.28	0.61	0.00	0.00
40	43.50	0.00	0.00	1.15	0.00	0.83	0.19	0.00	45.71	6.69	0.50	0.68	0.00	0.00
41	45.21	0.00	0.44	2.86	2.73	5.80	1.79	0.19	27.03	9.76	1.99	0.47	0.00	0.00
43部位1	34.40	0.00	0.00	3.42	0.89	4.44	0.76	5.35	41.49	7.29	0.00	0.64	0.00	0.00
43部位2	52.33	0.00	0.00	3.63	0.95	5.85	1.39	1.51	26.99	3.26	3.38	0.47	0.00	0.00
48	70.53	3.61	0.32	0.88	1.54	12.85	1.03	0.00	5.20	7.31	0.16	0.25	1.31	0.00
49	59.20	0.00	0.00	2.22	1.47	7.87	2.74	0.70	17.58	6.10	2.49	0.46	0.00	0.00
50	46.35	0.00	0.00	1.94	0.47	3.43	0.33	1.13	28.37	14.20	1.78	0.66	0.00	0.00
51部位1	53.66	0.00	0.00	1.84	1.19	8.15	1.19	1.37	21.94	8.94	1.93	0.39	0.47	0.00
51部位2	54.35	0.00	0.00	3.07	1.45	4.55	0.42	0.75	32.69	0.00	2.43	0.00	0.00	0.00
52	53.43	8.63	0.00	1.11	0.55	1.71	0.23	0.70	24.62	8.64	1.29	0.44	0.00	0.00
54	49.80	0.00	0.32	1.68	1.28	6.60	0.00	0.83	31.00	7.04	1.03	0.88	0.00	0.00
54严重风化点	71.07	0.00	0.00	0.00	1.11	5.29	0.00	1.34	20.66	0.00	1.03	1.12	0.00	0.00
56	60.52	0.00	0.00	0.59	0.00	2.73	0.00	0.79	21.42	15.45	0.58	0.00	0.00	0.00
57	54.92	0.00	0.00	0.67	0.00	3.35	0.00	1.16	24.37	17.30	0.00	0.00	0.00	0.00
58	60.78	0.00	0.34	1.64	0.79	5.01	0.86	3.13	19.68	7.66	1.96	0.24	0.00	0.00

附录 2.1 随机森林模型中的 11 棵决策树







附录 4. 1. 1

	二氧化硅(SiO2) 高	氧化钠(Na2O) 高	氧化钾(K2O) 高	氧化钙(CaO) 高	氧化镁(MgO) 高	氧化铝(Al2O3) 高	氧化铁(Fe2O3) 高	氧化铜(CuO) 高	氧化铅(PbO) 高	氧化钡(BaO) 高	五氧化二磷(P2O5) 高	氧化锶(SrO) 高	氧化锡(SnO2) 高	二氧化硫(SO2) 高
二氧化硅(SiO2)	1.000(0.000***)	-0.397(0.045**)	-0.581(0.001***)	-0.616(0.000***)	-0.427(0.017**)	-0.673(0.000***)	-0.551(0.001***)	-0.262(0.129)	-0.343(0.069*)	-0.267(0.173)	-0.433(0.012**)	-0.425(0.026**)	0.059(0.772)	-0.233(0.239)
氧化钠(Na2O)	-0.397(0.045**)	1.000(0.000***)	0.471(0.018**)	0.515(0.010***)	-0.184(0.368)	0.257(0.195)	0.187(0.346)	-0.094(0.638)	0.233(0.284)	-0.229(0.332)	-0.176(0.377)	-0.108(0.623)	-0.105(0.656)	-0.188(0.414)
氧化钾(K2O)	-0.581(0.001***)	0.471(0.018**)	1.000(0.000***)	0.497(0.004**)	0.132(0.461)	0.343(0.048**)	0.298(0.087*)	0.139(0.425)	0.181(0.339)	-0.021(0.916)	0.060(0.732)	0.289(0.133)	0.059(0.772)	0.283(0.157)
氧化钙(CaO)	-0.616(0.000***)	0.515(0.010***)	0.497(0.004**)	1.000(0.000***)	0.076(0.670)	0.446(0.010***)	0.336(0.053*)	0.257(0.139)	0.221(0.241)	-0.052(0.793)	0.099(0.569)	-0.052(0.785)	-0.315(0.123)	0.328(0.099*)
氧化镁(MgO)	-0.427(0.017**)	-0.184(0.368)	0.132(0.461)	0.076(0.670)	1.000(0.000***)	0.606(0.001***)	0.504(0.005**)	0.090(0.615)	0.198(0.309)	0.411(0.042**)	0.476(0.008***)	0.538(0.006***)	0.186(0.377)	0.369(0.072*)
氧化铝(Al2O3)	-0.673(0.000***)	0.257(0.195)	0.343(0.048**)	0.446(0.010***)	0.606(0.001***)	1.000(0.000***)	0.525(0.002***)	0.144(0.404)	0.457(0.015**)	0.370(0.059*)	0.393(0.023**)	0.407(0.033**)	-0.137(0.500)	0.210(0.289)
氧化铁(Fe2O3)	-0.551(0.001***)	0.187(0.346)	0.298(0.087*)	0.336(0.053*)	0.504(0.005**)	0.525(0.002***)	1.000(0.000***)	0.428(0.014**)	0.221(0.241)	0.391(0.046**)	0.349(0.044**)	0.330(0.084*)	-0.315(0.123)	0.258(0.195)
氧化铜(CuO)	-0.262(0.129)	-0.094(0.638)	0.139(0.425)	0.257(0.139)	0.090(0.615)	0.144(0.404)	0.428(0.014**)	1.000(0.000***)	0.041(0.828)	0.422(0.031**)	0.349(0.044**)	0.087(0.649)	-0.334(0.101)	0.293(0.141)
氧化铅(PbO)	-0.343(0.069*)	0.233(0.284)	0.181(0.339)	0.221(0.241)	0.198(0.309)	0.457(0.015**)	0.221(0.241)	0.041(0.828)	1.000(0.000***)	0.641(0.003**)	0.090(0.633)	0.325(0.120)	-0.171(0.443)	-0.306(0.159)
氧化钡(BaO)	-0.267(0.173)	-0.220(0.332)	-0.021(0.916)	-0.052(0.793)	0.411(0.042**)	0.370(0.059*)	0.391(0.046**)	0.422(0.031**)	0.641(0.003**)	1.000(0.000***)	0.350(0.074*)	0.477(0.029**)	-0.123(0.597)	-0.220(0.332)
五氧化二磷(P2O5)	-0.433(0.012**)	-0.176(0.377)	0.060(0.732)	0.099(0.569)	0.476(0.008***)	0.393(0.023**)	0.349(0.044**)	0.349(0.044**)	0.090(0.633)	0.350(0.074*)	1.000(0.000***)	0.400(0.036**)	0.256(0.210)	0.211(0.289)
氧化锶(SrO)	-0.425(0.026**)	-0.108(0.623)	0.289(0.133)	-0.052(0.785)	0.538(0.006***)	0.407(0.033**)	0.330(0.084*)	0.087(0.649)	0.325(0.120)	0.477(0.029**)	0.400(0.036**)	1.000(0.000***)	0.286(0.207)	-0.015(0.944)
氧化锡(SnO2)	0.059(0.772)	-0.105(0.656)	0.059(0.772)	-0.315(0.123)	0.186(0.377)	-0.137(0.500)	-0.315(0.123)	-0.334(0.101)	-0.171(0.443)	-0.123(0.597)	0.256(0.210)	0.286(0.207)	1.000(0.000***)	-0.105(0.656)
二氧化硫(SO2)	-0.233(0.239)	-0.188(0.414)	0.283(0.157)	0.328(0.099*)	0.369(0.072*)	0.210(0.289)	0.258(0.195)	0.293(0.141)	-0.306(0.159)	-0.220(0.332)	0.211(0.289)	-0.015(0.944)	-0.105(0.656)	1.000(0.000***)

注：***、**、*分别代表1%、5%、10%的显著性水平

附录 4. 1. 2

	二氧化硅(SiO2) 铅	氧化钠(Na2O) 铅	氧化钾(K2O) 铅	氧化钙(CaO) 铅	氧化镁(MgO) 铅	氧化铝(Al2O3) 铅	氧化铁(Fe2O3) 铅	氧化铜(CuO) 铅	氧化铅(PbO) 铅	氧化钡(BaO) 铅	五氧化二磷(P2O5) 铅	氧化锶(SrO) 铅	氧化锡(SnO2) 铅	二氧化硫(SO2) 铅
二氧化硅(SiO2)	1.000(0.000***)	0.318(0.004***)	0.206(0.054*)	0.334(0.001***)	0.095(0.360)	0.295(0.003***)	0.031(0.765)	0.319(0.001***)	0.559(0.000***)	-0.198(0.045**)	0.360(0.000***)	0.398(0.000***)	0.081(0.489)	-0.258(0.027**)
氧化钠(Na2O)	0.318(0.004***)	1.000(0.000***)	0.060(0.616)	0.293(0.009***)	0.021(0.858)	0.110(0.324)	-0.177(0.136)	-0.080(0.474)	-0.272(0.015**)	0.084(0.454)	0.468(0.000***)	-0.116(0.306)	-0.082(0.535)	-0.191(0.148)
氧化钾(K2O)	0.206(0.054*)	0.060(0.616)	1.000(0.000***)	-0.017(0.875)	0.280(0.013**)	0.386(0.000***)	0.137(0.227)	-0.159(0.138)	-0.249(0.019**)	-0.010(0.926)	-0.084(0.435)	-0.133(0.219)	0.125(0.322)	-0.020(0.873)
氧化钙(CaO)	0.334(0.001***)	0.293(0.009***)	-0.017(0.875)	1.000(0.000***)	0.239(0.022**)	0.073(0.458)	0.297(0.005***)	-0.031(0.756)	0.239(0.016**)	-0.102(0.305)	0.425(0.000***)	0.218(0.030**)	0.224(0.056*)	0.066(0.575)
氧化镁(MgO)	0.095(0.360)	0.021(0.858)	0.280(0.013**)	0.239(0.022**)	1.000(0.000***)	0.521(0.000***)	0.189(0.086*)	-0.202(0.052*)	-0.063(0.545)	0.324(0.002***)	0.114(0.277)	0.054(0.605)	0.223(0.069*)	-0.312(0.011**)
氧化铝(Al2O3)	0.295(0.003***)	0.110(0.324)	0.386(0.000***)	0.073(0.458)	0.521(0.000***)	1.000(0.000***)	0.271(0.010***)	-0.234(0.018**)	0.274(0.006***)	0.265(0.007***)	-0.022(0.829)	-0.142(0.156)	0.273(0.019**)	0.308(0.008***)
氧化铁(Fe2O3)	0.031(0.765)	-0.177(0.136)	0.137(0.227)	0.297(0.005***)	0.189(0.086*)	0.271(0.010***)	1.000(0.000***)	0.280(0.008***)	0.044(0.677)	0.320(0.002***)	0.168(0.112)	-0.079(0.456)	0.311(0.012**)	-0.290(0.020**)
氧化铜(CuO)	0.319(0.001***)	-0.080(0.474)	-0.159(0.138)	-0.031(0.756)	-0.202(0.052*)	-0.234(0.018**)	0.280(0.008***)	1.000(0.000***)	0.055(0.581)	0.346(0.000***)	0.134(0.180)	0.120(0.232)	-0.294(0.012**)	0.359(0.002***)
氧化铅(PbO)	0.559(0.000***)	-0.272(0.015**)	-0.249(0.019**)	0.239(0.016**)	-0.063(0.545)	0.274(0.006***)	0.044(0.677)	0.055(0.581)	1.000(0.000***)	-0.062(0.529)	0.257(0.010***)	0.263(0.009***)	-0.073(0.531)	-0.085(0.468)
氧化钡(BaO)	-0.198(0.045**)	0.084(0.454)	-0.010(0.926)	-0.102(0.305)	0.324(0.002***)	0.265(0.007***)	0.320(0.002***)	0.346(0.000***)	-0.062(0.529)	1.000(0.000***)	-0.131(0.189)	0.122(0.222)	-0.019(0.869)	0.377(0.001***)
五氧化二磷(P2O5)	0.360(0.000***)	0.468(0.000***)	-0.084(0.435)	0.425(0.000***)	0.114(0.277)	-0.022(0.829)	0.168(0.112)	0.134(0.180)	0.257(0.010***)	-0.131(0.189)	1.000(0.000***)	0.181(0.074*)	-0.025(0.830)	0.167(0.156)
氧化锶(SrO)	0.398(0.000***)	-0.116(0.306)	-0.133(0.219)	0.218(0.030**)	0.054(0.605)	-0.142(0.156)	-0.079(0.456)	0.120(0.232)	0.263(0.009***)	0.122(0.222)	0.181(0.074*)	1.000(0.000***)	0.012(0.921)	0.190(0.109)
氧化锡(SnO2)	0.081(0.489)	-0.082(0.535)	0.125(0.322)	0.224(0.056*)	0.223(0.069*)	0.273(0.019**)	0.311(0.012**)	-0.294(0.012**)	-0.073(0.531)	-0.019(0.869)	-0.025(0.830)	0.012(0.921)	1.000(0.000***)	-0.109(0.433)
二氧化硫(SO2)	-0.258(0.027**)	-0.191(0.148)	-0.020(0.873)	0.066(0.575)	-0.312(0.011**)	0.308(0.008***)	-0.290(0.020**)	0.359(0.002***)	-0.085(0.468)	0.377(0.001***)	0.167(0.156)	0.190(0.109)	-0.109(0.433)	1.000(0.000***)

注：***、**、*分别代表1%、5%、10%的显著性水平

附录 4. 2. 1 不同纹饰的肯德尔等级相关系数的显著性检验结果

A	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)	0	0.792	0.058*	0.475	0.038**	0.119	0.491	0.751	0.000***	0.002***	0.17	0.002***	0.835	0.265
氧化钠(Na2O)	0.792	0.000***	0.169	0.017**	0.326	0.943	0.002***	1	0.719	0.012**	0.000***	0.102	0.409	0.333
氧化钾(K2O)	0.058*	0.169	0.000***	0.764	0.019**	0.078*	0.086*	0.842	0.001***	0.003***	0.81	0.003***	0.568	0.028**
氧化钙(CaO)	0.475	0.017**	0.764	0.000***	0.513	0.142	0.000***	0.843	0.984	0.035**	0.125	0.781	0.929	0.018**
氧化镁(MgO)	0.038**	0.326	0.019**	0.513	0.000***	0.002***	0.078*	0.827	0.027**	0.005***	0.578	0.211	0.929	0.045**
氧化铝(Al2O3)	0.119	0.943	0.076*	0.142	0.002***	0.000***	0.026**	0.968	0.017**	0.029**	0.952	0.108	0.326	0.167
氧化铁(Fe2O3)	0.491	0.002***	0.086*	0.000***	0.078*	0.026**	0.000***	0.655	0.362	0.054*	0.011**	0.36	0.692	0.132
氧化铜(CuO)	0.751	1	0.842	0.843	0.827	0.968	0.655	0.000***	0.109	0.751	0.537	0.041**	0.000***	0.211
氧化铅(PbO)	0.000***	0.719	0.001***	0.984	0.027**	0.017**	0.362	0.109	0.000***	0.003***	0.524	0.000***	0.789	0.026**
氧化钡(BaO)	0.002***	0.012**	0.003***	0.035**	0.005***	0.029**	0.054*	0.751	0.003***	0.000***	0.224	0.002***	0.976	0.040**
五氧化二磷(P2O5)	0.17	0.000***	0.81	0.125	0.578	0.952	0.011**	0.537	0.524	0.224	0.000***	0.842	0.928	0.893
氧化锶(SrO)	0.002***	0.102	0.003***	0.781	0.211	0.108	0.36	0.041**	0.000***	0.002***	0.842	0.000***	0.591	0.081*
氧化锡(SnO2)	0.835	0.409	0.568	0.929	0.929	0.326	0.692	0.000***	0.789	0.976	0.928	0.591	0.000***	0.505
二氧化硫(SO2)	0.265	0.333	0.028**	0.018**	0.045**	0.167	0.132	0.211	0.026**	0.040**	0.893	0.081*	0.505	0.000***
B	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)	0.003***	0.000***	0.444	0.003***	0.114	0.017**	1	1	0.000***	0.000***	0.272	0.000***	0.000***	0.000***
氧化钠(Na2O)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
氧化钾(K2O)	0.444	0.000***	0.000***	0.444	0.647	0.444	0.251	0.702	0.000***	0.000***	0.022**	0.000***	0.000***	0.000***
氧化钙(CaO)	0.003***	0.000***	0.444	0.003***	0.114	0.017**	1	1	0.000***	0.000***	0.272	0.000***	0.000***	0.000***
氧化镁(MgO)	0.114	0.000***	0.647	0.114	0.016**	0.042**	0.499	0.114	0.000***	0.000***	0.822	0.000***	0.000***	0.000***
氧化铝(Al2O3)	0.017**	0.000***	0.444	0.017**	0.042**	0.003***	1	1	0.000***	0.000***	0.469	0.000***	0.000***	0.000***
氧化铁(Fe2O3)	1	0.000***	0.251	1	0.499	1	0.003***	0.272	0.000***	0.000***	0.469	0.000***	0.000***	0.000***
氧化铜(CuO)	1	0.000***	0.702	1	0.114	1	0.272	0.003***	0.000***	0.000***	0.469	0.000***	0.000***	0.000***
氧化铅(PbO)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
氧化钡(BaO)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
五氧化二磷(P2O5)	0.272	0.000***	0.022**	0.272	0.822	0.469	0.469	0.469	0.000***	0.000***	0.003***	0.000***	0.000***	0.000***
氧化锶(SrO)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
氧化锡(SnO2)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
二氧化硫(SO2)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
C	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)	0.000***	0.025**	0.049**	0.173	0.552	0.085*	0.103	0.057*	0.000***	0.014**	0.004***	0.000***	0.599	0.145
氧化钠(Na2O)	0.025**	0.000***	0.289	0.679	0.182	0.32	0.121	0.291	0.238	0.146	0.035**	0.034**	0.576	0.139
氧化钾(K2O)	0.049**	0.289	0.000***	0.018**	0.395	0.001***	0.122	0.863	0.005***	0.165	0.557	0.141	0.414	0.806
氧化钙(CaO)	0.173	0.679	0.018**	0.000***	0.011**	0.011**	0.028**	0.188	0.515	0.181	0.014**	0.987	0.401	0.677
氧化镁(MgO)	0.552	0.182	0.395	0.011**	0.000***	0.001***	0.038**	0.396	0.552	0.008***	0.037**	0.428	0.136	0.095*
氧化铝(Al2O3)	0.085*	0.32	0.001***	0.011**	0.001***	0.000***	0.005***	0.877	0.085*	0.000***	0.721	0.166	0.172	0.121
氧化铁(Fe2O3)	0.103	0.121	0.122	0.028**	0.038**	0.005***	0.000***	0.286	0.309	0.002***	0.834	0.144	0.275	0.074*
氧化铜(CuO)	0.057*	0.291	0.863	0.188	0.396	0.877	0.286	0.000***	0.285	0.024**	0.411	0.098*	0.753	0.004***
氧化铅(PbO)	0.000***	0.238	0.005***	0.515	0.552	0.085*	0.309	0.285	0.000***	0.446	0.137	0.001***	0.834	0.095*
氧化钡(BaO)	0.014**	0.146	0.165	0.181	0.008***	0.000***	0.002***	0.024**	0.446	0.000***	0.619	0.097*	0.752	0.012**
五氧化二磷(P2O5)	0.004***	0.035**	0.557	0.014**	0.037**	0.721	0.834	0.411	0.137	0.619	0.000***	0.024**	0.248	0.172
氧化锶(SrO)	0.000***	0.034**	0.141	0.987	0.428	0.166	0.144	0.098*	0.001***	0.097*	0.024**	0.000***	0.749	0.335
氧化锡(SnO2)	0.599	0.576	0.414	0.401	0.136	0.172	0.275	0.753	0.834	0.752	0.248	0.749	0.000***	0.64
二氧化硫(SO2)	0.145	0.139	0.806	0.677	0.095*	0.121	0.074*	0.004***	0.095*	0.012**	0.172	0.335	0.64	0.000***