

古代玻璃制品的成分分析与鉴别

摘 要

本论文就部分玻璃文物样品的特征与成分进行了统计与分析，并据此得出了不同类型玻璃制品在风化前后的成分规律、对文物风化部位在风化前的成分含量进行了预测、细分样品亚类、对未知样品归类、分析了不同类别玻璃文物的化学成分之间的关联与差异性。

问题一中，进行了文物特征与风化的相关性分析、结合类型的成分含量的统计规律分析以及对风化部位风化前成分含量的预测；应用了描述性统计、卡方检验、kappa 一致性检验等方法来解决问题，并得出在若干特征中，表面风化与玻璃类型的相关程度最大；文物表面风化过程中若干成分有不同程度的改变；根据风化采样点预测出未风化文物的成分比例。

问题二中，分析了两种类型玻璃的化学成分统计规律，并分别进行亚类划分，评估了亚类划分的合理性与敏感性；应用了随机森林算法进行分类，并使用 K-means 聚类分析进行亚类划分，通过向测试集添加微扰数据检验其敏感性；得出了分类的主要特征成分、亚类划分的具体标准范围、亚类模型的敏感性合适。

问题三中，解决了未知类型玻璃文物的归类与敏感性分析问题；应用了随机森林算法等方法；将未知类型的玻璃文物分入高钾与铅钡两类并进行了敏感性分析。

问题四中，解决了不同类型文物的化学成分之间的关联性与差异性问题；应用了 Spearman 相关性分析、可视化分析等方法；得出了化学成分之间的关联性以及不同类型化学成分之间的差异性。

关键词：玻璃成分分析 聚类分析 相关性分析 随机森林

一、问题重述

古代的玻璃文物由于取材和制作工艺不同，通常在纹饰、类型、颜色上有所差异，且外观相似的玻璃制品也可能因为取材不同在化学成分组成上有较大差异。

玻璃的主要化学成分为二氧化硅，在制作工艺中常用石灰石作为稳定剂，当使用助熔剂进行加工时，成品玻璃中诸如氧化钾、氧化铅、氧化钡等成分的含量也会因助熔剂不同而产生差异。

古代玻璃制品的成分易受埋藏环境影响而产生改变，风化过程中玻璃元素与环境元素发生大量交换，影响其各个成分的比例从而影响对于类别的判断，且文物不同部位的风化程度也不尽相同。

给定部分古代玻璃文物的数据信息，其中包括纹饰、类型、颜色、表面风化等特征信息，以及对应文物各化学成分含量比例的信息。给定数据中，可能存在测量误差与数据缺失，现需要根据上述信息，进行数据过滤与归一化后，解决下列四个问题：

问题 1 分析玻璃文物的纹饰、类型、颜色三个特征与表面风化的关系；根据类型分析文物表面有无风化的化学成分的统计规律；通过风化点检测数据预测风化前的化学成分含量。

问题 2 结合数据分析高钾和铅钡玻璃的分类规律，根据合理的分类方法给出亚类划分方案，并分析其合理性与敏感性。

问题 3 针对给定未知类型的玻璃文物的化学成分进行分类，并分析分类结果的敏感性。

问题 4 对不同类别的玻璃文物分析其化学成分之间的关联性，并分析不同类型化学成分关联性的差异。

二、问题分析

对于本文所述的四个问题，团队进行了如下的分析：

2.1 问题一的分析

对于问题一，可以拆分为三个子问题。

第一个子问题是对表面风化与其余三个特征进行差异性分析与相关性分析。对于这个问题我们将表单一数据代入 SPSS 中进行卡方检验，分析模型是否呈现显著性，通过显著值 p 判断数据是否存在显著性差异，进而分析其相关性与差异性。

第二个子问题是分析表面是否风化的化学成分的统计规律。解决该问题时，团队分不同类型对文物成分使用了描述性统计分析、正态性校验等分析方法，从中归纳出风化前后化学成分变化的统计规律。

第三个子问题是通过风化点预测风化前的化学成分的含量，对于该问题，团队针对不同类型的样本，使用了基于均值与样本数量的加权估值的分析方法，通过代入风化点数据计算得预测值，并进行了合理性分析。

2.2 问题二的分析

对于问题二，团队将其拆分为三个子问题。

第一个子问题是分析两种类型玻璃的分类规律。团队根据通过合并比表单一和表二所得的训练集建立了随机森林分类模型，并通过建立的随机森林来计算特征重要性，取其中最为重要的特征化学成分作为分类依据。

第二个子问题是对于不同类型的玻璃文物进行有具体指标的亚类划分。对于该问题，我们首先将数据根据问题一模型进行归一化，得到未风化玻璃文物的成分比例，应用 K-means 算法进行聚类分析，基于化学成分特征将不同类型的玻璃文物分为若干亚类。

第三个子问题是评判亚类划分的合理性与敏感性。亚类划分的合理性根据特征显著性说明。检验亚类划分的敏感性本文使用了微扰检验法，通过不同幅度调整训练集数据并加入测试集中观察其结果变化情况，来检验分类模型的敏感性。

2.3 问题三的分析

对于问题三，可以拆分为两个子问题。

第一个子问题是鉴别未知玻璃文物的类型。团队运用了随机森林方法，通过对表四进行多次训练和测试，并将两种模型运用于表三，得出测试结果，进行比对和筛查数据，预测表三中未知玻璃文物的类型。

第二个子问题是对鉴别结果进行敏感性分析。分析过程与随机森林根据训练集生成过程同时进行，通过将训练集作为测试集进行召回测试进行敏感性检验。

2.4 问题四的分析

对于问题四，可以拆分为两个子问题。

第一个子问题是分析不同类别的玻璃文物的化学成分之间的关联性。团队运用 spearman 相关性分析，分别对两种类型的玻璃中的所有化学成分进行相关性计算，得到任意两种不同的化学成分的关系及其显著性。

第二个子问题是对两个类别分析上述关联性的差异性。对于该问题，团队取了不同类型玻璃的任意两种不同化学成分之间的相关系数，用 excel 分别做了相关系数可视化表格，以红-白-蓝三种色阶反映化学成分之间的关联性大小，并对关联性的正负相关性作出差异性解释。

三、模型假设

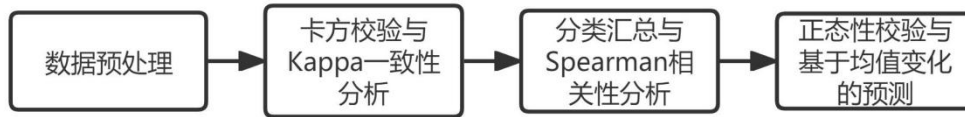
1. 假设讨论范围内文物仅有高钾与铅钡两类
2. 假设纹饰不会对采样点成分产生较大影响(相关性检验已验证)
3. 假设成分测量数据误差在可接受范围内
4. 假设成分数据留空代表未检测到该成分

四、符号说明

| 符号 | 说明 | 单位 |
|----------|-----|----|
| p | 显著值 | |
| χ^2 | 卡方 | |

五、模型的建立与求解

5.1 问题一模型的建立与求解



5.1.1 数据预处理

依据题目所述，本文首先对给定数据进行了如下方面的数据预处理，通过数据清洗、数据集成、数据变换获得有效数据。

1. 补全样本缺失信息

给定数据中，表单一的四例样本颜色缺失，表单二的部分化学成分含量为空白。据题述，表单二的空白处为未检测到对应成分，于是将未检测到的成分填补为 0。对于表单一缺失颜色的样本，本文采用最近邻插值法进行补差：19、40、48、58 四个文物样本均补差为浅蓝色^[3]。

2. 剔除测试失误样本

题目指出，表单二中各化学成分比例之和处于 85%~105% 之间时视为有效数据。经过求和算得，表单二中编号为 15 与 17 的样本各化学成分比例总和为 79.47% 和 71.89%，不满足 85% 的下限要求，属于测量失误数据，因此对其进行剔除^[3]。

3. 合并表单信息

表单二数据为表单一中已分类样本不同部位的化学成分比例。预处理的数据集成步骤中，本文将表单一文物编号与表单二文物采样点作为关键字，进行数据库 Join 操作后得到含有纹饰、类型、颜色、表面风化与化学成分比例的合并表格，便于后续交叉分析。

5.1.2 表面风化与纹饰、类型、颜色的关系分析

将表单一根据类型进行分组，并将分组后的数据引入 SPSS 软件中，对表面风化与纹饰、类型、颜色的关系通过卡方校验进行差异性分析，校验结果如下表所示：

| 题目 | 名称 | 表面风化 | | 总计 | X ² | 校正 X ² | P |
|----|----|------|----|----|----------------|-------------------|----------|
| | | 无风化 | 风化 | | | | |
| 颜色 | 浅绿 | 2 | 1 | 3 | 7.234 | 7.234 | 0.405 |
| | 浅蓝 | 8 | 16 | 24 | | | |
| | 深绿 | 3 | 4 | 7 | | | |
| | 深蓝 | 2 | 0 | 2 | | | |
| | 紫 | 2 | 2 | 4 | | | |
| | 绿 | 1 | 0 | 1 | | | |
| | 蓝绿 | 6 | 9 | 15 | | | |
| | 黑 | 0 | 2 | 2 | | | |
| 类型 | 铅钋 | 12 | 28 | 40 | 6.880 | 5.452 | 0.009*** |
| | 高钾 | 12 | 6 | 18 | | | |
| 纹饰 | A | 11 | 11 | 22 | 4.957 | 4.957 | 0.084* |
| | B | 0 | 6 | 6 | | | |
| | C | 13 | 17 | 30 | | | |

注：***、**、*分别代表 1%、5%、10%的显著性水平

表 1 卡方检验-差异性分析

由上述卡方检验分析的结果可知，对于表面风化，显著性 P 值为 0.009***，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。

而颜色对于表面风化，显著性 P 值为 0.405，纹饰对于表面风化，显著性 P 值为 0.084*，二者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、颜色数据不存在显著性差异^[1]。

随后对表单一数据进行 kappa 一致性检验，检验结果如下表所示：

| Kappa 一致性检验 | | | | |
|-------------|---------|-------|--------|----------|
| 配对项 | Kappa 值 | 标准误差 | Z 值 | p 值 |
| 表面风化配对颜色 | 0.046 | 0.08 | 0.571 | 0.568 |
| 表面风化配对类型 | 0.336 | 0.128 | 2.623 | 0.009*** |
| 表面风化配对纹饰 | -0.04 | 0.107 | -0.373 | 0.709 |

注：***、**、*分别代表 1%、5%、10%的显著性水平

表 2 kappa 检验-相关分析

由上述卡方检验分析的结果可知，基于表面风化和颜色，显著性 p 值为 0.568；基于表面风化和纹饰，可以看到显著性 p 值为 0.709，水平上不呈现显著性，不能拒绝原假设，说明两变量之间不存在一致性。同时，Kappa 系数的值为 0.046 和 -0.04，因此相关性的程度为极低的一致性。

基于表面风化和类型，可以看到显著性 p 值为 0.009***，水平上呈现显著性，拒绝原假设，说明两变量之间存在一致性。同时，Kappa 系数的值为 0.336，因此相关性的程度为一般的一致性^[2]。

综上所述，玻璃类型与表面风化的相关性与差异性最大。

5.1.3 分析文物表面是否风化的化学成分统计规律

对于该问题，团队采用描述性统计的方法进行分析。过程中对表二的数据统计了含有不同化学成分的样本数量、对应成分的极值、均值、标准差，并从中归纳出一定的统计规律。下列四表反映了分类汇总的结果，其中化学成分已按其含量均值降序排列：

| 高钾玻璃风化前描述统计 | | | | | |
|---------------------------------------|----|-------|-------|-------|------|
| | N | 最小值 | 最大值 | 均值 | 标准偏差 |
| 二氧化硅(SiO ₂) | 12 | 59.01 | 87.05 | 67.98 | 8.76 |
| 氧化钾(K ₂ O) | 11 | 5.19 | 14.52 | 10.18 | 2.72 |
| 氧化铝(Al ₂ O ₃) | 12 | 3.05 | 11.15 | 6.62 | 2.49 |
| 氧化钙(CaO) | 10 | 2.01 | 8.70 | 6.40 | 2.03 |
| 氧化钠(Na ₂ O) | 3 | 2.10 | 3.38 | 2.78 | 0.64 |
| 氧化铜(CuO) | 11 | 0.47 | 5.09 | 2.68 | 1.54 |
| 氧化锡(SnO ₂) | 1 | 2.36 | 2.36 | 2.36 | |
| 氧化铁(Fe ₂ O ₃) | 10 | 0.42 | 6.04 | 2.32 | 1.55 |
| 五氧化二磷(P ₂ O ₅) | 11 | 0.16 | 4.50 | 1.53 | 1.43 |
| 氧化钡(BaO) | 5 | 0.00 | 2.86 | 1.44 | 1.07 |
| 氧化镁(MgO) | 10 | 0.52 | 1.98 | 1.30 | 0.50 |
| 氧化铅(PbO) | 7 | 0.11 | 1.62 | 0.71 | 0.63 |
| 二氧化硫(SO ₂) | 3 | 0.36 | 0.47 | 0.41 | 0.06 |
| 氧化锶(SrO) | 6 | 0.04 | 0.12 | 0.08 | 0.03 |
| 总数 | 12 | | | | |

表 3 未风化高钾玻璃描述性统计

| 高钾玻璃风化后描述统计 | | | | | |
|---------------------------------------|---|-------|-------|-------|------|
| | N | 最小值 | 最大值 | 均值 | 标准偏差 |
| 二氧化硅(SiO ₂) | 6 | 92.35 | 96.77 | 93.96 | 1.73 |
| 氧化铝(Al ₂ O ₃) | 6 | 0.81 | 3.50 | 1.93 | 0.96 |
| 氧化铜(CuO) | 6 | 0.55 | 3.24 | 1.56 | 0.93 |
| 氧化钙(CaO) | 6 | 0.21 | 1.66 | 0.87 | 0.49 |
| 氧化钾(K ₂ O) | 4 | 0.59 | 1.01 | 0.82 | 0.19 |
| 氧化镁(MgO) | 2 | 0.54 | 0.64 | 0.59 | 0.07 |
| 五氧化二磷(P ₂ O ₅) | 5 | 0.15 | 0.61 | 0.34 | 0.18 |
| 氧化铁(Fe ₂ O ₃) | 6 | 0.17 | 0.35 | 0.27 | 0.07 |
| 二氧化硫(SO ₂) | 0 | | | | |
| 氧化锡(SnO ₂) | 0 | | | | |
| 氧化锶(SrO) | 0 | | | | |
| 氧化钡(BaO) | 0 | | | | |
| 氧化铅(PbO) | 0 | | | | |
| 氧化钠(Na ₂ O) | 0 | | | | |
| 总数 | 6 | | | | |

表 4 风化高钾玻璃描述性统计

由风化前后高钾玻璃各化学成分比例的描述性统计对照可知，风化后二氧化硅含量上升约 25%，风化后氧化钾、氧化钙、氧化铝、氧化铁、五氧化二磷含量有较大程度(3~10%)的流失，氧化铜与氧化镁含量有小幅降低。

推测出现上述情形的可能原因为：风化过程中二氧化硅流失较少，而如氧化钾

等可溶成分与土壤成分发生大量交换而流失，导致风化部位剩余成分中二氧化硅比例增加，其他氧化物成分相应减少。

| 铅钡玻璃风化前描述统计 | | | | | |
|---------------------------------------|----|-------|-------|-------|-------|
| | N | 最小值 | 最大值 | 均值 | 标准偏差 |
| 二氧化硅(SiO ₂) | 23 | 31.94 | 75.51 | 54.66 | 11.83 |
| 氧化铅(PbO) | 23 | 9.30 | 39.22 | 22.08 | 8.22 |
| 氧化钡(BaO) | 23 | 2.03 | 26.23 | 9.00 | 5.83 |
| 氧化铝(Al ₂ O ₃) | 23 | 1.42 | 14.34 | 4.46 | 3.26 |
| 氧化钠(Na ₂ O) | 10 | 0.92 | 7.92 | 3.87 | 2.08 |
| 二氧化硫(SO ₂) | 1 | 3.66 | 3.66 | 3.66 | |
| 氧化铜(CuO) | 21 | 0.11 | 8.46 | 1.57 | 2.01 |
| 氧化铁(Fe ₂ O ₃) | 11 | 0.17 | 4.59 | 1.54 | 1.25 |
| 氧化钙(CaO) | 20 | 0.38 | 4.49 | 1.52 | 1.26 |
| 五氧化二磷(P ₂ O ₅) | 17 | 0.08 | 6.34 | 1.42 | 2.03 |
| 氧化镁(MgO) | 15 | 0.51 | 1.67 | 0.98 | 0.33 |
| 氧化锶(SrO) | 17 | 0.12 | 0.91 | 0.36 | 0.21 |
| 氧化锡(SnO ₂) | 3 | 0.23 | 0.44 | 0.36 | 0.11 |
| 氧化钾(K ₂ O) | 15 | 0.11 | 1.41 | 0.34 | 0.33 |
| 有效个案数（成列） | 23 | | | | |

表 5 未风化铅钡玻璃描述性统计

| 铅钡玻璃风化后描述统计 | | | | | |
|---------------------------------------|----|-------|-------|-------|-------|
| | N | 最小值 | 最大值 | 均值 | 标准偏差 |
| 氧化铅(PbO) | 26 | 15.71 | 70.21 | 43.31 | 12.23 |
| 二氧化硅(SiO ₂) | 26 | 3.72 | 53.33 | 24.91 | 10.61 |
| 氧化钡(BaO) | 23 | 3.26 | 35.45 | 13.35 | 9.57 |
| 二氧化硫(SO ₂) | 4 | 1.96 | 15.95 | 8.88 | 7.65 |
| 五氧化二磷(P ₂ O ₅) | 24 | 0.07 | 14.13 | 5.72 | 4.06 |
| 氧化铝(Al ₂ O ₃) | 26 | 0.45 | 13.65 | 2.97 | 2.63 |
| 氧化钙(CaO) | 25 | 0.37 | 6.40 | 2.80 | 1.60 |
| 氧化铜(CuO) | 24 | 0.19 | 10.57 | 2.47 | 2.86 |
| 氧化钠(Na ₂ O) | 4 | 0.80 | 2.22 | 1.41 | 0.60 |
| 氧化镁(MgO) | 15 | 0.47 | 2.73 | 1.13 | 0.56 |
| 氧化铁(Fe ₂ O ₃) | 16 | 0.19 | 2.74 | 0.95 | 0.73 |
| 氧化锡(SnO ₂) | 2 | 0.47 | 1.31 | 0.89 | 0.59 |
| 氧化锶(SrO) | 23 | 0.19 | 1.12 | 0.47 | 0.23 |
| 氧化钾(K ₂ O) | 9 | 0.14 | 1.05 | 0.39 | 0.27 |
| 总数 | 26 | | | | |

表 6 风化铅钡玻璃描述性统计

由风化前后铅钡玻璃各化学成分比例的描述性统计对照可知，风化后二氧化硅含量下降约 30%，风化后氧化铅上升约 21%，氧化钠、氧化钾、氧化铝、氧化铁、五氧化二磷含量有小幅(<2%)的下降，二氧化硫、五氧化二磷、氧化钡、氧化钙呈现小幅度上升(1~8%)，其他成分基本持平。

与高钾玻璃样品不同，铅钡玻璃在风化后二氧化硅成分比例大多呈减小趋势，推测其可能原因为，在风化过程中，铅钡玻璃的二氧化硅成分相对氧化铅和氧化钡更易流失，且成分比例可能使其相对高钾玻璃更易失去二氧化硅，因而二氧化硅比例下降，氧化铅与氧化钡含量上升。

5.1.4 根据风化点预测风化前成分含量

对于该问题，本文首先采用频数直方图正态性检验法，观察各化学成分在风化前后比例的变化。将两种类型风化前后的数据引入 SPSS 软件中，并将不同化学成分的比例作为横坐标，统计对应区间样本的数量，并检验是否符合正态分布，观察风化前后化学成分比例的变化。以下为正态性检验直方图：

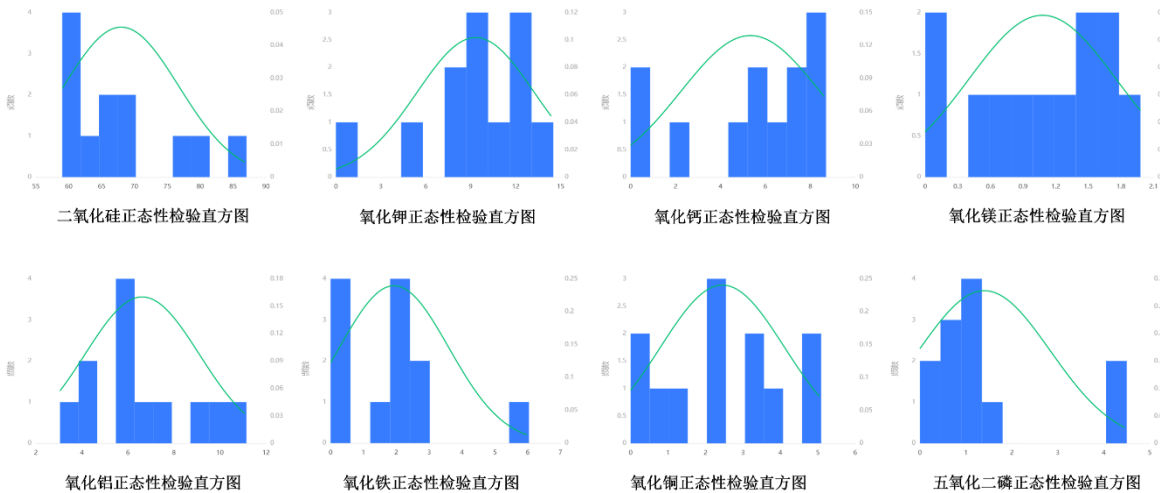


图 1 高钾玻璃(未风化)化学成分比例正态性检验直方图

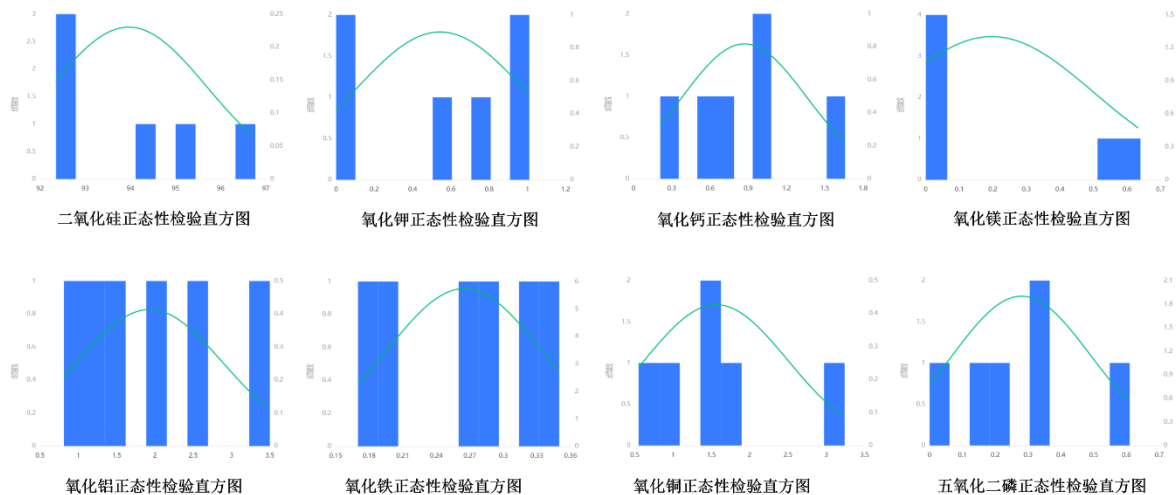


图 2 高钾玻璃(风化)化学成分比例正态性检验直方图

从正态性检验结果而言，通过累计概率(P)与正态累计概率(P)图(P-P 图)以及正态性检验 Quantile Quantile Plot(Q-Q 图) 的拟合程度判断得知，部分数据虽然不是绝对正态，但基本可接受为正态分布，但风化后样本的氧化铁与氧化铝含量不满足正态分布规律。

对于高钾玻璃样本，此处对于其中 8 个化学成分比例进行了统计规律的分析，由于高钾玻璃样本相对较少，风化前后难以形成对照，其他元素在表格中常体现为

缺失，通过此种分析方法无法较为精确地反映其含量，因此未进行正态性检验直方图分析。

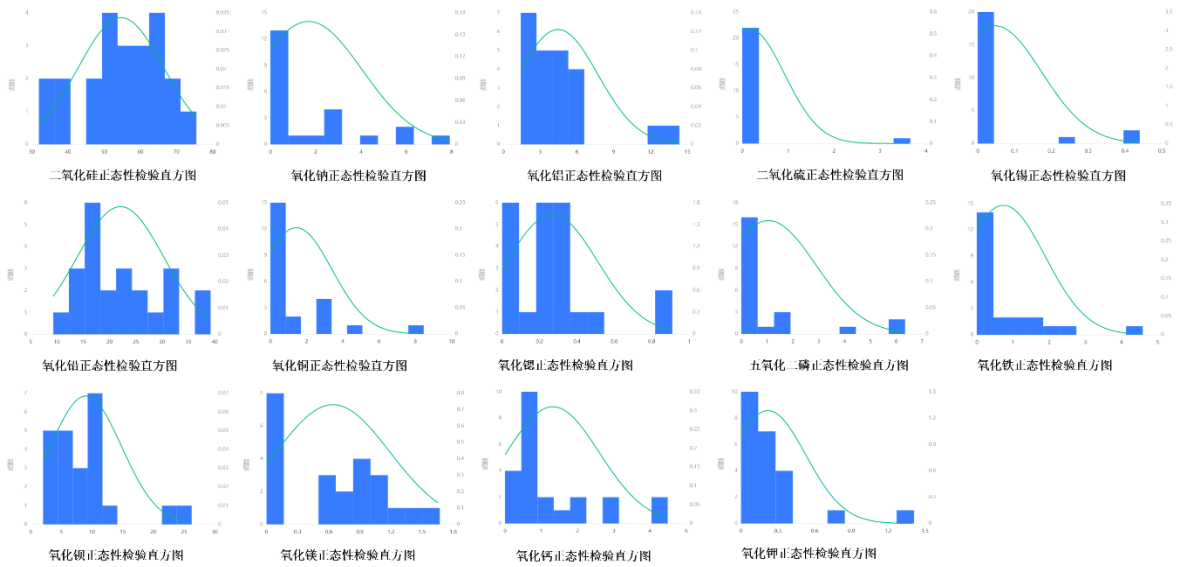


图 3 铅钡玻璃(未风化)化学成分比例正态性检验直方图

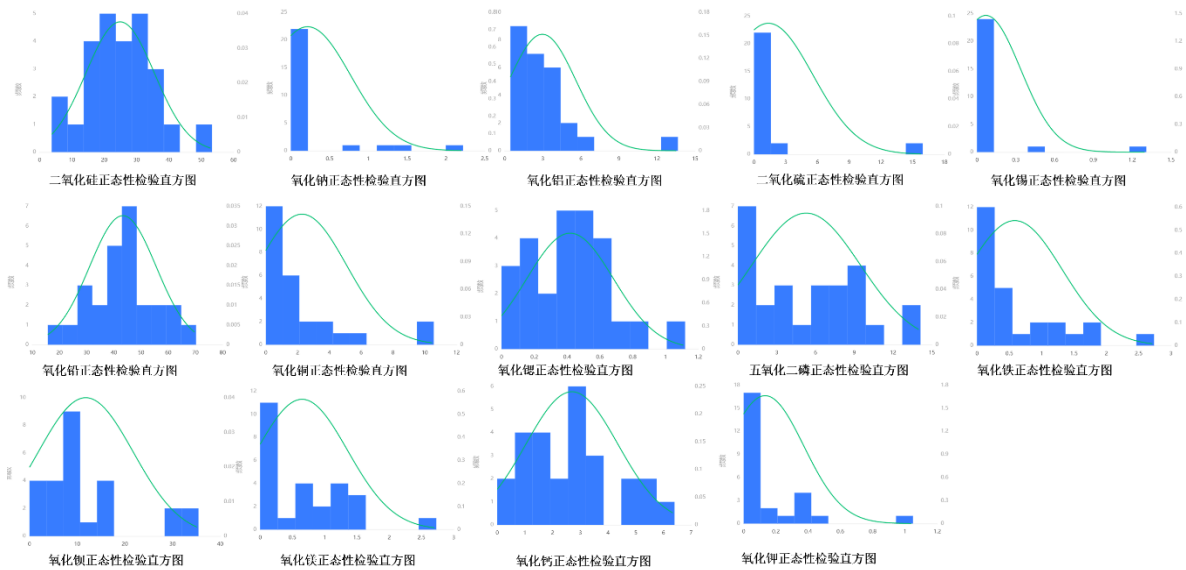


图 4 铅钡玻璃(风化)化学成分比例正态性检验直方图

对于铅钡玻璃样本，由于样本数量充足且成分缺失情况较少，此处对其进行了全部 14 种化学成分比例的分析。

从正态性检验结果而言，通过累计概率(P)与正态累计概率(P)图(P-P 图)以及正态性检验 Quantile Quantile Plot(Q-Q 图) 的拟合程度判断得知，上述数据虽然不是绝对正态，但基本可接受为正态分布。

在完成正态性检验后，本文采用均值的方法通过风化后的化学成分比例预测风化前比例，进行此预测的合理性在于上文进行的描述性分析中，计算所得标准差属于合理范围内，因此可以使用均值与基于样本数量的加权分析方法计算风化前化学成分。

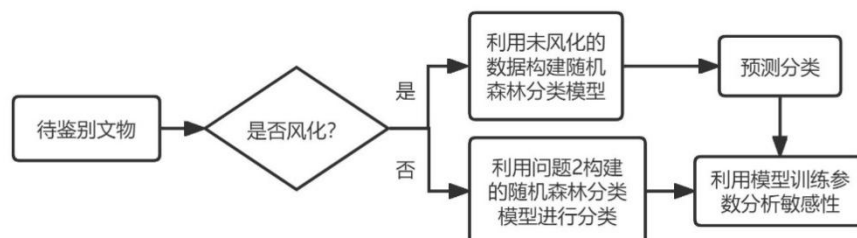
| 文物采样点 | 二氧化硅 (SiO ₂) | 氧化钠 (Na ₂ O) | 氧化钾 (K ₂ O) | 氧化钙 (CaO) | 氧化镁 (MgO) | 氧化铝 (Al ₂ O ₃) | 氧化铁 (Fe ₂ O ₃) |
|-------|-----------------------------|----------------------------|---------------------------|--------------|--------------|--|--|
| 02 | 66.03 | 1.47 | 1.14 | 0.97 | 1.17 | 7.22 | 2.01 |
| 07 | 66.65 | 0.70 | 8.79 | 5.53 | 0.88 | 6.67 | 1.84 |
| 09 | 69.04 | 0.70 | 9.38 | 5.08 | 0.88 | 6.01 | 1.99 |
| 10 | 70.79 | 0.70 | 9.71 | 4.67 | 0.88 | 5.50 | 1.93 |
| 11 | 63.34 | 1.47 | 0.30 | 2.14 | 0.70 | 4.18 | 0.15 |
| 12 | 68.31 | 0.70 | 9.80 | 5.18 | 0.88 | 6.15 | 1.96 |
| 19 | 59.39 | 1.47 | 0.09 | 1.56 | 0.58 | 5.06 | 1.48 |
| 22 | 66.37 | 0.70 | 9.53 | 6.12 | 1.52 | 8.19 | 2.02 |
| 27 | 66.74 | 0.70 | 8.79 | 5.40 | 1.42 | 7.20 | 1.87 |
| 34 | 65.53 | 1.47 | 0.34 | 0.00 | 0.00 | 3.11 | 0.62 |
| 36 | 69.32 | 3.69 | 0.23 | 0.00 | 0.00 | 3.09 | 0.47 |
| 38 | 62.68 | 2.85 | 0.09 | 0.00 | 0.00 | 4.06 | 0.44 |
| 39 | 56.00 | 1.47 | 0.09 | 0.00 | 0.00 | 1.99 | 0.15 |
| 40 | 46.46 | 1.47 | 0.09 | 0.50 | 0.00 | 1.94 | 0.34 |
| 41 | 48.21 | 1.47 | 0.53 | 3.59 | 2.72 | 4.82 | 1.94 |
| 48 | 83.08 | 2.27 | 0.41 | 1.45 | 1.53 | 15.14 | 1.18 |
| 52 | 55.49 | 2.69 | 0.09 | 0.90 | 0.54 | 2.65 | 0.38 |
| 56 | 58.90 | 1.47 | 0.09 | 0.00 | 0.00 | 3.34 | 0.15 |
| 57 | 55.17 | 1.47 | 0.09 | 0.00 | 0.00 | 3.67 | 0.15 |
| 58 | 60.14 | 1.47 | 0.43 | 2.12 | 0.78 | 5.01 | 1.01 |

| 文物采样点 | 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) |
|-------|--------------|--------------|--------------|---|--------------|----------------------------|----------------------------|
| 02 | 0.00 | 26.20 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 |
| 07 | 4.13 | 0.41 | 0.60 | 1.73 | 0.04 | 0.20 | 0.10 |
| 09 | 2.44 | 0.41 | 0.60 | 1.47 | 0.04 | 0.20 | 0.10 |
| 10 | 1.73 | 0.41 | 0.60 | 1.12 | 0.04 | 0.20 | 0.10 |
| 11 | 4.09 | 4.16 | 11.81 | 5.15 | 0.22 | 0.00 | 0.00 |
| 12 | 2.54 | 0.41 | 0.60 | 1.27 | 0.04 | 0.20 | 0.10 |
| 19 | 2.67 | 21.59 | 2.55 | 4.60 | 0.04 | 0.00 | 0.00 |
| 22 | 1.44 | 0.41 | 0.60 | 1.33 | 0.04 | 0.20 | 0.10 |
| 27 | 2.43 | 0.41 | 0.60 | 1.48 | 0.04 | 0.20 | 0.10 |
| 34 | 0.67 | 25.32 | 7.20 | 0.00 | 0.07 | 0.00 | 0.00 |
| 36 | 0.00 | 20.38 | 8.03 | 0.00 | 0.07 | 0.00 | 0.00 |
| 38 | 0.00 | 28.08 | 6.99 | 0.00 | 0.26 | 0.00 | 0.00 |
| 39 | 0.04 | 39.80 | 4.42 | 0.00 | 0.46 | 0.00 | 0.00 |
| 40 | 0.00 | 48.98 | 3.89 | 0.00 | 0.53 | 0.00 | 0.00 |
| 41 | 0.00 | 22.89 | 6.96 | 3.23 | 0.32 | 0.00 | 0.00 |
| 48 | 0.00 | 0.00 | 4.51 | 0.00 | 0.10 | 1.29 | 0.00 |

| | | | | | | | |
|----|------|-------|-------|------|------|------|------|
| 52 | 0.00 | 26.19 | 5.84 | 1.48 | 0.29 | 0.00 | 0.00 |
| 56 | 0.00 | 20.02 | 12.65 | 0.00 | 0.00 | 0.00 | 0.00 |
| 57 | 0.32 | 23.87 | 14.50 | 0.00 | 0.00 | 0.00 | 0.00 |
| 58 | 2.29 | 18.12 | 4.86 | 4.76 | 0.09 | 0.00 | 0.00 |

表 7 风化前样本化学成分比例预测值

5.2 问题二模型的建立与求解



5.2.1 分析高钾玻璃、铅钡玻璃的分类规律

为了得出两种类型玻璃的分类规律，团队将分类规律根据风化与未风化分为两种情况分别讨论，以排除表面风化与否对分类的影响。

对于未风化条件下的分类规律探究，此处本文沿用了问题一中预测风化前化学成分比例的模型，根据给定的已风化样本预测出未风化条件下的化学成分比例，使得作为训练集的数据全部为未风化数据，以排除表面风化这一特征对于分类的影响。

此处使用随机森林算法，将已知数据导入 SPSS 软件进行训练，分析得出化学成分在玻璃分类中作为特征的重要程度。

随机森林算法在决策树的训练过程中进一步引入了随机性，其随机性通过样本选取和特征选择 2 个角度体现。

选用随机森林进行分类的原因为：在使用决策树分类时，注意到易对氧化铅过拟合；使用 k 近邻分类时，注意到二氧化硅含量占比较大，轻微的变化就会对分类结果造成较大的影响，因此需要更强的抗过拟合以及抗噪能力。而随机森林是以决策树为基学习器通过集成方式构建而成的有监督机器学习方法，进一步在决策树的训练过程引入了随机性，具有优良的抗过拟合与抗噪能力，适合题设样本的情况。

下图展示了各化学成分作为分类特征的重要性比例：

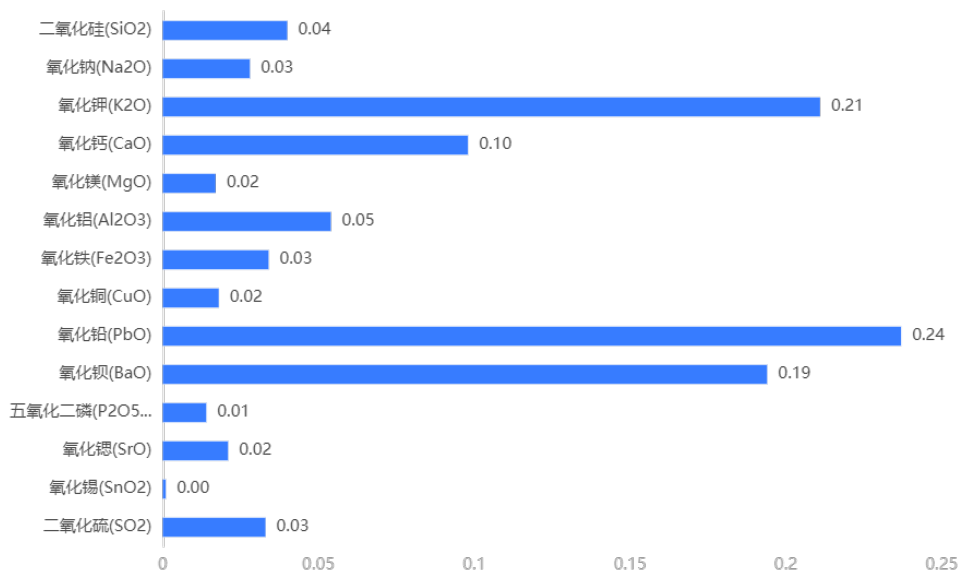


图 5 各化学成分作为分类特征的重要性比例(未风化)

对于已风化条件下的分类规律探究，相似地，随机森林的训练集仅导入风化后不同类型玻璃的化学成分比例，所得风化玻璃在分类过程中各化学成分作为特征的重要性比例如下图所示：

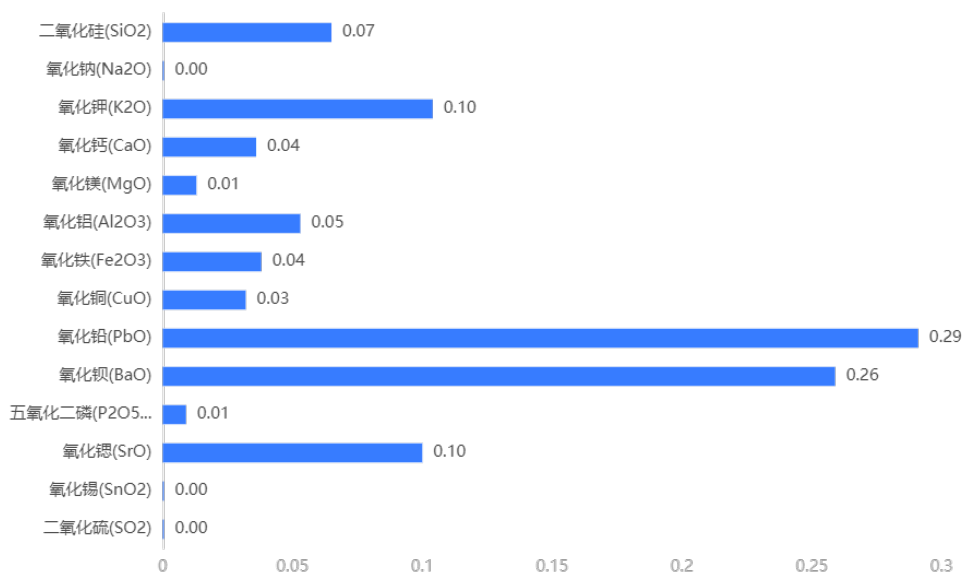


图 6 各化学成分作为分类特征的重要性比例(风化)

由此可得，在进行高钾玻璃与铅钡玻璃分类时，重要性比例最高、可以作为较为稳定的分类依据的三类化学成分依次为氧化铅、氧化钾、氧化钡。

推测得到该种结果的原因与题目背景中所述“助熔剂不同”有关，使用草木灰等高钾助熔剂生产的玻璃通常为高钾玻璃，含氧化钾较多；而使用铅矿石为助熔剂生产的玻璃含氧化钡、氧化铅较多，而作为在不同类型玻璃生产中均有使用的二氧化硅、氧化钙等化学成分则在分类中不具有较大重要性。

5. 2. 2 两种类型玻璃文物的亚类划分

对于该问题，团队使用 SPSS 对原数据进行 K-means 聚类分析，将化学成分比例作为字段进行聚类类别的差异性分析，并得出各亚类的划分标准。（两个类型的聚类树状图详见附件）

对于高钾玻璃，本文将其划分为 3 个亚类，三个亚类的具体划分标准如下表所示：

| | 聚类类别（平均值±标准差） | | | F | P |
|-------|---------------|--------------|--------------|--------|----------|
| | 类别 3 (n=9) | 类别 1 (n=6) | 类别 2 (n=3) | | |
| 二氧化硅 | 67.863±1.63 | 61.627±2.168 | 81.063±5.368 | 55.379 | 0.000*** |
| 氧化钠 | 0.463±0.348 | 1.39±1.576 | 0.0±0.0 | 2.706 | 0.099* |
| 氧化钾 | 9.224±0.813 | 11.722±2.287 | 4.87±4.718 | 9.292 | 0.002*** |
| 氧化钙 | 5.049±2.033 | 7.305±1.37 | 2.24±2.363 | 7.379 | 0.006*** |
| 氧化镁 | 1.209±0.419 | 0.965±0.704 | 0.917±0.809 | 0.447 | 0.648 |
| 氧化铝 | 6.804±2.004 | 7.437±1.844 | 4.433±1.603 | 2.577 | 0.109 |
| 氧化铁 | 1.976±0.184 | 2.437±2.055 | 0.79±1.368 | 1.629 | 0.229 |
| 氧化铜 | 2.586±0.887 | 2.802±1.897 | 1.353±1.714 | 1.123 | 0.351 |
| 氧化铅 | 0.297±0.182 | 0.582±0.737 | 0.417±0.52 | 0.622 | 0.550 |
| 氧化钡 | 0.552±0.404 | 0.638±1.155 | 0.657±1.137 | 0.028 | 0.973 |
| 五氧化二磷 | 1.617±0.997 | 1.262±1.656 | 1.04±0.354 | 0.317 | 0.733 |
| 氧化锶 | 0.04±0.032 | 0.053±0.05 | 0.023±0.04 | 0.578 | 0.573 |
| 氧化锡 | 0.131±0.099 | 0.0±0.0 | 0.787±1.363 | 2.601 | 0.107 |
| 二氧化硫 | 0.151±0.131 | 0.078±0.192 | 0.0±0.0 | 1.315 | 0.298 |

注：***、**、*分别代表 1%、5%、10%的显著性水平

表 8 高钾玻璃亚类划分标准

由表中显著性 P 值可知，在高钾玻璃的亚类划分中，二氧化硅、氧化钾、氧化钙水平上呈现显著性，在聚类分析划分的亚类之间存在显著性差异，即可以作为亚类划分的主要标准。根据上述标准划分亚类后，原数据划入各亚类的频数分别为 6、3、9，占比 33.33%、16.67%、50.00%，属于合理亚类划分。

对于铅钡玻璃，本文将其划分为 3 个亚类，三个亚类的具体划分标准如下表所示：

| | 聚类类别（平均值±标准差） | | | F | P |
|-------|---------------|--------------|--------------|--------|----------|
| | 类别 1 (n=18) | 类别 3 (n=11) | 类别 2 (n=8) | | |
| 二氧化硅 | 63.955±4.163 | 52.078±5.656 | 43.788±8.52 | 37.872 | 0.000*** |
| 氧化钠 | 1.183±1.035 | 2.826±2.747 | 0.994±1.139 | 3.735 | 0.034** |
| 氧化钾 | 0.25±0.237 | 0.258±0.213 | 0.198±0.491 | 0.106 | 0.900 |
| 氧化钙 | 1.249±0.841 | 0.793±1.127 | 1.822±1.831 | 1.733 | 0.192 |
| 氧化镁 | 0.595±0.484 | 0.882±0.823 | 0.254±0.377 | 2.651 | 0.085* |
| 氧化铝 | 5.065±3.387 | 4.079±1.499 | 2.654±1.237 | 2.431 | 0.103 |
| 氧化铁 | 0.841±1.099 | 0.492±0.721 | 0.863±1.108 | 0.486 | 0.619 |
| 氧化铜 | 1.329±1.403 | 1.395±1.619 | 1.397±2.889 | 0.006 | 0.994 |
| 氧化铅 | 17.124±6.082 | 21.357±4.842 | 36.289±6.526 | 30.089 | 0.000*** |
| 氧化钡 | 5.945±3.605 | 10.925±5.166 | 9.499±7.17 | 3.744 | 0.034** |
| 五氧化二磷 | 1.644±2.122 | 1.379±2.086 | 1.054±2.187 | 0.22 | 0.804 |
| 氧化锶 | 0.141±0.114 | 0.261±0.258 | 0.395±0.272 | 4.499 | 0.018** |
| 氧化锡 | 0.013±0.054 | 0.0±0.0 | 0.105±0.195 | 3.245 | 0.051* |
| 二氧化硫 | 0.203±0.863 | 0.0±0.0 | 0.0±0.0 | 0.514 | 0.603 |

注：***、**、*分别代表 1%、5%、10%的显著性水平

表 9 铅钡玻璃亚类划分标准

由表中显著性 P 值可知，在高钾玻璃的亚类划分中，二氧化硅、氧化铅水平上呈现较高显著性，氧化锶、氧化钡、氧化钠水平上呈现偏高显著性，在聚类分析划分的亚类之间存在显著性差异，即可以作为亚类划分的主要标准。根据上述标准划分亚类后，原数据划入各亚类的频数分别为 18、8、11，占比 48.65%、21.62%、29.73%，属于合理亚类划分^[5]。

5.2.3 亚类划分的敏感性分析

针对上文建立的亚类划分模型，以下进行敏感性分析，检验方法为将训练集数据的单个化学成分比例进行 5%、10%、20%的扰动调整后再投入测试集查看亚类划分是否产生变化，以下为测试结果：

对于二氧化硅、氧化钾、氧化钙逐一施加 5%增幅扰动，亚类划分受到影响的数

| 受扰成分 | 原始含量 | 原亚类 | 扰动因子 | 扰动后含量 | 扰动后亚类 |
|------|------|-----|------|-------|-------|
| 氧化钙 | 7.12 | 3 | 5% | 7.476 | 1 |

对于二氧化硅、氧化钾、氧化钙逐一施加 10%增幅扰动，亚类划分受到影响的

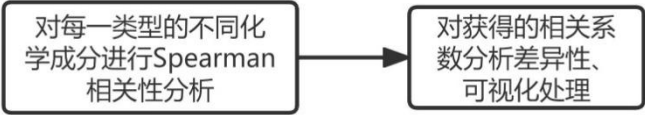
| 受扰成分 | 原始含量 | 原亚类 | 扰动因子 | 扰动后含量 | 扰动后亚类 |
|------|------|-----|------|-------|-------|
| 氧化钙 | 7.12 | 3 | 10% | 7.832 | 1 |

对于二氧化硅、氧化钾、氧化钙逐一施加 20%增幅扰动，亚类划分受到影响的

| 受扰成分 | 原始含量 | 原亚类 | 扰动因子 | 扰动后含量 | 扰动后亚类 |
|------|-------|-----|------|--------|-------|
| 二氧化硅 | 69.33 | 3 | 20% | 83.196 | 2 |
| 氧化钙 | 7.12 | 3 | 20% | 8.544 | 1 |

由上述扰动敏感性检验结果可见，在重要判别因子原本已处于较为偏离聚类中心值时，受到偏心扰动可使亚类划分发生变化，而对于距离聚类中心值较近的样本，当扰动达到了较高水平(如 20%)才会发生亚类划分的变化，说明模型的敏感性较为合适。

5.3 问题三模型的建立与求解



5.3.1 未知文物类型的预测

首先将待鉴别类型的文物根据是否风化分为两组，对于未风化的文物直接采用问题二中构建的随机森林分类模型进行分类，分类结果及预测概率如下表所示：

| 文物编号 | 表面风化 | 预测结果 | 铅钡概率 | 高钾概率 |
|------|------|------|------|------|
| A1 | 未风化 | 高钾 | 0.16 | 0.84 |
| A3 | 未风化 | 铅钡 | 0.85 | 0.15 |
| A4 | 未风化 | 铅钡 | 0.89 | 0.11 |
| A8 | 未风化 | 铅钡 | 0.92 | 0.08 |

表 10 未风化未知文物的类型预测

对于已风化的文物，利用表 2 中已风化的文物的各种化学成分占比的数据，再次构建随机森林分类模型，再利用构建的模型对文物进行分类，分类结果及预测概率如下表所示（模型参数、模型评估结果、测试数据预测评估结果见附录）：

| 文物编号 | 表面风化 | 预测结果 | 铅钡概率 | 高钾概率 |
|------|------|------|------|------|
| A2 | 风化 | 铅钡 | 0.64 | 0.36 |
| A5 | 风化 | 铅钡 | 0.67 | 0.33 |
| A6 | 风化 | 高钾 | 0.05 | 0.95 |
| A7 | 风化 | 高钾 | 0.02 | 0.98 |

表 11 风化未知文物的类型预测

5.3.2 分类结果的敏感性分析

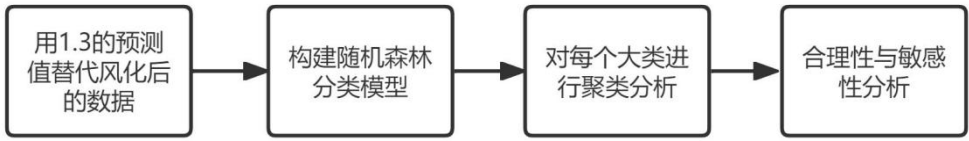
首先，根据随机森林分类模型的测试集数据预测评估结果（见附录）与模型评估结果，模型的准确率、召回率、精确率均较高，故模型的分类效果较好。

| | 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 |

表 12 随机森林模型评估结果

另外，观察表 10、表 11 不难发现每个样本两种类型的预测概率均相差悬殊，故分类结果的敏感性良好。其中高钾玻璃风化后的敏感性明显优于风化前，铅钡玻璃风化前的敏感性明显优于风化后。

5.4 问题四模型的建立与求解



5.4.1 不同类别玻璃文物化学成分之间的关联关系

此问题要求分别分析高钾玻璃和铅钡玻璃任意两种化学成分之间的关系，由于风化后的文物化学成分含量损失较大，对关联关系会产生一定影响，因此采用 Spearman 相关系数对未风化文物以及将风化文物预测前的数据进行分析。

运用 SPSS 软件分析，得出如下表格：

| | 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 | 氧化锆 | 氧化锡 | 二氧化硫 |
|-------|------------------|-----------------|-----------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 二氧化硅 | 1.000(0.000***) | -0.279(0.262) | -0.539(0.021**) | -0.771(0.000***) | -0.117(0.643) | -0.551(0.018**) | -0.553(0.017**) | -0.271(0.278) | 0.060(0.815) | 0.034(0.893) | 0.063(0.804) | -0.175(0.488) | 0.421(0.082*) | 0.038(0.880) |
| 氧化钠 | -0.279(0.262) | 1.000(0.000***) | 0.456(0.057*) | 0.437(0.070*) | -0.575(0.012**) | 0.146(0.564) | -0.221(0.379) | -0.136(0.590) | 0.299(0.228) | -0.216(0.388) | -0.096(0.706) | -0.310(0.211) | 0.260(0.298) | -0.066(0.794) |
| 氧化钾 | -0.539(0.021**) | 0.456(0.057*) | 1.000(0.000***) | 0.750(0.000***) | -0.371(0.129) | -0.183(0.468) | -0.045(0.858) | 0.124(0.624) | -0.158(0.530) | -0.441(0.067*) | -0.527(0.025**) | -0.162(0.521) | -0.186(0.459) | 0.121(0.634) |
| 氧化钙 | -0.771(0.000***) | 0.437(0.070*) | 0.750(0.000***) | 1.000(0.000***) | -0.283(0.254) | 0.239(0.340) | 0.159(0.530) | 0.205(0.416) | -0.170(0.500) | -0.448(0.062*) | -0.426(0.078*) | -0.374(0.126) | -0.435(0.071*) | 0.198(0.430) |
| 氧化镁 | -0.117(0.643) | -0.575(0.012**) | -0.371(0.129) | -0.283(0.254) | 1.000(0.000***) | 0.484(0.042**) | 0.504(0.033**) | 0.006(0.980) | -0.131(0.603) | 0.393(0.106) | 0.466(0.051*) | 0.666(0.003***) | 0.089(0.724) | 0.202(0.423) |
| 氧化铝 | -0.551(0.018**) | 0.146(0.564) | -0.183(0.468) | 0.239(0.340) | 0.484(0.042**) | 1.000(0.000***) | 0.488(0.040**) | -0.030(0.906) | 0.178(0.480) | 0.262(0.294) | 0.288(0.246) | 0.241(0.336) | -0.235(0.348) | -0.025(0.920) |
| 氧化铁 | -0.553(0.017**) | -0.221(0.379) | -0.045(0.858) | 0.159(0.530) | 0.504(0.033**) | 0.488(0.040**) | 1.000(0.000***) | 0.544(0.019**) | -0.039(0.877) | 0.472(0.048**) | 0.303(0.222) | 0.327(0.185) | -0.354(0.150) | 0.029(0.908) |
| 氧化铜 | -0.271(0.278) | -0.136(0.590) | 0.124(0.624) | 0.205(0.416) | 0.006(0.980) | -0.030(0.906) | 0.544(0.019**) | 1.000(0.000***) | 0.041(0.870) | 0.384(0.115) | 0.184(0.465) | 0.052(0.838) | -0.225(0.370) | 0.219(0.384) |
| 氧化铅 | 0.060(0.815) | 0.299(0.228) | -0.158(0.530) | -0.170(0.500) | -0.131(0.603) | 0.178(0.480) | -0.039(0.877) | 0.041(0.870) | 1.000(0.000***) | 0.623(0.006***) | -0.039(0.877) | 0.052(0.838) | 0.186(0.461) | -0.230(0.360) |
| 氧化钡 | 0.034(0.893) | -0.216(0.388) | -0.441(0.067*) | -0.448(0.062*) | 0.393(0.106) | 0.262(0.294) | 0.472(0.048**) | 0.384(0.115) | 0.623(0.006***) | 1.000(0.000***) | 0.447(0.063*) | 0.527(0.025**) | 0.108(0.670) | -0.216(0.388) |
| 五氧化二磷 | 0.063(0.804) | -0.096(0.706) | -0.527(0.025**) | -0.426(0.078*) | 0.466(0.051*) | 0.288(0.246) | 0.303(0.222) | 0.184(0.465) | -0.039(0.877) | 0.447(0.063*) | 1.000(0.000***) | 0.562(0.015**) | 0.489(0.039**) | 0.093(0.712) |
| 氧化锆 | -0.175(0.488) | -0.310(0.211) | -0.162(0.521) | -0.374(0.126) | 0.666(0.003***) | 0.241(0.336) | 0.327(0.185) | 0.052(0.838) | 0.052(0.838) | 0.527(0.025**) | 0.562(0.015**) | 1.000(0.000***) | 0.292(0.240) | -0.081(0.750) |
| 氧化锡 | -0.421(0.082*) | 0.260(0.298) | -0.186(0.459) | -0.435(0.071*) | 0.089(0.724) | -0.235(0.348) | -0.354(0.150) | -0.225(0.370) | 0.186(0.461) | 0.108(0.670) | 0.489(0.039**) | 0.292(0.240) | 1.000(0.000***) | 0.260(0.298) |
| 二氧化硫 | 0.038(0.880) | -0.066(0.794) | 0.121(0.634) | 0.198(0.430) | 0.202(0.423) | -0.025(0.920) | 0.029(0.908) | 0.219(0.384) | -0.230(0.360) | -0.216(0.388) | 0.093(0.712) | -0.081(0.750) | 0.260(0.298) | 1.000(0.000***) |

注：***、**、*分别代表1%、5%、10%的显著性水平

表 13 高钾玻璃化学成分相关系数

| | 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 | 氧化锆 | 氧化锡 | 二氧化硫 |
|-------|------------------|-----------------|-----------------|-----------------|-----------------|------------------|------------------|-----------------|------------------|------------------|-----------------|-----------------|-----------------|-----------------|
| 二氧化硅 | 1.000(0.000***) | -0.099(0.559) | 0.147(0.384) | -0.068(0.688) | 0.104(0.539) | 0.084(0.623) | 0.011(0.946) | -0.072(0.672) | -0.584(0.000***) | -0.389(0.017**) | -0.016(0.927) | -0.417(0.010**) | -0.174(0.304) | 0.047(0.783) |
| 氧化钠 | -0.099(0.559) | 1.000(0.000***) | -0.047(0.784) | -0.097(0.568) | 0.023(0.893) | -0.008(0.963) | -0.334(0.043**) | 0.166(0.327) | -0.073(0.670) | 0.235(0.162) | -0.363(0.027**) | 0.043(0.800) | -0.343(0.037**) | -0.193(0.253) |
| 氧化钾 | 0.147(0.384) | -0.047(0.784) | 1.000(0.000***) | 0.089(0.598) | 0.415(0.011**) | 0.565(0.000***) | 0.230(0.172) | 0.021(0.900) | -0.260(0.120) | 0.004(0.982) | -0.020(0.908) | -0.109(0.520) | 0.069(0.683) | 0.047(0.782) |
| 氧化钙 | -0.068(0.688) | -0.097(0.568) | 0.089(0.598) | 1.000(0.000***) | 0.399(0.014**) | 0.390(0.017**) | 0.387(0.018**) | -0.075(0.660) | -0.115(0.498) | -0.278(0.096*) | 0.505(0.001***) | 0.273(0.103) | 0.378(0.021**) | 0.016(0.927) |
| 氧化镁 | 0.104(0.539) | 0.023(0.893) | 0.415(0.011**) | 0.399(0.014**) | 1.000(0.000***) | 0.572(0.000***) | 0.024(0.888) | 0.008(0.962) | -0.280(0.093*) | -0.198(0.240) | 0.126(0.458) | 0.192(0.256) | 0.122(0.473) | -0.185(0.274) |
| 氧化铝 | 0.084(0.623) | -0.008(0.963) | 0.565(0.000***) | 0.390(0.017**) | 0.572(0.000***) | 1.000(0.000***) | 0.312(0.060*) | 0.037(0.828) | -0.419(0.010***) | -0.171(0.312) | 0.169(0.317) | -0.110(0.518) | 0.081(0.634) | -0.125(0.461) |
| 氧化铁 | 0.011(0.946) | -0.334(0.043**) | 0.230(0.172) | 0.387(0.018**) | 0.024(0.888) | 0.312(0.060*) | 1.000(0.000***) | -0.362(0.028**) | 0.067(0.693) | -0.430(0.008***) | 0.272(0.104) | -0.082(0.629) | 0.338(0.041**) | -0.199(0.239) |
| 氧化铜 | -0.072(0.672) | 0.166(0.327) | 0.021(0.900) | -0.075(0.660) | 0.008(0.962) | 0.037(0.828) | -0.362(0.028**) | 1.000(0.000***) | -0.382(0.020**) | 0.356(0.031**) | 0.349(0.034**) | -0.061(0.719) | -0.348(0.035**) | 0.219(0.192) |
| 氧化铅 | -0.584(0.000***) | -0.073(0.670) | -0.260(0.120) | -0.115(0.498) | -0.280(0.093*) | -0.419(0.010***) | 0.067(0.693) | -0.382(0.020**) | 1.000(0.000***) | -0.082(0.628) | -0.225(0.180) | 0.329(0.047**) | 0.225(0.181) | -0.109(0.520) |
| 氧化钡 | -0.389(0.017**) | 0.235(0.162) | 0.004(0.982) | -0.278(0.096*) | -0.198(0.240) | -0.171(0.312) | -0.430(0.008***) | -0.082(0.628) | -0.082(0.628) | 1.000(0.000***) | -0.133(0.434) | 0.094(0.579) | 0.037(0.826) | 0.109(0.520) |
| 五氧化二磷 | -0.016(0.927) | -0.363(0.027**) | -0.020(0.908) | 0.505(0.001***) | 0.126(0.458) | 0.169(0.317) | 0.272(0.104) | 0.349(0.034**) | -0.225(0.180) | -0.133(0.434) | 1.000(0.000***) | -0.064(0.708) | -0.009(0.958) | 0.112(0.510) |
| 氧化锆 | -0.417(0.010**) | 0.043(0.800) | -0.109(0.520) | 0.273(0.103) | 0.192(0.256) | -0.110(0.518) | -0.082(0.629) | -0.061(0.719) | 0.329(0.047**) | 0.094(0.579) | -0.064(0.708) | 1.000(0.000***) | 0.219(0.193) | 0.126(0.459) |
| 氧化锡 | -0.174(0.304) | -0.343(0.037**) | 0.069(0.683) | 0.378(0.021**) | 0.122(0.473) | 0.081(0.634) | 0.338(0.041**) | -0.348(0.035**) | 0.225(0.181) | 0.037(0.826) | -0.009(0.958) | 0.219(0.193) | 1.000(0.000***) | -0.049(0.771) |
| 二氧化硫 | 0.047(0.783) | -0.193(0.253) | 0.047(0.782) | 0.016(0.927) | -0.185(0.274) | -0.125(0.461) | -0.199(0.239) | 0.219(0.192) | -0.109(0.520) | 0.109(0.520) | 0.112(0.510) | 0.126(0.459) | -0.049(0.771) | 1.000(0.000***) |

注：***、**、*分别代表1%、5%、10%的显著性水平

表 14 铅钡玻璃化学成分相关系数

其中图表包括了相关系数，显著性 P 值，由图分析可知^[4]：

- 不同类别玻璃文物化学成分之间的存在关联关系，部分化学成分之间存在明显的关联关系，其中包括正相关和负相关；部分化学成分之间的关联关系不明显。
- 对于高钾玻璃，在 1%的显著性水平下，二氧化硅和氧化钙呈显著的负相关性，氧化钾和氧化钙呈显著的正相关性，氧化镁和氧化锆呈显著的正相关性；在 5%的显著性水平下，二氧化硅和氧化钾、氧化铝、氧化铁呈负相关性，氧化钠和氧化镁呈负相关性，氧化钾和五氧化二磷呈负相关性，氧化镁、氧化铝、氧化铁两两间呈正相关性，氧化铁和氧化铜、氧化钡呈正相关性，氧

化钡和氧化锶呈正相关性，五氧化二磷和氧化锶、氧化锡呈正相关性。

3. 对于铅钡玻璃，在 1%的显著性水平下，二氧化硅和氧化铅呈显著负相关性，氧化钾和氧化铝呈显著正相关性，氧化钙和五氧化二磷呈显著正相关性，氧化镁和氧化铝呈显著正相关性，氧化铝和氧化铅呈显著负相关性，氧化铁和氧化钡呈显著负相关性；在 5%的显著性水平下，二氧化硅和氧化钡、氧化锶呈负相关性，氧化钠和氧化铁、五氧化二磷、氧化锡呈负相关性，氧化钾和氧化镁呈正相关性，氧化钙和氧化镁、氧化铝、氧化铁、氧化锡呈正相关性，氧化铁和氧化铜呈负相关性，氧化铁和氧化锡呈正相关性，氧化铜和氧化铅、氧化锡呈负相关性，氧化铜和氧化钡、五氧化二磷呈正相关性，氧化铅和氧化锶呈正相关性。

5. 4. 2 不同类别化学成分关联关系的差异性

为比较化学成分关联关系的差异性，应针对不同类型玻璃的化学元素之间的关联系数，分析其关联系数的大小、正负反映差异性。

应用 excel 表格对相关系数进行色阶处理可视化，得到以下两个表：

| | 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 | 1.000 | -0.279 | -0.539 | -0.771 | -0.117 | -0.551 | -0.553 | -0.271 | 0.060 | 0.034 | 0.063 | -0.175 | 0.421 | 0.038 |
| 氧化钠 | -0.279 | 1.000 | 0.456 | 0.437 | -0.575 | 0.146 | -0.221 | -0.136 | 0.299 | -0.216 | -0.096 | -0.310 | 0.260 | -0.066 |
| 氧化钾 | -0.539 | 0.456 | 1.000 | 0.750 | -0.371 | -0.183 | -0.045 | 0.124 | -0.158 | -0.441 | -0.527 | -0.162 | -0.186 | 0.121 |
| 氧化钙 | -0.771 | 0.437 | 0.750 | 1.000 | -0.283 | 0.239 | 0.159 | 0.205 | -0.170 | -0.448 | -0.426 | -0.374 | -0.435 | 0.198 |
| 氧化镁 | -0.117 | -0.575 | -0.371 | -0.283 | 1.000 | 0.484 | 0.504 | 0.006 | -0.131 | 0.393 | 0.466 | 0.666 | 0.089 | 0.202 |
| 氧化铝 | -0.551 | 0.146 | -0.183 | 0.239 | 0.484 | 1.000 | 0.488 | -0.030 | 0.178 | 0.262 | 0.288 | 0.241 | -0.235 | -0.025 |
| 氧化铁 | -0.553 | -0.221 | -0.045 | 0.159 | 0.504 | 0.488 | 1.000 | 0.544 | -0.039 | 0.472 | 0.303 | 0.327 | -0.354 | 0.029 |
| 氧化铜 | -0.271 | -0.136 | 0.124 | 0.205 | 0.006 | -0.030 | 0.544 | 1.000 | 0.041 | 0.384 | 0.184 | 0.052 | -0.225 | 0.219 |
| 氧化铅 | 0.060 | 0.299 | -0.158 | -0.170 | -0.131 | 0.178 | -0.039 | 0.041 | 1.000 | 0.623 | -0.039 | 0.052 | 0.186 | -0.230 |
| 氧化钡 | 0.034 | -0.216 | -0.441 | -0.448 | 0.393 | 0.262 | 0.472 | 0.384 | 0.623 | 1.000 | 0.447 | 0.527 | 0.108 | -0.216 |
| 五氧化二磷 | 0.063 | -0.096 | -0.527 | -0.426 | 0.466 | 0.288 | 0.303 | 0.184 | -0.039 | 0.447 | 1.000 | 0.562 | 0.489 | 0.093 |
| 氧化锶 | -0.175 | -0.310 | -0.162 | -0.374 | 0.666 | 0.241 | 0.327 | 0.052 | 0.052 | 0.527 | 0.562 | 1.000 | 0.292 | -0.081 |
| 氧化锡 | 0.421 | 0.260 | -0.186 | -0.435 | 0.089 | -0.235 | -0.354 | -0.225 | 0.186 | 0.108 | 0.489 | 0.292 | 1.000 | 0.260 |
| 二氧化硫 | 0.038 | -0.066 | 0.121 | 0.198 | 0.202 | -0.025 | 0.029 | 0.219 | -0.230 | -0.216 | 0.093 | -0.081 | 0.260 | 1.000 |

表 15 高钾玻璃化学成分相关系数

| | 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 | 1.000 | -0.099 | 0.147 | -0.068 | 0.104 | 0.084 | 0.011 | -0.072 | -0.584 | -0.389 | -0.016 | -0.417 | -0.174 | 0.047 |
| 氧化钠 | -0.099 | 1.000 | -0.047 | -0.097 | 0.023 | -0.008 | -0.334 | 0.166 | -0.073 | 0.235 | -0.363 | 0.043 | -0.343 | -0.193 |
| 氧化钾 | 0.147 | -0.047 | 1.000 | 0.089 | 0.415 | 0.565 | 0.230 | 0.021 | -0.260 | 0.004 | -0.020 | -0.109 | 0.069 | 0.047 |
| 氧化钙 | -0.068 | -0.097 | 0.089 | 1.000 | 0.399 | 0.390 | 0.387 | -0.075 | -0.115 | -0.278 | 0.505 | 0.273 | 0.378 | 0.016 |
| 氧化镁 | 0.104 | 0.023 | 0.415 | 0.399 | 1.000 | 0.572 | 0.024 | 0.008 | -0.280 | -0.198 | 0.126 | 0.192 | 0.122 | -0.185 |
| 氧化铝 | 0.084 | -0.008 | 0.565 | 0.390 | 0.572 | 1.000 | 0.312 | 0.037 | -0.419 | -0.171 | 0.169 | -0.110 | 0.081 | -0.125 |
| 氧化铁 | 0.011 | -0.334 | 0.230 | 0.387 | 0.024 | 0.312 | 1.000 | -0.362 | 0.067 | -0.430 | 0.272 | -0.082 | 0.338 | -0.199 |
| 氧化铜 | -0.072 | 0.166 | 0.021 | -0.075 | 0.008 | 0.037 | -0.362 | 1.000 | -0.382 | 0.356 | 0.349 | -0.061 | -0.348 | 0.219 |
| 氧化铅 | -0.584 | -0.073 | -0.260 | -0.115 | -0.280 | -0.419 | 0.067 | -0.382 | 1.000 | -0.082 | -0.225 | 0.329 | 0.225 | -0.109 |
| 氧化钡 | -0.389 | 0.235 | 0.004 | -0.278 | -0.198 | -0.171 | -0.430 | 0.356 | -0.082 | 1.000 | -0.133 | 0.094 | 0.037 | 0.109 |
| 五氧化二磷 | -0.016 | -0.363 | -0.020 | 0.505 | 0.126 | 0.169 | 0.272 | 0.349 | -0.225 | -0.133 | 1.000 | -0.064 | -0.009 | 0.112 |
| 氧化锶 | -0.417 | 0.043 | -0.109 | 0.273 | 0.192 | -0.110 | -0.082 | -0.061 | 0.329 | 0.094 | -0.064 | 1.000 | 0.219 | 0.126 |
| 氧化锡 | -0.174 | -0.343 | 0.069 | 0.378 | 0.122 | 0.081 | 0.338 | -0.348 | 0.225 | 0.037 | -0.009 | 0.219 | 1.000 | -0.049 |
| 二氧化硫 | 0.047 | -0.193 | 0.047 | 0.016 | -0.185 | -0.125 | -0.199 | 0.219 | -0.109 | 0.109 | 0.112 | 0.126 | -0.049 | 1.000 |

表 16 铅钡玻璃化学成分相关系数

其中表采用的色阶为红-白-蓝，即相关系数越接近 1，颜色越偏向红色；相关系数越接近-1，颜色越偏向蓝色。由图分析可知：

1. 不同类型玻璃的化学元素之间的关联系数存在差异性。
2. 高钾玻璃中氧化钾和氧化钙的正相关性系数最大，为 0.75；其次是氧化镁和氧化锶、五氧化二磷和氧化锶。铅钡玻璃中氧化镁和氧化铝的正相关性系数最大，为 0.572；其次是氧化钾和氧化铝、氧化钙和五氧化二磷。
3. 高钾玻璃中二氧化硅和氧化钙的负相关性系数最小，为-0.771；其次是氧化钠和氧化镁、二氧化硅和氧化铁。铅钡玻璃中二氧化硅和氧化铝的负相关性系数最小，为-0.584；其次是氧化铁和氧化钡、氧化铝和氧化铅。

六、模型的评价、改进与推广

6.1 模型的优点

针对不同的问题(例如预测、相关性检验、亚类划分),基于 SPSS、Excel 等软件灵活使用了诸如卡方检验、kappa 检验、正态性检验直方图、K-means 聚类划分、随机森林算法等数据处理方法,对不同的问题有较强的针对性。

选用多种方法,最终择优建模。例如问题三对未知文物分类,可以建立 BP 神经网络、使用逻辑回归、随机森林等方法,通过横向比较择优,排除过拟合、明显错误的情况,最终选择随机森林作为最优建模方案。

对于构建的模型,进行了合理性与敏感性的分析,确保了模型的可推广性与普适性,可以适用于除给定数据集以外的情景使用。

对数据进行了合理的预处理,数据清洗剔除错误数据,数据集成将有效信息进行合并以便分析,数据变换将表面是否风化的数据进行归一化排除干扰。

6.2 模型的缺点

由于样本数量较少且存在不规范数据,对于模型的预测和估计并不完全准确,且风化前后可以形成严格对照的样本数量不足,由此建立起的模型可能会不够准确。

预测风化前含量的模型是通过均值法建立的,且仍然存在一定方差,因此可能会受到部分极端数据的影响而不够准确。

6.3 模型的改进

若能基于规模更大、更为规范的原始数据,则可以进行更多次迭代的机器学习,提升诸如随机森林算法的精确度,且风化前后数据充足则可以利用线性回归等方法更好地拟合预测风化前的数据。

七、参考文献

- [1]陆运清. 用 Pearson's 卡方统计量进行统计检验时应注意的问题[J]. 统计与决策, 2009 (15): 32-33.
- [2]唐万, 胡俊, 张晖,等. Kappa 系数:一种衡量评估者间一致性的常用方法(英文)[J]. 上海精神医学(Shanghai Archives of Psychiatry), 2015(1):62-67.
- [3]郭超,陆新建.工业过程数据中缺失值处理方法的研究[J].计算机工程与设计,2010,31(06):1351-1354.DOI:10.16208/j.issn1000-7024.2010.06.055.
- [4]Maurice G. Kendall, "The treatment of ties in ranking problems", Biometrika Vol. 33, No. 3, pp. 239-251. 1945.
- [5] Jain, A. K. , and R. C. Dubes . "Algorithms for clustering data." Technometrics 32.2(1988):227-229.

附录

附录 1

介绍：支撑材料的文件列表

支撑材料

└—代码

1.1 代码_Kappa 一致性检验.py

1.1 代码_卡方校验.py

1.2 代码_Spearman 相关性分析.py

1.2 代码_分类汇总.py

1.3 代码_正态性检验.py

2.1 代码_随机森林分类.py

2.2_代码_聚类分析.py

注：上述代码均为 python 语言编写，通过调用 SPSS 库实现数据分析。

└—图表

1.2_Spearman 相关系数（风化与化学成分）.xlsx

1.3_化学成分_预测后.xlsx

1.3_铅钡_未风化_正态性检验.xlsx

1.3_铅钡_风化_正态性校验.xlsx

1.3_高钾_未风化_正态性校验.xlsx

1.3_高钾_风化_正态性校验.xlsx

2.1_未风化_随机森林分类_模型参数.xlsx

2.2_铅钡_k-means 聚类分析.xlsx

2.2_铅钡_聚类树状图.png

2.2_高钾_k-means 聚类分析.xlsx

2.2_高钾_聚类树状图.png

3.1_风化_随机森林分类.xlsx

4.1_铅钡_相关系数.xlsx

4.1_高钾_相关系数.xlsx