

# 基于机器学习与关联网络的古代玻璃成分分析

## 摘要

古代玻璃的化学成分是鉴别其类别与产地的重要依据，然而风化作用会改变原始成分，为定量分析带来挑战。本文基于一批古代玻璃制品的化学成分数据，建立了一套多层次的数学模型，以解决风化影响下的玻璃分类、成分预测与内在关联分析问题。研究综合运用了统计检验、机器学习、智能优化算法与复杂网络理论，为通过化学成分数据理解古代玻璃制品提供了系统性的分析方案。

对于问题一，我们采用卡方检验分析了表面风化与玻璃类型、纹饰及颜色的关联性，确认了它们之间存在统计学关联。通过对风化前后样本的化学成分分布，发现风化主要导致高钾玻璃中的氧化钾与铅钡玻璃中的氧化铅和氧化钡流失。为预测风化前的成分，本文引入地球化学领域的质量平衡分析理论，构建了基于风化系数的预测模型，该模型在缺乏成对样本的情况下，能够有效恢复文物的原始化学成分。

对于问题二，为研究两大类玻璃的分类规律，我们分别构建了线性与非线性支持向量机模型。其中，非线性模型获得了 97.41% 的交叉验证准确率，我们引入博弈论中的 SHAP 值对其进行解释，而线性模型的结果则直接验证了氧化铅与氧化钾是分类的核心指标。在亚类划分中，我们依据变异系数筛选特征，并结合层次聚类与轮廓系数，确定铅钡玻璃存在两个亚类，高钾玻璃存在五个亚类。划分结果显示铅钡玻璃可分为高铅钡助熔剂型与高硅基质型，高钾玻璃的亚类则在多种成分上表现出不同特征。

对于问题三，我们构建了基于改进遗传算法优化的支持向量机分类器 IGA-SVM，以鉴别未知类别文物的所属类型。该模型通过智能化全局寻优确定最优超参数组合，避免了传统模型选择的局限性。应用此模型，我们完成了对全部未知样本的分类。为确保结果的可靠性，我们设计了基于蒙特卡洛模拟的数据扰动与基于参数网格搜索的模型扰动双重灵敏度分析，验证了分类结果对于数据测量误差和模型参数选择的高度稳健性。

对于问题四，为探寻不同类别玻璃配方的内在结构差异，我们构建了化学成分关联网络。为消除成分数据总和恒定带来的虚假相关，我们首先采用化学计量学中的中心化对数比变换对数据进行处理。随后利用图套索算法计算偏相关系数以构建网络，并应用鲁汶算法进行社群发现。网络分析结果表明，铅钡玻璃与高钾玻璃的化学成分关联结构存在显著不同，前者以二氧化硅和氧化铅为核心形成紧密社群，后者则结构较为分散，这些差异反映了两者在原料与烧制工艺上的区别。

**关键字：** 卡方检验 地球化学 质量平衡分析理论 SHAP 值 支持向量机 中心化对数比 图套索算法 鲁汶算法

## 一、问题背景

古代玻璃制品是早期中西方文化与贸易交流的重要物证。古代玻璃的主要成分为二氧化硅，为降低其熔点，制造过程中常需加入助熔剂。以草木灰为助熔剂得到高钾玻璃，而以铅矿石为助熔剂则得到铅钡玻璃，二者因制作工艺的差异导致化学成分构成不同，这也成为区分玻璃来源的依据。因此，通过分析玻璃文物样本的化学成分，能够为考古学研究提供有价值的参考信息。

玻璃制品在长期埋藏过程中会发生风化现象，其内部元素与环境发生交换，导致表面化学成分比例改变。风化作用的存在为依据化学成分鉴别玻璃类型带来了挑战。部分文物表面风化严重，其现有成分数据已不能直接反映其初始状态。如何利用已有数据，建立风化影响下玻璃分类的数学模型，并对风化过程进行定量分析，是本研究需要解决的核心科学问题。

## 二、问题重述

**问题一：**分析表面风化与玻璃物理属性的关联，研究风化对不同类型玻璃化学成分的影响规律，并构建模型预测文物风化前的成分含量。

**问题二：**建立高钾玻璃与铅钡玻璃的分类模型，并对两大类玻璃进行亚类划分，同时分析划分结果的合理性与敏感性。

**问题三：**应用已建立的分类模型鉴别未知类别玻璃文物的所属类型，并进行敏感性分析。

**问题四：**分析不同类别玻璃内部各化学成分间的关联模式，并比较这些模式在不同类别间的差异。

## 三、问题分析

针对问题一，该问题涉及两个层面，其一是定性关系的判断，其二是定量规律的分析与预测。表面风化与玻璃类型、纹饰、颜色均为分类变量，它们之间的关联性分析适合采用非参数的卡方检验。风化对化学成分含量的影响则需要分组进行统计比较，通过观察含量分布的变化来识别规律。由于缺乏同一文物风化前后的成对数据，直接建立回归预测模型存在困难，因此分析的重点在于依据风化机理，构建一个基于平均变化率的风化系数模型，用以反推风化前的化学成分。

针对问题二，该问题需要从监督学习与无监督学习两个角度展开。高钾与铅钡玻璃的分类规律研究是一个典型的二分类问题，可以基于探索性数据分析发现的关键化学成

分，构建支持向量机等分类模型，并对模型进行解释。亚类划分则是在已知大类的基础上进行更细致的探索，属于无监督聚类问题，需要先选择类内变异性高的化学成分作为特征，再运用层次聚类与轮廓系数等方法确定最优的亚类数量，最后对划分结果进行化学特征解释与敏感性分析。

针对问题三，该问题是分类模型的实际应用与可靠性验证。分析的核心在于构建一个泛化能力强且稳健的分类器。这需要比较多种模型架构，并论证选择支持向量机的合理性。为使模型性能最优化，需要采用改进遗传算法等智能优化算法进行超参数寻优。最终的鉴别结果需要通过数据扰动和模型参数敏感性分析的双重检验，以证明其结论并非偶然，而是具有高度的可靠性。

针对问题四，该问题要求探寻不同类别玻璃内部化学成分的关联结构。由于玻璃成分数据具有总和恒定的特性，直接计算相关系数会产生误导，因此必须首先采用中心化对数比变换来消除此统计约束。变换之后，应使用图套索等方法计算能反映直接关联的偏相关系数，并构建网络。通过对网络进行社群发现与核心节点分析，可以将复杂的关联关系简化为宏观结构，从而比较两类玻璃在配方原理上的差异。

## 四、模型假设

为保证模型构建的合理性与分析过程的严谨性，我们提出以下基本假设。

1. 所有用于分析的数据均为成分总和在百分之八十五至一百零五区间的有效数据。
2. 数据中的空白项表示该化学成分未被检测到其含量可视为零。
3. 同一类型玻璃的风化过程与化学成分变化规律具有统计上的一致性。
4. 文物的化学成分组合能够有效地区分高钾玻璃与铅钡玻璃两大类别。
5. 同一文物多个采样点的平均化学成分可以代表该文物的整体特征用于亚类划分。
6. 化学成分间的偏相关关系能够反映其在玻璃配方或制作工艺中的直接关联。

## 五、符号说明

本文后续章节中所使用的主要数学符号及其说明如表所示。

表 1 符号说明表

符号	说明
$k_{t,j}$	$t$ 类玻璃中 $j$ 成分的风化系数
$\bar{C}_{t,j,\text{weathered}}$	$t$ 类玻璃风化样本中 $j$ 成分的平均含量
$\bar{C}_{t,j,\text{unweathered}}$	$t$ 类玻璃未风化样本中 $j$ 成分的平均含量
$C'_{\text{unweathered},j}$	某一样本风化前 $j$ 成分的预测含量
$w, b$	支持向量机分类超平面的法向量与位移项
$C$	支持向量机的正则化系数
$\gamma$	径向基核函数的参数
$\xi_i$	支持向量机的松弛变量
$CV$	变异系数，用于衡量数据的相对离散程度
$\text{clr}(x)$	对样本向量 $x$ 进行中心化对数比变换
$S$	样本协方差矩阵
$\Theta$	精度矩阵，即逆协方差矩阵
$\lambda$	图套索算法的正则化参数
$\rho_{ij,\text{rest}}$	变量 $i$ 和 $j$ 之间的偏相关系数

## 六、数据预处理与分析

为了后续模型的建立与求解,本文进行如下数据预处理,即数据清洗、数据量化等。具体机制如图1所示。

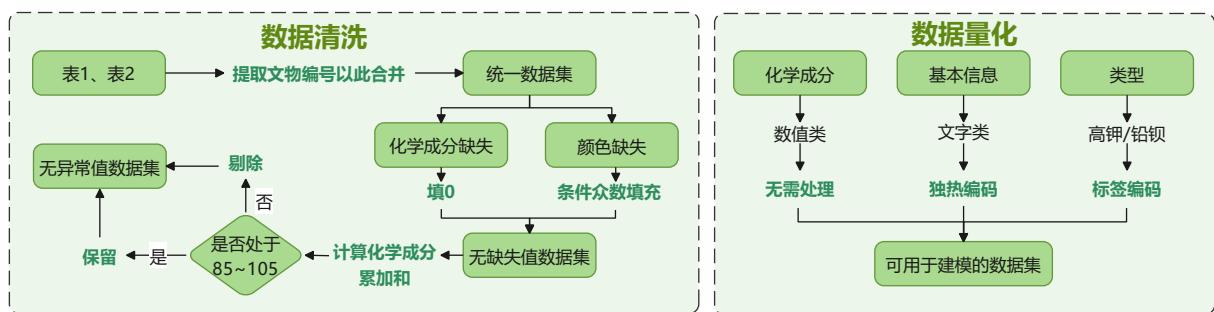


图 1 数据预处理机制

## 6.1 数据清洗

本文首先进行了数据清洗。题中给出的附件包含三个表单，表单 1 包含文物编号、纹饰、类型、颜色和表面风化情况，表单 2 则包含了文物采样点和化学成分比例。通过观察发现，表单 2 的“文物采样点”列前缀的数字部分与表单 1 的文物编号相对应，因

此我们通过正则表达式提取出其数字部分，创建出“文物编号”列，并基于此将表单1内容合并进来，形成包含完整信息的数据集。

随后我们对数据集中各变量进行缺失值统计，发现颜色和化学成分比例存在显著的缺失问题。题目中提到表单2中化学成分比例若存在空白处表示未检测到该成分，因此我们将这些缺失值填充为0，处理前后的化学成分频数如图2所示。

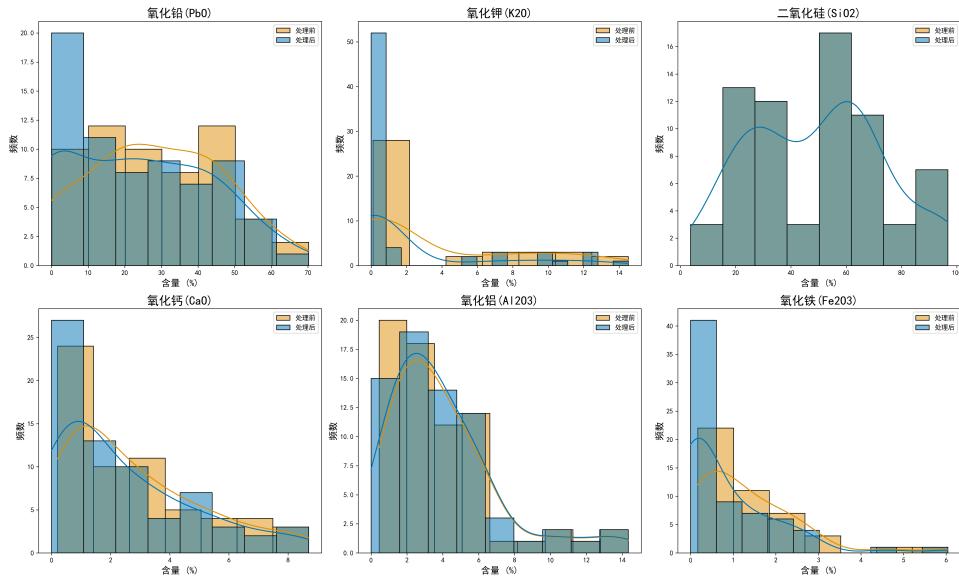


图2 化学成分缺失值填充

而对于颜色缺失，我们首先想到通过全局众数填充，但是其误差可能较大，于是我们采用了一种条件众数填充法。具体而言，对于一个颜色信息缺失的样本，系统会自动寻找数据集中所有与该样本在类型、表面风化和纹饰上完全一致的其他样本，然后用这个小群体中出现次数最多的颜色来填充缺失值。其填充内容如图??所示。

表2 按类型和纹饰划分的颜色众数及对应文物编号

类型	纹饰	颜色众数	对应文物编号
铅钡	A	浅蓝	2, 19, 20, 23, 28, 29, 30, 42, 44, 45, 46, 47, 48, 49, 50, 53
铅钡	C	浅蓝	8, 11, 24, 25, 26, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 51, 52, 54, 55, 56, 57, 58
高钾	A	蓝绿	3, 4, 5, 6, 18, 21
高钾	B	蓝绿	7, 9, 10, 12, 22, 27
高钾	C	浅蓝	1, 13, 14, 15, 16, 17

针对各种化学成分比例，本题中将成分比例累加和介于85%~105%之间的数据视为有效数据，因此我们对其进行成分性异常值检测。我们计算每个样本所有化学成分的累

加和，并筛选出总和介于 85% 到 105% 之间的数据作为有效样本，得到 67 条有效已分类样本、8 条有效未分类样本以及 2 条异常样本，然后剔除异常样本。

最后我们检测了高钾玻璃和铅钡玻璃以及它们中风化和未风化的占比，如图3所示：如图3所示，在有效样本中，高钾与铅钡玻璃的比例约为 1:2.7，风化与无风化的比例接

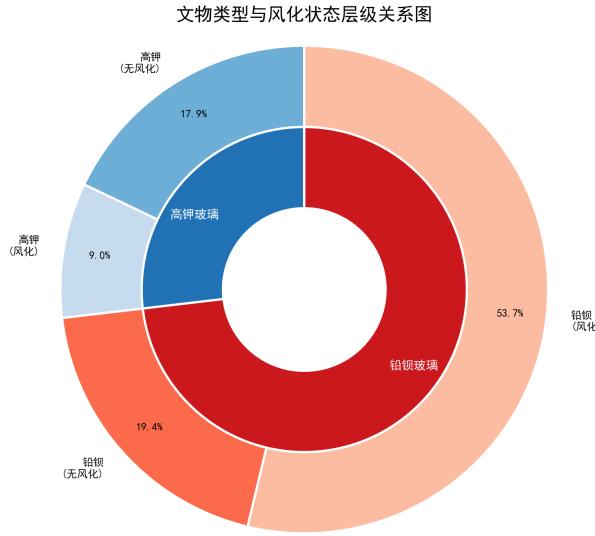


图 3 类型与风化嵌套饼图

近 1:1.4。总体来看，各类样本数量分布相对均衡，不存在极端情况，因此在后续建模中无需采用欠采样或过采样等特殊处理。

数据清洗的整体效果如图4所示。

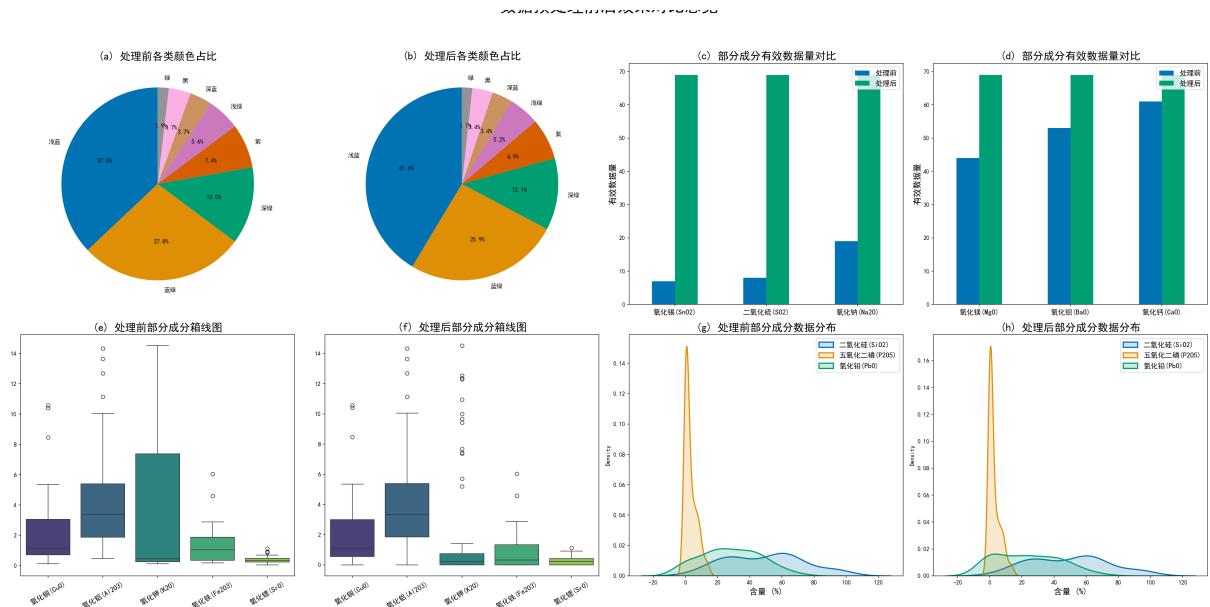


图 4 数据清洗效果

## 6.2 数据量化

为了便于后续相关数学模型的建立与求解，增强可理解性，本文进一步将合并后的文件中的名称进行数据量化。对于作为因变量的“类型”列，我们使用标签编码，将“铅钡”映射为0，“高钾”映射为1。对于作为自变量的各项分类特征列，为了避免引入虚假的、不存在的顺序和距离关系，我们排除了顺序编码和RGB向量的方案，最终选择了最能保留数据本身特性的独热编码。数据量化的具体细节如表3所示：

表3 数据量化方法示例

列名	原始数据形态	量化方法	量化后形态
类型	[高钾, 铅钡]	标签编码	[1,0]
表面风化	[风化, 无风化]	独热编码	[1,0]
颜色	[蓝绿, 浅蓝, ...]	独热编码	[00000001,00000010,...]
纹饰	[A,B,C]	独热编码	[001,010,100]
各类化学成分比例	如 69.33	无需处理	69.33

## 七、问题一：风化影响分析与成分恢复模型

古代玻璃文物在长期埋藏过程中会发生风化作用，导致其表面化学成分发生改变。为探究风化作用对不同类别玻璃文物的影响，并尝试恢复其风化前的成分含量，本章建立了分析与预测模型。我们首先检验了表面风化与文物其他物理属性之间的统计关联，然后分析了风化对高钾和铅钡两类玻璃化学成分含量的影响规律，最后构建了一个基于风化系数的预测模型，用于推断风化样品的原始化学成分，其框架如图5所示。

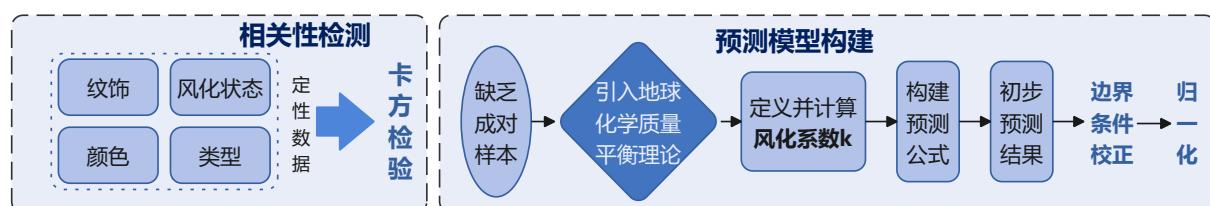


图5 风化影响分析与成分恢复模型框架

### 7.1 表面风化与文物物理属性的关联性检验

为了探究文物表面风化现象是否与其物理属性存在关联，我们首先需要分析表面风化状态与玻璃类型、纹饰及颜色这几个变量之间的关系。这些变量的共同特征是它们均为分类变量，其取值为离散的类别而非连续的数值。

这一数据特性决定了用于衡量连续变量间线性关系的皮尔逊相关系数或用于比较组间均值差异的方差分析等方法在此并不适用，因为对“浅蓝”、“纹饰 A”等类别进行数值运算不具备实际意义，需要采用一种能够处理定性数据频数的非参数检验方法来分析它们之间的关联性。因此，针对此问题，我们使用了卡方检验。

卡方检验是一种专门用于判断两个或多个分类变量之间是否存在关联的经典统计方法，其核心思想在于比较观测频数与期望频数之间的差异。其中，观测频数是样本数据中各组合的实际计数值，而期望频数则是在“变量间相互独立”这一零假设下，根据边际概率计算出的理论计数值。检验过程通过计算两者差异的卡方统计量，并将其转换为  $P$  值来进行判断。在本研究中，我们设定显著性水平为 0.05，若计算所得的  $P$  值小于该阈值，则拒绝变量间相互独立的零假设，认为它们之间存在显著的统计学关联。

为分析文物表面风化现象与其物理属性的关联，我们制作了关系分析的可视化图，如图6所示。图中第一部分展示了风化状态在两类玻璃中的分布情况。数据显示，铅钡玻璃的风化样本占其总数的 73.5%，这一比例远高于高钾玻璃的 33.3%。图中第二部分与第三部分则进一步展示了风化样本在不同纹饰和颜色类别下的数量分布。所有纹饰为 B 的样本均为风化样本，而纹饰为 A 的样本中风化与未风化数量相同。在不同颜色中，浅蓝色样本的风化数量为 20，远超其未风化数量 6，而蓝绿色样本中两者的数量则基本持平。这些在不同类别下风化比例与数量的显著差异直观地表明表面风化与文物的物理属性并非相互独立。

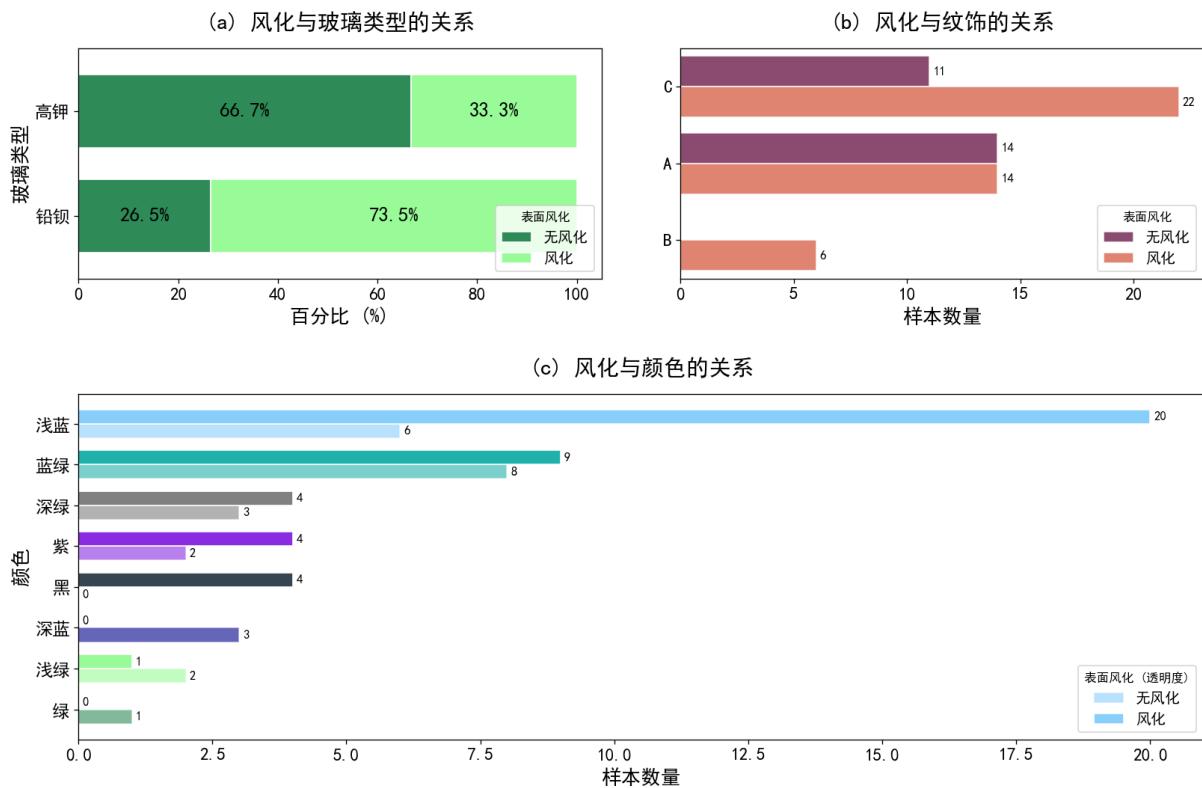


图 6 表面风化与类型、纹饰及颜色的关系可视化

## 7.2 风化对两类玻璃化学成分含量的影响规律

基于风化与玻璃类型存在关联的结论，我们进一步对风化在高钾和铅钡两类玻璃中引起的化学成分变化规律进行分析。我们将样本数据分为高钾未风化、高钾风化、铅钡未风化、铅钡风化四个组别，并对各组样本的化学成分含量分布进行了比较。

我们采用分面箱线图对两类玻璃在风化前后的化学成分分布进行可视化，如图7与图8所示。箱线图展示了数据的中位数、四分位距和离散程度。从图中可以观察到，对于高钾玻璃，风化作用导致氧化钾  $K_2O$  的含量中位数显著下降，而氧化硅  $SiO_2$  的含量则有上升趋势。对于铅钡玻璃，风化作用主要表现为氧化铅  $PbO$  与氧化钡  $BaO$  含量的大幅降低，同时氧化硅  $SiO_2$  含量相应增加。这种变化说明风化过程中发生了元素的选择性流失与富集。

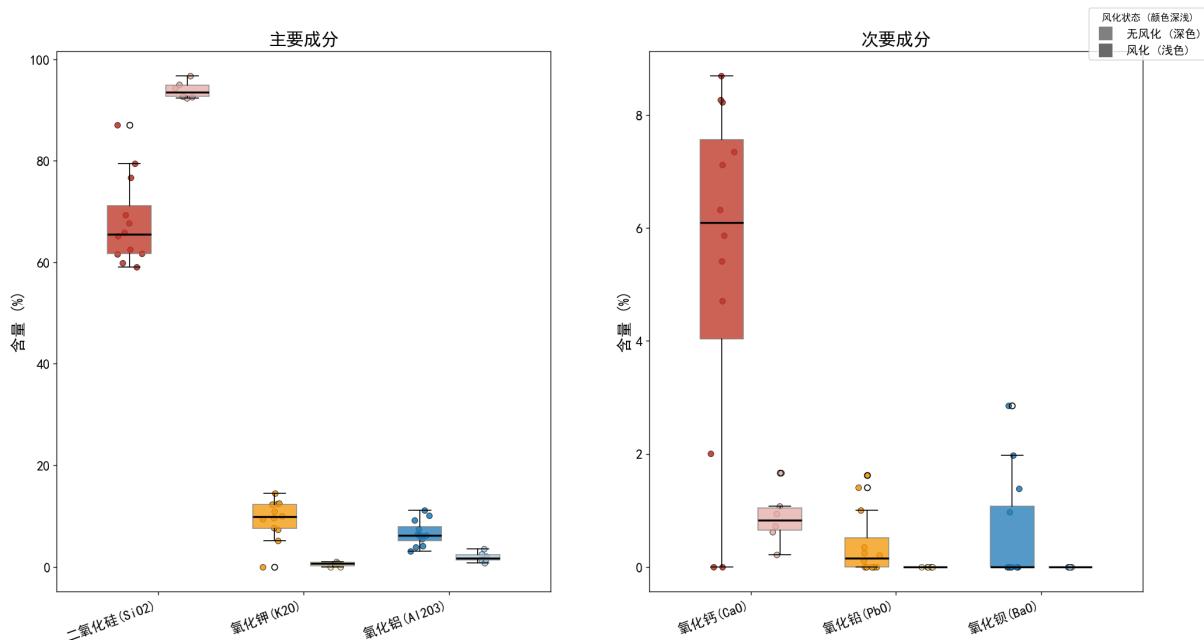


图 7 高钾玻璃在风化前后各化学成分含量分布

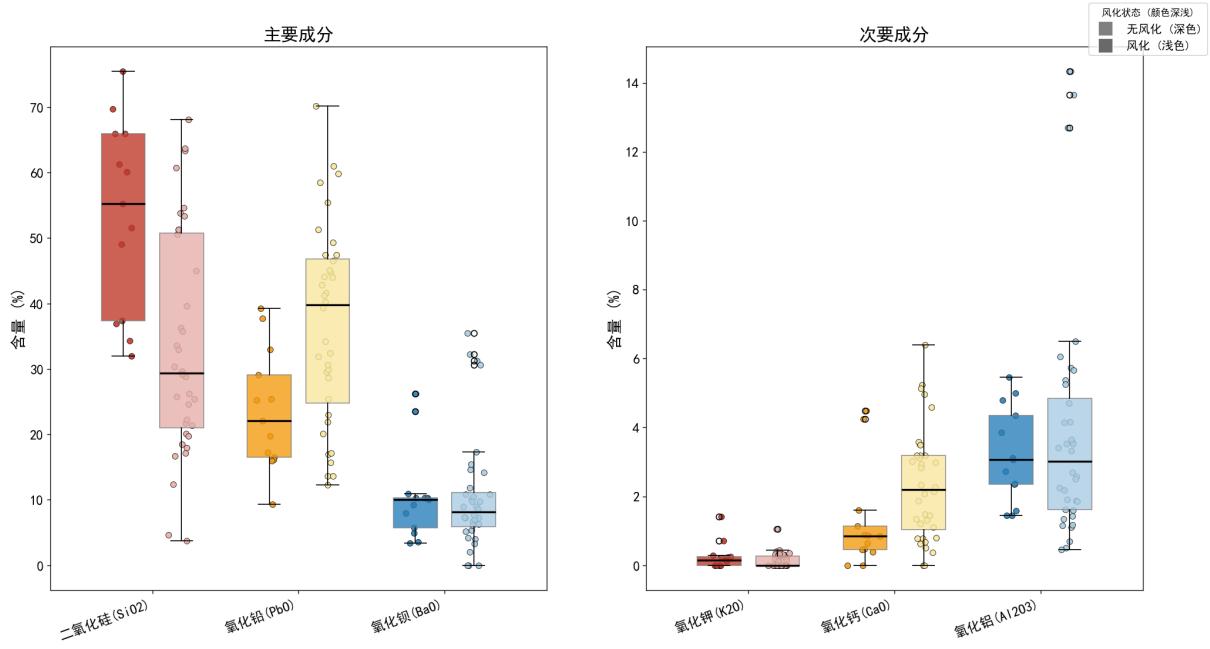


图 8 铅钡玻璃在风化前后各化学成分含量分布

为了更细致地观察关键化学成分的分布形态变化，我们绘制了部分核心化学成分，包括氧化铅  $PbO$ 、氧化钾  $K_2O$ 、氧化钡  $BaO$  以及二氧化硅  $SiO_2$  的分布图，如图9至图12所示。图中包含直方图与核密度估计曲线，它们共同描述了数据分布的集中趋势和形态。分析这些分布图可以发现，高钾玻璃在风化后，其氧化钾  $K_2O$  的含量分布从一个较宽的区间转化至接近零值的极低水平。对于铅钡玻璃，风化作用使其特征成分氧化铅  $PbO$  与氧化钡  $BaO$  的含量分布整体向低值区移动。与此相反，作为玻璃基体的二氧化硅  $SiO_2$ ，其含量分布在两类玻璃中均表现出向高值区偏移的趋势，说明在风化过程中其他元素的流失导致了二氧化硅的相对富集。

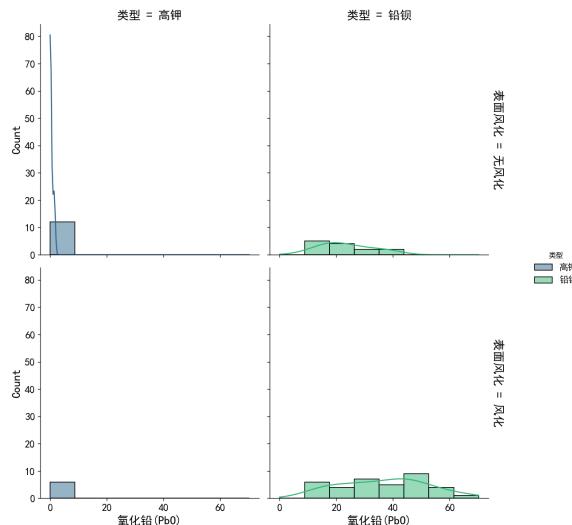


图 9 氧化铅  $PbO$  含量分布

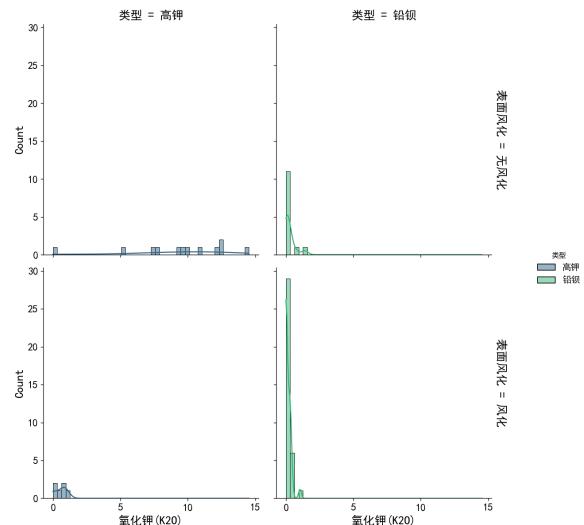


图 10 氧化钾  $K_2O$  含量分布

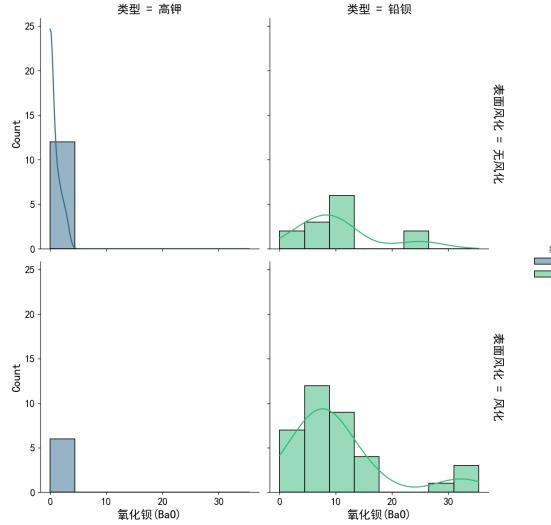


图 11 氧化钡  $BaO$  含量分布

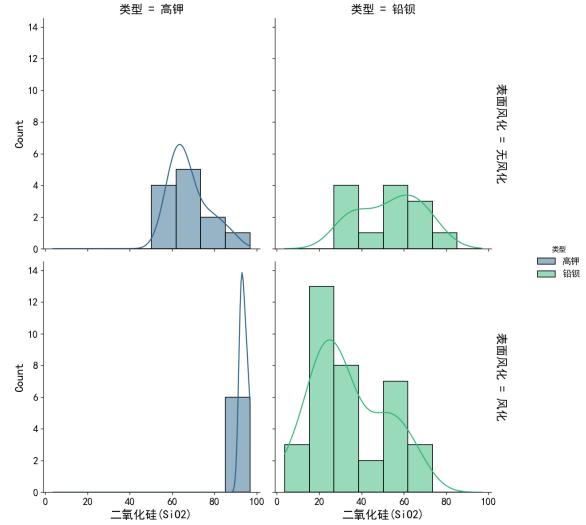


图 12 二氧化硅  $SiO_2$  含量分布

### 7.3 基于风化系数的化学成分含量预测模型

预测风化前化学成分的主要困难在于数据中缺乏源自同一文物的风化与未风化的观测样本。这使得依赖大量成对样本进行训练的传统监督学习模型，例如回归分析，难以直接应用且存在较高的过拟合风险。

于是，我们转向从风化过程的物理化学机理中寻求建模依据。玻璃的风化过程与地质学中岩石的蚀变过程在原理上具有相似性，均为长期化学环境作用下的元素迁移过程。因此，我们引入了地球化学领域成熟的质量平衡分析理论来构建预测模型，该理论能够在缺乏直接演变过程数据时，对成分变化进行有效推断。

我们的模型的核心假设为：特定化学成分在风化过程中的流失或富集比例，与其在未风化状态下的原始含量相关。此假设参考了地质学中分析交代蚀变作用的艾索康图法的原理，它将风化视为一个系统性的化学变化过程而非随机过程，从而建立了基于风化系数的预测方法。

我们首先为每种玻璃类型  $t$  和每种化学成分  $j$  定义一个风化系数  $k_{t,j}$ 。该系数由该类型玻璃中所有风化样本与未风化样本的平均含量计算得出，其数学表达式如下：

$$k_{t,j} = 1 - \frac{\bar{C}_{t,j,\text{weathered}}}{\bar{C}_{t,j,\text{unweathered}}} \quad (1)$$

其中， $\bar{C}_{t,j,\text{weathered}}$  表示  $t$  类玻璃风化样本中  $j$  成分的平均含量， $\bar{C}_{t,j,\text{unweathered}}$  表示  $t$  类玻璃未风化样本中  $j$  成分的平均含量。

利用计算得到的风化系数，我们可以对任意一个已知风化后成分含量  $C_{\text{weathered},j}$  的

样本，进行其风化前含量  $C'_{\text{unweathered},j}$  的初步预测，其预测公式为：

$$C'_{\text{unweathered},j} = \frac{C_{\text{weathered},j}}{1 - k_{t,j}} \quad (2)$$

考虑到测量误差和模型的局限性，初步预测得到的各成分总和可能偏离 100%。因此，我们对预测结果进行了边界条件校正。我们计算所有预测成分的总和，并进行条件归一化处理，以确保最终得到的预测成分总和落入题目要求的 85% 至 105% 的有效区间内，从而使预测结果在化学上更具合理性。模型的最终预测结果，以原始含量和预测含量并列对比的形式，被完整地保存至 Result/问题一\_ 预测结果.xlsx 文件中。

## 八、问题二：玻璃文物分类、亚类划分及可靠性分析

本部分的目标是根据文物的化学成分，建立高钾玻璃与铅钡玻璃的分类规律，并对两类玻璃进行内部亚类的划分，最后对所得结论的可靠性进行系统性分析。其框架如图13所示。

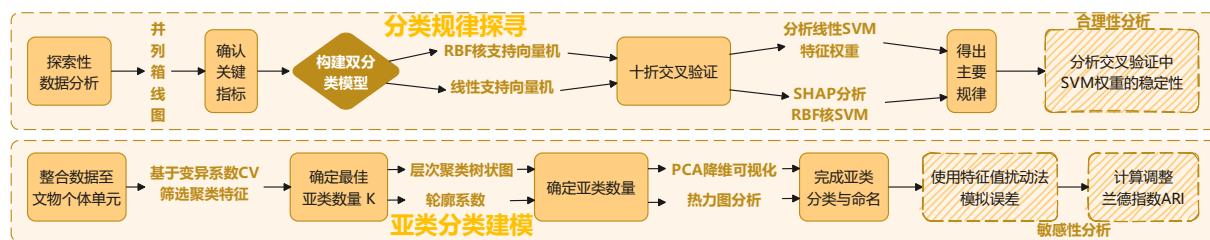


图 13 问题二框架

### 8.1 高钾玻璃与铅钡玻璃的分类规律

在构建分类模型之前，首先需要通过探索性数据分析对不同类别的数据在数值特征上的分布差异建立直观认知，这为后续的建模方向和假设形成提供依据。我们采用并列箱线图来直观对比高钾玻璃与铅钡玻璃在全部十四种化学成分含量上的分布差异。

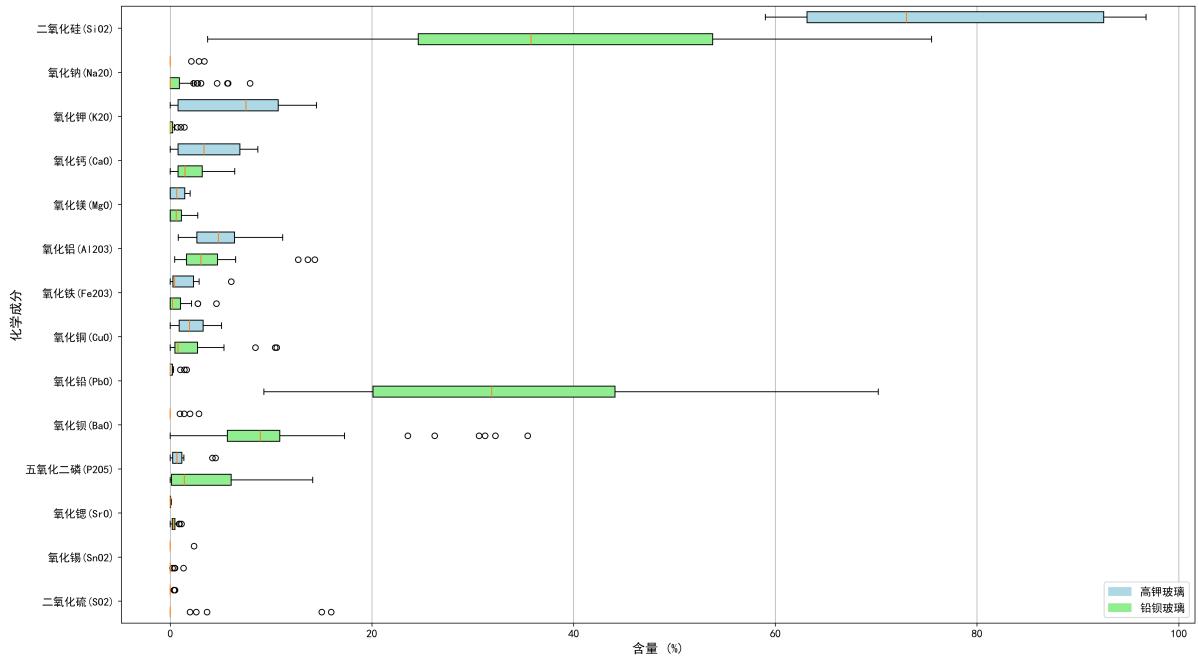


图 14 高钾玻璃与铅钡玻璃化学成分含量分布对比

如图14所示，图中纵轴为化学成分，横轴为含量百分比。对于每一种成分，均并排展示了两个箱体，分别代表高钾玻璃和铅钡玻璃的含量分布。由图可知，在氧化铅  $PbO$  与氧化钡  $BaO$  两种成分上，铅钡玻璃的含量分布远高于高钾玻璃，而在氧化钾  $K_2O$  上则呈现相反的模式。两个类别在这些关键成分上的分布区间几乎没有重叠，由此形成核心假设，即  $PbO$ 、 $BaO$  与  $K_2O$  是区分两类玻璃的关键指标。

基于前述探索性分析形成的假设，即特定化学成分含量可有效区分玻璃类型，我们构建监督学习分类模型以对该规律进行量化验证。为兼顾模型的解释性与对数据复杂关系的拟合能力，我们分别建立了线性支持向量机与使用径向基函数核的非线性支持向量机模型。

线性支持向量机的基本原理是在一个多维特征空间中，寻找一个最优的分类超平面。该超平面由法向量  $w$  和位移  $b$  定义，其数学形式为  $w^T x + b = 0$ 。最优化的目标是不仅能将两类样本分开，同时能最大化距离超平面最近的样本，即支持向量，到超平面的间隔。在允许部分样本被错误分类以增强模型泛化能力的情况下，该问题可表述为一个软间隔优化问题，其目标函数如下

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \quad (3)$$

式中  $C$  为正则化系数，用于平衡间隔最大化与分类误差， $\xi_i$  是松弛变量，允许样本点在一定程度上偏离其正确的分类边界。

当化学成分与玻璃类型的关系无法通过线性边界有效分离时，需要引入非线性模

型。非线性支持向量机通过核技巧将原始特征空间映射到一个更高维度的空间，并在新空间中构造线性超平面。我们选用径向基函数核，即高斯核，作为映射函数，其定义为

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (4)$$

其中  $\gamma$  是核函数的一个参数，它决定了单个训练样本影响范围的大小。通过使用该核函数，模型能够在原始空间中形成复杂的非线性决策边界。

为获得稳健的模型性能评估，避免因单次数据划分带来的偶然性，我们采用十折交叉验证方法。该方法将数据集随机分为十个互斥的子集，轮流使用其中九个作为训练集，一个作为测试集，最终以十次测试结果的平均值作为模型的性能度量。评估结果显示，线性支持向量机模型的平均分类准确率为 95.71%，而使用径向基函数核的非线性支持向量机模型平均分类准确率达到 97.41%。两个模型均表现出较高的分类能力。

为解释模型的分类机理，我们首先分析了线性支持向量机学习到的特征权重。

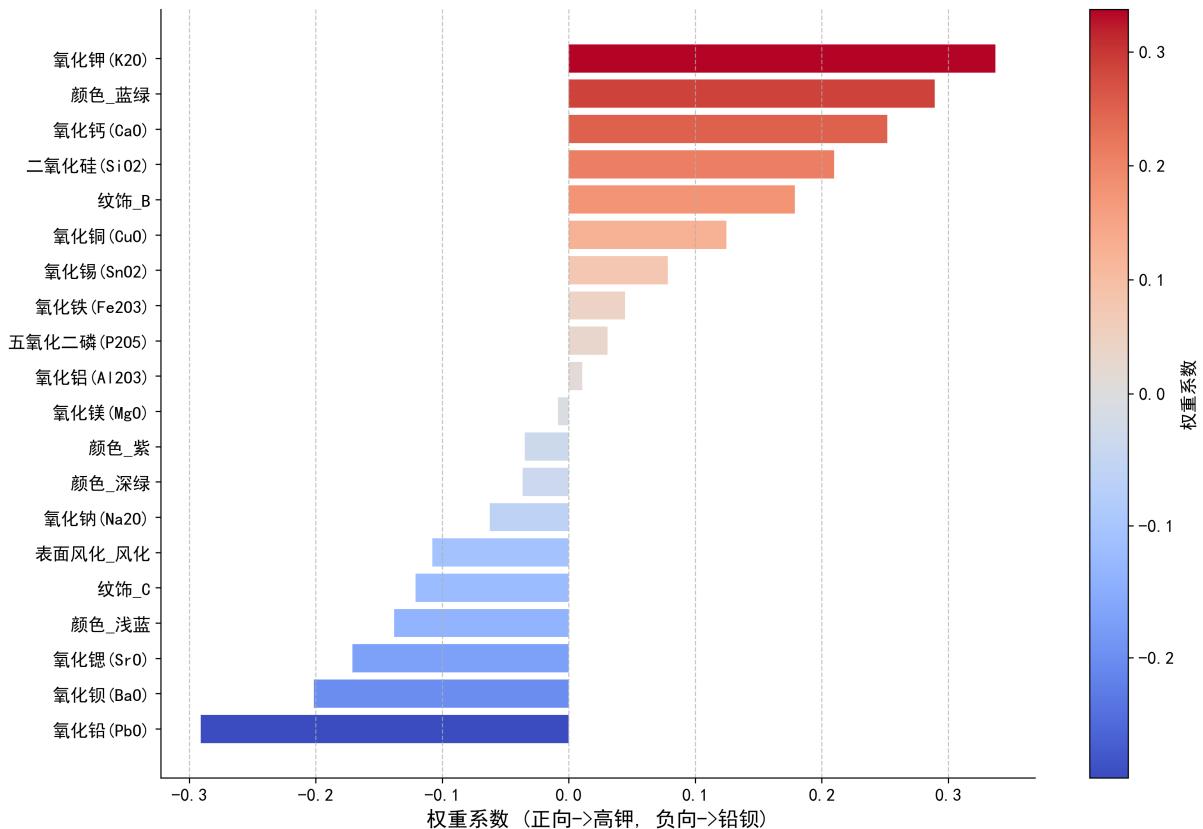


图 15 线性支持向量机模型学习的特征权重

图15展示了各化学成分的权重系数值，其中正向权重将预测推向高钾玻璃，负向权重则推向铅钡玻璃。图中可见，氧化钾  $K_2O$  与二氧化硅  $SiO_2$  获得了最大的正权重，而氧化铅  $PbO$  与氧化钡  $BaO$  获得了最大的负权重，这与探索性分析的发现一致。

对于性能更高但解释性较弱的非线性模型，我们采用 SHAP，即 SHapley Additive

exPlanations 框架进行分析。该方法基于合作博奕论，能够将模型的预测结果公平地归因到每个输入特征上。

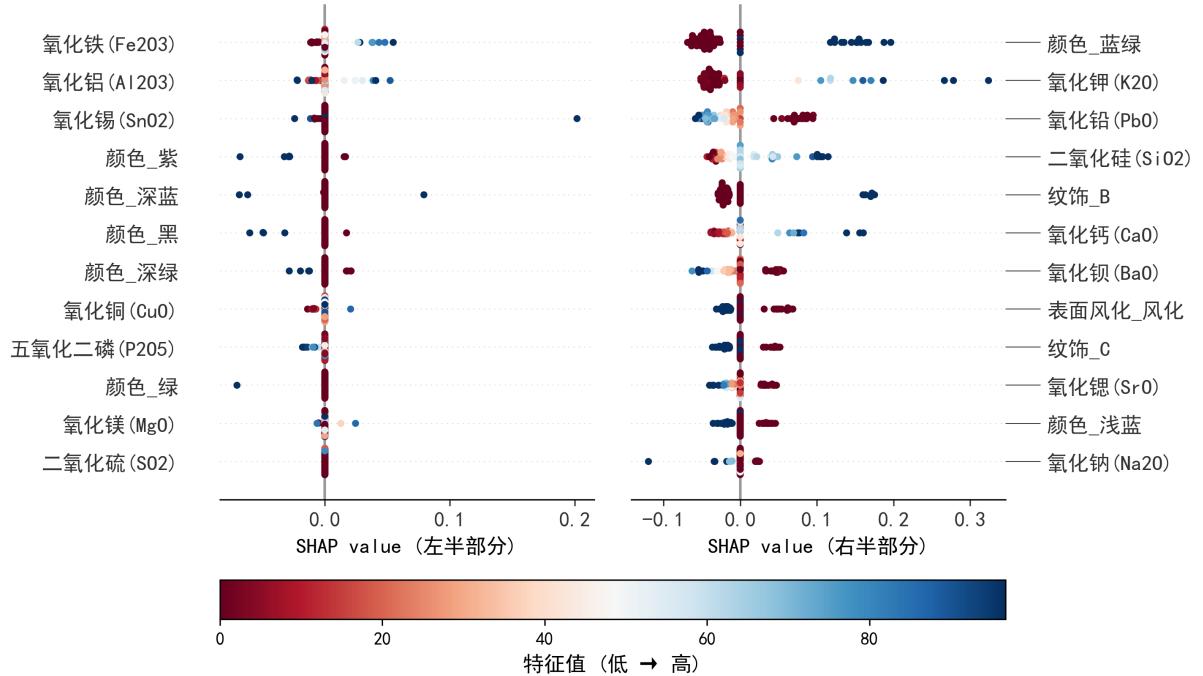


图 16 非线性支持向量机模型的 SHAP 全局特征重要性分析

图16展示了全局特征重要性，纵轴按重要性高低排列特征，横轴是 SHAP 值，点的颜色代表该特征原始值的大小。该图再次确认  $PbO$  与  $K_2O$  是最重要的特征，并且提供了更丰富的细节，高含量的  $PbO$  红点总是对应着强烈的负向 SHAP 值，而高含量的  $K_2O$  红点则总是对应着强烈的正向 SHAP 值。

## 8.2 各类别内部的亚类划分

亚类划分的目标是探索同一类别文物之间是否存在基于化学成分的更细微群组。因此，分析的最小单元必须是文物个体，而非单个采样点。我们首先整合数据，对同一未风化文物的多个采样点取化学成分平均值，以消除采样位置带来的随机误差。对于风化文物，则优先选取能代表其风化状态的采样点数据进行平均。若风化文物缺乏此类数据，则采用分级条件均值插补，依据类型、颜色、纹饰的相似性为其估算一个风化后的化学成分。此过程最终生成了一个包含三十四个独立文物，且成分状态一致的数据集。

在进行聚类前，需要筛选出适合用于划分亚类的化学成分。亚类划分旨在寻找类别内部的差异，因此变异系数  $CV$  成为一个合适的指标。它是一个无量纲的相对离散度量，其定义如下

$$CV = \frac{\sigma}{\mu} \quad (5)$$

其中  $\sigma$  是标准差， $\mu$  是均值。变异系数消除了不同化学成分自身含量均值大小的影响，

能更公平地比较各特征的真实变异程度。我们据此为高钾和铅钡玻璃分别筛选出类内变异程度最高的化学成分作为聚类特征。为确定各类玻璃内部可能存在的亚类数量，我们首先采用探索性的层次聚类方法。该方法无需预设类别数量，通过计算样本间的距离，将化学成分最相似的样本逐级合并，其过程可以通过树状图进行可视化。

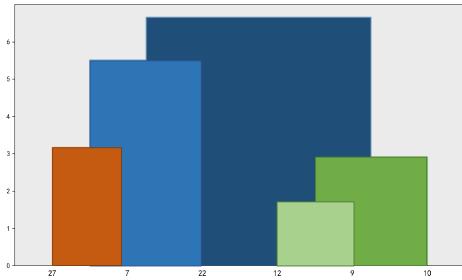


图 17 高钾玻璃层次聚类树状图

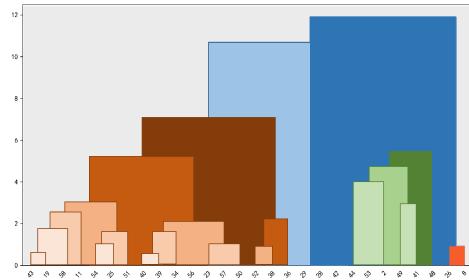


图 18 铅钡玻璃层次聚类树状图

图18展示的铅钡玻璃树状图呈现出清晰的两分支结构，两个主分支在较高的距离水平上才发生合并，这初步表明铅钡玻璃样本可自然地分为两个大类。相比之下，图17中高钾玻璃的树状图结构更为复杂，呈现出多层次的嵌套聚合模式，表示其内部可能存在更多数量的亚类。

为对亚类数量的选择提供定量依据，我们引入轮廓系数指标。该指标综合评估每个样本与其所属簇的相似度即内聚性，以及与其他簇的差异度即分离度。轮廓系数值越接近一，表示聚类结果的结构越合理。

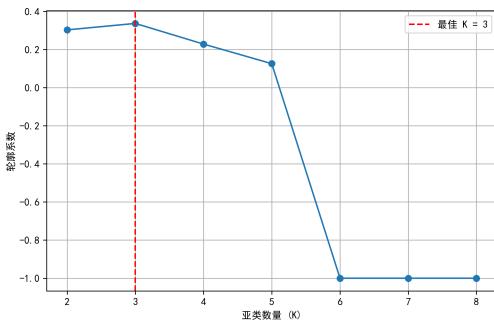


图 19 高钾玻璃轮廓系数随  $K$  值的变化

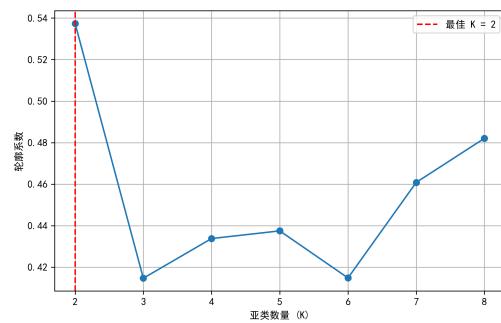


图 20 铅钡玻璃轮廓系数随  $K$  值的变化

如图20所示，对于铅钡玻璃，轮廓系数在  $K = 2$  时取得最大值，约为零点四八，当  $K$  值继续增加时，系数值出现明显下降。对于高钾玻璃，如图19所示，轮廓系数在  $K = 5$  时达到峰值，约为零点五二。综合定性观察与定量计算，我们将铅钡玻璃的亚类数量确定为二，高钾玻璃的亚类数量确定为五。

为验证划分结果的有效性并解读各亚类的化学特征，我们首先利用主成分分析方法将高维的化学成分数据投影到二维平面上。

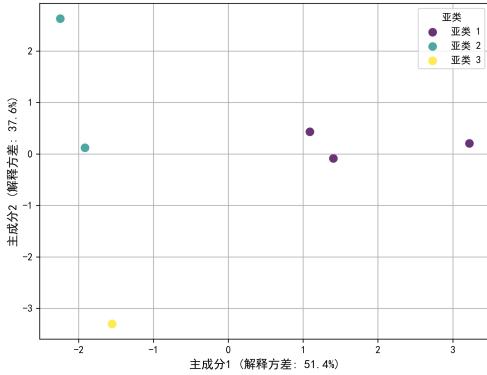


图 21 高钾玻璃亚类划分 PCA 可视化

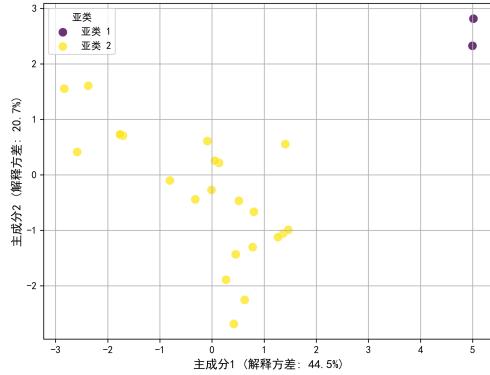


图 22 铅钡玻璃亚类划分 PCA 可视化

图21与图22的可视化结果显示，不同颜色的点代表不同亚类的样本。铅钡玻璃的两个亚类在第一主成分轴上具有显著的分离。高钾玻璃的五个亚类也在二维空间中占据了相对独立的区域，各亚类内部样本较为集中，而亚类之间存在明显界限。

为进一步阐释每个亚类所代表的化学成分模式，我们绘制了亚类化学特征热力图。

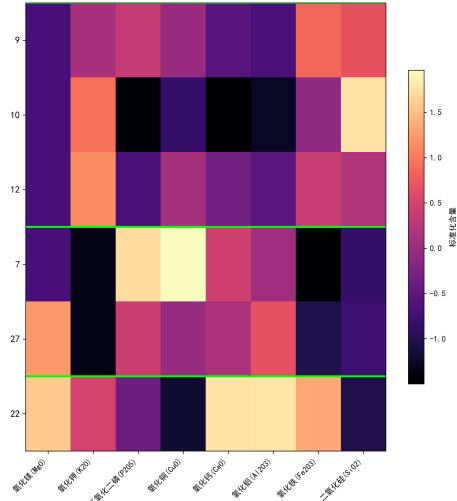


图 23 高钾玻璃亚类化学特征热力图

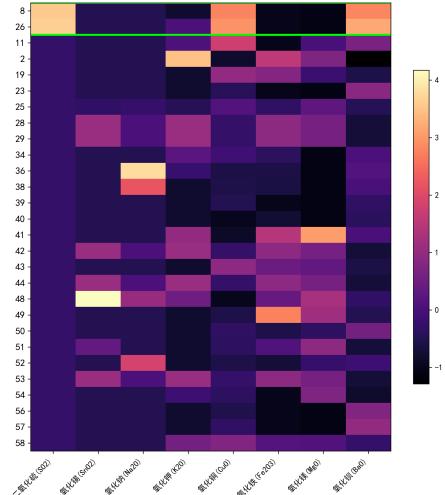


图 24 铅钡玻璃亚类化学特征热力图

图24展示了铅钡玻璃两个亚类的化学差异。亚类一的样本在氧化铅  $PbO$  与氧化钡  $BaO$  两种助熔剂成分上呈现深色，表明其含量普遍较高，而作为玻璃基体的二氧化硅  $SiO_2$  含量则相对较低。与此相反，亚类零的样本在二氧化硅  $SiO_2$  上呈现深色，含量普遍较高，而氧化铅  $PbO$  与氧化钡  $BaO$  含量则较低。基于此，可将亚类一命名为高铅钡助熔剂型，亚类零命名为高硅基质型。

图23则展现了高钾玻璃五个亚类更为细微的化学特征。亚类二的突出特征是其氧化钾  $K_2O$  含量极高，而其他成分含量较低。亚类一的氧化钙  $CaO$  含量相对突出。亚类零

则表现为氧化铝  $Al_2O_3$  与氧化铁  $Fe_2O_3$  含量较高，这可能与其他亚类使用了不同的矿物原料有关。亚类三与亚类四的差异主要体现在磷与硫等微量元素上，反映了更为精细的原料或工艺差别。通过热力图分析，我们明确了每个亚类独特的化学成分特征。

### 8.3 结果的敏感性分析

为验证上述分类与划分结果的稳健性，我们进行了敏感性分析。首先，我们检验分类规律的可靠性。我们在十折交叉验证的每一次折叠中，重新训练线性支持向量机模型并提取其权重，以评估规律本身的稳定性。

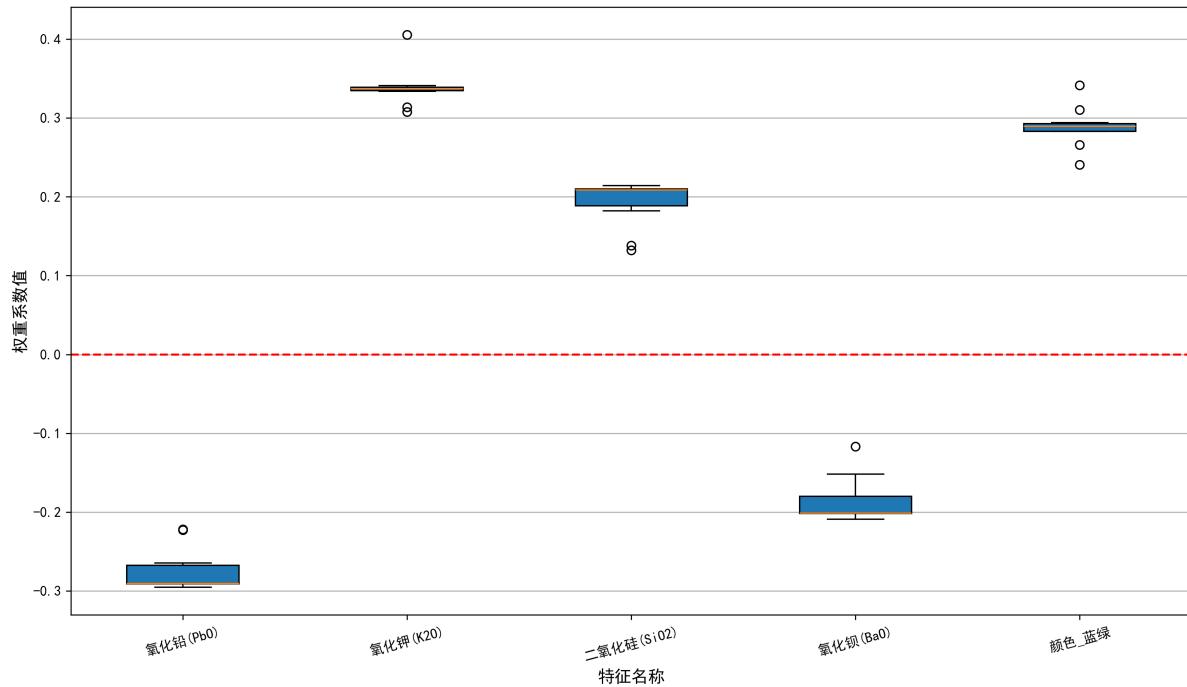


图 25 关键特征的 SVM 权重在十次交叉验证中的稳定性

图25通过箱线图展示了关键特征的权重在十次不同数据子集训练中的分布。图中显示，关键特征如  $PbO$ 、 $K_2O$  的权重符号在十次实验中从未改变，且波动范围很小。这证明了我们提炼出的分类规律是高度稳健的。

其次，我们检验亚类划分结果的敏感性。为检验划分结果对化学成分测量误差的容忍度，我们采用了特征值扰动法。该方法通过向数据中注入不同水平的随机噪声，来模拟测量误差，并检验聚类结构的稳定性。我们对筛选出的特征数据乘以一个范围在  $[1 - p, 1 + p]$  内的随机扰动因子，其中  $p$  为扰动水平。在此扰动数据上重新聚类，并使用调整兰德指数  $ARI$  来衡量该次聚类结果与原始基准结果的一致性。

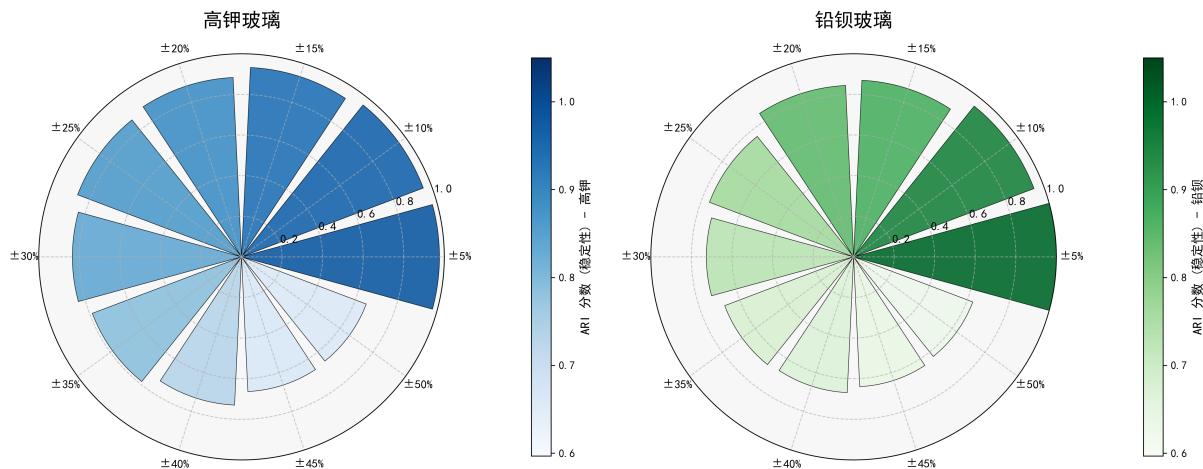


图 26 不同扰动水平下亚类划分结果的 ARI 分数

图26展示了在不同扰动水平下，高钾和铅钡玻璃亚类划分的平均 ARI 分数。对于高钾玻璃，即使在百分之十的扰动下，平均 ARI 分数依然高达 0.9513。对于铅钡玻璃，在百分之十的扰动下，平均 ARI 分数为 0.9596。在面临潜在的测量误差时，其核心划分依然能够保持高度一致。这证明了我们发现的亚类结构是真实且稳健的，而非随机产生的现象。

## 九、问题三：文物类别的鉴别与模型稳健性分析

本章的核心任务是应用分类模型对未知类别玻璃文物的化学成分进行分析，确定其所属类型，并对分类结果的稳健性与可靠性进行系统性验证。为完成此任务，我们首先进行了一系列对比实验以选择最合适的模型架构，随后采用改进遗传算法对所选模型的超参数进行寻优，构建最终的分类器。最后，我们使用该模型进行预测，并通过双重灵敏度分析来检验结论的稳定性。其整体框架如图27所示。

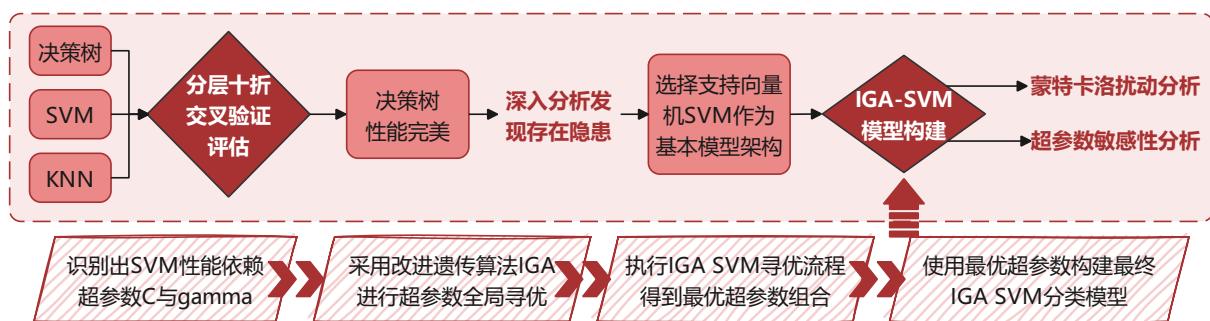


图 27 分类模型框架

## 9.1 分类模型的选择与论证

为建立有效的分类规律，需要从多种候选模型中筛选出最适宜本数据特性的算法。我们选取决策树，K 近邻算法以及支持向量机三种具有代表性的模型，在包含全部十四种化学成分及表面风化状况的完整特征集上进行初步性能评估。评估过程采用分层十折交叉验证方法，以保证每次训练与测试中样本类别的分布与原始数据保持一致。各模型的性能指标如表4所示。

表 4 初步模型性能评估

模型	准确率	F1 分数
决策树	1.0000	1.0000
SVM	0.9714	0.9576
KNN	0.9857	0.9788

评估结果表明，决策树模型在所有性能指标上均达到了 1.0 的满分。为探究决策树模型取得此结果的原因，我们对其内部结构进行了分析。通过在完整的已分类数据集上训练单个决策树模型，我们发现其特征重要性得分几乎全部集中于氧化铅  $PbO$  这一项上。模型的可视化结构进一步确认了此发现，如图28所示，该决策树仅在根节点依据氧化铅  $PbO$  含量是否大于一个特定阈值便完成了对所有样本的分类。

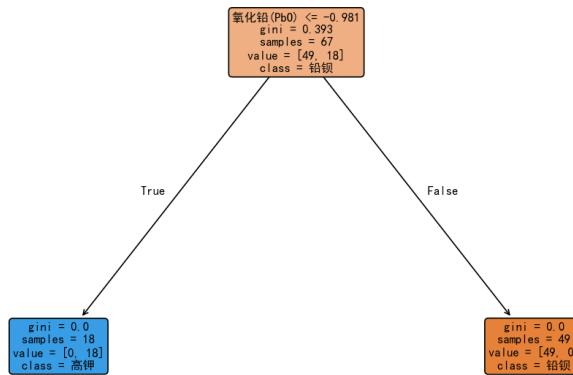


图 28 初步决策树模型结构

尽管决策树表现出很高的评分，但其过分依赖单一特征的决策模式存在稳健性隐患。一个仅凭氧化铅含量进行判断的模型，在面对不含氧化铅但仍属铅钡玻璃体系的未

知样本时可能失效，其泛化能力不足。同时，由于该特征的存在，不同模型间的性能差异被掩盖，使得模型对比失去参考价值。基于这些考量，我们排除了将决策树作为最终模型的方案，转而寻求更能利用多维度信息且结构更为稳健的模型。在移除氧化铅特征后进行的补充实验中，支持向量机表现出优越的适应能力，因此我们选择支持向量机作为最终模型的基本架构。

## 9.2 基于改进遗传算法的支持向量机模型构建

支持向量机是一种在高维空间中寻找最优分类超平面的算法，其性能表现高度依赖于正则化系数  $C$  和核函数系数  $\gamma$  这两个超参数的设定。为充分发挥支持向量机的性能，我们采用改进遗传算法 IGA 来代替传统的网格搜索，对其超参数组合进行智能化全局寻优。

支持向量机的基本原理可表述为求解一个软间隔优化问题，其目标函数如下

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \quad (6)$$

式中  $w$  与  $b$  定义了分类超平面， $\xi_i$  是允许样本偏离正确边界的松弛变量，而  $C$  则用以平衡间隔最大化与分类误差。

改进遗传算法 IGA 是一种模拟生物进化过程的全局优化算法。它将超参数寻优过程转化为一个适者生存的演化过程。算法首先在预设的参数范围内随机生成一个包含多个个体即多组  $C$  与  $\gamma$  参数组合的初始种群。随后，算法以分层十折交叉验证的 F1 分数为适应度函数，对种群中每个个体的优劣进行评估。适应度越高的个体，在后续的遗传操作中被选中作为父代产生后代的概率越大。通过选择，交叉和变异这三种模拟自然繁殖过程的核心算子，算法不断迭代产生适应度更高的新一代种群。此外，精英保留策略确保了每一代的最优个体都能直接进入下一代，保证了寻优过程的收敛性。当达到预设的进化代数或种群适应度不再提升时，算法终止，并输出整个进化过程中适应度最高的个体所对应的超参数组合作为全局最优解。具体机制如 algorithm 1 所示。

---

**Algorithm 1** 改进遗传算法 IGA 优化向量机超参数

---

**Require:**

数据集  $D$   
种群大小  $N$   
最大进化代数  $G_{max}$   
交叉概率  $p_c$   
变异概率  $p_m$   
超参数搜索空间  $S_C, S_\gamma$

**Ensure:**

最优超参数组合  $(C_{best}, \gamma_{best})$

```
1: function IGA-SVM-OPTIMIZATION( $D, N, G_{max}, p_c, p_m, S_C, S_\gamma$ )
2:   初始化种群  $P_0$ : 随机生成  $N$  个个体  $(C_i, \gamma_i)$ , 其中  $C_i \in S_C, \gamma_i \in S_\gamma$ 
3:   定义适应度函数  $Fitness(C, \gamma) \leftarrow$  使用  $(C, \gamma)$  配置的 SVM 在数据集  $D$  上进行分层十折交叉验证的 F1 分数
4:   for  $g = 1$  to  $G_{max}$  do
5:     计算当前种群  $P_{g-1}$  中每个个体的适应度
6:      $P_{new} \leftarrow \emptyset$                                  $\triangleright$  初始化新一代种群
7:     寻找当前种群中的最优个体  $elite \leftarrow \arg \max_{i \in P_{g-1}} Fitness(i)$            $\triangleright$  精英保留策略
8:      $P_{new} \leftarrow P_{new} \cup \{elite\}$                  $\triangleright$  精英保留策略
9:     for  $k = 1$  to  $N - 1$  do
10:     $\triangleright$  通过遗传算子生成新个体
11:     $parent_1, parent_2 \leftarrow \text{SELECT}(P_{g-1})$             $\triangleright$  根据适应度选择父代
12:    if  $\text{random}(0, 1) < p_c$  then
13:       $child \leftarrow \text{Crossover}(parent_1, parent_2)$          $\triangleright$  交叉操作
14:    else
15:       $child \leftarrow parent_1$                                  $\triangleright$  直接复制
16:    if  $\text{random}(0, 1) < p_m$  then
17:       $child \leftarrow \text{MUTATE}(child)$                        $\triangleright$  变异操作
18:     $P_{new} \leftarrow P_{new} \cup \{child\}$ 
19:     $P_g \leftarrow P_{new}$                                      $\triangleright$  更新种群
20:     $(C_{best}, \gamma_{best}) \leftarrow \arg \max_{i \in P_{G_{max}}} Fitness(i)$            $\triangleright$  找出最终的最优个体
21:  return  $(C_{best}, \gamma_{best})$ 
```

---

在具体实现中，我们配置了一个包含 15 个个体，进化 30 代的遗传算法。其寻优过程如图29所示，该三维瀑布图展示了在迭代过程中，种群的最高适应度，平均适应度与最低适应度的变化情况。图中可见，种群的整体适应度随着进化代数的增加而快速上升并趋于稳定，表明该算法能够高效地在参数空间中搜索到最优解区域。

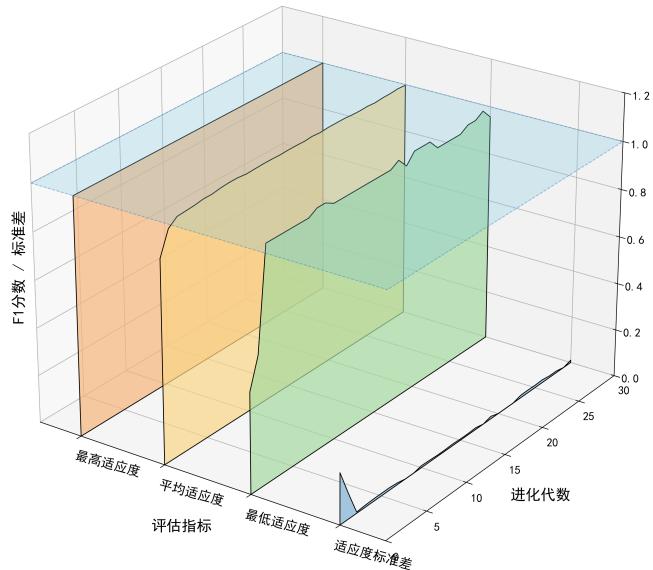


图 29 改进遗传算法寻优过程

### 9.3 最终鉴别结果与双重可靠性验证

利用前述流程寻得的最优超参数构建最终的 IGA-SVM 分类模型后，我们将其应用于 8 个未知类别样本的化学成分数据，完成了最终的类别鉴别。其具体的预测结果如表 5 所示。

表 5 未知样本最终预测结果

样本标识	0	1	2	3	4	5	6	7
预测类别	高钾	铅钡	铅钡	铅钡	铅钡	高钾	高钾	铅钡

以确保鉴别结果的可靠性，我们从数据与模型两个维度设计了双重灵敏度分析方案。第一重分析是蒙特卡洛扰动分析，用以检验预测结论对于数据测量误差的稳健性。我们进行了 1000 次模拟实验，在每次实验中，对样本的部分化学成分数据加入 0 至 5 个百分点的随机噪声，然后使用模型重新进行预测。我们通过计算每个样本预测结果在 1000 次扰动中发生改变的频率，来量化其预测不稳定性。实验结果显示，所有样本的预测不稳定性得分均为 0，这表明鉴别结论对于一定范围内的测量误差具有高度的稳定性。

第二重分析是支持向量机超参数敏感性分析，用以检验预测结论对于模型参数选择的稳健性。我们围绕寻得的最优  $C$  与  $\gamma$  值构建了一个 5 乘 5 的参数网格。对于网格中的每一对参数组合，我们都重新训练支持向量机模型并对未知样本进行预测，然后记录其预测结果与原始最优模型结果不一致的样本数量。分析结果通过热力图进行可视化，如图30所示。

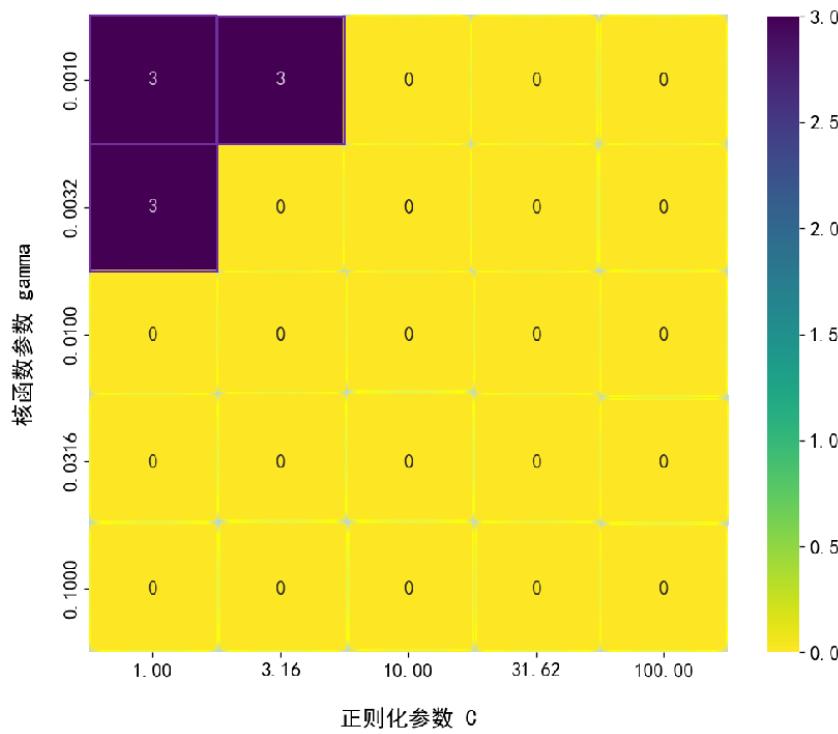


图 30 支持向量机超参数敏感性分析

图30中，每个方格的颜色代表在该参数对下预测结果发生改变的样本数量，颜色越深表示变化越大。图中可见，在以最优参数点为中心的广大区域内，颜色均为最浅色，对应数值为 0，这意味着预测结果没有发生任何改变。这一广阔的稳定区域证明，最终的鉴别结论对模型超参数的选择不敏感，并非是在某个特定参数点上偶然获得的结果，从而进一步确认了结论的可靠性。

## 十、问题四：化学成分关联网络与结构差异分析

古代玻璃的化学成分体系并非各种元素的随机组合，其背后反映了原料选择与烧制工艺的内在规律。为了探寻不同类别玻璃在配方层面的结构性差异，本章将各类玻璃的化学成分体系构建为网络模型。网络中的节点代表化学成分，节点间的连接则代表成分间在配方中存在的直接关联。通过比较高钾玻璃和铅钡玻璃网络的拓扑结构差异，可以探寻其原料与工艺的深层不同。分析过程始于已分类的化学成分数据，首先应用中心化对数比变换处理成分数据固有的统计约束，随后利用图套索算法计算反映成分间直接关

联的偏相关系数并构建网络。进一步理解网络的宏观结构，采用社群发现算法对网络节点进行聚类。最终，所有分析结果被整合至可视化网络图中，以呈现两类玻璃在配方结构上的分别特征。

## 10.1 成分数据变换与关联构建方法

玻璃成分数据属于一种特殊的成分数据，其各组分之和恒定。这一特性会在统计上产生闭合效应，引入虚假的负相关关系。例如，当主要成分二氧化硅  $SiO_2$  的含量增加时，其他所有成分的相对比例必然下降，即使它们在物理化学上并无关联。直接使用传统相关性分析会产生误导性结论。为解决此问题，本研究引入化学计量学领域处理此类数据的核心技术，即中心化对数比变换。该方法通过对数比值变换，将数据投影到一个无约束的几何空间中，从而消除产生虚假负相关的统计基础，保证后续关联性分析的有效性。

对于一个包含  $D$  种化学成分的样本向量  $\mathbf{x} = [x_1, x_2, \dots, x_D]$ ，首先计算其几何平均数  $g(\mathbf{x})$ 。

$$g(\mathbf{x}) = \left( \prod_{i=1}^D x_i \right)^{1/D} \quad (7)$$

随后对每个成分  $x_i$  进行变换，得到新向量  $\mathbf{y} = \text{clr}(\mathbf{x})$  的第  $i$  个元素  $y_i$ 。

$$y_i = \ln \left( \frac{x_i}{g(\mathbf{x})} \right) \quad (8)$$

在具体计算中，数据中的零值被一个极小的正数替代，以确保对数运算的有效性。

在变换后的数据基础上，需要探寻成分间排除了所有其他成分影响后的直接关联。常规相关系数无法区分直接关系与间接关系，因此需要计算偏相关系数。图套索算法是估计偏相关网络的常用方法，它通过求解一个带  $L1$  惩罚的优化问题，从数据中估计出一个稀疏的逆协方差矩阵，即精度矩阵，并由此推导出偏相关关系。其目标函数如下。

$$\hat{\Theta} = \arg \max_{\Theta > 0} (\log(\det \Theta) - \text{tr}(S\Theta) - \lambda \|\Theta\|_1) \quad (9)$$

其中， $S$  是变换后数据的样本协方差矩阵， $\lambda$  是控制稀疏度的正则化参数， $L1$  惩罚项会使精度矩阵  $\Theta$  中许多不重要的元素变为零。得到  $\Theta$  后，任意两个变量  $i$  和  $j$  之间的偏相关系数  $\rho_{ij\text{-rest}}$  可由以下公式算出。

$$\rho_{ij\text{-rest}} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}} \quad (10)$$

当且仅当  $\Theta_{ij}$  不等于零时，变量  $i$  和  $j$  之间存在直接关联。为避免人工设定正则化参数的主观性，本研究采用交叉验证版本的方法，使其能够自动为不同数据集找到最优的正

则化参数。

## 10.2 网络社群结构发现

充满连接的网络图虽然信息量大，但难以直接观察其核心结构。为从更高维度分析网络，本研究引入社群发现算法，其作用是自动识别网络中的模块或社群。一个社群是内部连接远比外部连接紧密的节点集群，在当前背景下可能代表了一组在化学功能或来源上高度相关的成分。本研究选用鲁汶算法，它通过优化模块度指标来迭代地寻找网络的最优社群划分。

鲁汶算法是一个多层次的迭代过程。第一阶段为节点移动，初始时每个节点是独立的社群，算法遍历所有节点，尝试将每个节点移动到其邻居所在的社群中，并计算此举带来的模块度增益，节点最终被分配到能使其模块度增益最大的邻居社群中。此过程反复进行，直到没有节点移动能再提升模块度。第二阶段为网络聚合，将第一阶段形成的每个社群压缩成一个超级节点，构建一个新的聚合网络，社群间的连接权重等于原始节点间连接权重的总和。算法重复这两个阶段，直到网络的模块度不再有显著提升，从而得到一个层次化的社群结构。该算法的具体步骤如算法2所示。

## 10.3 两类玻璃化学成分关联网络分析

最终的关联网络可视化图整合了偏相关强度，社群归属和节点重要性等多个维度的分析结果。图中每个节点代表一种化学成分，节点大小由连接数量与连接强度共同决定，连接广泛且作用强度高的成分由此成为网络的核心枢纽。节点颜色对应其所属的结构社群，颜色相同的节点在配方体系中构成一个功能或来源上相互关联的模块。节点间的连接代表成分间排除了其他变量影响后的直接关联，连接粗细正比于偏相关系数的绝对值，用以表现关联的强度，橙色连接为正相关，青色连接为负相关。

图31A展示了高钾玻璃的化学成分关联网络。其网络结构呈现两个主要社群。社群一为红色，由玻璃主干成分二氧化硅  $SiO_2$  和氧化铝  $Al_2O_3$  构成。社群二为青色，包含了氧化铜  $CuO$  和氧化钡  $BaO$  在内的多种金属氧化物。一个值得注意的现象是，作为高钾玻璃标志性成分的氧化钾  $K_2O$  并未归入任何主要社群，这表明在排除了其他成分的普遍影响后，它与其他成分的直接关联不集中。网络中，氧化铜  $CuO$ ，氧化钡  $BaO$  和二氧化硅  $SiO_2$  的节点尺寸较大，是网络中的关键节点。从关联上看，二氧化硅  $SiO_2$  与氧化铅  $PbO$  之间存在负相关，而氧化铜  $CuO$  与氧化钡  $BaO$  之间也表现出负相关关系。

图31B展示了铅钡玻璃的化学成分关联网络。该网络在结构上与高钾玻璃存在显著差异，共形成了三个社群。社群一为红色，是网络的主体，它同时包含了作为玻璃骨架的二氧化硅  $SiO_2$  和作为助熔剂的氧化铅  $PbO$  与氧化钾  $K_2O$ ，说明这些成分在配方中协同作用。社群二为青色，主要由氧化钡  $BaO$  和氧化铁  $Fe_2O_3$  等成分组成。社群三为蓝色，由氧化锡  $SnO_2$  和氧化钠  $Na_2O$  构成。网络的核心节点是氧化铅  $PbO$ ，二氧化硅

---

**Algorithm 2** 鲁汶社群发现算法

---

**Require:**带权重的图  $G = (V, E, w)$ **Ensure:**节点的社群划分  $P$ 

1: 初始化: 对于所有节点  $v \in V$ , 将其分配至独立的社群  $C_v \leftarrow \{v\}$

2: **repeat**

3:   设置标记  $improved \leftarrow \text{false}$

4:   **repeat** ▷ 第一阶段: 节点移动

5:     设置标记  $moved \leftarrow \text{false}$

6:     **for** 每一个节点  $v \in V$  **do**

7:        $C_{current} \leftarrow$  节点  $v$  当前所在的社群

8:        $\Delta Q_{max} \leftarrow 0$

9:        $C_{best} \leftarrow C_{current}$

10:      **for** 节点  $v$  的每一个邻居  $u$  **do**

11:         $C_{neighbor} \leftarrow$  节点  $u$  所在的社群

12:        **if**  $C_{neighbor} \neq C_{current}$  **then**

13:          计算将节点  $v$  从  $C_{current}$  移至  $C_{neighbor}$  的模块度增益  $\Delta Q$

14:          **if**  $\Delta Q > \Delta Q_{max}$  **then**

15:             $\Delta Q_{max} \leftarrow \Delta Q$

16:             $C_{best} \leftarrow C_{neighbor}$

17:          **if**  $C_{best} \neq C_{current}$  **then**

18:            将节点  $v$  从  $C_{current}$  移至  $C_{best}$

19:             $moved \leftarrow \text{true}$

20:             $improved \leftarrow \text{true}$

21:      **until**  $moved = \text{false}$  ▷ 第二阶段: 网络聚合

22:      **if**  $improved = \text{true}$  **then**

23:         $V' \leftarrow \emptyset, E' \leftarrow \emptyset$

24:        将当前划分中的每个社群  $C$  视为一个新的超级节点

25:         $V' \leftarrow$  所有超级节点的集合

26:        根据原始社群间的连接权重之和, 计算超级节点间的边和权重, 得到  $E'$

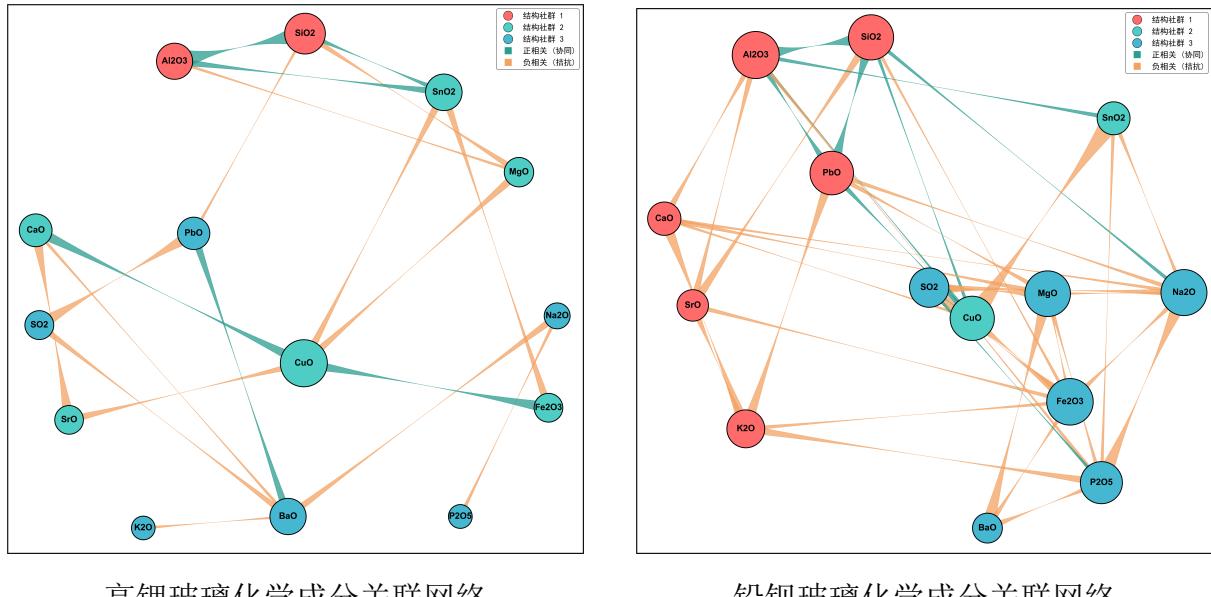
27:        构建新的聚合图  $G \leftarrow (V', E')$

28:      **until**  $improved = \text{false}$

31: **return** 当前图  $G$  中节点的最终社群划分  $P$

---

$SiO_2$  和氧化铝  $Al_2O_3$ 。网络中最突出的关联是氧化铅  $PbO$  与二氧化硅  $SiO_2$  之间强度很高的正相关, 这表明两者在铅钡玻璃的形成过程中有相互促进的作用。与此相对, 氧化铝  $Al_2O_3$  与二氧化硅  $SiO_2$  之间存在强度较高的负相关。两类玻璃关联网络在社群构成与核心关联上的不同, 系统地反映了它们在原料选择与烧制工艺上的区别。



高钾玻璃化学成分关联网络      铅钡玻璃化学成分关联网络

图 31 两类玻璃化学成分关联网络对比