

古代玻璃制品的成分分析和鉴别模型

摘要

玻璃在我国拥有悠久的发展历史,本文针对一批我国古代的玻璃制品进行数据处理和分析,构建了文物玻璃成分分析和类型鉴别的数学模型。

针对问题一,为研究玻璃表面风化与玻璃类型、颜色、纹饰之间的关系,本文采用斯皮尔曼相关性分析研究了数据之间的相关性,并通过卡方检验,寻找数据的差异性。接着,应用灰色关联分析方法,进一步研究玻璃化学成分与表面风化的关系,最终将 14 种化学成分划分为 6 种易风化成分和 8 种非风化成分;基于以上分析,本文通过查阅关于风化过程化学变化规律的文献,定义了成分变化系数,分别对易风化成分与非风化成分构建预测模型,建立了基于成分变化系数的风化前化学成分比例预测模型,成功预测了风化点风化前的化学成分比例,并使用附件数据进行验证,结果吻合程度高。

针对问题二,首先使用逻辑回归的方法,选出对玻璃类型影响大的变量作为决策变量。建立基于决策树的玻璃类型划分模型,研究了高钾和铅钡玻璃的分类规律,该方法对测试集的准确率达到 100%。之后,本文建立基于 K 均值聚类的亚类划分模型,分别对高钾、铅钡玻璃进行两级亚类划分,先将玻璃划分成风化、非风化两类,通过与已知情况比对,K 均值聚类方法的准确率为 91.3%。接着,对未风化样本再作亚类划分,分别选取聚类类别为 2、3、4、5 类,通过比较各分类的化学成分含量均值,结合文献,将未风化的玻璃分为具有明显化学成分含量差异的 6 类,其中未风化高钾玻璃分为 2 个亚类: $K_2O - Li_2O - SiO_2$, $Li_2O - SiO_2$ 。未风化铅钡玻璃分为 4 个亚类: $SiO_2 - PbO - BaO$, $PbO(\sim 25\%) - BaO - SiO_2$, $CaO - PbO(\sim 40\%) - BaO - SiO_2$, $PbO - BaO(\sim 25\%) - SiO_2$ 。之后,对分类结果进行了敏感性分析,包括剔除化学成分和扰动化学成分含量 2 种方式。

针对问题三,本文使用基于问题一、二的规律对其进行分析。首先使用了问题二的基于决策树的玻璃类型划分模型,对表单 3 中类型未知的玻璃文物进行鉴别。之后,使用支持向量机的方法对其再进行一次类别鉴定。最终二者的鉴定结果相同,说明问题二建立的分类模型效果优秀。文物玻璃 A1 ~ A8 的鉴定结果分别为:高钾、铅钡、铅钡、铅钡、铅钡、高钾、高钾、铅钡。最后,我们对化学成分比例这一数据进行了敏感性分析,得出当扰动幅度在 20% 以内时,模型的预测准确率为 100%。

针对问题四,首先使用主成分分析的方法对所有化学成分进行分析,初步给出了不同化学成分之间的关联关系。然后,将不同类别的玻璃划分为高钾风化、高钾非风化、铅钡风化、铅钡非风化四种类型,再通过斯皮尔曼相关性分析,寻找不同类别种化学成分之间的相关性。为比较不同类别之间的化学成分关联关系的差异性,我们定义差异系数,并给定临界值 $\alpha = 0.8$,进行了差异性分析。

关键词: 关联性分析 灰色关联分析 决策树 K 均值聚类 主成分分析

一、问题重述

1.1 问题背景

古代玻璃最初由丝绸之路传入我国，在与本土材料和技术相结合之后，发展出铅钡和高钾两类玻璃。古代玻璃存在风化问题，且风化过程中的化学反应，会引发玻璃的化学成分比例的改变。现有一批我国古代的铅钡、高钾玻璃制品，已知它们的分类信息和主要化学成分及比例，需要解决以下问题。

1.2 问题要求

问题一：根据表单一数据，首先需要分析表面风化与玻璃类型、纹饰和颜色的关系。再根据不同的玻璃类型，寻找表面风化与化学组成成分的关系，并预测风化前玻璃的化学成分含量。

问题二：基于附件数据，分析铅钡和高钾玻璃的分类规律。并根据表单二的数据，对每个类型的玻璃，选择合适的化学成分进行正分类，并对结果进行灵敏度分析。

问题三：根据表单3给出的几种玻璃的风化情况和化学成分，基于问题一、二得出的规律，对其玻璃类型进行预测，并对结果进行灵敏度分析。

问题四：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

2.1 问题分析

问题一可分为三步处理。第一步，根据表单一给出的数据，我们需要得到表面风化与玻璃类型、纹饰及颜色的关系。此关系包含相关性关系和差异性关系。考虑到玻璃风化、类型、颜色、纹饰四个变量都为定类变量，不能采用分析连续型变量的皮尔逊相关分析和方差分析等方法。因此，在进行相关性分析时，可以使用针对定类变量的斯皮尔曼相关系数进行分析；在进行差异性分析时，则可以采用卡方检验和效应量化分析的方法。

第二步，根据不同的玻璃类型，分析表面风化与玻璃化学成分含量的统计规律。首先，我们对数据进行归一化及缺失值处理，便于后续分析。由于表单二中的数据包含两种玻璃类型（铅钡玻璃和高钾玻璃），因此，我们可以首先根据玻璃种类将数据划分为两组，再采用灰色关联分析的方法进行分析。

第三步，根据风化情况，需要预测玻璃风化前的化学成分含量。要解决此问题，可以使用第二步得到的统计规律，并深入挖掘风化过程的化学变化规律，建立预测模型，从而得到风化点风化前化学成分比例。

2.2 问题二分析

对于问题二，首先需要寻找将玻璃划分为高钾和铅钡玻璃的分类方法和分类规律。由于影响两类玻璃分类结果的因素较多，共有 14 种化学成分，因此，可以先进行逻辑回归，确定一部分对玻璃类型影响效果最显著的化学成分。观察附件中化学成分的比例数据，可以发现对于铅钡玻璃和高钾玻璃，两者包含的 PbO 、 K_2O 、 BaO 等化学成分比例具有很大的差距，对于分类具有显著影响。基于此规律，并结合决策树的方法，可以建立玻璃的分类模型。

在此基础上，需要进一步对两种玻璃进行亚分类，即在高钾玻璃和铅钡玻璃中，细分出更多的亚类。可以通过 K 均值聚类的方法，对玻璃进一步划分亚类，并通过查阅文献来检验玻璃划分结果的合理性。

2.3 问题三分析

根据附件表单中所给的数据，我们已知八种玻璃样本的风化情况，以及化学成分所占比例。现在需要根据以上信息，推断玻璃的类型。基于问题一和问题二的模型，我们可以采用模型二决策树的方法，判断八个玻璃样本所属的类型。同时，采用支持向量机 (SVM) 的方法，对八个玻璃样本再一次鉴别类型，最后对比两种方法的分类结果。

2.4 问题四分析

问题四要求针对不同类型的玻璃文物样品，找到其化学成分之间的关联关系，并比较不同类型化学成分关联关系的差异性。

主成分分析可以用较少的变量来解释原来资料中的大部分信息，因此可以选用主成分分析法，初步分析化学成分之间的内在关系。再分别针对不同类别的玻璃进行斯皮尔曼相关性分析，得到不同类别的玻璃与化学成分之间的相关性关系。为比较不同类别玻璃的化学成分关联关系的差异性，可以定义关联性之差为差异系数，进行差异性分析。

三、模型假设

假设一：假设除风化作用外，不存在其他物理和化学作用对玻璃文物的化学成分含量产生影响。

假设二：在考虑风化前后化学成分含量的变化情况时，无需考虑时间的影响。

假设三：附件表单 3 中的空白处表示该含量未知。

四、符号说明

符号	定义	单位
ρ	斯皮尔曼相关系数	/
α_i	第 i 类化学成分的成分变化系数, $i = 1, 2, \dots, 14$	/
p_i	第 i 类化学成分风化前所占比例, $i = 1, 2, \dots, 14$	/
p'_i	第 i 类化学成分风化后所占比例, $i = 1, 2, \dots, 14$	/
acc	分类准确度	/
K_i	聚类结果的类别, $i = 1, 2, \dots, 6$	类
P_i	相关系数	/
β	差异系数	/

(注: 未列出符号及重复符号以出现处为准)

五、问题一模型

问题一要求根据附件数据，分析玻璃文物的风化与类型、颜色与纹饰之间的关系，并找到表面风化与化学成分组成比例的统计规律，并根据此规律求解玻璃风化之前的化学物成分比例。首先需要进行数据预处理，方便后续模型的建立。附件表单 1 中的数据为定类数据，因此我们使用斯皮尔曼相关性分析和卡方检验分析不同变量之间的关系。再通过灰色关联分析，分析玻璃风化与化学成分比例的关系，并运用此规律预测风化前的化学成分比例。

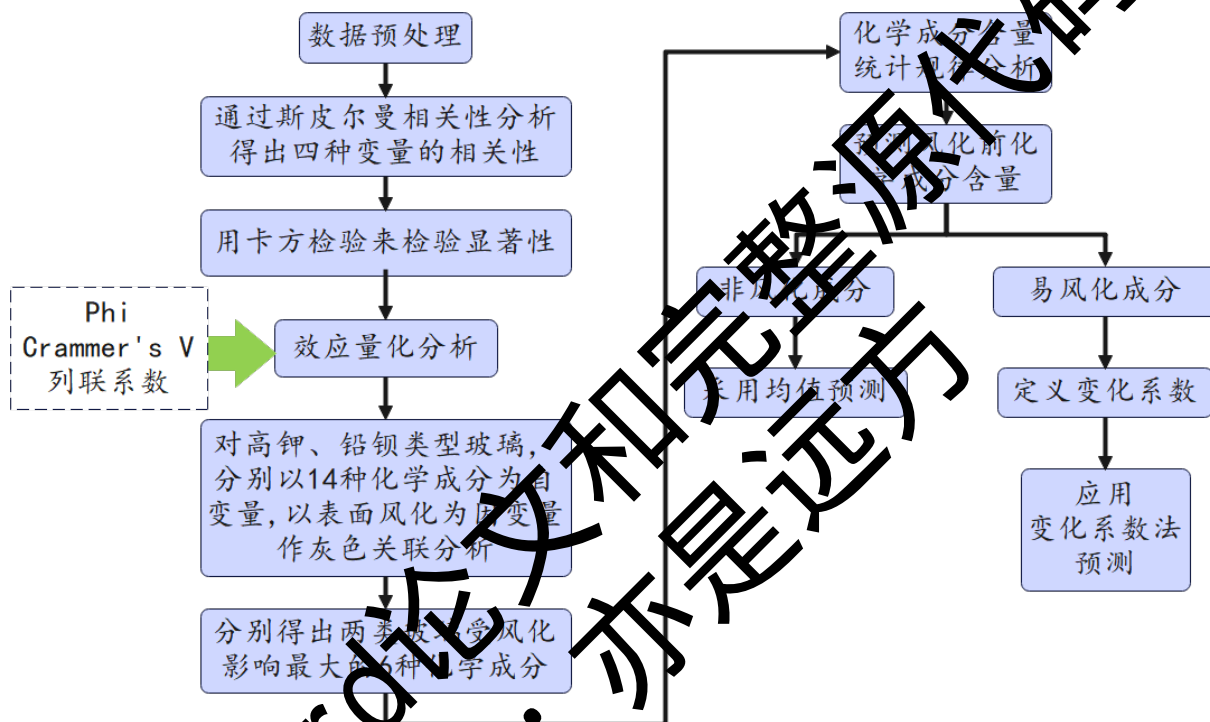


图1 问题一流程图

5.1 数据预处理

通过对附件数据的分析，我们可以发现表单中部分玻璃文物的数据存在缺失值与异常值。为方便后续模型的建立，我们对附件数据进行预处理。

1. 缺失值

对于附件中表单一的数据，我们可以发现对于部分编号的玻璃文物，其颜色数据缺失，例如文物编号为 19、40、48、58 的玻璃。在寻找玻璃文物是否风化与玻璃类型、颜色、纹饰关系的过程中，此类数据由于缺失颜色，且不易填补缺失值，因此对于求解过程没有实际意义，我们将这一小部分数据剔除。

2. 异常值

在附件表单 2 中，已知各文物主要化学成分所占的比例。这些数据具有成分性，因此各成分比例累加和应当为 100%。但由于检测手段的原因，可能导致其成分比例累和

不为 100% 的情况，因此，题目要求将化学成分比例累和介于 85% 至 105% 之间的数据视为有效数据。

根据此要求，我们将化学成分比例之和不在此范围内的数据认定为无效数据，并对附件表单 2 中的数据进行异常值检测，最后发现两组不符合要求的数据，其文物编号为 15 和 17，将之剔除。

3. 数据归一化

为了归纳统一样本的统计分布性，并消除奇异样本导致的不良影响，本文采用均值方差归一化的方法对数据进行预处理。归一化公式如下所示：

$$x' = \frac{x_i - \mu}{\sigma}, \quad (1)$$

其中， x' 表示归一化后的数据， x_i 为单个样本数据， μ 为样本数据均值， σ 为样本数据方差。

5.2 数据预分析

通过对表 1 数据的分析，可以发现对于某一玻璃文物，具有四种属性：表面风化、纹饰类型、颜色以及类型。

在数据预处理之后，对附件中给出的数据进行分析，我们可以得到玻璃不同特征与风化之间的数量关系。据此，我们将玻璃文物的是否风化、类型、颜色和纹饰的关系记录在表中，如表 1 所示：

表 1 风化关系分析表

玻璃类型	是否风化	纹饰	颜色	总含量范围
高钾	风化	B	蓝绿	99.81%-100%
	无风化	B、C	蓝绿、浅蓝、深蓝	97.25%-100%
铅钡	风化	A、C	黑、蓝绿、浅蓝、浅绿、深绿、紫	90.17%-99.89%
	无风化	A、C	深蓝、浅蓝、深绿、浅绿、紫、绿	88.41%-99.98%

通过表 1，我们可以发现一些数据间的规律，例如对于高钾玻璃，被风化的类型均具有 B 类纹饰；对于铅钡玻璃，颜色为黑色的玻璃均已风化，而 B 类纹饰的铅钡玻璃均未风化。

进一步对附件数据的分析，我们可以得到各类化学成分风化前后含量的均值，统计结果如表 2、表 3 所示：

表 2 高钾文物风化前后均值对比表

化学成分	未风化均值	风化后均值	化学成分	未风化均值	风化后均值
二氧化硅 (SiO_2)	67.9842	93.9633	氧化铜 (CuO)	2.4525	1.5617
氧化钠 (Na_2O)	0.695	0	氧化铅 (PbO)	0.4117	0
氧化钾 (K_2O)	9.3308	0.5433	氧化钡 (BaO)	0.5983	0
氧化钙 (CaO)	5.3325	0.87	五氧化二磷 (P_2O_5)	1.4025	0.28
氧化镁 (MgO)	1.0792	0.1967	氧化锶 (SrO)	0.04167	
氧化铝 (Al_2O_3)	6.62	1.93	氧化锡 (SnO_2)	0.1965	
氧化铁 (Fe_2O_3)	1.9317	0.265	二氧化硫 (SO_2)	0.101	0

表 3 铅钡文物风化前后均值对比表

化学成分	未风化均值	风化后均值	化学成分	未风化均值	风化后均值
二氧化硅 (SiO_2)	54.6596	24.9127	氧化铜 (CuO)	1.4517	2.2758
氧化钠 (Na_2O)	1.6826	0.2162	氧化铅 (PbO)	22.848	43.3138
氧化钾 (K_2O)	0.2187	0.1335	氧化钡 (BaO)	9.0017	11.8073
氧化钙 (CaO)	1.3204	2.6934	五氧化二磷 (P_2O_5)	1.0491	5.2773
氧化镁 (MgO)	0.6404	0.65	氧化锶 (SrO)	0.2683	0.4185
氧化铝 (Al_2O_3)	4.4561	2.97	氧化锡 (SnO_2)	0.0465	0.0685
氧化铁 (Fe_2O_3)	0.7365	0.7665	二氧化硫 (SO_2)	0.1591	1.3662

将上表结果可视化，我们得到图 2 所示结果：

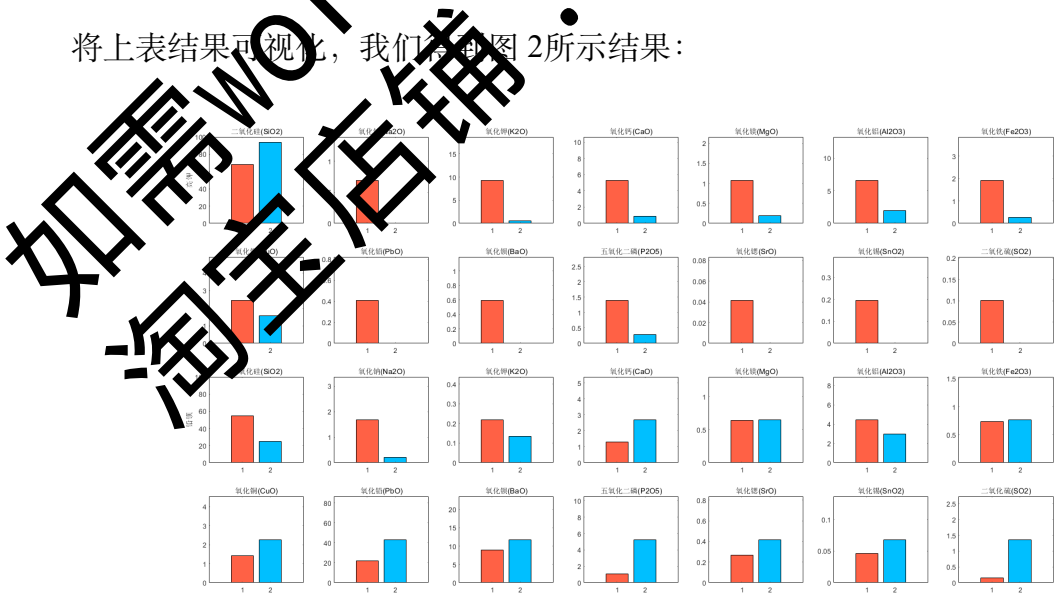


图 2 风化前后化学成分比例对比图

其中橙色为风化前数据，蓝色为风化后数据。

观察图 2，可以直观地得出风化前后化学成分含量的统计规律：风化后，高钾玻璃的 SiO_2 含量大幅升高，而其余化学成分的含量均呈下降趋势；铅钡玻璃的 SiO_2 含量大幅降低，而其余化学成分含量多数呈上升趋势。

5.3 基于斯皮尔曼相关系数与卡方检验的风化分析模型

由于表单数据类型为定类型数据，需要分别衡量玻璃文物的颜色、类型和纹饰对玻璃风化程度的影响，以及三个影响因素之间的关联程度，故本文采用斯皮尔曼相关性分析和卡方检验，分别得出不同影响因素之间的相关性和差异性。

5.3.1 斯皮尔曼相关性分析

斯皮尔曼相关系数是一个衡量两个变量之间依赖性的非参数指标，该系数对两个数据集是否属于同分布没有要求。斯皮尔曼相关系数分析范围从 -1 到 $+1$ 。当其取值为 0 时，表示两个参数之间没有相关性。当数据不重复，且两个变量完全单调相关时，斯皮尔曼相关系数为 $+1$ 或 -1 。当自变量和因变量变化方向相同时，斯皮尔曼相关系数为正，反之，当自变量和因变量变化方向不同时，斯皮尔曼相关系数为负。

斯皮尔曼相关系数定义如式 (2) 所示：

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2)$$

其中 d_i 表示第 i 个数据对的位次值之差， n 表示文物玻璃样本的总数目。

通过 SPSSPRO 软件计算斯皮尔曼相关系数，得到纹饰、类型、颜色及表面风化四个元素之间的相关性的大小，如图 3 所示：

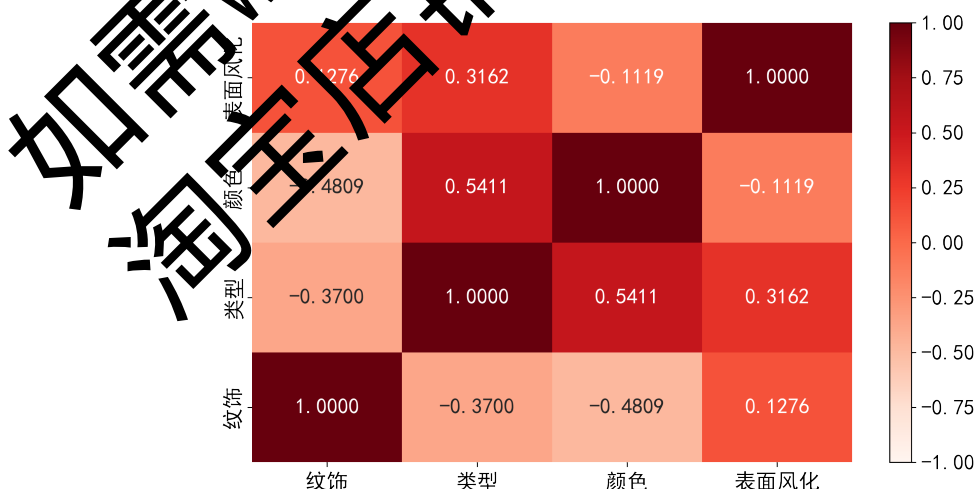


图 3 斯皮尔曼相关系数热力图

通过比较 ρ 值的大小，可以发现，与表面风化相关性最大的是玻璃类型，相关性系数 $\rho_{\text{类型}} = 0.37$ ；其次是纹饰，相关性系数 $\rho_{\text{纹饰}} = 0.1276$ ；最后是颜色，相关性系数 $\rho_{\text{颜色}} = 0.1119$ 。

5.3.2 卡方检验

为分析玻璃类型、颜色、纹饰之间的差异性，我们采用卡方检验进行分析。卡方值的大小由实际观测值与理论推断值之间的偏离程度决定，二者偏差程度越大，则卡方值越大；反之，卡方值越小。如两值完全相等，则卡方值为 0。

首先假设变量之间相互独立，并计算期望频次，代入卡方统计公式计算卡方值，公式如所示：

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}, \quad (3)$$

其中 χ^2 为卡方值， f_0 为观测频率， f_e 为期望频率。在此基础上，计算自由度。通过查表，得到并比较卡方值或 P 值。同过比较 P 值，可以得到玻璃颜色、纹饰、类型三个影响因素之间的关联程度。

通过 SPSSPRO 对附件数据进行处理，得到的卡方检验结果如图 4 所示（其中 NAN 表示数据缺失，且不参与计算）：

题目	名称	表面风化		总计	χ^2	校正 χ^2	P
		风化	无风化				
颜色	NAN	4	0	4	9.432	9.432	0.307
	浅绿	1	2	3			
	浅蓝	12	8	20			
	深绿	4	3	7			
	深蓝	0	2	2			
	紫	2	2	4			
	绿	0	1	1			
	蓝绿	9	6	15			
	黄	2	0	2			
类型	铝铂	12	28	40	6.880	5.452	0.009***
	高钾	12	6	18			
纹饰	A	5	11	22	4.957	4.957	0.084*
	B	0	6	6			
	总计	13	17	30			

注：***、**、* 分别代表 1%、5%、10% 的显著性水平

图 4 卡方检验结果

通过以上的卡方检验结果可知，对于玻璃的颜色和与风化，其显著性 P 值为 0.307，大于 0.05，不接受原假设，存在显著性差异。对于玻璃类型与风化，其显著性 P 为 0.009*，拒绝原假设，存在显著性差异，对于玻璃纹饰与风化，其显著性 P 值为 0.084*，接受原假设，不存在显著性差异。

在此基础上, 我们进一步对以上数据进行效应量化分析。效应量化分析包括 Φ 、Crammer's V 和列联系数, 用于分析玻璃风化与玻璃类型、纹饰、颜色的相关程度。现解释量化分析指标如下:

(1) **Φ** : 表示两样本之间的相关程度。若 $\phi < 0.3$, 则两样本相关性较弱; 当 $0.3 < \phi < 0.6$ 时, 表明两样本相关性中等; 若 $\phi > 0.6$, 则两样本相关性较强。

(2) **Crammer's V** : 用于衡量两样本之间的相关性大小。该系数取值范围为 $[0, 1]$, 其中 0 表示两个变量无关, 1 表示完全相关。

(3) **列联系数**: 简称 C 系数。当列联表中的两个变量相互独立时, 系数 $C = 0$, 但其不可能大于 1。

使用 SPSSPRO 进行求解, 得到的结果如表 4 所示:

表 4 效应量化分析表

分析对象	Φ	Crammer's V	列联系数
颜色	0.403	0.403	0.374
类型	0.344	0.344	0.326
纹饰	0.292	0.292	0.281

由表 4 中的数据可知: 颜色的 Crammer's V 的值为 0.403, 说明纹饰和玻璃表面风化的差异程度为中等差异程度。类型的 Crammer's V 值为 0.344, 同为中等差异程度。纹饰的 Crammer's V 值为 0.292, 为较弱差异程度。

对 Φ 的数值进行分析可知: 在玻璃纹饰、颜色、类型三种特征中, 纹饰的 Φ 小于 0.3, 表示纹饰与玻璃的风化差异程度不高; 颜色和类型的 Φ 大于 0.3, 表示颜色和类型与玻璃的风化具有中等的差异性。

4.4 基于灰色关联分析的化学成分与风化分析模型

根据附件中表单 2 的数据, 可知, 任一玻璃文物中都含有多种化学成分, 且每一种玻璃的化学成分组成与比例各不相同, 风化前后玻璃的化学成分比例会发生改变。

为衡量文物玻璃中不同化学成分对判断玻璃是否风化的影响程度的大小, 本文采用对样本数量的多少和样本有无规律并没有要求的灰色关联分析法, 对表单 2 数据进行处理, 将 14 种化学成分区分为 6 种受风化影响大的易风化成分以及 8 种受风化影响小的非风化成分。

5.4.1 灰色关联分析

灰色关联分析的步骤如下所示：

(1) 确定比较数列和参考数。其中反映系统行为特征的称为参考数列和影响系统行为的称为比较数列。在该模型中，玻璃是否风化得到参考数列，所有的化学成分比例得到比较数列。其中，比较数列如式 (4) 所示：

a_i = a_i(j)|j = 1, 2, ..., n, i = 1, 2, ..., m, (4)

其中，m 为评价对象的数目，题中评价的对象为是否风化。n 为评价指标的数目，题中的评价指标为各化学成分的比例。

(2) 计算灰色关联系数。根据以上确定的数列，计算关联分析系数，该系数即决定了不同化学组成成分对于是否风化影响程度的大小。灰色关联分析计算公式如所示：

ξ(j) = (min_{1≤s≤m} min_{1≤t≤n} |a_0(t) - a_s(t)| + ρ max_{1≤s≤m} max_{1≤t≤n} |a_0(t) - a_s(t)|) / (|a_0(j) - a_i(j)| + ρ max_{1≤s≤m} max_{1≤t≤n} |a_0(t) - a_s(t)|), i = 1, ..., m, j = 1, ..., n. (5)

ξ(j) 为比较数列 a_i 对参考数列 a_0 在第 j 个指标上的关联系数，其中 ρ ∈ [0, 1] 为分辨系数。其中，min_{1≤s≤m} min_{1≤t≤n} |a_0(t) - a_s(t)|、max_{1≤s≤m} max_{1≤t≤n} |a_0(t) - a_s(t)| 分别称为两级最小差及两级最大差。一般来讲，分辨系数 ρ 越大，分辨率越大，ρ 越小，分辨率越小。

5.4.2 模型求解

通过 SPSSPRO 软件，以是否风化作为母序列，14 种化学成分作为子序列，得到灰色关联度并排序，分别选取前 6 个化学成分作为易风化成分(已加粗)，其余成分作为非风化成分，如表 5 和 表 6 如下所示 (完整结果参见支撑材料):

表 5 高钾玻璃灰色关联分析结果

评价项	关联度	排名
二氧化硅 (SiO ₂)	0.974	1
氧化铝 (Al ₂ O ₃)	0.913	2
氧化铜 (CuO)	0.905	3
五氧化二磷 (P ₂ O ₅)	0.893	4
氧化钾 (K ₂ O)	0.883	5
氧化钙 (CaO)	0.879	6
氧化铁 (Fe ₂ O ₃)	0.871	7
氧化铅 (PbO)	0.847	8
...

表 6 铅钡玻璃灰色关联分析结果

评价项	关联度	排名
氧化铅 (PbO)	0.967	1
氧化钡 (BaO)	0.939	2
氧化钙 (CaO)	0.936	3
五氧化二磷 (P ₂ O ₅)	0.929	4
二氧化硅 (SiO ₂)	0.928	5
氧化铜 (CuO)	0.911	6
氧化铝 (Al ₂ O ₃)	0.906	7
氧化铁 (Fe ₂ O ₃)	0.893	8
...

5.4.3 结果分析

我们使用斯皮尔曼相关系数对 14 种化学成分各自与表面风化的相关性进行判断，检查所得结果是否符合要求。利用 MATLAB 程序，得到斯皮尔曼相关系数如表 7 所示：

表 7 对表面风化的相关系数

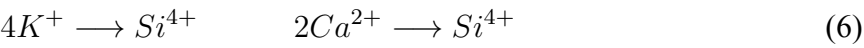
化学成分	高钾	铅钡	化学成分	高钾	铅钡
二氧化硅 (SiO_2)	0.6439	-0.6264	氧化铜 (CuO)	-0.3866	0.2435
氧化钠 (Na_2O)	-0.2379	-0.2933	氧化铅 (PbO)	-0.4182	0.5023
氧化钾 (K_2O)	-0.5420	-0.1829	氧化钡 (BaO)	-0.2829	0.1848
氧化钙 (CaO)	-0.4381	0.4245	五氧化二磷 (P_2O_5)	-0.4768	0.3732
氧化镁 (MgO)	-0.3673	-0.0714	氧化锶 (SrO)	-0.3682	0.1434
氧化铝 (Al_2O_3)	-0.5924	-0.2569	氧化钼 (MoO_3)	-0.1296	-0.0440
氧化铁 (Fe_2O_3)	-0.3608	0.0878	二氧化硫 (SO_2)	-0.2379	0.2088

根据表 7 结果，我们可以发现，利用灰色关联分析得到的易风化成分与表面风化的相关性比非风化成分基本都更高，说明基于灰色关联分析对易风化成分、非风化成分的区分是合理的。易风化成分与非风化成分的区分是风化前后化学成分含量统计规律的集中体现。

5.5 基于成分变化系数的风化前化学成分比例预测模型

5.5.1 风化过程化学规律分析

要对风化之前的化学成分进行预测，首先需要从化学的微观角度对风化过程进行分析。以高钾玻璃为例，在风化的过程中，氧化钾 K_2O 、氧化钙 CaO 、氧化铁 Fe_2O_3 等金属化学成分的含量下降，二氧化硅 SiO_2 的比例上升。从化学成分的化合价角度考虑，钾 K 、钙 Ca 、铁 Fe 的化合价分别为 1 价、2 价和 3 价，硅 Si 的化合价则为 4 价。因此，根据化合价守恒的原理，每损失 4 个钾原子，可以置换 1 个硅原子。同理，每损失 4 个钙原子，可以置换 2 个硅原子。原子之间的置换规律实例如式 (6) 所示：



5.5.2 预测模型的建立

通过以上分析,可以发现,在风化过程中,不同化学成分的含量有增有减,这种增减由元素的置换反应等化学反应引发,且不同化学成分之间的符合一定的比例关系。在此基础上,我们建立风化前化学成分比例的预测模型。

在之前的分析中,依据灰色关联度的大小,我们得到了六种易风化化学成分以及剩下的八种非风化化学成分。

1. 非风化成分的预测

对于非风化化学成分,其具有含量低,风化前后变化小的特性。例如氧化锶 SrO ,在风化前后比例都不超过 2%。对于此类非风化化学成分,我们以风化前该化学成分所占比例的均值来近似预测结果。

2. 易风化成分的预测

对于易风化化学成分,我们根据化学成分之间转化的比例关系,定义成分变化系数 α 为:

$$\alpha_i = \frac{P'_i - P_i}{P_i}, \tag{7}$$

其中, P_i 为第 i 种化学成分风化前所占的比例的均值, P'_i 为该化学成分风化后所占比例的均值。

因此,根据 (7) 我们得到风化后化学成分所占比例预测的公式如 (8) 所示:

$$p'_i = \frac{p_i}{1 - \alpha_i}, \tag{8}$$

其中, p_i 为第 i 种化学成分的预测值, p'_i 为该化学成分风化后所占比例。

结合风化前后的均值及预测模型,我们选择了文物编号为 49 号的文物进行测试。风化前的预测数据及实际数据如表 8 所示 (其余结果见支撑材料):

表 8-49 号文物风化前预测数据与实际数据对比表

化学成分	预测含量	实际含量	化学成分	预测含量	实际含量
二氧化硅 (SiO_2)	53.82	54.61	氧化铜 (CuO)	0.4169	0.45
氧化铜 (Cu_2O)	1.4752	0	氧化铅 (PbO)	21.9862	23.02
氧化钾 (K_2O)	0.2096	0.3	氧化钡 (BaO)	4.1678	4.19
氧化钙 (CaO)	2.3564	2.08	五氧化二磷 (P_2O_5)	4.5410	4.32
氧化镁 (MgO)	0.6612	1.2	氧化锶 (SrO)	0.3336	0.3
氧化铝 (Al_2O_3)	4.2156	6.5	氧化锡 (SnO_2)	0.0428	0
氧化铁 (Fe_2O_3)	0.6868	1.27	二氧化硫 (SO_2)	0.1464	0

从表中数据可知,实际数据与预测数据接近,预测效果优秀。

六、问题二模型

问题二需要寻找高钾玻璃和铅钡玻璃的分类规律，并在此基础上，对每类玻璃进一步划分亚类。本文采用逻辑回归的方法首先确定具有重要影响力的化学成分，再结合不同类型玻璃所含化学成分存在显著差异的规律，使用决策树给出两种玻璃的决策模型。并使用 K 均值聚类的方法，对子类进行划分。得出划分结果之后，我们对划分种类和化学成分比例含量进行改变，分析模型灵敏度。最后，结合化学文献，对分类的合理性进行说明。

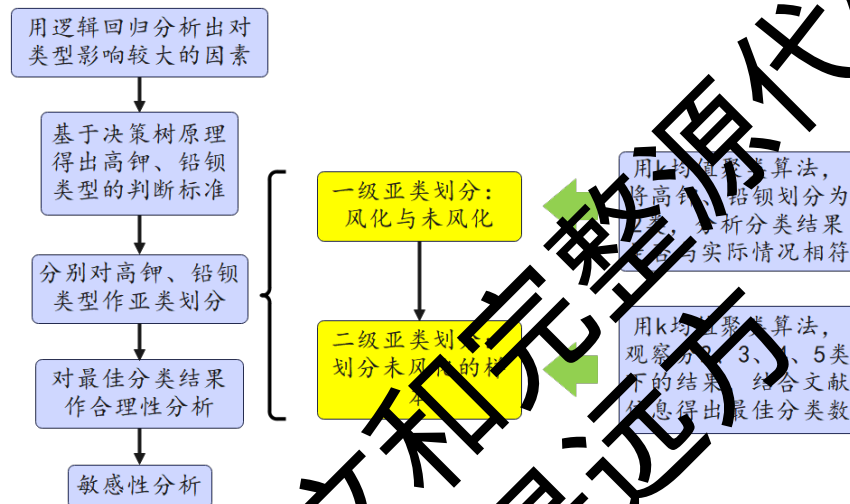


图5 问题二流程图

6.1 显著变量的选取

根据附件数据，每一种玻璃包含 14 种化学成分，以及 3 种属性特征（颜色、纹饰、表面风化）。每一种化学成分所占比例不同。其中部分化学成分在两种玻璃中的含量差别不大，而另一些化学成分则具有较大区别，对于玻璃种类的划分具有显著影响。

6.1.1 逻辑回归

Logistic regression 作为一种是分类模型，常用于二分类。分析步骤如下所示：

- (1) 对因变量分布状况进行描述。
- (2) 对模型进行似然比卡方检验，分析似然比卡方显著性，若拒绝原假设 ($P < 0.01$ 或 0.05)，说明模型有效，反之模型不成立；若设计多个模型，可以结合其他分类评价或者信息准则进行综合分析。
- (3) 根据模型参数表，分析 X 是否都呈显著性 (p 值小于 0.05 ，严格则需要小于 0.01)，用于探究 X 对 Y 的影响关系。
- (4) 分析回归系数 B 与 OR 值（优势比），对比分析 X 对 Y 的影响程度。

(5) 结合预测分类混淆矩阵与模型评价中的分类指标，分析模型预测。

6.1.2 显著变量

利用 SPSSPRO，采用逻辑回归的方法，我们逐一分析每一化学成分以及玻璃的 3 种属性（颜色、纹饰、风化）对于玻璃类型的划分是否有显著影响。最终得到如表 9 所示的几种具有显著影响的因素：

表 9 对玻璃类型影响显著因素种类表

颜色	表面风化	纹饰	二氧化硅	氧化钾
氧化钙	氧化铅	氧化钡	五氧化二磷	氧化锶

6.2 基于决策树的玻璃类型划分模型

根据之前逻辑回归的分析，我们找到了对玻璃类型划分具有显著影响的重要因素，在此基础上，我们构造了基于决策树的玻璃种类划分模型。

6.2.1 决策树构造

决策树是在已知各种情况发生概率的基础上，通过构成决策树来求取净现值的期望值大于等于零的概率，评价项目风险，判断其可行性的决策分析方法，是直观运用概率分析的一种图解法。由于这种决策分支画成图形很像一棵树的枝干，故称决策树。

结合上述的逻辑回归分析结果，我们选取上述 10 个对玻璃类型影响显著的因素作为决策变量构造决策树。

其中氧化钡、氧化铅和氧化钾在铅钡玻璃和高钾玻璃中具有很大的含量差异，从根本上反映了两种玻璃在化学成分组成上的差异。大部分样本，根据以上三组变量便足以区分属于何种玻璃。但是考虑到附件数据中存在数据缺失值，若进行补 0 操作，不能反映玻璃的真实组成成分，因此，若存在数据缺失，则将缺失的数据剔除，以剩下的化学成分指标构造决策树。

我们采用随机洗牌的方法对数据进行随机排序，按照 8 : 2 的比例将数据分为训练集和验证集，训练得到所需决策树模型。

6.2.2 决策树运行结果

经训练得到的决策树的运行流程图如图 6 所示，该决策树能够充分反映出高钾玻璃和铅钡玻璃的分类规律。

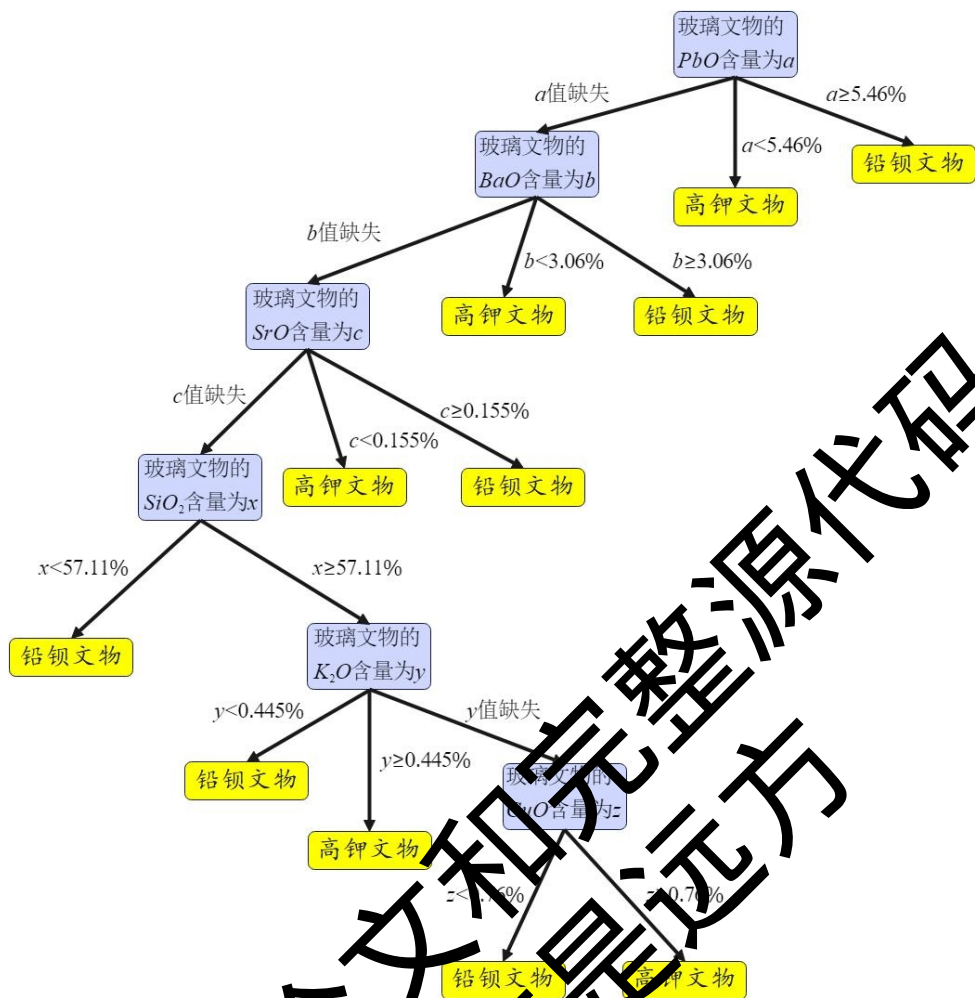


图 6 决策树分析过程

6.2.3 决策树模型检验

为了验证上述随机数的可行性，我们使用预留的 20% 验证集对随机数进行测试，表 10 列出预测结果：

表 10 验证集预测结果

文物编号	预测结果	实际结果	识别是否正确
3	高钾	高钾	是
8	铅钡	铅钡	是
...	是
42	铅钡	铅钡	是
50	铅钡	铅钡	是

为了量化分类效果，我们定义分类准确度：

$$acc = 1 - \frac{N_{\text{错误}}}{N_{\text{总数}}},$$

其中， $N_{\text{总数}}$ 是分类样本的总数， $N_{\text{错误}}$ 是被错误划分的样本数目。

测试结果显示，决策树的**预测准确度为 100%**。因此，该决策树模型可以根据氧化铅 PbO 、氧化钡 BaO 、氧化锶 SrO 、二氧化硅 SiO_2 、氧化钾 K_2O 、氧化铜 CuO 六个指标，无误差地判断出某一玻璃属于高钾玻璃还是铅钡玻璃。同时，该决策树也能充分反映出高钾玻璃和铅钡玻璃的分类规律。

6.3 基于 K 均值聚类的玻璃亚类划分模型

基于以上的分析，我们已经可以将玻璃文物分为高钾和铅钡玻璃两类，现在需要根据化学成分成分，对两个大类的玻璃进一步划分亚类。本文采用 K 均值分类的方法，实现亚类的划分。

6.3.1 K 均值聚类算法

K 均值聚类的思想是假定样本集中全体样本可分为 C 类，并选定 C 个初始聚类中心，然后根据最小距离原则将每个样本分配到某一类中，之后不断迭代计算各类的聚类中心，并依据新的聚类中心调整聚类情况，直到迭代收敛或聚类中心不再改变。

设 $m_i (i = 1, 2, \dots, C)$ 为 C 个聚类中心，记 $J_e = \sum_{i=1}^C \sum_{w \in G_i} \|w - m_i\|^2$ ，则使 J_e 最小的聚类使误差平方和准则下的最优结果。

K 均值聚类算法描述如下：

(1) 初始化. 设总样本集 $G = \{\omega_j, j = 1, 2, \dots, n\}$ 是 n 个样品组成的集合，聚类数为 $C (2 \leq C \leq n)$ ，将样本集 G 任意分为 C 类，记为 G_1, G_2, \dots, G_c ，计算对应的 C 个初始聚类中心，记为 m_1, m_2, \dots, m_c ，并计算 J_e 。

(2) $G_i \leftarrow \emptyset (i = 1, 2, \dots, C)$ ，按最小距离原则将样品 $\omega_j (j = 1, 2, \dots, n)$ 进行聚类，即若 $d(\omega_j, G_k) = \min_{1 \leq i \leq C} d(\omega_j, m_i)$ ，则 $\omega_j \in G_k, G_k = G_k \cup \{\omega_j\}, j = 1, 2, \dots, n$ 。重新计算聚类中心

$$m_i = \frac{1}{n_i} \sum_{\omega_j \in G_i} \omega_j, i = 1, 2, \dots, C,$$

式中， n_i 为当前 G_i 类中的样本数目，并重新计算 J_e 。

(3) 若连续两次迭代的 J_e 不变，则算法终止，否则算法转 (2)。

6.3.2 一级亚类划分

首先我们分别对高钾、铅钡进行亚类划分，在 SPSSPRO 软件中利用 K 均值聚类法分别把高钾、铅钡分为两类，根据上文分析可知，分类结果应当与已知的表面风化情况

相符合。我们在上文得到对类型有显著影响的变量中选择 7 个化学成分变量：氧化铅 PbO、氧化钡 BaO、氧化锶 SrO、二氧化硅 SiO₂、氧化钾 K₂O、氧化铜 CuO、五氧化二磷 P₂O₅，作为我们的划分依据。利用 SPSSPRO 软件，进行 K 均值聚类，得到的结果如图 7 所示：

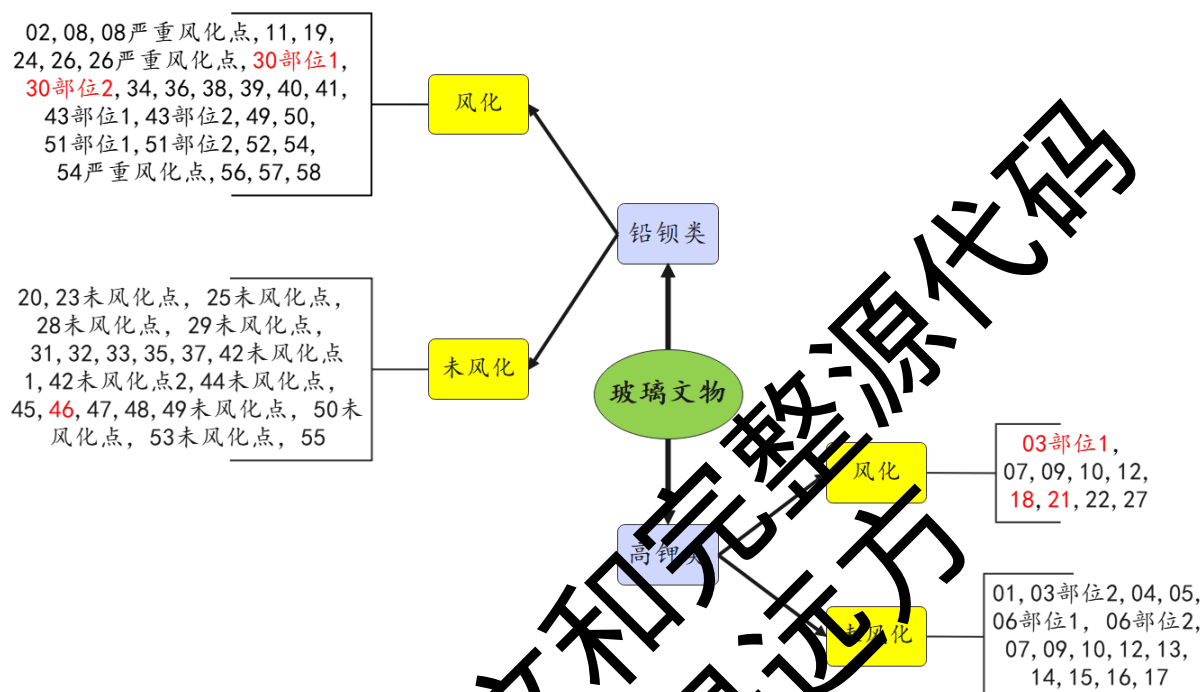


图 7 一级亚类划分图

可以发现，分类结果与已知的表面风化情况基本相符，高钾与铅钡中总共只有 3 个样本点被划分错误，准确率达到 91.3%，分类效果优秀。

6.3.3 二级亚类划分

利用 SPSSPRO 软件，同样在上文得到对类型有显著影响的变量中选择当中的 7 个化学成分变量，作为我们的划分依据，分别对高钾未风化类玻璃和铅钡未风化类玻璃进行 K 均值聚类。为了得到最佳的分类，聚类数目选取 2、3、4、5，得到每一种情况下各化学成分的中心值如图 8、图 9 所示：

聚类种类	二氧化硅(SiO ₂)	氧化钾(K ₂ O)	氧化铅(PbO)	氧化钡(BaO)	五氧化二磷(P ₂ O ₅)	氧化锶(SrO)
1	63.200	10.046	0.370	0.474	1.286	0.039
2	81.063	4.870	0.417	0.657	1.040	0.023
聚类种类	二氧化硅(SiO ₂)	氧化铅(PbO)	氧化钾(K ₂ O)	氧化钡(BaO)	五氧化二磷(P ₂ O ₅)	氧化锶(SrO)
1	63.594	0.449	11.759	0.409	0.719	0.029
2	81.063	0.417	4.870	0.657	1.040	0.023
3	62.510	0.232	7.050	0.587	2.280	0.058
聚类种类	二氧化硅(SiO ₂)	氧化钙(CaO)	氧化铅(PbO)	氧化钡(BaO)	五氧化二磷(P ₂ O ₅)	氧化锶(SrO)
1	76.680	4.710	1.000	1.970	1.100	0.000
2	62.510	1.353	0.232	0.587	2.280	0.058
3	63.594	7.409	0.449	0.409	0.719	0.029
4	83.255	1.005	0.125	0.000	1.010	0.035
聚类种类	二氧化硅(SiO ₂)	氧化钙(CaO)	氧化铅(PbO)	氧化钡(BaO)	五氧化二磷(P ₂ O ₅)	氧化锶(SrO)
1	83.255	1.005	0.125	0.000	1.010	0.035
2	66.797	7.237	0.037	0.000	1.553	0.013
3	60.916	7.112	0.676	0.766	1.514	0.056
4	63.410	0.000	0.193	0.410	1.540	0.037
5	76.680	4.710	1.000	1.970	1.100	0.000

图8 高钾玻璃聚类结果

聚类种类	二氧化硅(SiO ₂)	氧化钾(K ₂ O)	氧化钙(CaO)	氧化铅(PbO)	氧化钡(BaO)	五氧化二磷(P ₂ O ₅)	氧化锶(SrO)
1	65.422	0.157	1.205	16.447	5.911	0.525	0.174
2	46.381	0.266	2.409	26.422	12.379	1.452	0.341
聚类种类	二氧化硅(SiO ₂)	氧化钾(K ₂ O)	氧化钙(CaO)	氧化铅(PbO)	氧化钡(BaO)	五氧化二磷(P ₂ O ₅)	氧化锶(SrO)
1	65.422	0.157	1.205	16.447	5.911	0.525	0.174
2	35.143	0.530	2.300	27.855	17.605	1.825	0.435
3	51.376	0.149	1.013	25.342	8.612	1.287	0.299
聚类种类	二氧化硅(SiO ₂)	氧化钾(K ₂ O)	氧化钙(CaO)	氧化铅(PbO)	氧化钡(BaO)	五氧化二磷(P ₂ O ₅)	氧化锶(SrO)
1	65.422	0.157	1.205	16.447	5.911	0.525	0.174
2	51.376	0.149	1.013	25.342	8.612	1.287	0.299
3	35.143	0.705	2.300	38.480	10.320	0.705	0.415
4	24.650	0.355	0.235	19.220	24.890	2.945	0.455
聚类种类	二氧化硅(SiO ₂)	氧化钾(K ₂ O)	氧化钙(CaO)	氧化铅(PbO)	氧化钡(BaO)	五氧化二磷(P ₂ O ₅)	氧化锶(SrO)
1	65.422	0.157	1.205	16.447	5.911	0.525	0.174
2	52.957	0.123	0.707	22.108	9.448	0.783	0.377
3	48.213	0.000	1.627	31.810	6.940	2.293	0.143
4	24.650	0.355	0.235	19.220	24.890	2.945	0.455
5	32.655	0.705	4.365	38.480	10.320	0.705	0.415

图9 铅钡玻璃聚类结果

为了实现良好的聚类效果，并使结果更具合理性，我们需要选取合适的聚类数量。在查阅了相关论文资料后得知，我国古代的玻璃文物主要分为以下几类：

- (1) $PbO - BaO - SiO_2$ 玻璃系统。
- (2) $PbO - SiO_2$ 玻璃系统。
- (3) $K_2O - CaO - SiO_2$ 玻璃系统。

(4) $K_2O - SiO_2$ 玻璃系统。

根据图 8 和图 9 的结果，我们可以发现：对于高钾玻璃，当选取聚类数量为 2 的时候，聚类产生的两种亚类玻璃与 $K_2O - CaO - SiO_2$ 玻璃系统和 $K_2O - SiO_2$ 玻璃系统的特征一致；对于铅钡玻璃，当选取聚类数量为 4 或 5 时，聚类产生的亚类玻璃，与 $PbO - BaO - SiO_2$ 玻璃系统和 $PbO - SiO_2$ 玻璃系统的特性相符。

同时考虑到，当聚类数过多时，如铅钡类玻璃分为 5 种亚类时，第 2, 3 个亚类已经不具有明显的化学成分成分区别，此时出现分类过度的情况。因此综上所述，我们将未风化高钾玻璃聚类为 2 种亚类，未风化铅钡玻璃聚类为 4 种亚类。结果如 (9) 所示：

$$\left\{ \begin{array}{l} K1 : K_2O - CaO - SiO_2; \\ K2 : K_2O - SiO_2; \\ K3 : SiO_2(\sim 65\%) - PbO - BaO; \\ K4 : PbO(\sim 25\%) - BaO - SiO_2; \\ K5 : CaO - PbO(\sim 40\%) - BaO - SiO_2; \\ K6 : PbO - BaO(\sim 25\%) - SiO_2. \end{array} \right. \quad (9)$$

其中 $Ki(i = 1, 2, \dots, 6)$ 表示第 i 个亚类， $K1$ 与 $K2$ 属于未风化高钾玻璃细分的亚类， $K3$ 、 $K4$ 、 $K5$ 、 $K6$ 属于未风化铅钡玻璃细分的亚类。

对于未风化高钾玻璃，分成 $K1$ 、 $K2$ ：

$K1$ 具有较高含量的 $K_2O(\sim 10\%)$ ，以及较高含量的 $CaO(\sim 5.2\%)$ ，查阅文献可知，此类玻璃高含量的 K_2O 源于制造过程中使用了含钾量较高的助熔剂，如草木灰，而高含量的 CaO 来自烧制过程中使用较多的稳定剂石灰石。

$K2$ 为高硅高钾玻璃， $K2$ 的玻璃文物与 $K1$ 相比， CaO 的含量有明显下降，但是 SiO_2 的含量要大于 $K1$ 。查阅文献可知，在烧制过程中，加入的石英砂比 $K1$ 要多，而助熔剂草木灰、稳定剂石灰石比 $K1$ 少。

对于未风化铅钡玻璃，分成 $K3$ 、 $K4$ 、 $K5$ 、 $K6$

$K3$ 为高硅铅钡玻璃，在烧制过程中石英砂比其余铅钡玻璃都要多。

$K4$ 为中铅钡玻璃， PbO 的含量第二高，但 BaO 的含量不高。查阅文献可知，是由于在烧制过程中加入的助熔剂铅矿石中氧化铅的含量比较高。

$K5$ 为高铅钡玻璃， PbO 的含量最高， BaO 的含量第二高，同时 CaO 的含量也比较高。查阅文献可知，是由于在烧制过程中加入的助熔剂铅矿石比其余玻璃种类都要多，而作为稳定剂的石灰石也应当加入得多，且矿石中氧化铅的含量比较高。

$K6$ 为高钡铅钡玻璃， BaO 的含量最高，但 PbO 的含量不高。查阅文献可知，是由于在烧制过程中加入的助熔剂铅矿石中氧化钡的含量比较高。

6.3.4 敏感性分析

为验证模型的有效性，我们对上述 K 均值聚类的子类划分过程进行敏感性分析。

► 化学成分种类变化

从上述过程中，我们知道 K 均值聚类建立在七个显著影响因素的基础上，即氧化铅 PbO 、氧化钡 BaO 、氧化锶 SrO 、二氧化硅 SiO_2 、氧化钾 K_2O 、氧化铜 CuO 、五氧化二磷 P_2O_5 共七种化学成分。分别剔除其中一种化学成分，以剩下的六种化学成分作为聚类的决策变量，将剔除前后聚类的结果进行对比，计算剔除某一化学成分之后聚类的准确性。

最后，得到分别剔除七种化学成分之后聚类的准确率，如表 11 所示：

表 11 剔除一类化学成分之后的聚类准确率

被剔除种类	SiO_2	K_2O	CaO	PbO	BaO	P_2O_5	SrO_2
铅钡准确率	70.80%	95.65%	95.65%	100.00%	95.65%	100.00%	100.00%
高钾准确率	71.40%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

根据上表的结果可知，在大部分情况下，只剔除一种化学成分，模型也可以保持很高的准确度，但在剔除 SiO_2 的情况下，准确率仅有 70% 左右，说明玻璃种类的划分与 SiO_2 的关系很大，符合前文的分析与实际化学背景，分类是合理有效的。

► 化学成分含量扰动

对于不同的样本数据，我们对每一样本的七种化学成分比例添加一个扰动。以 SiO_2 为例，将样本中 SiO_2 的最大值、最小值分别进行一定比例的扰动，添加扰动后再进行分类，比较分类的准确率。结果展示在图 10 中。

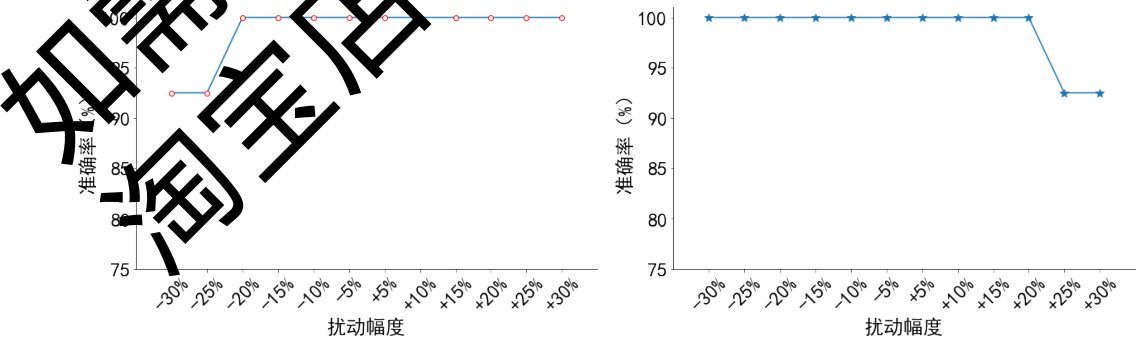


图 10 二氧化硅最大值（左）与最小值（右）变化与准确率的关系

可以发现，只有当 SiO_2 的扰动达到 25% 以上时，分类结果才会出现偏差，说明模型求解得到的结果中每种亚类之间是有显著差距，表明分类结果很好。

七、问题三模型

基于问题一、二的分析，可用上述的决策树对玻璃类型进行预测。为了说明决策树预测结果的有效性，再通过支持向量机 SVM 的方法预测玻璃类型。对比分析预测结果，从而得出决策树算法预测的准确性。

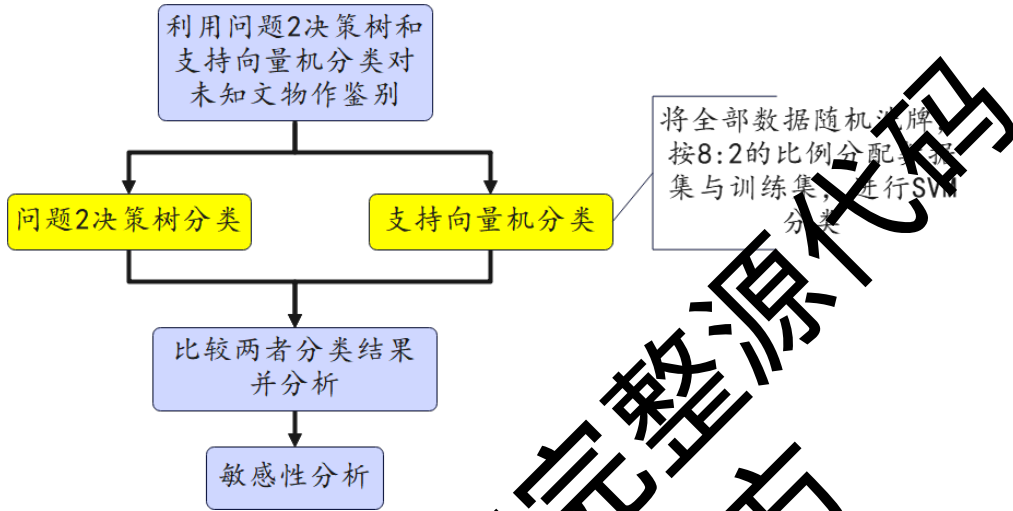


图 11 问题三流程图

7.1 支持向量机分类

支持向量机 (SVM) 的设计用意是找到各类样本点到超平面的距离最远，也就是找到最大间隔超平面。

任意超平面可用如下线性方程来描述：

$$w^T x + b = 0, \quad (10)$$

据此建立最优化问题如下：

$$\begin{cases} \min_{w,b} \left(\frac{1}{2} w^T \cdot w + C \sum_{i=1}^N \xi_i \right), \\ s.t. y_i (w^T \cdot x_i + b) \geq 1 - \xi_i, \\ \xi_i \geq 0, i = 1, 2, \dots, N, \end{cases} \quad (11)$$

式中, w 为法向量, b 为常数项, C 为惩罚因子, ξ_i 为松弛变量。

找出最优法向量 w 和常数项 b 的解就可以解得最优分类面，为了将上式转化为二次规划问题，引入相应拉格朗日函数，则分类问题可转化为

$$L(w, b, \lambda) = \frac{1}{2} w^T \cdot w + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^N \beta_i \xi_i, \quad (12)$$

式中, α_i, β_i 为拉格朗日乘子, $\alpha_i \geq 0, \beta_i \geq 0$ 。

根据对偶原理，上式可化为

$$\begin{cases} \max_{\alpha} L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^N \alpha_i \alpha_j y_i y_j (\Phi_{x_i} \times \Phi_{x_j}), \\ s.t. \sum_{i=1}^N y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, \end{cases} \quad (13)$$

其中， $\Phi_{x_i} \times \Phi_{x_j} = K(x_i, x_j)$ 为核函数。

对于线性不可分问题，为达到线性可分的目的，SVM 需引进核函数将数据从低维空间向高维空间投射。常用的核函数有线性核函数、多项式核函数、径向基核函数以及感知器核函数，不同核函数对应的支持向量机形式也不同。

7.2 模型的求解

通过 SPSSPRO 软件，进行支持向量机分类对 A1 到 A8 未知玻璃样本进行预测，得到的结果如表 12 所示：

表 12 未知类别玻璃文物的类型鉴别结果 (SVM)

文物编号	A1	A2	A3	A4	A5	A6	A7	A8
类型	高钾	铅钡	铅钡	铅钡	铅钡	高钾	高钾	铅钡

参照图 6 决策树，对附件表单 3 中未知类别玻璃文物的化学成分含量进行分析，可得其所属类型如表 13 所示：

表 13 未知类别玻璃文物的类型鉴别结果 (决策树)

文物编号	A1	A2	A3	A4	A5	A6	A7	A8
类型	高钾	铅钡	铅钡	铅钡	铅钡	高钾	高钾	铅钡

通过对以上结果进行分析，我们可以发现两种方法的预测结果一致，故问题 2 构建的决策树分类模型的准确度高且合理。

7.3 敏感性分析

题目要求对分类结果进行敏感性分析。因此，我们对需要预测的八个玻璃样本的 SiO_2 含量施加一个扰动，扰动的范围在 -20% 到 20% 之间以 5% 为间隔递增。受各化学成分含量总和需在 85% 至 105% 之间的条件约束，每一个扰动会使该样本的所有化学成分比例发生变化。我们采用决策树的方法预测不同扰动水平下的分类结果。得到的结果如表 14 所示：

表 14 玻璃文物推断灵敏度分析表

文物编号	-20% 扰动	-15% 扰动	-10% 扰动	-5% 扰动	5% 扰动	10% 扰动	15% 扰动	20% 扰动
A1	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变
A2	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变
A3	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变
A4	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变
A5	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变
A6	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变
A7	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变
A8	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变	结果不变

根据表 14 的结果，可得出结论：当扰动幅度在 20% 以内时，模型的预测结果均不变，即预测准确率为 100%。可知该模型能在一定干扰的情况下保持鉴别的准确度，具有较强的鲁棒性和广泛的适用性。

4. 问题四模型

问题四需要寻找不同类别玻璃文物样品的化学成分之间的关联关系，并比较它们之间的差异性。对此，我们可以首先采用主成分分析的方法，根据不同化学成分的因子载荷系数，初步挖掘其中的规律。在此基础上，对不同类别的玻璃进行关联性分析。对于不同类别之间的化学成分关联关系的差异性分析，通过定义差异系数，量化差异性并分析。

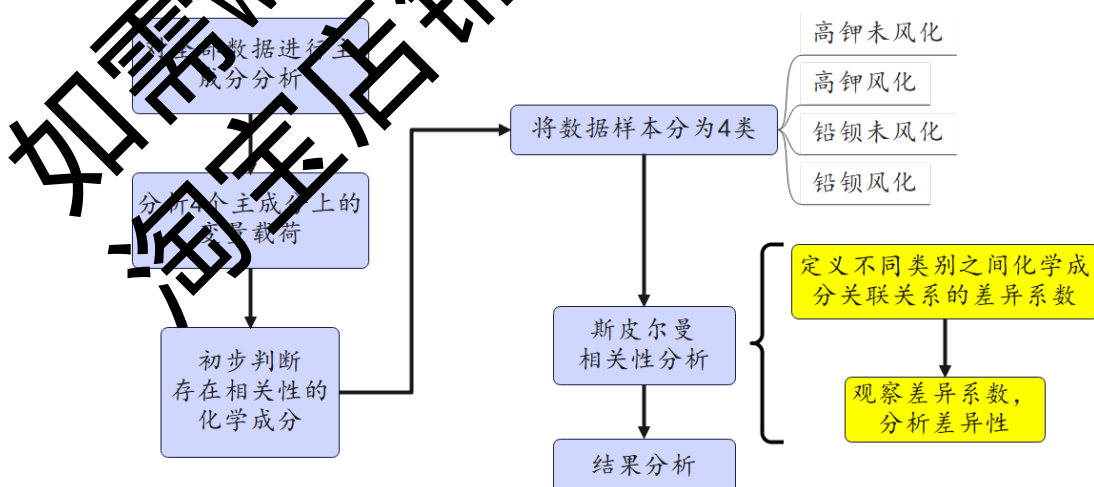


图 12 问题四流程图

8.1 基于主成分分析的关联关系初步挖掘模型

8.1.1 主成分分析算法

主成分分析是，以用较少的变量来解释原来资料中的大部分信息为目的，利用降维的方法，把多指标转化为几个综合指标的多元统计分析的方法。其基本步骤如下：

假设有 n 个研究对象， m 个指标变量 x_1, x_2, \dots, x_m ，第 i 个对象第 j 个指标取值 a_{ij} ，则构成数据矩阵 $A = (a_{ij})_{n \times m}$ 。

(1) 对原来的 m 个指标进行标准化

(2) 根据标准化后的数据矩阵 B 求得相关系数矩阵

(3) 计算相关系数矩阵 R 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 及对应的标准正交化特征向量 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ ，其中 $\mathbf{u}_j = [u_{1j}, u_{2j}, \dots, u_{mj}]^T$ ，由特征向量组成 m 个新指标向量

$$\begin{cases} F_1 = u_{11}y_1 + u_{21}y_2 + \dots + u_{m1}y_m, \\ F_2 = u_{12}y_1 + u_{22}y_2 + \dots + u_{m2}y_m, \\ \dots\dots\dots \\ F_m = u_{1m}y_1 + u_{2m}y_2 + \dots + u_{mm}y_m, \end{cases}$$

其中， F_1 是第 1 主成分， F_2 是第 2 主成分， \dots ， F_m 是第 m 主成分。

(4) 计算主成分贡献率及累计贡献率，一般取累计贡献率达 85% 以上的特征值 $\lambda_1, \lambda_2, \dots, \lambda_k$ 所对应的第 1, 第 2, \dots 第 k ($k \leq m$) 主成分。

(5) 最后利用得到的主成分 F_1, F_2, \dots, F_k 分析问题。

8.1.2 关联关系初步挖掘

使用 SPSSPRO 进行主成分分析，我们得到 14 种化学成分对应的四个主成分的因子载荷系数，如图 13 所示：

	因子载荷系数			
	主成分1	主成分2	主成分3	主成分4
二氧化硅(SiO ₂)	-0.804	-0.472	-0.008	-0.049
氧化钠(Na ₂ O)	-0.146	-0.343	-0.029	0.212
氧化钾(K ₂ O)	-0.658	0.232	0.473	0.078
氧化钙(CaO)	-0.278	0.684	0.4	-0.147
氧化镁(MgO)	-0.382	0.637	-0.261	0.276
氧化铝(Al ₂ O ₃)	-0.547	0.414	-0.007	0.317
氧化铁(Fe ₂ O ₃)	-0.385	0.629	0.094	-0.256
氧化铜(CuO)	0.318	-0.009	0.73	-0.005
氧化铅(PbO)	0.77	0.28	-0.419	-0.14
氧化钡(BaO)	0.779	-0.108	0.409	0.277
五氧化二磷(P ₂ O ₅)	0.435	0.659	-0.082	-0.001
氧化锶(SrO)	0.701	0.307	-0.195	0.274
氧化锡(SnO ₂)	-0.201	0.069	-0.228	0.702
二氧化硫(SO ₂)	0.468	0.019	-0.515	0.307

图 13 主成分分析结果

在图 13 中，化学成分与某一个主成分因子的因子载荷系数越大，该因子与变量的关系越近，对该化学成分的影响率越大。根据以上结果，可知：

(1) 在主成分 1 中，二氧化硅 SiO₂、氧化钾 K₂O、氧化铅 PbO、氧化钡 BaO、氧化锶 SrO 五种化学成分具有最大的载荷因子，且这五种化学成分对决定玻璃的类型具有较大影响，因此主成分 1 可以认定为玻璃的主要构成成分。

(2) 在主成分 2 中，氧化钙 CaO、氧化镁 MgO、氧化铁 Fe₂O₃、五氧化二磷 P₂O₅ 四种化学成分具有较高的载荷系数。经查找文献，以上四种化学成分在玻璃的制备过程中起着不同的作用。氧化钙 CaO 属于稳定剂，氧化镁 MgO 可以改善玻璃的折光性，五氧化二磷 P₂O₅ 是吸水剂和干燥剂的作用。因此，我们认定主成分 2 为玻璃的辅助剂成分。

(3) 在主成分 3 中，载荷系数较高的为氧化铜 CuO，氧化铜在调节玻璃的颜色中发挥着重要作用，因此，认定主成分 3 为玻璃的着色剂成分。

(4) 在主成分 4 中，载荷系数较高的为氧化锶 SrO，氧化锶对于着色剂在玻璃上的附着发挥着促进作用。虽然该物资并不影响玻璃颜色，但起着辅助上色的作用。因此，认定主成分 4 为玻璃的辅助着色成分。

因此可以认为，被划分到同一主成分下的化学成分之间是存在关联性的。

8.2 基于斯皮尔曼相关性的化学成分关联关系分析模型

为进一步分析不同类别的玻璃文物样品与化学成分的相关性，本文首先将不同类别的玻璃文物划分为高钾风化、高钾未风化、铅钡风化、铅钡未风化四类，并分别计算四类玻璃与化学成分成分之间的斯皮尔曼相关系数。

► 未风化：

对于高钾未风化类和铅钡未风化的玻璃，通过 *SPSSPRO* 计算其与化学成分之间的相关性系数，我们得到相关性系数表如图 14 所示（热力图中的具体数值参见支撑材料）：

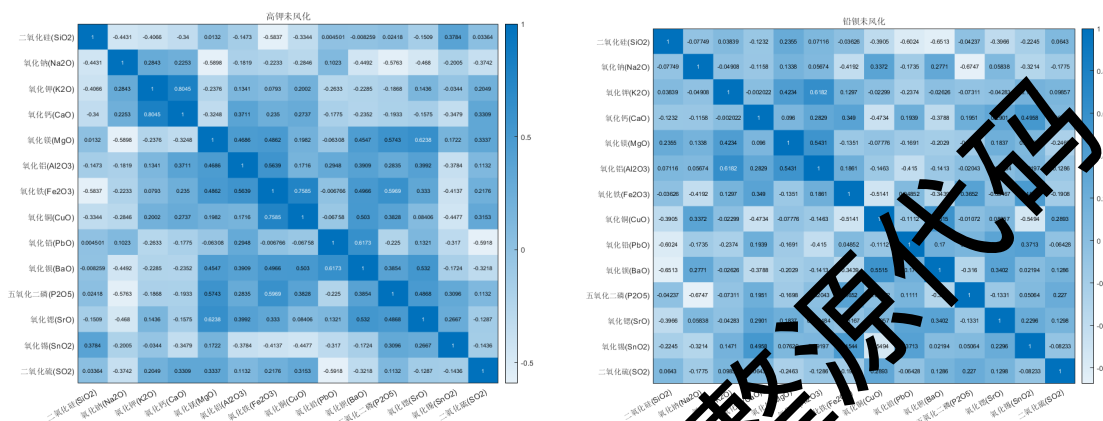


图 14 高钾（左）、铅钡（右）玻璃未风化的相关性系数热力图

从数据角度分析，我们可以发现：

对于未风化的高钾玻璃

氧化钾与氧化钙的相关系数 0.8941，氧化铁与氧化铜的相关系数 0.7585，氧化 Sr 与氧化镁的相关系数 0.6238，氧化铅与氧化钡的相关系数 0.6173，说明这几组关联关系大。

氧化钾与氧化钙的关系，是由于在烧制过程中草木灰作为助燃剂需要加入稳定剂石灰石来稳定。氧化铁与氧化铜的关系，是由于两者属于着色剂因子，在上色过程中相互作用来上色。氧化铅与氧化钡的关系，是由于两者在高钾玻璃的制备过程中基本不会出现。氧化 Sr 与氧化镁的关系，是由于两者属于调节玻璃折光程度因子，用于共同调节玻璃的折光性。

对于未风化的铅钡玻璃

氧化铅与二氧化硅的相关系数 -0.6024 较大，氧化钡与二氧化硅的相关系数 -0.6513 较大，说明这两组关联关系大。

从化学性质来看，氧化铅、氧化钡与二氧化硅的负向关系大，是由于铅矿石的加入会使得氧化铅、氧化钡的含量增大而二氧化硅的含量减少，例如问题 2 亚类分析中的 K4、K5。

► 风化：

对于高钾风化类的玻璃，通过 *SPSSPRO* 计算其与化学成分之间的相关性系数，我们得到相关性系数表如图 15 所示（热力图中的具体数值参见支撑材料）：

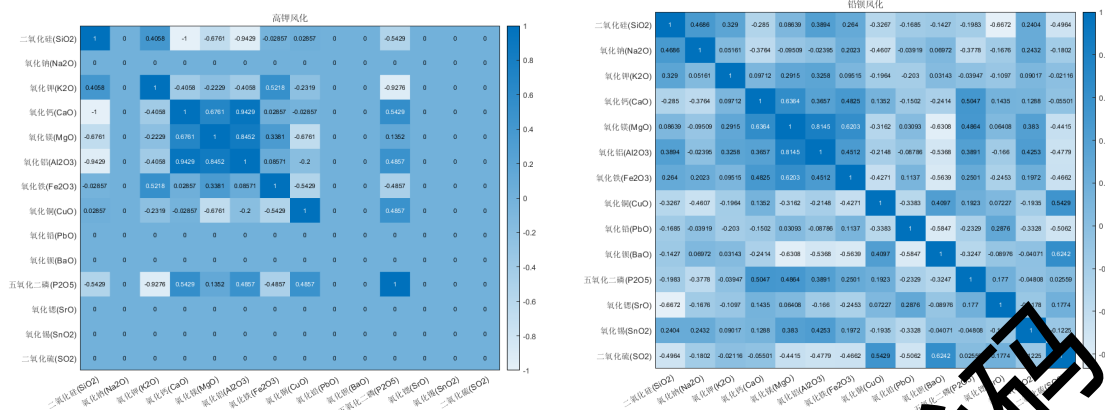


图 15 高钾（左）、铅钡（右）玻未风化的相关系数热力图

通过分析图 15，我们可以发现：

对于风化的高钾玻璃

氧化铝与氧化钙的相关系数 0.9429, 氧化铝与氧化镁的相关系数 0.8452, 氧化钙与氧化镁的相关系数 0.6761, 二氧化硅与氧化镁的相关系数 -0.6761 , 二氧化硅与氧化铝的相关系数 -0.9429 , 二氧化硅与氧化钙的相关系数 -1 , 说明这几组关联关系大。

从化学性质来看：氧化铝、氧化钙、氧化镁三者之间的正向关系大，且三者与二氧化硅的负向关系大，是由于高钾玻璃在风化过程中氧化铝、氧化钙、氧化镁转化为二氧化硅。

对于风化的铅钡玻璃

氧化铝与氧化镁的相关系数 0.8145, 氧化钙与氧化镁的相关系数 0.6364, 氧化铁与氧化镁的相关系数 0.6203, 二氧化硅与氧化钡的相关系数 -0.6672 , 氧化镁与氧化钡的相关系数 -0.6308 , 说明几组关联关系大。

8.3 差异分析

差异系数定义

为分析不同类别玻璃的化学成分关联性之间的差异性，我们定义差异系数 β 。对于两种不同类别的玻璃，记一对化学成分的相关系数分别为 $P1$ 和 $P2$ ，则有

$$\beta = ||P1| - |P2||$$

若差异系数 $\beta > 0.8$ 时，则说明不同类别之间该化学成分的相关关系差异较大，有必要对该对化学成分进行差异性分析。

风化前后两种玻璃各自的差异性：

下面我们分别对高钾风化前后的差异性、铅钡风化前后的差异性进行分析，分析玻璃化学组成成分关联关系的差异性。首先分别得到高钾和铅钡玻璃在风化前后的差异系数热力图，如图 16 所示（热力图中的具体数值参见支撑材料）：

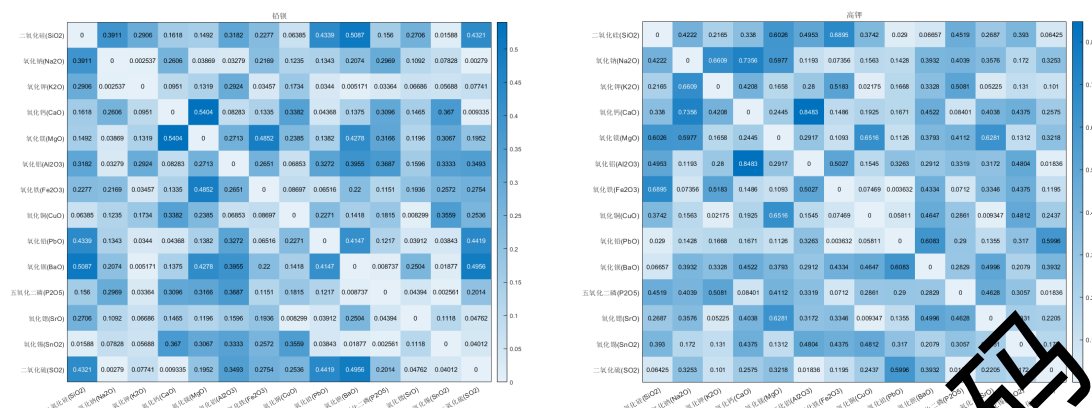


图 16 铅钡（左）、高钾（右）玻璃风化前后的差异系数热力图

对于高钾类玻璃，其中氧化铝与氧化镁的差异系数较大，为 0.8483，说明这一对化学成分的关系在风化前后差异较大。这是由于在未风化时，两者的关系仅与制备时加入的辅助剂含量有关，而氧化铝与氧化镁这两种辅助剂因子并没有显著的相互影响，两者关联小。而风化后，两者共同被转化，两者关联大，因此风化前后关系差异较大。

对于铅钡类玻璃，风化前后的化学成分差异系数差别不超过 0.8，不进行差异性分析。

► 未风化的高钾与铅钡的差异性

下面我们固定表面风化属性为未风化，对高钾与铅钡两个类别的差异性进行分析。得到未风化玻璃中铅钡和高钾的差异系数热力图，如图 17 所示（热力图中的具体数值参见支撑材料）：

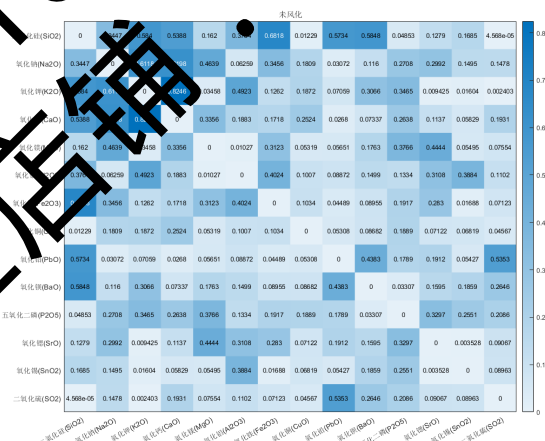


图 17 未风化的高钾与铅钡差异系数热力图

对于未风化类玻璃，其中氧化钙与氧化钾的差异系数较大，为 0.8246，说明这两者关系风化前后差异较大。这是由于高钾玻璃在制备过程中需要同时加入草木灰与石灰石，故在高钾玻璃中氧化钙与氧化钾关联大；而铅钡玻璃并不需要，故两者关联小。

九、模型的评价

9.1 模型的优点

- 问题一中，利用灰色关联分析，将 14 种化学成分区分为 6 种易风化成分与 8 种非风化成分。
- 问题一中，通过化学关系，对风化前后物质的变化定义成分变化系数，通过该系数对风化点进行风化前预测。
- 问题二中，基于决策树，给出了效果优秀的玻璃类型预测方法，准确率达到 100%。
- 问题二中，对未风化的玻璃进行二级亚类划分，将其分成具有明显化学成分量差异的 6 类，符合实际情况。

9.2 模型的缺点

- 问题一中构建的预测模型可以考虑更多约束条件，未做到，使模型更加符合实情。
- 数据预处理中删除了少部分数据，忽略了此数据带来的影响，导致分析不够全面。

9.3 模型的推广

考虑到古玻璃的风化作用本质上是一种化学反应，假定该批古代玻璃制品生产时间相同、所经风化时间相同，则其风化前后化学成分含量的变化应与化学反应速率相关。已知影响化学反应速率的因素有反应物本身的性质、温度、催化剂、浓度、压强、接触面积等。如能掌握这些信息，以及规模足够大的样本数据，可建立一个以时间和上述影响因素为自变量、以反应物和生成物质量为因变量的微分方程模型，对反应一定时间后玻璃中各化学成分的含量进行更加精确的预测。

参考文献

- [1] 黄启善. 广西古代玻璃制品的发现及其研究 [J]. 考古, 1988(03): 264-276+295.
- [2] 段浩, 干福熹, 赵虹霞. 实验室模拟过渡金属离子掺杂的中国古代玻璃的着色现象 [J]. 硅酸盐学报, 2009, 37(12): 1982-1989.
- [3] 王承遇, 陶瑛. 硅酸盐玻璃的风化 [J]. 硅酸盐学报, 2003(01): 78-85.
- [4] 段浩, 干福熹, 赵虹霞. 实验室模拟过渡金属离子掺杂的中国古代玻璃的着色现象 [J]. 硅酸盐学报, 2009, 37(12): 1982-1989.

[5] 伏修锋, 干福熹. 基于多元统计分析方法对一批中国南方和西南地区的古玻璃成分的研究 [J]. 文物保护与考古科学, 2006(04):6-13. DOI:10.16334/j.cnki.cn31-1652/k.2006.04.002.

如需word论文和完整源代码
淘宝店铺：亦是远方

附录 A 文件列表

表 15 支撑材料文件列表

文件名	文件描述
Question1.m	问题一代码
Question2.m	问题二代码
Question3.m	问题三代码
Question4.m	问题四代码
classify.m	问题二代码
drawp10.py	敏感分析绘图代码
xishu.py	斯图尔曼绘制代码
analysis.py	叠加柱状图代码
第四问不同类别的化学成分之间的相关系数.xlsx	第四问不同类别的化学成分之间的相关系数
风化点预测结果.xlsx	风化点预测结果
灰色关联度（高钾、铅钡）.xlsx	灰色关联度（高钾、铅钡）
问题 2 训练集模拟结果.xlsx	问题 2 训练集模拟结果
高钾风化.xlsx	高钾风化数据
高钾未风化.xlsx	高钾未风化数据
铅钡风化.xlsx	铅钡风化数据
铅钡未风化.xlsx	铅钡未风化数据
总表_编码.xlsx	附件数据的重新整理

附录 B 代码

```

Question1.m
NOT_FH_K=xlsread('高钾未风化.xlsx');
NOT_FH_K(isnan(NOT_FH_K))=0;
FH_K=xlsread('高钾风化.xlsx');
FH_K(isnan(FH_K))=0;
NOT_FH_Pb=xlsread('铅钡未风化.xlsx');
NOT_FH_Pb(isnan(NOT_FH_Pb))=0;
FH_Pb=xlsread('铅钡风化.xlsx');
FH_Pb(isnan(FH_Pb))=0;
NOT_FH_K(12,:)=[];NOT_FH_K(10,:)=[];

```



```

for j=1:14
    Percent_compare(1,j)=sum(NOT_FH_K(:,j))./12;
end
for j=1:14
    Percent_compare(2,j)=sum(FH_K(:,j))./6;
end
for j=1:14
    Percent_compare(3,j)=sum(NOT_FH_Pb(:,j))./23;
end
for j=1:14
    Percent_compare(4,j)=sum(FH_Pb(:,j))./26;
end

Percent_K=zeros(1,14);
Percent_Pb=zeros(1,14);
for i=1:14
    Percent_K(i)=(Percent_compare(2,i)-Percent_compare(1,i))/Percent_compare(1,i);
    Percent_Pb(i)=(Percent_compare(4,i)-Percent_compare(3,i))/Percent_compare(3,i);
end
feature_K=[1 3 4 6 8 11];
feature_Pb=[1 4 8 9 10 11];
Predict_K=[];Predict_Pb=[];
for i=1:6
    Predict_K=[Predict_K;Percent_compare(1,:)];
end
for i=1:26
    Predict_Pb=[Predict_Pb;Percent_compare(3,:)];
end
for i=1:6
    for j=1:6
        Predict_K(i,feature_K(j))=Percent_K(i,feature_K(j))./(1+Percent_K(feature_K(j)));
    end
end
for i=1:26
    for j=1:6
        Predict_Pb(i,feature_Pb(j))=Percent_Pb(i,feature_Pb(j))./(1+Percent_Pb(feature_Pb(j)));
    end
end

P1=corr([NOT_FH_K ones(12,1);FH_K 2.*ones(6,1)], 'type', 'Spearman');
P2=corr([NOT_FH_Pb ones(23,1);FH_Pb 2.*ones(6,1)], 'type', 'Spearman');
[~,id1]=sort(abs(P1(1:14,15)));
[~,id2]=sort(abs(P2(1:14,15)));

figure

```

```

title('高钾、铅钡风化前后各种化学成分含量的对比')
i=1;
subplot(4,7,i)
b=bar([Percent_compare(1,i) Percent_compare(2,i)]);
b.FaceColor = 'flat';
b.CData(1,:) = [1,0.38,0.27];
b.CData(2,:) = [0,0.75,1];
ylim([0 100])
title(Name(i))
ylabel('高钾')

for i=2:14
    subplot(4,7,i)
    b=bar([Percent_compare(1,i) Percent_compare(2,i)]);
    b.FaceColor = 'flat';
    b.CData(1,:) = [1,0.38,0.27];
    b.CData(2,:) = [0,0.75,1];
    ylim([0 2*max(Percent_compare(1,i),Percent_compare(2,i))])
    title(Name(i))
end
i=15;
subplot(4,7,i)
b=bar([Percent_compare(3,i-14) Percent_compare(4,i-14)]);
b.FaceColor = 'flat';
b.CData(1,:) = [1,0.38,0.27];
b.CData(2,:) = [0,0.75,1];
ylim([0 100])
title(Name(i-14))
ylabel('铅钡')
for i=16:28
    subplot(4,7,i)
    b=bar([Percent_compare(3,i-14) Percent_compare(4,i-14)]);
    b.FaceColor = 'flat';
    b.CData(1,:) = [1,0.38,0.27];
    b.CData(2,:) = [0,0.75,1];
    ylim([0 2*max(Percent_compare(3,i-14),Percent_compare(4,i-14))])
    title(Name(i-14))
end

```

Question2.m

```

%决策树
Total=xlsread('总表_编码.xlsx');
t=fitctree(Total(1:55,[5 7:20]),Total(1:55,3));
view(t,'Mode','graph')
t=fitctree(Total(1:55,[5 7:14 16:20]),Total(1:55,3));
view(t,'Mode','graph')

```

```

t=fitctree(Total(1:55,[5 7:14 17:20]),Total(1:55,3));
view(t,'Mode','graph')
t=fitctree(Total(1:55,[5 7:14 17 19 20]),Total(1:55,3));
view(t,'Mode','graph')
t=fitctree(Total(1:55,[5 7:8 10:14 17 19 20]),Total(1:55,3));
view(t,'Mode','graph')

```

%K均值聚类

```

NOT_FH_K=xlsread('高钾未风化.xlsx');
NOT_FH_K(isnan(NOT_FH_K))=0;
FH_K=xlsread('高钾风化.xlsx');
FH_K(isnan(FH_K))=0;
NOT_FH_Pb=xlsread('铅钡未风化.xlsx');
NOT_FH_Pb(isnan(NOT_FH_Pb))=0;
FH_Pb=xlsread('铅钡风化.xlsx');
FH_Pb(isnan(FH_Pb))=0;
NOT_FH_K(12,:)=[];NOT_FH_K(10,:)=[];

```

%一级亚类划分

```

K=[FH_K;NOT_FH_K];
Pb=[FH_Pb;NOT_FH_Pb];
C_Pb=Pb(:,[1 3 4 9 10 11 12]);
[IDX,C]=kmeans(C_Pb,2);
C_K=K(:,[1 3 4 9 10 11 12]);
[IDX,C]=kmeans(C_K,2);

```

%二级亚类划分

```

Classify_Pb=NOT_FH_Pb(:,[1 3 4 9 10 11 12]);
[IDX,C]=kmeans(Classify_Pb,2);
[IDX,C]=kmeans(Classify_Pb,3);
[IDX,C]=kmeans(Classify_Pb,4);
[IDX,C]=kmeans(Classify_Pb,5);

Classify_K=NOT_FH_K(:,[1 3 4 9 10 11 12]);
[IDX,C]=kmeans(Classify_K,2);
[IDX,C]=kmeans(Classify_K,3);
[IDX,C]=kmeans(Classify_K,4);
[IDX,C]=kmeans(Classify_K,5);

```

classify.m

```

function C=classify(Percent)
if ~isnan(Percent(9))
    if Percent(9)>=5.46
        C=2;
        return
    end
    if Percent(9)<5.46
        C=1;
    end
end

```

```

        return
    end
end
if ~isnan(Percent(10))
    if Percent(10)>=3.06
        C=2;
        return
    end
    if Percent(10)<3.06
        C=1;
        return
    end
end
if ~isnan(Percent(12))
    if Percent(12)>=0.155
        C=2;
        return
    end
    if Percent(12)<0.155
        C=1;
        return
    end
end
if Percent(1)<57.11
    C=1;
    return
end
if Percent(1)>=57.11
    if ~isnan(Percent(3))
        if Percent(3)>=0.445
            C=1;
            return
        end
        if Percent(3)<0.445
            C=2;
            return
        end
    end
end
if ~isnan(Percent(8))
    if Percent(8)>=0.76
        C=1;
        return
    end
    if Percent(8)<0.76
        C=2;
        return
    end
end

```

```

end
end
end

```

Question3.m

```

Percent=xlsread('附件.xlsx','表单3');
[m,~]=size(Percent);
Predict_Q3=zeros(m,1);
for i=1:m
    Predict_Q3(i)=classify(Percent(i,:));
end
Percent_ep=[];
for i=1:9
    Percent_ep=Percent;
    for j=1:8
        Percent_ep(j,11)=Percent_ep(j,11)*(1+(i-5)*0.05);
        C(j,i)=classify(Percent_ep(j,:));
    end
end
end

```

Question4.m

```

subplot(1,2,1)
P_FH_K=corr(FH_K,'type','Spearman');
P_FH_K(isnan(P_FH_K))=0;
xlabel_name = Name;
ylabel_name = Name;
hot_figure = heatmap(xlabel_name,ylabel_name,P_FH_K,'FontSize',10);
hot_figure.GridVisible = 'off';
hot_figure.Title = '高钾风化';

P_FH_Pb=corr(FH_Pb,'type','Spearman');
xlabel_name = Name;
ylabel_name = Name;
hot_figure = heatmap(xlabel_name,ylabel_name,P_FH_Pb,'FontSize',10);
hot_figure.GridVisible = 'off';
hot_figure.Title = '铅钨风化';
subplot(1,2,2)
P_NOT_FH_K=corr(NOT_FH_K,'type','Spearman');
xlabel_name = Name;
ylabel_name = Name;
hot_figure = heatmap(xlabel_name,ylabel_name,P_NOT_FH_K,'FontSize',10);
hot_figure.GridVisible = 'off';
hot_figure.Title = '高钾未风化';

P_NOT_FH_Pb=corr(NOT_FH_Pb,'type','Spearman');

```

```

xlabel_name = Name;
ylabel_name = Name;
hot_figure = heatmap(xlabel_name,ylabel_name,P_NOT_FH_Pb,'FontSize',10);
hot_figure.GridVisible = 'off';
hot_figure.Title = '铅钨未风化';

```

```

P_NOT_FH=abs(abs(P_NOT_FH_K)-abs(P_NOT_FH_Pb));
xlabel_name = Name;
ylabel_name = Name;
hot_figure = heatmap(xlabel_name,ylabel_name,P_NOT_FH,'FontSize',10);
hot_figure.GridVisible = 'off';
hot_figure.Title = '未风化';

```

```

P_FH=abs(abs(P_FH_K)-abs(P_FH_Pb));
xlabel_name = Name;
ylabel_name = Name;
hot_figure = heatmap(xlabel_name,ylabel_name,P_FH,'FontSize',10);
hot_figure.GridVisible = 'off';
hot_figure.Title = '风化';

```

```

P_K=abs(abs(P_FH_K)-abs(P_NOT_FH_K));
xlabel_name = Name;
ylabel_name = Name;
hot_figure = heatmap(xlabel_name,ylabel_name,P_K,'FontSize',10);
hot_figure.GridVisible = 'off';
hot_figure.Title = '高钾';

```

```

P_Pb=abs(abs(P_FH_Pb)-abs(P_NOT_FH_Pb));
xlabel_name = Name;
ylabel_name = Name;
hot_figure = heatmap(xlabel_name,ylabel_name,P_Pb,'FontSize',10);
hot_figure.GridVisible = 'off';
hot_figure.Title = '铅钨';

```

```

draw(10,py)

```

```

# encoding=utf-8
import matplotlib.pyplot as plt
from pylab import * #支持中文
mpl.rcParams['font.sans-serif'] = ['SimHei']

```

```

def remove_edge(plt):
    fig, ax = plt.subplots()
    ax.spines["top"].set_visible(False)
    #ax.spines["bottom"].set_visible(False)
    ax.spines["right"].set_visible(False)
    #ax.spines["left"].set_visible(False)

```

```

# 改变长宽比
remove_edge(plt);
names = ['-30%', '-25%', '-20%', '-15%', '-10%', '-5%', '+5%', '+10%', '+15%', '+20%', '+25%', '+30%']
x = range(len(names))

y = [92.5, 92.5, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100]
#y1= [100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 92.5, 92.5]

plt.plot(x, y, 'ro-')
plt.plot(x, y1, 'bo-')
plt.xlim(-1, 12) # 限定横轴的范围
plt.ylim(75, 101) # 限定纵轴的范围

plt.plot(x, y, marker='o', mec='r', mfc='w', label=u'二氧化硅最大值')
plt.plot(x, y1, marker='*', ms=10, label=u'二氧化硅最小值')

plt.legend() # 让图例生效
plt.xticks(x, names, rotation=45)
plt.margins(0)
plt.subplots_adjust(bottom=0.15)
plt.xlabel(u"扰动幅度", size=24) # X轴标签
plt.ylabel("准确率 (%) ", size=24) # Y轴标签
plt.title("A simple plot") # 标题
plt.rcParams["figure.figsize"] = (10, 6)
plt.tick_params(labelsize=24)

plt.savefig('sio2mx.png', dpi=20*12) # 保存图片
plt.savefig('sio2mn.png', dpi=20*12) # 保存图片

plt.show()

```

xishu.py

```

import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
pd.set_option('expand_frame_repr', False)
df = pd.read_csv('xishu.csv')
y_ticks = ['表面风化', '颜色', '类型', '纹饰']
x_ticks = ['纹饰', '类型', '颜色', '表面风化'] # 自定义横纵轴标签
plt.rcParams['font.sans-serif'] = ['SimHei'] # 可以正常显示中文
plt.rcParams['axes.unicode_minus'] = False # 正常显示负号
ax = sns.heatmap(df, annot=True, fmt='.4f', cmap="Reds", vmax=1, vmin=-1, xticklabels=x_ticks,
                 yticklabels=y_ticks, annot_kws={'size': 16}, cbar=False)
# annot=True表示每一个格子显示数字;fmt='.0f'表示保留0位小数,同理fmt='.1f'表示保留一位小数
# camp表示颜色,在另一个博主的文章中讲解的很清晰
# vmax=350, vmin=20表示右侧颜色条的最大最小值,在最大最小值外的颜色将直接以最大或最小值的颜色显示,

```

```
# 通过此设置就可以解决少数值过大从而使得大部分色块区别不明显的问题
# xticklabels=x_ticks, yticklabels=y_ticks, 横纵轴标签
#ax.set_title('分速度报警类型分布') # 图标题
ax.set_xlabel('纹饰',size=16) # x轴标题
ax.set_ylabel('风化程度',size=16) # y轴标题
```

```
#调整图片字体大小
```

```
font = {'family':'simhei',
        'weight': 'normal',
        'size': 16,
        };
```

```
#ax.legend(['未风化','风化'],prop=font,labelspace=1);#这样就可以设置标签字体大小
```

```
ax.tick_params(labelsize=16) #设置色标刻度字体大小
```

```
cb=ax.figure.colorbar(ax.collections[0]) #显示colorbar
```

```
cb.ax.tick_params(labelsize=16) #设置colorbar刻度字体大小。
```

```
plt.rcParams["figure.figsize"] = (12, 6)
```

```
plt.show()
```

```
figure = ax.get_figure()
```

```
figure.savefig('xishu.png',dpi=24*12) # 保存图片
```

analysis.py

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
plt.rcParams['font.sans-serif'] = ['SimHei']
```

```
plt.rcParams['axes.unicode_minus'] = False
```

```
#条形图的绘制
```

```
labels = ['浅绿', '浅蓝', '深绿', '浅紫', '紫', '绿', '蓝绿', '黑']
```

```
men_means = [2,8,5,2,2,1,6,0]
```

```
women_means = [2,12,4,0,2,0,6,2]
```

```
width = 0.35 # 条形图的宽度
```

```
fig,ax=plt.subplots(1)
```

```
ax.bar(labels, men_means,width,label='未风化')
```

```
ax.bar(labels, women_means,width,bottom=men_means,label='风化')
```

```
ax.set_ylabel('数目',size=16)
```

```
ax.set_xlabel('颜色',size=16)
```

```
ax.set_title('古代玻璃制品表面风化与其颜色的关系',size=16)
```

```
plt.yticks(np.arange(0,22,2))
```

```
#调整图片大小
```

```
font = {'family':'simhei',
        'weight': 'normal',
        'size': 15,
```



```
};  
ax.legend(['未风化','风化'],prop=font,labelsacing=1);#这样就可以设置标签字体大小  
ax.tick_params(labelsize=16) #设置色标刻度字体大小  
plt.rcParams["figure.figsize"] = (6, 5)  
  
plt.show()  
plt.savefig('C1(1).png',dpi = 1600); #保存图像, dpi可以调整图像的像素大小
```

如需word论文和完整源代码
淘宝店铺：亦是远方