

题目

摘要

摘要

关键字：关 键 字

一、问题重述与问题分析

1.1 问题重述

玻璃是早期丝绸之路的重要商品，从西亚和埃及地区传入我国后，古人因地制宜、就地取材制作出外观相似但化学成分不同的玻璃制品。玻璃炼制时需要助熔剂，不同的助熔剂会导致玻璃的化学成分不同，例如铅钡玻璃和钾玻璃分别以铅矿石和草木灰作为助熔剂。同时，古代玻璃易与外界环境进行元素交换而导致化学成分比例发生改变。

现有提供高钾玻璃和铅钡玻璃的相关数据，完成下列问题：

1. 分析玻璃文物表面风化和玻璃类型、纹饰和颜色的关系；对不同的玻璃类型分析表面风化与否和玻璃化学成分含量之间的统计规律；根据风化点的检测数据预测风化前的化学成分含量；
2. 根据附件数据分析铅钡玻璃和高钾玻璃的分类规律；对不同的玻璃类型，选择适当的化学成分划分亚类，并给出具体的方法和结果，并对分类结果进行合理性和敏感性分析；
3. 针对附件表单 3 的化学成分，鉴别其所属的类型，同样需要对结果进行敏感性分析；
4. 分析不同类别玻璃文物的化学成分之间的关系以及比较不同类别化学成分关系的差异性。

1.2 问题分析

1.2.1 数据预处理

1.2.2 问题一的分析

问题一要求分析附件表单 1 中文物表面风化与玻璃类型、颜色和纹饰之间的关系，随后分析表单 2 数据，得出风化与化学成分含量的统计规律，最后依据表单 2 风化点的检测数据，预测风化前的化学成分含量。本文首先对附件表单 1 的数据进行可视化，作出以玻璃类型、纹饰和颜色为自变量，是否风化为因变量的三维气泡图，再对高钾玻璃和铅钡玻璃分别作风化与纹饰和颜色的二维散点图得到四者之间的定性关系。未完待续

1.2.3 问题二的分析

问题二要求分析两类玻璃的分类规律，之后选取合适的化学成分划分亚类并给出方法和具体结果，并对亚类划分结果做合理性和敏感性分析。本文首先作出不同玻璃文物的不同化学成分含量的散点图，得出初步的高钾玻璃和铅钡玻璃的分类规律。未完待续

1.2.4 问题三的分析

问题三要求预测附件表单 3 中文物的类型，并对结果做敏感度分析。未完待续

1.2.5 问题四的分析

问题四要求分析同类玻璃化学成分的联系以及不同类别化学成分联系的差异性。未完待续

二、模型假设与符号说明

2.1 符号说明

三、数据预处理

3.1 数据清洗

3.1.1 处理缺失值

对于表单 1 的数据，我们首先找出缺失值，缺失值数据如表1所示

表 1 表单 1 缺失值数据

文物编号	纹饰	类型	颜色	表面风化
19	A	铅钡	NaN	风化
40	C	铅钡	NaN	风化
48	A	铅钡	NaN	风化
58	C	铅钡	NaN	风化

可以观察到，表单 1 的数据仅存在颜色属性的缺失，缺失数据量为 4，缺失量较少，可以忽略。

3.1.2 处理异常值

3.2 数据规约

由于表单 2 中不同的化学成分含量相差较大，故考虑首先对各列化学成分数据进行标准化处理，使处理后的数据更加能体现该化学成分含量的相对大小，使数据更加直观且具有可比性。此处使用了最小——最大规范化方法 (minmax-scale)，方法公式如下：

$$x_i = \frac{x - x_{min}}{x_{max} - x_{min}}$$

四、问题 1 模型的建立与分析

4.1 风化结果与玻璃类型、纹饰与颜色的关系的描述性分析

首先对数据的整体进行刻画，绘制以玻璃类型、纹饰和颜色为自变量，是否风化为因变量的三维散点气泡图，如图1所示。

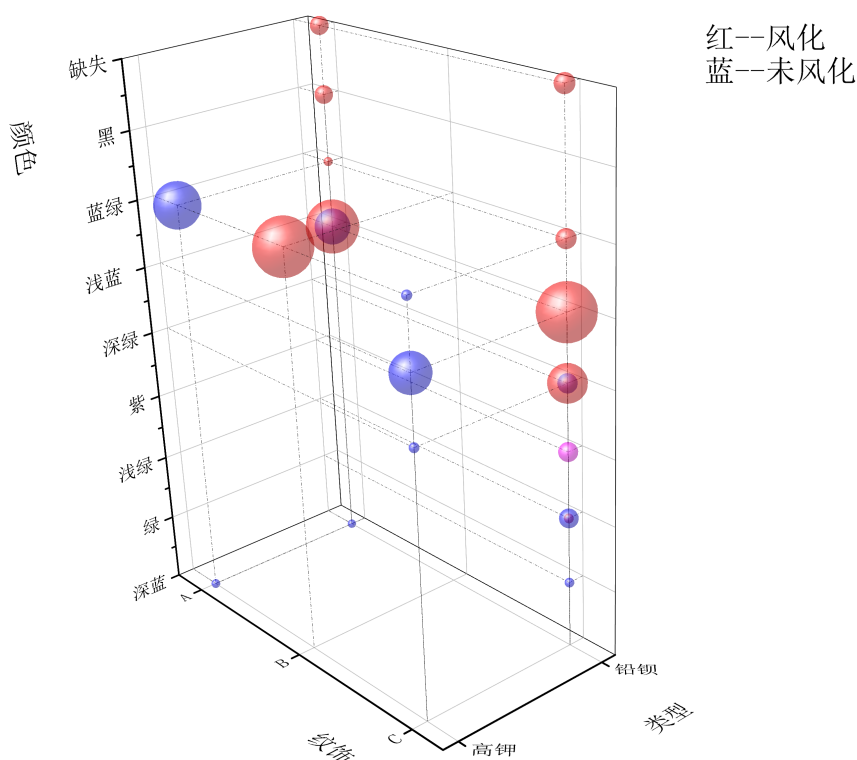


图 1 描述数据整体特征的三维散点气泡图

为了得到更加具体的关系特征，对图像进行降维处理，分别得到高钾玻璃和铅钡玻璃的纹饰、颜色与有无分化关系的二维散点图，如图2所示。

由图，可以得出如下的规律：

- 由高钾玻璃的二维图，可以观察到高钾玻璃中 B 纹饰——紫颜色的种类风化结果表现为风化，且数据样本较多，其他类型的纹饰与颜色均表现为未风化；
- 由铅钡玻璃的二维图，可以观察到 A 纹饰的浅蓝色同时表现出了未风化、风化两种结果，推测当 A 纹饰的铅钡玻璃介于风化与未风化的过渡地带时，往往呈现出浅蓝色，而 A 纹饰的其他颜色都具有较好的区分属性，如 A 纹饰——深蓝色表现为未风化，而 A 纹饰——蓝绿色/黑色/缺失颜色均表现为风化；
- 由铅钡玻璃的二维图，可以观察到 C 纹饰的深绿色、紫色、浅绿色同时表现出了未风化、风化两种结果，推测当 C 纹饰的铅钡玻璃介于风化与未风化的过渡地带时，往往呈现出这三种颜色，而 C 纹饰的其他颜色，如 C 纹饰——绿色表现为未风化，

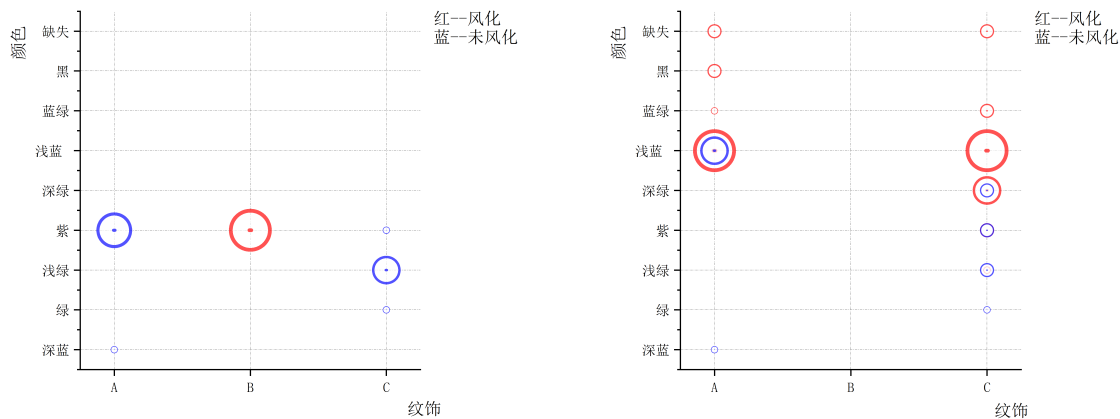


图2 描述高钾玻璃（左）和铅钡玻璃（右）数据特征的二维散点图

C 纹饰——蓝绿色/缺失颜色均表现为风化。

基于上述的规律，本文做出了三个维度数据与风化结果的关系特征的推导：

- 玻璃类型会对风化结果产生较显著的影响，铅钡玻璃风化比重比高钾玻璃大
- 风化过程中会存在颜色的渐变，这种渐变过程会因玻璃类型和纹饰存在差异，如铅钡玻璃的 C 纹饰，初始为绿色，经过风化由于化学成分改变逐渐变成浅绿、深绿的过度颜色，最终被完全风化，变成蓝绿色。

4.2 基于皮尔逊相关系数的按含量比例分类的分类模型

本文设计了基于皮尔逊相关系数的按含量比例分类的模型，具体流程如下所示：

首先针对高钾玻璃做具体的流程分析：

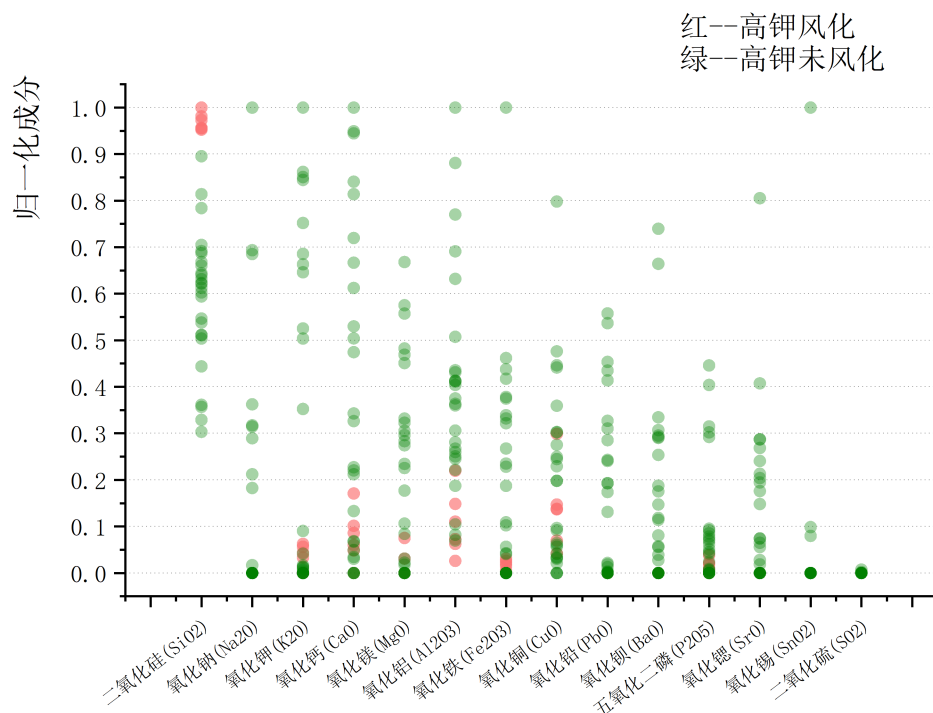
- **Step1:** 绘制高钾玻璃风化前后的各个归一化后的化学成分含量的散点图，观察数据特征。

由图3可以初步得出，高钾玻璃风化前后对部分化学成分含量有较大影响，其中，对 SiO_2 的含量有正向影响、对 K_2O 、 CaO 、 MgO 、 Al_2O_3 、 Fe_2O_3 、 CuO 、 P_2O_5 的含量有负向影响。

- **Step2:** 计算高钾玻璃各个化学成分的皮尔逊相关系数。皮尔逊相关系数的计算公式如下所示：

$$\rho_{X_i, X_j} = \frac{\text{cov}(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}} = \frac{E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})]}{\sigma_{X_i} \sigma_{X_j}}$$

若某两种皮尔逊相关性较大，则可以推测在风化前后这两种化学成分极可能存在某种化学转换关系，结合步骤 1，由于众多成分中只有 SiO_2 含量明显减少，故主要探究 SiO_2 与其他化学成分含量的皮尔逊系数，将较大的系数值标粗，得到的表如表2所示。



以皮尔逊相关系数 0.6 为分界, 选出与 SiO_2 相关性最强的五个化学成分, 分别为: K_2O 、 CaO 、 MgO 、 Al_2O_3 、 Fe_2O_3 , 此结果与步骤一的结果得到了相互印证。

$$p_1 = \frac{n(SiO_2)}{n(K_2O) + n(CaO) + n(MgO) + n(Al_2O_3) + n(Fe_2O_3)}$$

依照表格信息，可以选取 $p_1 = 10.7$ 作为区分高钾玻璃风化与否的依据。

当高钾玻璃中的 $p_1 = \frac{n(\text{SiO}_2)}{n(\text{K}_2\text{O})+n(\text{CaO})+n(\text{MgO})+n(\text{Al}_2\text{O}_3)+n(\text{Fe}_2\text{O}_3)}$ 大于 10.7 时, 高钾玻璃出现风化, 小于 10.7 时, 高钾玻璃不风化。

表 2 高钾各成分间相关性

	二氧化硅	氧化钠	氧化钾	氧化钙	氧化镁	氧化铝	氧化铁
二氧化硅	1.000	-0.457	-0.877	-0.823	-0.603	-0.831	-0.687
氧化钠	-0.457	1.000	0.554	0.621	-0.245	0.345	-0.023
氧化钾	-0.877	0.554	1.000	0.759	0.387	0.591	0.336
氧化钙	-0.823	0.621	0.759	1.000	0.205	0.536	0.457
氧化镁	-0.603	-0.245	0.387	0.205	1.000	0.698	0.595
氧化铝	-0.831	0.345	0.591	0.536	0.698	1.000	0.689
氧化铁	-0.687	-0.023	0.336	0.457	0.595	0.689	1.000
氧化铜	-0.465	-0.022	0.251	0.423	0.141	0.210	0.526
氧化铅	-0.398	0.359	0.275	0.368	0.166	0.422	0.151
氧化钡	-0.345	-0.214	0.093	0.078	0.427	0.381	0.452
五氧化二磷	-0.448	-0.186	0.171	-0.031	0.650	0.676	0.765
氧化锶	-0.525	-0.182	0.421	0.035	0.712	0.595	0.591
氧化锡	0.049	-0.106	0.142	-0.290	0.261	-0.163	-0.219
二氧化硫	-0.357	-0.194	0.332	0.429	0.408	0.147	0.232
	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	-0.465	-0.398	-0.345	-0.448	-0.525	0.049	-0.357
氧化钠	-0.022	0.359	-0.214	-0.186	-0.182	-0.106	-0.194
氧化钾	0.251	0.275	0.093	0.171	0.421	0.142	0.332
氧化钙	0.423	0.368	0.078	-0.031	0.035	-0.290	0.429
氧化镁	0.141	0.166	0.427	0.650	0.712	0.261	0.408
氧化铝	0.210	0.422	0.381	0.676	0.595	-0.163	0.147
氧化铁	0.526	0.151	0.452	0.765	0.591	-0.219	0.232
氧化铜	1.000	0.165	0.521	0.213	0.187	-0.361	0.301
氧化铅	0.165	1.000	0.632	-0.014	0.188	-0.133	-0.244
氧化钡	0.521	0.632	1.000	0.378	0.565	-0.118	-0.216
五氧化二磷	0.213	-0.014	0.378	1.000	0.741	0.065	-0.021
氧化锶	0.187	0.188	0.565	0.741	1.000	0.240	-0.049
氧化锡	-0.361	-0.133	-0.118	0.065	0.240	1.000	-0.108
二氧化硫	0.301	-0.244	-0.216	-0.021	-0.049	-0.108	1.000

参考文献

[1] 参考文献

表 3 高钾玻璃样品的 p_1 值

类型	表面风化	二氧化硅 (SiO2)	氧化钾 (K2O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al2O3)	氧化铁 (Fe2O3)	高钾的比例 (区分风化)
高钾	风化	96.77	0.92	0.21	0.00	0.81	0.26	43.99
高钾	风化	95.02	0.59	0.62	0.00	1.32	0.32	33.34
高钾	风化	92.63	0.00	1.07	0.00	1.98	0.17	28.77
高钾	风化	94.29	1.01	0.72	0.00	1.46	0.29	27.09
高钾	风化	92.72	0.00	0.94	0.54	2.51	0.20	22.13
高钾	风化	92.35	0.74	1.66	0.64	3.50	0.35	13.40
高钾	无风化	60.71	5.71	0.00	0.85	0.00	1.04	7.99
高钾	无风化	87.05	5.19	2.01	0.00	4.06	0.00	7.73
高钾	无风化	79.46	9.42	0.00	1.53	3.05	0.00	5.68
高钾	无风化	76.68	0.00	4.71	1.22	6.19	2.37	5.29
高钾	无风化	61.87	7.44	0.00	1.02	3.15	1.04	4.89

附录 A 附录

```
import math

def hix(x):
    if x > 0:
        return 1+1=2
    else:
        return 0
```