

综合运用 Critic 权重和 k-means 聚类的玻璃分类模型

摘要

玻璃制品经丝绸之路传入中国后，逐渐演变成两种类型：高钾玻璃和铅钡玻璃。受制于埋藏条件，许多玻璃文物受到了不同程度的风化，致使考古工作停滞。本文旨在建立玻璃文物基于化学成分的分类模型以及判断风化对化学成分影响。

针对问题一，我们首先分析玻璃文物的外观与是否风化的关系。利用**费舍尔精确检验**，和**卡方检验**，共同判断三个外观特征（纹饰、颜色、类型）和表面有无风化是否具有显著的相关性，结果显示纹饰、类型与表面有无风化显著相关。此外，三个外观特征对决定是否风化的贡献值大小也是分析外观与风化关系的重要指南。我们使用**决策树（CART）**对各外观指标在分类是否风化过程中的信息增益进行排序，贡献依次为类型、纹饰、颜色。

二分类的分析时，回归方法例如 logistic 回归或许更易解释，然而经分析发现玻璃文物中各个化学成分之间存在明显**内生性**，即所有化学成分比例总和大致等于 1，若采用回归的方式可能出现不一致性。利用统计软件 R 作出各类别（先按类型再按是否风化分类）的化学成分统计规律，并根据风化的实际作用过程，提出了**量化风化前后化学成分变化的方程**，并预测风化点风化前的化学成分。

针对问题二：我们首先通过 **Critic 权重法**对附件中的 14 中化学成分作权重排序，再按权重从大到小选取适当数量的化学成分作为**高钾和铅钡玻璃的分类基准**，它们依次是**二氧化硅、氧化铅、氧化钡、氧化钾和五氧化二磷**。基于 Critic 赋权重的思想，我们在高钾玻璃与铅钡玻璃中分别提取权重最高的五个化学成分作为亚类划分的基准，运用 **k-means 聚类**的方法进行分类，并使用**仿真实验和交叉检验**的方法对模型的合理性和敏感性做检验，仿真结果显示该分类模型在数据集量较小时仍有抗波动能力，且交叉检验也可达到 98%以上的正确性。

针对问题三，对附件表单 3 的数据进行分类实际是对我们上述建立的 Critic 权重模型和亚类划分模型的综合运用和外延。需要将表单三中的文物分类为高钾或铅钡两类，我们将表单 2 中的数据作为“训练集”¹，先用 Critic 提取在风化和无风化两种情况下权重靠前的化学成分作为分类的基准，一定程度上“**监督**”了 **k-means 聚类**的结果，随后进行 k-means 聚类，将其作为高钾和铅钡的分类模型，利用交叉检验分析其敏感性，随后导入表单 3 的数据并预测其类别。再者，我们比对了未风化数据使用随机森林分类、KNN 分类、SVM 分类等监督学习方法与我们模型的结果，也从侧面验证了模型的有效性。

针对问题四：样本量不足六十个的情况下，使用皮尔逊相关系数分析变量间的关联关系不合理，我们使用**斯皮尔曼相关系数**进行分析，制作相关系数矩阵进行类别间的比较。结果显示，在风化和未风化分类中氧化钾与二氧化硅以及氧化铝与氧化镁都具有显著相关性，在高钾玻璃和铅钡玻璃分类中的化学成分相关关系有明显的差异性，它们之间只有 11%化学成分的关联关系是大致一致的。

关键词 Critic 权重法，k-means 聚类，斯皮尔曼相关系数，亚类划分模型

¹ 该过程与普通拟合过程中对训练集训练相似，故称其为“训练集”，在 6.3.1 中详细说明模型建立原理

一、 问题背景与重述

1.1 问题的背景

我国古代从西亚以及埃及地区吸收玻璃制作工艺后，就地取材，制作出了外形相似，化学成分有所不同的玻璃制品。玻璃主要原料是石英砂，主要化学成分是二氧化硅(SiO_2)。由于该成分的熔点较高，在炼制时需要加入草木天然泡碱、硝石等助溶剂，同时需要加入石灰石作为稳定剂。古代玻璃因埋藏条件的原因，极容易风化。风化过程会造成玻璃制品内部化学成分的变化。对玻璃制品的化学成分比例的分析有助于对其类别辨认的正确判断。

1.2 问题的重述

本文将根据附件所给的数据进行统计性描述，分析数据特征并建立模型解决以下问题。

问题一： 根据附件的表单一分析玻璃三种外观参数（类型、纹饰和颜色）与是否表面风化进行相关性分析，进而建立二分类模型。依据表单二的数据通过进行对玻璃的内部参数（14 个化学成分）进行统计性描述，根据统计规律预测风化玻璃风化前的化学成分比例。

问题二： 分析各个化学成分对于玻璃类型定义的权重，找到能作为高钾玻璃和铅钡玻璃分类标准的化学成分。再在每个类型内部选取权重较高的化学成分作为亚类划分的根据作出具体的亚类划分并进行合理性和敏感性分析。

问题三： 利用问题一和问题二使用的模型原理对表单三中的玻璃进行化学成分分析与分类并敏感性分析。

问题四： 从不同类型的玻璃文物样品之间分析化学成分之间的相关性并且比较不同类别之间相关性的差异。

二、 问题分析

针对问题一：

玻璃文物表面风化和三个外观特征（纹饰、颜色和类型）的关系实际上就是分析这三个特征与表面风化与否的相关性，同时进一步分析其分别对决定表面是否风化的贡献大小。

对统计规律的讨论，实际上需要先将文物样品分成表面风化以及表面无风化

两类。分别在这两类中进行描述性统计。根据监测数据进行预测实际上就是对比风化前和风化后文物化学组成比例发生了那些显著的变化，以及量化具体的变化，随后根据量化的结果进行预测。我们使用费舍尔精确检验和卡方检验联合对相关性进行分析，利用决策树（CART）模型对其贡献大小进行排序。

针对问题二：

分析分类规律可以从高钾玻璃和铅钡玻璃分别讨论，对其中各个化学成分进行权重排序，比较两类玻璃文物的高权重化学成分比例的异同。

同样地，在分别对高钾和铅钡玻璃进行亚类划分，我们同理再分别在高钾玻璃和铅钡玻璃中在进行一次权重排序，将剩余化学成分中（剔除已经用作识别高钾、铅钡玻璃的化学成分）权重高的作为亚类划分的依据。模型合理性在数据集较小可以通过添加干扰数据和减少预测数据进行分析，模型的敏感性可通过交叉检验分析。

针对问题三：

对未知文物的化学成分分析并鉴别类型实际是问题二中建立模型的原理的实际运用，主要任务需要重新训练能完成对高钾、铅钡两种类型玻璃划分的聚类模型，再将其运用在表单 3 的数据中。由于数据量较小，采用监督学习可能会带来过拟合的影响，即使使用交叉验证也很难排除其影响，因此本文使用的 Kmeans 无监督算法对玻璃进行二分类，并利用聚类得出的两类与玻璃自身高钾、铅钡的分类结果进行比较，以此来证明该算法的合理性与正确性。分类结果敏感性分析我们将选择分类可信度最小的玻璃进行分析，尝试给各化学元素加入不同百分比成分的波动，分析其分类结果是否会出现波动。同时，我们也尝试使用主成分分析（PCA）方法得出前三名的主成分再进行聚类分析以及随机森林分类、KNN 分类、SVM 分类等监督学习方法对该问题进行分析，从侧面验证了我们模型算法以及结果分类的准确性。**针对问题四：**

分析不同类别的玻璃文物化学成分之间的关联关系，需要按类别将玻璃文物分类。顾名思义，高钾玻璃和铅钡玻璃是存在化学成分差异的，故而分别作出相关系数矩阵。之后再在高钾玻璃中比较表面风化和表面未风化的相关系数矩阵，铅钡玻璃表面风化和未风化的相关系数矩阵。然而数据的样本容量过少，同时在严谨的相关性分析时，不能假设其数据总体符合正态分布，因此我们使用斯皮尔

曼相关系数进行作出相关系数矩阵并比较。需要注意的是，需要先求出化学成分之间相关性是否显著，再可对比相关系数关系。

三、 模型假设

- 附件表单 2 中的化学成分数据总体符合正态分布；
- 风化是一致的，也即所有风化对玻璃化学成分产生的影响不变²；
- 附件中给出的化学成分特征具有代表性。

四、 符号说明

符号	解释
lx	类型，分别有高钾和铅钡
ys	颜色，指玻璃外表颜色
Ws	纹饰，分 A、B、C
X_i	第 i 的化学成分，总共有 14 个，依次对应表单 2 中的化学成分
$\overline{X_{ipo}}$	风化前第 i 个化学成分的均值
$\overline{X_{ipr}}$	风化后第 i 个化学成分的均值
Δ_i	风化发生时第 i 种化学成分的变化值

五、 数据预处理

我们首先去掉了附件表单二中的无效数据（各化学成分加总小于 85%和大于 105%的数据），分别是编号 15 和编号 17 的玻璃文物。

对于表单一的数据，我们分别将三个外观特征（纹饰、类型和颜色）转变成虚拟变量：纹饰的 A、B、C 分别为 1、2、3，类型的高钾、铅钡分别为 0、1，颜色分为绿、蓝绿、蓝、紫、黑分别为 1、2、3、4、5³，由于表单一的数据在颜色这一特征上有检测不到的数据，我们将这些数据去掉，以防止无根据赋值对后续计算和建模产生的偏差。

² 基于参考文献[6]

³ 表单 1 中会给出的颜色值太多，有一些颜色对应的编号比较少，故将浅绿、深绿都定义为绿，蓝色同理

六、 模型建立与求解

6.1 针对问题一：

问题一的总体流程图为：

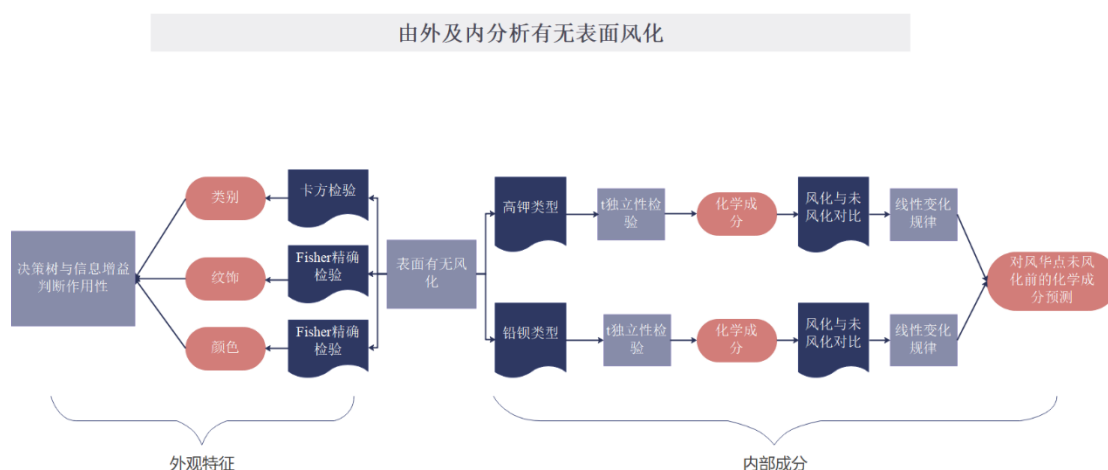


图 6-1 问题一处理流程图

6.1.1 表面风化与类型、纹饰和颜色关系分析

由于附件表单一中给出的三种外观特征的数据与表面风化与否的指标都是类型变量（或者说虚拟变量），使用费舍尔精确检验相比于皮尔逊相关分析更具有现实意义。因此，我们使用**费舍尔精确检验**对三个外观特征和表面风化与否分别分析其之间的相关性。原假设为三个特征分别与表面风化与否独立（即无相关性），备择假设他们之间不独立。同时，由于“类型”中，每一个类（高钾或铅钡）的频数都大于 5，且总样本数量大于 40，因此对类型与表面有无风化的相关分析我们采用**卡方检验**。故，我们用费舍尔精确检验对纹饰、颜色与表面有无风化作分析，用卡方检验对类型与表面有无风化做分析。我们用 R 进行该分析，给出的结果为⁴，

	类型	纹饰	颜色
p-value	0.01955	0.08361	0.3806

表 6-1-1 卡方检验的 p 值结果

⁴ 利用 R 给出的结果在附录

由上表，类型和纹饰与风化的相关性在 90%的置信水平上是相关的，而颜色与风化的相关性是不显著的。

此外，为了弄清楚这三个特征对表面风化与否的决定性作用的大小，我们通过 python 分析，使用决策树模型，找出三个外观特征对结果指标（表面风化与否）信息增益⁵大小。如下图：

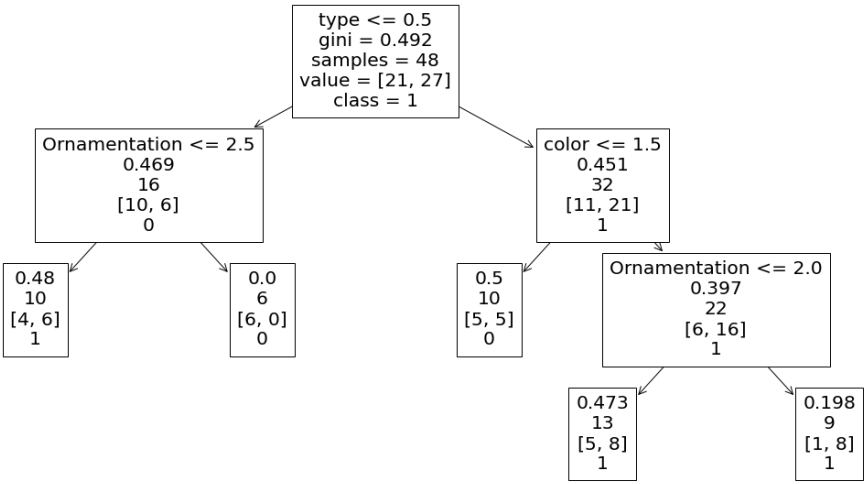


图 6-1-1 决策树分类结果

从决策树可以得到，类型（type）为决定是否表面风化提供了最高的信息增益，其次根据基尼指数（gini）公式⁶：

$$Gini(D) = 1 - \sum_{k=1}^{|y|} p_k^2$$

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

其含义系数含义为，以此纹饰（ornamentation）给是否风化的决定降低了更多的基尼指数，故称其比颜色（color）提供更高的信息增益。

6.1.2 根据类型与是否风化的化学成分统计规律

下图是几种化学成分在四种风化情况下（表面风化文物风华点、表面无风化文物、表面风化文物无风化点、表面风化文物严重风化点）的数据表现。

⁵ 一般而言，信息增益越大，意味着使用该属性（特征）进行划分所获得的“纯度提升”越大。

⁶ P 是属于 k 类的概率，a 为分类依据。

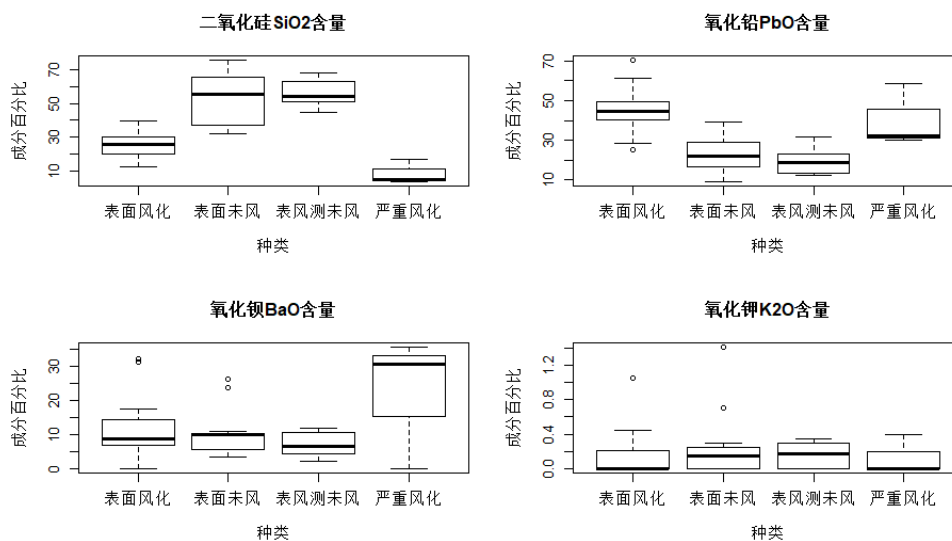


图 6-1-2_1 化学成分在不同风化情况下的箱式图

表面无风化文物的各种化学成分统计特征和表面风化文物无风化点的统计特征基本一致，故我们认为他们之间化学成分是无差异的，也即表面风化文物无风化点的化学成分在该文物经历风化前后没有发生变化。

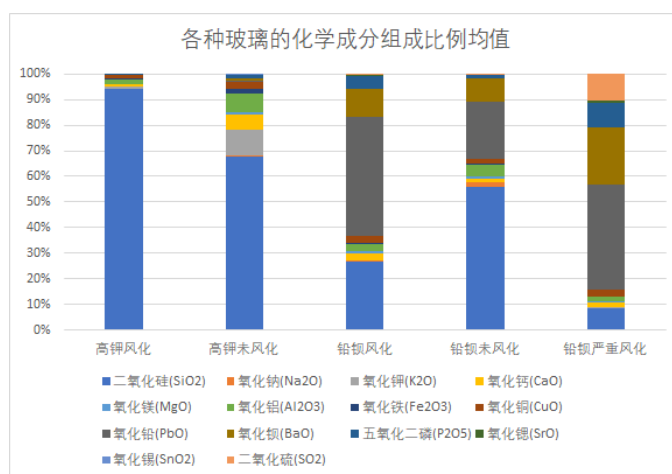


图 6-1-2_2 各种玻璃的化学成分组成和比例均值

上图是将玻璃文物分成四种状态（高钾风化、高钾未风化、铅钡风化、铅钡未风化和铅钡严重风化），显而易见各个状态下的化学成分比例差距明显。

6.1.3 风化前后化学成分预测模型与求解

我们通过运用 Excel 和 R 对表单二中的化学成分进行统计分析，依据类型和表面有无风化为根据计算其重要的统计特征，同时可视化这些化学成分的统计特征。在预测风化前化学成分前，我们进行**独立样本 t 检验**，将对是否风化相关性

弱的化学成分去掉。我们使用 SPSS 进行相关性分析，结果为：

高钾										铅钡									
变量名	变量值	样本量	平均值	标准差	t	P	平均值差值	Cohen's d值		变量名	变量值	样本量	平均值	标准差	t	P	平均值差值	Cohen's d值	
二氧化硅 (SiO ₂)	2	12	67.984	8.755	-7.095	0.000***	25.979	3.547		二氧化硅 (SiO ₂)	1	23	27.056	9.005	-8.905	0.000***	27.604	2.626	
	1	6	93.963	1.734							2	23	54.66	11.829					
	总计	18	76.844	14.467							总计	46	40.858	17.4					
氧化钠 (Na ₂ O)	2	12	0.695	1.287	1.303	0.211	0.695	0.651		氧化钠 (Na ₂ O)	1	23	0.244	0.587	-2.823	0.007***	1.439	0.832	
	1	6	0	0							2	23	1.683	2.372					
	总计	18	0.463	1.089							总计	46	0.963	1.857					
氧化钾 (K ₂ O)	2	12	9.331	3.92	5.391	0.000***	8.788	2.696		氧化钾 (K ₂ O)	1	23	0.133	0.246	-1.032	0.308	0.086	0.304	
	1	6	0.543	0.445							2	23	0.219	0.31					
	总计	18	6.402	5.308							总计	46	0.176	0.28					
氧化钙 (CaO)	2	12	5.333	3.092	3.461	0.003***	4.463	1.731		氧化钙 (CaO)	1	23	2.777	1.667	3.32	0.002***	1.457	0.979	
	1	6	0.87	0.488							2	23	1.32	1.285					
	总计	18	3.845	3.308							总计	46	2.049	1.645					
氧化镁 (MgO)	2	12	1.079	0.676	3.011	0.008***	0.882	1.505		氧化镁 (MgO)	1	23	0.687	0.719	0.245	0.808	0.047	0.072	
	1	6	0.197	0.306							2	23	0.64	0.547					
	总计	18	0.785	0.712							总计	46	0.663	0.632					
氧化铝 (Al ₂ O ₃)	2	12	6.62	2.492	4.393	0.000***	4.69	2.197		氧化铝 (Al ₂ O ₃)	1	23	3.099	2.747	-1.526	0.134	1.357	0.45	
	1	6	1.93	0.964							2	23	4.456	3.262					
	总计	18	5.057	3.077							总计	46	3.778	3.06					
氧化铁 (Fe ₂ O ₃)	2	12	1.932	1.667	2.411	0.028**	1.667	1.206		氧化铁 (Fe ₂ O ₃)	1	23	0.661	0.751	-0.263	0.793	0.076	0.078	
	1	6	0.265	0.069							2	23	0.737	1.155					
	总计	18	1.376	1.566							总计	46	0.699	0.964					
氧化铜 (CuO)	2	12	2.452	1.66	1.21	0.244	0.89	0.605		氧化铜 (CuO)	1	23	2.221	2.981	1.06	0.295	0.789	0.313	
	1	6	1.562	0.935							2	23	1.432	1.97					
	总计	18	2.156	1.492							总计	46	1.827	2.53					
氧化铅 (PbO)	2	12	0.412	0.589	1.686	0.111	0.412	0.843		氧化铅 (PbO)	1	23	43.71	12.077	7.1	0.000***	21.625	2.094	
	1	6	0	0							2	23	22.085	8.215					
	总计	18	0.274	0.514							总计	46	32.897	14.961					
氧化钡 (BaO)	2	12	0.598	0.982	1.47	0.161	0.598	0.735		氧化钡 (BaO)	1	23	10.475	7.966	0.716	0.478	1.473	0.211	
	1	6	0	0							2	23	9.002	5.825					
	总计	18	0.399	0.842							总计	46	9.738	6.94					
五氧化二磷 (P ₂ O ₅)	2	12	1.402	1.434	1.879	0.079*	1.122	0.94		五氧化二磷 (P ₂ O ₅)	1	23	4.76	3.989	4.049	0.000***	3.711	1.194	
	1	6	0.28	0.21							2	23	1.049	1.847					
	总计	18	1.028	1.281							总计	46	2.905	3.601					
氧化锶 (SrO)	2	12	0.042	0.048	2.077	0.054*	0.042	1.038		氧化锶 (SrO)	1	23	0.374	0.23	1.519	0.136	0.106	0.448	
	1	6	0	0							2	23	0.268	0.243					
	总计	18	0.028	0.044							总计	46	0.321	0.24					
氧化锡 (SnO ₂)	2	12	0.197	0.681	0.696	0.496	0.197	0.348		氧化锡 (SnO ₂)	1	23	0.077	0.286	0.473	0.639	0.03	0.139	
	1	6	0	0							2	23	0.047	0.127					
	总计	18	0.131	0.556							总计	46	0.062	0.219					
二氧化硫 (SO ₂)	2	12	0.102	0.186	1.322	0.205	0.102	0.661		二氧化硫 (SO ₂)	1	23	0.197	0.661	0.182	0.857	0.038	0.054	
	1	6	0	0							2	23	0.159	0.763					
	总计	18	0.068	0.157							总计	46	0.178	0.706					

注：***、**、*分别代表1%、5%、10%的显著性水平

图 6-1-3 风化前后化学成分独立样本 t 检验结果

在高钾玻璃中，我们去掉了 t 检验后 p 值大于 0.1 的化学成分，分别是：氧化钠、氧化铜、氧化铅、氧化钡、氧化锡和二氧化硫。

在铅钡玻璃中，我们去掉了 t 检验后 p 值大于 0.2⁷的化学成分，分别是：氧化钾、氧化镁、氧化铁、氧化铜、氧化钡、氧化锡和二氧化硫。

根据环境科学杂志中《赣南小流域的水文地球化学特征和主要风化过程》⁸文章的叙述，风化过程的化学反应原理为：含**主要元素**的分子或单质，通过与空气中的氧气、水、二氧化碳以及各种催化剂的作用下化学反应变为离子或其他化合物，从而损失或增加某种化学成分的比例。既然发生了化学反应，根据物料守恒定律，主要元素的变化是线性变化的，且当风化强度不变时，反应外界条件不变，元素的变化程度也不变。

⁷ 由于铅钡玻璃有数个化学成分的比例在风化前后的变化都比较不明显，为了后续分析预测的准确性，故提高了降低了置信度。

⁸ 参考文献[6]

因此，对于风化前后的化学成分差异，我们将风化后的某一化学成分在所有玻璃文物的比例平均值减掉风化前这一成分的比例均值作为差值：

$$\overline{X_{ipo}} - \overline{X_{ipr}} = \Delta_i$$

上述十四个化学成分在高钾和铅钡中风化前和风化后的 Δ 值为：

高钾玻璃：

	二氧化硅	氧化钾	氧化钙	氧化镁	氧化铝	氧化铁	五氧化二磷	氧化锶
Δ	27.71242424%	-9.163939394%	-4.764545455%	-0.980606061%	-4.922727273%	-1.842272727%	-1.19%	-0.045454545%

表 6-1-3_1 高钾玻璃风化前后 Δ 值

铅钡玻璃：

	二氧化硅	氧化钠	氧化钙	氧化铝	氧化铅	五氧化二磷	二氧化硫
Δ	-28.79774704%	-1.463517787%	1.455019763%	-1.836541502%	22.89794466%	3.877233202%	0.11173913%

表 6-1-3_1 铅钡玻璃风化前后 Δ 值

如此，我们对附件中表面风化的文物减去 Δ 值，得到我们的预测值：

文物编号 和采样点	S102	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化砷 (As ₂ O ₃)
02	65.0777	1.46352	1.05	0.88498	1.18	7.56654	1.86	0.26	24.5321	0	-0.3072	0.07826	0	0
07	64.9176	0	9.16394	5.83455	0.98061	6.90273	2.01227	3.24	0	0	1.8	0.04545	0	0
08	48.9377	1.46352	0	0.02498	0	3.17654	0	10.41	5.78206	31.23	-0.2872	0.25826	0	2.58
08严重风化点	48.9377	1.46352	0	0.02498	0	3.17654	0	10.41	5.78206	31.23	-0.2872	0.25826	0	2.58
09	67.3076	0	9.75394	5.38455	0.98061	6.24273	2.16227	1.55	0	0	1.54	0.04545	0	0
10	69.0576	0	10.0839	4.97455	0.98061	5.73273	2.10227	0.84	0	0	1.19	0.04545	0	0
11	62.3877	1.46352	0.21	2.05498	0.71	4.52654	0	4.93	2.49206	14.61	5.50277	0.25826	0	0
12	66.5776	0	10.1739	5.48455	0.98061	6.38273	2.13227	1.65	0	0	1.34	0.04545	0	0
19	58.4377	1.46352	0	1.47498	0.59	5.40654	1.33	3.51	19.9221	5.35	4.95277	0.07826	0	0
22	64.6376	0	9.90394	6.42455	1.62061	8.42273	2.19227	0.55	0	0	1.4	0.04545	0	0
26	48.5877	1.46352	0	-0.015	0	2.53654	0	10.57	6.63206	32.25	-0.7472	0.33826	0	1.96
26严重风化点	48.5877	1.46352	0	-0.015	0	2.53654	0	10.57	6.63206	32.25	-0.7472	0.33826	0	1.96
27	65.0076	0	9.16394	5.70455	1.52061	7.43273	2.04227	1.54	0	0	1.55	0.04545	0	0
34	64.5777	1.46352	0.25	-0.675	0	3.45654	0.47	1.51	23.6521	10	-3.5372	0.10826	0	0
36	68.3677	3.68352	0.14	-1.085	0	3.43654	0.32	0.68	18.7121	10.83	-3.8072	0.10826	0	0
38	61.7277	2.84352	0	-0.775	0	4.40654	0.29	0.73	26.4121	9.79	-3.3972	0.29826	0	0
39	55.0477	1.46352	0	-0.345	0	2.33654	0	0.88	38.1321	7.22	-2.7172	0.49826	0	0
40	45.5077	1.46352	0	0.41498	0	2.28654	0.19	0	47.3121	6.69	-2.1072	0.56826	0	0
41	47.2577	1.46352	0.44	3.50498	2.73	5.16654	1.79	0.19	21.2221	9.76	3.58277	0.35826	0	0
43部位1	41.2077	1.46352	0	3.78498	0.89	4.08654	0.76	5.35	36.9521	7.29	-3.8772	0.52826	0	0
43部位2	50.4977	1.46352	0	4.94498	0.95	5.24654	1.39	1.51	21.8521	3.26	8.95277	0.35826	0	0
48	82.1277	2.26352	0.32	1.36498	1.54	15.4865	1.03	0	-7.1879	7.31	-2.7772	0.13826	1.31	0
49	57.5877	1.46352	0	3.12498	1.47	7.21654	2.74	0.7	11.2821	6.1	7.22277	0.34826	0	0
50	46.7777	1.46352	0	1.73498	0.47	3.70654	0.33	1.13	21.1021	14.2	2.46277	0.54826	0	0
51部位1	53.4077	1.46352	0	2.12498	1.19	7.08654	1.19	1.37	17.3421	8.94	4.22277	0.27826	0.47	0
51部位2	50.1477	1.46352	0	3.67498	1.45	4.34654	0.42	0.75	28.4421	0	4.87277	-0.1117	0	0
52	54.5377	2.68352	0	0.81498	0.55	2.99654	0.23	0.7	24.5221	8.64	1.83277	0.32826	0	0
54	51.0777	1.46352	0.32	1.73498	1.28	5.98654	0	0.83	32.5621	7.04	0.36277	0.76826	0	0
54严重风化点	51.0777	1.46352	0.32	1.73498	1.28	5.98654	0	0.83	32.5621	7.04	0.36277	0.76826	0	0
56	57.9477	1.46352	0	-0.245	0	3.68654	0	0.79	18.3521	15.45	-1.3372	-0.1117	0	0
57	54.2177	1.46352	0	-0.145	0	4.01654	0	1.16	22.2021	17.3	-3.8772	-0.1117	0	0
58	59.1877	1.46352	0.34	2.03498	0.79	5.35654	0.86	3.13	16.4521	7.66	5.11277	0.12826	0	0

表 6-1-3_2 预测结果

值得一提的是，表单二中存在三个严重风化的监测点，这三个点所在的文物也都检测了普通风化点的化学成分比例。因此，要预测这个三个点在风化之前的化学成分比例，理论上只需要将普通风化的化学成分减去对应类型的 Δ 值，即可

得到风化前的化学成分比例。

6.2 针对问题二：

基于 Critic 的玻璃化学成分权重模型

要找出高钾玻璃和铅钡玻璃的分类基准化学成分，可以通过对表单 2 中提供的 14 种化学成分作权重分析，排序他们的权重并据此求出高钾玻璃和铅钡玻璃的分类基准化学成分。为了得到更加可观的权重值，我们采用 **Critic 权重法**。

基于指标的对比强度和指标之间的冲突性来综合衡量指标的客观权重。对比强度是指同一个指标取值差距的大小，以标准差 S_j ⁹的形式来表现。标准差越大，说明波动越大，即各方案之间的取值差距越大，权重会越高；指标之间的冲突性 R_j ，用相关系数 r_{ij} 进行表示，若两个指标之间具有较强的正相关，说明其冲突性越小，权重会越低。

$$S_j = \sqrt{\frac{\sum X_{ij} - \bar{X}_j}{n-1}}$$
$$R_j = \sum (1 - r_{ij})$$

上式分别是对比强度和冲突性的计算公式。跟这两个量化的数值计算每个指标（化学成分）的信息量 C_j ：

$$C_j = S_j * R_j$$

以及客观权重 W_j 据此可求：

$$W_j = \frac{C_j}{\sum_{j=1}^{14} C_j}$$

6.2.1 基于权重模型求解高钾玻璃、铅钡玻璃的分类规律

利用 SPSS，我们对所有有效数据的 14 种化学成分玻进行 Critic 权重分析：

⁹ 此节中下标 j 表示第 j 种化学成分

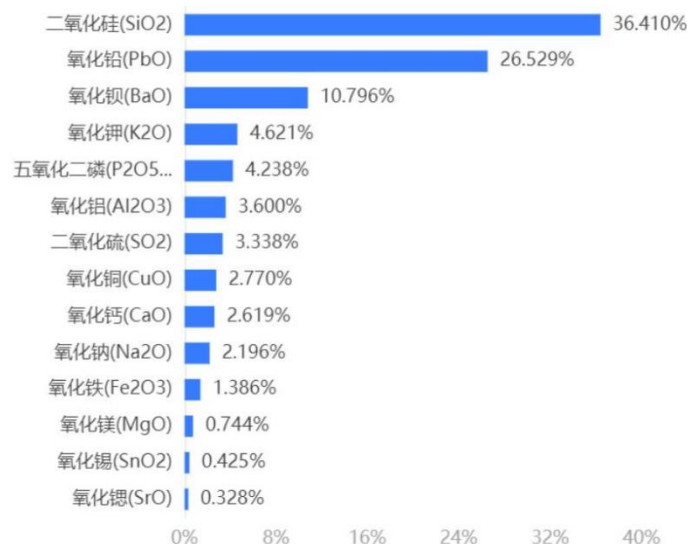


图 6-2-1 Critic 权重分析结果

由上图可知，在进行高钾和铅钡玻璃的分类时，作为基准化学成分的是：二氧化硅、氧化铅、氧化钡和氧化钾。

	高钾	铅钡
二氧化硅比例均值 (%)	75.60222222	38.8755102
氧化钾比例均值 (%)	6.195555556	0.173469388
氧化铅比例均值 (%)	0.285	33.34918367
氧化钡比例均值 (%)	0.398888889	10.49040816

表 6-2-1 二氧化硅、氧化铅、氧化钡和氧化钾的比例均值

可见，高钾玻璃的二氧化硅、氧化钾的含量比例远高于铅钡玻璃，而铅钡玻璃的氧化铅、氧化钡的含量比例远高于高钾玻璃。

6.2.2 基于 k-means 聚类的亚类划分模型以及求解

由于要分别对高钾玻璃和铅钡玻璃进行亚雷亚类划分，我们首先将高钾玻璃和铅钡玻璃分组，分别使用 Critic 权重法给 14 个化学成分赋予权重。与上文寻找分类规律同理，找到权重靠前的几个化学成分作为亚类划分的判别指标。

由于对亚类数目没有明确的获得方式，我们采用 **K-means 聚类法**，先根据肘部法则判断亚类数，进而聚类得到亚类。

肘部法则是一种利用误差平方和 SSE 和 K 值的关系图确认最优 K 值的方式。其算法思想为：数据集在 K-means 算法聚类下，随着 K 的不断增大，数据被分割的

更加详细，聚类中心不断增多，SSE 逐渐减小。当 k 值小于真实聚类数时，随着 K 值的增大，SSE 值的变化比较大，关系图显示两点之间的连线会比较陡峭。当 k 与真实聚类数相等时，随着 K 值的增大，SSE 值的变化较小，关系图显示两点之间的连线会变得平缓。所以 SSE 值和 K 值关系图是一个“手肘型”的折线图，“肘部”为最优的 K 值。

高钾玻璃的亚类：

根据 Critic 权重法，利用 SPSS，高钾玻璃的文物中化学成分的权重为：

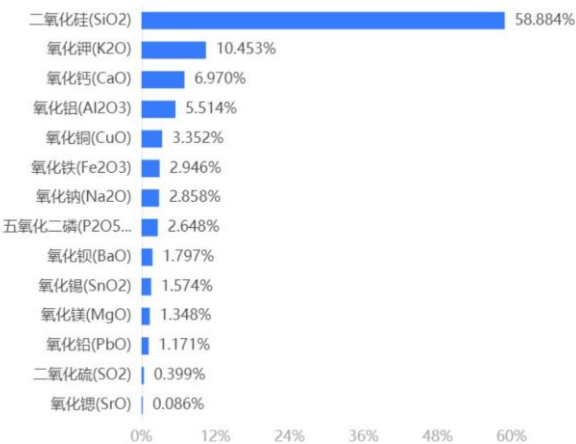


图 6-2-2 高钾玻璃中化学成分的权重

根据 85%的解释性选出权重前五的化学成分作为该亚类划分的判别成分：二氧化硅、氧化钾、氧化钙、氧化铝、氧化铜，并将这五种化学成分作为聚类的变量，进行 K-means 聚类，分别以 k=1，2，3，…，10 进行聚类，得到的 SSE 与 k 值的关系图如下：

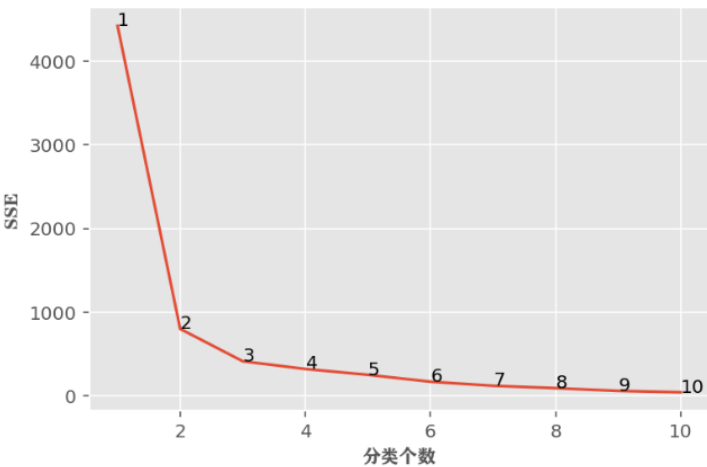


图 6-2-2_1 高钾玻璃 SSE 与 k 值

基于肘部法则，我们选择 k 值为 3 的 k-means 聚类。得到的三个类别（亚类）的中心点为：

	二氧化硅	氧化钾	氧化钙	氧化铝	氧化铜
A	81.06333333%	4.87%	2.24%	4.43333333%	1.35333333%
B	63.62444444%	10.81777778%	6.36333333%	7.34888889%	2.81888889%
C	93.96333333%	0.54333333%	0.87%	1.93%	1.56166667%

表 6-2-2_1 高钾玻璃各亚类中心点

A 类的文物编号为：3 部位 1、18、21

B 类的文物编号为：1、3 部位 2、4、5、6 部位 1、6 部位 2、13、14、16

C 类的文物编号为：7、9、10、12、22、27

高钾玻璃合理性分析和灵敏性分析：

由于高钾玻璃的总体数量小，在统计上可能存在较大的估计误差。因此对于高钾玻璃亚类划分的灵敏分析我们采用了三种方式来分析其合理性和灵敏性。

设计仿真实验：

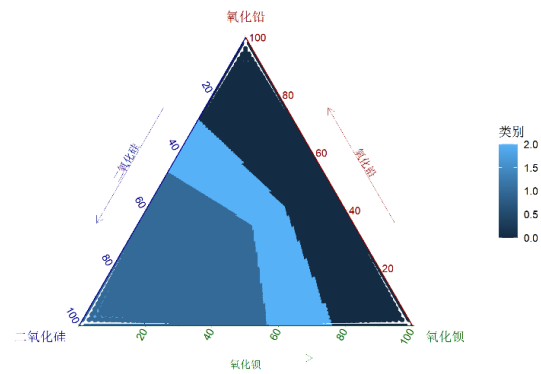
我们在权重前三的化学成分¹⁰（二氧化硅、氧化钾、氧化钙）最大值与最小值之间随机生成各 50 个数据，并随机组合，形成数量为 $50*50*50=125000$ 的合成数据来更好地可视化。

步骤为：先用原始数据进行 k-means 聚类并拟合，将生成的 125000 个数据带入到原聚类模型中，绘制三元相图。

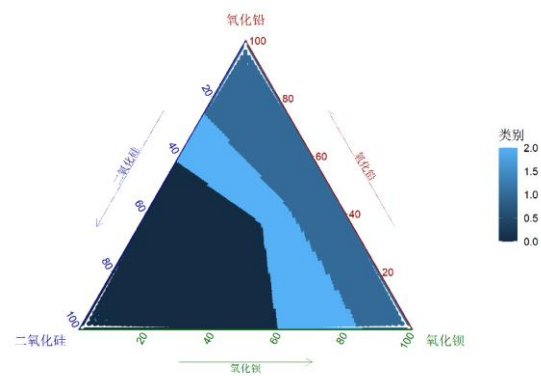
通过减少原始数据和增加干扰数据（每次改变 10%）来改变 k-means 聚类的结果，绘制每次的三元相图并观察差异来验证其合理性。

¹⁰ 便于可视化，原有的五个变量只选择前三个进行仿真，剩下的两个变量取原数据的均值。

高钾风化分类图



增加10%高钾风化敏感性分类图



剔除10%高钾风化敏感性分类图

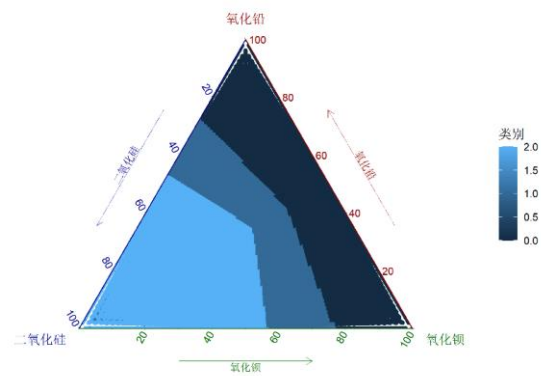


图 6-2-2_2 模型聚类数据量变化时仿真变化的三元相图

由上三图可得出，数据量的轻微改变（10%）对最终聚类结果的变化程度并不大。

类交叉检验：

首先利用高钾玻璃的原始数据根据亚类划分模型进行聚类作为正确的分类。随后随机减少 10%的数据量进行聚类得到模型，再将剩下的 10%的数据用模型进行预测其亚类，最后将这一分类结果和正确的分类结果进行比较，计算其

准确性。进行 10 次减少数据的操作，将 10 次的正确率均值作为交叉检验的正确率。

十次类交叉检验的结果我们的准确性依旧达到了 98.3%，说明我们针对问题二提出的基于 k-means 聚类的高钾玻璃亚类划分模型在原始参数（参与聚类数据量）发生轻微变化时，仍然能作出比较准确的分类效果。

铅钡玻璃的亚类：

同理，根据 Critic 权重法，利用 SPSS 对铅钡玻璃中的 14 种化学成分赋予权重值：

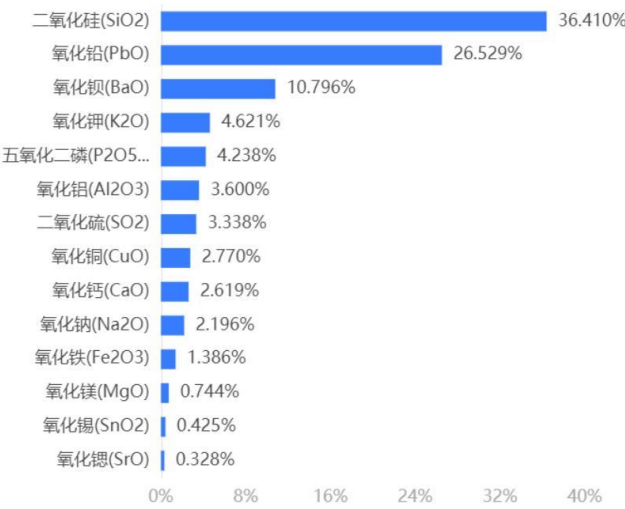


图 6-2-2_3 铅钡玻璃中化学成分的权重

根据解释性 85%选出权重前五的化学成分作为该亚类划分的判别成分：二氧化硅、氧化铅、氧化钡、氧化钾、五氧化二磷。用这五种化学成分进行 k-means 聚类，得到的 SSE 与 k 值的关系图：

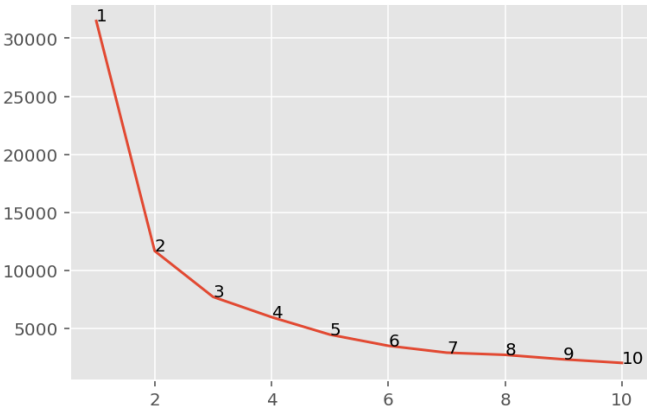


图 6-2-2_4 铅钡玻璃 SSE 与 k 值

基于肘部法则，我们选择 k 值为 3 的 k-means 聚类。得到的三个类别（亚类）的中心点为：

	二氧化硅	氧化铅	氧化钡	氧化钾	五氧化二磷
D	26.53727273%	47.33818182%	8.00727273%	0.17954545%	4.90090909%
E	21.59285714%	26.34428571%	27.70571429%	0.18857143%	5.08428571%
F	58.4965%	20.413%	7.1965%	0.1615%	0.8965%

表 6-2-2_2 铅钡玻璃各亚类中心点

D 类的文物编号为：2、19、30 部位 1、30 部位 2、34、36、38、39、40、41、43 部位 1、43 部位 2、49、50、51 部位 1、51 部位 2、52、54、54 严重风化点、56、57、58

E 类的文物编号为：8、8 严重风化点、11、20、24、26、26 严重风化点

F 类的文物编号为：23 未风化点、25 未风化点、28 未风化点、29 未风化点、31、32、33、35、37、42 未风化点 1、42 未风化点 2、44 未风化点、45、46、47、48、49 未风化点、50 未风化点、53 未风化点、55

合理性分析和灵敏性分析：

由于铅钡玻璃的数据量比高钾玻璃的数据量大，用相同原理的聚类得到的高钾玻璃亚类划分模型在已经证明其可靠性和稳定性的情况下（根据高钾玻璃亚类划分模型的仿真实验），铅钡玻璃的亚类划分模型可以直接使用与上文同理的交叉验证来分析灵敏性。最终结果为：

十次类交叉检验的结果我们的准确性依旧达到了 98.3%，说明我们针对问题二提出的基于 k-means 聚类的铅钡玻璃亚类划分模型在原始参数（参与聚类数据量）发生轻微变化时，仍然能作出比较准确的分类效果。

6.3 针对问题三：综合 Critic 权重模型与 k-means 聚类的类别划分模型

该任务实际是基于 Critic 的玻璃化学成分权重模型和基于 K-means 聚类的亚类划分模型的外延和实际运用，在本节中，我们将称之为基于 Critic 权重与 k-means 聚类的类别划分模型。由于目标要求识别表单 3 中的数据所属类型（高

钾或者铅钡), 在运用 k-means 聚类时 k 值为 2, 不需要再进行肘部法则来判断。

由于数据量较小, 采用监督学习可能会带来过拟合的影响, 即使使用交叉验证也很难排除其影响, 而由于 k-means 聚类是无监督学习的过程, 理论上是无法确定它的分类结果和表单 2 分成高钾和铅钡的结果一致的。

然而在通过基于 Critic 的玻璃化学成分权重模型分析高钾玻璃、铅钡玻璃的分类规律 (6.2.1) 中, 我们得到了这两者的分类依据化学成分。反过来说, 按照以上权重选取化学成分来进行分类会极大概率得到高钾和铅钡的分类。归根结底, 利用 Critic 权重模型得到权重次序靠前的变量来进行分类 (聚类) 可以得到相似的结果。

因此要是 k-means 聚类的结果与表单 2 中的类型一致, 需要先使用 Critic 权重模型确定需要使用的变量 (化学成分)。因此, 无监督学习 k-means 聚类在一定程度上成为了监督学习。同时也可以提供较高的正确率。

6.3.1 模型建立

根据要求要将表单 2 中的数据分成风化和无风化两组。

风化组:

利用权重模型, 我们对表面风化玻璃的化学成分权重进行排序:



图 6-3-1_1 风化玻璃的化学成分权重

选取权重前五的化学成分: 二氧化硅、氧化铅、氧化钡、五氧化二磷、二氧化硫来拟合后续的类型判别模型。将这五个化学成分作为聚类的变量, 进行 k=2 的 k-means 聚类, 将表面风化的玻璃分成两类, 作为风化组的类型判别模型。以

下是风化组类型判别模型在表单 2 中数据的结果与正确率¹¹：

聚类的中心点：

	二氧化硅	氧化铅	氧化钡	五氧化二磷	二氧化硫
高钾	23.78%	44.42%	11.99%	5.44%	1.42%
铅钡	88.16%	2.24%	1.04%	0.39%	0

表 6-3-1_1 风化组类型的聚类的中心点

分类的结果和原结果只有一个不相同，也即正确率为 $31/32 \times 100\% = 96.875\%$ 未风化组：

同理，对化学成分进行权重排序：



图 6-3-1_2 未风化玻璃的化学成分权重

选取权重前五的化学成分来拟合后续的类型判别模型。以下是风化组类型判别模型在表单 2 中数据的结果与正确率：

聚类的中心点：

	氧化铅	二氧化硅	氧化钡	氧化钾	氧化铝
高钾	7.70%	66.82%	3.01%	5.16%	6.10%
铅钡	26.42%	46.38%	11.38%	0.27%	3.67%

表 6-3-1_2 风化组类型的聚类的中心点

¹¹ 可以理解为我们将表单 2 中的数据作为模型的训练集，同时计算训练集的正确率。

分类结果与原结果相比正确率为 71.43%¹²。

6.3.2 利用模型对表单 3 数据进行分类

将表单三的数据带入到建立的模型中，判别他们的类型：

风化组，编号	类别
A2	铅钡玻璃
A5	高钾玻璃
A6	高钾玻璃
A7	高钾玻璃
无风化组，编号	类别
A1	高钾玻璃
A3	铅钡玻璃
A4	铅钡玻璃
A8	铅钡玻璃

表 6-3-2 表三分类结果

6.3.3 敏感性分析

基于表 6-3-2 中的分类结果，风化组的 A5 准确性最低仅为 67%，因此我们选择可信度最小的玻璃 A5 进行分析。我们风化元素 Critic 分析中二氧化硅、氧化铅和氧化钡前三个权重较高的元素，再，成分最大值 s%与最小值 s%之间随机生成各 20 个数据，形成数量为 20*20*20=8000 的合成数据，然后给出一个 s%与分类结果的曲线。

¹² 具体的结果在附录

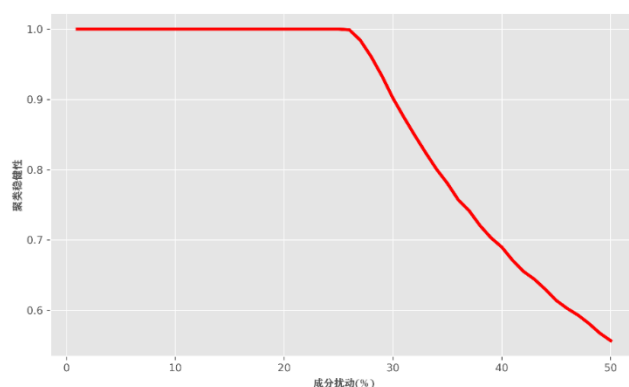


图 6-3-3 成分扰动与聚类稳健性曲线

可以看出，只有三个化学合成分同时扰动大概 25%以上才能对聚类结果产生较大的变化。模型分类结果总体是稳健的，敏感性低。

6.3.4 提升工作：使用所有化学成分进行聚类，不用 Critic 选择变量

为了使对比效果更明显，我们使用主成分分析 PCA 将所有 14 个化学成分综合并降维，将其转化成五个变量，进而进行在无风化组中进行聚类。正确率为 71.43%¹³，与我们经过 Critic 筛选变量后在聚类的结果（71.43%）基本一致，表明 Critic 筛选出的五个化学成分在聚类时基本上可以代表所有 14 个化学成分。

6.3.5 用机器学习方法验证分类结果

- ◆ 用**决策树模型**训练表单 2 中未风化数据的分类情况，再对表单 3 中的数据进行预测。此做法的原理在于验证我们利用 Critic 权重“监督”k-means 聚类的效果与真正的有监督学习效果的一致性。结果为：

无风化组，编号	原分类结果	决策树分类结果
A1	高钾玻璃	高钾玻璃
A3	铅钡玻璃	铅钡玻璃
A4	铅钡玻璃	铅钡玻璃
A8	铅钡玻璃	铅钡玻璃

图 6-3-5_1 决策树模型与原分类结果比对

¹³ 具体结果在附录

◆ 用随机森林、knn 算法、支持向量机训练模型，结果均为：

无风化组，编号	原分类结果	各训练模型分类结果
A1	高钾玻璃	高钾玻璃
A3	铅钡玻璃	铅钡玻璃
A4	铅钡玻璃	铅钡玻璃
A8	铅钡玻璃	铅钡玻璃

图 6-3-5_2 各模型与原分类结果比对

根据上述原结果和各种有监督的机器学习方法的比较，我们利用 Critic 权重对 k-means 聚类做到了较好的监督效果，做到了结果的可控性、可解释性。

6.4 针对问题四：

要分析化学成分之间的关联关系，可以采用相关系数的办法。然而数据的样本容量较少，且不能严谨说明两两化学成分之间的相关关系仅为线性关系，因此我们采用斯皮尔曼相关系数作为我们关联关系的分析基准。

斯皮尔曼系数：

$$\rho = \frac{\sum (R(X_i) - \overline{R(X)}) \cdot (R(Y_i) - \overline{R(Y)})}{d \left(\sum (R(X_i) - \overline{R(X)})^2 \right) \left(\sum (R(Y_i) - \overline{R(Y)})^2 \right)}$$

其中 $R(X)$ 和 $R(Y)$ 是 X 和 Y 的位次， $\overline{R(X)}$ 和 $\overline{R(Y)}$ 代表平均位次。

斯皮尔曼相关系数的具体算法为：

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

其中 n 是样本的数量， d 代表数据 x 和 y 之间的等级差

利用统计量：

$$r_s \sqrt{n-1} \sim N(0,1)$$

计算 p 值，先去判断两两化学元素之间是否具有相关性。

首先我们将数据分成高钾玻璃和铅钡玻璃两种，分别作出 14 种化学成分之间的相关系数矩阵¹⁴：

¹⁴ 标红表示该相关系数是有意义的，其相关性通过了显著性检验，各化学成分之间相关性的 p 值在附录

高钾总	二氧化硅 (SiO ₂)	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)
二氧化硅 (SiO ₂)	1	0.631971	0.472602	-0.54668	0.050193	0.617692	-0.265	-0.33263	-0.92846	0.230429	-0.51589	-0.46543	0.070275	0
氧化钠 (Na ₂ O)	0.631971	1	0.330025	-0.54343	-0.16714	0.234528	-0.37005	-0.11715	-0.65614	0.281129	-0.62062	-0.23755	-0.0038	0
氧化钾 (K ₂ O)	0.472602	0.330025	1	-0.19333	0.465266	0.506211	0.045961	-0.20123	-0.49169	0.160193	-0.24527	-0.01188	0.067879	0
氧化钙 (CaO)	-0.54668	-0.54343	-0.19333	1	0.174687	-0.14937	0.762913	0.071676	0.438114	-0.40316	0.520847	0.173981	0.179946	0
氧化镁 (MgO)	0.050193	-0.16714	0.465266	0.174687	1	0.531278	0.360566	-0.07652	-0.10656	-0.21085	0.28305	0.19271	0.311718	0
氧化铝 (Al ₂ O ₃)	0.617692	0.234528	0.506211	-0.14937	0.531278	1	0.18183	-0.2179	-0.70692	-0.20389	-0.12897	-0.12708	0.372621	0
氧化铁 (Fe ₂ O ₃)	-0.265	-0.37005	0.045961	0.762913	0.360566	0.18183	1	-0.15426	0.106	-0.43203	0.501479	0.193659	0.323951	0
氧化铜 (CuO)	-0.33263	-0.11715	-0.20123	0.071676	-0.07652	-0.2179	-0.15426	1	0.290279	0.142472	0.108938	-0.03518	-0.10715	0
氧化铅 (PbO)	-0.92846	-0.65614	-0.49169	0.438114	-0.10656	-0.70692	0.106	0.290279	1	-0.23928	0.458958	0.438397	-0.19285	0
氧化钡 (BaO)	0.230429	0.281129	0.160193	-0.40316	-0.21085	-0.20389	-0.43203	0.142472	-0.23928	1	-0.50553	-0.26348	-0.00327	0
五氧化二磷 (P ₂ O ₅)	-0.51589	-0.62062	-0.24527	0.520847	0.28305	-0.12897	0.501479	0.108938	0.458958	-0.50553	1	0.301829	0.095453	0
氧化锶 (SrO)	-0.46543	-0.23755	-0.01188	0.173981	0.19271	-0.12708	0.193659	-0.03518	0.438397	-0.26348	0.301829	1	-0.06565	0
氧化锡 (SnO ₂)	0.070275	-0.0038	0.067879	0.179946	0.311718	0.372621	0.323951	-0.10715	-0.19285	-0.00327	0.095453	-0.06565	1	0
二氧化硫 (SO ₂)	0	0	0	0	0	0	0	0	0	0	0	0	0	1

表 6-4_1 高钾玻璃相关系数表

铅钡总	二氧化硅 (SiO ₂)	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)
二氧化硅 (SiO ₂)	1	-0.47495	-0.78077	-0.75168	-0.54555	-0.83075	-0.78265	-0.42953	-0.44253	-0.34574	-0.52969	-0.52365	0.070122	-0.28433
氧化钠 (Na ₂ O)	-0.47495	1	0.606457	0.64048	-0.22328	0.312927	0.217731	-0.11284	0.265592	-0.23555	-0.22091	-0.10974	-0.10793	-0.19804
氧化钾 (K ₂ O)	-0.78077	0.606457	1	0.69012	0.255958	0.505689	0.457321	0.178997	0.241129	-0.00994	0.156234	0.404017	0.070267	0.307208
氧化钙 (CaO)	-0.75168	0.64048	0.69012	1	0.08729	0.521425	0.536674	0.344008	0.256022	-0.07301	0.022727	-0.05165	-0.37418	0.389374
氧化镁 (MgO)	-0.54555	-0.22328	0.255958	0.08729	1	0.734754	0.569492	0.204028	0.295313	0.521017	0.64311	0.666082	0.214272	0.428757
氧化铝 (Al ₂ O ₃)	-0.83075	0.312927	0.505689	0.521425	0.734754	1	0.747548	0.245741	0.602164	0.473264	0.518327	0.511353	-0.16362	0.246211
氧化铁 (Fe ₂ O ₃)	-0.78265	0.217731	0.457321	0.536674	0.569492	0.747548	1	0.642562	0.322964	0.538722	0.516529	0.408925	-0.37418	0.313088
氧化铜 (CuO)	-0.42953	-0.11284	0.178997	0.344008	0.204028	0.245741	0.642562	1	0.046977	0.482723	0.465909	0.140818	-0.39756	0.35441
氧化铅 (PbO)	-0.44253	0.265592	0.241129	0.256022	0.295313	0.602164	0.322964	0.046977	1	0.681739	0.120965	0.367709	-0.1861	-0.34148
氧化钡 (BaO)	-0.34574	-0.23555	-0.00994	-0.07301	0.521017	0.473264	0.538722	0.482723	0.681739	1	0.470673	0.564552	-0.12837	-0.23555
五氧化二磷 (P ₂ O ₅)	-0.52969	-0.22091	0.156234	0.022727	0.64311	0.518327	0.516529	0.465909	0.120965	0.470673	1	0.499934	0.30402	0.246339
氧化锶 (SrO)	-0.52365	-0.10974	0.404017	-0.05165	0.666082	0.511353	0.408925	0.140818	0.367709	0.564552	0.499934	1	0.306252	-0.0227
氧化锡 (SnO ₂)	0.070122	-0.10793	0.070267	-0.37418	0.214272	-0.16362	-0.37418	-0.39756	-0.1861	-0.12837	0.30402	0.306252	1	-0.10793
二氧化硫 (SO ₂)	-0.28433	-0.19804	0.307208	0.389374	0.428757	0.246211	0.313088	0.35441	-0.34148	-0.23555	0.246339	-0.0227	-0.10793	1

表 6-4_1 铅钡玻璃相关系数表

根据上两表，我们发现高钾玻璃中通过 P 检验化学成分两两相关的对数为 26，占总对数的 28. 57%，铅钡玻璃中通过 P 检验化学成分两两相关的对数为 40，占总对数的 43. 48%。通过两个相关系数矩阵，我们可以看出在两个类型中二氧化硅 SiO₂ 都与其他大部分元素都存在相关关系，且基本上为负相关，五氧化二磷 P₂O₅ 也与大部分元素都存在相关关系，在铅钡玻璃中，该元素与其他元素出二氧化硅之外基本为正相关，但是在高钾玻璃中正负相关的元素个数没有明显区

中，标绿表明该对相关系数在两类中都通过了显著性检验。

别。在铅钡玻璃中，氧化钡 BaO 与有相关性的化学元素都成正相关，而在高钾玻璃中氧化钡只与两个化学元素有相关关系。

在共同具有相关关系的化学元素中，二氧化硅与氧化钠、氧化钾、氧化铝、氧化钙，氧化钙与氧化钠，氧化铅与氧化铝，氧化钡与氧化铁的相关性正负性不同。在其他共同具有相关关系的化学元素中，相关的正负性基本上都是一致的。

根据有无风化，将两种玻璃的风化与无风化的样本分别再进行斯皮尔曼相关性检验，作出相关系数矩阵¹⁵。根据高钾玻璃有无风化分类所作的相关系数矩阵，我们发现同时在风化和未风化组中都具有显著相关性的化学成分对为：氧化钾与二氧化硅以及氧化铝与氧化镁，而且两者的正负性都相同。

根据铅钡玻璃有无风化分类的相关系数矩阵，同时都显著相关的化学成分对有七对，但是正负性都相同。

上述说明了高钾玻璃和铅钡玻璃的化学成分相关关系有明显的差异性，它们之间只有 10 对（总 91 对）化学成分的关联关系是大致一致（同时显著相关且相关方向一致）的。同时，再根据是否风化进行分类后，差异性再一次出现有无风化中，无论是高钾玻璃中还是铅钡玻璃中，同时显著相关的化学成分对只占总数的很少一部分，尽管方向都相同。

七、 模型的评价与推广

7.1 模型的优点

- ✓ 亚类划分模型分类准确且稳定，通过仿真检验和交叉检验验证；
- ✓ 化学成分权重模型客观性强，以对比强度与冲突性来建立权重赋值；
- ✓ 权重模型给基于 k-means 聚类的分类模型提供了监督，使其分类结果含义便于解释；
- ✓ 综合权重模型和分类模型的思想普适性强，可以用于各种领域的类型划分问题中。

7.2 模型的缺点

- ✓ 用于分析统计规律、训练模型的样本量太小，可能导致在进行 k-means 聚类

¹⁵ 具体的表格在附录中，矩阵为 14*14 在正文中不便展示。

等机器学习过程中出现欠拟合的现象¹⁶;

- ✓ 模型的建立过程大部分脱胎于数据的统计特征,对于现实情况(如风化现象的实际化学反应)的考虑不充分。

7.3 模型的改进

利用样本量更大的数据通过 Critic 权重模型和亚类划分模型作出训练与预测。此外,对风化现象的现实化学反应作更深入的探究也许可以使模型更贴近现实。

7.4 模型的推广

- ✓ 基于 Critic 的玻璃化学成分权重模型可以为其他领域分类问题中提供基准变量
- ✓ 于 k-means 聚类的亚类划分模型实际上可以作用到更广泛的分类问题中,机器学习方法(k-means)可以高效地完成分类任务,而通过 Critic 权重选择变量又完善了分类结果解释性与可控性。
- ✓ 验证亚类划分结果敏感性设计的仿真实验可为样本量小的模型提供敏感性检验的思路。

八、 参考文献

- [1] Wissler Clark. The Spearman Correlation Formula[J]. Science, 1905, 22(558) : 309-311.
- [2] 栾丽华,吉根林.决策树分类技术研究[J].计算机工程,2004(09):94-96+105.
- [3] Programming Fisher's Exact Method of Comparing Two Percentages[J]. Technometrics, 1960, 2(1) : 103-107.
- [4] 王千,王成,冯振元,叶金凤.K-means 聚类算法研究综述[J].电子设计工程,2012,20(07):21-24.DOI:10.14022/j.cnki.dzsjgc.2012.07.034.
- [5] 宋冬梅,刘春晓,沈晨,石学法,臧琳,冯文强.基于主客观赋权法的多目标多属性决策方法[J].山东大学学报(工学版),2015,45(04):1-9.
- [6] 孙媛媛,季宏兵,罗建美,江用彬,李甜甜.赣南小流域的水文地球化学特征和主

¹⁶ 尽管在本文中欠拟合现象不明显,甚至不存在。

要风化过程[J].环境化学,2006(05):550-557.

九、 附录

9.1Fisher's Exact Test 结果：

Fisher's Exact Test for Count Data	
data: weathering_and_ornamentation	
p-value = 0.08361	
alternative hypothesis: two.sided	

Fisher's Exact Test for Count Data	
data: yb	
p-value = 0.3806	
alternative hypothesis: two.sided	

9.2 斯皮尔曼相关系数显著性检验 p 值：

高钾	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	
x1		1	0.0007019	0.0170487	0.0046888	0.8116729	0.0010018	0.2004761	0.1042397	2.25E-11	0.2677949	0.0082973	0.0190469	0.7385315	1
x2	0.0007019		1	0.1071458	0.0049927	0.4245454	0.259138	0.068639	0.5770485	0.0003687	0.1734113	0.0009327	0.2528739	0.9856034	1
x3	0.0170487	0.1071458		1	0.3544665	0.0190961	0.0098247	0.8273071	0.3347785	0.0125487	0.444321	0.2373235	0.9550704	0.7471549	1
x4	0.0046888	0.0049927	0.3544665		1	0.4036238	0.476063	9.23E-06	0.7335011	0.028487	0.0456821	0.0075955	0.405555	0.3893891	1
x5	0.8116729	0.4245454	0.0190961	0.4036238		1	0.0062792	0.0766203	0.7161897	0.6121607	0.3116705	0.1703679	0.3560522	0.1292962	1
x6	0.0010018	0.259138	0.0098247	0.476063	0.0062792		1	0.3843598	0.2953899	7.79E-05	0.3282942	0.5389346	0.544967	0.0665913	1
x7	0.2004761	0.068639	0.8273071	9.23E-06	0.0766203	0.3843598		1	0.4615991	0.6140579	0.0310295	0.0106523	0.3536434	0.1141592	1
x8	0.1042397	0.5770485	0.3347785	0.7335011	0.7161897	0.2953899	0.4615991		1	0.1592457	0.4969041	0.6042141	0.8674019	0.6101969	1
x9	2.25E-11	0.0003687	0.0125487	0.028487	0.6121607	7.79E-05	0.6140579	0.1592457		1	0.2493359	0.0210133	0.0283727	0.3557034	1
x10	0.2677949	0.1734113	0.444321	0.0456821	0.3116705	0.3282942	0.0310295	0.4969041	0.2493359		1	0.0099401	0.2031744	0.9876259	1
x11	0.0082973	0.0009327	0.2373235	0.0075955	0.1703679	0.5389346	0.0106523	0.6042141	0.0210133	0.0099401		1	0.1425573	0.6499241	1
x12	0.0190469	0.2528739	0.9550704	0.405555	0.3560522	0.544967	0.3536434	0.8674019	0.0283727	0.2031744	0.1425573		1	0.7552053	1
x13	0.7385315	0.9856034	0.7471549	0.3893891	0.1292962	0.0665913	0.1141592	0.6101969	0.3557034	0.9876259	0.6499241	0.7552053		1	1
x14		1	1	1	1	1	1	1	1	1	1	1	1	1	1

高钾风化	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	
x1		1	0.1715281	0.0306759	0.0190195	0.8202845	0.1445863	0.0085393	0.1942017	0.9287016	0.8371459	0.7787254	0.3983556	0.2062558	0.8427454
x2	0.1715281		1	0.0192937	0.0063963	0.0401302	0.7118494	0.8202724	0.6275773	0.6580352	0.2060691	0.1929061	0.2537333	0.5930477	0.3021801
x3	0.0306759	0.0192937		1	0.000915	0.211643	0.6966831	0.9913804	0.5120853	0.6044192	0.2904551	0.174519	0.8718765	0.6848487	0.7548697
x4	0.0190195	0.0063963	0.000915		1	0.1612413	0.7700196	0.5816858	0.4899232	0.6037767	0.1399562	0.1335069	0.1929972	0.1549548	0.419166
x5	0.8202845	0.0401302	0.211643	0.1612413		1	0.0619578	0.1447716	0.9396045	0.7275539	0.2240272	0.0660406	0.0287276	0.6843214	0.3076746
x6	0.1445863	0.7118494	0.6966831	0.7700196	0.0619578		1	0.0441448	0.8882523	0.3006092	0.3584054	0.6330907	0.3149978	0.1139374	0.9548427
x7	0.0085393	0.8202724	0.9913804	0.5816858	0.1447716	0.0441448		1	0.0323849	0.9910622	0.1592344	0.0599836	0.2877047	0.1549548	0.7113635
x8	0.1942017	0.6275773	0.5120853	0.4899232	0.9396045	0.8882523	0.0323849		1	0.8576349	0.1279817	0.533867	0.9769998	0.1132134	0.4453532
x9	0.9287016	0.6580352	0.6044192	0.6037767	0.7275539	0.3006092	0.9910622	0.8576349		1	0.0358497	0.3604571	0.6746786	0.3154242	0.0393202
x10	0.8371459	0.2060691	0.2904551	0.1399562	0.2240272	0.3584054	0.1592344	0.1279817	0.0358497		1	0.3729219	0.0981507	0.5168224	0.2060691
x11	0.7787254	0.1929061	0.174519	0.1335069	0.0660406	0.6330907	0.0599836	0.533867	0.3604571	0.3729219		1	0.1297523	0.3338931	0.9548427
x12	0.3983556	0.2537333	0.8718765	0.1929972	0.0287276	0.3149978	0.2877047	0.9769998	0.6746786	0.0981507	0.1297523		1	0.4659658	0.4911258
x13	0.2062558	0.5930477	0.6848487	0.1549548	0.6843214	0.1139374	0.1549548	0.1132134	0.3154242	0.5168224	0.3338931	0.4659658		1	0.5930477
x14	0.8427454	0.3021801	0.7548697	0.419166	0.3076746	0.9548427	0.7113635	0.4453532	0.0393202	0.2060691	0.9548427	0.4911258	0.5930477		1

高钾未风化															
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	
x1	1		0.4246625		0.1403574	0.0048047	0.9571545	0.9571545		1	0.2657026		1	1	1
x2		1		1		1		1		1		1		1	1
x3	0.4246625		1	0.4246625	0.6711187	0.4246625	0.2883432	0.6583736		1	0.0076658		1	1	1
x4		0	0.4246625		0.1403574	0.0048047	0.9571545	0.9571545		1	0.2657026		1	1	1
x5	0.1403574	0.6711187	0.4246625	0.6711187		1	0.0341094	0.5122253	0.1403574		1	0.7983993		1	1
x6	0.0048047		0.4246625	0.0048047	0.0341094		1	0.8717434	0.704		1	0.328723		1	1
x7	0.9571545		0.2883432	0.9571545	0.5122253	0.8717434		1	0.2657026		1	0.328723		1	1
x8	0.9571545		0.6583736	0.9571545	0.1403574	0.704	0.2657026		1	1	0.328723		1	1	1
x9		1		1		1		1		1		1		1	1
x10			1		1		1		1		1		1	1	1
x11	0.2657026		0.0076658	0.2657026	0.7983993	0.328723	0.328723	0.328723		1	1	1	1	1	1
x12		1		1		1		1		1		1		1	1
x13			1		1		1		1		1		1		1
x14				1		1		1		1		1		1	1

铅钨	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	
x1		1	0.0463962	0.0001313	0.0003219	0.0191924	1.96E-05	0.0001233	0.0752517	0.065918	0.1599271	0.0237724	0.0257252	0.7821751	0.2528206
x2	0.0463962		1	0.0076243	0.0041896	0.3731458	0.2060972	0.3854379	0.6557456	0.2867871	0.3467373	0.3783653	0.6646758	0.6698911	0.4308378
x3	0.0001313	0.0076243		1	0.0015256	0.3052756	0.0322763	0.0563713	0.4772884	0.3350938	0.9687772	0.5358643	0.09635	0.7817344	0.2149535
x4	0.0003219	0.0041896	0.0015256		1	0.7305355	0.0264738	0.0216596	0.1621663	0.3051503	0.7734192	0.9286749	0.8387051	0.1260823	0.1102306
x5	0.0191924	0.3731458	0.3052756	0.7305355		1	0.0005144	0.013625	0.4167509	0.2341549	0.0266131	0.0039886	0.002545	0.3932174	0.0758365
x6	1.96E-05	0.2060972	0.0322763	0.0264738	0.0005144		1	0.0003621	0.3256448	0.0081847	0.0472877	0.0275463	0.0300845	0.5165101	0.3246896
x7	0.0001233	0.3854379	0.0563713	0.0216596	0.013625	0.0003621		1	0.0040298	0.1911397	0.0210692	0.0281843	0.0919937	0.1260823	0.2058505
x8	0.0752517	0.6557456	0.4772884	0.1621663	0.4167509	0.3256448	0.0040298		1	0.8531524	0.0424505	0.0513286	0.5772965	0.1023013	0.1490148
x9	0.065918	0.2867871	0.3350938	0.3051503	0.2341549	0.0081847	0.1911397	0.8531524		1	0.0018332	0.6325591	0.1332945	0.4596752	0.1654849
x10	0.1599271	0.3467373	0.9687772	0.7734192	0.0266131	0.0472877	0.0210692	0.0424505	0.0018332		1	0.0486828	0.0146523	0.6116982	0.3467373
x11	0.0237724	0.3783653	0.5358643	0.9286749	0.0039886	0.0275463	0.0281843	0.0513286	0.6325591	0.0486828		1	0.0346274	0.2199951	0.3244318
x12	0.0257252	0.6646758	0.09635	0.8387051	0.002545	0.0300845	0.0919937	0.5772965	0.1332945	0.0146523	0.0346274		1	0.2164569	0.9287466
x13	0.7821751	0.6698911	0.7817344	0.1260823	0.3932174	0.5165101	0.1260823	0.1023013	0.4596752	0.6116982	0.2199951	0.2164569		1	0.6698911
x14	0.2528206	0.4308378	0.2149535	0.1102306	0.0758365	0.3246896	0.2058505	0.1490148	0.1654849	0.3467373	0.3244318	0.9287466	0.6698911		1

铅钨风化	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
x1	1	0.7252687	0.8619272	0.575529	0.2792838	0.7469478	0.8695148	0.0654299	0.0023504	0.0007611	0.8477631	0.0609632	0.3030977	0.7706926
x2	0.7252687	1	0.8240291	0.598862	0.5428423	0.797072	0.0464815	0.1155688	0.4285658	0.2004515	0.0004134	0.7913243	0.1347651	0.4179201
x3	0.8619272	0.8240291	1	0.9926937	0.044113	0.001668	0.5552599	0.9170906	0.2754539	0.9053181	0.7402602	0.8461533	0.5031277	0.6545618
x4	0.575529	0.598862	0.9926937	1	0.6630169	0.1908951	0.1025937	0.0225065	0.3754243	0.0746471	0.3724196	0.1793839	0.016138	0.7705268
x5	0.2792838	0.5428423	0.044113	0.6630169	1	0.0073992	0.5388267	0.7243556	0.440504	0.3530824	0.4385383	0.401428	0.7293571	0.2572827
x6	0.7469478	0.797072	0.001668	0.1908951	0.0073992	1	0.3952728	0.5054018	0.0489278	0.5201463	0.9262659	0.9765754	0.6764169	0.5587992
x7	0.8695148	0.0464815	0.5552599	0.1025937	0.5388267	0.3952728	1	0.0120868	0.8259946	0.1081089	0.0866177	0.8148705	0.029405	0.3832531
x8	0.0654299	0.1155688	0.9170906	0.0225065	0.7243556	0.5054018	0.0120868	1	0.61349	0.0063715	0.9612879	0.7147705	0.0066175	0.1805354
x9	0.0023504	0.4285658	0.2754539	0.3754243	0.440504	0.0489278	0.8259946	0.61349	1	0.4381548	0.613652	0.1281301	0.0811378	0.7707478
x10	0.0007611	0.2004515	0.9053181	0.0746471	0.3530824	0.5201463	0.1081089	0.0063715	0.4381548	1	0.1418756	0.1122455	0.9208578	0.5587992
x11	0.8477631	0.0004134	0.7402602	0.3724196	0.4385383	0.9262659	0.0866177	0.9612879	0.613652	0.1418756	1	0.5449477	0.8184974	0.2976766
x12	0.0609632	0.7913243	0.8461533	0.1793839	0.401428	0.9765754	0.8148705	0.7147705	0.1281301	0.1122455	0.5449477	1	0.2920327	0.5550296
x13	0.3030977	0.1347651	0.5031277	0.016138	0.7293571	0.6764169	0.029405	0.0066175	0.0811378	0.9208578	0.8184974	0.2920327	1	0.7087949
x14	0.7706926	0.4179201	0.6545618	0.7705268	0.2572827	0.5587992	0.3832531	0.1805354	0.7707478	0.5587992	0.2976766	0.5550296	0.7087949	1

铅钨无风化	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
x1	1	0.0248387	0.0241829	0.0506682	0.8898941	0.1017959	0.5142074	0.3086354	0.1482534	0.9286653	0.6536619	0.0034925	0.308022	0.1626932
x2	0.0248387	1	0.7802445	0.0425517	0.5464872	0.7655203	0.8907456	0.0288985	0.8230895	0.5819028	0.0940379	0.6461501	0.2992023	0.5221007
x3	0.0241829	0.7802445	1	0.8136317	0.0588478	0.0327825	0.2037194	0.1985413	0.5204148	0.6729717	0.930774	0.56744	0.6194814	0.3161586
x4	0.0506682	0.0425517	0.8136317	1	2.59E-06	0.0075987	0.0082015	0.5740797	0.7829199	0.0190144	0.0004417	0.2001921	0.5439803	0.3404705
x5	0.8898941	0.5464872	0.0588478	2.59E-06	1	5.65E-06	0.0005165	0.2106669	0.594451	0.0045921	0.0088307	0.5680961	0.0677071	0.1176591
x6	0.1017959	0.7655203	0.0327825	0.0075987	5.65E-06	1	0.0005032	0.4964999	0.2029164	0.0288903	0.0408487	0.4974044	0.0335573	0.0824373
x7	0.5142074	0.8907456	0.2037194	0.0082015	0.0005165	0.0005032	1	0.2212445	0.719551	0.0031772	0.0725458	0.8395745	0.2342205	0.0761269
x8	0.3086354	0.0288985	0.1985413	0.5740797	0.2106669	0.4964999	0.2212445	1	0.2207746	0.1009198	0.4935121	0.8858038	0.390156	0.0183375
x9	0.1482534	0.8230895	0.5204148	0.7829199	0.594451	0.2029164	0.719551	0.2207746	1	0.0244878	0.1163653	0.1521986	0.0754481	0.0613673
x10	0.9286653	0.5819028	0.6729717	0.0190144	0.0045921	0.0288903	0.0031772	0.1009198	0.0244878	1	0.1084906	0.5438284	0.9052705	0.0184038
x11	0.6536619	0.0940379	0.930774	0.0004417	0.0088307	0.0408487	0.0725458	0.4935121	0.1163653	0.1084906	1	0.690881	0.8763147	0.8403388
x12	0.0034925	0.6461501	0.56744	0.2001921	0.5439803	0.4974044	0.8395745	0.8858038	0.1521986	0.5438284	0.690881	1	0.7410136	0.8905874
x13	0.308022	0.2992023	0.6194814	0.5439803	0.0677071	0.0335573	0.2342205	0.390156	0.0754481	0.9052705	0.8763147	0.7410136	1	0.6662049
x14	0.1626932	0.5221007	0.3161586	0.3404705	0.1176591	0.0824373	0.0761269	0.0183375	0.0613673	0.0184038	0.8403388	0.8905874	0.6662049	1

9.3 斯皮尔曼相关系数矩阵

高钾风化	二氧化硅 (SiO ₂)	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)
二氧化硅 (SiO ₂)	1	-0.42223	-0.62238	-0.662	-0.07356	-0.44755	-0.71804	-0.4028	-0.029	-0.06657	-0.09091	-0.26873	0.393042	-0.06425
氧化钠 (Na ₂ O)	-0.42223	1	0.660884	0.735603	-0.59768	0.119326	-0.07356	-0.15632	0.142767	-0.39319	-0.40387	-0.35763	-0.17197	-0.3253
氧化钾 (K ₂ O)	-0.62238	0.660884	1	0.826621	-0.38879	-0.12587	0.003503	0.210158	-0.16678	-0.33284	-0.41958	-0.05225	-0.13101	0.100968
氧化钙 (CaO)	-0.662	0.735603	0.826621	1	-0.43158	0.094571	0.177193	0.221053	-0.16707	-0.4522	-0.45884	-0.4038	-0.43748	0.257461
氧化镁 (MgO)	-0.07356	-0.59768	-0.38879	-0.43158	1	0.553416	0.447368	0.024561	-0.11259	0.379265	0.546411	0.628131	0.131244	0.321826
氧化铝 (Al ₂ O ₃)	-0.44755	0.119326	-0.12587	0.094571	0.553416	1	0.588442	-0.04553	0.326301	0.291232	0.153846	0.317248	-0.48038	-0.01836
氧化铁 (Fe ₂ O ₃)	-0.71804	-0.07356	0.003503	0.177193	0.447368	0.588442	1	0.617544	-0.00363	0.433446	0.556919	0.334629	-0.43748	0.119535
氧化铜 (CuO)	-0.4028	-0.15632	0.210158	0.221053	0.024561	-0.04553	0.617544	1	-0.05811	0.464704	0.19965	-0.00935	-0.48123	0.243669
氧化铅 (PbO)	-0.029	0.142767	-0.16678	-0.16707	-0.11259	0.326301	-0.00363	-0.05811	1	0.608279	-0.29005	0.135454	-0.31698	-0.59962
氧化钡 (BaO)	-0.06657	-0.39319	-0.33284	-0.4522	0.379265	0.291232	0.433446	0.464704	0.608279	1	0.282911	0.49962	-0.20786	-0.39319
五氧化二磷 (P ₂ O ₅)	-0.09091	-0.40387	-0.41958	-0.45884	0.546411	0.153846	0.556919	0.19965	-0.29005	0.282911	1	0.462809	0.305699	-0.01836
氧化锶 (SrO)	-0.26873	-0.35763	-0.05225	-0.4038	0.628131	0.317248	0.334629	-0.00935	0.135454	0.49962	0.462809	1	0.233084	-0.22046
氧化锡 (SnO ₂)	0.393042	-0.17197	-0.13101	-0.43748	0.131244	-0.48038	-0.43748	-0.48123	-0.31698	-0.20786	0.305699	0.233084	1	-0.17197
二氧化硫 (SO ₂)	-0.06425	-0.3253	0.100968	0.257461	0.321826	-0.01836	0.119535	0.243669	-0.59962	-0.39319	-0.01836	-0.22046	-0.17197	1

高钾末风	二氧化硅 (SiO ₂)	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)
二氧化硅 (SiO ₂)	1	0	0.40584	-1	-0.67612	-0.94286	-0.02857	0.028571	0	0	-0.54286	0	0	0
氧化钠 (Na ₂ O)	0	1	0	0	0	0	0	0	0	0	0	0	0	0
氧化钾 (K ₂ O)	0.40584	0	1	-0.40584	-0.22295	-0.40584	0.521794	-0.23191	0	0	-0.92763	0	0	0
氧化钙 (CaO)	-1	0	-0.40584	1	0.676123	0.942857	0.028571	-0.02857	0	0	0.542857	0	0	0
氧化镁 (MgO)	-0.67612	0	-0.22295	0.676123	1	0.845154	0.338062	-0.67612	0	0	0.135225	0	0	0
氧化铝 (Al ₂ O ₃)	-0.94286	0	-0.40584	0.942857	0.845154	1	0.085714	-0.2	0	0	0.485714	0	0	0
氧化铁 (Fe ₂ O ₃)	-0.02857	0	0.521794	0.028571	0.338062	0.085714	1	-0.54286	0	0	-0.48571	0	0	0
氧化铜 (CuO)	0.028571	0	-0.23191	-0.02857	-0.67612	-0.2	-0.54286	1	0	0	0.485714	0	0	0
氧化铅 (PbO)	0	0	0	0	0	0	0	0	1	0	0	0	0	0
氧化钡 (BaO)	0	0	0	0	0	0	0	0	0	1	0	0	0	0
五氧化二磷 (P ₂ O ₅)	-0.54286	0	-0.92763	0.542857	0.135225	0.485714	-0.48571	0.485714	0	0	1	0	0	0
氧化锶 (SrO)	0	0	0	0	0	0	0	0	0	0	0	1	0	0
氧化锡 (SnO ₂)	0	0	0	0	0	0	0	0	0	0	0	0	1	0
二氧化硫 (SO ₂)	0	0	0	0	0	0	0	0	0	0	0	0	0	1

铅钡风化	二氧化硅 (SiO ₂)	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)
二氧化硅 (SiO ₂)	1	-0.07749	0.038393	-0.12318	0.23554	0.071164	-0.03626	-0.39051	-0.60242	-0.65135	-0.04237	-0.39662	-0.22449	0.064298
氧化钠 (Na ₂ O)	-0.07749	1	-0.04908	-0.11577	0.133773	0.056738	-0.41918	0.337238	-0.17349	0.277143	-0.67471	0.05838	-0.32143	-0.17745
氧化钾 (K ₂ O)	0.038393	-0.04908	1	-0.00202	0.423375	0.618175	0.129714	-0.02299	-0.23737	-0.02626	-0.07311	-0.04283	0.147055	0.098565
氧化钙 (CaO)	-0.12318	-0.11577	-0.00202	1	0.096004	0.282888	0.349042	-0.47341	0.193868	-0.37883	0.195067	0.290075	0.495773	0.064346
氧化镁 (MgO)	0.23554	0.133773	0.423375	0.096004	1	0.543148	-0.13509	-0.07776	-0.1691	-0.20292	-0.16982	0.183707	0.076291	-0.24629
氧化铝 (Al ₂ O ₃)	0.071164	0.056738	0.618175	0.282888	0.543148	1	0.186078	-0.14628	-0.41502	-0.1413	-0.02043	0.006484	0.091969	-0.12856
氧化铁 (Fe ₂ O ₃)	-0.03626	-0.41918	0.129714	0.349042	-0.13509	0.186078	1	-0.51411	0.048519	-0.3439	0.365197	-0.05167	0.454356	-0.19077
氧化铜 (CuO)	-0.39051	0.337238	-0.02299	-0.47341	-0.07776	-0.14628	-0.51411	1	-0.11119	0.55152	-0.01072	0.080572	-0.54942	0.289342
氧化铅 (PbO)	-0.60242	-0.17349	-0.23737	0.193868	-0.1691	-0.41502	0.048519	-0.11119	1	0.16996	0.111143	0.326698	0.371251	-0.06428
氧化钡 (BaO)	-0.65135	0.277143	-0.02626	-0.37883	-0.20292	-0.1413	-0.3439	0.55152	0.16996	1	-0.31599	0.340165	0.021938	0.128565
五氧化二磷 (P ₂ O ₅)	-0.04237	-0.67471	-0.07311	0.195067	-0.16982	-0.02043	0.365197	-0.01072	0.111143	-0.31599	1	-0.13308	0.050643	0.226959
氧化锶 (SrO)	-0.39662	0.05838	-0.04283	0.290075	0.183707	0.006484	-0.05167	0.080572	0.326698	0.340165	-0.13308	1	0.229556	0.129789
氧化锡 (SnO ₂)	-0.22449	-0.32143	0.147055	0.495773	0.076291	0.091969	0.454356	-0.54942	0.371251	0.021938	0.050643	0.229556	1	-0.08233
二氧化硫 (SO ₂)	0.064298	-0.17745	0.098565	0.064346	-0.24629	-0.12856	-0.19077	0.289342	-0.06428	0.128565	0.226959	0.129789	-0.08233	1


```

#比较y_pred和原先分类结果高钾玻璃和铅钡玻璃的准确性
#聚类1可以看作铅钡玻璃的划分依据，聚类2可以看作高钾玻璃的划分依据
boli_ceshi = pd.read_csv('C_Q1_2.csv', encoding='UTF-8')
boli_leixing = boli_ceshi.iloc[:, 0]
boli_leixing = boli_leixing-1
#boli_leixing 中0是高钾玻璃，1是铅钡玻璃
#所以判断的依据改成
correct = np.zeros(len(boli_leixing))
for i in range(len(boli_leixing)):
    if(boli_leixing[i] != y_pred[i]):
        correct[i] = 1
print(correct)
zhengquelv = np.sum(correct)/len(correct)*100
print("正确率为")
print("%f%" % zhengquelv)

```

```

[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 1. 1. 0. 0. 1. 1. 1. 0. 1. 0. 1. 0. 0. 0. 0. 1. 1. 0. 0.
 1. 1. 1. 1. 1. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]

```

正确率为

```
%f% 25.37313432835821
```

9.5 代码（python 和 R）

从下页开始