

玻璃文物的类别鉴定与亚类划分

徐美萍*, 王行修*, 江萌萌, 乔斯昱, 由云川

北京工商大学数学与统计学院, 北京

收稿日期: 2023年6月14日; 录用日期: 2023年7月9日; 发布日期: 2023年7月19日

摘要

基于已知玻璃类型的数据, 研究玻璃文物表面风化的相关因素和由风化引起的主要化学成分的变化, 由此基于主要化学成分建立玻璃的分类模型和亚类划分模型, 并利用所得模型对未知类型的玻璃文物进行鉴别。研究发现: 玻璃表面风化与其类型相关, 与纹饰弱相关, 而与颜色基本无关; 与高钾玻璃相比, 铅钡玻璃表面更易受风化影响; 表面风化会引起高钾和铅钡玻璃一些成分比例的显著变化; 精选4个重要成分PbO、BaO、SiO₂、K₂O, 建立玻璃类型的随机森林(RF)和支持向量机(SVM)模型, 分类准确率均为100%, 使用这些模型对8个未知类型的玻璃文物进行鉴别, 结果一致, 均为: A1 A6 A7为高钾, 其余为铅钡; 对每个类型的玻璃各挑选8个重要成分, 使用k-means聚类算法对每个类型玻璃可以很好地细分为两个亚类; 最后, 通过对成分数据增加微小随机扰动来证实所建模型是稳健的。

关键词

随机森林, 支持向量机, K-Means聚类算法, 成分数据

Classification Identification and Sub-Classification of Glass Relics

Meiping Xu*, Xingxiu Wang*, Mengmeng Jang, Siyu Qiao, Yunchuan You

School of Mathematics and Statistics, Beijing Technology and Business University, Beijing

Received: Jun. 14th, 2023; accepted: Jul. 9th, 2023; published: Jul. 19th, 2023

Abstract

Based on glass relics data with known glass types, factors and important chemical compositions related to the weathering of glass surface are studied. And classification models of glass types and sub-classification models are established based on their important chemical compositions. Then

*第一作者。

these classification models are applied to identify glass relics with unknown types. Results show that the weathering of glass surface is related to its type, weakly related to its decoration, and unrelated to its color. The surface of PbO-BaO glass is more susceptible to weathering than that of high K₂O glass. Surface weathering can cause significant changes in the proportion of some compositions of high K₂O and PbO-BaO glass. Four important compositions, PbO, BaO, SiO₂ and K₂O, are selected to make models of glass types via random forest (RF) and support vector machine (SVM), with classification accuracy of 100%. These models are employed to identify 8 unknown types of glass relics, same results are presented that A1, A6 and A7 are high K₂O glass, and the rest are PbO-BaO glass. Then 8 important compositions are selected for each glass type, both of which are divided into two subclasses via k-means clustering very well. Finally, the robustness of the proposed models is verified by adding a small random disturbance to the composition data.

Keywords

Random Forest, Support Vector Machine, K-Means Clustering Algorithm, Composition Data

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

玻璃制品最初由丝绸之路传入中国, 主要类别有高钾、钠钙玻璃等。高钾玻璃是以含钾量高的物质如草木灰作为助熔剂烧制而成的, 主要流行于我国岭南以及东南亚和印度等区域。在中国本土发明制作的玻璃中, 常用铅矿石作助熔剂, 出土的玻璃文物往往铅钡含量较高。留存至今的古代玻璃制品往往长时间埋藏于墓穴等环境中, 极易受环境的影响而风化。在风化过程中, 内部元素与环境元素进行大量交换, 导致其化学成分所占比例发生变化, 从而影响对其类别的正确判断。另外, 由于被埋藏环境和时间等条件的不同, 不同类别玻璃文物风化的程度也有所差异, 也给分类问题增加较大辨别难度。

关于风化, 周良知[1]应用近代研究方法, 研究影响硅酸盐玻璃风化的主要因素。王承遇等[2]对不同类型玻璃风化作用的机理进行了深入研究。王婕等[3]利用分析测试方法对战国时期八棱柱状铅钡玻璃器的风化产物、风化形貌和元素迁移进行了分析和探讨。这些工作对后续研究风化玻璃的主要成分提供了依据和借鉴。

关于分类, 付强等[4]通过探讨风化对化学成分定量分析的影响, 对所占比例较高的硅钾酸盐玻璃进行了亚类划分。赵凤燕等[5]通过 pXRF 方法对玻璃器的化学成分进行无损分析来进行分类。但他们仅基于成分数据做了一些简单的分析就依此给出类型识别, 没有借助统计模型做更深入的探讨和分析, 提供一些可以按照玻璃成分进行较为精准分类的泛化模型。近期, 王祉皓等[6]基于机器学习的方法对风化硅酸盐原成分进行了预测及亚类划分, 但其主要使用的神经网络算法对文中所采用的小样本数据很容易过拟合, 容易导致模型的泛化能力和稳健性变差。

针对本文数据特点, 我们借鉴何秋妍等[7]使用的有监督模型, 选用随机森林(RF)和支持向量机(SVM)这两个既适用于小样本又有较好分类能力的模型对玻璃文物按高钾、铅钡两大类进行分类规律的讨论, 使用 k-means 这一最简单且收敛速度较快的硬聚类算法对高钾/铅钡玻璃进行亚类划分, 并对所建模型做稳健性分析以说明模型的可适用性, 最后探讨表面风化对分类的影响。研究结果表明我们的模型在保证分类准确类的前提下, 更简洁稳定, 且具有很好的泛化能力。

2. 思路流程

如图 1 所示, 本文首先对于所获取成分数据使用 CLR (对数中心比) 变换, 然后基于转换后的数据进行独立性检验与单因素方差分析, 选取最显著的 8 个变量, 并使用随机森林中算法进行分类, 同时使用具有线性和径向基核函数的 SVM 算法进行回判分析, 佐证稳健性, 最终得到高钾玻璃与铅钡两类玻璃的预测模型。然后使用 K-means 与多维尺度分析(cmd)对处理后的数据进行降维处理与亚类划分, 将分类结果分别命名为铅钡(高硅低铅)类玻璃、铅钡(高铅低硅)类玻璃、高钾(高硅)类玻璃与高钾(低硅)类玻璃。

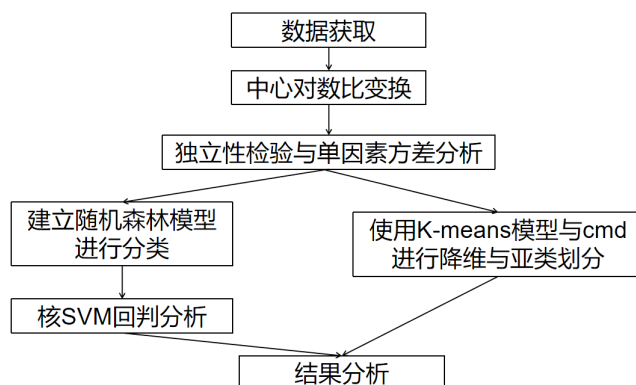


Figure 1. Research idea map

图 1. 研究思路图

3. 模型建立

3.1. 数据来源与预处理

通过查阅资料[8], 获得玻璃类型、纹饰、颜色与 14 个主要成分比例的玻璃文物相关数据。具体为: 表单 1 中列出的玻璃文物已经知道其玻璃类型、颜色、纹饰和表面风化与否, 表单 2 是表单 1 中文物在一些采样点的成分数据, 表单 3 给出的是未分类玻璃文物的成分数据。表单 2 和 3 的数据具有成分性, 但考虑到受限于当前的检测手段, 可能存在检测得到成分比例和不等 100% 的情况, 项目方认为成分比例累加和位于 85% 至 105% 之间的数据有效, 其余数据视作无效。

对表单 2 中编号 18 和 20 的两行无效数据进行删除, 共获得 67 行有效观测数据; 对表单 2 和 3 的未检测到的成分数据填补一个很小的正数 $1e-3$, 方便后续做几何平均; 对由于采集误差等原因造成的 14 个主要成分数据占比和不为 100% 的问题进行归一化处理。由于这些成分数据取值范围是 14 维空间中的一个单纯形, 我们通过式(1)的中心对数比变换(CLR)将其映射到 14 维实数空间。

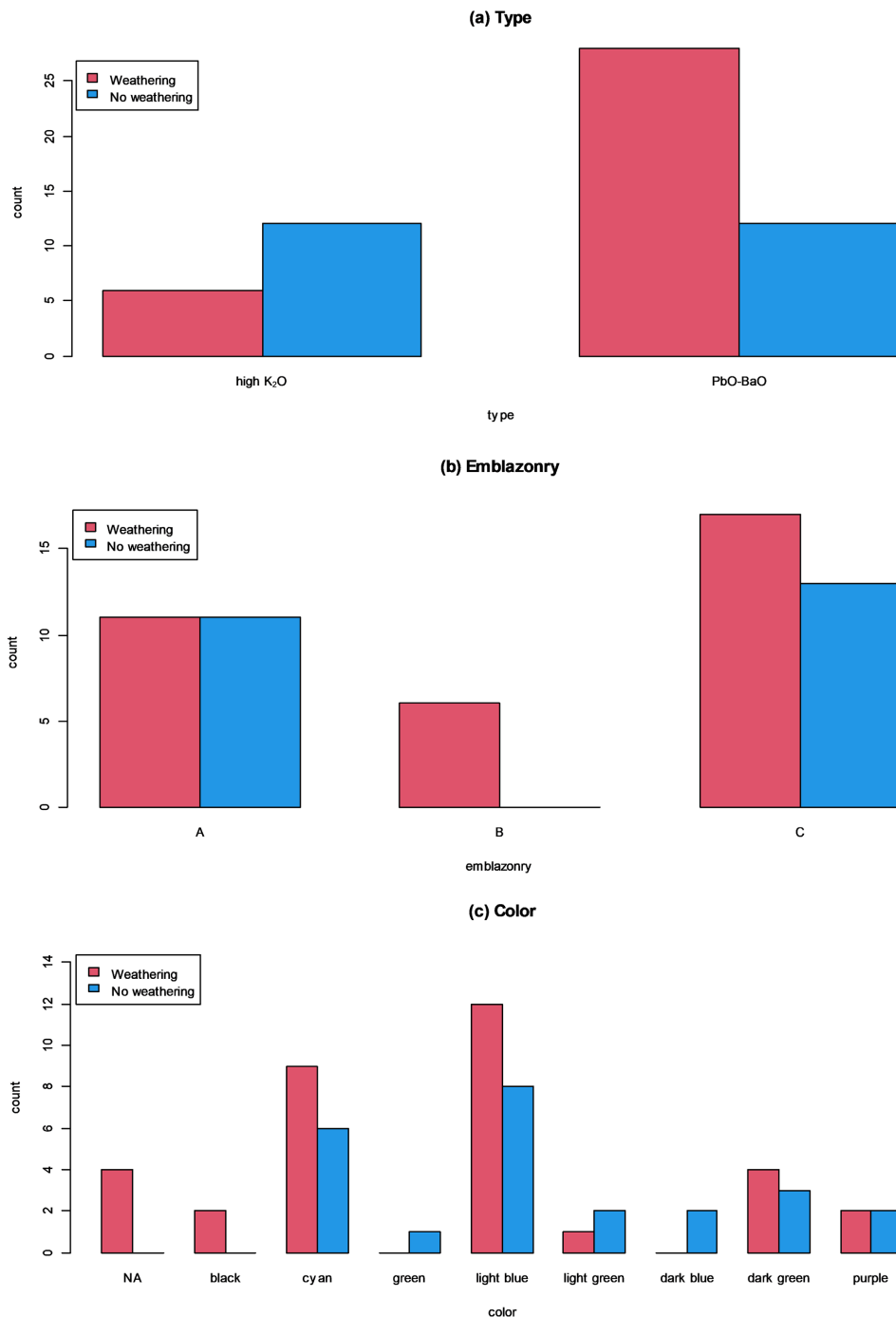
$$y_{ij} = \ln \frac{x_{ij}}{g(x_i)}, \quad g(x_i) = \left(\prod_{j=1}^d x_{ij} \right)^{1/d} \quad (1)$$

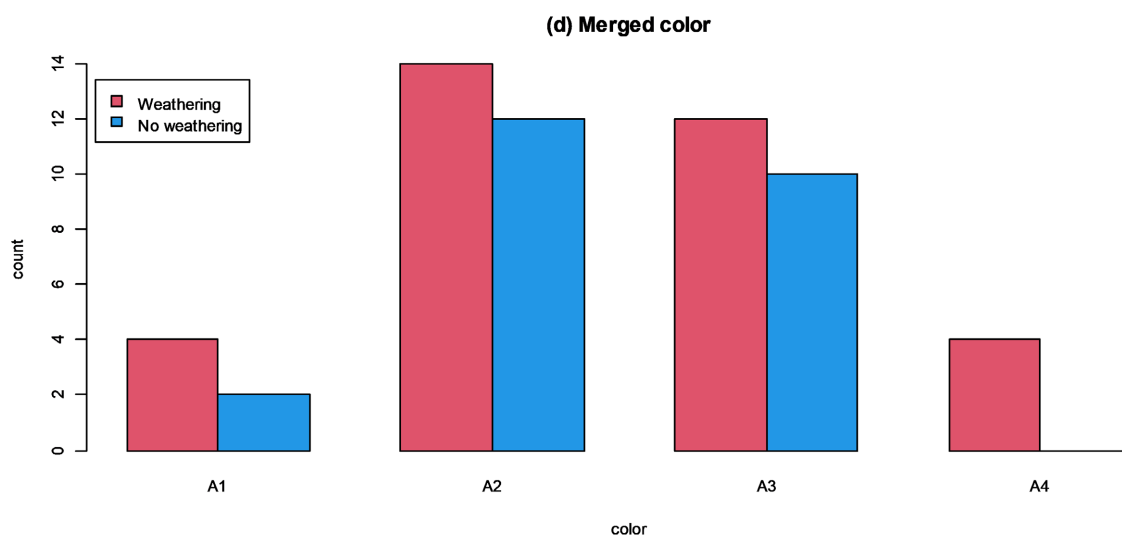
其中, $x_{ij} (i=1, \dots, n; j=1, \dots, d)$ 表示第 j 列成分数据的第 i 个观测, n 和 d 分别为有效观测个数和成分个数。由于这些主要成分之间具有完全多重共线性, 为此在对训练数据建模前需要对其进行特征筛选, 选出其中对区分不同类别玻璃的重要成分, 方便后续分类模型的建立和亚类划分。本文所有计算及图形均使用 R 软件完成。

3.2. 风化与类型、纹饰和颜色的独立性检验

通过对玻璃的表面风化状态与玻璃类型、纹饰和颜色绘制并列条形图(图 2)进行可视化分析, 观察其

差异。我们发现：与高钾玻璃相比，铅钡玻璃表面更易受风化影响；与其它纹饰相比，好像纹饰 B 易受风化影响；而风化状态和颜色的相关性从图 2 看不出。考虑到样本量较小，采用 χ^2 独立性检验分别对风化状态与类型、纹饰和颜色的 3 个二维列联表结果进行检验，给出相应的 χ^2 值和相伴 p 值。考虑到风化与纹饰、颜色的列联表中出现观测频数小于 5 的情况，使用卡方检验容易出现偏差，还给出了 Fisher 精确检验的模拟 p 值(表 1)。针对颜色分组太细的情况，把频数小于 5 的组按相近色合并为 4 个组后再次绘制其并列条形图并作上述检验，相应结果在图 2(d)和表 1 的最后一行给出。





注: A1 = {颜色为“黑”或“紫”}, A2 = {颜色为“蓝绿”、“绿”、“浅绿”或“深绿”}, A3 = {颜色为“浅蓝”或“深蓝”}, A4 = {未给出颜色}。

Figure 2. Parallel bar chart of surface weathering state and glass type, decoration and color

图 2. 表面风化状态和玻璃类型、纹饰和颜色绘制并列条形图

Table 1. Results from both chi-square and Fisher exact tests

表 1. 卡方和 Fisher 精确检验结果

H_0	χ^2 value	p value	Simulated p value of Fisher test
Weathering and type are independent	5.4518	0.0196	0.0113*
Weathering and emblazonry are independent	4.9565	0.0838	0.0846#
Weathering and color are independent	9.4324	0.3071	0.3528
Weathering and merged color are independent	3.3786	0.3368	0.3838

注: 表 1 中*和#分别表示在 0.05 和 0.1 的水平下显著。

从表 1 的检验结果可得到结论: 基于现有数据, 认为玻璃表面风化与其类型不独立, 与纹饰弱相关, 而与颜色基本无关, 与图 2 的可视化分析结果一致。

3.3. 成分含量与风化的统计规律

对高钾/铅钡玻璃经过 CLR 变换后的数据绘制其各成分数据在不同风化状态下的均值的并列条形图进行可视化分析, 图 3 呈现的是其中有显著变化的成分均值对比。

从图 3 可发现风化后高钾玻璃成分比例显著增加的有 SiO_2 、 CuO 、 CaO 、 Fe_2O_3 、 Al_2O_3 , 显著下降的有 K_2O 、 MgO 、 PbO 、 BaO ; 铅钡玻璃成分比例显著下降的有 SiO_2 、 Na_2O 、 K_2O 、 Al_2O_3 , 显著增加的有 P_2O_5 、 CaO 、 Fe_2O_3 、 SrO 。为更好地评价上述有显著变化的成分对表面风化的显著性, 我们也对它们做了箱型图及单因素方差分析进行比较, 表 2 列出了相应的 F 值和相伴 p 值。需要说明的是, 此处及下文 SnO_2 和 SO_2 的显著性我们均未予以考虑, 原因是接近 90% 的采样点未检测出这两种成分。

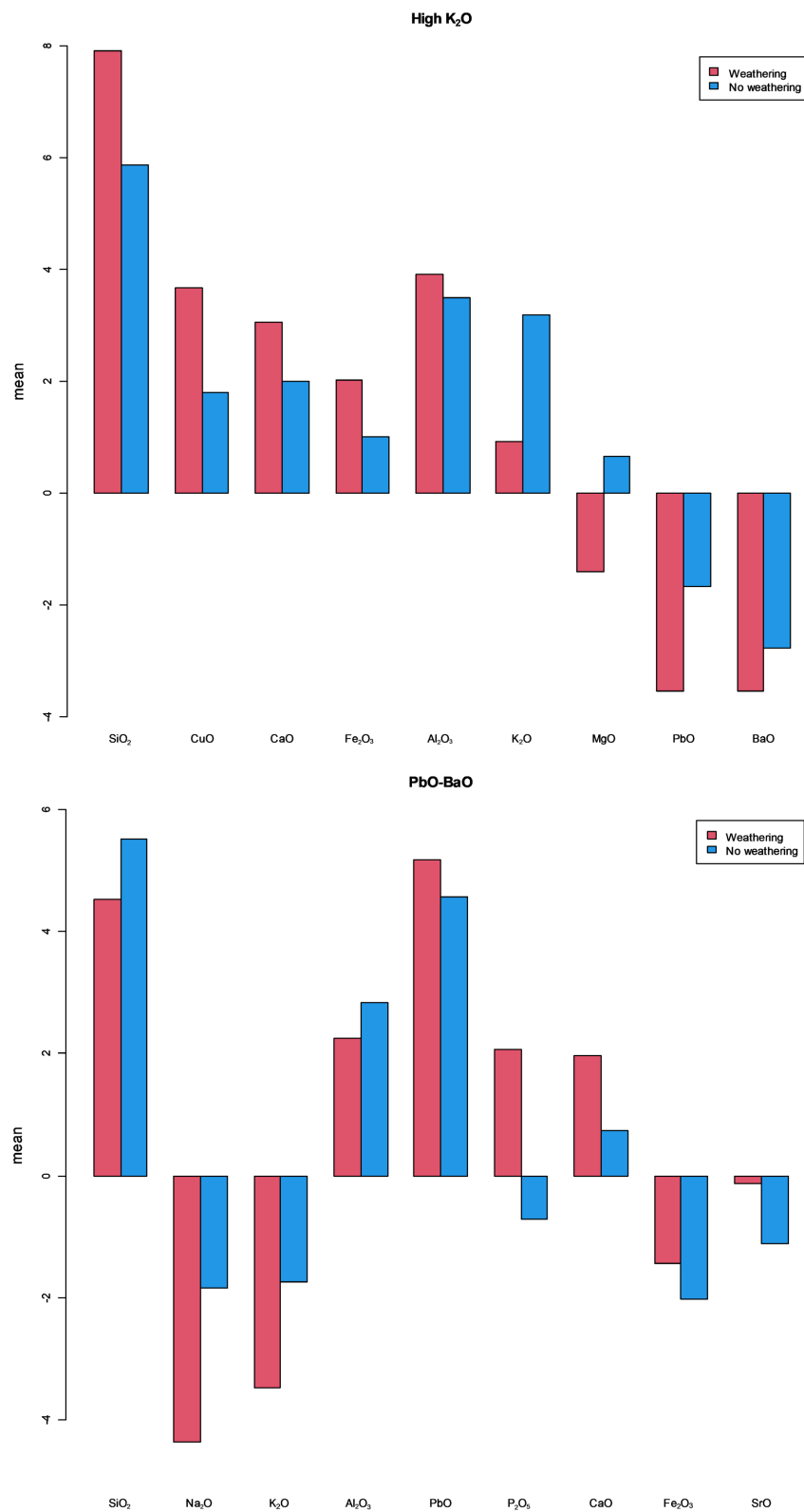


Figure 3. Mean comparison of some compositions of high K₂O (upper)/PbO-BaO (lower) glasses under different weathering states

图 3. 高钾(上)/铅钡(下)玻璃在不同风化状态下部分成分的均值比较

Table 2. Analysis of variance of some composition of high potassium/lead barium glass on surface
表 2. 高钾/铅钡玻璃部分成分对表面风化的方差分析

Composition	high K ₂ O		PbO-BaO	
	F value	p value	F value	p value
SiO ₂	31.3509	4e-05^{***}	19.3865	6.1e-05^{***}
Na ₂ O	0.0324	0.8595	6.9239	0.0115[*]
K ₂ O	2.3449	0.1452	5.3306	0.0254[*]
CaO	0.6230	0.4415	4.9538	0.0309[*]
Al ₂ O ₃	3.3550	0.0857[#]	9.0118	0.0043^{**}
CuO	4.8978	0.0418[*]	0.4645	0.4989
PbO	2.1582	0.1612	6.1300	0.0169[*]
P ₂ O ₅	0.0026	0.9602	12.7208	0.0008^{***}
SrO	0.2808	0.6034	2.9280	0.0936 [#]

注: ^{***}、^{**}、^{*}和[#]分别表示在 0.001、0.01、0.05 和 0.1 水平下显著。

从表 2 可看出: 风化会引起高钾玻璃 SiO₂ 成分比例的显著变化, 一定程度上引起 CuO、Al₂O₃ 成分比例的变化; 而铅钡玻璃由风化引起显著变化的成分比例较多, 按显著性由强到弱排序, 依次为 SiO₂、P₂O₅、Al₂O₃、Na₂O、PbO、K₂O、CaO、SrO。这些结果与图 3 的可视化分析结论是相互印证的。

3.4. 分类模型的建立

为识别能够明显区分两类玻璃的重要成分, 对 14 个 CLR 变换后的成分数据绘制了它们对类型的箱线图进行比较, 并做各成分对类型的单因素方差分析, 从中挑选出 8 个有显著差异的成分 SiO₂、K₂O、Al₂O₃、Fe₂O₃、CuO、PbO、BaO、SrO (图 4 和表 3)。

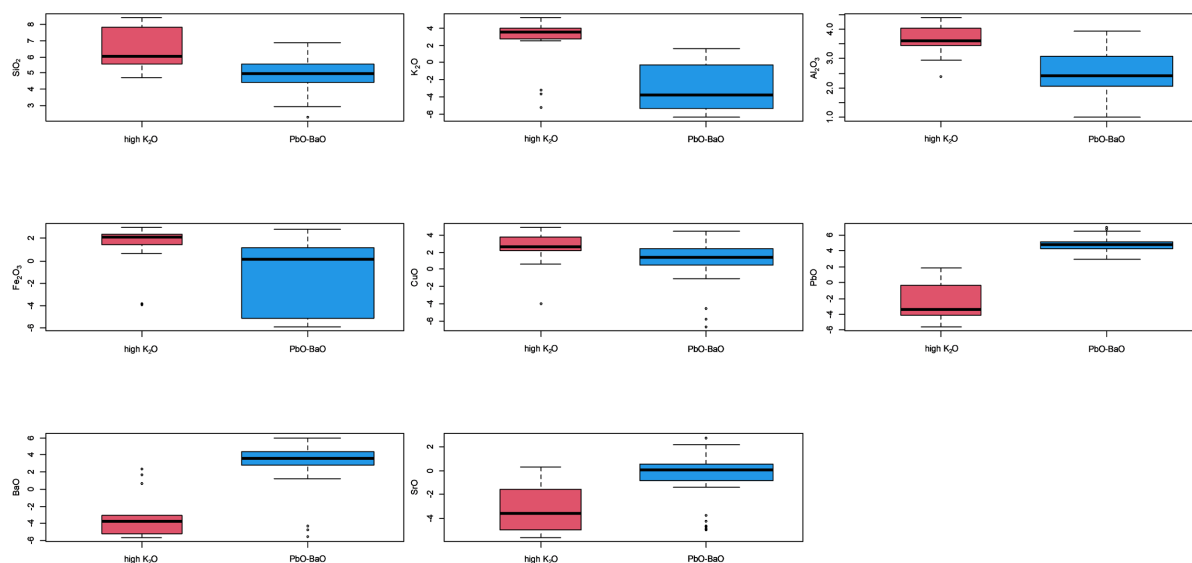


Figure 4. Comparison of box diagram of important compositions to glass types

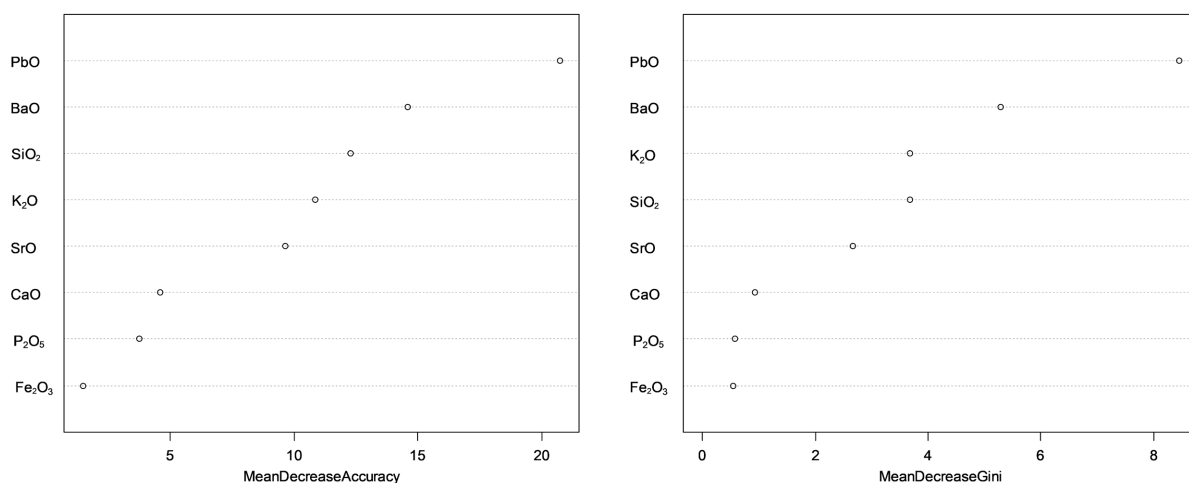
图 4. 重要成分对玻璃类型的箱线图比较

Table 3. Analysis of variance of important compositions on glass type**表 3.** 重要成分对玻璃类型的方差分析

Composition	F value	p value	Composition	F value	p value
SiO ₂	31.9268	3.83e-07 ^{***}	PbO	283.0483	2.26e-25 ^{***}
K ₂ O	43.1445	9.97e-09 ^{***}	BaO	87.2497	1.25e-13 ^{***}
Al ₂ O ₃	34.8645	1.41e-07 ^{***}	CuO	4.9262	2.99e-02 [*]
Fe ₂ O ₃	13.9721	3.94e-04 ^{***}	SrO	23.4977	8.12e-06 ^{***}

注: 表 3 中显著性标注同表 2。

使用表 3 中所列的 8 个重要成分, 建立玻璃类型对它们的归一化数据的 RF 模型[9]。每次有放回地抽 n 个观测, 建了 500 棵树, 每个节点使用 2 个变量, 算得 2 个特征重要性筛选的指标值: OOB 数据作为测试集所做的交叉验证得到的平均减少精度(图 5 左)和平均减少 Gini 值(图 5 右), 从中精选出 4 个足以区分两类玻璃的重要成分: PbO、BaO、SiO₂、K₂O。使用此 4 个重要成分建立对玻璃分类的 RF 模型, 误判率的 OOB 估计值为 0%, 分类准确率为 100%。不考虑风化影响, 直接运用该模型对表单 3 中 8 个未知类别的玻璃文物进行鉴别, 给出其所属类型, 结果见表 4。

**Figure 5.** Ranking of important characteristics under mean decrease accuracy and mean decrease Gini**图 5.** 平均减少精度和平均减少 Gini 值下的重要特征排序**Table 4.** Identification results of unknown types of glass relics**表 4.** 未知类型的玻璃文物进行鉴别结果

label	A1	A2	A3	A4	A5	A6	A7	A8
type	high K ₂ O	PbO-BaO	PbO-BaO	PbO-BaO	PbO-BaO	high K ₂ O	high K ₂ O	PbO-BaO

再分别应用具有线性和径向基核函数的 SVM 建立玻璃类型对上述 4 个重要成分的分类模型, 分类准确率也是 100%, 对表单 3 中未知类别的玻璃文物进行鉴别, 给出的结果与表 4 一致。这说明 RF 和 SVM 模型对本文数据的识别度很高, 且只需通过 4 个重要成分(PbO、BaO、SiO₂、K₂O)的占比就可以对玻璃文物类别做出可靠的判别。

实际上, 仔细观察表单 3 中被鉴别为高钾玻璃的编号为 A1, A6, A7 的化学成分, 会发现它们不论

风化与否, 都有明显高于铅钡玻璃的 SiO_2 占比, 且都未检测到 PbO 、 BaO 成分。事实上, 较高的 SiO_2 占比也是高钾玻璃区别于铅钡玻璃的一个显著特点, 而铅钡玻璃的 PbO 、 BaO 占比要明显高一些, 这一点从 2.5 中图 6 容易看出。

另外, 2.3 的研究表明表面风化会对玻璃主要化学成分的占比有一定影响, 但在筛选重要变量环节, 我们发现表面风化对玻璃类型的影响排在上述 4 个重要成分之后, 因而使用我们的模型鉴别玻璃类型无需特别关注表面风化因素, 但这并不意味着它对鉴别玻璃类型不起作用。事实上, 由于风化会引起 SiO_2 、 Al_2O_3 在高钾玻璃中明显富集, 而在铅钡玻璃中显著贫乏; PbO 在高钾玻璃中则显著减少、在铅钡玻璃中显著增加, 从而使得玻璃类型的鉴别更加容易。

3.5. 亚类划分

亚类划分是一个无监督分类问题, 同样由于 14 个化学成分的相关性, 在使用聚类模型前, 需要先对每个类型的玻璃选择合适的成分。通过比较高钾/铅钡玻璃各成分均值大小的方法给出比例由高到低的前 8 个重要成分的排序结果: 高钾玻璃为 SiO_2 、 K_2O 、 Al_2O_3 、 CaO 、 CuO 、 Fe_2O_3 、 P_2O_5 、 MgO , 铅钡玻璃为 SiO_2 、 PbO 、 BaO 、 Al_2O_3 、 P_2O_5 、 CaO 、 CuO 、 Na_2O (图 6)。

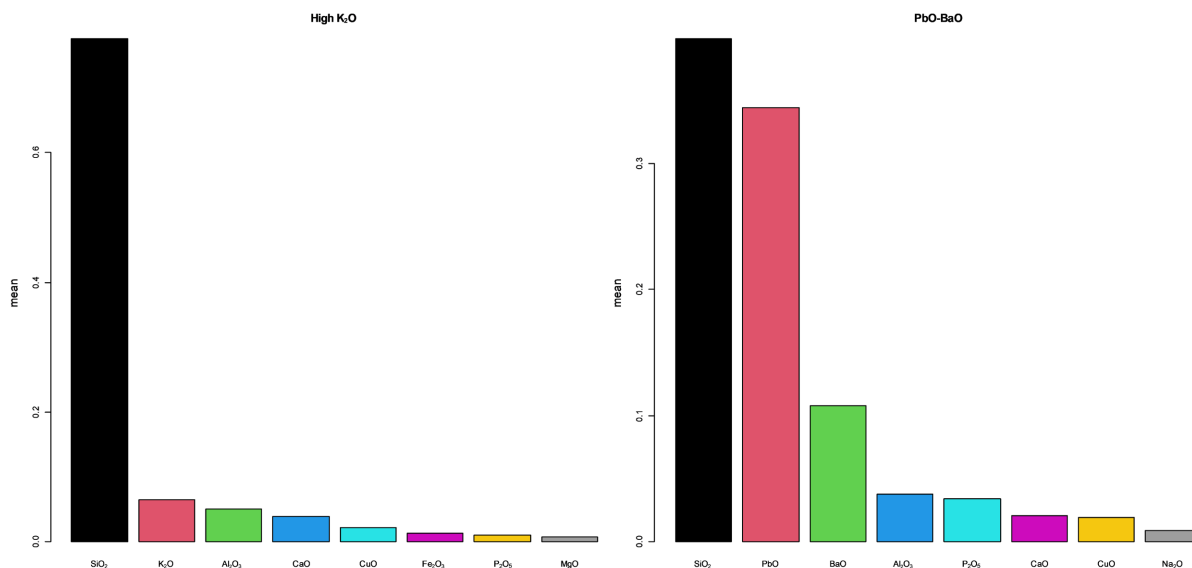


Figure 6. Mean ranking of the first eight important compositions of high K_2O /PbO-BaO glass

图 6. 高钾/铅钡玻璃前 8 个重要成分的均值排序

使用 k -means 算法对高钾/铅钡进行亚类划分[9]。使用类内总平方和(WSS)准则, 高钾、铅钡各分两类时, WSS 的下降基本平缓, 其占总平方和的比例分别达到 80.6%和 61.9%。考虑到高钾、铅钡玻璃分别只有 18 和 49 个采样点, 样本量偏小, 我们只对各个类型玻璃再细分为两个亚类。仍以高钾玻璃为例阐述具体聚类过程: 先对所选出的 8 个重要成分比例数据的每一列 x_j 使用 0-1 正规化方法 $[x_j - \min(x_j)] / [\max(x_j) - \min(x_j)]$ 做同一尺度化处理; 再利用高维空间中的欧氏距离算出数据的相似性矩阵; 然后使用 cmd 进行数据降维, 图 7 呈现的是降到二维的相应分类散点图, 其中同一形状的点表示在同一类。

从图 7 (左)可看出, 从 cmd 的第一个维度判断是否小于 0 就可以把高钾玻璃清晰地分为两类, 通过对它们各自所含重要成分的均值比较, 发现 SiO_2 的含量足以区分这两类, 因而按照 SiO_2 含量高或低依次

命名为高钾(高硅)类和高钾(低硅)类, 分别对应图 7 (左)左侧的 8 个点和右侧的 10 个点, 详见表 5。

对铅钡玻璃, 从图 7 (右)可看出, 从 cmd 的第一个维度判断是否大于 0.1 可以把它们清晰地分为两类, 再通过对它们各自所含重要成分的均值比较, 发现右侧 21 个点的 SiO_2 含量偏高, PbO 含量偏低, 而左侧 28 个点的 PbO 含量高, SiO_2 含量偏低, 故将它们分别命名为铅钡(高硅低铅)类和铅钡(高铅低硅)类, 详见表 5。

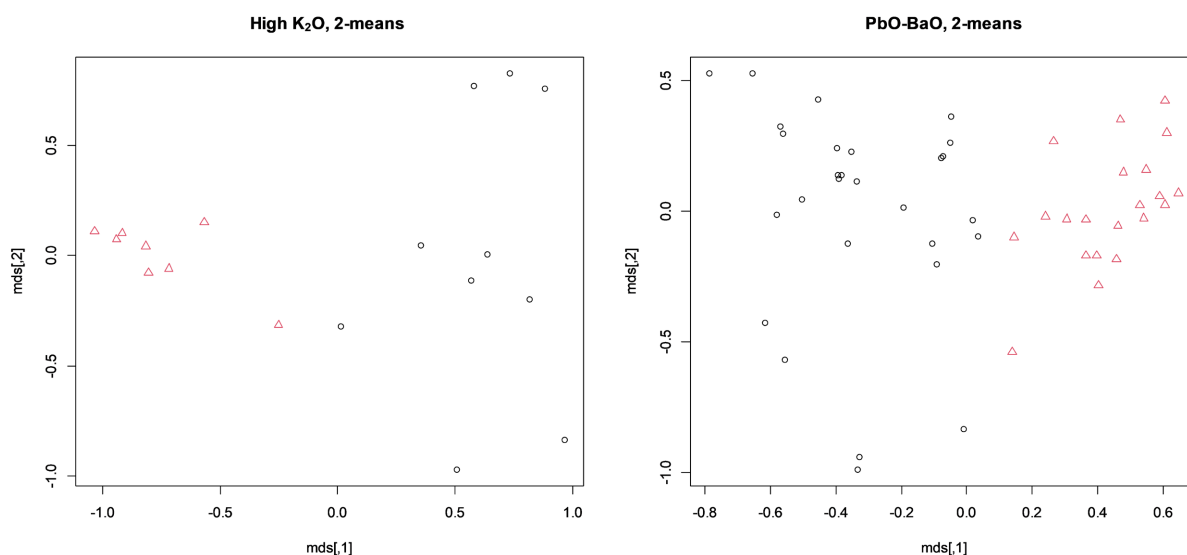


Figure 7. K-means clustering results of high K_2O / PbO - BaO glasses

图 7. 高钾/铅钡玻璃的 k-means 聚类结果

Table 5. Sub-classification results of high K_2O / PbO - BaO glasses

表 5. 高钾/铅钡玻璃的亚类划分结果

Type	Character	Subclass naming	Observation No.
High K_2O	High SiO_2	K_2O -high SiO_2	3, 9, 12, 13, 15, 19, 23, 29
	Low SiO_2	K_2O -low SiO_2	1, 4, 5, 6, 7, 8, 16, 17, 18, 22
PbO-BaO	High SiO_2 and low PbO	PbO-BaO-high SiO_2	21, 24, 26, 30, 31, 34, 35, 36, 38, 39, 40, 45, 46, 49, 50, 51, 52, 53, 55, 61, 64
	High PbO and low SiO_2	PbO-BaO-high PbO	2, 10, 11, 14, 20, 25, 27, 28, 32, 33, 37, 41, 42, 43, 44, 47, 48, 54, 56, 57, 58, 59, 60, 62, 63, 65, 66, 67

3.6. 模型的稳健分析

为考察 2.4 所建分类模型和 2.5 亚类划分模型的稳健性, 我们通过对数据增加微小随机扰动对模型的稳健性进行分析。具体为: 在预处理后的表 2 和表 3 的成分比例数据的每个数值上, 加上一个均值为 0 的任意正态随机数, 标准差依次取 $1e-4$, $1e-3$, $1e-2$, $1e-1$, 以表示对数据的干扰程度逐渐加强, 然后对扰动后的数据重复 2.4 和 2.5 的工作共 4 次, 结果均显示 2.4 所给几个模型的分类型准确率、对表 3 的分类鉴别结果和 2.5 的亚类划分结果没有改变, 这说明我们的模型是稳健的, 在后续玻璃文物的鉴别和亚类划分中均可放心使用。

4. 补充说明

我们也根据风化点检测成分数据, 先预测其风化前的成分含量。以高钾玻璃为例, 对由风化引起比

例有显著变化的成分, 先通过 QQ 图和 KS 检验检查其风化与未风化组各自中心尺度化后的分布匹配情况, 此处选择中位数和 MAD 分别作尺度化的中心和尺度, 主要是因为它们较均值和标准差更稳健, 分布的匹配程度也更好。我们发现大多数成分数据的分布是匹配的, 对风化成分数据利用式(2)的配对方法预测其无风化的数值。

$$\hat{y}_{ij} = m(Y_{nofen,j}) + \frac{s(Y_{nofen,j})}{s(Y_{fen,j})} [y_{ij} - m(Y_{fen,j})] \quad (2)$$

其中 $Y_{fen,j} = (y_{1j}, \dots, y_{n_j})^T$, n_j , $m(Y_{fen,j})$, $s(Y_{fen,j})$ 分别表示风化玻璃的第 j 个成分向量及其样本个数、中位数和 MAD, 未风化玻璃的相应记号为 $Y_{nofen,j}$, $m(Y_{nofen,j})$, $s(Y_{nofen,j})$, $j = 1, \dots, d$ 。铅钡玻璃的分析和预测类似, 故略去。

接着, 对高钾/铅钡玻璃的风化成分数据使用其预测数据, 通过 CLR 的逆变换将其还原为比例数据后与未风化成分比例数据合并, 作为 2.4 和 2.5 的建模数据进行分析, 发现对于那些由检测误差造成的成分比例之和明显偏低的观测值, 由于样本观测太少及匹配变换产生的估计误差使得成分比例估计偏差较大, 从而造成模型误判。同样采用 4 个成分 PbO、BaO、CaO、Fe₂O₃, 建立对玻璃分类的 RF 模型, 误判率的 OOB 估计值为 4.48%, 有 3 个观测被误判, 运用该模型对表 3 的玻璃文物进行鉴别的结果同表 4, SVM 模型所给结论类似, 但亚分类的结果与 2.5 有较大差异。这意味着在检测和分析的各环节尽量减少误差对玻璃文物类别的准确识别是非常必要的。

5. 结束语

本文基于玻璃文物的主要化学成分, 应用 RF 和 SVM 这两种适用于小样本的分类模型对其进行类型鉴定, 给出了仅使用 4 个重要成分、分类准确率 100% 的简约模型, 并对未知类型的玻璃给出分类结果; 使用 k-means 聚类算法对两类玻璃进行亚类划分, 分别基于 8 个重要成分对所有样本给出了清晰的划分; 通过对成分数据增加微小随机扰动证实了所建模型的稳健性; 最后, 通过补充说明对另一种可行的方法阐述了我们的观点。对于本文这种小样本数据, RF 和 SVM 的泛化能力和稳健性一般优于神经网络等复杂机器学习模型, 这是我们采用它们的重要原因。由于文物本身具有稀缺性的特点, 考古学与其它学科的交叉研究明显偏少, 本文把统计模型这一便捷工具应用于文物鉴定方面, 旨在抛砖引玉, 希望可以在一定程度上填补这方面的不足。本文的研究方法也适用于其他行业的样本分类问题。

基金项目

2023 年北京市大学生创新创业训练计划(B008)。

参考文献

- [1] 周良知. 影响硅酸盐玻璃风化的主要因素[J]. 大连轻工业学院学报, 1984(1): 34-44.
- [2] 王承遇, 陶瑛, 陈敏, 等. 钠钙铝镁硅酸盐玻璃和碱铅硅酸盐玻璃的风化[J]. 硅酸盐通报, 1989(6): 1-9.
- [3] 王婕, 李沫, 马清林, 等. 一件战国时期八棱柱状铅钡玻璃器的风化研究[J]. 玻璃与搪瓷, 2014, 42(2): 6-13.
- [4] 付强, 邝桂荣, 吕良波, 等. 广州出土汉代玻璃制品的无损分析[J]. 硅酸盐学报, 2013, 41(7): 994-1003.
- [5] 赵凤燕, 陈斌, 柴怡, 等. 西安出土若干玻璃器的 pXRF 分析及相关问题探讨[J]. 考古与文物, 2015(4): 111-119.
- [6] 王祉皓, 赵梦澈, 李智群, 等. 基于机器学习的风化硅酸盐玻璃原成分预测及亚分类方法[J]. 硅酸盐学报, 2023, 51(2): 416-426.
- [7] 何秋妍, 邓明华. 中药材鉴别的统计建模与分析[J]. 数学建模及其应用, 2022, 11(1): 91-105.

- [8] 中国工业与应用数学学会. 2022 年高教社杯全国大学生数学建模竞赛赛题[EB/OL].
http://www.mcm.edu.cn/html_cn/node/388239ded4b057d37b7b8e51e33fe903.html, 2023-05-12.
- [9] 吴喜之. 多元统计分析——R 与 Python 的实现[M]. 北京: 中国人民大学出版社, 2019: 109-144, 200-204.