



古代玻璃分类与成分复原模型

刘苏霆¹, 周彤¹, 李皓², 覃思义^{3*}

(1. 电子科技大学信息与通信工程学院, 成都 611731; 2. 电子科技大学计算机科学与工程学院, 成都 611731;
3. 电子科技大学数学科学学院, 成都 611731)

摘要: 分析丝绸之路古代玻璃的成分, 有利于了解历史的脉络。由于恶劣环境条件的影响, 玻璃会出现风化的情况, 这对玻璃成分分析和类别确定带来困难。该文提出了一种综合鉴定方法: 对风化前的玻璃成分进行预测, 并利用投票法和 K-Means 方法建立类别的划分模型。经过验证, 类别划分模型达到了 98.44% 的准确度, 且有较好的鲁棒性。

关键词: 方差分析; 投票法; K-Means; 鲁棒性

中图分类号: TP75

文献标志码: A

DOI: 10.12179/1672-4550.20230166

Classification and Composition Restoration Model of Ancient Glass

LIU Suting¹, ZHOU Tong¹, LI Hao², QIN Siyi^{3*}

(1. School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China;
2. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China;
3. School of Mathematics, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: The analysis of the composition of ancient glass on the Silk Road is helpful to understand the context of history. Due to the influence of adverse environmental conditions, the glass will be weathered, which brings difficulties to the composition analysis and classification of glass. In this paper, a comprehensive identification method is proposed, the glass composition before weathering is predicted, and the classification model is established by using voting method and K-Means method. After validation, the category classification model achieves an accuracy of 98.44% with good robustness.

Key words: analysis of variance; voting; K-Means; robustness

我国具有悠久的玻璃制作历史, 玻璃是丝绸之路历史的重要物证, 在考古工作中, 需要对玻璃的成分进行分析, 并对其种类进行鉴别^[1]。玻璃以石英砂(主要成分为 SiO₂)作为主要原料, 在制造过程中为了降低熔点, 添加不同的助熔剂进行炼制, 并加入氧化钙等作为稳定剂。根据所加入助熔剂的不同, 玻璃化学成分产生差异, 这使得古代玻璃产生了不同的种类。如添加铅矿石作为助熔剂的玻璃具有较高的氧化钡和氧化铅含量, 常被称为铅钡玻璃; 添加草木灰等作为助熔剂的玻璃, 具有较高的钾含量, 被称为钾玻璃。

古代玻璃经由漫长的埋藏过程而导致风化。风化的本质是玻璃与外界环境之间产生了元素交

换, 导致成分发生变化。风化程度的不同, 导致玻璃的成分、颜色发生较大变化, 给考古工作带来了一定困难。我们需要利用古代玻璃采样的化学检测数据, 探究风化与物质成分的变化关系, 以及分类标准与成分之间的相关关系。

利用一组古代玻璃样本数据, 首先对不满足成分要求的数据进行剔除。为了分析风化与类型、纹饰、颜色之间的关系, 考虑使用方差分析和卡方检验^[2], 分析 3 个因素与风化之间的关系。

在得到相关关系后, 进行风化前成分预测工作。利用两类玻璃风化前相对于风化后的均值变化程度, 以及原始数据中高钾玻璃的对照组, 即同一样本上风化前和风化后的部位, 找到第二个

收稿日期: 2023-03-24; 修回日期: 2024-04-05

基金项目: 国家自然科学基金(12071062)。

作者简介: 刘苏霆(2001-), 男, 本科生, 信息与通信工程专业。

*通信作者: 覃思义(1971-), 男, 硕士, 教授, 主要从事应用建模方面的研究。E-mail: qinsi yi@uestc.edu.cn

改变程度参考值。最后将两个改变量通过赋权的方式结合在一起,预测高钾玻璃风化前成分组成,因铅钡玻璃数据没有给出对照组,所以只使用第一个变化程度预测。

为了探寻两类玻璃的分类规律,我们先将风化后的数据还原为风化前的数据,然后计算各个成分在两类玻璃中的相对差值,并依据相对差值为各个成分打出分数,作为此成分区分两类玻璃的评分。

为了在每个大类下继续划分出亚类,挑选出对划分亚类有重要影响的指标,再用 K-Means++ 进行聚类,利用卡方分布的相关性系数检验聚类的合理性,采用合理性最强的聚类结果作为最终结果,并依据文献 [3-4] 中亚类的命名方式选取每一类中含量较大的一或两个指标对亚进行命名。最后,通过上下浮动原始数据重新进行分类来分析模型灵敏度。

1 玻璃成分相关性的分析

为了探究古代玻璃的纹饰种类、类型与颜色等因素的差异对于玻璃风化程度的影响,对这 3 个因素进行相关性分析。首先使用方差分析^[5] 检验相关性。若将纹饰视作考察的因素,那么不同的纹饰 {A, B, C} 则可以视为 r 个不同的水平,本文 $r = 3$ 。若用文献 [6] 所给 58 个古代玻璃样本数据进行分析,在各个样本独立同分布的条件下,记录第 k 种纹饰中第 i 个样本的风化程度为 X_{ik} , 每种纹饰的样本数量分别为 n_1, n_2, \dots, n_r 。根据上述信息,可以得到水平项离差平方和 (SS_A)、误差项离差平方和 (SS_E) 的统计量分别为:

$$SS_A = \sum_{i=1}^r n_i (\bar{X}_i - \bar{X})^2 \quad (1)$$

$$SS_E = \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad (2)$$

方差分析的基本思想是通过水平项离差平方和、误差项离差平方和的统计量,判断假设 $H_0: \mu_1 = \mu_2 = \dots = \mu_r$ 是否成立,这里的 μ 代表风化程度的均值。当原假设成立时满足:

$$\frac{SS_A}{\sigma^2} \sim \chi^2(r-1), \quad \frac{SS_E}{\sigma^2} \sim \chi^2(n-r)$$

式中: $n_1 + n_2 + \dots + n_r = n$ 。以 F 的显著性大小来检验纹饰种类同风化之间的关联程度:

$$F = \frac{SS_A/(r-1)}{SS_E/(n-1)} = \frac{MS_A}{MS_E} \quad (3)$$

式中: MS_A 为组间均值, MS_E 为组内均方。

在给定的显著性水平 α 下,原假设 H_0 的拒绝域为 $F \geq F_\alpha(r-1, n-r)$ ^[7]。而对于颜色和类型的水平与风化程度的分析方面,可以使用相似的方法进行分析。3 个指标相关性分析的结果如表 1 所示。

表 1 根据类型、纹饰和颜色的方差分析结果

因素	指标		
	类型	纹饰	颜色
SS_A	1.6690	1.2023	2.2880
SS_E	12.4000	12.8667	11.7810
MS_A	1.6690	0.6011	0.2542
MS_E	0.2214	0.2339	0.2454
F	7.5373	2.5697	1.0358

为了验证所得结果的正确性,使用卡方分布验证 3 个因素同风化程度之间是否呈现显著性。卡方分析结果如表 2 所示。

表 2 3 个因素的卡方分析结果表

表面风化因素	χ^2	p	显著性
类型	6.880	0.009**	高显著性
纹饰	4.957	0.084	无显著性
颜色	9.432	0.307	无显著性

*: 表示 $p < 0.05$, 说明具有显著性; **: 表示 $p < 0.01$, 说明具有高显著性。

分析结果,可以观察到在 3 个因素中只有类型产生了高显著性,而纹饰、颜色两个因素并未得出显著性结论,故可以认为我们所使用的方法具有合理性。

2 玻璃样品风化前含量预测模型

如果要预测样本风化前数据的成分占比,在理想情况下,需要通过同一样本的风化数据与未风化数据构成一对正反样本。在若干对正反样本中分析风化前后的数据变化,得到稳定结果。鉴于实际情况中难以形成成对数据,我们提出一种基于均值的原始成分预测模型。

首先分析不同样本风化前后的元素对比,这是因为该部分可利用的数据较多。为了减弱不同古代玻璃样本的成分含量差异,使用均值来计算某一类型的玻璃元素含量。若使用 c_{ij} 代表某一类别中第 i 条记录中的第 j 项化学成分的占比,那么对于该指标的均值可以计算为:

$$\mu_j = \frac{\sum_{j=1}^{14} c_{ij}}{n} \quad (4)$$

这里 n 代表采样点记录的数量。如果使用 (μ, μ') 分别代表风化前与风化后的均值, 那么可以计算出跨样本变化比例 Δ_1 。对于同一样本的不同风化程度数据, 同样可以计算其化学成分变化比例, 令风化前后的比例分别为 (c, c') , 那么可以计算样本内变化比例为:

$$\Delta_2 = \frac{\sum_{i=1}^m (c_{ij} - c'_{ij})}{mc'_{ij}} \quad (5)$$

综合上述两个比例, 加权相加后就可以得到综合变化比例为:

$$\Delta_0 = \omega_1 \Delta_1 + \omega_2 \Delta_2 \quad (6)$$

式中: ω 代表权重。在得到综合变化比例 Δ_0 后, 给出风化前成分的计算方法为:

$$c_{ij}^k = \begin{cases} (1 + \Delta_{K0})c_{ijk}, k = K \\ (1 + \Delta_{B0})c_{ijk}, k = B \end{cases} \quad (7)$$

经过计算, 在给定 $\omega_1 = 0.7$ 和 $\omega_2 = 0.3$ 的情况下, 计算出高钾玻璃和铅钡玻璃各个化学元素的综合变化比例, 如表 3 所示。表中 D_{K0} 为高钾玻璃, D_{B0} 为铅钡玻璃。

表 3 各化学成分的变化趋势

化学成分	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)
D_{K0}	-0.2765	Inf	16.1842	5.1287	4.4772
D_{B0}	1.0203	5.8980	0.6466	-0.5247	-0.0684
化学成分	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)
D_{K0}	2.4301	6.2906	0.5704	Inf	Inf
D_{B0}	0.4379	0.1150	-0.3552	-0.4947	-0.1406
化学成分	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)	
D_{K0}	4.0071	Inf	Inf	Inf	
D_{B0}	-0.7796	-0.2834	-0.3896	-0.1929	

对结果进行分析发现, 由于古代玻璃样品的特殊性, 每件样品的制作工艺不同, 导致成分具有差异, 流失情况也各有不同, 因此不能随意搭配预测的数据同原始未风化的数据验证正确性。在这一前提条件下, 我们提出一种检验相关性的方法, 将文献 [6] 玻璃数据中原始未风化的化学成分记录 r 与预测出的数据中的化学成分记录 r' 相混合, 得到未风化数据集 R 。并且使用“是否预测”这一 0-1 标签 f 分别标注数据来源为 r 或是 r' 。通过探究各个化学成分占比同数据来源的相关性, 判断该预测模型的性能。

使用相关性分析的理论依据在于: 若相关各个化学成分的占比 c_j 这一指标不能够揭示数据是否被预测而来, 那么就相当于该预测数据“隐藏”在原始数据中, 同是否预测之间没有显著的相关关系。这样可以推知预测出的数据同原始数据之间的差异不大, 认为预测有效。若所有指标均同“是否预测”之间没有显著性关联, 那么就

可以认为预测效果较好。经过权衡, 使用皮尔逊相关系数^[8]来检测 R 中各化学成分占比 c_j 与是否预测指标 f 之间的相关性进行分析, 最终可知各化学成分占比同是否预测之间相关系数值 p 均大于 0.05, 说明没有相关性。

3 玻璃样品类别划分模型

本文提出一种古代玻璃的类型划分方法, 即给出铅钡玻璃和高钾玻璃之间的分类关系; 在这一基础上, 对两种玻璃进一步划分更细的亚类, 并做敏感性分析, 解释其合理性。针对大类划分模型, 利用预测风化前的样本数据, 统计高钾玻璃与铅钡玻璃间各个化学元素的分界点, 以投票方式决定一个样本的种类。针对亚类划分问题, 使用专家知识以及变异系数^[9], 挑选重要指标, 使用 K -means++ 聚类方法, 利用相关系数确定最佳类中心数目。在得出聚类结果后, 对原始数据的化学组成变化进行敏感度分析。

3.1 大类划分

由于风化后的玻璃样本元素含量变化,不易于分类,我们对风化样本的含量数据利用前文所述的含量预测模型对原始含量预测后做类别划分。

根据指标 c_{ij} 与划分界限 h_j 相对大小,判断玻璃在这一指标上倾向于哪一类 v_j 。确定划分界限 h_j 的方法根据该指标在不同大类下的分布差异而有所不同。部分指标间有种类间的明显分界,如氧化钾、氧化钙、氧化铅和氧化锡等。其余指标虽然种类之间没有明显间隔,但是仍体现出相对的大小规律。

当相对含量的大小差异明显时,我们给出划分标准的计算式:

$$h_j = \begin{cases} [c_{bj,\max} + c_{aj,\min}] / 2, \mu_{0j} < \mu_{1j} \\ [c_{bj,\min} + c_{aj,\max}] / 2, \mu_{0j} > \mu_{1j} \end{cases} \quad (8)$$

式中: c 代表含量, μ_0 、 μ_1 分别代表铅钡玻璃和高钾玻璃在第 j 指标上的均值。对于含量的大小差异不明显情况,使用两类均值的中点来作为分界 h_j 。

综上,在得到界限 h_j 后可对第 i 条记录在第 j 项成分上的倾向性 v_j 进行判断,根据指标同界限 h_j 的相对大小进行倾向划分。 v_j 是一个 0-1 变量,1 代表样本为高钾玻璃,0 代表铅钡玻璃。随后赋权对指标倾向的结果求和,便可得到最终的分类结果。首先计算样本在指标 j 上的均值距离,与样本均值中点的商,并将其归一化后作为权重:

$$D_j = \frac{2|\mu_{0j} - \mu_{1j}|}{\mu_{0j} + \mu_{1j}} \quad (9)$$

$$w_j = \frac{D_j}{\sum_{j=1}^{14} D_j} \quad (10)$$

最终可以得到分类函数:

$$C = \sum_{j=1}^{14} w_j v_{ij} \quad (11)$$

当输出结果大于 0.5 时认定为高钾玻璃,结果为 0 时认定为铅钡玻璃。利用 64 条数据进行分类测试,高钾玻璃 18 项全部分类正确,46 项铅钡玻璃仅有 1 处错误,分类正确率为 98.44%。

3.2 亚类划分

我们首先进行亚类指标筛选。查阅相关资

料,一些文献常常基于相对丰度的大小,进行亚类的细分^[3-4]。这一划分方式以某一指标的绝对含量进行分析,满足某个条件时便属于某类玻璃。但由于古代玻璃制作工艺的特性,这些指标可能仅对特定地区的玻璃样品具有意义。因此我们仅参考上述文献所提及的成分指标,包括 Na、Al、Ca、Cu、Fe 等元素。

此外,利用样本的统计规律数据,通过变异系数 c_v 筛选出变化幅度较大的若干指标作为聚类指标。考虑到均值过小时变异系数产生异常,对含量均值小于 1% 的指标也不予考虑。鉴于在区分高钾玻璃和铅钡玻璃的过程中,氧化铅、氧化钾和氧化钡发挥主要作用,在亚类分析时可不考虑上述 3 个因素;二氧化硅为玻璃制品的基础元素,不因为类别不同而成分有所不同,且其占据的成分比例远高于其余成分,会影响其余指标的距离计算,所以我们不分析二氧化硅。经过上述分析,最终确定高钾玻璃的亚类划分指标为氧化铝、氧化钙、氧化铜、氧化铁、铅钡玻璃的划分指标为氧化铝、氧化钙、氧化铜、五氧化二磷和氧化钠。在筛选重要指标后进行聚类,使用 K-Means++ 算法^[10-11]进行聚类。聚类类别数量的选择依据聚类结果的 F 检验相关系数,选择相关系数之和最小的类别为聚类类别数^[12-13]。最终将铅钡玻璃划分为“高 Ca-Al 玻璃”“Ca-Al 玻璃”“高 Cu-Fe 玻璃”“高 Al-P 玻璃”和“高 Al-Fe 玻璃”5 类,将高钾玻璃划分为“高 Cu 玻璃”“Al-Ca 玻璃”“Al-Na 玻璃”“高 Al 玻璃”“Al-P 玻璃”5 类。

最后,我们就亚类划分的情况做了敏感性分析,对于元素的占比数据添加噪声,并记录了分类的变化情况。在数据量大的情况下,某一成分含量波动 10%,分类准确率仅仅下降到 93.48%。

4 结束语

本文针对古代玻璃风化后难以确定原始含量和类别的问题,提出了一种风化成分复原以及分类模型。该方法具有较好的稳定性,有利于分析玻璃样本的原始成分和分类工作。

参考文献

- [1] 干福熹. 中国古代玻璃技术的发展[M/OL]. 上海科学技术出版社, 2005. [2022-09-16]. <http://book.ucdrs>.

- superlib.net/views/specific/2929/bookDetail.jsp?dxNumber=000005333381&d=6A48A360DFED20083A546741E97EAAAB&fenlei=18180802.
- [2] 卡方检验[M/OL]//维基百科,自由的百科全书. [2022-09-17]. <https://zh.wikipedia.org/wiki/%E5%8D%A1%E6%96%B9%E6%A3%80%E9%AA%8C>.
- [3] 干福熹, 承焕生, 李青会. 中国古代玻璃的起源: 中国最早的古代玻璃研究[J/OL]. 中国科学 (E 辑: 技术科学) (3): 382–391 [2022-09-18]. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2007&filename=JEXK200703007&uniplatform=NZKPT&v=B760xYkhXPu73zRrkQUN56HS55jHffIP7RdbV82quAb_ljFkNkJA1g5E0bUswCE7.
- [4] 斯琴毕力格, 李青会, 干福熹. 激光剥蚀-电感耦合等离子体-原子发射光谱/质谱法分析中国古代钾玻璃组分[J]. 分析化学, 2013, 41(9): 1328–1333.
- [5] 方差分析[M/OL]//维基百科,自由的百科全书. [2022-09-16]. <https://zh.wikipedia.org/w/index.php?title=%E6%96%B9%E5%B7%AE%E5%88%86%E6%9E%90&oldid=73116586>.
- [6] 中国工业与数学应用会. 2022 年高教社杯全国大学生数学建模竞赛赛题 C 题[EB/OL]. [2022-09-16]. http://www.mcm.edu.cn/html_cn/block/8579f5fce999cdc896f78bca5d4f8237.html.
- [7] F-distribution tables[EB/OL]. [2022-09-16]. http://www.socr.ucla.edu/Applets.dir/F_Table.html.
- [8] HAUKE J, KOSSOWSKI T. Comparison of values of pearson's and spearman's correlation coefficientson the same sets of data[J/OL]. Quaestiones Geographicae, 30(2): 87–93. [2022-09-17]. <https://sciendo.com/article/10.2478/v10117-011-0021-1>. <https://zh.wikipedia.org/w/index.php?title=%E5%8F%98%E5%BC%82%E7%B3%BB%E6%95%B0&oldid=72175861>.
- [9] 变异系数[M/OL]//百度百科. [2022-09-17]. <https://baike.baidu.com/item/%E5%8F%98%E5%BC%82%E7%B3%BB%E6%95%BD/6463621>.
- [10] K 均值聚类[EB/OL]. [2022-09-18]. https://ww2.mathworks.cn/help/stats/kmeans.html?s_tid=srchtitle_kmeans_1#bueft14-1.
- [11] ARTHUR D, VASSILVITSKII S. K-means++: the advantages of careful seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Louisiana, USA: ACM, 2007.
- [12] F 检验[EB/OL]. [2022-09-18]. <https://baike.baidu.com/item/F%E6%A3%80%E9%AA%8C/9910842?qq-pf-to=pcqq.group>.
- [13] F 检验[M/OL]//维基百科,自由的百科全书. [2022-09-18]. <https://zh.wikipedia.org/wiki/F%E6%A3%80%E9%AA%8C>.

编辑 张俊