

赵华振, 焦志豪, 杨润林, 等. 基于层次聚类的古代玻璃的亚类划分[J]. 智能计算机与应用, 2026, 16(4): 91-95. DOI: 10.20169/j. issn. 2095-2163. 24060701

基于层次聚类的古代玻璃的亚类划分

赵华振, 焦志豪, 杨润林, 薄树奎, 刘 华

(郑州航空工业管理学院 计算机学院, 郑州 450046)

摘要: 古代玻璃制品是人类历史和文化的重要载体, 对其化学成分进行分析可为人们研究其制作工艺、原料来源等提供宝贵信息。研究古代玻璃化学成分对古代玻璃制品的鉴定具有重要意义。本研究收集到一批已被划分为高钾和铅钡两种类型的古代玻璃制品及其各种化学成分的含量, 现对其进行亚类划分。首先建立 CRITIC 权重模型, 通过计算方差、冲突性和信息量, 确定权重占比最高的几种化学成分, 然后建立层次聚类模型针对这几种化学成分进行层次聚类, 得到亚类划分结果。依据亚类划分结果, 将未知玻璃类型的化学成分加入样本中, 再次进行层次聚类以确定其类型。CRITIC 权重模型是一种客观赋权法, 可以避免实验中实验者的主观因素对实验结果的干扰。基于层次聚类得到的亚类划分结果, 可生成多种分类方案, 为文物工作者提供了多种分类选择。

关键词: 古代玻璃; 亚类划分; CRITIC 权重法; 层次聚类

中图分类号: TP391.41

文献标志码: A

文章编号: 2095-2163(2026)04-0091-05

Subclass division of ancient glass based on hierarchical clustering

ZHAO Huazhen, JIAO Zhihao, YANG Runlin, BO Shukui, LIU Hua

(School of Computer Science, Zhengzhou University of Aeronautics, Zhengzhou 450046, China)

Abstract: Ancient glassware is an important carrier of human history and culture, and the analysis of its chemical composition can provide valuable information for us to understand its production process and raw material source. It is important to study the chemical composition of ancient glass for the identification of ancient glass products. In this study, a group of ancient glass products divided into two types of high potassium and lead barium and their contents of various chemical components were collected, and now they are subdivided. First, a CRITIC weight model is established to calculate variance, conflict and information content, and then several chemical components with the largest weight proportion are determined. Then a hierarchical clustering model is established to perform hierarchical clustering for these chemical components, and the result of sub-classification can be obtained. According to the result of subclass classification, the chemical composition of the unknown glass type is added, and the type of the unknown glass type can be obtained by hierarchical clustering again. CRITIC weight model is an objective weighting method, focusing on objectivity, so as to avoid the interference of experimental results caused by subjective factors of experimenters in the experiment. Based on hierarchical clustering, there can be a variety of subclass classification results, which provides a variety of classification options for historical relic workers.

Key words: ancient glass; subclass division; CRITIC weight method; hierarchical clustering

0 引言

在古代玻璃制品化学成分分析的相关研究方面已有不少成果, 多数研究依赖化学分析方法鉴别玻璃类型, 通过定量分析化学成分, 结合 K-means 聚类、随机森林模型实现种类的划分。成分数据分析方法可对玻璃制品的化学成分进行分析, 通过研究

其分类规律, 为未知玻璃文物所属类型^[1-6]提供依据。李公全^[7]曾指出针对古代玻璃的成分分析与鉴别的问题, 首先对古代玻璃制品的成分数据进行清洗和统计分析, 在此基础上建立随机森林模型分析高钾玻璃和铅钡玻璃的分类规律, 并通过聚类分析确定两类玻璃的亚分类。随机森林算法对于数据的预测已有多方面的应用, 在研究古代玻璃方面也

作者简介: 赵华振(2001—), 男, 硕士研究生, 主要研究方向: 目标检测; 焦志豪(2004—), 男, 本科生, 主要研究方向: 目标检测; 杨润林(2005—), 男, 本科生, 主要研究方向: 目标检测; 薄树奎(1976—), 男, 博士, 教授, 主要研究方向: 图像处理, 模式识别。

通信作者: 刘 华(1980—), 男, 硕士, 讲师, 主要研究方向: 项目管理。Email: calflyok@163.com。

收稿日期: 2024-06-07

被广泛应用^[8-11]。为解决古代玻璃种类不易鉴别与分析的问题,可以对传统的 K-means 算法进行优化,在古代玻璃的亚类划分中可以取得很好的分类效果^[12-14]。陈志豪等^[15]曾建立层次聚类模型对不同种类的玻璃制品进行亚类划分,并对分类标准进行合理性和灵敏性检验。层次聚类^[16]是一种常用的聚类分析方法,在需要对多组数据进行聚类的时候往往展示出较高的准确性。对于聚类划分有多种不同的方式,包括线性回归^[17]、AHP - CRITIC 法^[18]、广义 Shapley 函数^[19]等多种方法,都可以展示出很好的分类结果^[20]。

本文勇夺 CRITIC (Criteria Importance Through Intercriteria Correlation) 权重法确定主要化学成分,进而依据这几种化学成分来建立层次聚类模型,相较于同类问题的其它研究方法,CRITIC 权重法可最大限度地克服人为因素造成的实验误差;同时,采用层次聚类对古代玻璃进行亚类划分,为文物工作者在不同研究场景下提供多元化的分类研究提供了新的思路。

1 数据集与预处理

1.1 数据来源

本文数据集来源于 2022 年全国大学生数学建模比赛 C 题附件,其中包括 58 种已风化和未风化的玻璃和玻璃检测点,化学成分的具体含量见表 1。

表 1 化学成分数据集

Table 1 Chemical Composition Dataset wt%

文物编号	二氧化硅 (SiO ₂)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化铅 (PbO)	氧化钡 (BaO)
01	69.33	9.99	6.32	0	0
02	36.28	1.05	2.34	47.43	0
03	87.05	5.19	2.01	0.25	0
04	65.88	9.67	7.12	0	0
05	61.58	0	7.35	0	0
06	67.65	0	0	0.20	1.38
07	92.63	0	1.07	0	0
08	20.14	0	1.48	28.68	31.23
51	24.61	0	3.58	40.24	8.94
52	25.74	1.22	2.27	47.42	8.64
53	63.66	3.04	0.78	13.66	8.99
54	22.28	0	3.19	55.46	7.04
55	49.01	2.71	1.13	32.92	7.95
56	29.15	0	1.21	41.25	15.45
57	25.42	0	1.31	45.10	17.30
58	30.39	0	3.49	39.35	7.66

1.2 数据预处理

首先检查数据中是否存在缺失值,对于关键特

征的缺失,采用合适的方法进行填补。受限于当时的检测手段,未检测出的化学成分数据在本研究中视为含量为 0(即该化学成分不存在)。本研究将基于这些数据开展实验。

2 研究方法

2.1 特征选择与聚类方法

古代玻璃数据包含 14 个化学成分特征,而有些特征在分类辨别中的作用较小,因此需要进行特征选择。CRITIC 权重法,是一种确定各评价指标权重的方法,是一种客观评价法。可以最大程度的减小由主观因素造成的误差,该方法考虑了各评价指标之间的相关性,并通过计算确定具体步骤,如图 1 所示。

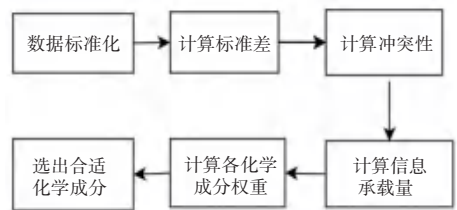


图 1 选择合适的化学成分流程图

Fig. 1 Flowchart for selecting the right chemical composition

层次聚类无需事先指定数量,而是生成类间的层次嵌套结构,通常以树状图的形式呈现。层次聚类有两种基本方法:凝聚和分裂。在本文中选择使用的是凝聚层次聚类,即自底向上的聚类。具体步骤如图 2 所示。

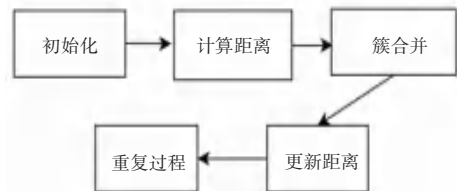


图 2 层次聚类流程图

Fig. 2 Flowchart of hierarchical clustering

2.2 建立 CRITIC 权重模型

在 CRITIC 权重模型中,标准差是衡量准则评分变异性的一种统计量,反映的是评分的一种离散程度,较高的标准差意味着区分度较高,可能需要较高的权重来反映其重要性,这对于模型的准确性和可靠性至关重要。冲突性是指不同准则之间的潜在矛盾与不一致性。其分析可以帮助决策者识别和解决潜在的矛盾,从而提升决策的质量与一致性。权重用于反映某个准则在决策过程中的相对重要性,这也是 CRITIC 权重模型的核心,通过计算得到的

合理权重可帮助决策者在决策过程中更加客观和公正。在本研究中权重可以帮助得到对古代玻璃亚类划分结果有重要影响的几种化学成分。

(1) 数据标准化

设标准化后的数据矩阵元素为 x'_{ij} , 由统计数据可知, 所有数据均为正向指标(越大越优型指标), 标准化公式为:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (1)$$

其中, x'_{ij} 为第 i 个样本第 j 个指标的标准化值; x_{ij} 为第 i 个样本第 j 个指标的原始值; $\max(x_j)$ 为第 j 个指标的原始最大值; $\min(x_j)$ 为第 j 个指标的原始最小值。

采用 CRITIC 权重法对评价指标赋权前, 需要计算标准差与冲突性。

(2) 标准差

用标准差 σ_j 表示 j 项指标的对比性:

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^m (x'_{ij} - \bar{x}_j)^2}{m - 1}} \quad (2)$$

其中, \bar{x}_j 是数据均值。

(3) 冲突性

冲突性反映的是不同指标之间的相关程度, 若呈现显著正相关性, 则冲突性数值越小。设指标 j 与其余指标冲突性大小为 f_j :

$$f_j = \sum_{i=1}^m (1 - r_{ij}) \quad (3)$$

其中, r_{ij} 表示第 i 个指标与第 j 个指标之间的相关系数, 在此使用的是皮尔逊相关系数, 为线性相关系数。

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X] \text{Var}[Y]}} \quad (4)$$

其中, $\text{Cov}(X, Y)$ 为指标 X 与 Y 的协方差。

(4) 计算信息承载量

$$C_j = \sigma_j f_j \quad (5)$$

(5) 计算各化学成分权重

$$w_j = \frac{C_j}{\sum_{j=1}^n C_j} \quad (6)$$

(6) 结果

计算所得各化学成分的标准差、冲突性、信息量和权重, 并按照权重降序排序(见表2)。

按照权重排序, 选出前5个化学成分, 分别为: 二氧化硅(SiO_2)、氧化铅(PbO)、氧化钾(K_2O)、氧

化钡(BaO)和氧化钙(CaO), 则依据这化学成分的含量进行亚类划分。

表2 各化学成分的标准差、冲突性、信息量和权重(降序排列)

Table 2 Standard deviation, conflict, information content and weight of each chemical component (descending order)

	标准差	冲突性	信息量	权重
二氧化硅(SiO_2)	0.261 3	15.536	4.060 1	0.098 4
氧化铅(PbO)	0.277 9	14.107	3.920 6	0.095 0
氧化钾(K_2O)	0.267 1	12.361	3.302 5	0.080 0
氧化钡(BaO)	0.237 6	13.296	3.159 8	0.076 6
氧化钙(CaO)	0.267 2	11.688	3.123 3	0.075 7
氧化锡(SnO_2)	0.142 9	13.065	1.867 3	0.045 2

2.3 建立层次聚类模型

层次聚类是一种直观且易于理解的聚类方法, 基本思想是将最相似的个体组合在一起, 并且这个过程是逐步的, 递归进行的。具体过程如下:

(1) 初始化。首先对玻璃文物进行重新编号, 将每个样本都看作是一个单独的簇。

(2) 计算距离。计算簇与簇之间的距离。本研究中主要使用的是中心链接的方法来计算, 即计算两个簇中5种化学成分含量差异的平均值来实现的。

(3) 簇合并。将距离最近的簇合并成一个新的簇。本研究中, 这里的“最近”指的是两个簇的中心点之间的距离, 即5种化学成分差值的平均数, 平均数相差最小的两个簇合并为一个簇。

(4) 更新距离。重新计算新生成的簇与其他所有簇之间的距离。即将其中的5种化学成分的值更新为得到的新簇内样本的化学成分的平均值。

(5) 重复过程。重复步骤3和步骤4, 直到所有的数据点都被合并成为一个簇结束。

利用 MATLAB 进行层次聚类分析, 得出文物所属亚类的层次聚类图, 如图3所示。

最后, 可以根据实验需要, 决定簇的数量, 也就是决定要分成几个亚类。比如, 需要分成5个亚类, 那么就需要在层次聚类图中找到一个层次, 这个层次有5个簇, 然后这5个簇就是层次聚类的最终结果。

3 实验结果与分析

(1) 聚类结果可靠性

由表2中二氧化硅(SiO_2)、氧化钾(K_2O)、氧化钙(CaO)、氧化铅(PbO)和氧化钡(BaO)含量规律, 进而得出这类的划分标准, 见表3。

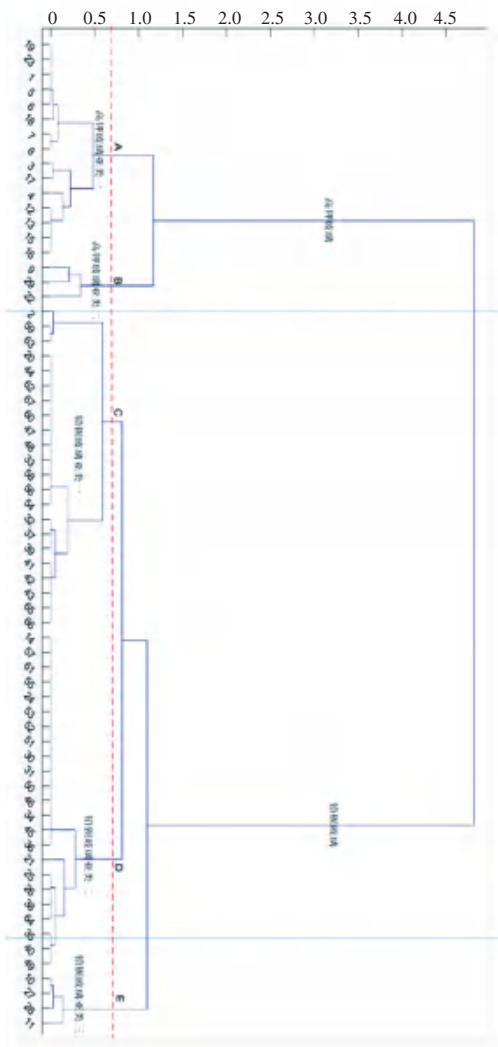


图3 玻璃文物层次聚类图

Fig. 3 Hierarchical clustering diagram of glass cultural relics

表3 分类标准

Table 3 Classification criteria

类别	分类依据(化学成分组合)
高钾硅类	K_2O-SiO_2
高钾硅钙类	$K_2O-CaO-SiO_2$
铅钡硅类	$PbO-BaO-SiO_2$
铅硅类	$PbO-SiO_2$
铅钡钙硅类	$PbO-BaO-CaO-SiO_2$

将图3得出的结果与实验收集到的文物所属类别进行对照,结果显示所有分类均与实际数据完全一致,据此可认定本次亚类分析具有合理性。

(2) 对未知玻璃类型进行分类

继续使用 MATLAB 对 1~75 号文物进行层次聚类,其中 68~75 号是未知类型的玻璃文物(即 A1~A8),得到层次聚类图后,通过查找 68~75 号文物的聚类位置,即可确定这 8 个未知玻璃文物的所属类型。层次聚类图的结果如图 4 所示。

由图 4 可以得到这 8 个未知类别的玻璃文物的所属类型,总结结果见表 4。

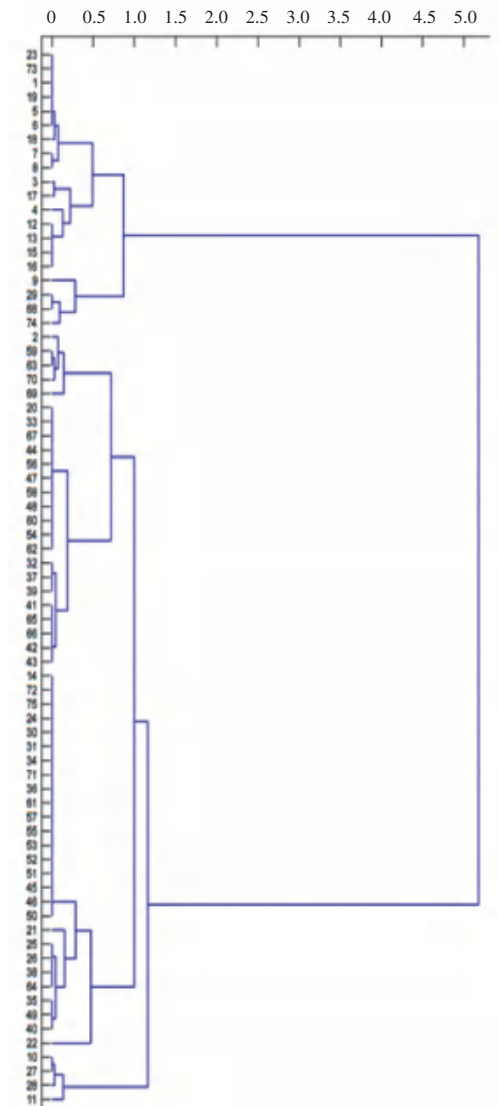


图4 层次聚类结果图

Fig. 4 Result of hierarchical clustering

表4 未知类别玻璃文物的所属类型

Table 4 Types of glass artifacts belonging to unknown categories

编号	文物编号	所属类型	所属亚类
68	A1	高钾	$K_2O-CaO-SiO_2$
69	A2	铅钡	$PbO-BaO-SiO_2$
70	A3	铅钡	$PbO-BaO-SiO_2$
71	A4	铅钡	$PbO-SiO_2$
72	A5	铅钡	$PbO-SiO_2$
73	A6	高钾	K_2O-SiO_2
74	A7	高钾	$K_2O-CaO-SiO_2$
75	A8	铅钡	$PbO-SiO_2$

4 结束语

本研究的主要目的是在已知古代玻璃被划分为高钾和铅钡两种类型前提下,实现古代玻璃的亚类划分并对未知类别的玻璃进行分类。首先,使用CRITIC权重法来选取5种占比最大的化学成分,尽可能的避免了人为因素造成的实验误差。然后,使用层次聚类中凝聚的层次聚类的方式进行聚类分析,可以得到多种亚类划分结果,在本研究中,最终确定为5个亚类,经过与原数据对比可得此次分类结果完全合理。最后运用层次聚类对未知类别的玻璃进行亚类划分,可得到未知玻璃类型的分类结果以及亚类划分结果。

古代玻璃制品作为文化遗产的重要组成部分,承载着丰富的历史信息,对研究古代社会、科技、文化等方面具有重要意义。通过对古代玻璃制品进行分类分析,可深入了解其原料来源、生产工艺、与年代特征,为文物保护和修复提供科学依据。而层次聚类可依据实际需要得到多种亚类划分方案,可供文物工作者基于不同的场景进行选择。

参考文献

- [1] 王保乾, 蒋剑军. 基于成分数据分析与模糊模式识别的古代玻璃种类鉴别[J]. 科技创新与应用, 2024, 14(7): 41-46. DOI: 10.19981/j.cn23-1581/G3. 2024. 07. 010.
- [2] 解慧, 张建国, 郭磊, 等. 古代玻璃制品属性及化学成分数学建模的研究[J]. 德州学院学报, 2023, 39(6): 6-14.
- [3] 马佩莹, 韩雁来, 李德兰, 等. 基于成分数据的古代玻璃制品分析与分类[J]. 数学建模及其应用, 2023, 12(2): 63-73. DOI: 10.19943/j.2095-3070.jmmia.2023.02.06.
- [4] 杨磊, 魏彤, 宗毅恒. 古代玻璃制品的成分分析与鉴别问题的数学模型[J]. 北京印刷学院学报, 2023, 31(9): 61-66. DOI: 10.19461/j.cnki.1004-8626.2023.09.010.
- [5] 宛惠, 邓明华. 古代玻璃制品成分分析与鉴别的统计建模[J]. 数学建模及其应用, 2023, 12(2): 27-40. DOI: 10.19943/j.2095-3070.jmmia.2023.02.03.
- [6] 楼阳, 窦雷, 卓朝阳, 等. 古代玻璃成分分析与亚类划分方法研究[J]. 数学建模及其应用, 2023, 12(4): 73-83. DOI: 10.19943/j.2095-3070.jmmia.2023.04.10.
- [7] 李公全, 林旭旭, 彭梅, 等. 基于随机森林和支持向量机的古代玻璃制品成分分析与鉴别[J]. 科技创新与应用, 2023, 13(34): 109-113. DOI: 10.19981/j.cn23-1581/G3. 2023. 34. 025.
- [8] 曹晋云, 许天宇, 刘勇, 等. 基于RBF和SVM的古代玻璃制品成分预测与分类[J]. 科技创新与应用, 2023, 13(18): 41-43. DOI: 10.19981/j.cn23-1581/G3. 2023. 18. 009.
- [9] 路佳佳. 基于集成特征选择和随机森林的古代玻璃分类模型[J]. 硅酸盐学报, 2023, 51(4): 1060-1065. DOI: 10.14062/j.issn.0454-5648.20220790.
- [10] 张倩, 吕坤山, 赵先庆, 等. 基于随机森林分类模型的古代玻璃制品成分分析[J]. 河南科技, 2023, 42(9): 19-22. DOI: 10.19968/j.cnki.hnkj.1003-5168.2023.09.004.
- [11] 吕飞, 付航玮, 刘成林. 基于机器学习的古代玻璃制品的成分分析与鉴别[J]. 信息与电脑(理论版), 2023, 35(4): 98-102.
- [12] 邵光明, 夏贤齐, 殷何杰. 基于CART和聚类分析的古代玻璃分类预测模型研究[J]. 通化师范学院学报, 2024, 45(2): 31-35. DOI: 10.13877/j.cnki.cn22-1284.2024.02.005.
- [13] 陈小君, 叶子, 石怀旺. 基于K-means聚类与SVM算法对古代玻璃文物的分类与鉴别[J]. 哈尔滨师范大学(自然科学学报), 2023, 39(4): 70-79.
- [14] 康澜. 基于改进GDBT与K-means聚类的玻璃文物亚类划分研究[J]. 信息技术与信息化, 2023(7): 149-152.
- [15] 陈志豪, 季晶敏. 基于层次聚类模型的古代玻璃制品成分分析与鉴定[J]. 现代信息技术, 2023, 7(8): 122-125. DOI: 10.19850/j.cnki.2096-4706.2023.08.031.
- [16] 陈斌, 谢文波, 付勋, 等. 基于改进局部密度的可扩展层次聚类算法[J]. 南京大学学报(自然科学), 2024, 60(3): 370-382. DOI: 10.13232/j.cnki.jnju.2024.03.002.
- [17] 钱鸿羽, 岳思情, 许雷, 等. 基于多元线性回归的古代玻璃体系判别方法[J]. 甘肃科技, 2024, 40(2): 56-61. DOI: 10.20156/j.cnki.2097-2490.2024.02.012.
- [18] 王春燕, 史可, 王梦珂, 等. 基于AHP-CRITIC法醋酸青皮有效成分与颜色值的相关性研究[J]. 中华中医药学刊, 2023, 41(11): 208-214. DOI: 10.13193/j.issn.1673-7717.2023.11.039.
- [19] CAI N N, YIN Y Y, HAN Q. Prediction and classification of chemical composition of ancient glass objects based on generalized Shapley functions[J]. Frontiers in Chemistry, 2024, 12: 13511.
- [20] YANG X, TIAN P, WENG J. A study of ancient glass subclassification based on K-means algorithm[J]. Advances in Computer, Signals and Systems, 2024, 8(2): 45-52.