

基于系统聚类算法和Fisher判别模型的古代玻璃成分分析与类别鉴定

李明宇, 李汶蓉, 王谋新, 杨文龙

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2022年10月31日; 录用日期: 2022年12月31日; 发布日期: 2023年1月4日

摘要

古代玻璃体现了我国古代高度发达的工艺技术, 新时代, 玻璃成为了人们生产生活的必需品, 研究古代玻璃有助于加深人们对玻璃的认识, 唤醒我国人民对工艺研究的热情。本文对古代玻璃的实验数据进行分析, 首先使用描述性统计分析和卡方检验分析研究玻璃文物样品表面有无风化化学成分含量的统计规律, 得到玻璃表面风化与类型的关系; 然后依据玻璃表面风化与类型的关系建立系统聚类算法模型, 研究不同类别玻璃文物样品的化学成分之间的关联关系; 最后依据化学成分之间关联关系建立Fisher判别模型, 得到玻璃类别鉴定的Fisher判别公式, 可以对未知类别的玻璃文物进行类别鉴定。本文对古代玻璃的成分分析与类别鉴定提供一定的借鉴意义。

关键词

古代玻璃, 系统聚类算法, Fisher判别模型

Composition Analysis and Category Identification of Ancient Glass Based on Systematic Clustering Algorithm and Fisher Discriminant Model

Mingyu Li, Wenrong Li, Mouxin Wang, Wenlong Yang

School of Opto-Electronic and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Oct. 31st, 2022; accepted: Dec. 31st, 2022; published: Jan. 4th, 2023

文章引用: 李明宇, 李汶蓉, 王谋新, 杨文龙. 基于系统聚类算法和 Fisher 判别模型的古代玻璃成分分析与类别鉴定[J]. 建模与仿真, 2023, 12(1): 1-15. DOI: 10.12677/mos.2023.121001

Abstract

The ancient glass reflected the highly developed technology of Chinese ancient times. In the new era, glass has become a necessity of people's production and life. The study of ancient glass is helpful to deepen people's understanding of glass and awaken the enthusiasm of Chinese people for the research of craft. In this paper, the experimental data of ancient glass are analyzed. Firstly, descriptive statistical analysis and Chi-square test were used to study the statistical rule of weathering chemical components on the surface of glass cultural relics samples, and the relationship between weathering and types of glass surface was obtained. Then, according to the relationship between glass surface weathering and glass types, a systematic clustering algorithm model was established to study the correlation between the chemical components of different types of glass cultural relics samples. Finally, the *Fisher* discriminant model was established according to the correlation between chemical components, and the *Fisher* discriminant formula of glass category identification was obtained, which can be used for category identification of glass relics of unknown category. This paper provides some reference for the composition analysis and category identification of ancient glass.

Keywords

Ancient Glass, Systematic Clustering Algorithm, *Fisher* Discriminant Model

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

古代玻璃是中西方通过丝绸之路进行文化交流的宝贵物证。古代玻璃易受环境的影响而风化，在风化过程中，由于内部元素与环境元素进行大量交换，导致其化学成分比例发生变化，影响对玻璃类别的判断，从而影响对其产地、原料的来源以及制作工艺的考古研究[1]。因此，对古代玻璃的化学成分分析和类别鉴定具有重要研究意义。

早期研究指出，古代玻璃的成分分析与类别鉴定易受环境的影响。文献[2]研究了古代玻璃的化学元素组成；文献[3] [4]研究了古代玻璃的化学成分组成。文献[5]研究了古代玻璃经历风化后的化学成分衰变。文献[6]研究了古代玻璃的类别鉴定。这些研究表明，研究古代玻璃的化学成分分析和类别鉴定是十分必要的。

本文在前人研究的基础上，对高教社杯数学建模 C 题中提供的实验数据[1]进行建模，使用描述性统计分析和卡方检验分析研究玻璃文物样品表面有无风化化学成分含量的统计规律，得到玻璃表面风化与类型的关系；再依据玻璃表面风化与类型的关系的结论，使用系统聚类算法模型研究不同类别玻璃文物样品的化学成分之间的关联关系；最后依据关联关系的结论使用 *Fisher* 判别模型，得到玻璃类别鉴定的 *Fisher* 判别公式，对未知类别的玻璃文物进行类别鉴定。

2. 玻璃的表面风化与类型的关系

由于实验数据量较大，无法直接得出玻璃的表面风化与类型是否存在相关性。在无明显的相关性时，

直接使用卡方检验分析可能会导致结论误差较大，所以先进行玻璃表面有无风化的描述性统计分析来初步确定是否存在相关性，再通过卡方检验分析验证相关性存在并且研究相关性的强弱。

2.1. 数据预处理

实验数据[1]给出了这些文物的分类信息和相应的主要成分所占比例。因检测手段等原因可能导致其成分比例的累加和非 100%的情况。根据实验要求有效数据是成分比例累加和介于 85%~105%之间，为了方便之后数据的使用，首先对数据进行数据预处理。附件表单 2 [1]中 15 号文物采样点的化学成分比例累加和为 79.47，17 号文物采样点的化学成分比例累加和为 71.89，在之后的计算过程中都去除无效数据 15 号和 17 号，未检测到数据的空白处记作 0，其余实验数据可以正常使用。

2.2. 描述性统计分析

为了结合玻璃类型，分析文物样品有无风化化学成分含量的统计规律，进而分析出玻璃的表面风化与类型的关系，使用实验数据[1]，划分高钾有无风化和铅钡有无风化进行描述性统计分析，分别求出 14 种化学成分含量的平均值、标准差、方差、峰度和偏度。在玻璃类型固定的情况下，在某一种化学成分下，比较其在风化和无风化下的成分含量统计规律。

2.2.1. 峰度系数

峰度系数(β_k)用来反映峰部的尖度程度。若 $\beta_k > 0$ ，说明该分布相比与正态分布顶端更加尖锐或者尾部更加粗；若 $\beta_k < 0$ ，则说明说明该分布相比与正态分布顶端更加平坦或者尾部更加细；若 $\beta_k = 0$ ，说明该分布与标准正态分布顶端尖锐程度和尾部粗细程度相当。所以峰度可以作为衡量偏离正态分布形状的尺度[7]。峰度的计算公式：

$$\beta_k = \frac{v_4}{v_2^2} = \frac{E(X - E(X))^4}{|Var(X)|^2} - 3$$

其中， $E(X)$ 为均值， $Var(X)$ 为方差， v_n 表示 n 阶中心矩。

2.2.2. 偏度系数

偏度系数(β_s)常用于衡量样本数据分布偏离对称性的程度。若 $\beta_s > 0$ ，则认为处于均值右边的数据偏多；若 $\beta_s < 0$ ，则认为处于均值左边的数据较多；当 β_s 接近与 0 时，认为样本数据是对称分布的[8]。偏度的计算公式：

$$\beta_s = \frac{v_3}{\frac{v_2^3}{v_2^2}} = \frac{E(X - E(X))^3}{Var(X)^{\frac{3}{2}}}$$

其中， $E(X)$ 为均值， $Var(X)$ 为方差， v_n 表示 n 阶中心矩。

2.2.3. 描述性统计分析求解

利用 SPSS 分别求解高钾玻璃和铅钡玻璃表面有无风化化学成分含量的统计规律。

由表 1 知，在高钾玻璃条件下，通过分析 14 种化学成分含量的平均值、标准差、方差、峰度和偏度，我们发现：二氧化硅在风化的情况下含量更高；五氧化二磷、氧化钾、氧化铜、氧化铁、氧化钙、氧化镁和氧化铝在无风化的情况下含量更高；二氧化硫、氧化锡和氧化钠在风化的情况下不存在，在无风化的情况下，存在含量也较少；氧化锶、氧化钡和氧化铅在两种情况下含量都较小，而且含量变化不明显。

Table 1. The statistics of high potassium glass surface weathering chemical composition content
表 1. 高钾玻璃表面有无风化化学成分含量的统计

变量名	表面风化	平均值	标准差	方差	峰度	偏度
二氧化硅	风化	88.559	12.226	149.473	5.618	-2.326
	无风化	70.067	10.131	102.644	1.655	1.372
氧化锶	风化	0.014	0.038	0.001	7	2.646
	无风化	0.04	0.048	0.002	-1.059	0.723
二氧化硫	风化	0	0	0	0	0
	无风化	0.122	0.198	0.039	-0.866	1.108
氧化锡	风化	0	0	0	0	0
	无风化	0.236	0.746	0.557	10	3.162
氧化钠	风化	0	0	0	0	0
	无风化	0.548	1.194	1.426	3.268	2.051
五氧化二磷	风化	0.347	0.258	0.067	-1.062	0.304
	无风化	1.481	1.568	2.458	0.895	1.456
氧化钾	风化	2.974	4.487	20.132	3.795	1.997
	无风化	8.188	4.789	22.938	0.205	-0.941
氧化铜	风化	1.714	1.552	2.408	5.305	2.21
	无风化	2.207	1.308	1.71	-0.913	-0.577
氧化钡	风化	0.409	1.081	1.169	7	2.646
	无风化	0.432	0.735	0.54	0.698	1.429
氧化铅	风化	0.237	0.526	0.276	6.276	2.487
	无风化	0.328	0.55	0.303	2.934	1.895
氧化铁	风化	0.511	0.736	0.542	6.476	2.503
	无风化	1.831	1.791	3.208	2.877	1.438
氧化钙	风化	1.719	1.933	3.735	4.805	2.112
	无风化	4.848	3.306	10.928	-1.329	-0.662
氧化镁	风化	0.327	0.444	0.197	-0.212	1.011
	无风化	1.184	0.646	0.417	-0.716	-0.6
氧化铝	风化	2.737	1.701	2.895	-0.825	0.557
	无风化	6.57	3.014	9.085	-0.946	-0.005

由表 2 知, 在铅钡玻璃条件下, 通过分析 14 种化学成分含量的平均值、标准差、方差、峰度和偏度, 可知, 五氧化二磷、氧化铜、氧化钡、氧化铅和氧化钙在风化的情况下含量更高; 二氧化硅、二氧化硫、氧化钠和氧化铝在无风化的情况下含量更高; 氧化锶、氧化锡、氧化钾、氧化铁和氧化镁在两种情况下含量都较小, 而且含量变化不明显。

Table 2. The statistics of lead barium glass surface weathering chemical composition content
表 2. 铅钡玻璃表面有无风化化学成分含量的统计

变量名	表面风化	平均值	标准差	方差	峰度	偏度
二氧化硅	风化	25.897	10.195	103.935	1.638	0.494
	无风化	55.692	10.995	120.881	-0.405	-0.335
氧化锶	风化	0.391	0.234	0.055	-0.365	-0.03
	无风化	0.239	0.204	0.042	2.538	1.047
二氧化硫	风化	0.891	3.348	11.211	20.997	4.518
	无风化	0.166	0.78	0.609	22	4.69
氧化锡	风化	0.077	0.286	0.082	17.379	4.094
	无风化	0.049	0.13	0.017	5.396	2.558
氧化钠	风化	0.244	0.587	0.345	5.566	2.457
	无风化	1.759	2.398	5.752	0.585	1.222
五氧化二磷	风化	4.632	3.893	15.153	-0.86	0.487
	无风化	1.09	1.88	3.533	3.396	2.097
氧化钾	风化	0.136	0.249	0.062	7.758	2.544
	无风化	0.229	0.314	0.098	9.773	2.845
氧化铜	风化	2.242	2.989	8.932	3.687	2.065
	无风化	1.112	1.267	1.606	2.057	1.642
氧化钡	风化	11.683	9.483	89.921	1.66	1.465
	无风化	8.219	4.558	20.773	5.179	1.692
氧化铅	风化	43.3	12.388	153.469	0.422	-0.088
	无风化	21.764	8.26	68.224	-0.272	0.737
氧化铁	风化	0.623	0.762	0.58	1.204	1.332
	无风化	0.77	1.17	1.37	4.495	2.012
氧化钙	风化	2.757	1.66	2.757	-0.536	0.525
	无风化	1.359	1.301	1.693	0.842	1.28
氧化镁	风化	0.652	0.733	0.537	1.146	1.079
	无风化	0.67	0.541	0.293	-1.177	0.002
氧化铝	风化	2.997	2.774	7.693	9.811	2.766
	无风化	4.586	3.277	10.741	4.104	1.967

筛选出铅钡玻璃和高钾玻璃风化前后相对重要的化学成分，利用 *Matlab* 分析其频率分布直方图。

由图 1 可以看出高钾玻璃在风化后主要化学成分含量呈下降趋势；铅钡类玻璃在风化后主要化学成分含量呈上升趋势。

通过玻璃表面有无风化化学成分统计规律的描述性分析和不同玻璃类型在风化前后主要化学成分含量的变化分析知，玻璃的表面风化与类型存在一定的相关性。

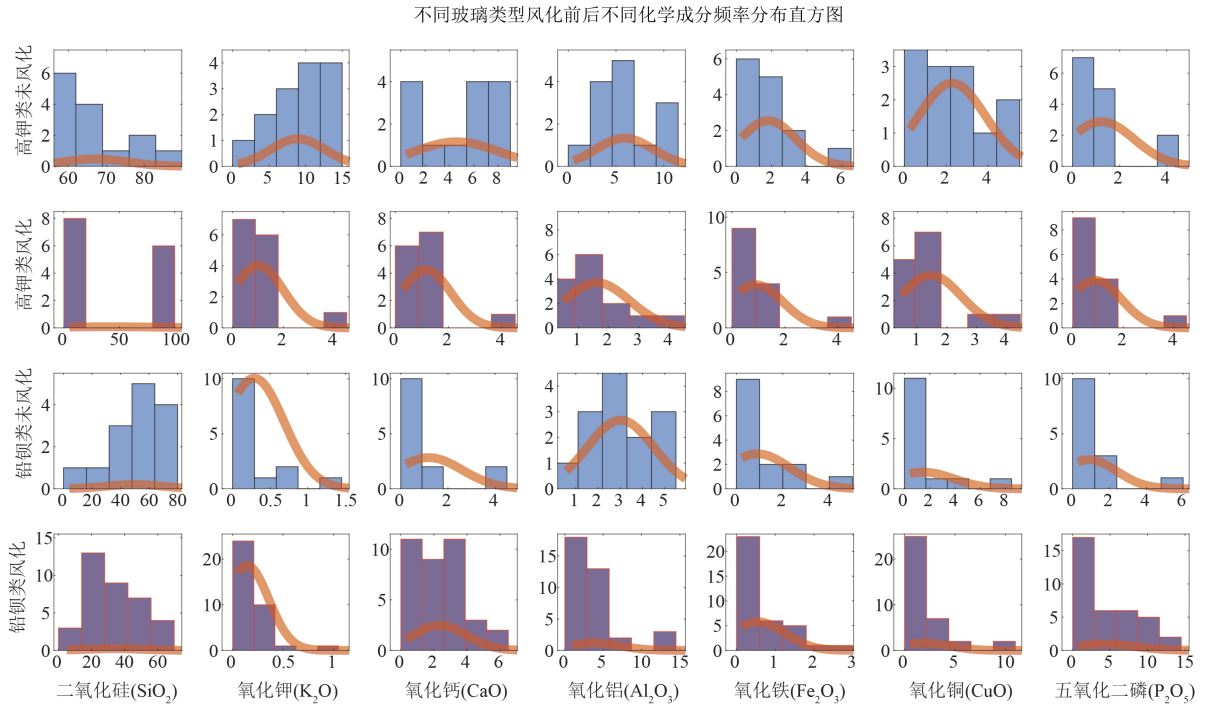


Figure 1. Distribution histograms of chemical composition frequencies before and after weathering for different glass types
图 1. 不同玻璃类型风化前后化学成分频率分布直方图

2.3. 卡方检验分析

由描述性统计分析和频率分布直方图可以看出，玻璃的表面风化与类型具有相关性，为了进一步研究二者的相关性强弱，对玻璃的表面风化与类型进行卡方检验分析。

2.3.1. 卡方检验

为了对这些玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系进行分析，由于此时附件表单 1 [1] 中的纹饰、类型、颜色和表面风化均为定类变量，所以，表面风化与其玻璃类型、纹饰和颜色的关系分析使用卡方检验。

卡方检验主要是比较定类变量与定类变量之间的差异性分析。通过统计样本的实际观测值与理论推断值之间的偏离程度，实际观测值与理论推断值之间的偏离程度就决定卡方值的大小[9]。

卡方检验公式：

$$\chi^2 = \sum \frac{(A - E)^2}{E} = \sum_{i=1}^k \frac{(A_i - E_i)^2}{E_i} = \sum_{i=1}^k \frac{(A_i - np_i)^2}{np_i}$$

其中， A 代表颜色、类别或纹饰的观察频数， E 代表基于假设 H_0 计算出的期望频数， A 和 E 的差称为残差。

提出假设：

H_0 : 表面风化与玻璃类型、纹饰或颜色相互独立不相关。

H_1 : 表面风化与玻璃类型、纹饰或颜色有关联。

2.3.2. 卡方检验求解

使用 SPSSPRO 软件进行交互分析，得出如下表所示的卡方检验分析结果。模型检验的结果包括数据

的频数、卡方值、显著性 P 值。

变量 X : 表面风化; 变量 Y : 颜色, 类型, 纹饰。

表 3 的结果显示: 对于表面风化, 颜色的显著性 P 值为 0.268, 水平上不呈现显著性, 接受原假设; 纹饰的显著性 P 值为 0.085* (10%显著性水平), 水平上不呈现显著性, 接受原假设; 类型的显著性 P 值为 0.024** (5%显著性水平), 水平上呈现显著性, 拒绝原假设。因此, 玻璃的表面风化与类型存在相关性且相关性较强。

Table 3. Chi-square test results

表 3. 卡方检验结果

变量	名称	表面风化		总计	X^2	校正 X^2	P
		风化	无风化				
颜色	无色	4	0	4	9.962	9.962	0.268
	浅绿	1	2	3			
	浅蓝	12	6	18			
	深绿	4	3	7			
	深蓝	0	2	2			
	紫	2	2	4			
	绿	0	1	1			
	蓝绿	9	6	15			
	黑	2	0	2			
类型	铅钡	12	28	40	5.061	3.790	0.024**
	高钾	10	6	16			
纹饰	A	11	11	22	4.941	4.941	0.085*
	B	0	6	6			
	C	11	17	28			

注: **、*、*分别代表 1%、5%、10%的显著性水平。

3. 玻璃文物的化学成风之间的关联关系

由卡方检验分析的结论知, 玻璃的表面风化与类型相关, 且相关性较强。所以当研究玻璃文物的化学成分之间的关联时, 将玻璃文物划分为高钾风化、高钾无风化、铅钡风化和铅钡无风化四大类进行分析时结论更加准确。针对不同类别的玻璃文物样品, 为了研究其化学成分之间的关联关系, 以及关联关系之间的差异性, 采用系统聚类算法模型来分析化学成分之间的关联关系, 选择轮廓系数比较分析高钾玻璃与铅钡玻璃之间的化学成分关联关系的差异性。

3.1. 系统聚类算法模型的建立

系统聚类分析的原理是按照距离远近, 将距离相近的变量先聚成类, 距离较远的变量后聚成类, 依次进行, 直到每个变量都归入合适的类中[10]。

系统聚类过程见图 2。

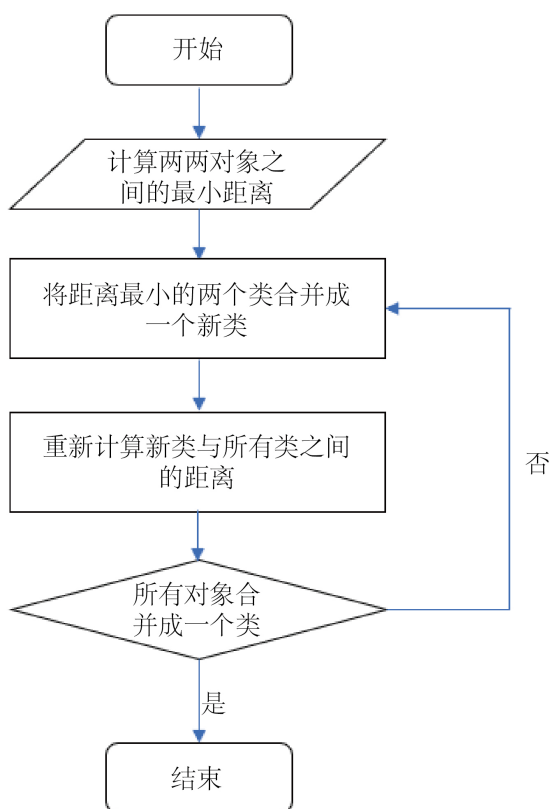


Figure 2. System clustering algorithm
图 2. 系统聚类算法

3.2. 系统聚类算法模型的求解

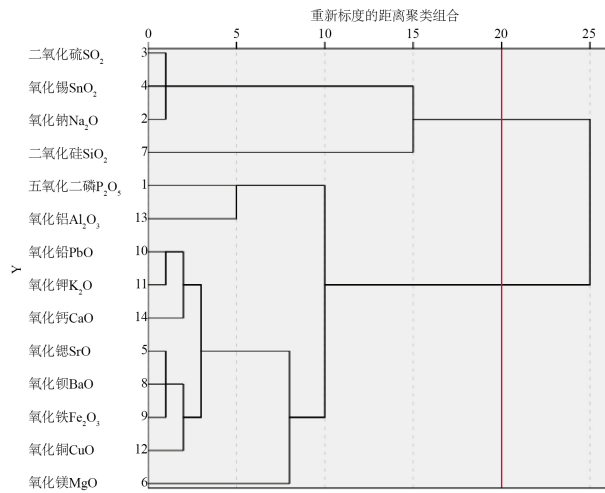
为了得到不同类别玻璃文物的化学成分之间的关联关系，根据系统聚类算法模型的步骤，把高钾风化、高钾无风化、铅钡风化、铅钡无风化四种类别玻璃文物的化学成分数据用 *SPSS* 分析，得到四种类别状态下的谱系图。

由图 3 分析可知，在高钾风化的谱系图 3(a) 中，化学成分主要分为两类，二氧化硫、氧化钙、氧化钠和二氧化硅分为一类，其他十种化学成分分为第二类；在高钾无风化的谱系图 3(b) 中，化学成分主要分为三类，二氧化硅和氧化锡分为第一类，氧化钠、氧化铅、氧化钾和氧化钙分为第二类，其他八种化学成分分为第三类；在铅钡风化的谱系图 3(c) 中，化学成分主要分为四类，氧化铅和氧化镉分为第一类，氧化钡、氧化铜和二氧化硫分为第二类，氧化镁、氧化钙、氧化铁和五氧化二磷分为第三类，其他五种化学成分分为第四类；在铅钡无风化的谱系图 3(d) 中，化学成分主要分为四类，二氧化硅、氧化铝和氧化镁分为第一类，五氧化二磷分为第二类，二氧化硫、氧化钠、氧化铜和氧化钡分为第三类，其他六种化学成分分为第四类。即高钾风化、高钾无风化、铅钡风化和铅钡无风化的玻璃文物的化学成分的最优分类数目分别为 2、3、4、4。

3.3. 玻璃文物的化学成分之间的关联关系的差异性

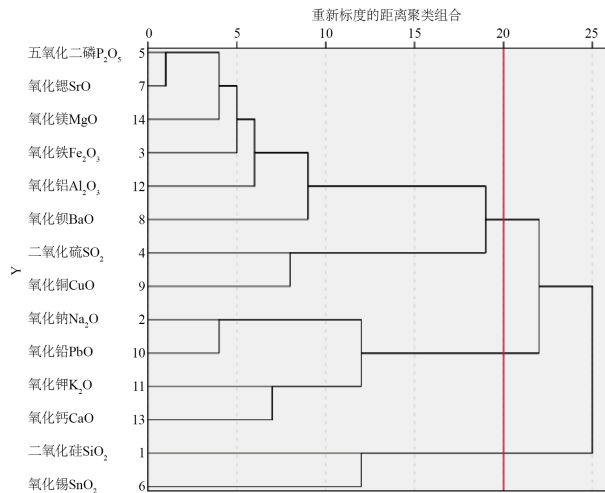
由系统聚类算法模型的研究知，高钾风化、高钾无风化、铅钡风化和铅钡无风化的玻璃文物的化学成分的最优分类数目分别为 2、3、4、4。在系统聚类结果的基础上选择合适的指标进行分类结果差异性的判断，本次选择轮廓系数进行比较分析。

使用平均联接(组间)的谱系图



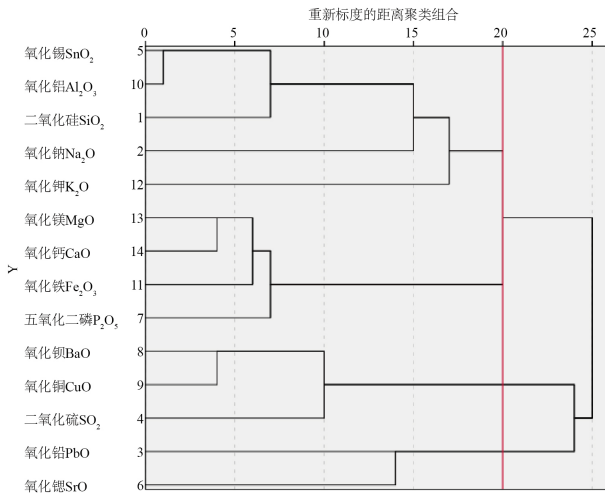
(a)

使用平均联接(组间)的谱系图



(b)

使用平均联接(组间)的谱系图



(c)

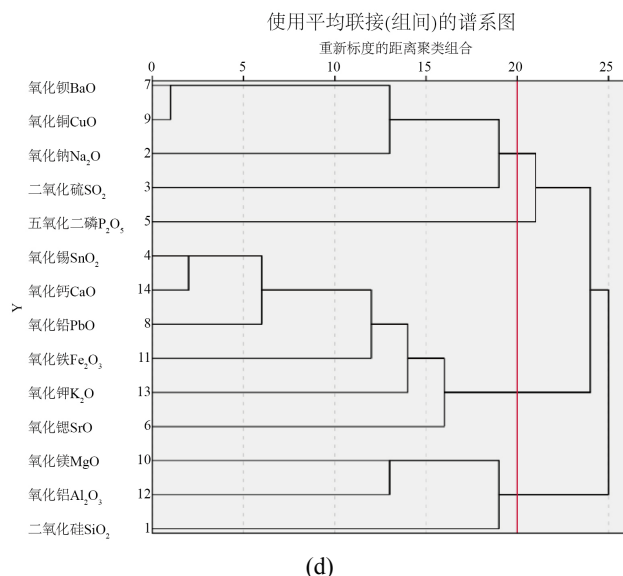


Figure 3. Pedigree map of high potassium glass and lead-barium glass with or without weathering
图 3. 高钾玻璃和铅钡玻璃有无风化的谱系图

3.3.1. 轮廓系数

轮廓系数：轮廓系数评价指标说明，轮廓系数的取值在[-1, 1]之间，越趋近于 1 代表聚类度和分离度都相对较优，即聚类效果越好[11]。

轮廓系数(s_i)计算公式：

$$s_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

其中， $a(i)$ 表示样本 i 到同簇其他样本的平均值，称为簇内不相似度； $b(i)$ 表示样本 i 到相邻簇其他样本的平均值，称为簇间不相似度。

3.3.2. 轮廓系数求解

使用 *Matlab* 求解得出了不同玻璃类型风化前后的轮廓系数指标变化情况：

由表 4 分析可知，高钾玻璃风化后的最优分类数目减少，铅钡玻璃风化后的最优分类数目不变。对于轮廓系数评价指标，高钾玻璃风化后更远离+1，而铅钡玻璃风化后更接近于+1，说明铅钡玻璃风化后化学成分之间的关联关系加强，而高钾玻璃风化后化学成分之间的关联关系减弱。

Table 4. Analysis table of evaluation index of system clustering

表 4. 系统聚类评价指标分析表

类别	最优分类数目	轮廓系数
高钾风化	2	0.78836
高钾无风化	3	0.83888
铅钡风化	4	0.93844
铅钡无风化	4	0.88246

4. 玻璃文物的类别鉴定

玻璃文物易受埋藏环境的影响而风化。能从玻璃文物的表面明显看出是否风化，在部分风化的文物中，其表面也有未风化的区域[1]。实验中已经测出玻璃文物的化学成分比例数据，能从玻璃表面观察到是否风化，但是玻璃的类型未知。由不同玻璃类别之间的化学成分关联关系的研究中，得出铅钡玻璃风化后化学成分关联关系加强，而高钾玻璃风化后化学成分关联关系减弱。所以为了减小有无风化对玻璃类型鉴定的影响，在建立玻璃的类别鉴定模型时，选用 *Fisher* 线性判别模型，将表面风化与玻璃类别用数字来表示：未风化计为 0，风化计为 1；高钾玻璃计为 0，铅钡玻璃计为 1。

使用经过预处理的实验数据，先建立玻璃类别的 *Fisher* 线性判别模型，利用回代分析模型的准确率。通过计算出高钾玻璃与铅钡玻璃各个化学成分的含量，选择出含量差别比较大的元素作为修改数据，对分类结果的合理性和敏感性进行分析。

4.1. Fisher 判别模型的建立

Fisher 判别的基本思想是投影，即将 K 组 p 维数据投影到某一个方向，使得组间距离最大，然后借助一元方差分析的思想构造一个线性判别函数，其系数确定原则是根据类间距离最大、类内距离最小的原则确定线性判别函数，再根据建立的线性判别方程结合合适的判别规则来判断待判样品的类别。经判别方程划分后，同类样品在空间上的分布集中，而不同类之间距离较远，差别明显。由于此时玻璃类别只有高钾玻璃与铅钡玻璃两种，所以仅考虑两总体的情况。

设有 p 维总体为 X_1, X_2 ，它们都有二阶矩存在。此时在 p 维的情况下， x 的线性组合为 $y = a^T x$ ，其中 a 为 p 维实向量。又有 X_1, X_2 的均值向量分别为 μ_1, μ_2 ，它们也都是 p 维，都是公共的协方差矩阵 $\Sigma (\Sigma > 0)$ 。

那么线性组合 $y = a^T x$ 的均值为：

$$\begin{cases} \mu_{y1} = E(y | y = a^T x, x \in X_1) = a^T \mu_1 \\ \mu_{y2} = E(y | y = a^T x, x \in X_2) = a^T \mu_2 \end{cases},$$

方差为：

$$\sigma_y^2 = \text{Var}(y) = a^T \Sigma a。$$

此时设置一个比值为：

$$\frac{(\mu_{y1} - \mu_{y2})^2}{\sigma_y^2} = \frac{[a^T (\mu_1 - \mu_2)]^2}{a^T \Sigma a} = \frac{(a^T \delta)^2}{a^T \Sigma a}$$

其中： $\delta = \mu_1 - \mu_2$ 。此时，根据 *Fisher* 判别的思想，要选择 a 使得比值达到最大。有以下定理可以确定 a 值与判别函数：

定理 1 x 为 p 维随机向量，设 $y = a^T x$ ，当选取 $a = c \Sigma^{-1} x$ ， $c \neq 0$ 且为常数时，比值达到最大。特别地，当 $c = 1$ 时，线性函数：

$$y = a^T x = (\mu_1 - \mu_2)^T \Sigma^{-1} x$$

把它叫做 *Fisher* 线性判别函数。

令：

$$K = \frac{1}{2}(u_{y1} + u_{y2}) = \frac{1}{2}(a^T u_1 + a^T u_2) = \frac{1}{2}(u_1 - u_2)^T \Sigma^{-1} (u_1 + u_2)$$

定理 2 根据定理 1, 取 $a^T = (\mu_1 - \mu_2)^T \Sigma^{-1}$, 则有: $u_{y1} - K > 0, u_{y2} - K < 0$ 。
此时, 由定理 3.2 可得到如下的 *Fisher* 判别规则:

$$\begin{cases} x \in X_1, \text{ 当 } x \text{ 使得 } (\mu_1 - \mu_2)^T \Sigma^{-1} x \geq K, \\ x \in X_2, \text{ 当 } x \text{ 使得 } (\mu_1 - \mu_2)^T \Sigma^{-1} x < K. \end{cases}$$

则定义判别函数为:

$$W(x) = (\mu_1 - \mu_2)^T \Sigma^{-1} x - K = \left[x - \frac{1}{2}(\mu_1 + \mu_2) \right]^T \Sigma^{-1} (\mu_1 - \mu_2),$$

最后判别规则可改写为[12]:

$$\begin{cases} x \in X_1, \text{ 当 } x \text{ 使得 } W(x) \geq K, \\ x \in X_2, \text{ 当 } x \text{ 使得 } W(x) < K. \end{cases}$$

建立玻璃文物类别鉴定的 *Fisher* 判别分析模型首先需要确定评价指标, 这是模型准确性的关键。根据题目所给条件, 选定 14 种化学成分与表面风化共 15 个数据, 即二氧化硅(S1), 氧化钠(S2), 氧化钾(S3), 氧化钙(S4), 氧化镁(S5), 氧化铝(S6), 氧化铁(S7), 氧化铜(S8), 氧化铅(S9), 氧化钡(S10), 五氧化二磷(S11), 氧化锶(S12), 氧化锡(S13), 二氧化硫(S14), 表面氧化(S15)共 15 个影响因子作为玻璃判别分析的评价指标。再将玻璃的两种类型作为 *Fisher* 模型的输出参数。

4.2. Fisher 判别模型的求解

接下来, 将实验数据的附件表单 2 [1]作为 *Fisher* 模型的学习样本进行训练, 并且应用上文所介绍的原理, 利用 SPSS 软件与 *Fisher* 判别函数求得线性系数向量见表 5:

Table 5. Vector of linear coefficients
表 5. 线性系数向量

名称	线性系数	名称	线性系数
二氧化硅	0.017	氧化铅	0.137
氧化钠	0.211	氧化钡	0.170
氧化钾	-0.342	五氧化二磷	0.154
氧化钙	-0.014	氧化锶	-1.378
氧化镁	0.053	氧化锡	0.240
氧化铝	0.205	二氧化硫	-0.051
氧化铁	-0.110	表面风化	-2.422
氧化铜	-0.104	常数	-4.447

求得玻璃类别鉴定的 *Fisher* 判别函数为:

$$\left\{ \begin{array}{l} \text{高钾玻璃: } y_0 = -1361.792 + 27.428S1 + 30.246S2 + 38.367S3 + 27.235S4 + 38.254S5 \\ \quad + 23.234S6 + 24.986S7 + 11.565S8 + 27.899S9 + 35.163S10 \\ \quad + 36.565S11 - 58.381S12 + 2.176S13 + 11.732S14 + 7.737S15 \\ \text{铅钡玻璃: } y_1 = -1381.660 + 27.546S1 + 31.734S2 + 35.956S3 + 27.135S4 + 38.626S5 \\ \quad + 24.678S6 + 24.209S7 + 10.832S8 + 28.868S9 + 36.364S10 \\ \quad + 37.655S11 - 68.114S12 + 3.872S13 + 11.373S14 - 9.366S15 \end{array} \right.$$

此时高钾玻璃与铅钡玻璃的分类规律为：将 $S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15$ 带入计算出 y_0, y_1 ， y_0 的数值比较大时就为高钾玻璃， y_1 的数值比较大时就为铅钡玻璃。

4.3. Fisher 判别模型的灵敏度分析

为了进行 Fisher 判别模型的灵敏度分析，首先利用实验数据的附件表单 2 [1]，分别通过 Excel 算出高钾玻璃、铅钡玻璃的化学成分均值。

Table 6. Average chemical composition of high potassium glass and lead-barium glass

表 6. 高钾玻璃与铅钡玻璃化学成分均值

化学成分	高钾均值	铅钡均值
二氧化硅(SiO ₂)	76.64389	39.58938
氧化钠(Na ₂ O)	2.78	3.165714
氧化钾(K ₂ O)	7.682	0.354167
氧化钙(CaO)	4.325625	2.210455
氧化镁(MgO)	1.1775	1.054333
氧化铝(Al ₂ O ₃)	5.056667	3.720833
氧化铁(Fe ₂ O ₃)	1.001429	1.19037
氧化铜(CuO)	2.282353	2.021818
氧化铅(PbO)	0.705714	33.36792
氧化钡(BaO)	1.436	10.74244
五氧化二磷(P ₂ O ₅)	1.156875	3.8445
氧化锶(SrO)	0.083333	0.42359
氧化锡(SnO ₂)	2.36	0.57
二氧化硫(SO ₂)	0.406667	6.0375

由表 6 知，在高钾玻璃与铅钡玻璃中化学成分含量差别较大的成分是：二氧化硅、氧化钾、氧化铅、氧化钡、二氧化硫。换言之，两种玻璃中其它元素含量类似，而这五种元素的含量决定了玻璃的类型。当改变这四种元素的含量时，也就会改变玻璃的类型。

此时我们随机选择了十组高钾玻璃的数据，将其二氧化硅、氧化钾、氧化铅、氧化钡、二氧化硫的数据修改为铅钡玻璃的平均值，如果玻璃类型的 Fisher 判别模型的灵敏性较好，那么应该此时预测这十组修改过后的数据，得到的结果应该都是铅钡玻璃。

Table 7. The prediction result after the modification of the value**表 7.** 修改数值后预测结果

文物采样点	原类别	修改后类别	文物采样点	原类别	修改后类别
01	0	1	06 部位 1	0	1
03 部位 1	0	1	06 部位 2	0	1
03 部位 2	0	1	07	0	1
04	0	1	12	0	1
05	0	1	13	0	1

观察表 7 的原类别与预测结果发现：修改数据之后，十组数据都从高钾玻璃预测得到铅钡玻璃，符合预期，模型判别的准确率为 100%，说明此时的玻璃类型的 *Fisher* 判别模型的灵敏度是较好的。

5. 结论

本文通过对已有的实验数据进行分析后建立数学模型，基于实验数据探究了玻璃文物样品表面有无风化化学成分含量的统计规律、不同类别玻璃文物样品的化学成分之间的关联关系和给定的玻璃文物的类别鉴定。首先，通过玻璃表面有无风化化学成分的统计规律的描述性统计分析和卡方检验分析知，高钾玻璃在风化后主要化学成分含量呈下降趋势，铅钡类玻璃在风化后主要化学成分含量呈上升趋势；玻璃的表面风化与类型存在相关性且相关性较强。

然后，基于玻璃的表面风化与类型存在相关性且相关性较强这个结论，建立玻璃文物化学成分的无监督的系统聚类算法模型，研究表明，高钾风化、高钾无风化、铅钡风化和铅钡无风化的玻璃文物的化学成分的最优分类数目分别为 2、3、4、4。再通过轮廓系数评价指标说明铅钡玻璃风化后化学成分关联关系加强，而高钾玻璃风化后化学成分关联关系减弱。

最后，依据玻璃文物的化学成分之间的关联关系研究的结论，建立了玻璃类别鉴定的 *Fisher* 判别模型，得到玻璃类别鉴定的 *Fisher* 判别公式，研究表明，玻璃类别鉴定的 *Fisher* 判别模型的判别准确率较高、灵敏度较好。本文对于古代玻璃的成分分析与类别鉴定具有一定的指导价值。

参考文献

- [1] 中国工业与应用数学学会. 2021 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/5267fe3e6a512bec793d71f2b2061497.html, 2022-09-15.
- [2] 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.
- [3] 胡志中, 李佩, 蒋璐蔓, 王通洋, 杜谷, 杨波. 古代玻璃材料 LA-ICP-MS 组分分析及产源研究[J]. 岩矿测试, 2020, 39(4): 505-514.
- [4] 刘松, 李青会, 干福熹. 古代玻璃样品表面因素对便携式 X 射线荧光定量分析的影响[J]. 光谱学与光谱分析, 2011, 31(7): 1954-1959.
- [5] 薛吕. 玻璃文物保护与修复[C]//中国文物保护技术协会. 中国文物保护技术协会第七次学术年会论文集. 北京: 科学出版社, 2012: 89-92.
- [6] 李军, 罗海明. 安徽当涂东晋墓发掘简报——兼论出土的玻璃碗残片类别和来源[J]. 东南文化, 2006(2): 33-39.
- [7] 周晓波, 王洪, 周郭飞. 一种基于峰度系数的脉冲噪声检测算法的研究及其 FPGA 实现[J]. 电子与信息学报, 2021, 43(7): 1816-1820.
- [8] 胡松, 江亚群, 黄纯. 基于偏度系数的变压器励磁涌流识别方法[J]. 电网技术, 2018, 42(6): 1954-1959.
- [9] 李濛, 包蕾, 胡毅, 成嵩, 胡晓波, 高鹰. 基于卡方检验的随机数在线检测方法的实现[J]. 微电子学, 2022, 52(3):

388-392.

- [10] 杨真真, 李泓瑶, 杨可易, 刘琳. 基于系统聚类算法与肘部准则的各省市居民消费水平聚类分析[J]. 中国科技信息, 2021(12): 121-122.
- [11] Peter, J.R. (1987) Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, **20**, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- [12] 司守奎, 孙玺倩. 数学建模算法与应用[M]. 第三版. 北京: 国防工业出版社, 2021: 306-307.