

基于数学模型的古代玻璃制品的鉴定与分类

沈佳, 郭智蕊

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2023年2月13日; 录用日期: 2023年3月8日; 发布日期: 2023年3月15日

摘要

古代玻璃种类繁多且结构复杂, 因此对于古代玻璃制品的鉴别与分类成为值得关注的方面。本文以2022年高教杯全国大学生数学建模C题——古代玻璃制品的成分分析与鉴别为基础, 通过构建数学模型来探究两类玻璃样品的统计规律, 并在预测样品风化前化学成分含量以及两类玻璃的亚类划分问题中分别应用统计学方法, 提供新的玻璃分析方法, 从而更精准的分析玻璃成分, 为玻璃生产领域注入新的生机。

关键词

聚类分析法, 斯皮尔曼相关系数, 决策树, 玻璃制品

Identification and Classification of Ancient Glass Products Based on the Mathematical Model

Jia Shen, Zhirui Guo

College of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: Feb. 13th, 2023; accepted: Mar. 8th, 2023; published: Mar. 15th, 2023

Abstract

There are many kinds of ancient glass and the structure is complex, so the identification and classification of ancient glass products become worthy of attention. Based on the composition analysis and identification of ancient glass products in 2022, this paper explores the statistical rules of the two types of glass samples, and applies statistical methods in predicting the problems of chemical composition content of samples before weathering and the subclass division of the two types of glass, and provides new glass analysis methods, so as to more accurately analyze the glass composition, and inject new vitality into the field of glass production.

Keywords

Clustering Methodology, Spearman Correlation Coefficient, Decision-Making Tree, Glass Products

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

随着我国丝绸之路的发展, 早期贸易的物品越来越受关注, 这使得古代玻璃成为考古人员的重要研究内容。古代丝绸之路是古玻璃发展的重要媒介, 我国古代玻璃主要以含氧化铅和氧化钡较高的铅钡玻璃为特色, 铅钡玻璃是我国最早制造出的本土玻璃。由于判别玻璃类型等信息主要依据其化学成分, 因此本文将利用决策树模型对古代玻璃制品的化学成分进行数据分类, R 型聚类分析法研究不同化学成分之间的关系, 进而对古代玻璃制品进行分类以及预测出风化前各元素含量。

2. 预备知识

1) 决策树是一种用于数据分类的方法, 通过建立一种树形结构来进行问题决策。它是直观运用概率分析的一种图解法, 每个内部节点表示某种特征的测试, 每个分支代表一个测试输出, 每个叶节点是输出结果[1]。在建立决策树模型时, 通常采用 ID3 算法, 即利用信息增益来度量。

信息熵(H)以及信息增益(I)定义为:

$$\begin{aligned} H(p) &= -\sum p \times \log p \\ H(Y|X) &= \sum_{i=1}^n p_i H(Y|X=X_i) \\ I(D, N) &= H(D) - H(D|N) \end{aligned} \quad (1)$$

其中 p 表示随机变量的概率, N 表示变量的特性, D 表示数据集合, $H(D)$ 定义为经验熵, $H(Y|X)$ 为条件熵, $H(D|N)$ 表示 N 在 D 条件下的经验熵。

2) R 型聚类是聚类分析的一种, 一般对样本指标进行分类[2]。在对指标进行聚类时, 需计算出各变量间的相关系数, 这里给出最长距离法。

记变量 f_p 的取值为 $[f_{1p}, f_{2p}, \dots, f_{14p}] \in R^{14}$, ($p=1, 2, \dots, 14$), 则计算两个变量 f_p 与 f_q 的相关系数, 从而确定其为相似性度量, 即

$$R_{pq} = \frac{\sum_{i=1}^{14} (f_{ip} - \bar{f}_p)(f_{iq} - \bar{f}_q)}{\left[\sum_{i=1}^{14} (f_{ip} - \bar{f}_p)^2 \sum_{i=1}^{14} (f_{iq} - \bar{f}_q)^2 \right]^{\frac{1}{2}}}, p, q = 1, 2, \dots, 14 \quad (2)$$

3. 玻璃文物表面有无风化的量化对比与统计分析

由于题中所给有效数值范围为 85%~105%, 根据附表单 1 和单 2 的数据, 通过对各样本成分含量累加进行数据的合理性分析, 得到 67 组有效数据。利用斯皮尔曼相关系数法来研究分类信息之间的相关性, 构建统计模型, 进而分析研究文物样品表面有无风化化学成分含量的相关统计规律。

3.1. 分类信息的相关性计算

斯皮尔曼相关系数是用来衡量变量间非参数指标之间的相关性, 其主要适用于其中一组变量为定序变量的情况[3]。

1) 定义变量

分别对四组原始变量进行定量处理, 即各组变量间不同指标所占比重按照从小到大的顺序进行赋值(若不同指标所占比重相同, 则对其所赋值求取平均), 具体赋值情况如下表 1 所示:

Table 1. Argument Table

表 1. 变量表

变量名称	变量符号	变量定义
表面风化	f	$f = \begin{cases} 1 & \text{无风化} \\ 2 & \text{风化} \end{cases}$
纹饰	w	$w = \begin{cases} 1 & B \\ 2 & A \\ 3 & C \end{cases}$
类型	x	$x = \begin{cases} 1 & \text{高钾} \\ 2 & \text{铅钡} \end{cases}$
颜色	y	$y = \begin{cases} 1 & \text{绿} \\ 2.5 & \text{深蓝} \\ 4 & \text{浅绿} \\ 5.5 & \text{紫、空白} \\ 7 & \text{深绿} \\ 8 & \text{蓝绿} \\ 9 & \text{浅蓝} \end{cases}$

2) 计算差值

结合上表, 分别计算纹饰、类型和颜色与表面风化对应差值, 即将上表中 w, x, y 与 f 对应元素差值组成的集合分别记为 d_1, d_2, d_3 , 记各集合的第 j 个元素为 $d_{j1}, d_{j2}, d_{j3}, (j=1, 2, \dots, 58)$, 公式如下:

$$\begin{aligned} d_{j1} &= x_j - f_j \\ d_{j2} &= w_j - f_j \\ d_{j3} &= y_j - f_j \end{aligned} \quad (3)$$

注: 如果两个元素的数值相同, 则将取算术平均值进行差值计算。

3) 计算斯皮尔曼相关系数

利用式(3), 分别计算 w, x, y 与 f 之间的斯皮尔曼系数, 分别用 $R_1 \sim R_3$ 表示。其中斯皮尔曼相关系数 $R_i \in [-1, 1]$, 当 R_i 绝对值越接近于 1 时, 两变量之间相关性越强。

$$R_i = 1 - \frac{6 \sum_{i=1}^{58} d_{ij}^2}{N(N^2 - 1)}, i = (1, 2, 3) \quad (4)$$

4) 对斯皮尔曼相关系数进行假设检验

由于原始样本数 $n > 30$, 选用适合多样本点数的 p 值判断法对三组相关系数分别进行检验。首先确

定原假设 $r_s = 0$ 、备择假设 $r_s \neq 0$ ，其中 $r_s = 1, 2, 3$ 。在原假设成立的条件下，进一步构造服从标准正态分布的统计量，即

$$r_s \sqrt{n-1} \sim N(0,1) \quad (5)$$

5) 模型的求解

将差值集中的元素 d_{ij} 代入式(3)，求解出对应的 R_i ，即 w, x, y 与 f 之间的相关系数。最终得出的结果如表 2:

Table 2. Correlation coefficient

表 2. 相关系数

变量	纹饰(w)	类型(x)	颜色(y)
相关系数(R)	-0.1164	0.3444	0.0583

将样本数据代入式(4)中的统计量中，得出正态检验值 m ，利用标准正态分布密度函数求得 p 值，求得结果如表 3:

Table 3. Hypothesis test

表 3. 假设检验

p_1	p_2	p_3
0.3840	0.0081	0.6636

在置信水平为 95% 的基础上，进行 P 值检验，结果得出只有 $p_2 < 0.05$ ，通过显著性检验。

6) 结果分析

由于玻璃类型与表面风化的相关系数 $R_2 = 0.344$ ，相关系数的检验值 $p_2 = 0.0081$ ； $R_3 = 0.0583$ ，相关系数的检验值 $p_1 = 0.6636$ 。因此，在显著性检验值为 0.05 的基础上，表面风化与否与玻璃类型呈显著性关系，与玻璃颜色呈不显著关系。

3.2. 统计规律

为了探究不同玻璃类型表面有无风化化学成分含量的统计规律，对附件表单 2 样本进行分类，分成高钾类与铅钡类，同时利用决策树模型对原始数据进行分类，将其拆分成风化点数据与未分化点数据。

1) 高钾玻璃

统计得高钾玻璃表面有无风化时其他化学成分占比的最值与平均值，通过对图 1 表面有无风化的化学成分含量的变化趋势，发现无风化时各个化学成分占比，较风化来说均是高的，并且氧化铝在风化时平均含量值接近 2%，在未风化时平均含量值接近 11%，较其余成分相比均为最高。

2) 铅钡玻璃

通过对图 2 的直观分析，发现氧化铝、氧化铜、氧化钠、氧化钙最大值含量在无风化时达到峰值，而氧化锡的最大值则随着风化程度加深的过程中逐渐降低。

3) 针对风化点与未风化点的分类

针对未风化点数据，取 70% 数据作为训练集，30% 作为测试集。由结果所得，未分化的玻璃的分类规律主要受 PbO 影响，当 PbO 含量小于等于 8.488 时，该样本归为高钾玻璃类；当 PbO 含量大于 8.488 时，该样本归为铅钡玻璃类。

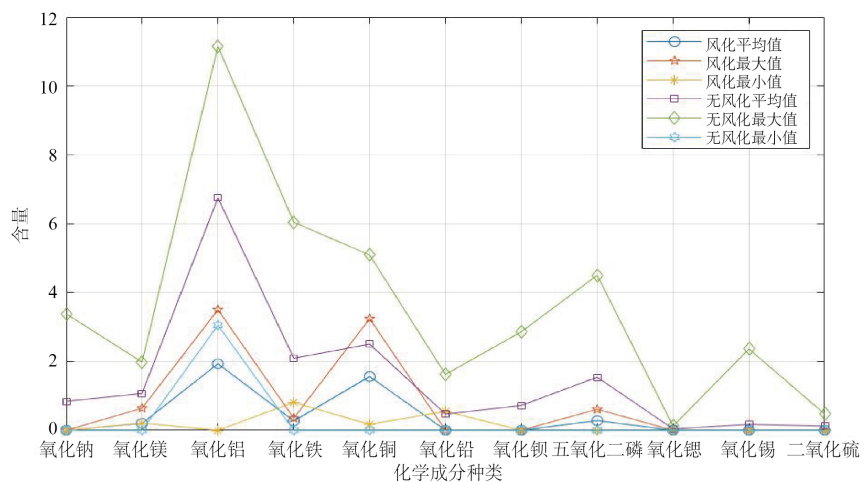


Figure 1. The content of other chemical components of high potassium glass
图 1. 高钾玻璃其他化学成分含量

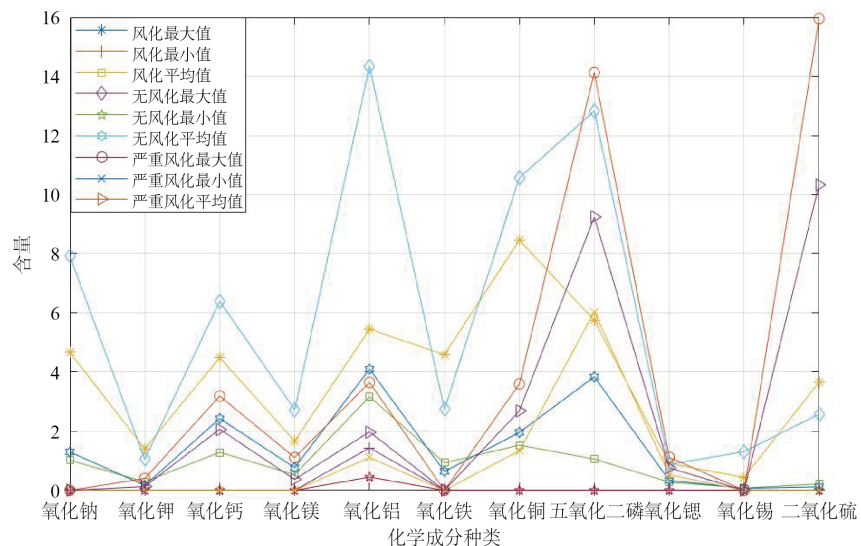


Figure 2. Other chemical components of lead and barium glass
图 2. 铅钡玻璃其他化学成分含量

针对风化点数据, 取 70%数据作为训练集, 30%数据作为测试集。由结果所得, 风化的玻璃的分类规律主要受 PbO 影响, 当 PbO 含量小于等于 6.255 时, 该样本归为高钾玻璃类; 当 PbO 含量大于 6.255 时, 该样本归为铅钡玻璃类。

3.3. 两类玻璃的亚类划分

通过查阅资料可知, 样本所含的各化学成分含量可看作分类的重要指标。为了将两类玻璃进行亚类的划分, 对不同样本的化学成分进行聚类, 通过数学统计模型单独研究每个类别的性质[4]。

利用式(4)得两类玻璃各化学成分间的相关系数, 以铅钡玻璃为例, 计算相关系数, 具体见图 3。

基于玻璃类型 14 种化学成分比例的不同, 根据其相关性建立 R 型聚类分析模型来研究各化学成分之间的关系, 按最长距离法对玻璃类型进行亚类划分, 将相似的化学成分分成一类。

对两个样本间的距离进行计算, 把每个化学成分看作一个类, 任选一个模式样本作为第一类的聚类中心 Z_1 , 选择离 Z_1 距离最远的模式样本作为第二类聚类中心 Z_2 , 逐个计算每个模式样本与已知确定所有聚类之间的距离, 选出其中的最小值。

	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14
f1	1	0.0707	0.1358	0.2617	0.1664	0.2393	0.0867	0.2645	0.6135	0.74	0.0331	0.5999	0.5074	0.0773
f2	0.0707	1	0.0115	0.2009	0.0332	0.0372	0.4726	0.2009	0.1003	0.0111	0.4	0.0459	0.2304	0.1565
f3	0.1358	0.0115	1	0.1528	0.2735	0.5595	0.043	0.0679	0.1074	0.1752	0.2405	0.0698	0.0719	0.1587
f4	0.2617	0.2009	0.1528	1	0.1866	0.0055	0.323	0.4187	0.4814	0.0908	0.1212	0.2971	0.6211	0.1545
f5	0.1664	0.0332	0.2735	0.1866	1	0.4392	0.0157	0.3386	0	0.2659	0.4167	0.2408	0.2205	0.284
f6	0.2393	0.0372	0.5595	0.0055	0.4392	1	0.2867	0.0303	0.2637	0.4286	0.0055	0.2035	0.2795	0.0772
f7	0.0867	0.4726	0.043	0.323	0.0157	0.2867	1	0.6131	0.0836	0.1493	0.4157	0.086	0.5603	0.2516
f8	0.2645	0.2009	0.0679	0.4187	0.3386	0.0303	0.6131	1	0.2779	0.4512	0.2558	0.5313	0.6202	0.1543
f9	0.6135	0.1003	0.1074	0.4814	0	0.2637	0.0836	0.2779	1	0.0934	0.2558	0.5313	0.6202	0.1543
f10	0.74	0.0111	0.1752	0.0908	0.2659	0.4286	0.1493	0.4512	0.0934	1	0.0413	0.4465	0.2359	0.1543
f11	0.0331	0.4	0.2405	0.1212	0.4167	0.0055	0.4157	0.2558	0.2558	0.0413	1	0.2349	0.105	0.3039
f12	0.5999	0.0459	0.0698	0.2971	0.2408	0.2035	0.086	0.5313	0.5313	0.4465	0.2349	1	0.3594	0.0794
f13	0.5074	0.2304	0.0719	0.6211	0.2205	0.2795	0.5603	0.6202	0.6202	0.2359	0.105	0.3594	1	0.1227
f14	0.0773	0.1565	0.1587	0.1545	0.284	0.0772	0.2516	0.1543	0.1543	0.1543	0.3039	0.0794	0.1227	1

Figure 3. Correlation coefficient among various chemical components of high potassium glass

图 3. 高钾玻璃各化学成分间的相关系数

在所有最小距离中选出一个最大距离, 如果该值与 $|Z_1 - Z_2|$ 的比值在一定分数以上, 则产生最大距离的那个模式样本为新增的聚类中心, 重复以上步骤, 直到没有新的聚类中心为止, 从而对各类型玻璃进行亚类划分。

1) 高钾玻璃的亚类划分(如图 4 所示):

第一类包含的化学成分有: 氧化钙、氧化铜、氧化锡、二氧化硫

第二类包含的化学成分有: 二氧化硅、氧化钠、氧化钾、氧化镁、氧化铝、氧化铅、氧化钡、五氧化二磷、氧化锶

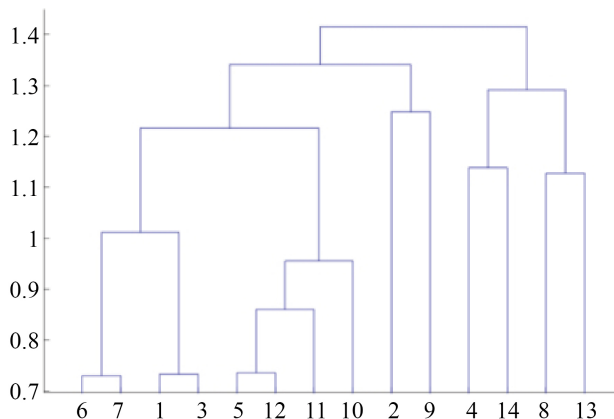


Figure 4. Subclass division of the high-potassium chemical composition

图 4. 对高钾化学成分的亚类划分

2) 铅钡玻璃的亚类划分(如图 5 所示):

第一类包含的化学成分有: 二氧化硅、氧化钙、氧化铜、氧化铅、氧化钡、氧化锶、氧化锡

第二类包含的化学成分有: 氧化钠、氧化钾、氧化镁、氧化铝、五氧化二磷、氧化硫

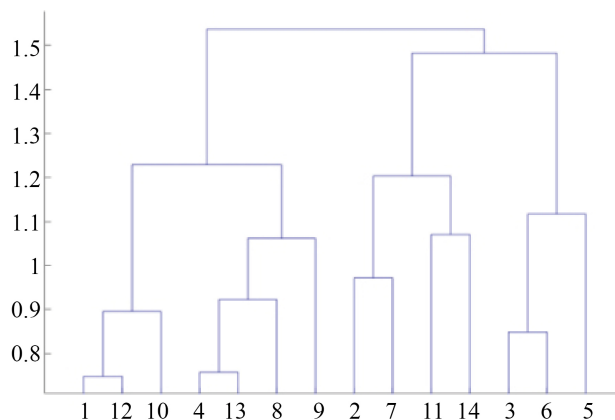


Figure 5. Subclass division of the chemical composition of lead and barium
图 5. 对铅钡化学成分的亚类划分

3.4. 预测风化前各元素含量

由于高钾玻璃和铅钡玻璃预测风化前的元素含量所用方法等同, 因此本文将以铅钡玻璃为例, 计算其风化前的各元素含量。

1) 首先, 依据玻璃类型以及表面是否风化将原始样本分为四类: 铅钡风化类(A)、铅钡未风化类(C)。为保证预测风化前成分值的合理性, 利用平均值作为衡量各类别样本的标准, 即分别求出上述两类中样本各化学指标的均值, 分别记为

$$A' = (A'_1, A'_2, \dots, A'_{14}), C' = (C'_1, C'_2, \dots, C'_{14}) \tag{6}$$

2) 其次, 利用移动平均法求预测值[5]。设上述四类中第*i*个样本的第*j*个化学指标分别为 A_{ij}, C_{ij} 。同时, 设定 A 两类中的第 *i* 个样本的第 *j* 个未风化前的化学指标为 A'_{ij} , 其中 $j = (1, 2, \dots, 14)$ 。由于需要预测风化样本风化前的成分含量, 所以只需计算 A 类样本风化前的成分含量。根据距离相等, 有以下公式成立:

$$A_{ij} - A'_{ij} = A'_j - C'_j \tag{7}$$

根据上述内容求解出风化前的预测值, 计算得铅钡玻璃风化前的部分化学成分含量值如下表 4。

Table 4. Prediction of partial chemical content of lead-barium glass before weathering
表 4. 铅钡玻璃预测风化前部分化学含量值

文物采样点	类型	(SiO ₂)风化前	(Na ₂ O)风化前	(K ₂ O)风化前	(CaO)风化前	(MgO)风化前
02	铅钡	56.11	0.00	1.17	1.22	0.97
08	铅钡	39.97	0.00	0.12	0.36	0.00
08 严重风化点	铅钡	24.44	0.00	0.12	2.07	0.00
11	铅钡	53.42	0.00	0.33	2.39	0.50
19	铅钡	49.47	0.00	0.12	1.81	0.38
26	铅钡	39.62	0.00	0.12	0.32	0.00
26 严重风化点	铅钡	23.55	0.00	0.52	1.89	0.00
34	铅钡	55.61	0.00	0.37	0.00	0.00
36	铅钡	59.40	2.04	0.12	0.00	0.00

Continued

38	铅钡	52.76	1.20	0.12	0.00	0.00
39	铅钡	46.08	0.00	0.12	0.00	0.00
40	铅钡	36.54	0.00	0.12	0.75	0.00
41	铅钡	38.29	0.00	0.56	3.84	2.52
43 部位 1	铅钡	32.24	0.00	0.12	4.12	0.68

4. 结论

古代玻璃制品主要分为高钾玻璃与铅钡玻璃, 对于分化前后的玻璃, 其所含的化学成分有很大区别。通过计算斯皮尔曼相关系数以及决策树的构建, 我们发现无风化时高钾玻璃中各个化学成分含量最高, 并且氧化铝、氧化钾是高钾玻璃中的主要化学成分。利用聚类分析法对两大类玻璃进行了亚类划分, 此外, 通过移动平均法对这两种玻璃制品进行风化前的化学元素含量的预测, 这为玻璃的化学成分研究提供了新的方法, 加强了玻璃成分分析, 为玻璃生产领域注入了新的生机。

基金项目

大创项目编号: X202210165176。

参考文献

- [1] 吕飞, 陈明洁, 魏晓芳. 基于决策树模型的医疗设施可达性影响因素分析——以苏州市为例[J]. 上海城市规划, 2022, 166(5): 40-44.
- [2] 卓金武, 王鸿钧. MATLAB 数学建模方法与实践[M]. 第 3 版. 北京: 北京航空航天大学出版社, 2018.
- [3] 于群, 霍筱东, 何剑, 李琳, 张建新, 冯煜尧. 基于斯皮尔曼相关系数和系统惯量的中国电网停电事故趋势预测[J/OL]. 中国电机工程学报, 1-12, 2023-03-11. <http://kns.cnki.net/kcms/detail/11.2107.TM.20220824.1625.012.html>
- [4] 史美光, 何欧里, 吴宗道, 周福征. 一批中国古代铅玻璃的研究[J]. 硅酸盐通报, 1986(1): 17-23. <https://doi.org/10.16552/j.cnki.issn1001-1625.1986.01.004>
- [5] 姜启源, 谢金星, 叶俊. 数学模型[M]. 第 5 版. 北京: 高等教育出版社, 2018.