

引文格式:高国云,王青芸.基于支持向量机(SVM)的古代玻璃制品分类[J].赣南师范大学学报,2024,45(3):19-22.

基于支持向量机(SVM)的 古代玻璃制品分类^{*}

高国云,王青芸[†]

(赣南师范大学 数学与计算机科学学院,江西 赣州 341000)

摘要:古代玻璃制品是古丝绸之路交易的商品之一,一般依据化学成分对玻璃制品进行分类.但是风化会改变玻璃制品化学成分的含量,从而影响玻璃制品类型的鉴别.本文尝试先预测风化前的化学成分以消除风化的影响,再采用灰色关联分析化学成分的关联关系以及差异,最后建立支持向量机(SVM)模型对古代玻璃制品进行分类.

关键词:支持向量机(SVM);系统聚类;灰色关联分析;古代玻璃;玻璃风化

中图分类号:O29

文献标志码:A

文章编号:2096-7659(2024)03-0019-04

0 引言

古丝绸之路是东西方的经济、政治、文化交流的主要通道,古代玻璃制品是当时古丝绸之路上交易的商品之一.我国考古工作者依据玻璃制品的化学成分和其他检测手段将玻璃文物分类.例如赵凤燕等人认为按化学成分划分西安出土的若干玻璃器可分为四种成分体系^[1],赵德云认为按制造技术划分中国的蜻蜓眼式玻璃珠能分成九型^[2].但是古代玻璃很容易因为埋藏环境的影响导致风化,在风化过程中,玻璃的化学成分也会产生变化.王承遇和陶瑛认为钠和钙的碳酸盐是钠钙硅酸盐玻璃风化主要产物^[3],宋燕和马清林认为玻璃风化严重的主要原因是外界环境中的二氧化碳、水分和表面附着物等的长期影响^[4].风化会影响古代玻璃类别的判断,但目前通过预测风化前化学成分进行玻璃分类的研究比较少.

为了提高玻璃制品分类的准确性,本文通过预测玻璃风化前的化学成分对玻璃制品分类.本文收集了我国最常见的两种古代玻璃:高钾玻璃和铅钡玻璃的采样数据,数据包括文物的风化情况、类型、纹饰等信息以及采样点的主要化学成分比例.先对数据进行 clr 变换预测风化前的化学成分,再运用灰色关联得到部分化学成分的关联关系,最后构建支持向量机(SVM)模型对玻璃制品进行分类.本文的分析方法能对考古工作者的研究有帮助.

1 方法及分析

1.1 预测风化前数据

高钾玻璃和铅钡玻璃的采样数据^①包括文物的表面风化、类型、颜色等信息和采样点的主要化学成分比例.其中确定类型的有 67 个,未确定类型的有 8 个.首先确定风化对玻璃制品化学成分占比的影响,先借助 R 语言软件进行 fisher 精确检验确定表面风化与玻璃类型、纹饰、颜色的相关关系.由表 1 可知,表面风化与纹饰、颜色是不相关的.而表面风化与玻璃类型是相关的,且其置信区间 $[0.054, 0.811]$ 整体小于 1,说明优势比小于 1,表示负相关,即铅钡玻璃比高钾玻璃更容易风化.铅钡玻璃的风化过程^[5]是铅与外界的水蒸气

^{*} 收稿日期:2024-04-20

DOI:10.13698/j.cnki.cn36-1346/c.2024.03.004

基金项目:江西省自然科学基金(20224BAB201014);赣州市科技计划项目(2023PNS16940)

作者简介:高国云(2001-),女,江西赣州人,赣南师范大学数学与计算机科学学院 2020 级应用统计学专业本科生.

[†] 通讯作者:王青芸(1983-),女,江西赣州人,赣南师范大学数学与计算机科学学院副教授,研究方向:生物数学.

^① 数据来源为 http://www.mcm.edu.cn/html_cn/node/388239ded4b057d37b7b8e51e33fe903.html.

和二氧化碳之间接触,生成白色粉末状且不溶于水的碳酸铅,碳酸铅积累的越多,玻璃风化程度越高。

由 fisher 精确检验可知不同类型的玻璃受风化的影响程度不同,故需要预测风化前的化学成分含量。但收集的数据中有些文物采样点风化与否未确定,借助 R 语言软件进行系统聚类^[6]将采样点分为未风化点和风化点。部分聚类结果见表 2。

所采集的化学成分数据是比例数据,需进行变换。clr 变换^[7]可将数据变得线性可分,因此使用 clr 变换后的数据去预测。将数据分成高钾玻璃和铅钡玻璃,再分别考虑风化数据 CLR(A)与未风化数据 CLR(B)。其中 A 指风化点数据,CLR(A)指作 clr 变换的风化点数据,B 指未风化点数据,CLR(B)指作 clr 变换的未风化点数据。认为作 clr 变换的风化点数据呈正态分布集中在一起,作 clr 变换的未风化点数据也呈正态分布集中在一起,故构建

风化前数据的预测方程式(1)。其中 μ_a 为风化数据均值, μ_b 为未风化数据均值, σ_a^2 为风化数据方差, σ_b^2 为未风化数据方差。

$$\begin{cases} CLR(A) \sim N(\mu_a, \sigma_a^2) \\ CLR(B) \sim N(\mu_b, \sigma_b^2) \\ CLR(B) - \mu_b = \frac{\sigma_b}{\sigma_a} (CLR(A) - \mu_a) \end{cases} \quad (1)$$

为预测风化前的化学成分,先计算其均值与方差,再求出方程的系数,最后将风化数据代入预测方程得出风化前的推测结果。式(2)为高钾玻璃中 SiO_2 的风化预测方程,其他的风化预测方程以此类推。用风化预测方程得到了作 clr 变换的风化前数据 CLR(B),再进行 clr 逆变换得到风化前的成分数据 B。式(3)为 clr 变换和 clr 逆变换的公式(上:clr 变换,下:clr 逆变换)。部分预测结果见表 3。

$$CLR(B) = 0.831 \times CLR(A) - 1.627 \quad (2)$$

$$\begin{aligned} clr(x_i) &= \ln \left(\frac{x_i}{\sqrt{\prod_{j=1}^D x_j}} \right) = y_i, i = 1, 2, \dots, D \\ x_i &= \frac{e^{y_i}}{\sum_{j=1}^D e^{y_j}}, i = 1, 2, \dots, D \end{aligned} \quad (3)$$

表 3 部分预测结果

文物采样点	SiO ₂	Na ₂ O	K ₂ O	CaO	MgO	Al ₂ O ₃	Fe ₂ O ₃
02	57.245%	0.000%	17.580%	0.510%	4.205%	5.606%	0.176%
08	59.037%	0.000%	0.000%	0.616%	0.000%	2.971%	0.000%
08 严重风化点	27.787%	0.001%	0.000%	2.572%	0.000%	4.599%	0.000%
03 部位 1	25.028%	0.000%	9.389%	6.037%	0.002%	5.782%	0.083%
文物采样点	CuO	PbO	BaO	P ₂ O ₅	SrO	SnO ₂	SO ₂
02	0.182%	14.322%	0.072%	0.081%	0.021%	0.000%	0.000%
08	5.709%	16.176%	15.204%	0.179%	0.100%	0.000%	0.007%
08 严重风化点	3.904%	32.236%	27.805%	0.788%	0.268%	0.000%	0.039%
03 部位 1	1.227%	51.570%	0.000%	0.882%	0.000%	0.000%	0.000%

1.2 化学成分含量关系

化学成分数目较多,为减少冗余信息,降低计算复杂程度,提高识别精度,需要对数据进行降维。为找到化学成分中更重要的成分,先借助 MATLAB 软件进行灰色关联^[8]探究化学成分之间的关联关系。由表 4 可

知,高钾玻璃 SiO_2 与 Al_2O_3 有较强的关联度. 铅钡玻璃 SiO_2 与 Al_2O_3 、 BaO 也有较强的关联度. 两类别关联关系的差异是 SiO_2 与 PbO 、 BaO 的关联度不同,铅钡玻璃中关联度高,而高钾玻璃中关联度低.

表 4 灰色关联度及排名

高钾玻璃 关联度 排名	铅钡玻璃 关联度 排名
K_2O 0.891 1	PbO 0.922 1
Al_2O_3 0.867 2	BaO 0.859 2
CaO 0.820 3	Al_2O_3 0.827 3
CuO 0.813 4	CuO 0.733 4
Fe_2O_3 0.801 5	CaO 0.728 5
MgO 0.750 6	MgO 0.662 6
P_2O_5 0.744 7	P_2O_5 0.649 7
PbO 0.600 8	SrO 0.630 8
SrO 0.562 9	K_2O 0.620 9
Na_2O 0.555 10	Na_2O 0.564 10
BaO 0.547 11	Fe_2O_3 0.564 11
SO_2 0.531 12	SnO_2 0.446 12
SnO_2 0.438 13	SO_2 0.436 13

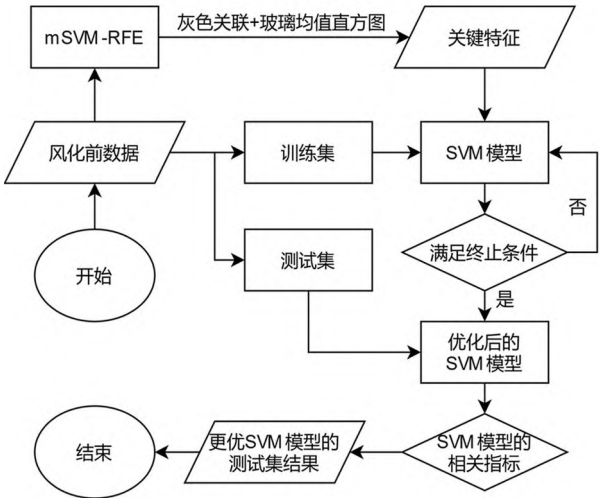


图 1 SVM 模型算法流程图

1.3 SVM 模型

为消除风化对分类结果的影响,用风化前的化学成分数据对玻璃制品进行分类. 考虑到玻璃数据样本量较少且自变量维数较高,而 SVM 模型可以做到低维计算,等效高维表现,所以决定用支持向量机(SVM)^[9]对玻璃制品进行分类. 图 1 为 SVM 模型的算法流程图.

由于化学成分数目较多,先利用 R 实现 mSVM-RFE^[10]判别并选择出关键特征. 表 5 为各成分排名. 由表 5 可知,关键特征是 CuO 、 BaO 、 PbO 、 Fe_2O_3 和 K_2O . 可是由图 2 可知, SiO_2 和 Al_2O_3 在玻璃类型的分类中差别不太明显,故决定先剔除这两个化学成分. 而 K_2O 、 Fe_2O_3 、 PbO 、 BaO 在高钾玻璃与铅钡玻璃中差别较大. 同时灰色关联分析辅助了 SVM 降维,1.2 节分析中在两种玻璃类型中 SiO_2 与 PbO 、 BaO 的关联度有一定的差异. 所以在玻璃分类中优先考虑元素钾、铁、铅、钡.

先构建 PbO 与 BaO 的 SVM 模型,使用线性核函数. 再构建 PbO 与 K_2O 的 SVM 模型,也使用线性核函数. 两个模型的相关指标见表 6. 由表 6 可知, PbO 与 BaO 的 SVM 模型是优于 PbO 与 K_2O 的 SVM 模型,所以采用 PbO 与 BaO 的 SVM 模型来分类风化前的玻璃制品.

表 5 各特征的排序

FeatureName	AvgRank	FeatureName	AvgRank
CuO	2.5	SO_2	7.4
BaO	3.6	P_2O_5	7.8
PbO	4.4	SrO	9.9
Fe_2O_3	4.6	Al_2O_3	10
K_2O	4.8	MgO	11
Na_2O	6.1	SnO_2	12.2
SiO_2	7.1	CaO	13.6

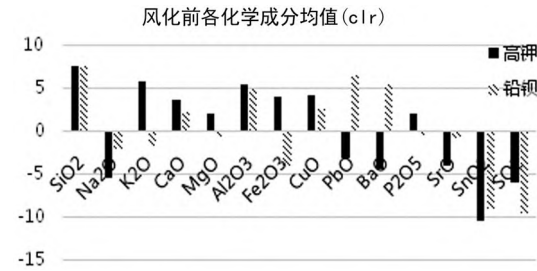


图 2 高钾玻璃与铅钡玻璃均值的直方图

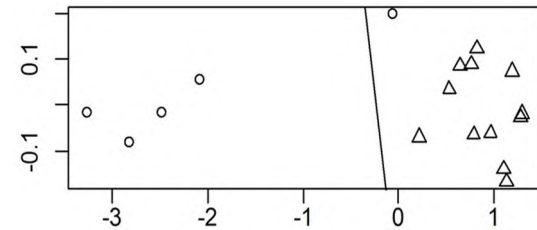


图 3 测试集的分类二维散点图

表 6 两线性 SVM 模型的相关指标

模型	accuracy	P-value	kappa	sensitivity	specificity	AUC
PbO 与 BaO	0.9412	0.065	0.8496	0.9231	1.00	0.900
PbO 与 K_2O	0.8824	0.200	0.6731	0.9231	0.75	0.837

图 3 是铅、钡 SVM 模型的二维散点图,其中三角形的为铅钡玻璃样本点,圆形的为高钾玻璃样本点. 将

已知分类的采样点数据分为训练集和测试集,用训练集数据对 SVM 模型进行训练,将训练好的 SVM 模型用测试集进行检验.由图 3 可知,测试集中高钾玻璃分对 4 个,错分 1 个;铅钡玻璃分对 12 个,错分 0 个.分错的样本是 03 部位 2.测试集 SVM 模型的准确率为 94%.

最后用此模型对 8 个未分类的玻璃制品进行分类,结果是 3 个样本分为高钾玻璃,5 个样本分为铅钡玻璃.

2 结论

古代玻璃类型和化学成分有关,风化会对分类结果有影响.本文预测了风化前的化学成分,用风化前的化学成分对古代玻璃制品进行了分类.首先通过系统聚类区分出风化点,并认为 clr 变换后数据服从正态分布,由此预测出风化前化学成分含量.然后通过灰色关联得知高钾玻璃中有较强关联度的是 SiO_2 与 Al_2O_3 ,铅钡玻璃中有较强关联度的为 SiO_2 与 Al_2O_3 、 PbO .其差异是在 SiO_2 与 PbO 、 BaO 的关联度方面,铅钡玻璃关联度高,而高钾玻璃低.最后,通过构建 SVM 模型对未知分类的玻璃制品进行分类.这种方法对考古工作中文物类型鉴别有一定的参考意义.该方法除了可以区分高钾玻璃和铅钡玻璃,还能推广应用到其他类型玻璃的分类上.将来的工作考虑进一步细分玻璃类型,例如结合偏最小二乘法判别分析^[11]对古代玻璃制品进行亚类划分.

参考文献:

- [1] 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.
- [2] 赵德云.中国出土的蜻蜓眼式玻璃珠研究[J].考古学报,2012(2):177—216.
- [3] 王承遇,陶瑛.硅酸盐玻璃的风化[J].硅酸盐学报,2003(1):78—85.
- [4] 宋燕,马清林.宁夏固原北周田弘墓出土玻璃残片研究[J].玻璃与搪瓷,2008(2):35—42+45.
- [5] 程金树,刘启明.提高浮法玻璃化学稳定性的研究与进展[J].玻璃,2008(2):9—11.
- [6] 薛毅,陈立萍. R 语言实用教程[M].北京:清华大学出版社,2014:306—318.
- [7] HRON K, MENAFOGLIO A, TEMPL M, et al. Simplicial principal component analysis for density functions in Bayes spaces[J]. Computational Statistics and Data Analysis, 2016(94):330—350.
- [8] 刘思峰,蔡华,杨英杰,等.灰色关联分析模型研究进展[J].系统工程理论与实践,2013,33(8):2041—2046.
- [9] 牟丹,王祝文,黄玉龙,等.基于 SVM 测井数据的火山岩岩性识别——以辽河盆地东部拗陷为例[J].地球物理学报,2015,58(5):1785—1793.
- [10] 王君.基于 SVM-RFE 的特征选择方法研究[D].大连:大连理工大学,2015.
- [11] 孙立丽,王萌,任晓亮.化学模式识别方法在中药质量控制研究中的应用进展[J].中草药,2017,48(20):4339—4345.

Classification of Ancient Glass Products Based on Support Vector Machine (SVM)

GAO Guoyun, WANG Qingyun

(School of Mathematics and Computer Science, Gannan Normal University, Ganzhou 341000, China)

Abstract: Ancient glass products are one of the commodities traded on the ancient Silk Road. Ancient glass products were generally classified based on their chemical composition, but weathering can affect the identification of glass types. Therefore, this article attempts to predict the chemical composition before weathering to eliminate the influence of weathering, then use grey correlation analysis to analyze the correlation and differences of chemical composition, and finally establish a support vector machine (SVM) model to classify ancient glass products.

Key words: Support Vector Machine(SVM); hierarchical clustering; GR; ancient glass; glass weathering