

基于集成特征选择和随机森林的古代玻璃分类模型

路佳佳

(山西工商学院计算机信息工程学院, 太原 030006)

摘 要: 将机器学习中的特征选择方法和分类算法融入古代玻璃制品成分分析和类别鉴定问题研究, 以准确率和 AUC 作为分类性能度量指标, 尝试构造古代玻璃制品化学成分选择的集成特征选择模型和鉴定分类的随机森林模型。对不同特征选择方法的结果进行集成, 选择重要的化学成分, 对选出的重要特征结合随机森林模型、 k 近邻学习和 Naive Bayes 模型等方法进行实验分析。结果表明, 采用集成特征选择分析出氧化铅、氧化钡、氧化钾等成分对玻璃表面风化影响比较显著, 且高钾玻璃中这 3 种成分两两关联很大, 对选出的重要特征应用基于 k 折交叉验证的随机森林模型进行分类得到的准确率较高, 模型稳定性强。该方法可以为我国古代玻璃制品的成分分析和类别鉴定提供理论参考, 对其它玻璃的相似研究也有一定程度的借鉴意义。

关键词: 机器学习; 随机森林; k 近邻学习; 分类算法

中图分类号: TQ172.75 文献标志码: A 文章编号: 0454-5648(2023)04-1060-06

网络出版时间: 2023-03-07



Classifying Model of Ancient Glass Products Based on Ensemble Feature Selection and Random Forest

LU Jiajia

(Computer Information and Engineering Academy, Shan Xi Technology and Business College, Taiyuan 030006, China)

Abstract: This paper was to construct ensemble feature selection and random forest for the identification of ancient glass products via integrating the machine learning algorithm into the identification and analysis of ancient glass products and taking accuracy rate and AUC as the measurement indexes of classification performance. The results of different feature selection methods were analyzed, and the important chemical components were selected. The selected important features with random forest, k -nearest neighbor learning and naive Bayesian were investigated. The results show that lead oxide, barium oxide, and potassium oxide have an impact on the weathering of the glass surface by using ensemble feature selection. In high potassium glass, three components are closely related, the accuracy of classification by the random forest based on the k -fold cross validation for selected important features is great, and the model is stable. This method can provide a theoretical reference for the composition analysis and category identification of ancient glass products, and other glasses.

Keywords: machine learning; random forest; k -nearest neighbor; classification algorithm

早在战国时期我国已经能够制造花纹非常精致的玻璃, 并且古代丝绸之路作为中西方文化交流的重要通道, 促进了中国古代玻璃的发展和传播^[1]。我国最早的玻璃类型有钾硅酸盐玻璃、铅钡硅酸盐玻璃、高钾硅酸盐玻璃等, 但由于古代玻璃在埋藏过程中容易受到环境变化的影响而风化^[2], 导致玻

璃的成分比例发生变化, 因此影响对其所属类别的准确判断。

目前关于古代玻璃的研究有的侧重在具体出土, 详细出土器物, 制造技术^[3]等方面, 有的研究关注不同玻璃中某种重要化学成分的鉴定和分析^[4-6], 以及在工业生产中的应用^[7]。统计中的回归

收稿日期: 2022-09-24。 修订日期: 2022-10-26。

基金项目: 山西省教育科学“十四五”规划课题(GH-21400)。

第一作者: 路佳佳(1988—), 女, 硕士, 讲师。

Received date: 2022-09-24. Revised date: 2022-10-26.

First author: LU Jiajia (1988—), female, Master, Lecturer.

E-mail: luyoucan1988@163.com

分析^[8], 线性模型^[9], 主成分分析法^[5]等方法也被应用到玻璃制品研究中, 而机器学习^[10]中的分类模型, 如随机森林^[11], 特征选择方法^[12]在古代玻璃中的研究很少, 但文献[13–14]利用特征选择对玻璃性能进行预测研究, 表明特征选择方法有一定优势, 为此, 应用特征选择方法对古代玻璃主要成分进行研究, 并且提出基于不同学习器的集成特征选择方法, 通过对古代玻璃文物表面风化化学成分进行主成分分析, 并且结合玻璃类型, 建立 logistic^[15]模型分析古代玻璃有无风化表面化学成分中的主要影响因素, 再应用决策树中的变量重要性值和随机森林中表示变量重要性的 MeanDecreaseGini 的结果, 集成特征选择。根据选择的主要化学成分, 利用机器学习中的基于 k 折交叉验证^[16]的方法构建随机森林, 再结合 k 近邻学习, Naive Bayes 等方法对古代玻璃进行分类。并且针对类别未知的古代玻璃文物, 结合其化学成分对其进行正确分类。因此该方法对于我国古代玻璃制品的成分分析研究和玻璃类别鉴定有重要的现实意义。

1 模型构造

1.1 主成分分析

为更好地对玻璃表面化学成分含量进行特征选择, 首先引入主成分分析法, 该方法是从 p 个原始的化学成分中选择 m 个主成分, 各个主成分相互独立, 并且每个主成分都是原来某些成分的线性组合。

对提取的主成分根据特征值的大小进行排序, 特征值最大的主成分对玻璃类别的解释性强, 当选择的 m 个主成分的化学成分含量的方差或累计贡献率能够超过 80%, 说明选择的 m 维主成分构成的空间可以最大限度地保留原始 p 个化成成分的信息^[17]。具体过程如下。

1.1.1 将原始数据标准化 假设有 M 个古代玻璃类型的样本, 每个样本有 N 个特征, 则数据集 $\mathbf{X}$

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ x_{M1} & x_{M2} & \cdots & x_{MN} \end{pmatrix} = (\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N) \quad (1)$$

一般对原始数据应用 Z-score 进行标准化处理, 记标准化后的由古代玻璃表面化学成分构成的矩阵为 $\mathbf{V}$ 。

1.1.2 计算协方差矩阵的特征值和特征向量 主

成分分析需要分析各个化学成分构成的协方差矩阵, 由于标准化后协方差矩阵与相关系数矩阵 $\mathbf{R}$ 相等, 记 $\mathbf{R} = (r_{ij})_{N \times N}$ 。使用 Jacobi 方法对特征方程 $|\lambda \mathbf{E} - \mathbf{R}| = 0$ 进行求解, 得到特征值, 应用 $|\mathbf{R} - \lambda \mathbf{E}| \mathbf{U} = 0$ 可得特征向量矩阵 $\mathbf{U}$ 。因子得分矩阵 $\mathbf{F} = \mathbf{U}^T \mathbf{V}$ 。

$$\mathbf{F} \mathbf{F}^T = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \quad (2)$$

1.1.3 计算主成分贡献率和累计贡献率 第 m 个主成分的方差贡献率为:

$$c_m = \frac{\lambda_m}{\sum_{i=1}^N \lambda_i}$$

前 m 个主成分的累计方差贡献率为:

$$c(m) = \frac{\sum_{j=1}^m \lambda_j}{\sum_{i=1}^N \lambda_i}$$

第 m 个成分的方差贡献率可以表示该成分的方差在总方差中所占的比例, 累计方差贡献率越大说明前 m 个成分包括的信息量越多。

1.1.4 选出主成分 为尽可能保留原始化学成分含量更多的信息, 确定主成分的个数应该尽量使得累计贡献率更大。

选择的 m 个主成分可以看成将特征向量矩阵 $\mathbf{U}$ 分成 2 部分, 即

$$\begin{aligned} \mathbf{U}_{N \times N} &= (\mathbf{U}_{N \times m}^{(1)}, \mathbf{U}_{N \times (N-m)}^{(2)}) \\ \mathbf{V} &= \mathbf{U} \mathbf{F} = (\mathbf{U}^{(1)}, \mathbf{U}^{(2)}) (\mathbf{F}^{(1)}, \mathbf{F}^{(2)})^T = \\ &\quad \mathbf{U}_{N \times m}^{(1)} \mathbf{F}_{m \times M}^{(1)} + \mathbf{U}_{N \times (p-m)}^{(2)} \mathbf{F}_{(p-m) \times M}^{(2)} \end{aligned}$$

这里 $\mathbf{U}^{(1)} \mathbf{F}^{(1)}$ 表示主成分可以解释的信息, 而 $\mathbf{U}^{(2)} \mathbf{F}^{(2)}$ 表示其余的信息。根据因子得分矩阵 $\mathbf{F}$, 化学成分 $\mathbf{V}_t$ 可以表示为:

$$\mathbf{V}_t = \mathbf{U}_{t1} \mathbf{F}_1 + \mathbf{U}_{t2} \mathbf{F}_2 + \cdots + \mathbf{U}_{tm} \mathbf{F}_m \quad (3)$$

1.2 logistic 模型

logistic 模型^[15]是基于概率建模, 通过 Sigmoid 函数将线性回归模型的输出值压缩到 (0,1) 之间进行预测。设 x_i 为自变量, 表示古代玻璃表面化学成分中的第 i ($i=1, 2, \cdots, n$) 个成分; y 是因变量表示玻璃的类型; β_i 表示各自变量的系数; β_0 为截距项, 可建立如下线性模型。

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (4)$$

为使得线性关系更加显著, 突出模型能够更好的解释随机因素的变化导致 y 也发生显著的变化, 对上式的因变量 y 进行对数变换, 右面依然保持线性关系, 可以得如下形式。

$$\log(y) = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (5)$$

考虑到概率和几率存在正向关系, 用几率 odd 取代概率来描述随机事件发生的可能性。设 A 表示古代玻璃类型 $y=1$, P 描述 A 发生的概率, 并且将对数变为自然对数, 可得式(6)和式(7):

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (6)$$

$$P = \frac{\exp(\beta_0 + \sum_{i=1}^n \beta_i x_i)}{1 + \exp(\beta_0 + \sum_{i=1}^n \beta_i x_i)} \quad (7)$$

1.3 集成特征选择

由于单个学习器的结果的不鲁棒性, 结合多个学习器的集成学习的方法被提出, 基本思想是应用一个分类模型(回归模型)来融合多个分类(回归)模型。在分类问题中, 采用投票法选择分类结果。对于回归问题, 对输出结果加权平均可得最后输出值。一般集成学习具有比个体学习器更好的泛化性能^[10]。

应用不同特征选择方法进行特征提取会得到不同的选择结果, 本文研究的集成特征选择是对不同方法得到的特征选择结果进行集成可以得到更加合理的特征选择结果。具体集成特征选择的思路如下。

假设古代玻璃文物表面各个化学成分作为特征是否被选择构成的向量为 $(te_1, te_2, \dots, te_N)^T$, 其中 te_i 表示第 i 个特征, 如果 te_i 被选为重要的特征, te_i 取 1, 即 $te_i = 1$, 反之 $te_i = 0$ 。假设有 g 种特征选择方法, 记第 s 种方法的特征选择结果为 J_s 。 J_s 为一个列矩阵, 那么将所有特征选择的方法综合到一起可以构成一个特征选择矩阵 T 。

$$T = (J_1 \ J_2 \ \cdots \ J_s) = \begin{pmatrix} te_{11} & te_{21} & \cdots & te_{s1} \\ te_{12} & te_{22} & \cdots & te_{s2} \\ \vdots & \vdots & \ddots & \vdots \\ te_{1N} & te_{2N} & \cdots & te_{sN} \end{pmatrix} \quad (8)$$

T 的某一行表示某种特征用不同方法得到的特征选择结果, 对特征选择矩阵 T 的各行进行求和, 比如第 i 行求和为 $te_{1i} + te_{2i} + \cdots + te_{si}$, 因此可得 N 个值, 对这 N 个结果按照从大到小的顺序进行排列, 前 m 个即为集成特征选择的结果。

1.4 随机森林模型

为对不同的玻璃类别选择合适的方法进行分类, 首先考虑应用随机森林模型进行分类。随机森林是包含多个分类树的集成算法, 一般情况下对数据集进行随机采样, 选用 k 折交叉验证^[18]的方法, 将数据集分为不相交的训练集和测试集。

对于离散型随机变量而言, 随机森林进行分类的方法是假设输入的数据集为 D 。通过对各分类树的结果进行投票, 把多数的投票结果作为最终分类结果, 即

$$L(x) = \arg \max \sum_{i=1}^n I[P_i(x) = y] \quad (9)$$

1.5 应用 k 折交叉验证的随机森林

k 折交叉验证^[18]是把容量大小为 n 数据集 D_n 划分为 k 份, 每份互不相交且数据集大小几乎相等。假设划分后的 k 个数据集表示为 $F_1, F_2, \dots, F_k$, 用 $|F_t| = m$ 表示第 t 个数据集的大小, 则 $n = km$, 依次将经过切分的 k 折中的每一折作为验证集, 其余 $k-1$ 折作为训练集, 建立 k 个随机森林模型。设每个随机森林模型分类的结果为 $L_k(x)$, 把 k 个分类结果记为 $R(x)$ 。

$$R(x) = \{L_1(x), L_1(x), \dots, L_k(x)\}$$

最后用 $R(x)$ 的众数 $\text{mode}(R(x))$ 表示 k 折交叉验证的随机森林模型预测结果。

$$\text{pred}_k(x) = \text{mode}(R(x)) \quad (10)$$

2 结果与讨论

2.1 样本采集与数据处理

此次研究数据来自 2022 年全国大学生数学建模竞赛官网提供的一批我国古代玻璃制品的相关数据。考古工作者通过专业的技术手段已经将这些文物样品的化学成分含量比例和玻璃类型给出, 玻璃类型为高钾玻璃和铅钡玻璃。对收集到的数据中各化学成分比例累加介于 85%~100% 之间的数据视为有效数据, 将数据中没有检测到的化学成分记比例记为 0。

2.2 实证过程和结果分析

首先应用 python 和统计软件 R 对数据采用 4 种

方法进行特征选择，分别为主成分分析法，logistic 模型，决策树，随机森林，对得到的特征选择结果进行集成特征选择，再结合分类模型进行分类。表 1 给出各变量与化学物对应关系。

表 1 各变量与化学成分对应表
Table 1 Corresponding table of each variable and chemical composition

Variable	Composition
x_1	Silicon dioxide
x_2	Sodium oxide
x_3	Potassium oxide
x_4	Calcium oxide
x_5	Magnesium oxide
x_6	Alumina
x_7	Ferric oxide
x_8	Copper oxide
x_9	Lead oxide
x_{10}	Barium oxide
x_{11}	Phosphorus Pentoxide
x_{12}	Strontium oxide
x_{13}	Stannic oxide
x_{14}	Sulfur dioxide

2.2.1 主成分分析结果 图 1 是主成分分析的特征根图，图 2 是具体的因子分析图。从图 1 和图 2 可以看出不同特征的重要程度。因子分析图中第一行对应的-0.01 表示第 1 个化学成分 x_1 ，则第 13 行对应的 0.014 表示第 14 个化学成分 x_{14} ，根据因子分析图可选出的重要特征为 x_8 、 x_9 、 x_{10} 、 x_{13} 、 x_{14} ，对应的化学成分分别为氧化铜、氧化铅、氧化钡、氧化锡、二氧化硫。

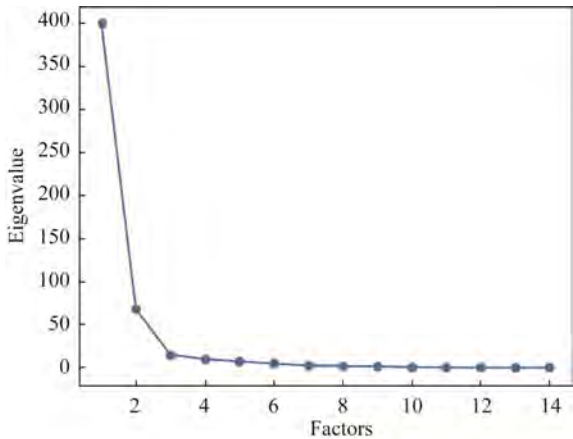


图 1 特征根图
Fig.1 Eigenvalue plot

2.2.2 Logistic 模型估计 表 2 给出具体的 Logistic 模型参数估计值，在应用 Logistic 模型时，结合逐步回归方法选择重要影响因素，变量 x_1 、

x_7 、 x_9 、 x_4 的 P 值都小于 0.05。因此化学成分二氧化硅，氧化铁，氧化铅，氧化钙对风化的影响都比较显著。

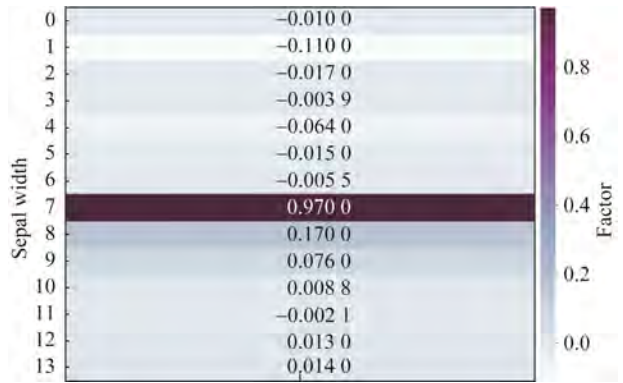


图 2 因子分析图
Fig. 2 Factor analysis plot

表 2 Logistic 模型估计
Table 2 Logistic model estimation

	Estimated value	Standard deviation	Pr(> z)
Constant	-16.030 0	6.635 7	0.015 7*
x_1	0.176 4	0.071 2	0.013 2*
x_7	2.329 8	1.044 8	0.025 8*
x_9	0.525 4	0.196 2	0.007 4**
x_4	-0.989 6	0.476 8	0.037 9*
x_5	-0.340 3	0.929 4	0.714 2

2.2.3 随机森林 (MeanDecreaseGini) MeanDecreaseGini 表示变量改变后 Gini 系数降低，它能够用来比较变量的重要性，该值越大表示该变量的重要性越大。用 R 软件中的 RandomForest 包实现，表 3 给出具体的值。从表 3 中可以发现， x_9 (氧化铅)、 x_{10} (氧化钡)、 x_3 (氧化钾)MeanDecreaseGini 值相比其它变量更大，说明选出的重要特征符合实际。

表 3 MeanDecreaseGini 值
Table 3 MeanDecreaseGini value

Variable	MeanDecreaseGini
x_1	1.817 691 97
x_2	0.043 135 53
x_3	3.706 863 24
x_4	1.351 050 60
x_5	0.449 696 75
x_6	1.073 914 60
x_7	0.540 867 55
x_8	1.381 556 57
x_9	3.498 564 97
x_{10}	2.296 866 35
x_{11}	0.065 679 64
x_{12}	1.360 197 85
x_{13}	0.109 217 47
x_{14}	0.290 549 79

2.2.4 决策树中变量重要性值 再应用 R 中的 rpart 包计算每个变量的重要性值, 具体如表 4。由表 4 可见, 变量 x_9 (氧化铅), x_{10} (氧化钡), x_3 (氧化钾)变量重要性值影响最大。

表 4 变量重要性值
Table 4 Variable importance value

Variable	Importance value
x_9	6.379 598 6
x_1	3.673 548 4
x_8	1.451 247 2
x_{10}	5.929 408 1
x_{12}	2.835 377 2
x_3	4.253 065 7
x_4	2.480 950 0
x_6	0.161 249 7
x_{14}	0.483 749 1

2.2.5 集成特征选择结果 对应用 4 种方法选出的重要特征进行集成特征选择, 得到的特征选择矩阵 T , 其中第 1 列为主成分分析法对应的 14 种化学成分是否被选择的结果, 第 2, 3, 4 列分别为 logistic 模型, 随机森林和决策树进行特征选择的结果。

$$T = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad (11)$$

对 T 中的各行元素求和, 选择得到的值大于等于 2 的为集成特征选择的结果, 每一行与各个化学成分相对应, 比如第一行对应为化学成分 x_1 , 第 i 行对应为化学成分 x_i 。通过对各行求和可得 x_9 (氧化铅)对应行的和为 4, 也就是为选出的化学成分中最为重要的化学成分。 x_{10} (氧化钡), x_1 (二氧化硅), x_4 (氧化钙), x_8 (氧化铜)对应行的和为 3, 说明这些化学成

分对玻璃的类型比较重要。 x_3 (氧化钾), x_6 (氧化铝), x_{12} (氧化锶)对应行的和为 2, 说明这 4 个化学成分对玻璃有一定程度的影响。

2.2.6 不同分类模型的准确率比较 通过应用 python 中的 sklearn 包, 采用 k 折交叉验证对数据进行切分, 在随机森林模型和 k 近邻学习模型上进行训练, 得到分类准确率如下表 5。从模型分类准确率比较可以发现, 随机森林分类结果准确率更高, 模型更加稳定, 但是 k 近邻和随机森林模型分类结果容易受到参数变化的影响。

表 5 不同分类器准确率比较

Table 5 Comparison of accuracy of different classifiers

Model	Accuracy			
Random forest	0.95	0.90	0.85	0.80
k -Nearest neighbor	0.86	0.79	0.93	0.71

AUC 是通过对 ROC 曲线下面各部分的面积求和得到。改变随机森林模型中的参数 Ntree 和 Mtry, 得到不同的 ROC 曲线, 不同的 AUC 值, 结合不同的 AUC 分析这些属性变化对随机森林模型的影响。表 6 考虑随着参数 Ntree 和 Mtry 的变化模型 AUC 的变化情况。通过结果可以发现, 参数 Ntree 和 Mtry 对分类模型都比较敏感。

表 6 随着 Ntree 和 Mtry 变化导致 AUC 的变化

Table 6 Changes with Ntree and Mtry lead to changes in AUC

Ntree	AUC	Mtry	AUC
10	0.788	2	0.800
20	0.833	3	0.667
30	0.732	4	0.875
50	0.750	5	0.912
80	0.780	6	0.567
100	0.775	7	0.900
250	0.778	8	0.804
300	0.833	9	0.694
400	0.867	10	0.960
500	0.750	11	0.867

2.2.7 应用分类模型进行预测 把未知类别玻璃文物的化学成分数据作为测试集, 输入已经建立随机森林模型, k 近邻学习模型, Naive Bayes 分类器分别进行分类结果的预测。具体预测结果如下表 7。随机森林模型和 KNN 模型的分类结果相同, 并且准确率为 100%, Naive Bayes 分类器的预测准确率为 62.5%。

表 7 分类预测结果
Table 7 Classification forecast results

Cultural relic number	Predict results		
	Random Forest	Naive Bayes	k-Nearest Neighbor
A1	High potassium	High potassium	High potassium
A2	Lead barium	Lead barium	Lead barium
A3	Lead barium	Lead barium	Lead barium
A4	Lead barium	Lead barium	Lead barium
A5	Lead barium	Lead barium	Lead barium
A6	High potassium	Lead barium	High potassium
A7	High potassium	Lead barium	High potassium
A8	Lead barium	High potassium	Lead barium

3 结论

古代玻璃制品表面是否风化与其玻璃类型、颜色、纹饰、化学成分含量等数据进行分析的研究中很多基于传统的实验手段展开, 本工作充分利用计算机软件 python^[19], 统计软件 R, 有效结合机器学习方法中的主成分分析, 决策树特征选择, logistic 等方法对古代玻璃中的化学成分进行集成特征选择, 再应用随机森林, k 近邻学习, Naive Bayes 分类器等方法进行古代玻璃类型的分类。选择合适的化学成分对高钾玻璃和铅钡玻璃进行分类能够更好的为古代玻璃的成分和鉴定研究提供一些模型和方法。通过几种不同分类模型的对比, 发现随机森林模型的准确率更高。该研究对于我国古代玻璃制品的成分分析研究和玻璃类别鉴定有重要的现实意义。

参考文献:

- [1] 干福熹. 古代丝绸之路和中国玻璃[J]. 自然杂志, 2006, 28(5): 253–260.
GAN Fuxi. Chin J Nat (in Chinese), 2006, 28(5): 253–260.
- [2] 王承遇, 陶瑛. 硅酸盐玻璃的风化[J]. 硅酸盐学报, 2003, 31(1): 78–85.
WANG Chengyu, TAO Ying. J Chin Ceram Soc, 2006, 28(5): 253–260.
- [3] 程朱海. 试探我国古代玻璃的发展[J]. 硅酸盐学报, 1981, 9(1): 96–106.
CHENG Zhuhai. J Chin Ceram Soc, 1981, 9(1): 96–101.
- [4] 李青会, 周虹志, 贵教珍, 等. 一批中国古代玻璃珠化学成分的检测报告[J]. 江汉考古, 2005, 4: 81–88.
LI Qinghui, ZHOU Hongzhi, GUI Jiaozhen, et al. Jiangnan Archaeol (in Chinese), 2005, 4: 81–88.
- [5] 崔剑锋, 吴小红, 谭远辉, 等. 湖南沅水流域战国时期楚墓出土古代玻璃器的成分分析[J]. 硅酸盐学报, 2009, 37(11): 1910–1918.
CUI Jianfeng, WU Xiaohong, TAN Yuanhui, et al. J Chin Ceram Soc, 2009, 37(11): 1910–1918.
- [6] 李飞, 李青会, 干福熹, 等. 一批中国古玻璃化学成分的质子激发 X 射线荧光分析[J]. 硅酸盐学报. 2005, 33(5): 581–586.
LI Fei, LI Qinghui, GAN Fuxi, et al. J Chin Ceram Soc, 2005, 33(5): 581–586.
- [7] 侯宝晰. 窗玻璃结石岩相鉴定及其在生产上的应用[J]. 硅酸盐学报, 1965, 4(2): 105–108.
HOU Baoxi. J Chin Ceram Soc, 1965, 4(2): 105–108.
- [8] 张谦, 何涌, 吕彦杰, 等. 微晶玻璃中物相含量的回归分析计算[J]. 分析测试学报, 2008, 27(4): 415–418.
ZHANG Qian, HE yong, LYU Yanjie, et al. J Instrum Anal (in Chinese), 2008, 27(4): 415–418.
- [9] 吴继祖. 含铅钡玻璃的分析稀土湿法冶金浸出液的分析[J]. 理化检验通讯(化学分册), 1977(Z3): 79–81.
WU Jizu. Phys Testing Chem Analysis: Chem Anal (in Chinese), 2005, 33(5): 581–586.
- [10] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 33–37.
- [11] 张家伟, 郭林明, 杨晓梅. 针对不平衡数据的过采样和随机森林改进算法[J]. 计算机工程与应用, 2020, 1(11): 39–45.
ZHANG Jiawei, GUO Linming, YANG Xiaomei. Comput Eng Appl (in Chinese), 2020, 1(11): 39–45.
- [12] 吴喜之. 应用回归及分类——基于 R[M]. 北京: 中国人民大学出版社, 2016: 50–70.
- [13] LIU Yue, ZHAO Tianlu, YANG Guang. The onset temperature (T_g) of As_xSe_{1-x} glasses transition prediction: A comparison of topological and regression analysis methods[J]. Comput Mater Sci, 2017, 140: 315–321.
- [14] LIU Yue, WU Junming, YANG Guang. Predicting the onset temperature (T_g) of Ge_xSe_{1-x} glass transition: A feature selection based two-stage support vector regression method[J]. Sci Bull, 2019, 64: 1195–1203.
- [15] 韦丞婧, 李国东. 基于超混沌系统和 Logistic 映射的视频图像加密设计[J]. 计算机工程, 2022, 48(5): 263–271.
WEI Chengjing, LI Guodong. Comput Eng (in Chinese), 2022, 48(5): 263–271.
- [16] RODRIGUEZ J D, PEREZ A, LOZANO J A. Sensitivity analysis of k-fold cross validation in prediction error estimation[J]. IEEE Trans Pattern Anal Mach Intell, 2010, 32(7): 569–575.
- [17] 张文彤. SPSSII 统计分析教程[M]. 北京: 北京希望电子出版社, 2002: 190–202.
- [18] YOSHUA Bengio, YVES Grandvalet. No unbiased estimator of the variance of k-fold cross-validation[J]. J Mach Learn Res, 2004, 5: 1089–1105.
- [19] 吴喜之. Python-统计人的视角[M]. 北京: 中国人民大学出版社, 2018: 81–88.