

基于近红外光谱技术的茶叶新旧鉴别及产地溯源研究

孙晓荣^{1,2}, 胡毅然^{1,2*}, 刘翠玲^{1,2}, 张善哲^{1,2}, 王子涵^{1,2}

(1. 北京工商大学人工智能学院, 北京 100048; 2. 北京工商大学, 食品安全大数据技术北京市重点实验室, 北京 100048)

摘要: **目的** 建立基于近红外光谱的定性分析模型, 实现对茶叶的新旧分类和产地溯源。**方法** 首先采用傅立叶近红外光谱仪采集茶叶样品的漫反射光谱数据, 然后使用卷积(Savitzky-Golay, S-G)平滑算法和数据标准化(Normalization)对光谱数据进行预处理, 最后基于遗传优化算法(genetic algorithm, GA)和粒子群优化算法(particle swarm optimization, PSO)分别建立了优化向量机模型(support vector machine, SVM), 从而实现新旧茶叶的分类以及产地溯源。**结果** 与 GA-SVM 模型相比, PSO-SVM 模型的建模效果较好, 且分类时间更短, 在新旧鉴别和产地溯源实验中都达到了 100% 的预测精度。**结论** 基于近红外光谱建立的 PSO-SVM 模型可以实现茶叶新旧的判别以及产地溯源, 为鉴别茶叶年份和追踪茶叶产地提供了理论支撑和技术指导。

关键词: 近红外光谱技术; 茶叶新旧; 产地溯源

Identification of new and old tea and origin traceability based on near infrared spectroscopy

SUN Xiao-Rong^{1,2}, HU Yi-Ran^{1,2*}, LIU Cui-Ling^{1,2}, ZHANG Shan-Zhe^{1,2}, WANG Zi-Han^{1,2}

(1. School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China; 2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China)

ABSTRACT: Objective To establish the qualitative analysis model by near infrared spectroscopy, realize the new and old classification and the origin traceability of the tea. **Methods** Firstly, Fourier near-infrared spectrometer as used to acquire diffuse reflectance spectral data of tea samples, and then the spectral data was pretreated using Savitzky-Golay (S-G) smoothing and data Normalization. Finally, based on genetic optimization algorithm (GA) and particle swarm optimization (PSO), support vector machine (SVM) models were established, respectively. Thus to realize the classification of old and new tea and origin traceability. **Results** Compared to GA-SVM model, PSO-SVM model had better modeling effects and shorter classification times, which displayed as 100% prediction accuracy in both the old and new identification and origin traceability experiments. **Conclusion** The PSO-SVM

基金项目: 北京市自然科学基金资助项目(4222043)、2021 年教育部高教司产学研合作协同育人项目(202102341023)、2022 年北京工商大学研究生教育教学改革专项

Fund: Supported by the Beijing Natural Science Foundation (4222043), the 2021 Industry-university Cooperative Education Project of Higher Education Department of Ministry of Education (202102341023), and the 2022 Beijing Technology and Business University Postgraduate Education Teaching Reform Special

*通信作者: 胡毅然, 硕士研究生, 主要研究方向为食品安全检测。E-mail: hyr18801449369@163.com

*Corresponding author: HU Yi-Ran, Master Student, School of Artificial Intelligence, Beijing Technology and Business University, No.11, Fucheng Road, Haidian District, Beijing 100048, China. E-mail: hyr18801449369@163.com

model set by the near infrared spectroscopy can realize the new and old classification and the origin traceability of the tea, which provides theoretical support and technical guidance for the identification of tea years and traceability of tea origin.

KEY WORDS: near infrared spectroscopy; new and old tea; origin traceability

0 引言

中国的茶叶消费群体非常庞大,人们对茶叶的品质要求也在向多元化发展,如茶叶的年份、产区等。然而国内市场缺乏对有关方面的有效监管,冒充老茶、名茶的事件时有发生。有一些黑心商家为了谋取暴利,把廉价的陈旧茶叶更换包装,成为新产的茶来销售,谋求非法利润。还有的商贩通过将散装劣质茶叶更换包装来冒充名茶,甚至回收废茶进行简单加工再高价卖出,不但威胁着消费者的切身利益,也损害了中国茶叶的品牌声誉^[1]。因此,研究快速、准确的鉴别方法,对于整治市场乱象、提高茶叶品质具有重要现实意义。

国内外对于茶叶品质的鉴定有多种分析方法,我国评定茶叶等级的国家标准主要遵循感官评审法,但茶叶感官评审法受到评审者主观好恶影响缺乏客观性,且耗时耗力,难以规模化应用于生产实践;而化学检测技术受限于其成本高、操作复杂、损伤样品的特点,难以在市场推广普及。现行的传统检测方法大都自身具有局限性,检测效率低下、检测成本高昂且大部分为有损检测^[2-3]。

近红外光谱技术(near infrared spectroscopy, NIRS)是一项成熟的检测技术,近年来凭借其操作简单、对样品无损等优点快速在市场普及。NIRS 逐步应用于茶叶品质鉴定,并凭借其独特的优势获得了良好的市场反馈^[4-5]。例如孟维君^[6]利用 NIRS 结合多变量分析不同产地的铁观音样本,建立偏最小二乘判别分析(partial least squares discriminant analysis, PLS-DA)模型,正确度 85.5%还有继续提升的空间,该方法结合核磁共振技术与多变量分析应用于白茶年份鉴定,可以具体找到多种主要差别物质,但检测时间长、受外界影响大;周健等^[7]扫描了西湖龙井和浙江龙井,主成分分析(principal component analysis, PCA)聚合度效果不理想,但偏最小二乘回归(partial least squares, PLS)模型预测精度可达 100%,说明模型建立对于产地溯源具有重要的决定意义;周昌海等^[8]在近红外波段结合 PLS,建立 20 种不同浓度的茶叶近红外光谱和咖啡碱浓度的数学模型,结果表明近红外光谱法可作为一种快速有效检测茶叶品质的方法。国外也有很多相关的研究成果。印度 DAS 等^[9]基于 NIRS 使用 PCA 队伍中不同的克隆红茶进行了区分,但研究局限于同一品种的茶叶分类;巴西 CARDOSO 等^[10]使用近红外光谱和支持向量机进行数据建模,可靠区分了 4 种商业绿茶混合物,并且对比了两种

不同光谱仪对分类实验的影响,但使用模型比较单一。现有的研究普遍存在对样品和检测环境要求严格、检测方法复杂、检测时间长的不足。而现行的农产品存储运输机制存在漏洞,茶叶的仓储陈化受温度、湿度、氧气、光照等环境因素以及微生物代谢强度的综合影响,很难根据组分或品质成分完成仓储陈化时间的判定^[11-12]。

本研究拟采用 NIRS,以新茶旧茶、不同产地和品种的茶叶为研究对象,采集光谱信息,通过建立机器学习模型的方法进行分类预测,并对比了不同优化算法对模型性能的影响。与之前的研究把预处理和分类模型作为研究重点不同,着重对比了不同优化算法对分类结果的影响,以达到无损、快速、可靠分类鉴别的目的,为机器学习方法用于茶叶分类预测提供了研究思路和技术手段。

1 材料与方法

1.1 材料说明

实验从年份和产地两方面对茶叶进行了品质鉴定。在鉴别茶叶新陈的实验中,采用的 72 个样品分别为不同品牌和批次,新茶样本和老茶样本各 36 个,其中新茶均为当年采摘的茶叶,老茶则至少有 5 年的存放历史。在茶叶产地溯源实验中,采用 82 个茶叶样本为不同批次不同品牌,其中山东茶叶 63 个,福建茶叶 10 个,江浙茶叶 9 个,包括绿茶、红茶、白茶等。

1.2 近红外光谱的采集

实验采用德国布鲁克光学仪器公司 VERTEX70 型傅立叶变换近红外光谱仪,漫反射样品杯附件。将上述茶叶样本逐一放置在旋转样品台的样品杯中,然后进行近红外光谱采集。波数范围 12500~4000 cm^{-1} ,扫描次数 16 次,分辨率 8 cm^{-1} ,波长间隔 8 cm^{-1} ,环境温度 23~25°C。

1.3 方法

1.3.1 支持向量机法

研究采用支持向量机(support vector machine, SVM)进行分类。SVM 是一种机器学习的方法,其基本思想是在样本空间构造一个最优决策面,使该平面与不同类样本集到达最大距离。该方法是一个凸二次优化问题,能够得到全局最优解。此外, SVM 较传统的神经网络不需要预设网络结构,拥有收敛速度快、易于训练等优点。因此在模式识别、数据挖掘、函数逼近、图像处理方面都得到了广泛的应用^[13-15]。

SVM 算法的核心思想是线性化, 把样本空间从低维非线性映射到高维空间, 目标就从分类变成了解决凸规划问题, 最终求解要在特征空间内部结合线性学习机实现。SVM 分类识别预测模型需要划分训练集和预测集, 先划分样本空间训练模型, 完成后预测剩余数据的类别, SVM 决策方程如公式(1):

$$f(x) = \sum_{i=1}^n \alpha_i k(x, x_i) + b \quad (1)$$

其中, 实变量用 α_i 表示, 训练对象特征变量用 x_i 表示, 影响决策的还有模型训练过程中的被优化对象。支持向量机的核函数可以在非线性映射表达未知的情况下进行使用, 这就能有效避免计算复杂度高的缺点, 在该公式中 $k(x, x_i)$ 就是核函数。核函数可分为线性核函数、多项式核函数、径向基核函数以及两层感知机核函数^[16]。为了提升分类性能, 本研究采用径向基核函数如公式(2):

$$K(x, x_i) = \exp(-\gamma \|x, x_i\|^2) \quad (2)$$

其中, 核函数参数 γ 要提前设定, x_i 是空间中的任意一点, x 表示为空间中心点预测模型的性能受核函数参数 γ 和惩罚参数 C 影响较大, 建模时要着重考虑^[17]。核函数参数 γ 作为支持向量样本的影响半径的倒数, 其数值大小与单个样品对模型影响大小成反比。惩罚参数 C 用来根据误分类样本调整决策面, C 值越大, 越有利于模型分类性能, 反正有利于分界面平滑性。

1.3.2 粒子群算法

粒子群优化算法(particle swarm optimization, PSO)属于进化计算技术的重要组成部分, 其基本思想是通过种群中个体之间信息共享寻求最优解。粒子群算法设置粒子模拟随机解, 粒子只有速度、位置, 没有质量和其他任何属性, 位置属性并非粒子当前点位, 而是决定位移方向。粒子会作为单独的个体各自寻找最优解, 其寻找结果会和其他所有粒子实现信息共享, 拥有最优解的个体粒子相当于决定了整个种群的最优解, 其他粒子根据当前种群的最优解与自己找到的最优解调节自己的速度和位置, 也即移动速度和移动方向。开始的粒子群作为随机解, 在找寻最优解的过程中会根据 p_{best} 和 g_{best} 两个参数调整自身属性, 具体表达式如公式(3)~(4)^[18-19]:

$$v_i = v_i + c_1 \times \text{rand}() \times (p_{best_i} - x_i) + c_2 \times \text{rand}() \times (g_{best_i} - x_i) \quad (3)$$

$$x_i = x_i + v_i \quad (4)$$

其中 i 的取值从 1 到 N 代表粒子数, v_i 表示该粒子速度, x_i 表示粒子所在位置, $\text{rand}()$ 取 0 到 1 之间的随机数。公式(3)中 v_i 为记忆项, 反应调整前上一次粒子的属性; $c_1 \times \text{rand}() \times (p_{best_i} - x_i)$ 为自身认知项, 表示当前位置指向该粒子已知最优解的矢量, 反应自身寻优; $c_2 \times \text{rand}() \times (g_{best_i} - x_i)$ 为群体认知项, 表示当前位置指向整个种群已知最优解的矢量, 反应群体中个体间的信息共享。粒子就是通过后两项, 即自身经验和群体信息共享决定下一步粒子移动方向和速度,

其具体实现包括首先初始化 N 个粒子的属性, 评价每个粒子的初始适应值作为初始最优解并记录其位置, 所有粒子中的最优适应值作为种群初始最优值并记录其位置, 就可以依据公式(3)~(4)进行迭代计算, 期间如果有更优解就会实时进行更新, 每次迭代结束都会更新当前粒子群的种群最优解^[20-21]。

1.3.3 遗传算法

遗传算法(genetic algorithm, GA)通过模拟生物进化产生体现解空间中最优解的优良新个体, 一方面提高了算法效率, 另一方面避免了早熟收敛现象。GA 是一种自适应概率优化算法, 可以通过随机选择、交叉和变异在局部区域繁衍、进化, 目的是进化出最能适应环境的个体, 常用于复杂系统优化。本研究选择用 GA 算法对 SVM 参数进行寻优调整, 关键在于设定 GA 编码规则、建立初始种群、设计自适应函数^[22-23]。支持向量机核函数参数 γ 和惩罚参数 C 代表染色体, 且染色体都被编码为特定序列, 数学表达式如公式(5):

$$f(x) = (\sum_{i=1}^k \alpha_i 2^{i-1})(l_1 - l_2) / (2^k - 1) + l_1 \quad (5)$$

其中, α_i 表示染色体优化参数的二进制基因位数, $[l_1, l_2]$ 限定了优化参数区间, k 表示单基因二进制编码的长度。为保证 GA 对支持向量机算法的参数进行快速寻优, 寻求最优解, 将 GA 种群的操作方式设为随机方式, 既可以提高种群多样性, 也可以将最优个体直接带入下一个种群, 继而通过单点交叉法执行交叉操作和多点变异法执行离散、变异操作, 最终实现全种群遗传进化。GA 算法通过判断是否满足遗传运算的终止进化代数来确定 SVM 分类预测模型的最优参数。GA 中的种群数量一定程度上决定着算法运行速度, 不能随意制定, 而是要靠样本集决定初始种群^[24-25]。

1.3.4 卷积平滑算法

经近红外光谱仪器扫描后, 得到的样品的光谱数据包括体现样品数据的光谱信息和噪声。卷积(Savitzky-Golay, S-G)平滑算法的本质是对光谱信息进行低通滤波, 去掉高频成分, 有效的保留低频信息。因此, 平滑算法可以有效的降低噪声。其基本思路是在平滑点附近随机取点模拟该点, 以求得平滑点的最佳估计值, 这一方法的基本前提是随机噪声在处理“窗口”内的均值为 0^[26]。常用的移动平均法, 其中最为重要的参数是平滑窗口的宽度。若窗口宽度太大, 在做求均值计算时, 会忽略掉一些有用信息, 使得光谱信号产生一定的失真。若窗口宽度太小, 噪声不能被有效的去除掉。本研究采用 S-G 法, 不是计算数学上的简单平均, 而是利用加权平均通过多项式对移动窗口内的数据进行多项式最小二乘拟合, 更加突出了中心点的中心作用^[27]。

1.3.5 数据标准化

数据标准化(Normalization)的作用是消除或减少量纲的干扰, 使影响建模的各要素间具有可比性。在茶叶的新

旧与产地预测实验中,影响分类精度的茶叶理化性质有很多种,但是不同性质对最终模型预测的影响大小是不同的,如果不进行标准化处理就很有可能出现一些有价值的信息没有办法被利用的情况发生^[28]。因此,对数据进行标准化处理,可以增强数据可比性,提高模型性能。Normalization 的本质是数据特征的缩放,实验中采取的支持向量机算法属于基于数据距离的算法,标准化处理有利于模型性能的提升^[29]。

2 结果与分析

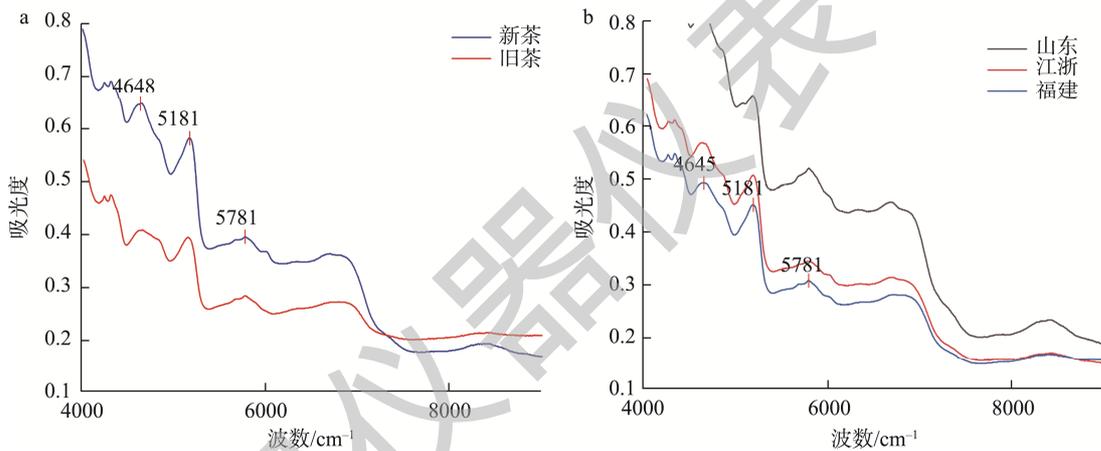
2.1 光谱数据采集

本研究采集茶叶的近红外光谱数据,如图 1 所示。近红外光谱主要是由于分子振动的非谐振性使分子从基态向激发态跃迁产生的^[30]。从图 1 中可知茶叶光谱图吸收峰主要集中在 4645、5181、5781 cm^{-1} 附近,根据之前取得的研究成果,4000~5000 cm^{-1} 波数之间的峰可归因于 N-H 组合、C-H+C-H 组合、C-H=C-C 组合以及 N-H 和 O-H 组合的拉

伸吸收;5181 cm^{-1} 处的峰可归因于 O-H 组合和 C-O 拉伸;5781 cm^{-1} 处的吸收峰可归因于亚甲基官能团($-\text{CH}_2$)^[31]。

2.2 光谱预处理

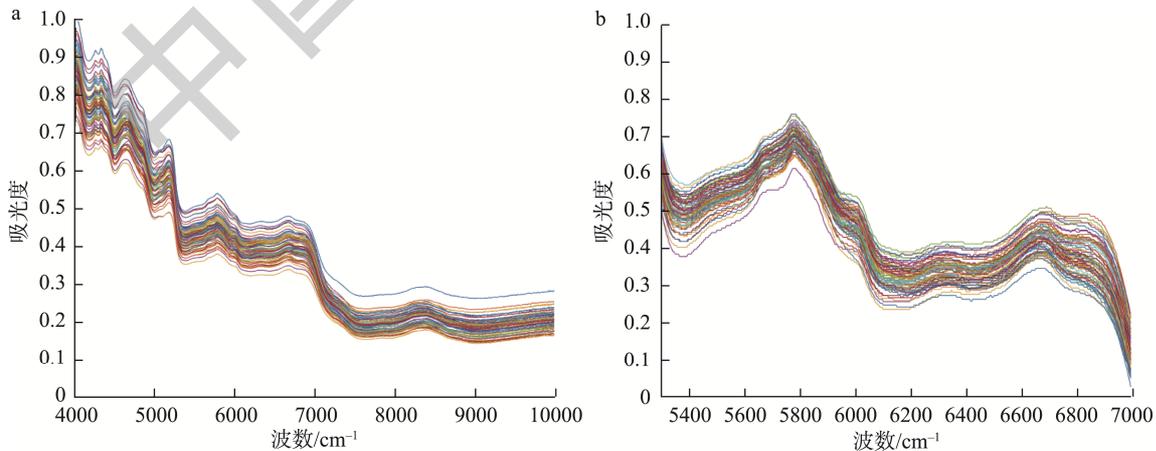
由于茶叶原始近红外光谱吸收峰区分度不高,样本之间的差异性不强,不适合用来训练分类模型,因此使用 S-G 平滑+Normalization 的方法对光谱数据进行预处理。经过预处理后,光谱峰值更加明显,样本之间的差异性也得到增强,可以有效提升分类精度。调用 $y=\text{sgolayfilt}(x,\text{order},\text{framelen})$ 函数进行 S-G 平滑,其中 x 代表矩阵,sgolayfilt 对每一列数据进行处理,其中多项式阶次 order 必须小于框长度 framelen,framelen 必须是奇数。在反复对比不同参数组合下的平滑模式后,最终设置 order=7、framelen=25,即多项式阶次设为 7,平滑窗口中平滑点数为 25,以达到最好的平滑效果。图 2 以山东茶叶为例,对比 63 个山东茶叶样本的原始光谱与预处理后的光谱,预处理后去噪效果明显,光谱峰值更加集中,有利于研究进一步建模分析。



注: a 为新旧茶叶近红外平均光谱图; b 不同产区茶叶近红外平均光谱图。

图 1 茶叶近红外光谱图

Fig.1 Near-infrared spectrograms of tea



注: a 为原始茶叶近红外光谱图; b 为预处理后的茶叶近红外光谱图。

图 2 预处理前后的茶叶近红外光谱图

Fig.2 Near infrared spectrograms of tea before and after pretreatment

2.3 新旧茶叶分类

在新旧茶叶分类实验中, 预处理采用了 S-G 平滑和 Normalization 的方法, 在 GA-SVM 分类算法中还对数据进行了 MSC 即多元散射校正处理, 主要是要消除茶叶内部散射水平不同引起的光谱差异, 使光谱与数据间相关性更加紧密^[32-33]。采用不同算法获取径向基函数参数的数值会有所不同, 直接影响分类精度和效率^[34]。采用 GA-SVM 方法设置参数: 最大进化代数 maxgen 设为 100, sizepop 种群最大数量设为 20, 交叉概率 pCrossover 交叉概率设为 0.4, pMutation 变异概率设为 0.1, c 参数和 g 参数取值范围分别设为 0~100 和 0.01~100.00。

给新旧茶叶分别设置标签, 新茶为 1, 旧茶为 2。通过人工划分样本的方式, 使训练集和预测集的比例达到 3:1, 红色圆圈表示实际测试集分类, 蓝色米字表示经 GA-SVM 模型预测测试集分类。结果显示 18 个预测集样本中有 6 个被错分, 准确率为 66.6667%, 分类可靠性一般。分类结果如图 3 所示。

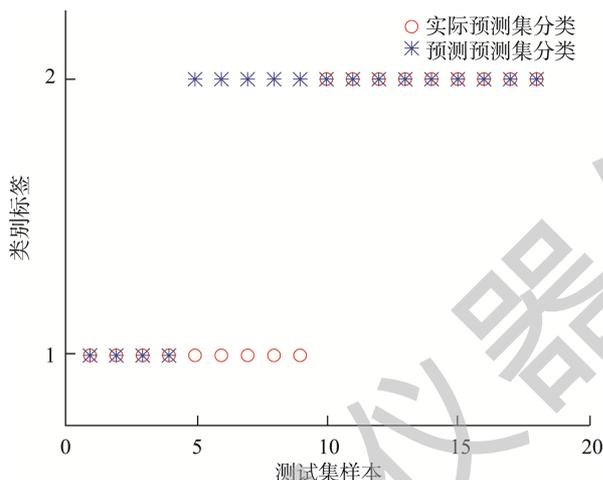


图 3 GA-SVM 判别茶叶新旧结果

Fig.3 GA-SVM results of distinguishing new and old tea

采用 PSO-SVM 方法设置参数如下: c_1 初值设为 1.5, c_2 初值设为 1.7, 最大进化代数 maxgen 设为 200, 种群最大数量 sizepop 设为 20, 速度系数 k 初值设为 0.6, 弹性系数 w_v 和 w_p 初值均设为 1, c 参数和 g 参数取值范围分别设为 0~100 和 0.01~100.00。

样本划分上选取和 GA-SVM 相同的训练集和测试集训练模型。红色圆圈表示实际测试集分类, 蓝色米字表示经 PSO-SVM 模型预测测试集分类。结果显示 18 个预测集样本全部被划分到了正确的类别, 准确率 100.0000%, 分类可靠性很高。分类结果如图 4 所示。

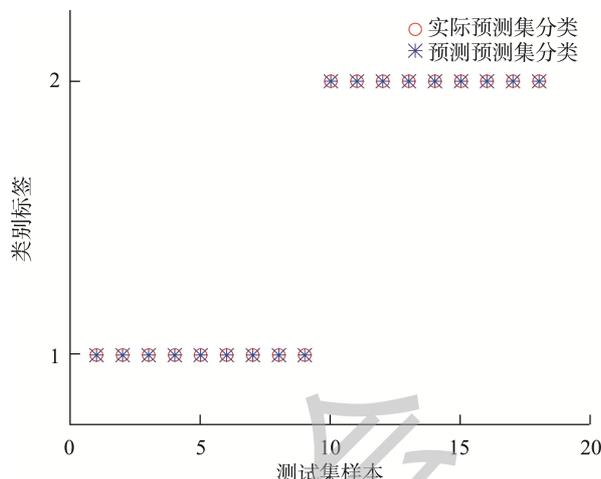


图 4 PSO-SVM 判别茶叶新旧结果

Fig.4 Results of PSO-SVM for distinguishing new and old tea

由图 3、图 4 可知 PSO-SVM 模型预测新旧茶叶分类精度非常高, 在对实验样本的分类中没有出现错分。相比之下 GA-SVM 模型对预测集 18 个样本进行分类测试, 有 5 个样本被分到了错误的类别。为了提高 GA-SVM 模型分准确性, 在预处理中增加了多元散射校正, 增强光谱与数据之间的紧密型。为验证两种模型分类速度的差异, 单独进行多元散射校正处理数据, 记录预处理时间为 3.18 s, 与整体分类用时 388.72 s 的差异相比可以忽略不计。由于 GA 模型的中止条件是动态的, 当解趋近收敛时循环会自动终止, 在一定程度上加快了模型的训练效率。但是 PSO 模型解的收敛速度高于 GA 模型, 因此在实际预测过程中 PSO 模型的运算速度远高于 GA 模型。寻找最优解过程中 c 参数和 g 参数及分类用时如表 1 所示。

2.4 茶叶溯源分类

在茶叶溯源分类实验中, 预处理采用了 S-G 平滑和 Normalization 的方法。采用 GA-SVM 方法设置参数与茶叶新旧分类实验中设置相同。给来自不同产地的 3 种茶叶分别设置标签并划分训练集和预测集, 山东茶叶共 63 个样本, 其中 53 个做训练集, 10 个做预测集, 标签为 1; 福建茶叶共 10 个样本, 其中 7 个做训练集, 3 个做预测集, 标签为 2; 江浙茶叶共 9 个样本, 其中 5 个做预测集, 4 个做训练集, 标签为 3。红色圆圈表示实际测试集分类, 蓝色米字表示经 GA-SVM 模型预测测试集分类。结果显示 19 个预测集样本中 5 个被错误分类, 模型分类正确率为 82.3529%, 结果较为可靠。分类结果如图 5 所示。

表 1 新旧茶叶分类结果

Table 1 Classification results of old and new tea

建模算法	预处理	分类用时/s	最佳参数 c	最佳参数 g	分类预测精度/%
GA-SVM	S-G-MSC-Normalization	520.19	0.4589	150.8962	66.6667
PSO-SVM	S-G-Normalization	131.72	0.1000	0.0100	100.0000

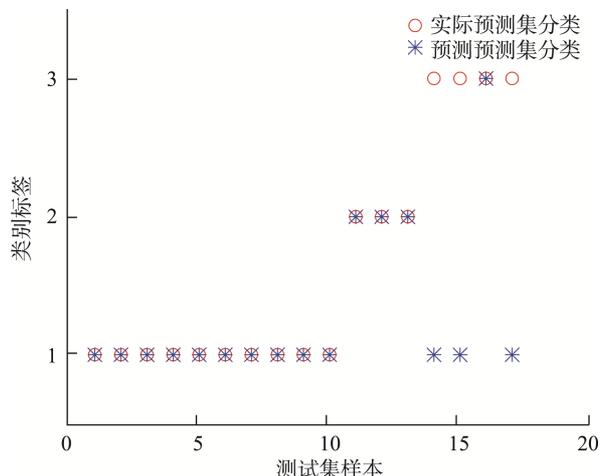


图 5 GA-SVM 判别茶叶产地结果

Fig.5 GA-SVM results for identifying tea origin

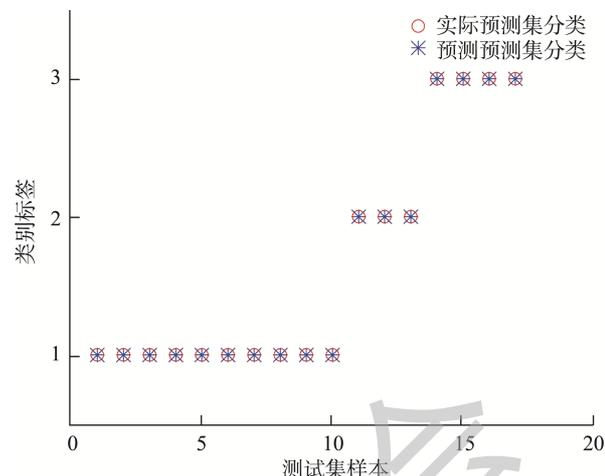


图 6 PSO-SVM 判别茶叶产地结果

Fig.6 PSO-SVM results for identifying tea origin

采用 PSO-SVM 方法设置参数与茶叶新旧分类实验中设置相同, 样本划分上设置训练测试比 trainingRate 为 0.8, 自动划分训练集和测试集训练模型, 分类结果如图 6 所示。红色圆圈表示实际测试集分类, 蓝色米字表示经 PSO-SVM 模型预测测试集分类。结果显示 19 个预测集样本全部被正确分类, 模型分类正确率为 100.0000, 分类精度很高。

由图 5、图 6 可知 GA-SVM 模型进行溯源三分类预测精度较高, 在实验所用的 19 个样本中, 有 3 个来自江浙地区的茶叶被错分上了山东标签, 而 PSO-SVM 模型预测下 3 种产地的所有样本都被分到了正确的类别。与新旧茶叶分类实验类似, 得益于更快的求解收敛速度, PSO-SVM 模型的耗时要少于 GA-SVM 模型, 这种差异随着样本和分类类别的增多更为明显。具体参数及分类用时如表 2 所示。

表 2 茶叶产地分类结果

Table 2 Tea origin classification results

建模算法	预处理	分类用时/s	最佳参数 c	最佳参数 g	分类预测精度/%
GA-SVM	S-G-Normalization	1034.39	92.8412	87.6715	82.3529
PSO-SVM	S-G-Normalization	520.19	2.2372	0.0100	100.0000

3 结论

本研究基于茶叶的近红外光谱对茶叶新旧判别及产地溯源进行了探究。采用 NIRS 获取茶叶样本的光谱数据, 利用 Normalization 方法及 S-G 平滑方法对光谱数据进行预处理, 并根据遗传优化算法和粒子群优化算法建立模型。其中, GA-SVM 模型的分时用时为 1034.39 s, 预测精度为 82.3529%, 而 PSO-SVM 模型的分时用时为 520.19 s, 预测精度为 100.0000%。与 GA-SVM 模型相比, PSO-SVM 模型的分时用时更短, 预测精度更高。结果表明, 利用 NIRS 可以实现茶叶新旧的判别及产地溯源。

参考文献

- [1] 张钦部. 茶叶品质危害因素及预防机制分析[J]. 福建茶叶, 2022, 44(9): 36-38.
ZHANG QB. Analysis of tea quality hazards and prevention mechanisms [J]. Tea Fujian, 2022, 44(9): 36-38.
- [2] 刘奇, 欧阳建, 刘昌伟, 等. 茶叶品质评价技术研究进展[J]. 茶叶科学, 2022, 42(3): 316-330.

- LIU Q, OUYANG J, LIU CW, *et al.* Research progress of tea quality evaluation technology [J]. J Tea Sci, 2022, 42(3): 316-330.
- [3] 肖明霖, 庄映菁, 肖梦暄, 等. 中国茶叶感官审评的发展与溯源[J]. 茶业通报, 2021, 43(2): 62-66.
- XIAO MQ, ZHUANG YJ, XIAO MX, *et al.* Development and traceability of sensory evaluation of tea in China [J]. J Tea Bus, 2021, 43(2): 62-66.
- [4] LIU JX, XIN JY, GAO TT, *et al.* Effect of variable selection and rapid determination of total tea polyphenols contents in Fuzhuan tea by near-infrared spectroscopy [J]. Cyta-J Food, 2022, 20(1): 236-243.
- [5] 朱盼, 万欢, 姚敬, 等. 指纹图谱技术在茶叶产地溯源中的应用[J]. 食品安全质量检测学报, 2021, 12(7): 2520-2525.
ZHU P, WAN H, YAO J, *et al.* Application of fingerprint technology in tracing tea origin [J]. J Food Saf Qual, 2021, 12(7): 2520-2525.
- [6] 孟维君. 核磁共振波谱和近红外光谱应用于茶叶产地溯源及年份分析[D]. 厦门: 厦门大学, 2017.
MENG WJ. Nuclear magnetic resonance spectroscopy and near-infrared spectroscopy are used in tea origin traceability and year analysis [D]. Xiamen: Xiamen University, 2017.
- [7] 周健, 成浩, 贺巍, 等. 基于近红外的 PLS 量化模型鉴定西湖龙井真伪的研究[J]. 光谱学与光谱分析, 2009, 29(5): 1251-1254.
ZHOU J, CHENG H, HE W, *et al.* Research on identification of the authenticity of West Lake Longjing by PLS quantitative model based on near-infrared measurement [J]. Spectrosc Spect Anal, 2009, 29(5):

- 1251–1254.
- [8] 周昌海, 吴畅, 林其斌, 等. 近红外光谱分析(NIR)无损快速检测茶叶中咖啡碱含量模型研究[J]. 安徽科技学院学报, 2018, 32(3): 54–58.
ZHOU CH, WU C, LIN QB, *et al.* Nondestructive and rapid determination of caffeine in tea by near infrared spectroscopy (NIR) [J]. J Anhui Sci Technol Univ, 2018, 32(3): 54–58.
- [9] DAS D, NAG S, NASKAR H, *et al.* Discrimination of various clones of black tea using NIR spectroscopy [C]. 2020 International Conference on Emerging Frontiers in Electrical and Electronic Technologies (Icefeet 2020), 2020.
- [10] CARDOSO VGK, POPPI RJ. Non-invasive identification of commercial green tea blends using NIR spectroscopy and support vector machine [Z]. 2021.
- [11] 黄俐舒. 鲜活农产品物流仓储问题研究[J]. 物流工程与管理, 2021, 43(11): 99–101, 88.
HUANG LS. Research on logistics and warehousing of fresh agricultural products [J]. Logist Eng Manage, 2021, 43(11): 99–101, 88.
- [12] 马冰滢, 徐成成, 任小盈, 等. 普洱茶(生茶)0至10年仓储陈化过程中的化学成分变化[J]. 食品研究与开发, 2022, 43(5): 156–162.
MA BS, XU CC, REN XY, *et al.* Changes in chemical composition of Pu'er tea (raw tea) during storage and aging for 0 to 10 years [J]. Food Res Dev, 2022, 43(5): 156–162.
- [13] JEONG ET, LE CH. A study on disturbance classification of unmanned vehicle data using svm [J]. J Inst Control, 2022, 28(4): 304–312.
- [14] AKBARI MG, KHORASHADIZADEH S, MAJIDI MH, *et al.* Support vector machine classification using semi-parametric model [J]. Soft Comput, 2022, 26(19): 10049–10062.
- [15] CHEN GY, KRZYŻAK A, QIAN SE. Hyperspectral imagery classification with minimum noise fraction, 2D spatial filtering and SVM [J]. Int J Wavelets Multi, 2022, 20(6): 2250025.
- [16] 尹嘉鹏. 支持向量机核函数及关键参数选择研究[D]. 哈尔滨: 哈尔滨工业大学, 2016.
YIN JP. Research on support vector machine kernel function and key parameter selection [D]. Harbin: Harbin Institute of Technology, 2016.
- [17] 刘方园, 王水花, 张煜东, 等. 支持向量机模型与应用综述[J]. 计算机系统应用, 2018, 27(4): 1–9.
LIU FY, WANG SH, ZHANG YD, *et al.* A review of support vector machine models and applications [J]. Computer Syst Appl, 2018, 27(4): 1–9.
- [18] VERMA P, AWASTHI VK, SAHU SK, *et al.* Coronary artery disease classification using deep neural network and ensemble models optimized by particle swarm optimization [J]. Int J Appl Metaheur, 2022, 13(1): 1–25.
- [19] BALASUBRAMANIAN K, ANANTHAMOORTHY NP, RAMYA K. An approach to classify white blood cells using convolutional neural network optimized by particle swarm optimization algorithm [J]. Neural Comput Appl, 2022, 38(18): 16089–16101.
- [20] SUN SL. Shield tunneling parameters matching based on support vector machine and improved particle swarm optimization [J]. Sci Programm-Neth, 2022, (2022): 6782947.
- [21] ZHANG BX. Pork quality identification based on principal component analysis and improved support vector machine [J]. Food Mach, 2022, 1: 146–151.
- [22] POONGODI K, SABARI A. Identification of bio-markers for cancer classification using ensemble approach and genetic algorithm [J]. Intell Autom Soft Comput, 2022, 33(2): 939–953.
- [23] LUND JL, DUBERSTEIN PR, LOH KP, *et al.* Life expectancy in older adults with advanced cancer: Evaluation of a geriatric assessment-based prognostic mode [J]. J Geriatr Oncol, 2022, 13(2): 176–181.
- [24] BHATIA A, CHUG AS, AMIT P, *et al.* A hybrid approach for noise reduction-based optimal classifier using genetic algorithm: A case study in plant disease prediction [J]. Intell Data Anal, 2022, 26(4): 1023–1049.
- [25] SHAFAT A, MARBOUTI F, MAHFOUZ T. Early warning system for highway construction projects using GA-SVM [J]. Int J Constr Manage, 2022, (N):1–10.
- [26] WANG L, PANG L, YAN L. Nondestructive rapid identification of soybean varieties using hyperspectral imaging technology [J]. J Appl Spectrosc, 2022, 89(1): 84–91.
- [27] NING HW, WANG JW, JIANG H, *et al.* Quantitative detection of zearalenone in wheat grains based on near-infrared spectroscopy [J]. Spectrochim Acta A, 2022, 280(2022): 121545.
- [28] CHENG TD, LI P, MA JC, *et al.* Identification of four chicken breeds by hyperspectral imaging combined with chemometrics [J]. Processes, 2022, 10(8): 1484.
- [29] ZHANG XD, DUAN ZH, MAO HP, *et al.* A lettuce moisture detection method based on terahertz time-domain spectroscopy método de detecção de umidade de alface baseado em THz-TDS [J]. Cienc Rural, 2022, 52(6): e20210002.
- [30] 于伶俐, 谭亚军, 赵甲慧, 等. 近红外光谱技术在食品领域的研究进展[J]. 食品安全导刊, 2022, (29): 177–180.
YU LL, TAN YJ, ZHAO JH, *et al.* Research progress of near-infrared spectroscopy in the field of food [J]. China Food Saf Magaz, 2022, (29): 177–180.
- [31] MENG WJ, XU XN, CHENG K, *et al.* Geographical origin discrimination of oolong tea (TieGuanYin, *Camellia sinensis* (L.) O. Kuntze) using proton nuclear magnetic resonance spectroscopy and near-infrared spectroscopy [J]. Food Anal Method, 2017, 10(11): 3508–3522.
- [32] ZHANG F, CUI XH, DING K, *et al.* Study on the influence of different pretreatment methods on gender determination of multiposition [J]. Spectrosc Spect Anal, 2022, 42(2): 434–439.
- [33] ZHANG LX, ZHANG NN, ZHANG X. Discriminant analysis of apple origin based on machine learning algorithm [J]. Laser Optoelectron P, 2022, 59(4): 0430001.
- [34] HUANG WC, LIU HY, ZHANG Y, *et al.* Railway dangerous goods transportation system risk identification: Comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM [J]. Applied Soft Comput, 2021, 109(C): 107541.

(责任编辑: 韩晓红 黄周梅)

作者简介



孙晓荣, 教授, 主要研究方向为智能测量技术与数据处理、系统建模与仿真方法、智能控制方法。

E-mail: srxrchy@sohu.com



胡毅然, 硕士研究生, 主要研究方向为食品安全检测。

E-mail: hyr18801449369@163.com