

DOI: 10.13595/j.cnki.issn1000-0720.2023121205

基于近红外光谱和集成学习的烤烟叶部位快速鉴别

何昕¹, 王鹏², 陈昆燕², 李秋潼², 唐杰², 杨俊¹, 邵利民^{*1}

(1. 中国科学技术大学化学与材料科学学院, 合肥 230026; 2. 重庆中烟工业有限责任公司技术中心, 重庆 400060)

摘要: 本研究基于近红外光谱技术并结合线性判别分析(LDA)和随机子空间方法(RSM), 构建了RSM-LDA集成学习模型, 用于实现烤烟叶部位的高效、准确鉴别。对RSM-LDA模型的子空间维度和分类器数量2个关键参数进行了优化。比较了不同预处理方法、样本集组成和训练集大小对建模的影响。结果表明, 基于原始光谱数据的RSM-LDA模型即可达到很好的鉴别效果, 训练集和测试集准确率分别为98.89%和98.33%。相较于基于单个分类器的鉴别方法, 该方法鉴别准确率高、稳定性好, 可为烤烟叶部位的高效、准确鉴别提供参考。

关键词: 近红外光谱; 烤烟; 鉴别; 集成学习

中图分类号: O657.33 文献标识码: A 文章编号: 1000-0720(2025)03-0396-07

Rapid identification of flue-cured tobacco parts based on near-infrared spectroscopy and ensemble learning

HE Xin¹, WANG Peng², CHEN Kunyan², LI Qiutong², TANG Jie², YANG Jun¹, SHAO Limin^{*1} (1. College of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026, China; 2. China Tobacco Chongqing Industrial Co., Ltd., Chongqing 400060, China)

Abstract: Linear discriminant analysis (LDA) and random subspace method (RSM) were combined to construct an RSM-LDA ensemble learning model for the efficient and accurate identification of flue-cured tobacco leaf parts based on near-infrared spectroscopy technology. Two key parameters of the RSM-LDA model, the dimension of subspace and the number of classifiers, were optimized. Effects of different pretreatment methods, sample set composition, and training set size on modeling were investigated. Results indicated that the RSM-LDA model built on raw spectral data could yield the best identification, with accuracies of 98.89% and 98.33% for the training set and the test set, respectively. In contrast to single-classifier-based identification methods, this approach offered higher accuracy and better stability, and could provide a reference for efficient and accurate identification of flue-cured tobacco leaf parts.

Keywords: near infrared spectroscopy; flue-cured tobacco; identification; ensemble learning

烤烟作为烟草产品的重要原料, 不同叶位的烤烟叶在品质上存在较大差异。由于上部烟叶的价格更高, 市场上常存在以次充好的情况。鉴别烤烟叶部位的传统方法主要有外观评价和化学成分分析等^[1-2]。但是, 这些方法工作量大且耗时, 专家的评判又存在主观差异从而影响鉴别结果。因此, 需要一种更为高效、客观的方法来确保烤烟叶部位的准确鉴别。

近红外光谱技术具有分析速度快、样品前处理简单、无损样品、重复性和再现性好等优点, 在烟草行业发挥着重要作用^[3]。目前, 基于近红外光谱技术对烤烟的研究主要集中在对烤烟香型^[4]、等级^[5]和产区^[6]的鉴别, 对烤烟部位的研究较少。束茹欣等^[7]基于近红外光谱实现了对烤烟叶部位的鉴别, 但模型的训练集和测试集准确率较低。谢娟等^[8]证明近红外光谱结合主成分分析马氏距离判

收稿日期: 2023-12-12; 修回日期: 2024-01-12

基金项目: 烟草近红外光谱在烤烟烟叶分类及果胶含量测定中的应用研究(2020500000300361)基金资助

作者简介: 何昕, 女(1998-), 硕士研究生。研究方向: 近红外光谱与化学计量学, 414402519@qq.com

*通信作者: 邵利民, 男(1972-), 博士, 副教授。研究方向: 化学计量学方法的研究和应用, lshao@ustc.edu.cn

别准则(PCA-MD)可以用于烤烟叶部位鉴别,但模型的稳健性较低,鉴别准确率随样本集改变波动较大。因此,需要探究新的方法以实现烤烟叶部位的准确、可靠鉴别。集成学习在提升模型预测精度、增强模型对噪声和异常值的鲁棒性方面具有显著优势,近年来受到广泛关注^[9-11]。集成学习可以将多个不同烤烟叶部位鉴别模型组合,利用不同模型的差异性来提升模型的泛化能力,提高鉴别的准确性和稳定性。

本研究基于近红外光谱技术,结合线性判别分析(LDA)和随机子空间(RSM)方法构建了RSM-LDA集成学习模型,用于烤烟叶部位鉴别。对RSM-LDA模型的子空间维度和分类器数量2个关键参数进行了优化。比较了不同光谱预处理方法、样本集组成和训练集大小对建模的影响。结果表明,近红外光谱结合RSM-LDA是一种鉴别烤烟叶部位的有效方法,相较于在近红外光谱分析领域取得很好分类效果的LDA^[12]、簇类独立软模式(SIMCA)^[13]、支持向量机(SVM)^[14]、K最近邻(KNN)^[15]、偏最小二乘判别分析(PLS-DA)^[16]方法,该方法鉴别准确率高、稳定性好,可为烤烟叶部位的高效鉴别提供参考。

1 实验部分

1.1 仪器与材料

Antaris II 型傅里叶变换近红外光谱仪(美国 Thermo Scientific 公司);数据分析由 Matlab R2021a 软件(美国 Mathworks 公司)完成。

收集同一产区植株上部烤烟叶 120 份,中部烤烟叶 180 份,在室温 22 °C、湿度 60% 环境下平衡水分 48 h。按 YC/T 31-1996^[17] 烘干烟叶、测定含水量。对样本进行研磨后,通过 450 μm 筛网筛选,取大于 150 g 烟末置于密封袋中避光保存。

1.2 光谱采集

温度 22~26 °C 和湿度 50%~60% 条件下,预热近红外光谱仪 2 h。采集样本光谱前,采集一次背景。设定光谱采集范围为 4000~10000 cm⁻¹,光谱分辨率为 8 cm⁻¹,扫描次数为 72 次。每个样本重复测定 3 次,取平均值作为该样本的原始光谱。

1.3 样本集划分

采用随机划分方法,按照训练集和测试集样本数量比为 3:2 进行样本集划分。划分后训练集样本数量为 180,其中,上部烟叶数量为 72,中部烟叶数量为 108;测试集样本数量为 120,其中,上部烟叶数量为 48,中部烟叶数量为 72。

1.4 光谱预处理

为消除近红外光谱中与鉴别无关的背景干扰和噪声,从基线校正、散射校正、平滑校正、尺度缩放 4 个角度选择了具有代表性的 8 种方法对光谱进行预处理^[18-19]。这些方法是去趋势算法(DT)、多元散射校正(MSC)、标准正态变量变换(SNV)、一阶导数(FD)、二阶导数(SD)、卷积平滑(SG)、移动平均平滑(MA)以及最大最小归一化(MMS)。

1.5 基于 RSM-LDA 的烤烟叶部位鉴别模型

1.5.1 LDA 方法 LDA 通过将数据投影到低维的空间中保证类内差异最小化而类间差异最大化进行分类^[20]。在新的空间中,同一类别的数据点会聚集在一起,而不同类别的数据点则相隔较远,从而达到良好的分类效果^[21]。为了得到新空间的投影向量,定义类间散度矩阵 S_b^{LDA} 为:

$$S_b^{LDA} = \sum_{i=1}^L P_i (m_i - m)(m_i - m)^T \quad (1)$$

式中, L 为类别数, P_i 为第 i 类别的先验概率, m_i 为第 i 类别均值, m 为样本集均值。

类内散度矩阵 S_w^{LDA} 定义为:

$$S_w^{LDA} = \sum_{i=1}^L P_i \sum_{l=1}^{n_i} (x_l^{(i)} - m_i)(x_l^{(i)} - m_i)^T \quad (2)$$

式中, n_i 为第 i 类别样本数, $x_l^{(i)}$ 为第 i 类别的第 l 样本。

LDA 方法根据类间散度矩阵 S_b^{LDA} 和类内散度矩阵 S_w^{LDA} 求得最佳投影向量,从而实现分类。

1.5.2 RSM 方法 RSM 是一种集成学习方法,在原始特征空间中随机选择新的特征子集组合多个分类器对数据集进行分类,综合不同分类器的预测结果从而提高整体模型的预测精度。RSM 保留所有样本,仅对特征进行采样,可有效处理高维数据。不同子空间提供了观察数据的不同视角,有助于提高模型的稳定性^[22]。

1.5.3 RSM-LDA 模型 将 RSM 和 LDA 相结合构建 RSM-LDA 模型,以 LDA 作为 RSM 集成中的分类器,模型训练过程如图 1 所示。RSM-LDA 结合了 LDA 和 RSM 的优势,在建立模型时,关键在于优化分类器的数量以及新特征的数量,也就是子空间维度。

2 结果与讨论

2.1 原始及预处理后的光谱图

图 2a 为不同部位烤烟叶样本原始近红外光谱图,原始光谱中存在明显的基线漂移和背景干扰,为消除背景干扰和基线漂移现象,分别采用 1.4 节中的预处理方法对原始光谱进行预处理。预处理后

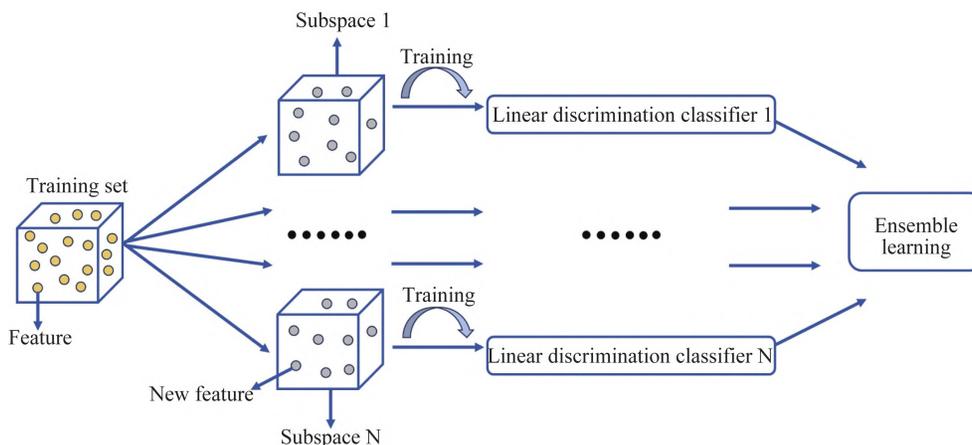


图 1 RSM-LDA 模型训练过程图

Fig.1 Training process diagram of the RSM-LDA model

的光谱图如图 2b~2i 所示。图 2b 和 2c 为 MSC 和 SNV 预处理后的光谱图，预处理后谱图的重合度提高，消除了固体颗粒大小产生的光谱散射影响。图 2d 和 2e 为 FD 和 SD 预处理后的光谱图，可以看出，FD 和 SD 消除了谱峰重叠的影响，突出了特征峰，并且在 5200 波数处均有明显的吸收峰，这可能与样本

中水的吸收有关。图 2f 体现出经过 DT 预处理可以一定程度消除背景干扰但仍然存在基线漂移的情况。MMS 预处理后的光谱吸光度值在 0~1 范围内 (图 2g)。图 2h 和 2i 由 MA 和 SG 预处理得到，MA 和 SG 的使用保留了光谱中的信息并滤除了噪声。从预处理后的光谱图可以看出，预处理方法均在不同

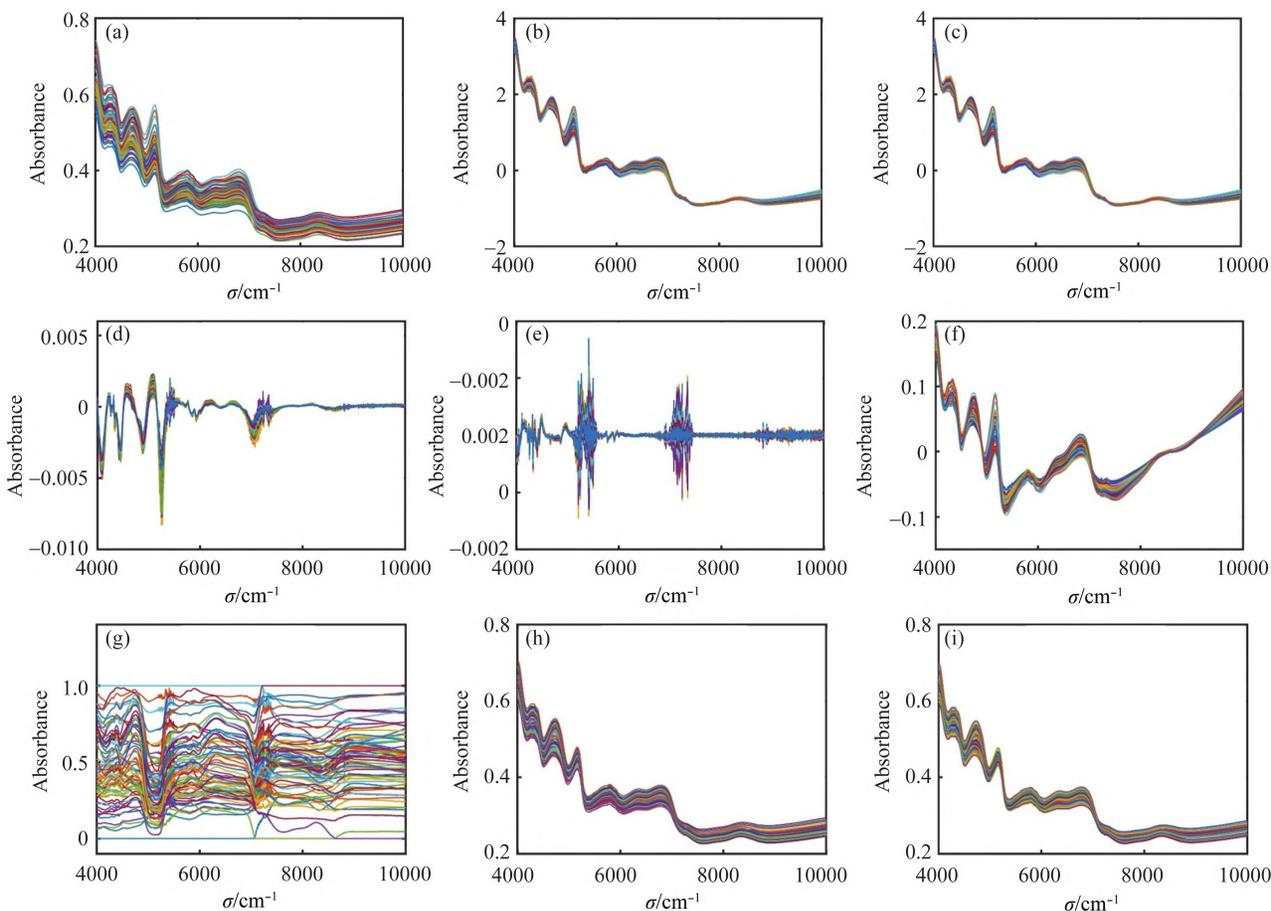


图 2 原始及预处理后的烤烟光谱图

Fig.2 Spectra of untreated and preprocessed flue-cured tobacco

(a) Raw spectra; (b) MSC; (c) SNV; (d) FD; (e) SD; (f) DT; (g) MMS; (h) MA; (i) SG

程度上消除了背景、噪声等因素的影响,但是对于建模的影响还需要进一步探究。

2.2 RSM-LDA 关键参数的选择

子空间维度和分类器数量是构建 RSM-LDA 的关键参数。以原始光谱数据为例,先将子空间维度在 5~100 范围内以 5 为增量进行优化,再将分类器数量在 5~100 范围内以 1 为增量进行优化。图 3 表示十折交叉验证错误率和子空间维度的关系。开始时随着子空间维度的增加,错误率迅速下降,这是因为较低维度的子空间会丢失一些重要的分类信息;随着维度的增加,更多的特征被包含在内,这些特征含有区分不同类别的关键信息。在子空间维度为 85 和 90 时错误率达到最低,维度为 95 和 100 时错误率反而上升,这意味着包含更多的子空间特征并不一定是优势。子空间维度过高可能会导致模型复杂度过高,计算成本增加,甚至引起维数灾难。基于上述分析,将 RSM-LDA 的子空间维

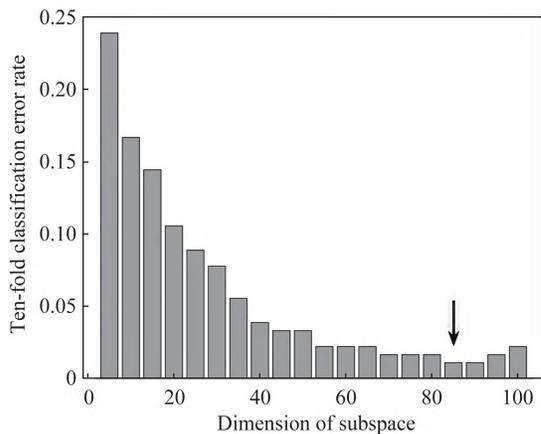


图 3 十折交叉验证错误率和子空间维度的关系

Fig.3 The relationship between ten-fold cross-validation error rate and subspace dimension

度设置为 85。

图 4 表示十折交叉验证错误率和分类器数量的关系。在分类器数量小于 36 时错误率会随分类器数量改变而波动,分类器数量大于 36 时错误率趋于稳定。这是因为增加分类器数量可以增加集成中的多样性,但是当集成达到一定的大小后,即使增加更多的分类器,也无法增强模型对数据内在规律的识别和学习,因此错误率不再降低。过多的分类器数量会导致建模时间增加,因此选择分类器数量为 40 建立 RSM-LDA 模型。

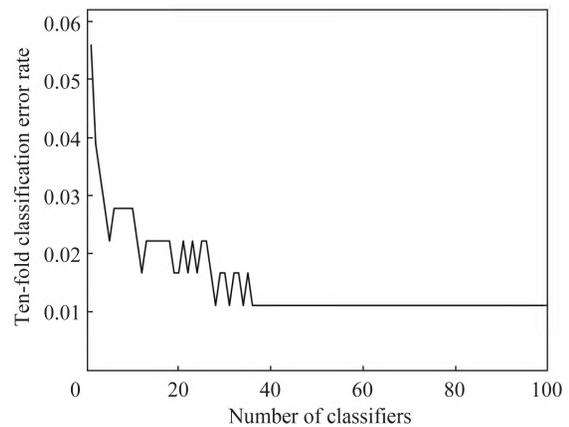


图 4 十折交叉验证错误率和分类器数量的关系

Fig.4 The relationship between ten-fold cross-validation error rate and the number of classifiers

2.3 光谱预处理对建模的影响

分别采用原始光谱数据和预处理后的光谱数据建立 RSM-LDA 模型,以 2.2 节的方法进行参数选择。不同预处理方法下模型对烤烟叶部位的鉴别准确率如表 1 所示。由表 1 可知,基于原始光谱数据的 RSM-LDA 模型就能达到很好的鉴别效果,训练集和测试集准确率分别达到 98.89% 和 98.33%。预处理

表 1 不同预处理方式的 RSM-LDA 模型的鉴别结果

Table 1 Identification results of the RSM-LDA model with different pretreatment methods

Pretreatment method	Training set accuracy/%			Test set accuracy/%		
	Upper	Middle	Total	Upper	Middle	Total
None	100.0	98.15	98.89	97.92	98.61	98.33
FD	95.83	97.22	96.67	93.75	97.22	95.83
SD	91.67	92.59	92.22	89.58	91.67	90.83
SG	98.61	98.15	98.33	97.92	98.61	98.33
MA	98.61	98.15	98.33	95.83	98.61	97.50
DT	95.83	97.22	96.67	93.75	97.22	95.83
MMS	91.67	92.59	92.22	89.58	93.06	91.67
MSC	100.0	98.15	98.89	97.92	98.61	98.33
SNV	95.83	98.15	97.22	95.83	97.22	96.67

虽然具有消除光谱中背景干扰和噪声的作用,但是对于提升RSM-LDA建模准确率并无效果。不恰当的预处理方法,如FD和SD等,还会降低模型的鉴别能力。这可能是因为RSM-LDA本身就具有很好的鉴别能力,它能够处理原始数据中的噪声和变异,有效挖掘不同部位烤烟叶光谱数据中的隐藏差异。

2.4 不同方法建模结果比较

为验证RSM-LDA模型对烤烟叶部位的鉴别能力,比较了在高维近红外光谱分析领域取得很好分类效果的LDA, SIMCA, SVM, KNN, PLS-DA模型。与RSM-LDA不同,这些模型均基于单个分类

器进行分类。采用2.3节方法对数据进行预处理后,选择不同模型的最高鉴别准确率进行比较。不同模型的鉴别效果及对应的最佳预处理方法如表2所示。从表2可以看出,相较于单个分类器模型,RSM-LDA模型鉴别效果最佳。基于单个分类器的模型均需要对光谱进行预处理后才能达到最佳的鉴别效果。这可能是因为单个分类器模型对数据中的噪声和异常值更敏感,而集成模型能够在多个子模型的帮助下平衡这些影响。RSM-LDA模型通过整合多个分类器,能够更好地捕捉数据的复杂结构,进而提高鉴别效果。

表2 不同模型的鉴别效果及对应的最佳预处理方法

Table 2 Identification effects of different models and corresponding optimal pretreatment methods

Modeling method	Optimal pretreatment method	Training set/%			Test set/%		
		Upper	Middle	Total	Upper	Middle	Total
RSM-LDA	None	100.0	98.15	98.89	97.92	98.61	98.33
LDA	SNV	92.59	91.67	92.22	93.75	91.67	92.50
SIMCA	SNV	91.67	92.59	92.22	93.75	90.28	91.67
SVM	MSC	94.44	96.30	95.56	95.83	94.44	95.00
KNN	FD	72.22	77.78	75.56	72.92	79.17	76.67
PLS-DA	MSC	93.05	92.59	92.78	93.75	91.67	92.50

2.5 样本集组成对建模的影响

为了更全面的探究RSM-LDA模型的分性能类和泛化能力,研究了样本集组成对建模的影响。将所有样本随机划分为训练集和测试集200次,以在2.4节中表现较好的SVM模型为参考模型,2种模型的训练集和测试集准确率结果箱型图如图5所示。直观上看,RSM-LDA模型200次训练结果分布更集中,RSM-LDA模型优于SVM模型。计算了RSM-LDA和SVM模型200次训练的测试集准确率

平均值,分别为98.67%和94.66%。以上结果均表明,相较于SVM模型,RSM-LDA模型更稳定,预测准确率更高。

2.6 训练集大小对建模的影响

除了样本集组成,训练集的大小也会对模型产生影响。为探究训练集大小对模型的影响,在4~100范围内,以4为增量增加训练集样本数量,保持测试集样本不变,建立RSM-LDA和SVM模型。训练集大小对不同模型测试集准确率的影响如图6所示。

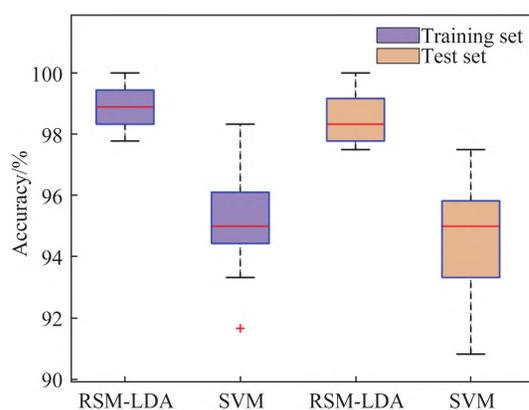


图5 不同模型训练集和测试集结果比较箱型图

Fig.5 Box diagram of comparison of the results of different model training sets and test sets

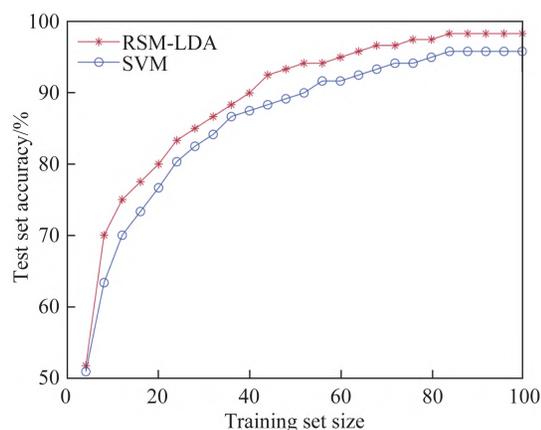


图6 训练集大小对模型准确率的影响

Fig.6 Influence of training set size on the model accuracy

3 结论

建立了RSM-LDA模型并应用于烤烟叶部位鉴别,结果表明,对RSM-LDA模型进行参数选择后,仅需使用原始光谱数据就能实现烤烟叶部位的精准鉴别。在建模过程中,数据预处理是一项非常复杂且庞大的工作,预处理的方法多样,不同数据的最优预处理方法不同,往往需要根据经验选择最优的预处理方法。理想的方法不是对光谱进行预

处理,而是通过巧妙的建模策略达到同样的效果。RSM-LDA模型无需预处理就能达到很好的鉴别效果,这在实际应用中将显著节省建模时间。与基于单个分类器的模型相比,RSM-LDA模型的稳定性好、泛化能力强、鉴别准确率高。因此,基于RSM-LDA模型可以实现烤烟叶部位的准确、可靠鉴别,为近红外光谱技术实现烤烟叶部位高效鉴别提供了参考。

参考文献

- [1] Li Z L, Qin N. Analysis of tobacco leaf quality of different flue-cured tobacco varieties [J]. Journal of Anhui Agricultural Sciences, 2022, 50(11): 32.
李自林, 秦娜. 不同烤烟品种烟叶质量分析 [J]. 安徽农业科学, 2022, 50(11): 32.
- [2] Wang Y J, Zhou J H, Li Q, Zhang Y Y, Zeng M, Xie Y, Mao J S. Chemical component usability of Qujing tobacco leaves and its influences on sensory quality [J]. Tobacco Science & Technology, 2014, 47(11): 67.
王育军, 周冀衡, 李强, 张一扬, 曾敏, 解燕, 毛建书. 曲靖烟叶化学成分可用性及其对感官评吸质量的影响 [J]. 烟草科技, 2014, 47(11): 67.
- [3] Bin J, Zhou J H, Fan W, Li X, Liang Y Z, Xiao Z X, Li C S. Automatic grading of flue-cured tobacco leaves based on NIR technology and extreme learning machine algorithm [J]. Acta Tabacaria Sinica, 2017, 23(2): 60.
宾俊, 周冀衡, 范伟, 李鑫, 梁逸曾, 肖志新, 李春顺. 基于NIR技术和ELM的烤烟烟叶自动分级 [J]. 中国烟草学报, 2017, 23(2): 60.
- [4] Sha Y F, Huang W, Wang L, Liu T A, Yue B H, Li M J, You J L, Ge J, Xie W Y. Merging MIR and NIR spectral data for flavor style determination [J]. Spectroscopy and Spectral Analysis, 2021, 41(2): 473.
沙云菲, 黄雯, 王亮, 刘太昂, 岳宝华, 李敏杰, 尤静林, 葛炯, 谢雯燕. 中红外和近红外数据融合的香型风格判别 [J]. 光谱学与光谱分析, 2021, 41(2): 473.
- [5] Shen H C, Geng Y R, Ni H F, Wang H, Wu J Z, Liao F, Chen Y, Liu X S. Grade determination of flue-cured tobacco by near infrared spectroscopy combined with teaching-learning-based optimization algorithm optimized extreme learning machine [J]. J Instrum Anal, 2022, 41(7): 1052.
沈欢超, 耿莹蕊, 倪鸿飞, 王辉, 吴继忠, 廖付, 陈勇, 刘雪松. 近红外光谱技术结合教与学算法优化极限学习机实现烤烟等级判定 [J]. 分析测试学报, 2022, 41(7): 1052.
- [6] Geng Y R, Shen H C, Ni H F, Chen Y, Liu X S. Support vector machine optimized by near-infrared spectroscopic technique combined with grey wolf optimizer algorithm to realize rapid identification of tobacco origin [J]. Spectroscopy and Spectral Analysis, 2022, 42(9): 2830.
耿莹蕊, 沈欢超, 倪鸿飞, 陈勇, 刘雪松. 近红外光谱结合灰狼算法优化支持向量机实现烟叶产地快速鉴别 [J]. 光谱学与光谱分析, 2022, 42(9): 2830.
- [7] Shu R X, Wang G D, Zhang J P, Ni L J. NIRS pattern recognition of domestic flue-cured tobacco leaves [J]. Tobacco Science & Technology, 2006, 39(8): 12.
束茹欣, 王国东, 张建平, 倪力军. 国产烤烟烟叶的NIRS模式识别 [J]. 烟草科技, 2006, 39(8): 12.
- [8] Xie J, Luo J Q, Yao H M, Ni L J, Zhang L G. Pattern recognition of growing area and stalk position of domestic flue-cured tobacco based on NIR and chemical components [J]. Tobacco Science & Technology, 2008, 41(7): 42.
谢娟, 罗建群, 姚鹤鸣, 倪力军, 张立国. 基于NIR和化学指标的国产烤烟烟叶产地、部位模式识别 [J]. 烟草科技, 2008, 41(7): 42.
- [9] He S H, Li L Q, Liu T, Liu Z B, Yang H H. Near-infrared spectroscopy based ensemble modeling method for Huanglongbing detection [J]. J Anal Sci, 2020, 36(2): 287.
贺胜晖, 李灵巧, 刘彤, 刘振丙, 杨辉华. 柑橘黄龙病检测的近红外光谱集成建模方法 [J]. 分析科学学报, 2020, 36(2):

- 287.
- [10] Ju W, Lu C H, Zhang Y J, Chen X J, Jiang W W. Research on quantitative regression method of IR spectra of organic compounds based on ensemble learning with wavelength selection [J]. *Spectroscopy and Spectral Analysis*, 2023, 43(1): 239.
鞠 薇, 鲁昌华, 张玉钧, 陈晓静, 蒋薇薇. 集成学习结合波长选取的有机物红外光谱定量回归方法研究 [J]. *光谱学与光谱分析*, 2023, 43(1): 239.
- [11] Li Q L, Chen Z G, Jiao F. Prediction of oil content in oil shale by near-infrared spectroscopy based on Stacking ensemble learning [J]. *Spectroscopy and Spectral Analysis*, 2023, 43(4): 1030.
李泉伦, 陈争光, 焦 峰. 基于Stacking集成学习的近红外光谱油页岩含油率预测 [J]. *光谱学与光谱分析*, 2023, 43(4): 1030.
- [12] Li J Y, Yu M, Zheng Y, Li P. Nondestructive identification of *Poria cocos* blocks from different origins based on near infrared spectroscopy [J]. *Chin J Anal Lab*, 2021, 40(12): 1381.
李嘉仪, 余 梅, 郑 郁, 李 跑. 基于近红外光谱技术的不同产地茯苓块无损鉴别 [J]. *分析实验室*, 2021, 40(12): 1381.
- [13] Li C B, Niu C W, Su L, Wu S J, Sun X Q. Identification and variance analysis of Chinese yam from different origins by near infrared spectroscopy [J]. *Food Research and Development*, 2022, 43(15): 175.
李长滨, 牛畅炜, 苏 丽, 吴圣江, 孙宪秋. 不同产地山药的近红外鉴别和差异分析 [J]. *食品研究与开发*, 2022, 43(15): 175.
- [14] Hui Y T, Wang D C, Tang X, Peng Y Q, Wang H D, Zhang H F, You Y. Detection of Sorghum-Sudan grass seed germination rate based on near infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2022, 42(2): 423.
惠云婷, 王德成, 唐 欣, 彭要奇, 王红达, 张海凤, 尤 泳. 基于近红外光谱分析的高丹草种子发芽率检测研究 [J]. *光谱学与光谱分析*, 2022, 42(2): 423.
- [15] Liu T K, Hu Z K, Long W J, Lei G H, Wang X T, He J L, Yang X L, Yang J, Fu H Y. Geographical origin traceability of *Eucommiae cortex* based on near and mid infrared spectroscopy [J]. *Chemical Reagents*, 2022, 44(7): 952.
刘庭恺, 胡子康, 龙婉君, 雷光华, 王芯恬, 何杰灵, 杨小龙, 杨 健, 付海燕. 基于近红外和中红外光谱的杜仲产地溯源 [J]. *化学试剂*, 2022, 44(7): 952.
- [16] Xia Z Z, Zheng D, Xia H, Yao J J, Wang S P, Qiu J F. Rapid identification of shiitake geographical origins based on near infrared spectroscopy [J]. *J Instrum Anal*, 2020, 39(11): 1371.
夏珍珠, 郑 丹, 夏 虹, 姚晶晶, 王胜鹏, 仇建飞. 基于近红外光谱分析技术的栽培香菇产地快速鉴别 [J]. *分析测试学报*, 2020, 39(11): 1371.
- [17] YC/T 31—1996 Tobacco and tobacco products—Preparation of test sample and determination of water content—Oven method [S].
YC/T 31—1996 烟草及烟草制品试样的制备和水分的测定烘箱法 [S].
- [18] Wang Y, Liu Z D, Li W X, Li N N, Wang X C. Influence of online near-infrared spectroscopy preprocessing on qualitative identification of waste textiles [J]. *Chin J Anal Lab*, 2023, 42(11): 1449.
王 悦, 刘正东, 李文霞, 李宁宁, 王笑宸. 在线近红外光谱预处理对废旧纺织品定性识别的影响 [J]. *分析实验室*, 2023, 42(11): 1449.
- [19] Rinnan A, Berg F S, Engelsen S B. Review of the most common pre-processing techniques for near-infrared spectra [J]. *TrAC-Trends Anal Chem*, 2009, 28(10): 1201.
- [20] Yang L, Huang J H, Chen X N, Lyu H S, Wang L Q, Su B M. Nondestructive analysis of cultural relics pigments by chemometrics combined with diffuse reflectance spectroscopy [J]. *J Instrum Anal*, 2020, 39(7): 844.
杨 璐, 黄建华, 陈欣楠, 吕兹树, 王丽琴, 苏伯民. 漫反射光谱结合化学计量学方法无损分析彩绘文物颜料 [J]. *分析测试学报*, 2020, 39(7): 844.
- [21] Wang S Y, Yang Y Z, He W W, Li R K. Identification of black signing-pen ink based on hyperspectral imaging technique [J]. *J Instrum Anal*, 2021, 40(10): 1489.
王书越, 杨玉柱, 何伟文, 李润康. 基于高光谱的黑色签字笔墨种类鉴别方法研究 [J]. *分析测试学报*, 2021, 40(10): 1489.
- [22] Zha W, Zhao X N, Luo B, Bai W W, Kang K, Hou P C, Zhang H. Identification of wheat seed endosperm texture using hyperspectral imaging combined with an ensemble learning model [J]. *J Food Compost Anal*, 2023, 121(4): 105398.