

激光与光电子学进展

基于中红外和远红外数据融合的油茶籽成熟度鉴别

马鑫, 王标, 李春, 马卿效, 滕燕, 蒋玲*

南京林业大学信息科学技术学院, 江苏 南京 210037

摘要 针对油茶籽采收过程中缺乏判断其成熟度的依据, 导致茶油的质量和产量不佳等问题, 提出一种基于中红外和远红外光谱数据融合检测油茶籽成熟度的方法。采用傅里叶变换红外光谱仪测试了在不同成熟阶段, 不同含油率油茶籽的中红外和远红外光谱数据, 利用不同特征提取方法(主成分分析法、连续投影算法、无信息变量消除法)对原始光谱数据进行特征提取, 再结合支持向量机算法(SVM)建立了油茶籽成熟度的鉴别模型。结果表明: 在中红外波段范围内, 采用连续投影算法结合遗传算法优化的 SVM 模型, 获得最优的鉴别精度为 93.33%; 在远红外波段范围内, 利用主成分分析法结合遗传算法优化的 SVM 模型, 实现了 96.67% 的鉴别精度。将中红外波段数据与远红外波段进行数据融合, 结合优化后的 SVM 算法能将鉴别精度提高到 100%。该研究表明, 红外光谱技术结合优化后的 SVM 模型可以实现对油茶籽含油率的精确鉴别, 数据融合技术能够有效地增加光谱信息并且去除单一光谱的冗余信息。该结果可为油茶的最佳采摘时间提供参考, 并可拓展应用到其他农林产品成熟度的鉴别中。

关键词 光谱学; 中红外光谱; 远红外光谱; 油茶籽成熟度检测; 数据融合; 支持向量机

中图分类号 O433.4

文献标志码 A

DOI: 10.3788/LOP202259.1930001

Maturity Identification of Camellia Seeds Based on Mid- and Far-Infrared Data Fusion

Ma Xin, Wang Biao, Li Chun, Ma Qingxiao, Teng Yan, Jiang Ling*

College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, Jiangsu, China

Abstract To solve the problems of poor quality and yield of camellia oil owing to the lack of basis for determining the maturity of camellia seeds during the process of harvesting, a method for detecting the maturity of camellia seeds based on mid- and far-infrared spectral data fusion was proposed herein. A Fourier transform infrared spectrometer was used to test the mid- and far-infrared spectroscopy data of camellia seeds with different oil contents at different maturity stages. Various feature extraction methods (principal component analysis, successive projection algorithm, and noninformation variable elimination method) were used to extract the original spectral data, and the methods were combined with a support vector machine algorithm (SVM) to develop a model for identifying the maturity of camellia seeds. The results show that the best discrimination accuracy in the mid-infrared band is 93.33% when using the successive projection algorithm combined with the genetic algorithm to optimize the SVM model. In the far-infrared band, nine variables extracted using the principal component analysis are used as input variables, and when combined with the SVM model optimized by the genetic algorithm, the identification accuracy of 96.67% is attained. The identification model of camellia seed maturity is established using the SVM algorithm after parameter optimization. The experimental results show that the accuracy of intermediate data fusion combined with the optimized SVM algorithm can reach 100%. The results of this study show that when combined with an improved SVM model, the infrared spectroscopy can accurately determine the oil content of camellia seeds. Data fusion can effectively increase the spectral information and remove redundant information from a single spectrum. The results can provide a reference to determine the best picking time of camellia and can be extended to determine the maturity of other agricultural and forestry products.

Key words spectroscopy; mid-infrared spectra; far-infrared spectra; maturity identification of camellia seeds; data fusion; support vector machine

收稿日期: 2021-08-05; 修回日期: 2021-08-23; 录用日期: 2021-09-24

基金项目: 国家自然科学基金(62001235)、江苏省自然科学基金(BK20161526)

通信作者: *jiangling@njfu.edu.cn

1 引言

油茶是我国特有的木本油料树种,与橄榄、棕榈和椰子并称为世界四大木本食用油料^[1]。油茶籽中含油量丰富,含不饱和脂肪酸高达 90%,且含有多种活性成分,是一种营养丰富的食用油,被誉为“东方橄榄油”。油茶籽含油率是茶油产量的重要因素,油茶的采摘时间直接影响着茶油的产量和品质,适时的采收能够保证茶油的质量和品质,从而提高出油率^[2-3]。目前茶油的提取方法有压榨法、浸出法、超声波提取法、水酶法等化学方法,但这些方法存在费时费力、出油率低、有溶剂残留等缺点^[4],不利于简便快速地检测油茶籽的含油率。因此需要一种能够快速准确地检测油茶籽含油率,从而鉴别油茶籽成熟度的方法,为油茶的采摘时间以及提高茶油产量提供帮助和参考。

红外光谱技术可以对样品进行定性或定量分析,具有分析速度快、操作方便、绿色环保等特点,适合于农林产品品质检测、分级鉴别等方面。目前大部分针对油茶籽的红外光谱研究主要侧重在油茶籽内部营养成分的检测^[5-7],针对成熟度的研究较少,而红外光谱技术已经在水果、谷物等农产品的成熟度检测中得到了广泛应用^[8]。2018年,赵娟等^[9]利用近红外光谱结合成熟度评价指数对三种成熟阶段的苹果进行分类鉴别,分类准确率可达到 85.17%。2020年,邵圆圆等^[10]采用高光谱技术对不同成熟度的肥桃进行检测,准确率可达到 98.3%。2021年,Minas 等^[11]采用近红外光谱技术,结合非线性迭代偏最小二乘回归模型预测了桃子的内部质量及成熟度,桃子在成熟过程中其干物质容量和可溶固形物浓度的预测误差在 0.58% 以内,识别准确率达到 96% 以上。以上研究表明,利用某一波段的红外光谱技术分析鉴别农林产品的成熟度具有较高的准确率。但是利用单一来源的光谱信息无法完全表征样品中的化学信息,而数据融合可以综合利用不同来源光谱的有效信息,从而增强模型的稳定性,提高识别准确率。沙云菲等^[12]融合了近红外与

中红外光谱数据,使用支持向量机(SVM)成功鉴别清香型、中间香型和浓香型三种香型风格的烟叶。李亚惠等^[13]融合了近红外和中红外光谱数据,使用偏最小二乘-支持向量机(LS-SVM)成功鉴别 5 种不同产地的枸杞,识别准确率能达到 99.17%。以上研究表明,红外光谱数据融合技术能更精确地进行样品品质分类,油茶籽的红外吸收光谱包含了大量其内部有机物官能团信息,不同成熟时期的油茶籽所携带光谱信息不同,通过合适的数据处理可以有效地应用于油茶籽的成熟度鉴别。

到目前为止,国内外对于油茶的检测研究主要以近红外光谱为主,大多数应用于油茶籽内部成分和品质的检测,基于中红外和远红外光谱在油茶籽成熟过程中含油率的检测研究鲜有报道。因此,本文采用中红外(MIR)和远红外(FIR)光谱技术结合化学计量学方法对油茶籽的含油率进行检测,从而建立油茶籽成熟度的鉴别模型,以期开发一种快速鉴别油茶籽成熟度的方法,为油茶的采摘时间以及提高茶油产量提供帮助和参考。

2 样品制备和实验方法

2.1 样品制备和实验方法

本次实验的油茶籽样品来自江苏省南京市江宁区禄口街道南京金航油茶专业合作社的油茶林。图 1 为油茶果和油茶籽样品图,分别于 2020 年的 9 月 25 日、9 月 30 日、10 月 6 日、10 月 13 日和 10 月 20 日分批采摘油茶鲜果共计 100 个,去壳后置于 60 °C 烘箱烘干至恒重。图 1(a)为生长于树上的成熟油茶果。图 1(b)为去壳后的油茶籽。图 1(c)为中红外光谱测试用的油茶籽切片样本,厚度控制为 1 mm。图 1(d)为远红外光谱测试采用的油茶籽粉末与聚乙烯粉末的混合样品。油茶籽利用高速粉碎机粉碎,过 200 目标标准筛盘,将聚乙烯粉末与不同成熟时期的油茶籽粉末按照 7:1 的比例研磨混合均匀,最后压制直径 13 mm,厚度在 1.0~1.2 mm 之间的圆形薄片。



图 1 油茶果和油茶籽样品图。(a)成熟的油茶果;(b)油茶果去除壳后的油茶籽;(c)用于中红外光谱测试的油茶籽切片样品;(d)用于远红外光谱测试的油茶籽与聚乙烯粉末混合研磨后的压片样品

Fig 1 Camellia fruit and camellia seed sample diagrams. (a) Ripe amellia fruit; (b) camellia seeds after removing the shell; (c) camellia seed slice sample for mid-infrared spectroscopy test; (d) compressed samples of camellia seeds mixed with polyethylene powder for far-infrared spectroscopy

仪器使用的是 Bruker 公司的 Vertex80v 傅里叶变换红外光谱仪,远红外吸收光谱的测试采用 25 μm 厚 Mylar 分束器,工作波数范围为 16~120 cm^{-1} ,分辨率为 2 cm^{-1} 。同时使用衰减全反射(ATR)附件测试中红外吸收光谱,采用 KBr 分束器,波数范围为 400~4000 cm^{-1} ,分辨率为 2 cm^{-1} 。样品腔和光谱仪系统均被抽真空,以减少空气中水蒸气对吸收光谱的影响,数据采集过程中保持室内温度、湿度基本不变。

2.2 含油率测定

油茶籽含油率采用 NAI-ZFCDY-6Z 型脂肪测定仪(上海那艾精密仪器有限公司)按照 GB5009.6—2016 规定的方法测定。首先称取充分混匀后的油茶仁 2.000 g 放入滤纸筒内,然后将滤纸筒放入索氏抽提器的抽提筒内,连接到已经充分干燥的接收瓶,从抽提器冷凝管上端加入 60 mL 无水乙醚至瓶内容积的 2/3 处,于水浴上加热,使无水乙醚不断回流抽提 6 h。提取结束时,用磨砂玻璃棒接取一滴提取液,磨砂玻璃棒上无油斑表明提取完毕。取下接收瓶,回收无水乙醚,等到接收瓶内溶剂剩余 1~2 mL 时放在水浴上蒸干,再于 100 $^{\circ}\text{C}$ 干燥箱内干燥 1 h,放干燥器内冷却 0.5 h 后称量。重复以上操作直至质量恒定(直至两次称量差不超过 2 mg)。最后油茶籽含油率可按式(1)进行计算。

$$R = \frac{m_1 - m_0}{m_2} \times 100\%, \quad (1)$$

式中: R 为油茶籽含油率; m_1 为恒重后接收瓶和脂肪的质量(单位为 g); m_0 为接收瓶的质量(单位为 g); m_2 为油茶仁的质量(单位为 g)。

3 化学计量学方法

物质的红外光谱反映了其分子中化学键或官能团的振动吸收,不同的化学键或官能团吸收频率不同,在红外光谱上将处于不同位置,从而可获得分子中含有化学键或官能团的信息^[14-17]。在对物质进行鉴别时,其光谱不仅包含对建立鉴别模型的有效信息也包括一些会对建模引起干扰的无效信息,从而影响所建立模型的预测精度。为了提取原始光谱数据的有效信息,本文首先采用三种不同特征变量筛选方法主成分分析法(PCA)、连续投影算法(SPA)和无信息变量消除(UVE)进行光谱特征数据提取^[18-19]。由于油茶籽内部成分复杂,测得的光谱数据与成熟度的线性关系易受到干扰,在特征选取的基础上,采用非线性的支持向量机模型进行油茶籽成熟度的预测。

支持向量机可用于定性分析和非线性回归^[20-21],其主要思想是建立一个分类超平面使得正例和反例之间的距离最远。在利用支持向量机建立分类模型中,有两个参数(惩罚参数 c 和核函数参数 g)会影响分类的准确精度,通过对参数的寻优可以有效地改善模型的精度。支持向量机也适用于样本较少的分类,其计算的复杂性取决于支持向量的数目,而不是样本空间

的维数^[22],因此适用于本文油茶籽样本量较少的情况。同时,本文利用三种寻优算法网格搜索法^[23](GS)、遗传算法^[24](GA)和粒子群算法^[25](PSO)对惩罚参数 c 和核函数参数 g 进行优化,以进一步提高预测精度。

4 结果与讨论

4.1 油茶籽含油率统计

油茶籽油脂转化累积期一般在 8 月下旬至 10 月果实成熟前,在这个时期,油茶籽的体积和质量增长较少,逐渐趋于稳定,油茶果外观变化不大,而体内的油脂量呈直线增加,含油率也在提升,因此可以通过含油率的变化来区分油茶籽的成熟度。表 1 为不同采摘时期油茶籽的含油率统计。将 5 个时间批次的所有油茶籽样本分开进行粉碎,各等分为 10 份,分别进行含油率的统计。由表 1 可知,不同时期的油茶籽含油率有一定差异,其中采摘间隔时间越长,含油率差异越明显,不同批次油茶籽的含油率还出现交叉的现象,可能是因为地理环境因素影响,即使同一天采摘的油茶籽的含油率也会出现差异。但是,整体来说,不同批次的油茶籽含油率变化范围在 13.12%~24.33% 之间,含油率的差异会带来油茶籽吸收光谱的不同,有助于鉴别模型的建立。

表 1 油茶籽含油率
Table 1 Oil content of camellia seed

Picking time	Minimum / %	Maximum / %	Average value / %
September 25th	13.12	18.02	15.23
September 30th	15.29	19.31	17.36
October 6th	18.26	23.50	20.65
October 13th	19.22	23.54	22.03
October 20th	20.60	24.33	23.08

4.2 光谱分析

图 2 为不同含油率油茶籽的中红外光谱图,可以看出,不同时期采摘的油茶籽的中红外光谱峰形相似,

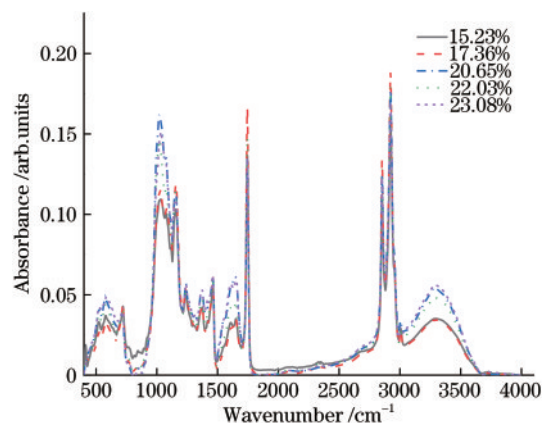


图 2 不同含油率油茶籽的中红外吸收光谱图

Fig. 2 Mid-infrared absorption spectra of camellia seeds with different oil contents

但是存在部分交叉重叠的现象,体现在吸收峰处的频率位置没有变化,只是强度上存在一定改变,无法直观地区分不同含油率的油茶籽。图 3 为远红外光谱图,不同含油率的油茶籽在远红外波段并无明显的吸收峰,但是吸收系数存在一定差异,这是由于油茶籽成分复杂包含多种不同的营养物质,不同物质对远红外波段的吸收互相干扰导致无明显吸收峰。

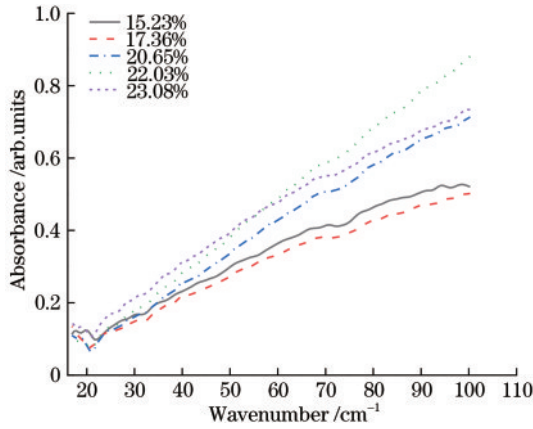


图 3 不同含油率油茶籽的远红外吸收光谱图

Fig. 3 Far-infrared absorption spectra of camellia seeds with different oil contents

4.3 基于中红外和远红外光谱数据的模型建立

在建立模型前,随机选取不同含油率油茶籽的 14 个数据作为训练集,其余 6 个数据作为测试集,共计 70 个训练集,30 个预测集。将含油率为 15.23% 的油

茶籽定义为标签“1”,将含油率为 17.36%、20.6%、22.03%、23.08% 的油茶籽分别定义为标签“2、3、4、5”,来表征相应阶段油茶籽的成熟度,含油率越高,油茶果成熟度越高。整个建模过程分两步进行,首先利用默认参数的 SVM 模型确定最佳特征变量提取方式,其次利用三种寻优算法(GS、GA、PSO)对 SVM 模型进行参数寻优,提出通过油茶籽的含油率大小反映油茶籽成熟度的最佳鉴别模型。

表 2 为对于远红外和中红外光谱数据,不同特征变量选取方法建立的 SVM 模型预测结果。中红外波段经过主成分分析后前 9 个主成分的贡献率之和为 98.07%,远红外波段经过主成分分析后的第 1 个主成分的贡献率达到 85.06%,前 9 个主成分的累计贡献率达到了 99.52%,证明前 9 个主成分包含了样本原始光谱的大量信息。连续投影算法中设定选取的最大变量数为 60,最小变量数为 10,中红外光谱筛选出 32 个变量,远红外光谱筛选出 23 个变量。中红外光谱经过 UVE 挑选的阈值为 35.1418,对上下阈值之间的光谱变量剔除后挑选出了 430 个变量。远红外光谱经过 UVE 挑选的阈值为 21.1676,挑选出 163 个变量。对于中红外波段,SPA-SVM 模型的准确率最高能达到 83.33%,对于远红外波段,PCA 与 SPA 的效果相当,但 PCA 筛选出的变量数更少,PCA-SVM 的准确率能达到 90%。经特征选取后的预测精度相比于原始光谱建模都有所改善,因此选择 SPA 和 PCA 分别作为中红外与远红外波段数据的最佳变量选择方法。

表 2 中红外与远红外光谱不同特征提取方法下的 SVM 模型预测结果

Table 2 Prediction results of SVM model under different feature extraction methods of mid-infrared and far-infrared spectra

Spectral type	Feature extraction method	Number of variables	Number of misjudgments in the training set	Number of misjudgments in the prediction set	Training set accuracy /%	Prediction set accuracy /%
Mid-infrared	NONE	1868	14	9	80.00	70.00
	PCA	9	5	6	92.86	80.00
	SPA	32	5	5	92.86	83.33
	UVE	430	7	8	90.00	73.33
Far-infrared	NONE	174	3	4	95.71	86.67
	PCA	9	2	3	97.14	90.00
	SPA	23	2	3	97.14	90.00
	UVE	163	3	4	95.71	86.67

确定最佳变量选择方法后,分别利用 GS、GA 和 PSO 等优化算法对 SVM 模型里的惩罚参数 c 和核函数参数 g 进行优化,采用 GA 和 PSO 算法对参数优化时,首先要初始化算法的参数,其中 GA 算法的种群最大数量为 20,最大的进化代数设为 200,惩罚参数 c 和核函数参数 g 的范围设置为 (0, 100),交叉验证折数为 5。PSO 算法的种群最大数量为 20,最大进化数量为 200,参数局部搜索能力为 1.5,参数全局搜索能力为 1.7,惩罚参数 c 的范围为 (0.01, 100),核函数参数 g 的

范围为 (0.01, 1000),利用优化后的最佳参数 c 、 g 分别建立对应的 SVM 模型,优化后的模型结果如表 3 所示。使用算法优化后的 SVM 模型的预测准确率与默认参数相比有明显提高,基于中红外波段数据建立的模型经 GA 算法优化后的准确率最高为 93.33%,此时最佳参数 (c, g) 确定为 (2.8523, 7.0050)。远红外波段下经 GA 和 PSO 算法优化后的准确率均能达到 96.67%,由于过高的惩罚系数 c 会导致过拟合的产生,使得模型的泛化能力变差,在同等效果的模型下,

表 3 优化后的 SVM 模型分类结果对比

Table 3 Comparison of classification results of optimized SVM model

Spectral type	Parameter optimization method	SVM parameter		Number of misjudgments	Prediction set accuracy / %
		c	g		
Mid-infrared	GS	2.2974	6.9644	3	90.00
	GA	2.8523	7.0050	2	93.33
	PSO	5.8397	9.7162	3	90.00
Far-infrared	GS	1.0000	2.8284	2	93.33
	GA	3.3453	0.4628	1	96.67
	PSO	60.1000	0.0100	1	96.67

应选择惩罚参数较小的那组数据,因此最佳参数(c, g)确定为(3.3453, 0.4628)。

4.4 基于中级数据融合的模型建立

中红外波段与远红外波段所包含的光谱信息不同,如果仅使用单一波段光谱技术对样品进行检测,则只能获得样品的部分信息,特别是像油茶籽这种成分比较复杂的样品,获得的信息可能具有片面性。与低级数据融合策略不同,中级数据融合是对不同的数据源先进行特征提取后,再将提取后的变量进行拼接,这样可提取有效的光谱信息。本文把 SPA 提取后的中红外光谱数据的特征变量与 PCA 提取后的远红外光

谱特征变量融合成新的变量矩阵,结合经参数优化的 SVM 模型进行油茶籽含油率的鉴别。与单一光谱波段数据处理过程一样,随机选取不同含油率油茶籽的 14 个数据作为训练集,其余 6 个数据作为测试集,共计 70 个训练集,30 个预测集。参数寻优及交叉验证过程如图 4 所示,三种优化算法的模型交叉验证(5 折交叉验证)准确率都达到了 98.5714%。图 4(b)是由图 4(a)投影在二维平面上得到的,随着交叉验证准确率的不断升高,不同 c, g 值构成的网格颜色以及等高线颜色逐渐由深蓝色变成明黄色,当惩罚参数 $c=0.7579$ 和核函数参数 $g=1.3195$ 时,交叉验证的准确率达到最高。

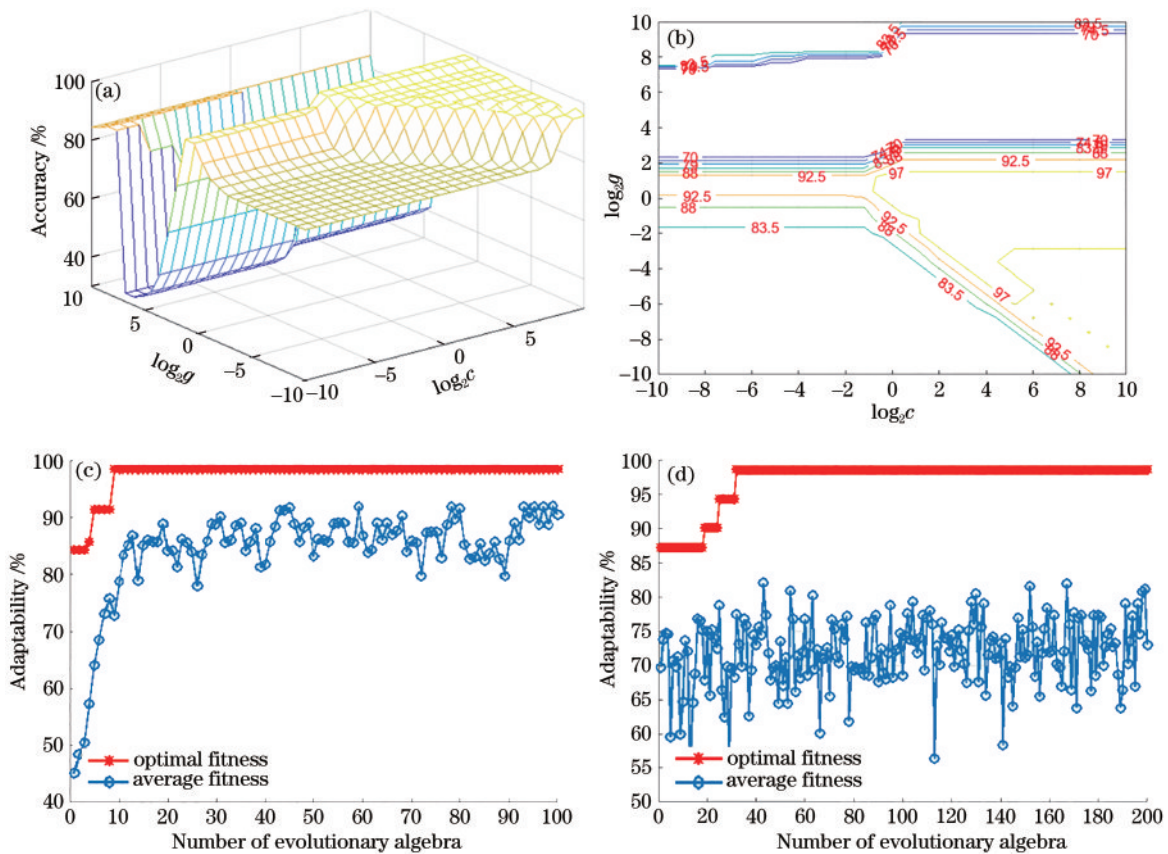


图 4 基于中级数据融合的 SVM 模型 c, g 参数优化结果图。(a) GS 算法优化后的三维视图;(b) GS 算法优化后的等高线图;(c) GA 算法优化后的准确率曲线;(d) PSO 算法优化后的准确率曲线

Fig. 4 Optimization results of c and g parameters of SVM model based on intermediate data fusion. (a) Three-dimensional view optimized by GS algorithm; (b) contour map optimized by GS algorithm; (c) accuracy curve optimized by GA algorithm; (d) accuracy curve optimized by PSO algorithm

图 4(c)是遗传算法优化参数的结果图,可以看出,遗传算法优化的准确率经过 10 次左右的迭代稳定在最佳准确率,优化后最佳的惩罚参数 $c=4.3098$,核函数参数 $g=1.3823$ 。图 4(d)是经过粒子群算法优化的结果图,由其交叉验证的准确率在经过 40 次的迭代后稳定在最优,优化后最佳的惩罚参数 $c=1.1067$,核函数参数 $g=1.2319$ 。

把每种算法搜索到的最优惩罚参数 c 和核函数参数 g 代入到 SVM 模型中并根据预测集的准确率来比较模型的好坏,所建立的 SVM 模型结果如表 4 所示。可以看出,经过 GA 和 PSO 算法优化后的模型预测集

的判别准确率达到 100%,其中采用 PSO 算法优化后的惩罚参数 c 最小。故基于中级数据融合的油茶籽含油率鉴别模型中的最佳惩罚参数 $c=1.1067$,核函数参数 $g=1.2319$,此时的预测准确率达到 100%,这是因为不同波段所包含的信息量不同,所以不同特征提取方法的效果也不一样,中红外波段的油茶籽吸收峰较多,通过 SPA 算法提取的特征变量很好地覆盖了吸收峰附近的区域,远红外波段并无明显吸收峰,通过 PCA 算法有效地降低了变量的维数,而中级数据融合利用 SPA 和 PCA 两种特征选取方法,在尽可能保留两个波段有效特征信息的同时去除了干扰信息。

表 4 基于中级数据融合的参数优化后 SVM 模型分类结果

Table 4 Results of SVM model after parameter optimization based on intermediate data fusion

Optimization algorithm	Optimal parameters		Number of misjudgments	Prediction set accuracy / %
	c	g		
Grid search	0.7579	1.3195	1	96.67
Genetic algorithm	4.3098	1.3823	0	100
Particle swarm algorithm	1.1067	1.2319	0	100

5 结 论

本文利用傅里叶变换红外光谱仪测试了不同成熟时期油茶籽的中红外和远红外光谱,以油茶籽含油率作为成熟度指标,探讨了不同光谱类型及中级数据融合策略对油茶籽成熟度鉴别的可行性。中红外和远红外光谱均能实现油茶籽成熟度的鉴别,通过对比不同的特征提取方法可以发现:对于中红外光谱,SPA-SVM 模型的精度最高,准确率能达到 83.33%;对于远红外光谱,PCA-SVM 与 SPA-SVM 模型的效果相同,准确率都能达到 90%,但经 PCA 算法提取的变量数更少,所以 PCA-SVM 模型更优。通过采用三种不同优化算法对 SVM 模型里的惩罚参数 c 和核函数参数 g 进行优化发现:对于中红外光谱,经 GA 算法优化后的 SVM 模型的精度最高,准确率为 93.33%,此时惩罚参数 $c=2.8523$,核函数参数 $g=7.0050$;对于远红外光谱,经 GA 和 PSO 算法优化后的 SVM 模型的精度最高,准确率为 96.67%,但 GA 优化后的惩罚参数较小,模型泛化能力更强,此时惩罚参数 $c=3.3453$,核函数参数 $g=0.4628$ 。采用中级数据融合策略把中红外光谱和远红外光谱数据融合,结合 SVM 建立油茶籽成熟度鉴别模型,最佳模型为 PSO-SVM,准确率能达到 100%。对比数据融合后的光谱与单一光谱,使用数据融合后的光谱数据建立的 SVM 模型精度要比单一光谱高,实现了对油茶籽成熟度的精准预测。因此,基于中级数据融合的优化方法可以在增加有效光谱信息的同时去除单一光谱里的冗余信息,有利于油茶籽成熟度的快速精准鉴别。本文中用于建模的样本总数较少(100 个),后期可通过增加采摘时间及频次提高样本数,进一步提升模型的预测精度和泛化能

力。另外,虽然含油率是评价油茶籽成熟度的重要指标,但在含油率变化较小的情况下,预测准确率可能会降低,可通过构建多元化的油茶籽成熟度评价体系,优化模型算法,提高建模精度,为油茶采摘时间提供有效指导。

参 考 文 献

- [1] 庄瑞林. 中国油茶[M]. 2 版. 北京: 中国林业出版社, 2008.
Zhuang R L. Chinese camellia[M]. 2nd ed. Beijing: China Forestry Publishing House, 2008.
- [2] 王亚萍, 费学谦, 石晓丽, 等. 采收期和处理方法对油茶籽及其油脂营养物质积累的影响[J]. 中国油脂, 2017, 42(4): 20-23, 46.
Wang Y P, Fei X Q, Shi X L, et al. Effects of harvest time and treatment methods on nutrient accumulation of oil-tea camellia seed and its oil[J]. China Oils and Fats, 2017, 42(4): 20-23, 46.
- [3] 曾奥, 陈元堃, 罗振辉, 等. 茶籽油不同提取工艺的特点分析及研究进展[J]. 广东药科大学学报, 2020, 36(1): 151-154.
Zeng A, Chen Y K, Luo Z H, et al. Characteristics and research progress of extraction technology of tea seed oil [J]. Journal of Guangdong Pharmaceutical University, 2020, 36(1): 151-154.
- [4] 吴雪辉, 王泽富. 红外光谱快速测定油茶籽油中主要功能活性成分含量的模型建立与评价[J]. 中国油脂, 2018, 43(10): 140-143.
Wu X H, Wang Z F. Establishment and evaluation of models for rapid detection of functional active ingredients in oil-tea camellia seed oil by infrared spectroscopy[J]. China Oils and Fats, 2018, 43(10): 140-143.
- [5] 王泽富, 吴雪辉, 江盛宇, 等. 红外光谱快速检测油茶籽油有关指标的研究[J]. 中国粮油学报, 2018, 33(3):

- 119-125.
Wang Z F, Wu X H, Jiang S Y, et al. Rapid determination on relevant indexes of camellia oil by infrared spectroscopy[J]. Journal of the Chinese Cereals and Oils Association, 2018, 33(3): 119-125.
- [6] 文韬, 郑立章, 龚中良, 等. 基于近红外光谱技术的茶油原产地快速鉴别[J]. 农业工程学报, 2016, 32(16): 293-299.
Wen T, Zheng L Z, Gong Z L, et al. Rapid identification of geographical origin of camellia oil based on near infrared spectroscopy technology[J]. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32(16): 293-299.
- [7] 马本学, 喻国威, 王文霞, 等. 光谱分析在西甜瓜内部品质无损检测中的研究进展[J]. 光谱学与光谱分析, 2020, 40(7): 2035-2041.
Ma B X, Yu G W, Wang W X, et al. Recent advances in spectral analysis techniques for non-destructive detection of internal quality in watermelon and muskmelon: a review[J]. Spectroscopy and Spectral Analysis, 2020, 40(7): 2035-2041.
- [8] 李雪, 王督, 白艺珍, 等. 我国油料产品品质的近红外光谱快速检测技术研究进展[J]. 分析测试学报, 2020, 39(10): 1189-1195.
Li X, Wang D, Bai Y Z, et al. Research progress of near infrared spectroscopy in rapid determination on quality of oilseed products in China[J]. Journal of Instrumental Analysis, 2020, 39(10): 1189-1195.
- [9] 赵娟, 全朋坤, 马敏娟, 等. 富士苹果采收成熟度光谱无损预测模型对比分析[J]. 农业机械学报, 2018, 49(12): 347-354.
Zhao J, Quan P K, Ma M J, et al. Comparative analysis of harvest maturity model for fuji apple based on visible/near spectral nondestructive detection[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(12): 347-354.
- [10] 邵园园, 王永贤, 玄冠涛, 等. 基于高光谱成像的肥城桃品质可视化分析与成熟度检测[J]. 农业机械学报, 2020, 51(8): 344-350.
Shao Y Y, Wang Y X, Xuan G T, et al. Visual detection of SSC and firmness and maturity prediction for Feicheng peach by using hyperspectral imaging[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 344-350.
- [11] Minas I S, Blanco-Cipollone F, Sterle D. Accurate non-destructive prediction of peach fruit internal quality and physiological maturity with a single scan using near infrared spectroscopy[J]. Food Chemistry, 2021, 335: 127626.
- [12] 沙云菲, 黄雯, 王亮, 等. 中红外和近红外数据融合香型风格判别[J]. 光谱学与光谱分析, 2021, 41(2): 473-477.
Sha Y F, Huang W, Wang L, et al. Merging MIR and NIR spectral data for flavor style determination[J]. Spectroscopy and Spectral Analysis, 2021, 41(2): 473-477.
- [13] 李亚惠, 李艳肖, 谭伟龙, 等. 基于近、中红外光谱法融合判定黑果枸杞产地及品质信息[J]. 光谱学与光谱分析, 2020, 40(12): 3878-3883.
Li Y H, Li Y X, Tan W L, et al. Rapid identification and evaluation of lycium ruthenicum Murr. by near-infrared and Fourier transform infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2020, 40(12): 3878-3883.
- [14] Ballabio D, Robotti E, Grisoni F, et al. Chemical profiling and multivariate data fusion methods for the identification of the botanical origin of honey[J]. Food Chemistry, 2018, 266: 79-89.
- [15] Chen T, Zhang Q, Li Z, et al. Experimental and theoretical investigations of tartaric acid isomers by terahertz spectroscopy and density functional theory[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 205: 312-319.
- [16] Ren G H, Zong S Q, Zhu Z J, et al. Far-infrared terahertz properties of L-cysteine and its hydrochloride monohydrate[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 225: 117476.
- [17] 刘俊秀, 杜彬, 邓玉强, 等. 基于差分-主成分分析-支持向量机的有机化合物太赫兹吸收光谱识别方法[J]. 中国激光, 2019, 46(6): 0614039.
Liu J X, Du B, Deng Y Q, et al. Terahertz-spectral identification of organic compounds based on differential PCA-SVM method[J]. Chinese Journal of Lasers, 2019, 46(6): 0614039.
- [18] 刘燕德, 张雨, 徐海, 等. 基于近红外光谱检测不同产地石榴的糖度[J]. 激光与光电子学进展, 2020, 57(1): 013002.
Liu Y D, Zhang Y, Xu H, et al. Detection of sugar content of pomegranates from different producing areas based on near-infrared spectroscopy[J]. Laser & Optoelectronics Progress, 2020, 57(1): 013002.
- [19] 刘燕德, 徐振, 胡军, 等. 太赫兹光谱检测技术对农产品品质检测的研究进展[J]. 激光与光电子学进展, 2021, 58(1): 0100005.
Liu Y D, Xu Z, Hu J, et al. Research on quality of agricultural products by terahertz spectroscopy[J]. Laser & Optoelectronics Progress, 2021, 58(1): 0100005.
- [20] 张文涛, 李跃文, 占平平, 等. 基于太赫兹时域光谱技术与 PCA-SVM 的转基因大豆油鉴别研究[J]. 红外与激光工程, 2017, 46(11): 1125004.
Zhang W T, Li Y W, Zhan P P, et al. Recognition of transgenic soybean oil based on terahertz timedomain spectroscopy and PCA-SVM[J]. Infrared and Laser Engineering, 2017, 46(11): 1125004.
- [21] 张晓雪, 杨志辉, 曹珊珊, 等. 基于支持向量机的甘薯冷害光谱检测方法[J]. 农业机械学报, 2020, 51(S2): 471-477.
Zhang X X, Yang Z H, Cao S S, et al. Spectral detection method for chilling damage of sweet potato based on support vector machine[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(S2): 471-477.
- [22] Devos O, Ruckebusch C, Durand A, et al. Support vector machines (SVM) in near infrared (NIR) spectroscopy: focus on parameters optimization and model interpretation

- [J]. Chemometrics and Intelligent Laboratory Systems, 2009, 96(1): 27-33.
- [23] Liu J J, Li Z, Hu F R, et al. A THz spectroscopy nondestructive identification method for transgenic cotton seed based on GA-SVM[J]. Optical and Quantum Electronics, 2015, 47(2): 313-322.
- [24] Li R D, Zhang X B, Li K Q, et al. Nondestructive and rapid grading of tobacco leaves by use of a hand-held near-infrared spectrometer, based on a particle swarm optimization-extreme learning machine algorithm[J]. Spectroscopy Letters, 2020, 53(9): 685-691.
- [25] 王书涛, 展书杰, 刘诗瑜, 等. 三维荧光光谱结合 ICSO-SVM 对性激素的分类鉴别[J]. 光学学报, 2021, 41(10): 1030004.
- Wang S T, Zhan S J, Liu S Y, et al. Classification and identification of sex hormones by three-dimensional fluorescence spectroscopy combined with ICSO-SVM[J]. Acta Optica Sinica, 2021, 41(10): 1030004.