

基于可见-近红外光谱法无损检测赣南脐橙中总酸含量

章海亮^{1*}, 任众财², 刘雪梅³, 罗微³, 詹白勺³, 黄海华⁴, 陈宏⁴

(华东交通大学电气与自动化工程学院, 南昌 330000)

摘要: **目的** 基于可见-近红外光谱法建立一种无损测定赣南脐橙总酸含量的技术。**方法** 利用设计的可见-近红外光谱检测系统检测 168 个赣南脐橙总酸含量。以给定赣南脐橙的 126 个样品作为校正集, 42 个未知样品作为预测集。本研究以去除首尾处噪声后的 400~880 nm 范围的光谱波段, 共 481 个波长点进行研究分析。结合 SG (Savitzky-Golay)平滑法、多元散射校正法、变量标准化法、基线校正法 4 种预处理方法处理原始光谱数据, 通过 PLSR 数学模型确定最佳预处理模型; 再利用竞争性自适应重加权算法(competitive adaptive reweighted sampling, CARS)、随机蛙跳算法(random frog, RF)、遗传算法(genetic algorithm, GA)、连续投影算法(successive projections algorithm, SPA)和主成分分析法(principal component analysis, PCA) 5 种算法对预处理后的数据提取特征变量, 降低维度, 随后分别建立基于特征变量的总酸偏最小二乘回归(partial least squares regression, PLSR)、主成分回归(principal component regression, PCR)、最小二乘支持向量机(least square support vector machine, LS-SVM)及多元线性回归(multiple linear regression, MLR)预测模型。**结果** 通过 PLSR 数学模型确定 SG 平滑预处理模型效果为最佳, 基于 SG+GA+LS-SVM 模型对总酸含量预测效果最佳, 预测集均方根误差(root mean square error of prediction, RMSEP)值为 0.016, 预测集决定系数(prediction set coefficient of determination, R_p^2)值为 0.9834, 相对分析误差(residual predictive deviation, RPD)值为 7.76。**结论** 基于可见-近红外光谱法实现赣南脐橙中总酸含量的无损检测是可行的, 结合 SG+GA+LS-SVM 预测模型可以实现赣南脐橙总酸含量的定量检测, 可用于评价赣南脐橙总酸含量。

关键词: 可见-近红外光谱法; 无损检测; 总酸含量; 赣南脐橙

Non-destructive detection of total acid content in Gannan navel orange based on visible-near infrared spectroscopy

ZHANG Hai-Liang^{1*}, REN Zhong-Cai², LIU Xue-Mei³, LUO Wei³, ZHAN Bai-Shao³, HUANG Hai-Hua⁴, CHEN Hong⁴

(School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330000, China)

ABSTRACT: Objective To establish a non-destructive method for the determination of total acid content in

基金项目: 国家自然科学基金项目(41867020)、江西省科技厅项目(20212ABC03A17、20212ABC03A32、20203BBF63031)、江西省教育厅项目(GJJ200609)

Fund: Supported by the National Natural Science Foundation of China (41867020), the Project of Science and Technology Department of Jiangxi Province (20212ABC03A17, 20212ABC03A32, 20203BBF63031), and the Project of Department of Education of Jiangxi Province (GJJ200609)

***通信作者:** 章海亮, 博士, 教授, 主要研究方向为智能检测以及光谱检测等。E-mail: 2242526723@qq.com

***Corresponding author:** ZHANG Hai-Liang, Ph.D, Professor, School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330000, China. E-mail: 2242526723@qq.com

Gannan navel orange based on visible-near infrared spectroscopy. **Methods** The designed visible-near infrared spectroscopy detection system was used to detect the total acid content in 168 Gannan navel oranges. The 126 samples in the given Gannan navel orange were used as the modeling set, and the 42 unknown samples served as prediction sets. In this experiment, a total of 481 wavelength points were studied and analyzed in the 400-880 nm range of the spectrum band after removing the noise at the head and tail. The original spectral data were processed by 4 kinds of preprocessing methods including SG (Savitzky-Golay) smoothing method, multivariate scattering correction method, standard normal variable method, and baseline offset correction method, and the optimal pretreatment model was determined by PLSR mathematical model; then 5 kinds of algorithms including competitive adaptive reweighted sampling (CARS), random frog (RF), genetic algorithm (GA), successive projections algorithm (SPA) and principal component analysis (PCA) were used to extract characteristic variables and reduce the dimensions of the data information, the partial least squares regression (PLSR)、principal component regression (PCR)、least square support vector machine (LS-SVM) and multiple linear regression (MLR) prediction models of total acid content in Gannan navel orange based on the characteristic variables were established. **Result** The prediction effect of total acid content based on SG+GA+LS-SVM model was the best. The root mean square error of prediction (RMSEP) value was 0.016, the prediction set coefficient of determination (R_p^2 value was 0.9834, and the residual predictive deviation (RPD) value was 7.76. **Conclusion** It is feasible to realize the non-destructive detection of the total acid content in the Gannan navel orange based on visible-near infrared spectroscopy, which can realize the quantitative detection of the total acid content of Gannan navel orange combining with the SG+GA+LS-SVM prediction model and can be used to evaluate the total acid content of Gannan navel orange.

KEY WORDS: visible-near infrared spectroscopy; non-destructive examination; total acid content; Gannan navel orange

0 引言

柑橘属水果是世界上最重要的水果之一,包括柚、橘(桔)、橙等。目前,世界上有超过 130 个国家种植,其在中国和世界贸易中都有着巨大的市场。水果中总酸、可溶性固形物、矿物质等理化指标是衡量水果质量的关键指标,其中总酸含量是评价农产品的一个常用参数,对水果的口感、销售价格等具有重要的影响。

可见-近红外光谱法具有无损、快速、低成本的检测特点,如今已被广泛运用于农业、食品等领域。目前,有不少学者基于可见-近红外光谱法对柑橘属水果中总酸含量、可溶性固形物等理化值的无损检测进行了相关研究,例如, LIAO 等^[1]采用脐橙近红外光谱结合偏最小二乘法(partial least square, PLS)建立预测模型测定了脐橙中可滴定酸度,预测集决定系数(prediction set coefficient of determination, R_p^2)到达了 0.9085;章海亮等^[2]基于一阶微分结合偏最小二乘回归(partial least squares regression, PLSR)成功实现脐橙总酸含量的无损快速检测, R_p^2 到达了 0.9562;江水泉等^[3]基于遗传算法-竞争性自适应重加权算法(genetic algorithm-competitive adaptive reweighted sampling, GA-CARS)实现了脐橙中可溶性固形物含量在线检测, R_p^2 到

达了 0.824;刘燕德等^[4-7]实现了赣南脐橙以及南丰蜜桔等柑橘属水果的可溶性固形物无损检测,其中近红外光谱结合最小二乘支持向量机(least square support vector machine, LS-SVM)预测模型效果最优, R_p^2 到达了 0.9059;李俊杰等^[8]利用近红外光谱结合 PLS 实现了塔罗科血橙总酸的研究, R_p^2 到达了 0.699;吕强等^[9]采用联合区间偏最小二乘法(synergy interval partial least square, siPLS)对班菲尔脐橙的可溶性固形物进行了无损研究, R_p^2 到达了 0.8789;薛龙等^[10]基于 GA-PLS 实现了脐橙可溶性固形物的研究, R_p^2 到达了 0.9132。

以上研究为本研究总酸含量检测提供了思路,但是以上学者所创建的预测模型较少,也就导致了总酸含量与光谱数据之间的相关性较低,本研究通过 CARS^[11]、随机蛙跳算法(random frog, RF)^[12]、GA^[13]、连续投影算法(successive projections algorithm, SPA)^[14]、主成分分析法(principal component analysis, PCA)^[15]对脐橙光谱特征变量进行提取,再结合 PLSR^[16]、多元线性回归(multiple linear regression, MLR)^[17]、主成分回归(principal component regression, PCR)^[18]、LS-SVM^[19]建立多种预测模型,增强总酸含量与近红外光谱之间的相关性,探求一种快捷、准确的预测模型,为评价柑橘属水果总酸含量提供更加准确的检测方法。

1 材料与方法

1.1 材料与试剂

脐橙样本均来自江西省某大型农贸批发市场, 挑选出 168 个大小相近, 形状差异较小且无损伤的赣南脐橙样品。以 3:1 比例将脐橙样本分为校正集和预测集, 并对样品逐一排序, 对 126 个建模样品进行训练, 42 个未知样品进行验证。

试剂: 氢氧化钠(分析纯, 中国西安鼎天化工有限公司); 酚酞(分析纯, 中国江西洛特化工有限公司); 无水乙醇(分析纯, 中国江西雪奥化工有限公司)。

1.2 仪器与设备

ImSpectorV10E+QE 型号成像光谱仪(芬兰 Spectral Imaging 公司); HR1895 型榨汁机、100W 卤素灯光源(荷兰 Philips 公司); V23+f/2.4 型号透镜(芬兰 Specim 公司); DCR III 型号控制器(美国 Schoot 公司); EZHR17EN 型号的样品平台(美国 AllMotion 公司); 台式计算机(中国联想公司)。

1.3 设备检测原理

为了避免光谱采集过程中周围环境杂散光的影响, 整个系统安装在涂有哑光黑漆的铁柜中, 可见-近红外光谱检测系统的原理如图 1 所示。调整透镜到输送台的距离为 400 mm。两个卤素灯光源分别安装在距离水平位置 45° 角的位置上。采集阶段的移动速度和扫描时间分别设定为 0.85 mm/s 和 20 ms/次。对脐橙样品依次编号, 避开存在果皮缺陷的部分, 在样品赤道线等间距连续做 3 次光谱扫描, 最后取 3 幅图谱平均值作为最终光谱数据。后续利用 The Unscramble X10.1, Matlab 2010 以及 Origin 2018 对光谱数据进行处理分析。

1.4 总酸含量的测定

将赣南脐橙样品剥皮后, 清理掉橘络, 取脐橙果肉放入榨汁机中, 榨汁并且过滤。取上清液, 采用传统滴定法^[20]测定样品中的总酸含量, 重复取样 3 次, 最后取均值作为样品的总酸测定值。

1.5 光谱的预处理

以消除首尾处噪声后的原始数据, 取 400~880 nm 的波段范围内, 共计 481 个波长点作为输入变量 X , 以吸光度作为输出变量 Y 。基于 4 种预处理方法, 分别包括 SG 平滑法、多元散射校正法、变量标准化法、基线校正法。

1.6 特征光谱的选择

原始数据中存在大量的冗余信息, 使得模型的建立和数值的计算较为烦琐。基于适当的算法处理光谱数据, 以此来减少低效率和冗余数据对实验的影响, 简化数学模型, 同时减少计算量。本研究采用了 CARS、RF、GA、SPA

和 PCA 5 种算法简化数据计算。

CARS 用来消除光谱数据中的冗余的、涵盖信息量少的数据, 利用交叉验证提取出当均方根误差较低时的子集, 选出良好的变量组合, 替代原始光谱数据。以此来建模, 简化数学模型, 提高计算结果的准确性^[11]; RF 广泛应用于光谱数据特征波长提取, 通过算法的运行结果来显示每个波长迭代的概率, 然后根据概率大小提取出迭代概率高的波长。并以挑选出的少量的波长作为特征变量进行建模^[12]; GA 是基于物种进化论, 是进化算法的一种。通过遗传、变异、交叉、复制, 来寻找全局最优解^[13]; SPA 是一种常用的经典降维算法, 广泛运用于降低高维度数据, 简化数学计算。在本次赣南脐橙总酸含量检测实验中, 以样本的光谱数据和总酸含量作为输入, 以 5~20 为选取的波长数量范围输出特征波长^[14]; PCA 广泛应用于数据处理, 是一种常用的高维度数据降维方法, 从大数据库中选取少量的涵盖主特征信息的数据代替原始的, 大量的数据^[15]。

1.7 模型的建立

本研究赣南脐橙总酸含量的无损检测实验, 选择 PLSR、MLR、PCR、LS-SVM 4 种数学模型^[16-19], 并对对建立的各个数学预测模型进行比较分析, 验证总酸含量预测模型的准确性和稳定性, 从而选择最优的预测模型完成总酸含量的定量检测。

基于 PLSR 将所提取的光谱数据与总酸含量进行线性回归分析, 对预处理模型进行评价, 通过均方根误差以及决定系数评价预处理模型; MLR 同样用来线性回归分析, 广泛应用于处理大量数据, 在实际的生产生活中, 它能够更加接近真实的机器学习应用; PCR 通常来说是 MLR 模型的替代方法, 相比较于 MLR 模型适用更加广泛, 具有很大的优势, 同样用于构建未知的预测变量与已知的预测变量之间的线性回归; LS-SVM 是支持向量机(SVM)的改进, 比较适用于样本容量小的数据, 能够提高非线性信息数据的准确度以及模型的运算速度和性能。

1.8 数学模型评价参数

分别采用校正集均方根误差(root mean square error of calibration, RMSEC)、预测集均方根误差(root mean square error of prediction, RMSEP)、校正集决定系数(calibration set coefficient of determination, R_c^2)、 R_p^2 以及相对分析误差(residual predictive deviation, RPD)评价预测模型性能。当 RMSEC 与 RMSEP 之间的差值越小且均趋向于 0 时, 说明此模型的检测效果越稳健; R_c^2 和 R_p^2 值越趋向于 1 时, 说明此模型越具有较高的检测精度; 当 RPD>2.0 且 $R_p^2 \geq 0.80$ 时, 说明此模型对总酸含量的预测效果较优, 适用于对样品进行定量预测; 当 $1.4 \leq RPD \leq 2.0$ 且 $0.50 \leq R_p^2 < 0.80$ 时, 说明

此模型对总酸含量的预测效果较良,适用于近似地对样品进行预测;当 $RPD < 1.4$ 且 $R_p^2 < 0.50$ 时,说明此模型对总酸含量的预测效果较差,无法对样品进行预测^[21]。

2 结果与分析

2.1 样本集的划分

样本集总酸含量统计结果如表1所示,校正集样本总酸含量的极差、平均值、标准偏差分别为0.64、3.52、0.12;预测集样本总酸含量的极差、平均值、标准偏差分别为0.49、

3.54、0.11。预测集总酸含量的统计范围处于校正集的统计范围之内,表明实验样本的划分相对合理;校正集与预测集总酸含量的离散系数分别为3.57%、3.01%,可以看出样本离散程度较小,具有很好的稳定性。

2.2 最佳预处理模型的确定

图2显示基于PLSR回归模型对赣南脐橙中总酸含量的校正结果以及预测结果,可以明显看出,总酸测量值与预测值均向线性回归线靠拢,表明脐橙总酸含量的测量值与预测值具有很高的线性相关性。

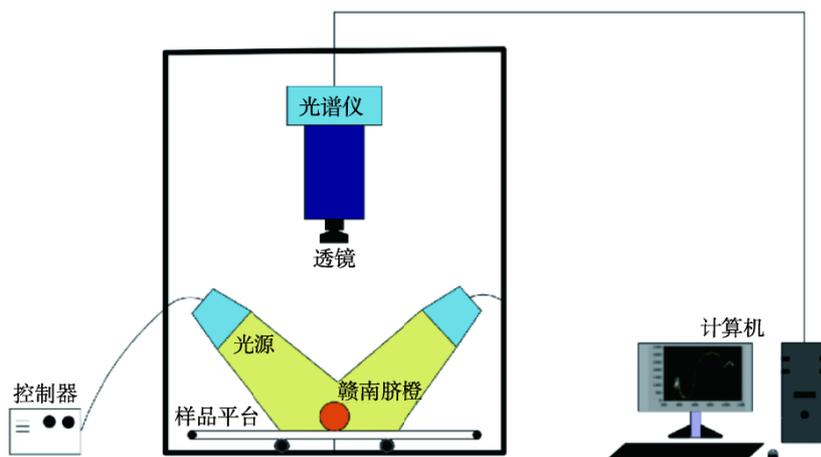


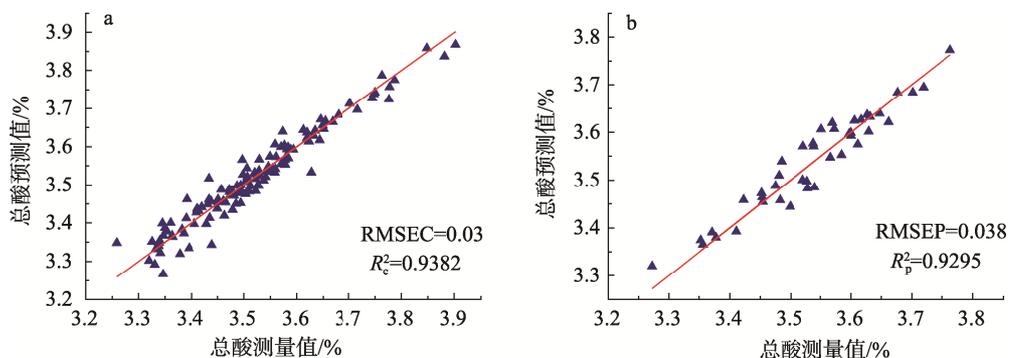
图1 可见-近红外光谱检测原理

Fig.1 Detection principle of visible-near infrared spectroscopy

表1 赣南脐橙样品总酸结果的统计

Table 1 Statistics of total acid results in Gannan navel orange samples

	变量数	最大值/%	最小值/%	平均值/%	极差/%	标准偏差	离散系数/%
校正集	126	3.90	3.26	3.52	0.64	0.12	3.57
预测集	42	3.76	3.27	3.54	0.49	0.11	3.01



注: a 为总酸含量的校正结果; b 为总酸含量的预测结果。

图2 PLSR 数学模型对脐橙总酸含量的校正和预测结果

Fig.2 Correction and prediction results of total acid content in navel orange using PLSR mathematical model

基于不同的预处理模型分别建立 PLSR 数学模型, 不同预处理模型的校正和预测结果如表 2 所示。通过分析可知, 原始光谱数据经过 SG 平滑预处理后, 总酸含量的校正以及预测效果最优。SG 平滑预处理模型中的 R_c^2 和 R_p^2 的值分别是 0.9530 和 0.9401。RMSEC 和 RMSEP 的值分别是 0.027 和 0.028。其预处理效果优于其他 3 种预处理模型的效果, 因为 SG 平滑算法的基本原理是假设光谱噪声是以零为中心产生的均匀随机白噪声, 通过多次测量求平均值的方法来消除噪声, 以此来提高样品的信号, 该算法相比于其他算法更稳定、误差更小。因此本研究确定 SG 平滑法为最佳预处理模型。

表 2 基于不同预处理方法对脐橙总酸含量的 PLSR 模型预测效果

Table 2 Effects of PLSR model prediction of total acid content in Gannan navel orange based on different pretreatment methods

预处理方法	变量数	R_c^2	RMSEC	R_p^2	RMSEP
Original spectra	481	0.9382	0.03	0.9295	0.038
SG	481	0.9530	0.027	0.9401	0.028
SNV	481	0.8510	0.047	0.7743	0.054
MSC	481	0.8510	0.047	0.7740	0.054
Baseline	481	0.8451	0.048	0.8730	0.041

2.3 光谱数据的优化

2.3.1 基于 CARS 算法提取特征波长

在 CARS 算法中, 设置蒙特卡罗采样次数 $N=50$, 由图 3a 和图 3b 可知, 当采样次数 $N=34$ 时, RMSE 值最小, 表明在第 1~33 次采样中, 大量的无关信息被剔除, 第 34 次采样之后, RMSE 值快速上升, 表明在第 34 次采样之后, 不断有涵盖主特征信息的数据被剔除, 使模型变差; 观察图 3c 各个波长回归系数的趋势变化, 竖线所在的位置表示此时的波长变量子集中所提取的波长是最优的, 包含 12 个特征变量, 分别包括 418、420、427、445、447、462、

492、500、517、679、693、855 nm, 涵盖了原始数据中大部分特征信息, 作为提取的特征波长。

2.3.2 基于 RF 算法提取特征波长

RF 算法的运行结果如图 4 所示。由图 4a 可知, 当数目为 6 时, RMSE 值最小, 同时在图 4b 中找到被选择概率靠前的 6 个波长变量, 即 462、693、749、750、766、791 nm, 作为预测脐橙总酸含量的特征波长。

2.3.3 基于 GA 算法提取特征波长

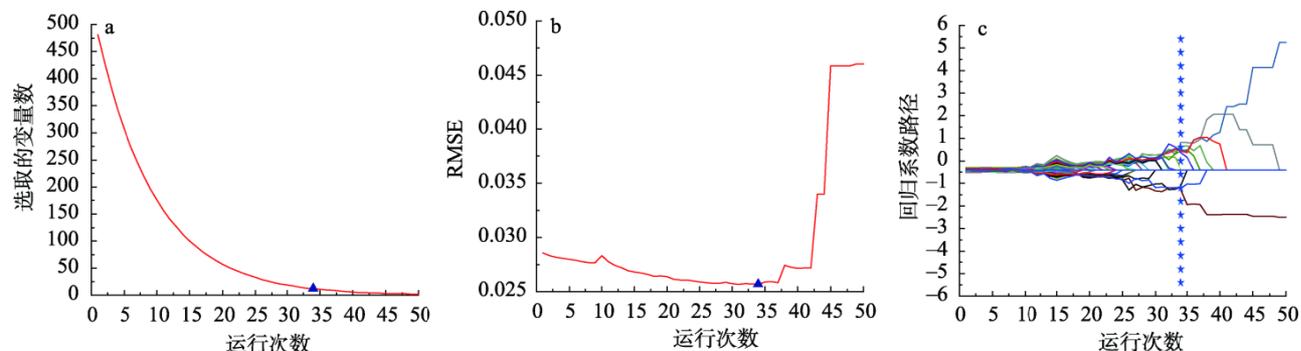
GA 算法运行结果如图 5 所示。由图 5a 和图 5b 可知, 当选取前 39 个波长作为特征波长时, 预测精度为 94.97%, RMSE 值为 0.0272。此时波长数目所对应的预测精度相对较高并且预测误差相对较小, 同时包含的波长数目也较少, 能够达到减少计算量, 简化模型的效果。波长迭代效果如图 5c 所示, 算法在运行过程中对数据迭代了 100 次, 波长与频率一一对应, 波长迭代次数越多, 所选的波长越稳定, 前 39 个波长变量迭代次数均超过 6 次。因此以提取的前 39 个波长作为预测脐橙总酸含量的特征波长。

2.3.4 基于 SPA 算法提取特征波长

基于 SPA 算法提取特征波长效果图如图 6 所示。由图 6a 可知, 当最终选定波长数量为 5 时, RMSE 值变化基本达到稳定, 只随波长数增加发生轻微变化。图 6b 表示经 SPA 选定的 5 个波长在第一个校准光谱上的分布。因此, 以 SPA 选取的 5 个光谱波长 441、481、506、657、856 nm 作为特征波长建立预测模型。

2.3.5 基于 PCA 算法降低数据维度

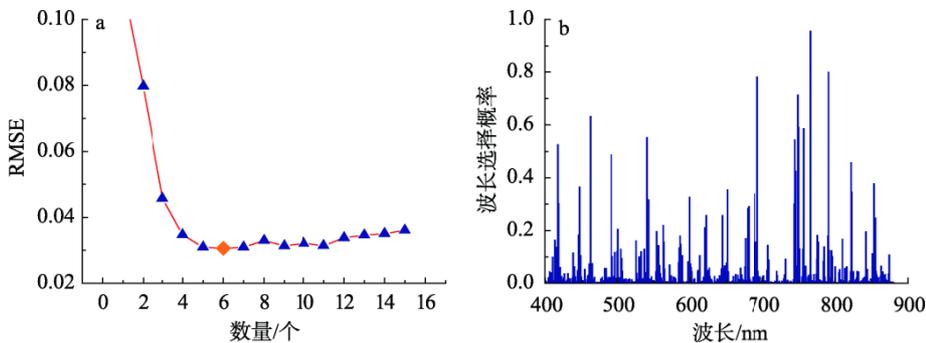
由图 7a 可知, 主成分方差贡献率从第 1 主成分到最后主成分呈现下降趋势, 并且下降幅度非常大; 同时可以明显看到, 方差贡献率值从第 1 主成分到第 2 主成分下降最迅速, 差值甚至超过了 70%, 由此可见, 第 1 主成分中数据包含了绝大部分主特征信息。观察图 7b, 主成分贡献率曲线在第 7 个主成分位置处逐渐趋向于平稳, 表明前 7 个主成分贡献率之和在 99% 以上, 涵盖的有效信息最全面, 可以替代数据降维前的原始数据。



注: a 为选取的变量数随运行次数的变化规律; b 为 RMSE 值随运行次数的变化规律; c 为回归系数路径随运行次数的变化规律。

图 3 基于 CARS 算法的特征波长提取

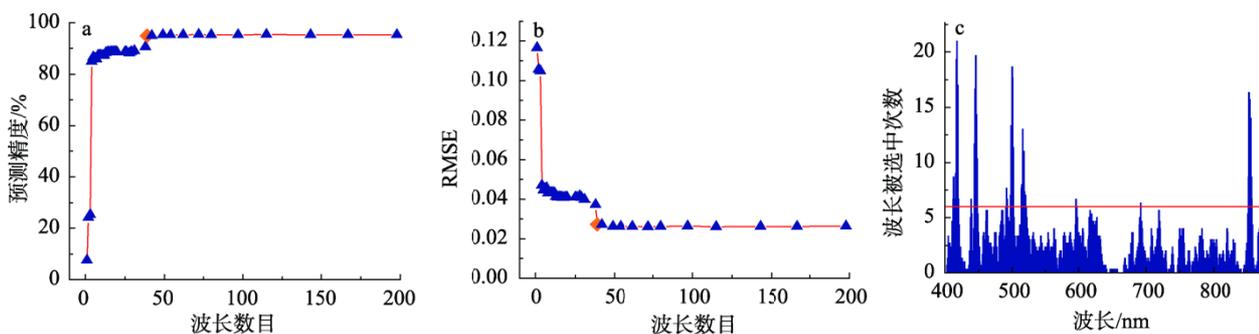
Fig.3 Extraction of characteristic wavelengths based on the CARS algorithm



注: a 为 RMSE 值随波长数目的变化规律; b 为各波长变量被选择概率。

图 4 基于 RF 算法的特征波长提取

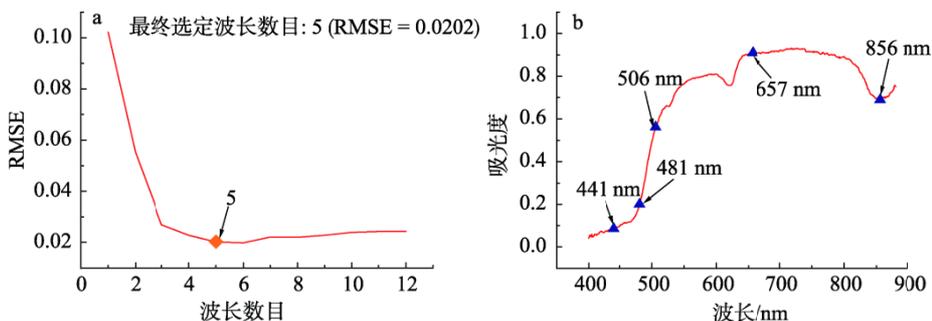
Fig.4 Extraction of characteristic wavelengths based on the RF algorithm



注: a 为 GA 算法预测精度; b 为 RMSE 值随波长数目的变化规律; c 为各波长变量被选中次数。

图 5 基于 GA 算法的特征波长提取

Fig.5 Extraction of characteristic wavelengths based on the GA algorithm



注: a 为 RMSE 值随波长数目的变化规律; b 为提取的特征波长。

图 6 基于 SPA 算法的特征波长提取

Fig.6 Extraction of characteristic wavelengths based on the SPA algorithm

2.4 不同数学模型比较及分析

基于不同算法提取的特征变量所建立不同数学模型的校正结果与预测结果如表 3 所示。

根据实验结果可知,所建立预测模型的准确度与稳定性相比原始光谱都有所提升,预测效果较为理想。由表 3 中模型的对比可知,基于 RF 和 SPA 所创建模型的预测效果相对较差,究其原因,上述两种算法所提取特征波长数目较少,导致在建立预测模型的过程中丢失了一些涵盖特

征信息的数据。对各个预测模型进行比较,其中基于 SG+GA+LS-SVM 预测模型的预测效果最优,其对应的 RMSEC 与 RMSEP 的值分别为 0.022 和 0.016, R_c^2 与 R_p^2 的值分别为 0.9689 和 0.9834, RPD 的值为 7.76。由 RMSEP 和 R_p^2 的值可以判断出上述预测模型的预测效果比较稳健且检测准确度相对较高,同时通过校正集参数与预测集参数的比较,可知校正与预测结果相近,不存在过拟合和欠拟合情况。

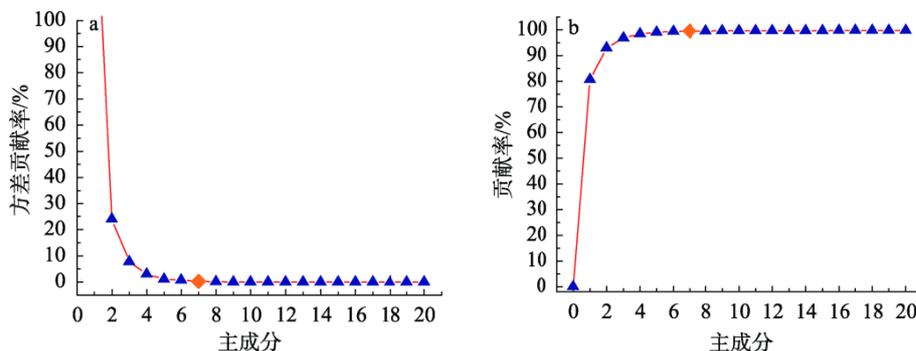
3 结 论

本研究提出一种基于可见-近红外光谱法实现赣南脐橙总酸含量的无损检测, 并建立了赣南脐橙总酸含量检测的预测模型, 得到以下结论:

(1)在本研究中, 基于 SG 平滑法的预处理效果最好, 能够很好地消除外界噪声对实验的影响; 基于 GA 算法所提取的特征波长寻优功能最佳; 在相同的变量下, 基于 LS-SVM 所创建的数学模型的预测效果最佳。

(2)SG+GA+LS-SVM 预测模型的预测效果最佳, RMSEP 的值为 0.016, R_p^2 的值为 0.9834, RPD 的值为 7.76, 优于其他模型的预测效果, 适用于定量地对赣南脐橙样品进行检测。

(3)本研究所采用的模型有效地提高了赣南脐橙中总酸含量的检测精度, 也证明了基于可见-近红外光谱无损检测技术评价赣南脐橙中总酸含量的可行性, 同时为其他柑橘属水果中总酸含量以及可溶性固形物等理化值的研究提供了理论参考。



注: a 为方差贡献率变化曲线; b 为主成分贡献率变化曲线。

图 7 方差贡献率曲线以及前 n 个主成分贡献率曲线

Fig.7 Variance contribution rate curve and cumulative contribution rate curve of the first n principal components

表 3 基于不同数学模型对赣南脐橙总酸含量的预测结果

Table 3 Prediction results of total acid content in Gannan navel orange based on different mathematical models

模型类型	变量数	R_c^2	RMSEC	RPD	R_p^2	RMSEP	RPD
Raw+PLSR	481	0.9382	0.030	4.02	0.9295	0.038	3.77
SG+CARS+PLSR	12	0.9552	0.026	4.72	0.9620	0.023	5.13
SG+RF+PLSR	6	0.9406	0.030	4.10	0.9560	0.025	4.77
SG+GA+PLSR	39	0.9611	0.024	5.07	0.9730	0.019	6.09
SG+SPA+PLSR	5	0.9363	0.030	3.96	0.9564	0.023	4.79
SG+PCA+PLSR	7	0.9472	0.028	4.35	0.9504	0.027	4.49
Raw+PCR	481	0.9441	0.029	4.23	0.9347	0.029	3.91
SG+CARS+PCR	12	0.9560	0.025	4.77	0.9579	0.023	4.87
SG+RF+PCR	6	0.9430	0.029	4.19	0.9540	0.026	4.66
SG+GA+PCR	39	0.9572	0.025	4.83	0.9640	0.026	5.27
SG+SPA+PCR	5	0.9423	0.029	4.16	0.9534	0.022	4.63
SG+PCA+PCR	7	0.9440	0.029	4.22	0.9552	0.023	4.72
Raw+LS-SVM	481	0.9604	0.025	5.02	0.9540	0.025	4.66
SG+CARS+LS-SVM	12	0.9609	0.024	5.06	0.9684	0.020	5.63
SG+RF+LS-SVM	6	0.9414	0.030	4.13	0.9543	0.025	4.68
SG+GA+LS-SVM	39	0.9689	0.022	5.67	0.9834	0.016	7.76
SG+SPA+LS-SVM	5	0.9421	0.030	4.15	0.9584	0.022	4.90
SG+PCA+LS-SVM	7	0.9536	0.026	4.64	0.9579	0.023	4.87
Raw+MLR	481	0.9480	0.028	4.38	0.9424	0.030	4.17
SG+CARS+MLR	12	0.9581	0.025	4.88	0.9626	0.022	5.17
SG+RF+MLR	6	0.9430	0.031	4.19	0.9526	0.025	4.59
SG+GA+MLR	39	0.9689	0.022	5.67	0.9684	0.020	5.63
SG+SPA+MLR	5	0.9463	0.029	4.31	0.9566	0.023	4.80
SG+PCA+MLR	7	0.9502	0.027	4.48	0.9523	0.026	4.58

参考文献

- [1] LIAO YB, HUANG LS, YANG WX, *et al.* Measurement of titratable acidity in Gannan navel orange by near infrared spectroscopy [J]. *Adv Mater Res*, 2013, 781-784: 1485-1488.
- [2] 章海亮, 孙旭东, 郝勇, 等. 近红外漫反射无损检测赣南脐橙中可溶性固形物和总酸[J]. *食品科学*, 2011, 32(6): 151-154.
- ZHANG HL, SUN XD, HAO Y, *et al.* Determination of soluble solids and total acidity in Gannan navel orange by near infrared diffuse reflection spectroscopy [J]. *Food Sci*, 2011, 32(6): 151-154.
- [3] 江水泉, 孙通. 基于可见/近红外光谱和变量选择的脐橙可溶性固形物含量在线检测[J]. *食品与机械*, 2020, 36(2): 89-93.
- JIANG SQ, SUN T. Online detection of soluble solid content in navel orange based on visible/near infrared spectroscopy and variable selection [J]. *Food Mach*, 2020, 36(2): 89-93.
- [4] 刘燕德, 翟建龙. 脐橙可溶性固形物的在线近红外光谱检测[J]. *西北农林科技大学学报(自然科学版)*, 2014, 42(3): 186-190.
- LIU YD, ZHAI JL. Online detection of soluble solid content in navel orange using near-infrared diffuse spectroscopy [J]. *J Northwest Agric Forest Univ (Nat Sci Ed)*, 2014, 42(3): 186-190.
- [5] 刘燕德, 罗吉, 陈兴苗. 可见/近红外光谱的南丰蜜桔可溶性固形物含量定量分析[J]. *红外与毫米波学报*, 2008, (2): 119-122.
- LIU YD, LUO J, CHEN XM. Analysis of soluble solid content in Nanfeng mandarin fruit with visible/near infrared spectroscopy [J]. *J Infrared Millim Waves*, 2008, (2): 119-122.
- [6] 刘燕德, 胡军, 欧阳玉平, 等. 赣南脐橙可溶性固形物近红外光谱在线无损检测[J]. *广东农业科学*, 2016, 43(9): 105-111.
- LIU YD, HU J, OUYANG YP, *et al.* Online non-destructive detection of soluble solids of Gannan navel orange with near-infrared spectroscopy [J]. *Guangdong Agric Sci*, 2016, 43(9): 105-111.
- [7] 刘燕德, 陈兴苗, 欧阳爱国. 可见/近红外光谱法无损检测赣南脐橙可溶性固形物[J]. *光学学报*, 2008, (3): 478-481.
- LIU YD, CHEN XM, OUYANG AIG. Non-destructive measurement of soluble solid content in Gannan navel oranges by visible/near-infrared spectroscopy [J]. *Acta Optic Sin*, 2008, (3): 478-481.
- [8] 李俊杰, 张绩, 汪小伟, 等. 近红外光谱大数据分析塔罗科血橙内在品质的研究[J]. *食品研究与开发*, 2020, 41(20): 193-197.
- LI JJ, ZHANG J, WANG XW, *et al.* Study on internal quality of tarocco blood orange based on large data analysis of near infrared spectroscopy [J]. *Food Res Dev*, 2020, 41(20): 193-197.
- [9] 吕强, 何绍兰, 刘斌, 等. 班菲尔脐橙可溶性固形物近红外光谱特征谱区选择[J]. *农业机械学报*, 2012, 43(S1): 211-214.
- LV Q, HE SL, LIU B, *et al.* Selection of NIR efficient regions for total soluble solids of Bamfield navel orange [J]. *Trans Chin Soc Agric Mach*, 2012, 43(S1): 211-214.
- [10] 薛龙, 黎静, 刘木华, 等. 基于遗传算法的脐橙可溶性固形物的可见/近红外光谱无损检测[J]. *激光与光电子学进展*, 2010, 47(12): 109-113.
- XUE L, LI J, LIU MH, *et al.* Nondestructive detection of soluble solids content on navel orange with vis/NIR based on genetic algorithm [J]. *Las Optoe Prog*, 2010, 47(12): 109-113.
- [11] 郭城, 吴卫成, 湛迪, 等. 基于ATR-FTIR结合CARS算法测定蜂王浆中水溶性蛋白质和总糖含量[J]. *中国食品学报*, 2021: 1-10. [2021-11-19]. <http://kns.cnki.net/kcms/detail/11.4528.TS.20211025.1353.002.html>
- GUO C, WU WC, CHEN D, *et al.* Determination of water-soluble protein and total sugar in royal jelly based on attenuated total reflection mid-infrared spectroscopy and CARS algorithm [J]. *J Chin Inst Food Sci Technol*, 2021: 1-10. [2021-11-19]. <http://kns.cnki.net/kcms/detail/11.4528.TS.20211025.1353.002.html>
- [12] 孙晶京, 杨武德, 冯美臣, 等. 基于随机蛙跳和支持向量机的冬小麦叶面积指数估算[J]. *山西农业大学学报(自然科学版)*, 2020, 40(5): 120-128.
- SUN JJ, YANG WD, FENG MC, *et al.* Estimation of winter wheat leaf area index based on random leapfrog and support vector regression approach [J]. *J Shanxi Agric Univ (Nat Sci Ed)*, 2020, 40(5): 120-128.
- [13] 都月, 孟晓辰, 祝连庆. 遗传算法和神经网络的重叠光谱解析[J]. *光谱学与光谱分析*, 2020, 40(7): 2066-2072.
- DU Y, MENG XC, ZHU LQ. Overlapping spectral analysis based on genetic algorithms and BP neural networks [J]. *Spectrosc Spect Anal*, 2020, 40(7): 2066-2072.
- [14] 赵静远, 熊智新, 宁井铭, 等. 小波变换结合连续投影算法优化茶叶中咖啡碱的近红外分析模型[J]. *分析科学学报*, 2021, (5): 611-617.
- ZHAO JY, XIONG ZX, NING JM, *et al.* Wavelet transform combined with SPA to optimize the near-infrared analysis model of caffeine in tea [J]. *J Anal Sci*, 2021, (5): 611-617.
- [15] 刘俊安, 李嘉铭, 赵楠, 等. 主成分分析结合支持向量机辅助激光诱导击穿光谱对塑料快速分类识别[J]. *光谱学与光谱分析*, 2021, 41(6): 1955-1960.
- LIU JAN, LI JM, ZHAO N, *et al.* Rapid classification and identification of plastic using laser-induced breakdown spectroscopy with principal component analysis and support vector machine [J]. *Spectrosc Spect Anal*, 2021, 41(6): 1955-1960.
- [16] 程介虹, 陈争光. 基于迭代保留信息变量和连续投影的近红外光谱波长选择方法[J]. *分析化学*, 2021, 49(8): 1402-1409.
- CHENG JH, CHEN ZG. Wavelength selection method for near infrared spectroscopy based on iteratively retains informative variables and successive projections algorithm [J]. *Chin J Anal Chem*, 2021, 49(8): 1402-1409.
- [17] 武国良, 祖光鑫, 杨志军, 等. 基于MLR和LSTM神经网络的短期负荷预测方法[J]. *黑龙江电力*, 2021, 43(4): 297-301.
- WU GL, ZU GX, YANG ZJ, *et al.* Short-term load prediction method based on MLR and LSTM neural networks [J]. *Heilongjiang Electr Pow*, 2021, 43(4): 297-301.
- [18] 万国玲, 刘贵珊, 何建国, 等. 贮藏期内灵武长枣果糖含量的高光谱预测[J]. *光谱学与光谱分析*, 2019, 39(10): 3261-3266.
- WAN GL, LIU GS, HE JG, *et al.* Prediction of fructose content of Lingwu long jujube during storage using hyperspectral imaging technique [J]. *Spectrosc Spect Anal*, 2019, 39(10): 3261-3266.
- [19] 陈素彬, 胡振. 优化的近红外光谱LS-SVM模型测定小麦蛋白质[J]. *食品工业*, 2019, 40(12): 329-333.
- CHEN SB, HU Z. Determination of wheat protein by optimized near infrared spectrum LS-SVM model [J]. *Food Ind*, 2019, 40(12): 329-333.
- [20] 尚建疆, 蔡立新. 酸度计连续滴定法测定葡萄汁中总酸和氨基氮的质量浓度[J]. *现代盐化工*, 2020, 47(3): 24-26.
- SHANG JJ, CAI LX. Determination of the mass concentration of total acid and amino nitrogen in grape juice by continuous titration with acidity meter [J]. *Mod Salt Chem Ind*, 2020, 47(3): 24-26.
- [21] 郑文瑞, 李绍稳, 韩亚鲁, 等. 土壤速效磷近红外迁移学习预测方法研究[J]. *分析测试学报*, 2020, 39(10): 1274-1281.
- ZHENG WR, LI SW, HAN YL, *et al.* Study on transfer learning prediction methods for soil available phosphorus NIR [J]. *J Instrum Anal*, 2020, 39(10): 1274-1281.

(责任编辑: 郑丽 张晓寒)

作者简介

任众财, 硕士研究生, 主要研究方向为智能检测等。

E-mail: 1379955946@qq.com

章海亮, 博士, 教授, 主要研究方向为智能检测以及光谱检测等。

E-mail: 2242526723@qq.com