

基于 CARS-MIV-SVR 的库尔勒香梨可溶性固体含量预测方法

朱晓琳^{1,2}, 李光辉^{1,2*}, 张萌^{1,2}

1. 江南大学物联网工程学院, 江苏 无锡 214122

2. 物联网技术应用教育部工程技术研究中心, 江苏 无锡 214122

摘要 为了实现库尔勒香梨依据可溶性固体含量(SSC)分级定等和按质论价, 推动采后处理向标准化、产业化方向健康发展, 利用高光谱成像技术研究出了一种快速、有效、无损检测库尔勒香梨 SSC 的方法。以表面无损伤的 157 个库尔勒香梨作为研究样本, 应用高光谱成像采集系统获取 400~1 000 nm 波长范围内高光谱图像并用 ENVI5.3 软件提取感兴趣区域(ROI), 获得高光谱数据。采用 Kennard-Stone(KS)样本集划分方法将全部样本按照 2:1 的比例划分为校正集(105)和预测集(52)。对比标准变量变换(SNV)、多元散射校正(MSC)、一阶导数(FD)和二阶导数(SD)等数据预处理方法对建模精度的影响, 最终选用 SNV 方法对光谱曲线进行平滑去噪。该研究提出竞争性自适应重加权算法与平均影响值算法的组合算法(CARS-MIV)选择特征波长。在竞争性自适应重加权算法(CARS)方法中, 建模样本由蒙特卡罗算法随机选择生成, 变量回归系数会随之发生变化, 因而回归系数的绝对值不能全面反映变量重要性, 从而影响模型检测精度。为降低这种影响, 应用平均影响值(MIV)算法对选出的自变量进行二次筛选, 筛选出相关性较大的变量用以建模分析, 并与 CARS、连续投影算法(SPA)、蒙特卡罗无信息变量消除算法(MCUVE)等经典特征波长选择算法进行比较。最后分别以全波长(FS)光谱信息和四种特征波长选择方法得出的光谱信息作为输入矢量, 应用支持向量回归(SVR)建立库尔勒香梨可溶性固体含量定量预测数学模型, 以校正集相关系数(R_c)、校正集均方根误差(RMSEC)、预测集相关系数(R_p)和预测集均方根误差(RMSEP)四个参数来评估模型的预测精度。比较分析发现, CARS-MIV-SVR 模型效果最佳, 校正集相关系数(R_c)为 0.985 94, 预测集相关系数(R_p)达到 0.946 31, 校正集和预测集均方根误差分别为 0.185 85 和 0.403 33。结果证明: CARS-MIV 特征波长选择方法能够有效增强库尔勒香梨光谱数据特征波长选择的稳定性和精确性, 提高模型的预测精度。利用高光谱技术结合 CARS-MIV-SVR 模型能够满足库尔勒香梨可溶性固体含量测定需求, 实现库尔勒香梨的分级定等和按质论价。

关键词 光谱分析; 可溶性固体含量; 变量选择; 竞争性自适应重加权算法与平均响应值算法的组合; 支持向量回归

中图分类号: TP391

文献标识码: A

DOI: 10.3964/j.issn.1000-0593(2019)11-3547-06

引言

库尔勒香梨是新疆特色水果, 果肉质地细腻, 酸甜可口, 富含维生素 C。可溶性固体含量(soluble solid content, SSC)是香梨品质优劣的重要衡量指标, 可溶性固体含量的快速无损检测有助于香梨品质检测与分级。

高光谱成像技术(hyper spectral imaging, HSI)融合了光谱信息和图像信息, 能反映样本的化学成分和微观结构。国

内外许多学者对水果的无损检测进行了研究, 研究表明利用 HSI 能够有效检测水果常见缺陷^[1]、品质指标(如硬度、糖度、水分等)^[2-4]和成熟度^[5-6]等。Li 等^[7]采用长波近红外光谱(930~2 548 nm)检测鸭梨的可溶性固体含量, 使用连续投影算法和蒙特卡罗无信息变量消除的组合算法(SPA-MCUVE)筛选出 18 个特征波长建立偏最小二乘(partial least squares, PLS)预测模型, 预测集相关系数(R_p)达 0.88, 校正集和预测集均方根误差分别为 0.49 和 0.35。何洪巨等^[8]研究不同糖度的西瓜、甜瓜的光谱差异以及西瓜、甜瓜糖度在

收稿日期: 2018-10-09, 修订日期: 2019-02-10

基金项目: 国家自然科学基金项目(61472368)资助

作者简介: 朱晓琳, 女, 1995 年生, 江南大学物联网工程学院硕士研究生 e-mail: 17851309205@163.com

* 通讯联系人 e-mail: ghli@jiangnan.edu.cn

400~1 000 nm 波长范围的响应, 研究表明对糖度响应最高的波长是 639.3 nm, 相关系数为 0.951。Dong 等^[9]在 900~1 700 nm 波长范围内对苹果的内部品质(包括糖度、硬度、含水量、pH 值)进行无损检测, 通过组合不同的波长选择方法和建模方法探究预测苹果内部品质的可行性。研究发现 SPA-LSSVM 模型对糖度、含水量、pH 值具有较好的预测性能。

纵观现有库尔勒香梨可溶性固体含量的检测研究, 部分学者采用竞争性自适应重加权算法(competitive adaptive re-weighted sampling, CARS)提取特征波长, 在 CARS 方法中, 变量回归系数会因建模样本的随机选择而发生变化, 回归系数的绝对值不能全面反映变量重要性, 因而影响模型检测精度。基于此, 本文引入平均影响值(mean impact value, MIV)反映自变量对输出神经元的影响大小, 进一步筛选出相关性较大的变量用以建模分析, 提高模型预测精度。

1 实验部分

1.1 材料

人工挑选大小适中、表面无损伤香梨共 157 个并依次编号。通过人工控制, 室内温度保持 21 °C 恒温, 将香梨样本置于恒温室内 24 h, 然后进行高光谱图像采集, 用于消除温度对最终结果的影响。

1.2 图像采集与图像校正

高光谱成像采集系统使用美国产出的 SOC710VP 高光谱成像仪, SOC710VP 光谱分辨率为 4.687 5 nm, 获得 400~1 000 nm 波长, 共计 128 个波长的高质量数据。

高光谱成像采集系统如图 1 所示。由高光谱成像仪、双 CCD 摄像机、4 个卤素灯以及一套数据采集软件和计算机组成。为减少外部光线对实验数据的影响, 图像采集操作均在暗箱中进行, 由光纤卤素灯提供光源。将用于实验的香梨水平放置于升降台中央并且被测区域对准摄像机镜头, 调整摄像头焦距使成像清晰, 设置升降台高度为 45 cm, 积分时间为 25 ms。

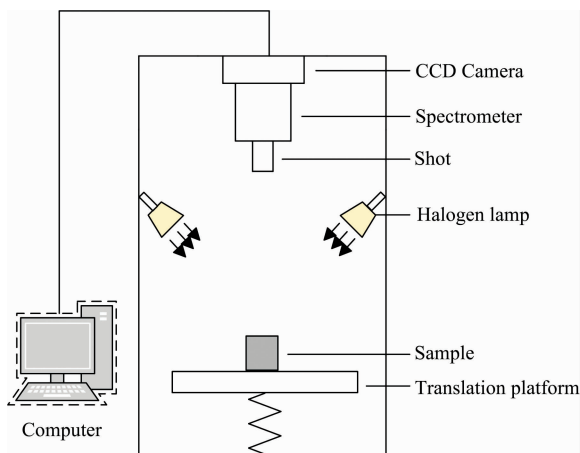


图 1 高光谱图像采集系统示意图

Fig. 1 Schematic of hyperspectral imaging system

由于高光谱摄像头中存在许多暗电流以及光照的不均匀性, 获取的高光谱图像不能直接用于数据分析, 需要先对其进行黑白校正消除暗电流的影响, 黑白校正公式如式(1)

$$R = \frac{R_0 - R_b}{R_w - R_b} \quad (1)$$

样本采集结束后, 保持各项系统参数设置不变, 扫描美国 NIST 溯源校准参考板得到白标定图像 R_w , 然后盖上摄像机镜头盖采集黑板校正图像 R_b 。 R_0 为原始噪声图像, 依据式(1)计算, 得到校正后的高光谱图像 R 。

1.3 可溶性固体含量测定与样本集划分

使用手持式数字折光仪(digital hand-held "Pocket" refractometer)测量样本的 SSC 值, 初次测量前用蒸馏水对折光仪进行零值校正。在每个香梨样本赤道部位带皮切下标记区域果肉然后人工挤压汁液滴在折光仪溶液盛放区域, 记录香梨样本的 SSC 值。

Kennard-Stone(KS)方法能够保证训练集中样本按照空间距离均匀分布, 提高预测模型稳定性与精确性。使用 KS 方法按照 2 : 1 的比例划分校正集和预测集, 样本集划分结果如表 1 所示。

校正集的 SSC 范围为 8.9~13.2, 预测集的 SSC 范围为 9~13.2, 校正集的 SSC 范围较好的覆盖了预测集, 两个数据集的 SSC 平均值和标准差相近, 这些特征有助于建立一个稳定有效的预测模型。

表 1 训练集与预测集可溶性固体含量分布统计

Table 1 Summary statistics of SSC for calibration and prediction sets/(°Brix)

数据集	样本数量 /个	最大值 /%	最小值 /%	平均值 /%	标准差 /%
校正集	105	13.20	8.90	11.54	0.85
预测集	52	13.20	9.00	11.47	0.91

1.4 数据处理分析软件

校正后的高光谱图像中包含了整个香梨样本表面的光谱信息, 使用 ENVI5.3 软件选取香梨赤道位置 10×10 像素的光谱信息作为感兴趣区域(region of interest, ROI)。后续数据预处理、特征波长选取及建立预测模型是使用 Matlab R2016a 软件编程实现。

1.5 建模方法

实验中使用支持向量回归^[10](support vector regression, SVR)建立模型对库尔勒香梨 SSC 进行预测。SVR 引入了核函数, 将低维空间映射到高维空间, 将非线性问题转化为线性可分问题求解, 用以解决小样本的高维空间建模问题, 避免了维数灾难, 选用径向基函数(radial basis function, RBF)作为核函数, 使用网格搜索(grid search, GS)算法寻找最佳的惩罚参数 c 与核函数的参数 g 。

1.6 模型评价指标

使用校正集相关系数(R_c)、校正集均方根误差(RM-SEC)、预测集相关系数(R_p)和预测集均方根误差(RMSEP)四个参数来评估模型的预测精度^[18]。相关系数表征了样本可溶性固体含量真实值与预测值的相关程度, 而均方根误差

用于衡量可溶性固体含量真实值与预测值之间的偏差, 各项参数计算公式如下

$$R_c = \sqrt{\frac{\sum_{i=1}^{N_c} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N_c} (\hat{y}_i - y_c)^2}} \quad (2)$$

$$R_p = \sqrt{\frac{\sum_{i=1}^{N_p} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N_p} (\hat{y}_i - y_p)^2}} \quad (3)$$

$$RMSEC = \sqrt{\frac{1}{N_c} \sum_{i=1}^{N_c} (\hat{y}_i - y_i)^2} \quad (4)$$

$$RMSEP = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (\hat{y}_i - y_i)^2} \quad (5)$$

式中, N_c 为校正集样本个数, N_p 为预测集样本个数, \hat{y}_i 表示第 i 个样本的 SSC 预测值, y_i 代表第 i 个样本的 SSC 测量值, y_c 和 y_p 分别为校正集样本、预测集样本的 SSC 均值。

2 结果与讨论

2.1 光谱特性与光谱预处理

黑白校正后得到的光谱数据中除包括了香梨本质信息以外, 还存在仪器噪声、环境杂散光等与香梨 SSC 值预测无关的信息, 需要对原始光谱数据进行预处理, 最大程度净化数据^[11]。

实验中使用了标准正态变量 (standard normal variate, SNV)、多元散射校正 (multiplicative scatter correction, MSC)、一阶导数 (first derivative, FD)、二阶导数 (second derivative, SD) 方法对原始光谱预处理^[12], 结果表明 SNV 能够有效去除数据噪声信息, 建模效果最佳, 因此把经过 SNV 处理后的光谱数据作为研究数据。

图 2 为原始光谱数据经过 SNV 方法处理后的光谱反射率曲线图。157 个样本的光谱曲线趋势相似, 没有明显异常样本。在 680 nm 处光谱吸收峰是库尔勒香梨表皮中叶绿素吸收所致, 820 nm 处吸收峰与库尔勒香梨含糖量有关, 920 nm 处峰值变化是由水分吸收引起^[2]。

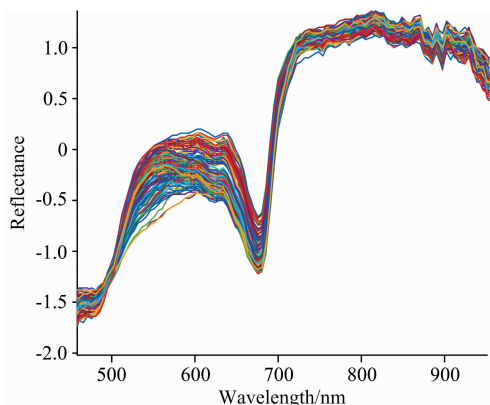


图 2 SNV 预处理后光谱反射率曲线

Fig. 2 Reflectance spectra preprocessed by SNV

2.2 CARS-MIV 方法选择特征波长

高光谱图像光谱波长比较连续, 相邻波长间相似性很

高, 存在大量数据冗余, 会影响多变量分析的时效性和准确性^[13]。因此, 选择能够充分表征全部波长信息的波长变量子集尤为重要。

CARS 方法首先采用蒙特卡罗采样选择校正集样本建立偏最小二乘 (PLS) 模型, 然后通过自适应重加权采样技术 (adaptive reweighted sampling, ARS) 和指数衰减函数 (exponentially decreasing function, EDF) 选取 PLS 模型中对应回归系数较大的波长, 利用交叉验证得到 RMSECV 值最低的子集^[14]。图 3 表示随着采样次数的增加十折交叉验证均方根误差的变化曲线。可以看出, RMSECV 曲线整体呈现出先下降再上升的趋势, 采样次数为 14 时, RMSECV 值达到最小值 0.489 4, 对应的变量数为 42。

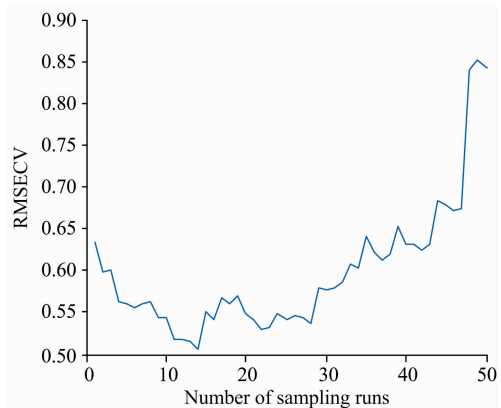


图 3 RMSECV 变化趋势图 (CARS)

Fig. 3 The 10-fold RMSECV values with the increasing of sampling runs by CARS

在 CARS 方法中, 建模样本由蒙特卡罗采样随机选择生成, 变量回归系数会因样本的随机选择而发生变化, 回归系数的绝对值无法全面反映变量的重要性, 因而影响模型精度。为降低这种影响, 引入 MIV 算法对自变量进行二次筛选, 在保证模型预测精度前提下进一步精简数据规模, 提高模型健壮性。

MIV 算法^[15]是神经网络中输入神经元对输出神经元的影响重要性的最好的评价指标。图 4 为 MIV 算法流程图, N 表示自变量数目, IV_i 表示第 i 个自变量的影响值, MIV_i 表示第 i 个自变量的平均影响值。利用已训练好的预测模型 `svr_train` 计算每个自变量 x_i ($1 \leq i \leq N$) 的平均影响值 MIV_i , 该值正负符号代表相关的方向, 绝对值代表对模型影响重要性。对所有自变量的 MIV 值进行降序排序, 得到自变量对网络输出影响相对重要性的位次表, 从而判断输入特征对模型结果的影响程度。

图 5(a) 为 CARS 方法筛选出的 42 个特征波长对应的 MIV 值, 可以看出有些特征波长 MIV 值较小, 例如第 11—15 个特征波长 MIV 值小于 0.1, 对模型输出结果影响不大。图 5(b) 将 42 个特征波长按照其 MIV 值降序排列, 图中曲线代表了前 i ($1 \leq i \leq 42$) 个特征波长的 MIV 累积贡献率。视图可知前 33 个变量 MIV 累积贡献率大于 95%, 因此二次筛选后特征波长数量减少为 33。

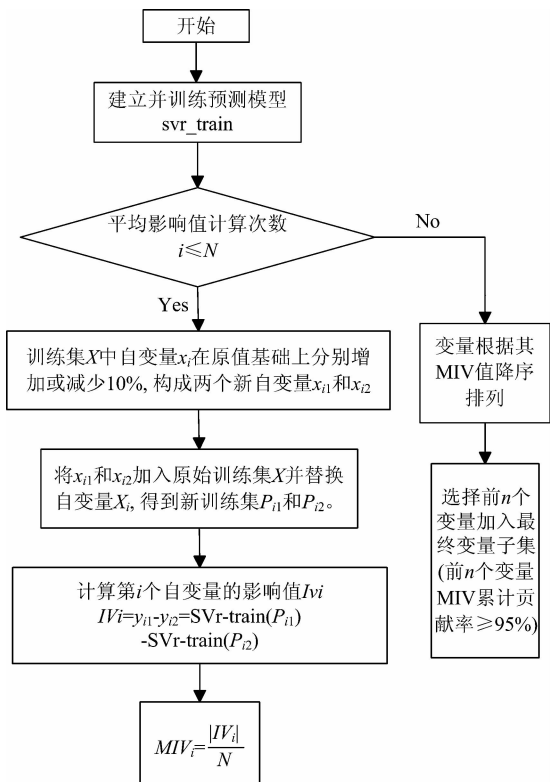


图 4 MIV 算法筛选变量流程图

Fig. 4 Flow chart of variables selected using MIV

2.3 其他特征波长选择方法

2.3.1 连续投影算法

连续投影算法 (successive projection algorithm, SPA) 是一种前向变量选择算法^[16], 能够消除波长数据间共线性影响, 提取出具有最低冗余度和最小共线性的波长子集。图 6 反映了选择出的特征波长分布情况。红色方框代表其所对应的波长被选为特征波长, 共计有 20 个波长被选择出来, 大多数分布在第 105 到第 128 个波长之间。

2.3.2 蒙特卡罗无信息变量消除

蒙特卡罗无信息变量消除 (Monte-Carlo uninformative variable elimination, MCUVE) 算法是一种基于变量稳定性

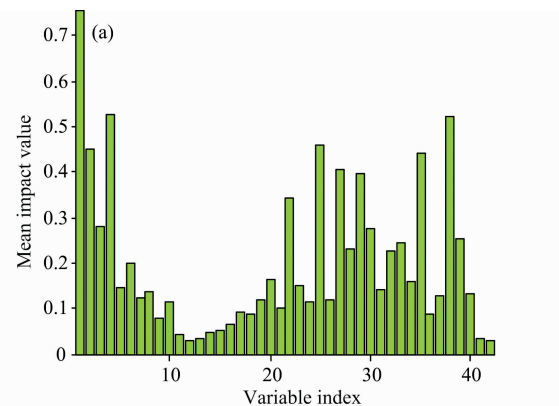


图 5(a) 特征波长 MIV 值 (CARS-MIV)

Fig. 5(a) Mean impact value of selected variables

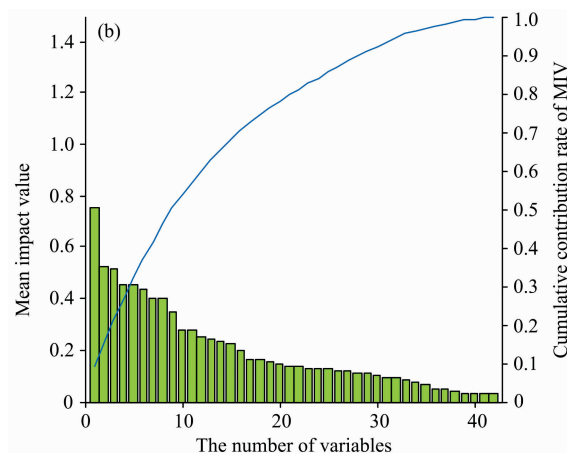


图 5(b) MIV 累积贡献率 (CARS-MIV)

Fig. 5(b) Cumulative contribution rate of MIV

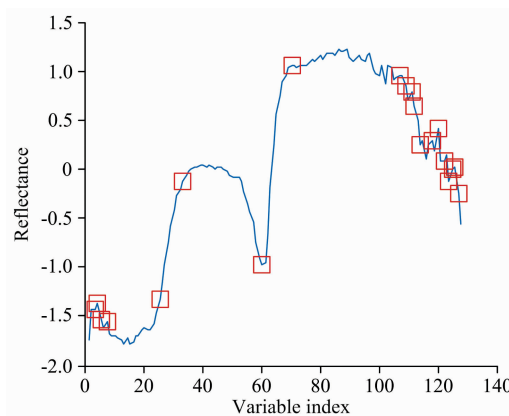


图 6 特征波长分布 (SPA)

Fig. 6 Distribution of variables selected by SPA

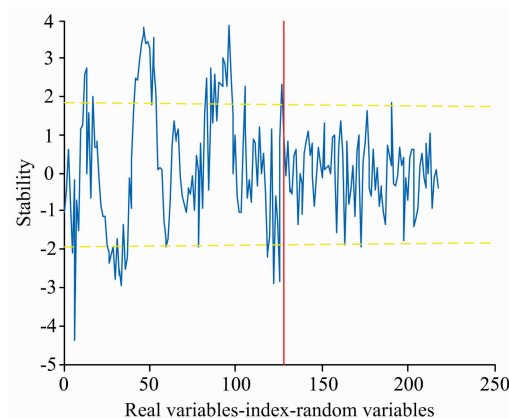


图 7 变量稳定性 (MCUVE)

Fig. 7 Stability distribution of variables by MCUVE

的特征波长选择方法^[17]。使用 MCUVE 进行特征波长选择如图 7 所示, 平行于横坐标的两条黄色阈值虚线刻画了可被选为特征波长的变量稳定性范围, 两条虚线中间区域表示原始光谱中与香梨可溶性固体含量无关的信息, 分布在两虚线

外的稳定性值所对应的波长被选入特征波长集合。共有 39 个特征变量被选出,用以建模分析。

2.4 基于特征波长的模型分析

分别将全光谱(full spectrum, FS)数据、选出的特征波长数据作为输入,香梨可溶性固体含量测量值作为输出建立 SVR 预测模型并对结果进行比较分析。使用不同特征波长选择方法建模结果如表 2 所示。

表 2 不同特征波长选择方法建模结果比较

Table 2 The SVR prediction results of SSC based on different wavelength selection methods

模型	特征波长数量	校正集		预测集	
		R_c	RMSEC	R_p	RMSEP
FS-SVR	128	0.973 06	0.208 51	0.904 29	0.402 67
SPA-SVR	20	0.899 01	0.384 12	0.825 44	0.532 44
MCUVE-SVR	39	0.934 60	0.185 85	0.901 35	0.518 89
CARS-SVR	42	0.982 41	0.171 60	0.908 41	0.416 45
CARS-MIV-SVR	33	0.985 94	0.185 85	0.946 31	0.403 33

由表 2 可知,使用 CARS 方法建立 SVR 预测模型 R_c 值为 0.982 41, R_p 值为 0.908 41,相比于 FS-SVR, SPA-SVR, MCUVE-SVR 模型, R_c 和 R_p 值较优,同时 RMSEC 和 RMSEP 值都比较小。虽然经 CARS 方法筛选后特征波长数量较多(是 SPA 选择的特征波长数量的 2.1 倍),但鉴于样本数据量较小,计算时间差别不大,因此在 CRAS, SPA, MCUVE 这 3 种基本的特征波长选择方法中,基于 CARS 算法的 SVR 预测模型具有相对较好的预测性能。

对比 CARS-SVR 与 CARS-MIV-SVR 建模结果,相比于

CARS-SVR, CARS-MIV-SVR 方法特征波长数减少 21.43%,两者 R_c 值相差不大, R_p 值由 0.908 41 增长为 0.946 31,同时 RMSEP 值有所减少。证明 CARS-MIV 特征波长提取方法能够有效筛选出表征光谱大部分信息的波长,去除原始数据中的冗余信息,减少计算时间,提升模型预测精度。

3 结 论

使用可见-近红外高光谱技术结合不同的光谱预处理方法、特征波长选择方法对库尔勒香梨可溶性固体含量进行定量分析,主要结论如下:

(1)利用高光谱图像采集系统获取高光谱图像,通过 ENVI5.3 软件手动选取感兴趣区域,得到高光谱数据。针对每一个香梨样本,使用手持式数字折光仪测量其可溶性固体含量。比较无数据预处理、SNV 预处理、MSC 预处理等几种数据预处理方式建模结果,分析发现使用 SNV 方法对数据进行预处理后建模效果最佳。

(2)与全波段建模结果相比,提取特征波长后建立预测模型,能够减少计算量,模型预测精度也有一定程度提升。其中 CARS-MIV-SVR 模型达到最优预测效果,校正集相关系数(R_c)为 0.985 94,预测集相关系数(R_p)达到 0.946 31, RMSEC 和 RMSEP 分别为 0.185 85 和 0.403 3。

(3)通过特征波长筛选并且基于特征波长建立预测模型,结果表明 CARS-MIV 方法能够有效增强库尔勒香梨光谱数据特征波长选择的稳定性和精确性,提高模型的预测精度,实现香梨 SSC 的有效预测。

References

- [1] Tan W, Sun L, Yang F, et al. Optik, 2018, 154: 581.
- [2] Gabriel A Leiva-Valenzuela, RenfuLu José Miguel Aguilera. Innovative Food Science and Emerging Technologies, 2014, 24: 2.
- [3] JIE Deng-fei, LI Ze-hai, ZHAO Jun-wei, et al(介邓飞,李泽海,赵竣威,等). Chinese Journal of Luminescence(发光学报), 2017, 38(5): 685.
- [4] GU Wen-jun, TIAN You-wen, ZHANG Fang, et al(古文君,田有文,张芳,等). Journal of Shenyang Aricultural University(沈阳农业大学学报), 2017, 48(5): 584.
- [5] Rajkumar P, Wang N, Elmasry G, et al. Journal of Food Engineering, 2012, 108: 194.
- [6] Wang A, Fu X, Xie L. Food Analytical Methods, 2015, 8(6): 1403.
- [7] Li J, Tian X, Huang W, et al. Food Analytical Methods, 2016, 9(11): 3087.
- [8] HE Hong-ju, HU Li-ping, LI Wu, et al(何洪巨,胡丽萍,李武,等). Food and Nutrition in China(中国食物与营养), 2016, 22(10): 53.
- [9] Dong J, Guo W. Food Analytical Methods, 2015, 8(10): 2635.
- [10] Hui L, Liang L, Zhang L, et al. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(11): 139.
- [11] FENG Di, JI Jian-wei, ZHANG Li, et al(冯迪,纪建伟,张莉,等). Chinese Journal of Luminescence(发光学报), 2017, 38(6): 799.
- [12] Hu M H. Postharvest Biology & Technology, 2016, 115: 122.
- [13] Liu D, Sun D W, Zeng X A. Food and Bioprocess Technology, 2014, 7(2): 307.
- [14] WANG Hai-long, YANG Guo-guo, ZHANG Yu, et al(王海龙,杨国国,张瑜,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(7): 2115.
- [15] CHEN Xin, LIU Fei(陈鑫,刘飞). Computers and Applied Chemistry(计算机与应用化学), 2012, 29(7): 812.
- [16] Fan S, Huang W, Guo Z, et al. Food Analytical Methods, 2015, 8(8): 1936.

- [17] Hao Y, Sun X, Zhang H, et al. *Sensor Letters*, 2011, 9(3): 1229.
- [18] SUN Jun, CONG Sun-li, MAO Han-ping, et al(孙俊, 丛孙丽, 毛罕平, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2017, 33(5): 178.

Prediction of Soluble Solid Content of Korla Pears Based on CARS-MIV

ZHU Xiao-lin^{1,2}, LI Guang-hui^{1,2*}, ZHANG Meng^{1,2}

1. School of IOT Engineering, Jiangnan University, Wuxi 214122, China

2. Engineering Research Center of IOT Technology Applications, Ministry of Education, Wuxi 214122 China

Abstract In order to classify and set different prices on the basis of soluble solid content (SSC) of korla pears and promote the development of post-harvest processing healthily in standardization and industrialization, a fast, precise and nondestructive method to detect soluble solid content of korla pears was determined by applying hyperspectral reflectance imaging technology. 157 korla pears freshly and with no surface damage were collected as samples. Hyperspectral images with a spectral range of 400~1000 nm of pears were acquired by hyperspectral imaging system. Then the region of interest (ROI) function of ENVI 5.3 software was used to conduct spectral data extraction from each hyperspectral image of pear. Totally, 157 pear samples were divided into calibration set (105) and prediction set (52) based on the Kennard-Stone(KS)sample set partitioning method. The research compared the influence of accuracy of modeling in terms of the spectrum pretreatment methods of original spectrum, standard normal variate (SNV), multiplicative scatter correction (MSC), first derivative (FD) and second derivative (SD). The SNV was applied for smoothing and denoising of the original hyperspectral data. A variable selection method combining competitive adaptive reweighted sampling and mean impact value (CARS-MIV) was utilized to extract the characteristic variables from full spectrum (FS). The modeled samples of competitive adaptive reweighted sampling (CARS) are generated by random selection of Monte Carlo sampling, and the regression coefficients of variables will change accordingly. The absolute value of regression coefficients cannot fully reflect the importance of variables, and affect the accuracy of the model. To lower the impact, the mean impact value (MIV) algorithm is applied to select the independent variables for secondary screening, and the variables with bigger correlation are selected for modeling and analysis. In this paper, the variables selected by CARS, successive projection algorithm (SPA) and Monte-Carlo uninformative variable elimination (MCUVE) were used for comparison. Finally, the spectral information selected from full wavelength and the spectral information selected from four characteristic wavelength selection method were taken as input vector to build support vector regression(SVR)model to predict soluble solid content of korla pears. The performances of the models were evaluated by the root of mean square of calibration (RMSEC), the root of mean square of prediction (RMSEP), the correlation coefficient of calibration (R_c) and the correlation coefficient of prediction (R_p). By means of comparison, the CARS-MIV-SVR models achieved the optimal performance with the R_c reaching 0.985 94 and R_p up to 0.946 31. The RMSEC and RMSEP are 0.185 85 and 0.403 33 respectively. These experimental results demonstrated that CSRS-MIV method can efficiently improve the stability and accuracy of wavelength selection, and optimize the precision of prediction model. The hyperspectral technique combined with CARS-MIV-SVR model can meet the needs of determination of soluble solid content and be used to classify and set different prices on the basis of SSC of korla pears.

Keywords Spectral analysis; Soluble solid content; Variable selection; Competitive adaptive reweighted sampling-mean impact value; Support vector regression

(Received Oct. 9, 2018; accepted Feb. 10, 2019)

* Corresponding author