

不同品种苹果糖度近红外光谱在线检测通用模型研究

刘燕德, 徐海, 孙旭东, 姜小刚, 饶宇, 张雨

华东交通大学机电与车辆工程学院, 水果智能光电检测技术与装备国家地方联合工程研究中心, 江西 南昌 330013

摘要 由于果实内部细胞结构、组成成分和光学传输特性的不同, 品种差异会对近红外建模分析果实内部品质时产生较大的影响, 以致原有模型无法高精度地预测果实品质参数。探讨开发不同品种近红外通用模型用于在线检测苹果内部品质的可行性。采用水果动态在线分选设备, 设置运行参数为: 积分时间 100 ms, 运动速度 5 s^{-1} , 采集包括冰糖心, 红富士及水晶富士三个品种苹果的近红外漫透射光谱。分析了三个品种近红外漫透射光谱的响应特征, 其光谱曲线走势基本一致, 在 650, 709 和 810 nm 附近存在突出吸收峰, 而在 670, 750 与 830 nm 附近存在波谷, 其差异主要表现为光谱吸收强度的差异。采用多元散射校正, Savitzky-Golay 卷积平滑及归一化处理, 减少了不同品种引起的光谱信息差异。混合三个品种各校正集样本, 采用偏最小二乘回归算法建立了不同品种糖度的通用模型, 并利用无信息变量消除法(UVE)对建模变量进行筛选, 最终得到的有效变量个数为 155。所建立的 UVE-PLS 模型对验证集的决定系数, 均方根误差以及残留预测偏差分别为 0.80, 0.61% 与 2.21。在 UVE 筛选变量的基础上, 采用连续投影算法再对建模变量进行选择, 最终选出的变量个数为 22。采用多元线性回归(MLR)方法建立了简化后的通用模型, 对验证集的决定系数与均方根误差分别为 0.78 与 0.64%。测试集用于评估最佳的不同品种糖度通用模型的实际性能, 模型对每个品种测试集的潜变量数, 决定系数与均方根误差分别为 6~10, 0.77~0.79 与 0.45~0.75%。结果表明水果动态在线分选设备对不同品种苹果内部品质检测的潜力。通过建立通用模型, 扩大了单一品种模型的预测范围, 提高了模型在不同品种间的预测稳健性。并且采用合适的变量选择方法能够减少模型变量个数, 降低模型复杂程度, 并最终提高模型速率。开发不同品种水果内部品质通用模型在波长有限的近红外光谱设备中具有良好的潜在应用。

关键词 在线检测; 近红外光谱; 通用模型; 偏最小二乘法; 无信息变量消除; 苹果; 糖度

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)03-0922-07

引言

水果内部参数, 如糖度(soluble solid content, SSC), 是水果的主要品质属性^[1]。传统检测方法为破坏性方法, 且分析数量有限, 而近红外(near infrared, NIR)光谱结合化学计量学是一种快速、准确、易于实现的无损检测技术。目前已经研究了各种水果, 包括苹果^[2-4]、石榴^[5]、樱桃^[6]、圣女果^[7]等水果种类。

在新样本测定前, 必须建立 NIR 模型以建立光谱和属性值之间的函数关系。因为果实物化特性的不同, 导致其内部细胞结构、组成成分和光学传输特性的差异, 故一般针对不同品种建立不同的近红外模型。但这种方法增加了模型建立及维护的成本。此外, 对于波长有限的近红外仪器, 由于校

正模型的限制, 品种的适用范围也受到限制。因此, 建立近红外通用模型将有助于推广近红外技术。Liu 等^[8]建立了富士苹果, 大久保桃子和丰水梨糖度通用模型。Wang 等^[9]建立了糖度与硬度的多品种洋梨模型。前期研究中建立了新梨 7 号, 砀山酥梨, 玉露香梨和红富士苹果的糖度在线检测通用模型。前两项研究都是基于便携式检测设备建立的数学模型, 未能探讨建立动态在线检测设备通用模型的可行性。而后一项研究重点研究并比较了单一品种检测模型与在线检测通用模型的优劣势, 结果表明混合模型在减小品种差异对预测模型稳健性的影响上表现出良好的性能。然而, 由于大量的建模数据, 增加了模型的潜变量个数, 从而降低了模型的效率。因此, 为了在保证预测精度的前提下, 提高通用模型的效率, 以三种苹果(冰糖心、红富士及水晶富士)为研究对象, 采用合适的预处理方法以减少不同品种的光谱差异, 并

收稿日期: 2019-03-07, 修订日期: 2019-06-26

基金项目: 国家自然科学基金项目(31760344), 水果光电检测技术能力提升项目(S2016-90)资助

作者简介: 刘燕德, 女, 1967年生, 华东交通大学机电与车辆工程学院教授 e-mail: jxliuyd@163.com

采用无信息变量消除法以及连续投影算法筛选特征波长以减少建模变量,开发不同品种苹果糖度的近红外在线检测通用模型。研究结果将对水果商品化处理环节具有实际应用价值。

1 实验部分

1.1 样品

样品选自陕西省某商业果园,经过手工分拣后立即放于22℃的实验室环境下,且整个实验阶段,样品都保持相同的贮藏状态。其中,冰糖心75个,红富士50个,水晶富士40个,共165个样品果。在口感上,红富士与冰糖心均属于脆苹果,而水晶富士属于粉苹果。口感的粉与脆实质上是苹果在成熟期果肉细胞的不溶性果胶及纤维素的含量和性质差异导致细胞间粘连性下降的程度和速率不同造成的^[10]。每个批次按样品比例使用15~20个果实,并定期评估。每个样品均在果实赤道附近的4个位置进行测量,各测量部位相距90°并垂直于茎轴。在同一位置进行光谱采集与糖度测定,且每个观测点均作为一个独立样本。

1.2 动态在线检测设备与光谱采集

水果动态在线检测设备采用漫透射检测方式(图1),将一组呈圆环状配置的光源对称排列在待测样品两侧,其发出的光均匀照射样品表面,透过样品内部的光由果杯下方光纤探头收集后传回光谱仪,这种方式得到的样品果肉信息更多,避免了果实表面反射光及外部光干扰,同时能有效消除由果实不同部位造成的糖、酸等成分的测定误差。设备采用QE65Pro型光谱仪(Ocean optics INC. USA),其波长范围为350~1150 nm,光源为10盏12 V,100 W的卤钨灯,采集时参数设定为:积分时间100 ms,运动速度5 s⁻¹。

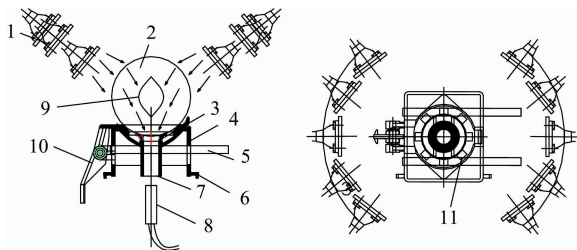


图1 漫透射检测机构

1: 光源; 2: 样品; 3: 密封圈; 4: 果杯; 5: 拨杆; 6: 支持端; 7: 限位装置; 8: 光纤探头; 9: 果核; 10: 弹跳爪; 11: 橡皮垫圈

Fig. 1 Mechanism of diffuse transmission detection

1: Light source; 2: Samples; 3: Gasket ring; 4: Cups; 5: Driving lever; 6: Stand; 7: Stop block; 8: Probe; 9: Kernel; 10: Bouncing claw; 11: Rubber hood

设备预热30 min后,以聚四氟乙烯球作为参比,校正参比后采集光谱。采用硬件触发方式采集光谱,其结构简图如图2所示,齿轮每4个齿对应安装一个果杯,果杯和分度盘齿一一对应。分度盘下方2 mm处安装有光电开关,每转一齿,即传动一个果杯的行程,触发传感器,使后续电路发出信号,从而触发光谱仪采集并保存一条光谱。

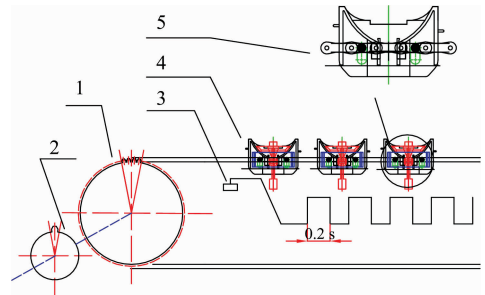


图2 光谱采集触发装置

1: 链轮; 2: 分度盘; 3: 光电开关; 4: 果杯; 5: 链条

Fig. 2 Trigger of spectral acquisition

1: Gear wheel; 2: Dividing plate;

3: Photoelectric switch; 4: Cups; 5: Chain

1.3 SSC含量测定

从每个样品的测量位置取出一个圆形楔形物(最大直径50 mm),通过粗棉布压制,并使用温度补偿糖度计(型号PAL-1; Atago Co., Tokyo, Japan)测量果汁的SSC含量。重复测量三次,取其平均值作为对应测量位置的SSC含量真值。

1.4 偏最小二乘回归

偏最小二乘(PLS)回归广泛应用于近红外光谱分析,将光谱矩阵 X 与浓度矩阵 Y 相关联。PLS回归模型可以写为

$$Y = bX + e \quad (1)$$

式(1)中, b 表示回归系数的向量, e 表示模型残差。

在PLS算法中,留一法交叉验证通常用于分别避免由于使用太小或太大的潜变量(latent variables, LVs)而导致的欠拟合或过拟合。LVs的最佳数量由交叉验证值的最低均方根误差确定。

1.5 数据统计分析

通过校正集的决定系数(R_c^2)与均方根误差(RMSEC)来评价SSC与光谱数据之间的校正方程。通过验证集的决定系数(R_p^2)与均方根误差(RMSEP)来评价校正模型的预测精度。残留预测偏差(residual prediction deviation, RPD),用来判断模型的质量。各统计指标公式如式(2)~式(4)

$$R_c^2, R_p^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_m)^2} \quad (2)$$

$$\text{RMSEC}, \text{RMSEP} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

$$\text{RPD} = \frac{SD}{\text{RMSEP}} = \frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - y_m)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}} \quad (4)$$

式(2)~式(4)中, n 表示校正集或验证集中的样本数, y_i 和 \hat{y}_i 分别表示校正集和验证集中第*i*个样本的测量值和预测值, y_m 表示样本集所有测量值的平均值。

RPD值的增加表明模型质量的提高,Chang等^[11]定义了三个质量类别以考虑模型的可靠性:优秀模型,RPD>2;

可靠模型, RPD 为 1.4~2; 不可靠模型, RPD<1.4。

2 结果与讨论

2.1 样本集的定义

在整个实验阶段, 将最后一批的样本作为测试集, 而其余批次样本进行建模分析。校正建模之前, 对样本进行肖维勒准则测试。其中, 冰糖心, 红富士与水晶富士各剔除异常样本 4 个, 共计 12 个。将冰糖心、红富士与水晶富士依次定义为品种 1, 2 与 3, 每个品种的剩余样本通过 Kennard-

Stone(K-S)法按约 3:1 的比例分成 2 组, 校正集与验证集。

全部剩余样本的校正与验证集分别由各个品种的校正和验证子集组成。校正集仅用于构建校正模型, 验证集仅用于决定模型的参数以评估校正模型, 而测试集用于模型的外部验证。如表 1 所示, 校正集与验证集的 SSC 范围分别为 8.5%~17.5% 与 8.9%~16.7%, 标准偏差分别为 1.48% 与 1.35%, 样本集覆盖了足够大的范围, 且校正集范围大于每个品种的验证集, 这些特征有利于建立稳健的近红外模型。

表 1 样本集糖度含量统计

Table 1 Statistical values of the SSC(%) of data sets

品种	校正集				验证集				测试集			
	数量	范围	平均值	标准差	数量	范围	平均值	标准差	数量	范围	平均值	标准差
1	200	9.2~17.5	13.82	1.72	64	10.5~16.7	13.87	1.43	32	9.9~16.3	13.82	1.55
2	135	8.5~16.4	13	1.27	41	8.9~15.1	12.93	1.23	20	11.4~14.8	13.12	0.94
3	108	10.9~16.2	13.22	1.02	32	11~15.6	13.21	1.05	16	11.6~14.6	13.11	0.82
总	443	8.5~17.5	13.42	1.48	137	8.9~16.7	13.43	1.35	68	9.9~16.3	13.45	1.28

2.2 光谱特征及预处理

每个品种的平均原始光谱如图 3(a)所示, 光谱下部(350~594 nm)与上部(900~1 150 nm)的光谱能量值较低, 有效信息较少, 主要涵盖了无价值的信号, 故将有效波长范围定义为 594~900 nm(400 个数据点)。总体上看, 三种光谱主要存在吸收强度的差异, 光谱走势基本一致。红富士光谱较另外两种光谱吸收强度较高, 而水晶富士强度最低, 其可能与果型及果肉软化有关。光谱曲线分别在 650, 709 和 810 nm 附近存在突出吸收峰, 而在 670, 750 与 830 nm 附近存在波谷。其中, 650 nm 的吸收主要受表皮颜色的影响, 670 nm 附近可能为叶绿素^[12]的吸收, 750 nm 附近的波谷与 O—H 三倍频伸缩振动有关, 830 nm 附近的较弱波谷与 N—H 三倍频伸缩振动^[13]有关。经过比较分析, 使用多元散射校正(MSC)对光谱进行预处理, 其主要作用为消除颗粒分布不均匀及颗粒大小产生的散射影响。并结合 Savitzky-Golay(S-G)卷积平滑(平滑点数为 3)来有效减少光谱的噪声, 提高光谱信噪比。

2.3 光谱差异性分析和标准化处理

为了进一步消除不同品种的影响, 首先进行光谱差异性分析, 对光谱进行主成分分析(PCA), 以光谱前 3 个主成分的得分计算样本的空间距离(前 3 个主成分已表征了 99% 的光谱信息), 三个品种的三维得分分布图如图 4(a)所示。

由图所示, 三个品种的主成分空间分布存在差异, 这些差异可能反映了品种之间的差异, 例如果皮, 果肉的色泽与厚度, 果胶含量等。通过比较分析, 在光谱预处理的基础上结合归一化处理以减小差异。将未处理的光谱与经过归一化处理及光谱预处理的光谱的 PC 得分图[图 4(b)]进行比较, 后者的分布明显更加集群, 即光谱差异性减小, 且平均光谱[图 3(b)]差异明显缩小, 因而可以进一步研究三个品种模型的通用性。

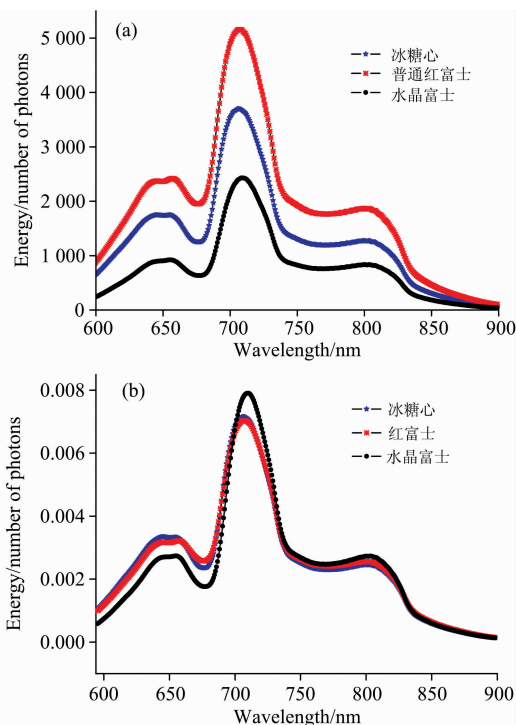


图 3 样品平均光谱

(a): 原始光谱; (b): 处理后光谱

Fig. 3 Average spectra of the samples

(a): Original spectra; (b): Processed spectra

2.4 PLS 模型建立及优化

采用 PLS 回归算法建立三个品种的通用模型, 即混合三个品种的校正模型, 建立不同品种混合模型, 用于预测所有品种的 SSC。采用基于 PLS 回归系数的无信息变量消除法^[14](uninformative variable elimination, UVE)筛选特征变

量,其基本思想是把变量对应的回归系数平均值与其标准差的比值作为变量选择的衡量依据,其表达式如式(5)

$$CV = \frac{\text{mean}(b_i)}{\text{Std}(b_i)} \quad (5)$$

式(5)中, CV 为回归系数的平均值与标准差之比, i 为光谱矩阵中第 i 列向量。

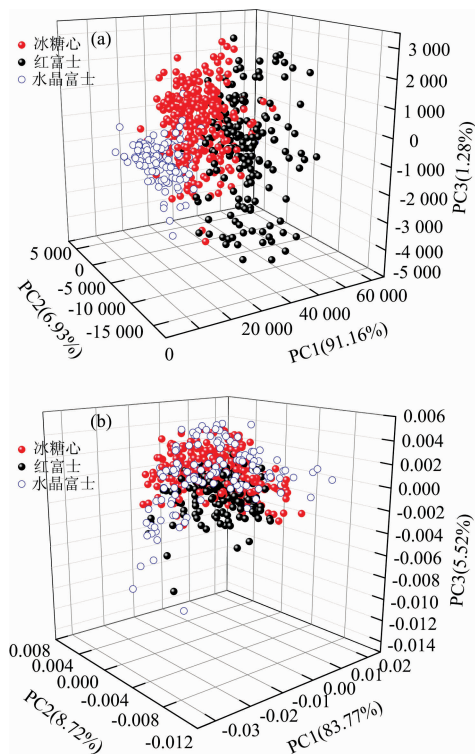


图 4 三个品种的前 3 个主成分得分分布

(a): 原始光谱的主成分得分;

(b): 经过归一化及多元散射校正处理的光谱的主成分得分

Fig. 4 The first three PCs score plots of spectra of all cultivars

(a): PC scores of original spectral data;

(b): PC scores of MSC spectral data with normalized processing

最终的判断方法是将一定数量的随机变量矩阵加入光谱矩阵中,然后通过交叉验证建立 PLS 模型,得到回归系数矩阵,将随机变量矩阵的最大 CV 值(CV_{max})作为阈值,当变量对应的 CV 值低于 CV_{max} 时,该变量被剔除。

采用 UVE 对经过预处理的全谱变量(398 个数据点)进

行有效变量筛选,控制随机变量的大小为原变量的 10^{-10} ,使得扩展矩阵的特征值基本保持不变以最大程度地降低随机矩阵对原变量回归系数的影响,随机变量个数设为 200,最终的处理结果如图 5 所示。

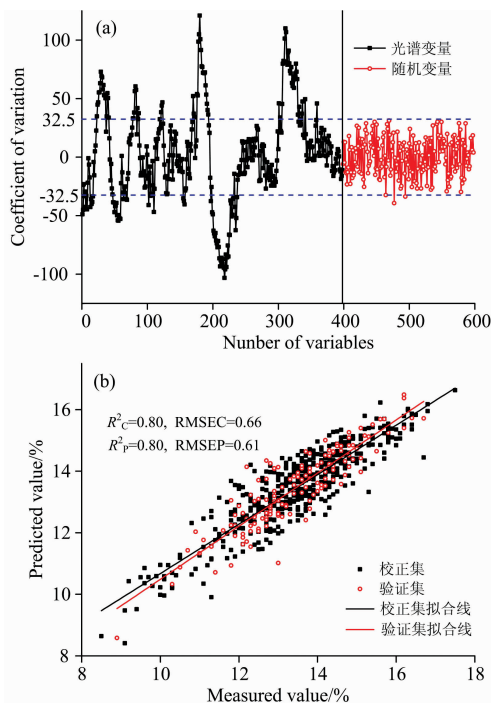


图 5 UVE 处理结果

(a): UVE 筛选变量的结果;

(b): 通过 UVE 选择波长后的 PLS 模型预测结果

Fig. 5 Results of UVE processing

(a): Result of variables selected by UVE;

(b): PLS model in the wavelengths selected by UVE

由图 5(a)所示,黑色竖线为光谱变量与随机变量分界线,其左侧黑色曲线表示光谱变量的 CV 分布情况,右侧红色曲线表示随机变量的 CV 分布情况。2 条蓝色水平虚线表示所得的阈值上下限,即随机矩阵的 CV_{max} 为 32.5,且当随机变量的 CV 值达到其最大值的 99% 时,阈值被触发。虚线之间的变量为被剔除的变量,而虚线外的变量即为选中的光谱变量。最终经过 UVE 筛选出的剩余光谱变量个数为 155,其验证模型对比结果如表 2 所示。

表 2 苹果糖度的 PLS 模型验证结果

Table 2 Results of PLS modeling for the SSC of apples

项目	光谱区间	变量数	潜变量数	R^2_c	RMSEC/%	R^2_v	RMSEP/%	RPD
全局模型	原始全谱	400	12	0.80	0.64	0.72	0.73	1.85
	预处理后全谱	398	12	0.82	0.62	0.80	0.61	2.21
	UVE 法后光谱区域	155	10	0.80	0.66	0.80	0.61	2.21

由表 2 可知,经过预处理, RMSEP 由 0.73% 降低至 0.61%, 而 RPD 由 1.85 增至 2.21, 可见前述所用的预处理方法是可行有效的。并且经过 UVE 筛选后,大大减少了建

模变量,其潜变量个数也有所降低,最终降低了模型的复杂程度,提高了模型的运算速度。而其 RPD 并没有降低,说明 UVE 筛选变量后,光谱的有效信息并没有丢失,模型表现出

优秀的质量。将模型定义为 UVE-PLS, 其预测散点图如图 5 (b) 所示。

2.5 模型的简化

为了进一步去除光谱的冗余信息, 在 UVE 筛选变量的基础上, 再采用连续投影算法 (successive projections algorithm, SPA) 选择变量, 其基本思想是利用向量的投影分析^[15], 找出包含最低限度的冗余信息的变量组, 并且使变量之间达到最小的共线性。图 6(a) 显示当 SPA 所选变量数为 22 时, RMSEP 最低, 并且随着所选变量数的进一步增加, RMSEP 几乎不再降低。因此这 22 个变量作为 SPA 选定变量数。由图 6(b) 所示, 在 UVE 处理的基础上, 建模变量个数被 SPA 减少至 22 个, 且涵盖了由 UVE 筛选波段的大部分范围。最终简化的多元线性回归 (MLR) 模型的预测结果如图 7 所示, 其 R^2_c 与 R^2_p 分别为 0.77 与 0.78, 而 RMSEC 与 RMSEP 分别为 0.71% 与 0.64%。对比 UVE-PLS 模型, 其预测精度下降了 4.6%, 但是 MLR 较 PLS 计算简单, 模型易于理解, 并且在实际生产要求不太高时, 这种方式能够大大减少模型变量个数, 降低模型复杂程度, 并最终提高模型的速率。结果表明简化的 MLR 通用模型能够在波长有限的近红外光谱仪中具有潜在的良好的应用。

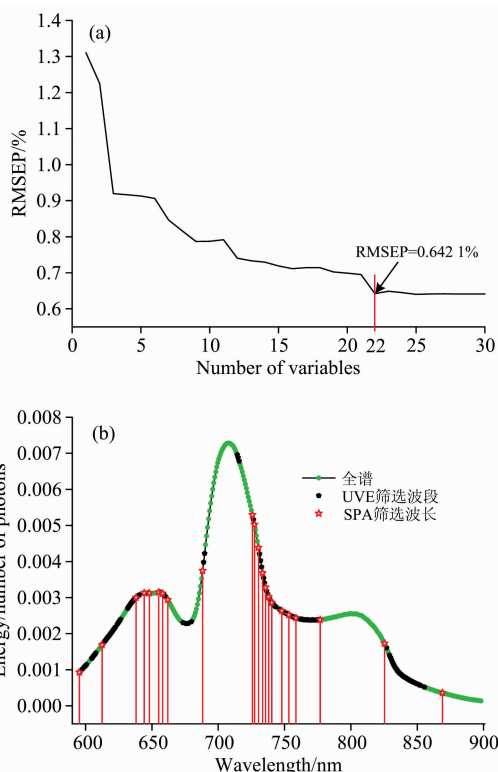


图 6 SPA 变量筛选结果

(a): SPA 选择变量数; (b): SPA 所选变量在光谱中的位置分布

Fig. 6 Results of variable selection using SPA

(a): Number of selected variables;

(b): Locations of the selected variables in processed spectra

2.6 模型验证

根据上述分析, 确定 UVE-PLS 模型为不同品种苹果糖

度的最佳预测模型, 利用测试集数据评估模型的实际性能。将每个品种测试集按照相同的预处理方法进行处理, 并放入模型进行验证, 其结果如表 3 所示。其中, 对于冰糖心, 红富士与水晶富士测试集的 R^2_p 和 RMSEP 分别为 0.79, 0.77 与 0.78, 和 0.75%, 0.63% 与 0.45%。而模型的潜变量数分别为 10, 10 和 6, 与文献[8]的研究相比 (潜变量数为 11~13), 模型得到了简化, 并最终提升了模型的效率。

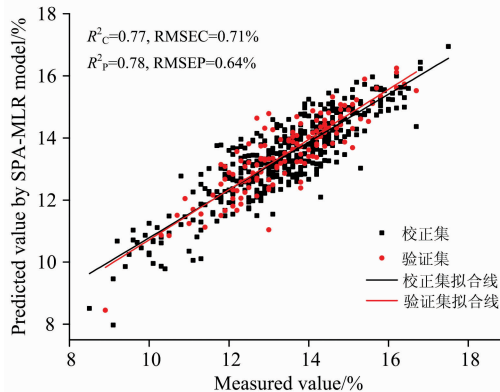


图 7 基于 22 个 SPA 优选变量的 MLR 模型结果

Fig. 7 Result of MLR model based on 22 variables selected by SPA

表 3 苹果糖度最优模型的实际性能

Table 3 The practical performance of the optimal model for SSC of the apples

测试集品种	LVs	R^2_p	RMSEP/%
冰糖心	10	0.79	0.75
红富士	6	0.77	0.63
水晶富士	10	0.78	0.45

3 结论

近红外动态在线检测设备可以用于开发包括冰糖心, 红富士及水晶富士三种苹果的糖度通用模型。采用多元散射校正结合 S-G 卷积平滑的光谱预处理方法, 结合归一化处理, 能够有效去除随机噪声以及基线偏差, 并减小由样品品种引起的光谱差异。采用 PLS 回归算法建立了糖度通用模型, 结合 UVE 波长筛选方法能够有效剔除光谱的无用信息, 并且在保证预测精度的基础上, 模型预测效率由于建模数据的减少从而得到了提升。利用 SPA 进一步筛选建模变量, 并且采用 MLR 算法建立了简化后的糖度通用模型, 其结果表明 MLR 通用模型在波长有限的近红外光谱仪中具有潜在的良好的应用。选定的 UVE-PLS 模型的实际性能表现出对糖度可靠的预测能力。本研究表明水果动态在线分选设备用于建立不同品种水果糖度模型是可行的, 进一步研究多品种建模是必要的。研究简化了水果商品化加工的过程, 降低了建立糖度分选模型的维护成本, 对多品种水果商品化处理具有指导意义。

References

- [1] NIE Ji-yun, LI Zhi-xia, LI Hai-fei, et al(聂继云, 李志霞, 李海飞, 等). *Scientia Agricultura Sinica(中国农业科学)*, 2012, 45(14): 2895.
- [2] Yuan L, Cai J, Sun L, et al. *Food Analytical Methods*, 2016, 9(3): 785.
- [3] Daniela E, Alessia P, Peter R, et al. *Molecules*, 2015, 20(8): 13603.
- [4] Fan S, Guo Z, Zhang B, et al. *Food Analytical Methods*, 2016, 9(5): 1333.
- [5] Arendse E, Fawole O A, Magwaza L S, et al. *Biosystems Engineering*, 2017, 159: 22.
- [6] Escribano S, Biasi W V, Lerud R, et al. *Postharvest Biology and Technology*, 2017, 128: 112.
- [7] Jun S, Yating L, Xiaohong W, et al. *Journal of Food Process Engineering*, 2018, 41(8): e12884.
- [8] Liu R, Qi S, Han D. *Journal of Near Infrared Spectroscopy*, 2015, 23(5): 301.
- [9] Wang J, Wang J, Chen Z, et al. *Postharvest Biology and Technology*, 2017, 129: 143.
- [10] GAO Zi-yi, FAN Xian-guang, YANG Hui-juan, et al(高滋艺, 范献光, 杨惠娟, 等). *Food Science(食品科学)*, 2016, 37(19): 70.
- [11] Chang C W, Laird D A, Mausbach M J, et al. *Soil Science Society of America Journal*, 2001, 65(2): 480.
- [12] Zude-Sasse M, Truppel I, Herold B. *Postharvest Biology and Technology*, 2002, 25(2): 123.
- [13] YAN Yan-lu(严衍禄). *Basic and Application of Near Infrared Spectroscopy(近红外光谱分析基础与应用)*. Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005. 8.
- [14] Yun Y H, Li H D, Deng B C, et al. *TrAC Trends in Analytical Chemistry*, 2019, 113: 102.
- [15] de Araújo Gomes A, Galvão R K H, de Araújo M C U, et al. *Microchemical Journal*, 2013, 110: 202.

Development of Multi-Cultivar Universal Model for Soluble Solid Content of Apple Online Using Near Infrared Spectroscopy

LIU Yan-de, XU Hai, SUN Xu-dong, JIANG Xiao-gang, RAO Yu, ZHANG Yu

School of Mechatronics & Vehicle Engineering, East China Jiaotong University, National and Local Joint Engineering Research Center of Fruit Intelligent Photoelectric Detection Technology and Equipment, Nanchang 330013, China

Abstract Cultivar variability influences the near-infrared modeling and analysis of the internal quality of the fruit due to the different cell structure, composition and optical transmission characteristics of the fruit so that the original model can not predict the fruit quality parameters with high precision. The feasibility of a multi-cultivar model's development for the online determination of the internal quality of apple including "Candy Heart", "Red Fuji" and "Crystal Fuji" was investigated. Near infrared diffuse transmittance spectra of each cultivar were collected by the fruit sorting equipment under the condition of the interval time of 100 ms and motion speed of 5 s^{-1} . The spectral curves of all the cultivars were similar where the prominent absorption peaks were near 650, 709 and 810 nm, and troughs were near 670, 750 and 830 nm, and their variations were mainly reflected in the spectral absorption intensity. The spectral pre-process methods including multiplicative scatter correction, Savitzky-Golay smoothing and normalization were employed to filter out the variations in signals caused by the cultivars. Partial least squares regression method was used to establish the common model for the soluble solid content where the calibration sets of the total samples were combined. Uninformative variable elimination was used to select the modeling variables whose number of effective variables selected was 155, and the performance of the UVE-PLS model resulted in greater coefficient of determination for prediction of 0.80, lower root mean square error of 0.61% and higher residual prediction deviation of 2.21. Successive projections algorithm was employed to select the variables in the wavelengths selected by UVE and the number of variables selected was 22. Multivariable linear regression was used to establish the simplified model, which resulted in coefficient of determination for prediction of 0.78 and root mean square error of prediction of 0.64%. The test sets of all the cultivars were used to access the performance of best universal model, which resulted in latent variables of 6~10, coefficient of determination for prediction of 0.77~0.79 and root mean square error of 0.45%~0.75%. The results highlighted the potential of dynamic on-line sorting instruments for the testing of internal qualities of apples. The prediction range of the single cultivar model was expanded, and the robustness of prediction model among different cultivars were improved by establishing the common model. Appropriate variable selection methods can decrease the number of model variables, reduce the complexity of the model and ultimately increase the model rate. The development of the universal model of different cultivars for predicting internal quality has a good potential application in wavelength-limited near infrared spectroscopy equipment.

Keywords Detection online; Near infrared spectroscopy; Universal model; Partial least squares; Uninformative variable elimination; Apple; Soluble solid content

(Received Mar. 7, 2019; accepted Jun. 26, 2019)

《光谱学与光谱分析》投稿简则

《光谱学与光谱分析》是由中国科协主管,中国光学学会主办,钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的专业学术期刊。国内外公开发行人,从 2004 年起为月刊,大 16 开本,2020 年仍为月刊,每期 332 页。《光谱学与光谱分析》主要报道我国光谱学与光谱分析领域内具有创新性科研成果,及时反映国内外光谱学与光谱分析的进展和动态;发现并培育人才;推动和促进光谱学与光谱分析的发展。为科教兴国服务。读者对象为从事光谱学与光谱分析的科研人员、教学人员、分析测试人员和科研管理干部。

栏目设置和要求

1. 研究报告 要求具有创新性的研究成果,一般文章以 8000 字(包括图表、参考文献、作者姓名、单位和中文、英文摘要,下同)为宜。
2. 研究简报 要求在前人研究的基础上有重大改进或阶段性研究成果,一般不超过 5000 字。
3. 评述与进展 要求评述国内外本专业的发展前沿和进展动态,一般不超过 10000 字。
4. 新仪器装置 要求介绍新型光谱仪器的研制、开发、使用性能和应用,一般不超过 5000 字。
5. 来稿摘登 要求测试手段及方法有改进并有应用交流价值,一般以 3000~4000 字为宜。

稿件要求

1. 投稿者请经本刊编委(或历届编委)一人或本专业知名专家推荐,并附单位保密审查意见及作者署名顺序,主要作者介绍。文章有重大经济效益或有创新者,请说明,同时注明受国家级基金或国家自然科学基金资助情况。
2. 来稿要观点明确、数据真实可靠、层次分明、言简意明、重点突出。来稿必须是网上在线投稿(含各种符号和外文字母大写、小写、正体、斜体;希腊字母、拉丁字母;上角、下角标位置应标清楚)。中文摘要以 800 字为宜,英文摘要(建议经专业英语翻译机构润色)与中文摘要要对照;另附关键词。要求来稿应达到“齐、清、定”,中文、英文文字通顺,方可接受送审。
3. 为了进一步统一和完善投稿方式、缩短论文发表周期,本刊只接收网上在线投稿,不接收以邮寄方式或 e-mail 方式的投稿,严禁“一稿多投”,对侵权、抄袭、剽窃等学术不端行为,一经发现,取消三年投稿资格。
4. 文中插图要求完整,图中坐标、线条、单位、符号、图注等应标注准确、完整。如作者特殊要求需出彩色插图者,必须在投稿时事先加以说明,并承担另加的彩印费用。图幅大小:单栏图 7.5cm(宽)×6cm(高);双栏图:14cm(宽)×6cm(高);图中数字、图题、表题全部用中文、英文对照,图中数字、中文、英文全用 6 号字。电子文档中除实物图外,曲线图要用 Matlab, Excel, Visio 或 Origin 等软件制作,稿件中图片的原图并转成相应(可编辑)的文件格式(.fig, .xls, .vsd, .opj),非“.jpg”格式的文档,随电子版修改稿一同发送到本刊的修改稿专用邮箱。
5. 文中出现的单位必须按“中华人民共和国计量标准”及有关 GB 标准规定缮写。物理量符号一律用斜体,单位符号和词头用正体字母。
6. 名词术语,请参照全国科学技术名词规定缮写。
7. 参考文献,采用顺序编码制,只列主要文献;以 15~20 条为宜。内部资料、私人通讯、未经公开发表的一律不能引用。日文、俄文等非英文文献,请用英文表述;中文文献和中文图书采用中、英文对照表述,文献缮写格式请参照本刊。
8. 请在投稿第一页左下角写明投稿联系人的电话和两个 e-mail,以便及时联系。

稿件处理

1. 自收到稿件之日起,一个月内作者会收到编辑部的稿件处理意见。请根据录用通知中所提出的要求认真修改,希望修改稿在 30 天内寄回编辑部,并作为作者最终定稿(当作者接到校样时,以此修改稿为准进行校对,请勿再做大的改动),若二个月内编辑部没收到修改稿,将视为自行撤稿处理。
2. 有重大创新并有基金资助者可优先发表;不录用的稿件,编辑部将尽快通知作者,底稿一律不退,请自留底稿。
3. 来稿一经发表将酌致稿酬并送样刊 2 册。
4. 遵照《中华人民共和国著作权法》,投稿作者须明确表示,该文版权(含各种媒体的版权)授权给《光谱学与光谱分析》期刊社。国内外各大文献检索系统摘录本刊刊出的论文;凡不同意被检索刊物无稿酬摘引者,请在投稿时事先声明,否则,本刊一律认为已获作者授权认可。
5. 修改稿请寄:100081 北京市海淀区学院南路 76 号(南院西北门),《光谱学与光谱分析》期刊社(收)
电话:010-62182998 或 62181070 传真:010-62181070
e-mail: chngpxygpfx@vip.sina.com; 修改稿专用邮箱: gp2008@vip.sina.com 网址: http://www.gpxygpfx.com