

激光与光电子学进展

基于近红外光谱的苹果可溶性固形物特征波长
筛选建模对比分析张金富^{1,2,3}, 汤斌^{1,2,3}, 王建旭^{1,2,3*}, 传焱霏^{1,2,3}, 龙邹荣^{1,2,3}, 陈庆^{1,2,3}, 缪俊锋^{1,2,3}, 蔡林峰^{1,2,3}, 赵明富^{1,2,3}, 周密²¹光纤传感与光电检测重庆市重点实验室, 重庆 400054;²智能光纤感知技术重庆市高校工程研究中心, 重庆 400054;³重庆理工大学电气与电子工程学院, 重庆 400054

摘要 采用近红外光谱进行检测时,光谱波段包含了大量的噪声和散射,这些都影响了模型的稳定性。基于竞争性自适应重加权算法(CARS)和互信息算法(MI)的特征波长筛选方法来建立偏最小二乘(PLS)回归模型,探测苹果内部的可溶性固形物含量(SSC)。通过光谱仪获取 800~2400 nm 的 120 个样本的漫反射光谱数据,经过预处理之后的数据通过 Kennard-Stone(KS)算法随机选取 96 个作为校正集,24 个作为预测集,然后分别建立全波段 PLS 模型、CARS-PLS 模型和 MI-PLS 模型来对比分析。结果显示:利用全波段建立 PLS 模型,模型的决定系数 R^2 为 0.8511,模型均方根误差(RMSEC)以及预测均方根误差(RMSEP)分别为 0.9413 和 1.1915;CARS 算法筛选的特征波长点变量从 303 减少到了 12,下降了 96.03%,建立的 PLS 模型决定系数 R^2 为 0.8746,上升了 2.76%,RMSEC 和 RMSEP 分别为 0.864 和 0.9757;MI-PLS 模型包含了 56 个特征波长点,选用的波长占全波长的 18.49%, R^2 、RMSEC 和 RMSEP 分别为 0.9218、0.6822 和 0.8235,MI-PLS 与 CARS-PLS 相比特征波长数增长了 64.55%,决定系数 R^2 提高了 0.0472。因此 CARS 和 MI 算法都能很好地解决光谱数据本身的噪声、散射等问题,可以有效用于特征波段筛选,所建立的模型可以对苹果内部 SSC 含量进行测定。

关键词 光谱学; 互信息; 竞争性自适应重加权算法; 近红外光谱; 可溶性固形物

中图分类号 O657.33

文献标志码 A

DOI: 10.3788/LOP221956

Comparative Analysis of Characteristic Wavelength Screening of Apple
Soluble Solids Based on Near-Infrared SpectroscopyZhang Jinfu^{1,2,3}, Tang Bin^{1,2,3}, Wang Jianxu^{1,2,3*}, Chuan Yanfei^{1,2,3}, Long Zourong^{1,2,3},
Chen Qing^{1,2,3}, Miao Junfeng^{1,2,3}, Cai Linfeng^{1,2,3}, Zhao Mingfu^{1,2,3}, Zhou Mi²¹Chongqing Key Laboratory of Optical Fiber Sensing and Photoelectric Detection, Chongqing 400054, China;²Intelligent Optical Fiber Perception Technology, Chongqing University Engineering Research Center,
Chongqing 400054, China;³School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing 400054, China

Abstract When using near-infrared spectroscopy for detection, the spectral band contains significant noise and scattering, which affect the stability of the model. Based on the competitive adaptive reweighting (CARS) and mutual information (MI) algorithms, a partial least-squares (PLS) regression model was established to detect the soluble solid content (SSC) in apples. The diffuse reflectance spectrum data of 120 samples at 800–2400 nm were obtained using a spectrometer. After preprocessing, 96 samples were randomly selected as the calibration set for modeling, and 24 samples were selected as the prediction set for prediction using the Kennard-Stone (KS) algorithm. Next, the full-band PLS model, CARS-PLS model, and MI-PLS model were established for comparative analysis. The results show that the coefficient of determination R^2 of the PLS model is 0.8511, the root-mean-square error of calibration (RMSEC) of the model and the root-mean-square error of prediction (RMSEP) are 0.9413 and 1.1915, respectively. The number of characteristic wavelength point variables screened by the CARS algorithm reduces from 303 to 12, a decrease of 96.03%. The

收稿日期: 2022-06-30; 修回日期: 2022-08-11; 录用日期: 2022-09-05; 网络首发日期: 2022-09-15

基金项目: 国家自然科学基金(61805029)、重庆理工大学科研启动基金项目(0107210299, 0107200283)

通信作者: *wangjianxu@cqut.edu.cn

coefficient of determination R^2 of the PLS model is 0.8746, an increase of 2.76%. The RMSEC and RMSEP values are 0.864 and 0.9757, respectively. The MI-PLS model contains 56 characteristic wavelength points, and the selected wavelength accounts for 18.49% of the total wavelength. R^2 , RMSEC, and RMSEP are 0.9218, 0.6822, and 0.8235, respectively. Compared with CARS-PLS, the number of characteristic wavelengths of MI-PLS increases by 64.55%, and the coefficient of determination R^2 increases by 0.0472. Therefore, the CARS and MI algorithms can overcome the problems of noise and scattering of spectral data and can be effectively used for screening characteristic bands. The established model is suitable for determining the SSC in apples.

Key words spectroscopy; mutual information; competitive adaptive reweighting algorithm; near-infrared spectroscopy; soluble solids

1 引言

苹果被称为水果之王,含有一些多酚化合物和人体所必需的微量元素可以有效清除人体内的代谢物,降低血糖、血脂等,是心脑血管类疾病的天然克星,内部的膳食纤维可以有效促进肠道蠕动,加上其价格实惠可以长期食用,深受人们的喜爱。评价苹果品质的重要指标之一就是可溶性固形物的含量(SSC),它与苹果的口感有着直接的关系,同时也影响着苹果的后续销量^[1-2]。

随着光谱检测技术的发展,水果品质检测早期的化学计量等有损检测比例正在逐渐缩小,近红外光谱检测由于其无损性、高效性和稳定性,越来越受到人们的青睐。近红外光谱检测技术是基于化学计量法的二次判别技术,因为我们无法直接从光谱中看出待测物质的某种成分含量以及其他化学信息,所以需要借助化学检测去人为标定并建立模型。由于采集苹果的近红外光谱时不可避免地会含有大量噪声和干扰波长等无用信息,直接利用这类光谱所建的模型往往难以达到实际的检测要求,所以需要全波段光谱进行特征波长筛选,剔除无关波长,这样不仅可以简化模型,还能提高预测精度。石岩等^[3]使用竞争性自适应重加权采样(CARS)算法,分别从定性和各定量指标的角度,对近红外光谱的特征波长进行分析和研究,筛选出的特征波长数目仅为全变量的0.48%~4.44%,且模型精度得到改善。Xing等^[4]采用先进的数据融合方法,即通过CARS对变量选择步骤后的两种光谱技术的选定变量进行级联,得到较好的结果。胡建等^[5]把CARS和间隔偏最小二乘法(iPLS)两种算法相结合建立偏最小二乘(PLS)模型来检测物质鲜味强度,并取得了较高的相关系数。从上述数据可以看出,经过筛选后的特征波长作为输入集合建模的效果要优于全波段作为输入变量的模型。在众多特征波长筛选方法中:主成分筛选法只考虑自变量的信息,忽略了自变量和因变量的内在因素;间隔偏最小二乘法筛选的波段过少,容易忽略有用信息段;而后向间隔偏最小二乘法提取的信息又过于繁杂,仍然包含较多的无用变量信息。CARS是一种较为新颖的特征筛选算法,该算法能同时降低线性变量对模型的

影响和去除无用变量,最终得到与待测信息相关度最高的特征变量。互信息表示两随机变量之间依赖度的量度:当两变量存在一定相关性时,将会得到某一正值的互信息;当两变量相互独立时,它们的互信息为零。互信息的另一大特点便是不易受到初始化过程中数据传递以及噪声的影响^[6]。我们所需要测量的物质浓度与光谱的幅值存在一定关系,但噪声会干扰有用信息,甚至覆盖有用信息,因此互信息对于定量分析能产生积极作用。

本文对比分析PLS建模以及两种变量筛选算法分别与PLS结合所得出的结果,找寻适合苹果建模的特征波长采集方法,提高模型的准确率与泛化能力。

2 实验部分

2.1 样本

实验样本是从水果店购买的山东烟台栖霞红富士苹果,数量为30个,苹果的赤道直径范围是80~85 mm。在实验前会对每个苹果进行清洗处理,消除苹果表面的污渍和食用蜡。在苹果的赤道处做4个标记,且每个标记间隔90°。由于刚购买的苹果处于低温保鲜状态,需要于测试前在实验室静置24 h使苹果温度统一达到室温,从而降低样品的温差对建模精度的影响。

2.2 检测设备与光谱采集

光谱仪使用的是上海复享光学公司生产的NIR22+Px光谱仪,探测波长范围是800~2400 nm,分辨率为5.3 nm。光源使用的是Ocean Optics公司生产的HL-2000-HP卤素灯,可输出波长范围是360~2400 nm。检测方式如图1所示,采用漫反射的光学结构,光源的出射端口与苹果被测表面呈90°,探测头的角度和水平位置都是可调节状态,通过螺丝固定,这样在检测时可以通过自行调节来找到合适的采样位置,本次实验探头与光源出射端的夹角为45°。在采样之前使用标准白板为背景进行背景光谱采集。使用Morpho3.2(上海复享公司)光谱分析软件来进行光谱数据预处理与分析。

图2为通过NIR22+Px光谱仪对120个样本进行采样后得到的原始光谱图。

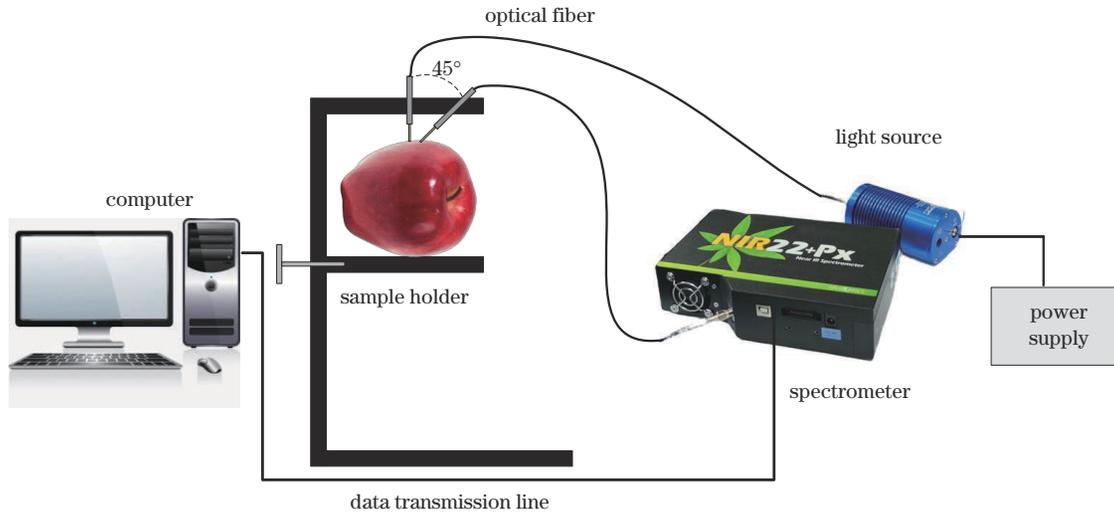


图 1 光谱采集平台示意图

Fig. 1 Schematic diagram of the spectrum acquisition platform

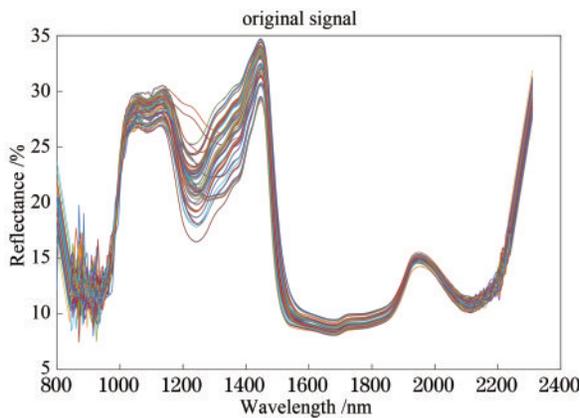


图 2 120 个样本的原始光谱图

Fig. 2 Original spectra of 120 samples

图 2 中的 120 条原始光谱数据在变化趋势上大致相同,局部地区会有细微的差距。每条光谱包含 303 个波长点。在 1000~1200 nm、1400~1500 nm 和 1900~2000 nm 处有较强的光谱反射程度,且样本之间的变化也较为明显,这说明此处包含信息较多。同时,原始光谱存在较多噪声,预处理后会得到更好的效果^[7]。褚小立等^[8]研究了不同光谱预处理方法的适应条件,对于复杂体系需要结合不同的预处理方法才能取得预期效果,再结合刘昊辰^[9]对于苹果光谱预处理的研究,本文先以小波去噪消除光谱背景及暗噪声,再利用多元散射校正(MSC)进行预处理。选取 4 个样本,处理结果如图 3 所示。

2.3 苹果的 SSC 的真实值

苹果的 SSC 真实值通过艾拓公司(Atago Co., 日本)生产的折光仪获取。每个样品在获取近红外光谱数据后马上从被测区域中切取一块带果皮的果肉,人工压汁并过滤后滴于折光仪镜面上,通过折光仪读取并记录实测 SSC 值。

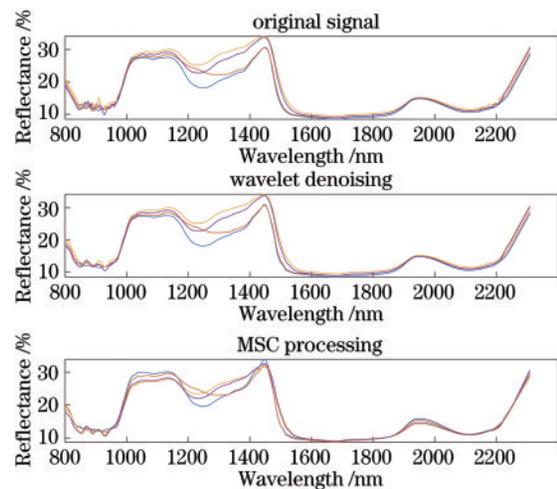


图 3 原始光谱与预处理光谱对比

Fig. 3 Comparison of original and preprocessed spectra

3 建模方法

3.1 建模数据集划分

建模前需要对数据集进行划分,这里我们选用 Kennard-Stone 算法,在 120 个样本集中选取 96 个作为训练集,剩余的 24 个作为测试集,样本集的 SSC 统计结果如表 1 所示。然后,人为地将苹果 SSC 的最值及对应的光谱数据调整到训练集中,以确保数据集的 SSC 跨度足够广,从而增强模型的泛化能力。

3.2 传统 PLS 模型

经过实验可知,传统 PLS 算法在光谱数据建立回归模型时效果较为突出。PLS 回归模型的公式为

$$Y = aX + b, \quad (1)$$

式中: Y 表示待测含量矩阵; X 表示光谱数据矩阵; a 为光谱数据的回归系数向量; b 为模型的偏差值向量。

建立 PLS 模型时,通常会发生过拟合和欠拟合的现象,这都和潜变量(LVs)的个数相关,常采用交叉验

表 1 数据集中 120 个样本的 SSC 统计

Table 1 Statistics of SSC of 120 samples in the data sets

Data set	Number of samples	Min / %	Max / %	Mean / %	Standard deviation / %
Calibration set	96	8.78	14.62	11.20	0.92
Prediction set	24	8.92	13.53	11.18	0.92
Overall	120	8.78	14.62	11.19	0.91

证法来确定 LVs 的个数,当交叉验证后的均方根误差最小时所确定的 LVs 的数量就是最适变量数。

3.3 竞争性自适应重加权算法

基于适者生存的思想,分别利用 PLS 系数的特征筛选方法和蒙特卡罗(MC)采样生成了一种竞争性自适应重加权采样法^[10]。蒙特卡罗采样公式为

$$W_i = |b_i| / \sum_{i=1}^m |b_i|, \quad (2)$$

式中: $|b_i|$ 为第 i 个变量的回归系数绝对值; W_i 为第 i 个变量的回归系数绝对值权重; m 为每次采样的剩余变量数。

再利用指数衰减函数(EDF)强行去除回归系数绝对值权重相对较小的波长。在第 i 次基于 MC 采样建立 PLS 模型时,根据 EDF 得到保留的波长点的比例 R_i 为

$$R_i = \mu \exp(-k_i), \quad (3)$$

式中: μ 和 k 是常数。

在每次采样时,都是从前一次采样时的变量数中采用自适应加权采样(ARS)选择数量为 $R_i \times n$ 个波长变量建立 PLS 回归模型,根据回归系数的权值来取舍数据,建立相关度较高的子集。多次迭代后,选择 PLS 模型的交叉验证均方根误差(RMSECV)最小的子集中波长作为特征波长^[11]。

3.4 互信息特征筛选

互信息(MI)作为 Shannon 信息论中的一个重要概念^[12],被定义为

$$I(X; Y) = G(X) - G(X|Y), \quad (4)$$

式中: $I(X; Y)$ 表示 X 和 Y 两变量之间的相互依赖性的程度,也可以表示两者间的共享信息,即互信息; $G(X)$ 表示变量 X 的不确定度的量度; $G(X|Y)$ 表示变量 X 未涉及到变量 Y 的不确定度的量度^[12]。

由于 $G(X) \geq G(X|Y)$, 因此互信息的值 $I(X; Y)$ 总是不小于零。根据互信息的对称性可得 $I(X; Y) = I(Y; X)$ 。根据熵的定义,两个离散随机变量 X 和 Y 的互信息可以写成

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left[\frac{p(x, y)}{p(x)p(y)} \right], \quad (5)$$

式中: $p(x)$ 表示 $X=x$ 的概率; $p(y)$ 表示 $Y=y$ 的概率; $p(x, y)$ 为两者同时发生的概率。

在互信息和信息熵的定义中,必须预先得知随机变量的概率分布情况,但是真实数据的分布概率往往是未知的,因此会采用熵的估计法或无参数概率密度

来求互信息的近似值。本文采用 Kraskov 等^[13]的最邻近方法来估计互信息,主要依据 Python 库中 Mutual_Info_Regression 函数包来实现。基于互信息的 PLS 建模步骤如图 4 所示。

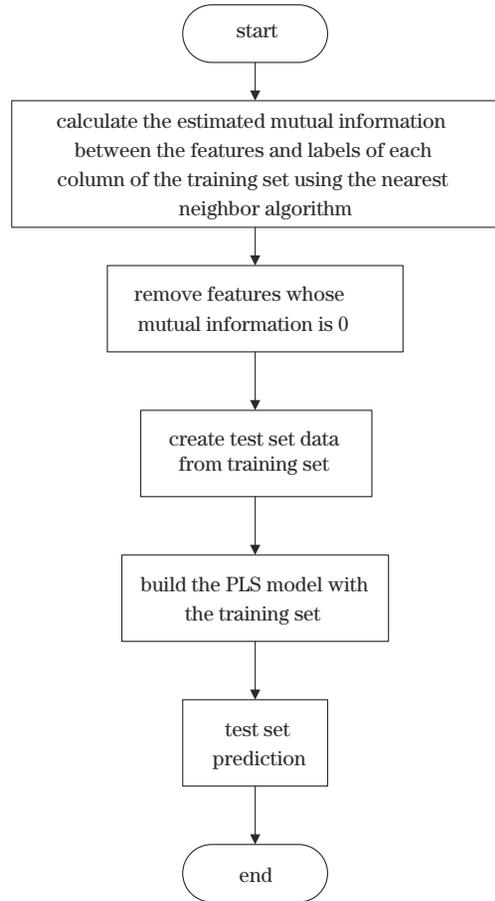


图 4 基于互信息 PLS 建模流程图
Fig. 4 Flow chart of PLS modeling based on mutual information

3.5 建模方法与评价

先建立全波段 PLS 回归模型,再利用 PLS 结合 MI 和 CARS 特征波长筛选算法获得的变量分别建立苹果的 SSC 预测的 PLS 校正模型,对三者进行对比分析。

本文分别采用决定系数 R^2 、均方根误差(MSE)、建模均方根误差(RMSEC)和预测均方根误差(RMSEP)来评价模型^[14]。通常一个好的模型的 R^2 会更接近于 1, MSE 接近于 0,同时 RMSEC 和 RMSEP 的值要尽可能相近^[15]。

4 结果与讨论

4.1 全谱 PLS 模型结果分析

依据决定系数 R^2 和 MSE 来综合考虑主成分的个数, 如图 5 所示。

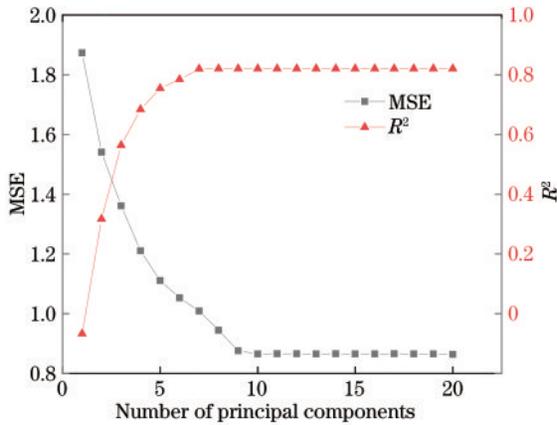


图 5 PLS 主成分个数与训练集 R^2 和 MSE 的关系

Fig. 5 Relationship between number of PLS principal components and training sets R^2 and MSE

从图 5 可以看出: 当主成分个数为 1~9 时, MSE 快速下降, R^2 快速上升; 当主成分个数为 10~20 时, MSE 和 R^2 都趋于稳定。因此, 选择 9 个主成分来建立传统 PLS 定量模型。

训练集和测试集的真实值与预测值的结果如图 6 所示。传统的全波长 PLS 模型的定量分析并未达到理想效果, 测试集的决定系数 R^2 仅为 0.8511, 且 RMSEC 和 RMSEP 分别为 0.9413 和 1.1915, 二者具有较大差异, 发生了一定的欠拟合。原始光谱维度过高, 其中包含了一定的有用和较多无用信息, 同时还有多余的噪声干扰, 在样本量较小时, 该方法不利于建模的稳定性^[16]。

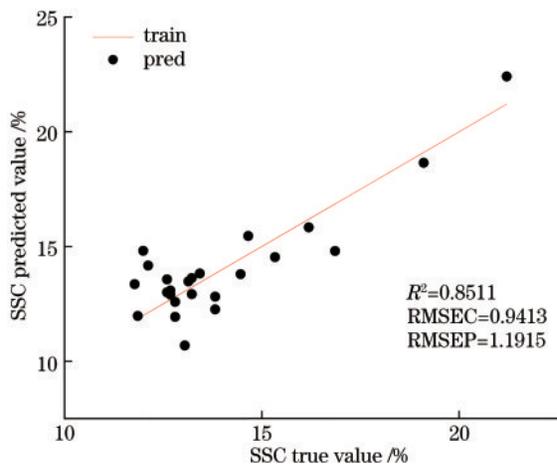


图 6 全光谱 PLS 分析结果

Fig. 6 Full spectrum PLS analysis results

4.2 CARS-PLS 模型结果分析

图 7 为设定 50 次蒙特卡罗采样次数后的特征波段筛选结果图。其中: 图 7(a) 为单次算法运行过程中, 保

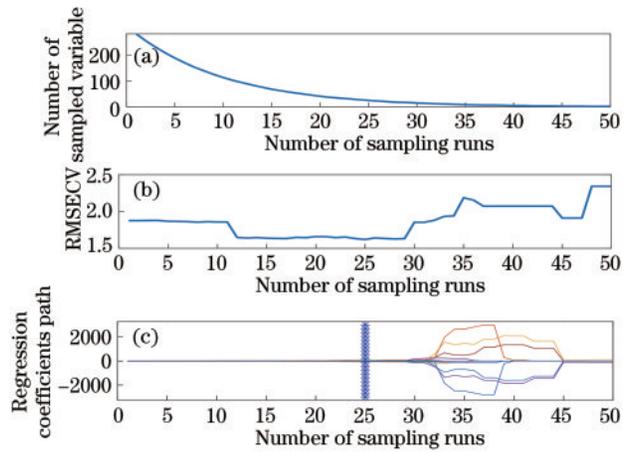


图 7 基于 CARS 的关键变量筛选结果。(a) 每次运行的保留变量; (b) 基于 CARS 建立的 PLS 模型的 RMSECV; (c) 每次运行过程中变量回归系数的变化, 星号表示最低的 RMSECV

Fig. 7 Screening results of key variables based on CARS. (a) Reserved variables per run; (b) RMSECV of PLS model based on cars; (c) change of variable regression coefficient during each operation, and the asterisk indicates the lowest RMSECV

留的变量数随采样次数的变化情况, 由于指数函数的衰减性, 保留变量在采样前期就会急速下降, 到达某一值后, 衰减速度会变缓慢, 表明了该算法具有“粗选”和“精选”两个过程; 图 7(b) 的横纵坐标分别表示采样次数和交叉验证的均方根误差, 采样次数为 1~25, 均方根误差逐渐降低, 表明此时红外光谱中的无用信息在逐渐被去除, 采样次数超过 25 次后, 均方根误差上升, 说明此时一些有用变量信息在被剔除; 图 7(c) 显示了 303 个变量在每一次的 MC 采样中的回归系数变化路径, 星号垂线标示的是所建立的 PLS 模型的十折交叉验证的均方差最小时的采样次数, 即第 25 次。因此选取 25 次为最佳采样次数, 此时筛选的子集作为测定苹果 SSC 的特征变量进行回归建模, 模型结果如图 8

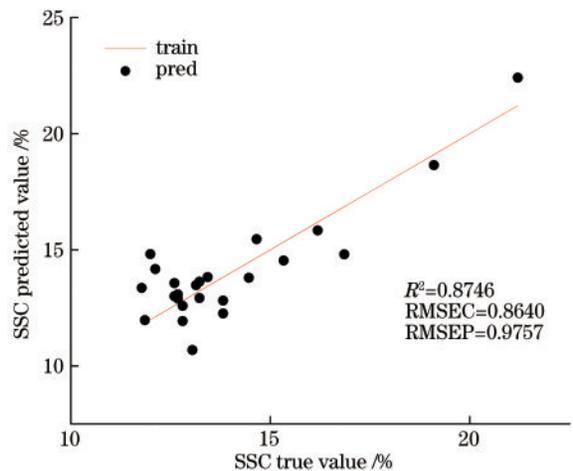


图 8 CARS-PLS 结果分析

Fig. 8 Analysis of CARS-PLS results

所示。

从图 8 可以看出,利用 CARS 算法筛选完自变量后的 PLS 模型精度有了细微的提高。对苹果 SSC 含量的测定的决定系数 R^2 从 0.8511 上升到 0.8746,上升了 2.76%,但是变量从 303 减少到了 12,下降了 96.03%,RMSEC 和 RMSEP 分别从 0.9413 和 1.1915 降到了 0.864 和 0.9757。

4.3 MI-PLS 模型结果分析

通过 MI 算法分别得出各个自变量和因变量的相关量度,去除相关性较弱的自变量,经特征筛选后剩余的有效特征个数为 56,用剩余的 56 个特征进行 PLS 建模,其表现效果如图 9 所示。

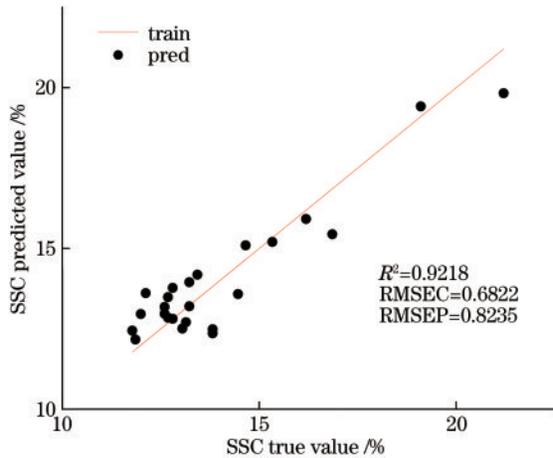


图 9 MI-PLS 结果分析

Fig. 9 Analysis of MI-PLS results

从图 9 可以看出,经过互信息变量筛选的 PLS 模型对苹果 SSC 含量的定量分析准确度有了显著提高,决定系数 R^2 从 0.8511 上升到 0.9218,上升了 8.3%,变量从 303 减少到了 56,下降了 81.51%,RMSEC 和 RMSEP 分别从 0.9413 和 1.1915 降到了 0.6822 和 0.8235。

4.4 对比分析

CARS 和 MI 两种方法选取的特征波长如图 10 所示。从图 10 可以看出,CCARS 和 MI 方法选取的特征波长变量在 1100 nm、1250 nm 和 1900 nm 处有重合,其他波长处并没有明显的重合。CARS 算法选出的 1000 nm、1050 nm、1179 nm、1450 nm 等特征波长分别位于 N—H 基团的三倍频、C—H 基团的三倍频以及 O—H 基团的两倍频吸收带附近。MI 算法选出的波长囊括了 CARS 算法的部分波长点,还反映了其他基团的相关信息。其中,C—H 基团和 O—H 基团反映了可溶性固形物的化学特征,而经由波长筛选算法选出的有关 N—H 等其他基团的特征波长需要作为校准波长来建模,以防止光谱偏移所造成的模型误差。

从表 2 中可以看出,全波段-PLS 模型的 R^2 、RMSEC、RMSEP 分别为 0.8511、0.9413、0.1915,

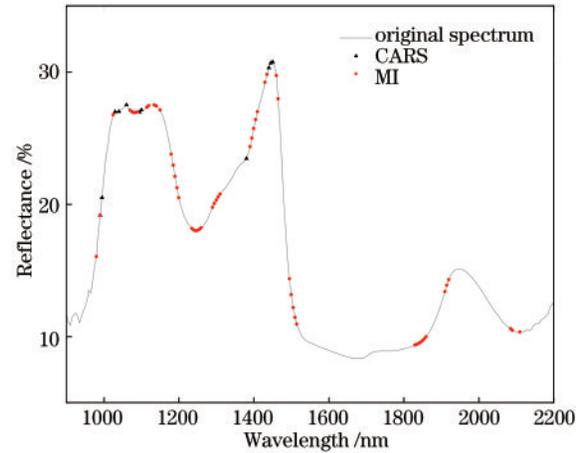


图 10 选取的变量

Fig. 10 Selected variables

表 2 使用不同变量选择算法得到的变量 PLS 模型对苹果 SSC 的预测结果

Table 2 SSC prediction results of apple by PLS models using the variables obtained by different variable selection algorithms

Modeling method	Number of wavelengths	LVs	R^2	RMSEC	RMSEP
Full band -PLS	303	16	0.8511	0.9413	1.1915
CARS-PLS	12	10	0.8746	0.8640	0.9757
MI-PLS	56	15	0.9218	0.6822	0.8235

CARS-PLS 模型和 MI-PLS 模型的 R^2 分别为 0.8746 和 0.9218,两种模型的均方根误差分别为 0.8640 和 0.6822,预测均方根误差分别为 0.9757 和 0.8235。三者的 R^2 都超过了 0.5,说明三个模型对于苹果的 SSC 都有较好的预测结果,但是利用互信息筛选的特征波长模型精度更高,选用的波长占全波长的 47.85%,MI-PLS 与 CARS-PLS 相比特征波长数增长了 64.55%,但是决定系数 R^2 却提高了 0.0472。

5 结 论

本文将近红外光谱技术分别结合竞争性自适应重加权特征波长筛选算法和互信息特征波长筛选算法应用于苹果内部 SSC 的预测。通过结论可以看出,采用 MI 算法获得的关键变量建立的 PLS 模型性能优于全波段建模,决定系数达到了 0.9218。CARS 算法大大减少了原始变量中的无用信息变量,但同时也删减了一些有用变量,因此,决定系数相较于全波长,仅由 0.8511 提高到了 0.8746。综上所述,CARS 和 MI 算法都能有效降低数据中的噪声,提高模型的稳定性和准确度,可以有效用于特征波段筛选,所建立的模型可以对苹果内部 SSC 进行测定,均剔除了大量的非相关性变量,降低了计算量,为开发多点单波长的无损检测仪器提供了理论支撑。

参 考 文 献

- [1] Pourdarbani R, Sabzi S, Jarolmasjed S, et al. Determination of the most effective wavelengths for prediction of fuji apple starch and total soluble solids properties[J]. *Applied Sciences*, 2020, 10(22): 8145.
- [2] Mogollón M R, Contreras C, de Freitas S T, et al. NIR spectral models for early detection of bitter pit in asymptomatic ‘Fuji’ apples[J]. *Scientia Horticulturae*, 2021, 280: 109945.
- [3] 石岩, 孙冬梅, 熊婧, 等. 近红外光谱结合竞争性自适应重加权采样算法用于人工牛黄的质量分析研究[J]. *中国药学杂志*, 2018, 53(14): 1216-1221.
- Shi Y, Sun D M, Xiong J, et al. Analysis of artificial cow-bezoar by near-infrared spectroscopy coupled with competitive adaptive reweighted sampling method[J]. *Chinese Pharmaceutical Journal*, 2018, 53(14): 1216-1221.
- [4] Xing Z, Du C W, Shen Y Z, et al. A method combining FTIR-ATR and Raman spectroscopy to determine soil organic matter: improvement of prediction accuracy using competitive adaptive reweighted sampling (CARS) [J]. *Computers and Electronics in Agriculture*, 2021, 191: 106549.
- [5] 胡建, 冯耀泽, 王益健, 等. 基于近红外光谱的鲜味物质与鲜味强度检测[J]. *光学学报*, 2022, 42(1): 0130002.
- Hu J, Feng Y Z, Wang Y J, et al. Detection of umami substances and umami intensity based on near-infrared spectroscopy[J]. *Acta Optica Sinica*, 2022, 42(1): 0130002.
- [6] Darbellay G A. An estimator of the mutual information based on a criterion for conditional independence[J]. *Computational Statistics & Data Analysis*, 1999, 32(1): 1-17.
- [7] 郝勇, 吴文辉, 商庆园, 等. 山茶油中油酸和亚油酸近红外光谱分析模型[J]. *光学学报*, 2019, 39(9): 0930004.
- Hao Y, Wu W H, Shang Q Y, et al. Analysis model of oleic and linoleic acids in camellia oil via near-infrared spectroscopy[J]. *Acta Optica Sinica*, 2019, 39(9): 0930004.
- [8] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. *化学进展*, 2004, 16(4): 528-542.
- Chu X L, Yuan H F, Lu W Z. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique[J]. *Progress in Chemistry*, 2004, 16(4): 528-542.
- [9] 刘昊辰. 苹果缺陷和糖度的近红外光谱技术动态在线检测研究[D]. 南昌: 华东交通大学, 2021.
- Liu H C. Dynamic on-line detection of apple defect and sugar content by near infrared spectroscopy[D]. Nanchang: East China Jiaotong University, 2021.
- [10] Li H D, Liang Y Z, Xu Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. *Analytica Chimica Acta*, 2009, 648(1): 77-84.
- [11] 刘爽, 于海业, 朴兆佳, 等. 基于高光谱技术反演大豆生理信息的特征波长提取方法研究[J]. *光谱学与光谱分析*, 2020, 40(11): 3542-3548.
- Liu S, Yu H Y, Piao Z J, et al. Study on extracting characteristic wavelength of soybean physiological information based on hyperspectral technique[J]. *Spectroscopy and Spectral Analysis*, 2020, 40(11): 3542-3548.
- [12] Zhu X. Semi-supervised learning with graphs[D]. Carnegie Mellon University, 2005.
- [13] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information[J]. *Physical Review E*, 2004, 69(6): 066138.
- [14] 何国康, 袁凯, 张志勇, 等. 基于二维相关近红外光谱术的小米含水率检测[J]. *激光与光电子学进展*, 2022, 59(8): 0830002.
- He G K, Yuan K, Zhang Z Y, et al. Millet moisture content detection based on two-dimensional correlation near infrared spectroscopy[J]. *Laser & Optoelectronics Progress*, 2022, 59(8): 0830002.
- [15] Pombo M A, Dotto M C, Martínez G A, et al. UV-C irradiation delays strawberry fruit softening and modifies the expression of genes involved in cell wall degradation [J]. *Postharvest Biology and Technology*, 2009, 51(2): 141-148.
- [16] Hou J J, Zhang L, Zhao Y, et al. Mechanisms and efficient elimination approaches of self-absorption in LIBS [J]. *Plasma Science and Technology*, 2019, 21(3): 129-143.