

基于遗传算法和间隔偏最小二乘的苹果硬度特征波长分析研究

屠振华¹, 籍保平¹, 孟超英², 朱大洲¹, 史波林³, 庆兆坤^{1*}

1. 中国农业大学食品科学与营养工程学院, 北京 100083
2. 中国农业大学信息与电气工程学院, 北京 100083
3. 中国标准化研究院食品与农业标准化研究所, 北京 100088

摘要 利用傅里叶近红外光谱(FT-NIRS)测定了苹果的硬度。通过使用几种基于遗传算法和间隔偏最小二乘法的特征波长选取方法,包括动态向后间隔偏最小二乘(dynamic backward version of interval PLS, dynamic biPLS)、动态向后间隔偏最小二乘结合遗传算法(dynamic biPLS & GA-PLS)和反复的遗传算法(iterative GA-PLS),分析了苹果硬度的特征波长。结果表明,运用遗传算法和间隔偏最小二乘选择特征波长后,不但可以降低模型的复杂度,同时能够达到提高模型预测精度的效果。在此基础上,研究分析了苹果硬度特征波长的物理化学意义。由于果胶是在苹果成熟过程中一种和硬度有很大关联的物质,通过比较苹果硬度的特征波长和果胶的特征吸收峰,发现两者具有很好的一致性。因此,采用遗传算法和间隔偏最小二乘法得到的苹果硬度的特征波长能够反映果胶的吸收信息,从而解释了近红外技术检测苹果硬度的机理。

关键词 近红外光谱; 硬度; 遗传算法; 间隔偏最小二乘法; 苹果; 果胶

中图分类号: O657.3, S37 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2009)10-2760-05

引言

近年来,对于苹果等水果的内部品质(可溶性固形物、硬度等)的无损检测得到了国内外学者的广泛关注和研究^[1]。苹果的硬度是一种反映细胞间结合力的物理指标,常被用于评判苹果成熟度及采集时间,并对水果的采后存储、保鲜具有很高的参考价值。目前常用的苹果硬度常规检测方法为果肉穿刺法。然而,此方法损伤水果品质,同时不适用于大批量实验^[2]。近红外光谱技术具有检测速度快、成本低、测试重现性好、无需样品预处理等特点^[3],利用该技术结合化学计量学算法提取有效光谱信息进行样品定性或定量分析,已被广泛应用于苹果硬度的测定中^[4,5]。但研究发现,现有的苹果硬度近红外检测精度尚难达到在商业上应用的要求。究其原因,是因为近红外技术是一种检测含氢基团在近红外谱区的合频和倍频信息的技术,而苹果的硬度是一个与苹果密度、细胞结构和组织结构相关的物理指标,近红外检测苹果硬度属于一种间接检测技术。可见,通过特征波长的选取以及分析其与硬度测量的关系,同时分析近红外苹果硬度的检测机理就显得尤为重要。

本研究通过遗传算法和间隔偏最小二乘等特征波长的算

法,在近红外波段选择苹果硬度的特征波长,同时通过分析特征波长的物理化学意义,从而研究苹果硬度的近红外检测机理。

1 材料与方法

1.1 苹果样品采集

从市场上购买山东烟台红富士苹果 170 个,依次做好标记,置于 4℃ 冰柜中贮藏。实验前,将冰柜中取出的苹果置于实验室中 12 h,以使苹果整体温度达到与环境温度一致。光谱采集和硬度测量实验均在 ~26℃ 的环境温度下进行。

1.2 光谱采集仪器及方法

采用 ANTARIS 傅里叶变换近红外光谱仪(Thermo Nicolet, 美国),使用自带的智能光纤枪采集漫反射光谱。所用参数设置如下:光谱采集范围 10 000~4 000 cm^{-1} ,采样点数 1 557 点,采样间隔 4.0 cm^{-1} ,扫描 64 次取平均,仪器使用 InGaAs 检测器。实验时,每个苹果进行 4 次光谱测量,分别位于苹果赤道等距离的 4 个位置,尽可能避免明显的表面缺陷(擦伤、伤疤等)。然后取 4 个点平均光谱为该样本的光谱。

1.3 苹果样本光谱采集和硬度标准值测定

收稿日期: 2008-04-16, 修订日期: 2008-07-22

基金项目: 国家“十一五”重大科技专项项目(2006BAD05A06-Z1)资助

作者简介: 屠振华, 1983 年生, 中国农业大学食品科学与营养工程学院在读博士 *通讯联系人 e-mail: qingzhaoshen@cau.edu.cn

本实验中苹果硬度根据 GB10651-89 标准, 采用 GY-1 型硬度压力计(牡丹江市机械研究所)测量。光谱采集后, 在对应的光谱采集位置, 用直径为 10 mm 的探针刺入对应位置去皮后果肉中 8 mm 深, 以得到果实硬度($N \cdot cm^{-2}$), 同样取 4 个点平均硬度为该样本的硬度值。

1.4 特征波长选择算法原理

研究采用的特征波长选取方法主要为遗传算法和间隔偏最小二乘法^[7-9]。遗传算法是以达尔文的适者生存和优胜劣汰的生物进化论为基础, 模拟生物界的遗传和进化过程而建立的一种优化方法。主要操作为选择(Selection)、交叉(Crossover)和变异(Mutation)算子。其基本思想为将“父代”中建模效果好的波长点组合作为个体进入下一代; 下一代中的个体通过交叉和变异等遗传操作, 产生新的个体, 即“后代”; 经过若干代后, 建模效果好的“后代”代替原来的个体, 最终算法收敛于最好的个体, 即波长选择的最优组合解。而间隔偏最小二乘是一种通过分析不同波段对 PLS 建立模型的影响来选择特征波长的方法。

本文使用的特征波长选择的方法主要有^[7-9]: (1) iterative GA-PLS, 针对近红外光谱采样点数较多的特点, 为防止过拟合现象的发生, 采用反复遗传算法, 对包含 1 485 个波长点波段, 将每 10 个连续波长点取平均值作为一个新变量, 总计 148 个新变量, 经过 5 次重复遗传算法后, 挑选原始波长点再进行遗传算法。其算法的具体参数设定初始群体大小为 30, 最大繁殖代数 100, 交叉概率 0.5, 变异概率 0.01。(2) dynamic biPLS, biPLS 将整个波段划分一定数量的谱区(间隔), 然后依次剔出一个谱区建立 PLS 模型, 通过比较各个建模结果来选择首先剔出的波段, 然后在剩下的波段中进行同样的工作, 直至剩下最后一个谱区。然而, 在运用 biPLS 时候如果间隔数量过少容易遗漏细节信息, 过多会导致最后结果过于复杂。因此, 对 dynamic biPLS 采用从 16 到 25 个间隔数, 连续运算 3 次来选择特征波长。(3) dynamic biPLS & GA-PLS, 在上述利用 dynamic biPLS 算法选出的特征谱区范围内结合利用遗传算法来进一步选择特征波长, 遗传算法的具体参数设定同上。

2 结果与讨论

2.1 近红外光谱分析及分析波段的选取

图 1(a) 为 10 个苹果样本在 1 000 ~ 2 500 nm 范围内的光谱图。通过图 1(a) 可以看到在长波区的光谱噪声较大, 因此去除 2 338 ~ 2 500 nm 范围内的光谱, 共 72 个点, 余下波长为 1 000 ~ 2 338 nm 光谱来建立苹果硬度的校正模型。去除波段后, 全部苹果的光谱的平均光谱图见图 1(b)。

2.2 异常样本的剔除和样本集划分

异常样本会影响模型的稳健性和预测精度, 为了建立可靠、稳定的近红外定量分析模型, 需在建模前进行异常样本的剔除。本研究利用外在学生化残差—杠杆值图来剔除异常样本。通过分析得到苹果的异常样本数为 3 个, 对余下 167 个样本按照 Kennard-Stone 法^[10]以 3 : 1 分为 2 组, 其中校正集 126 个, 预测集 41 个, 样本中苹果含量的统计结果见

表 1。

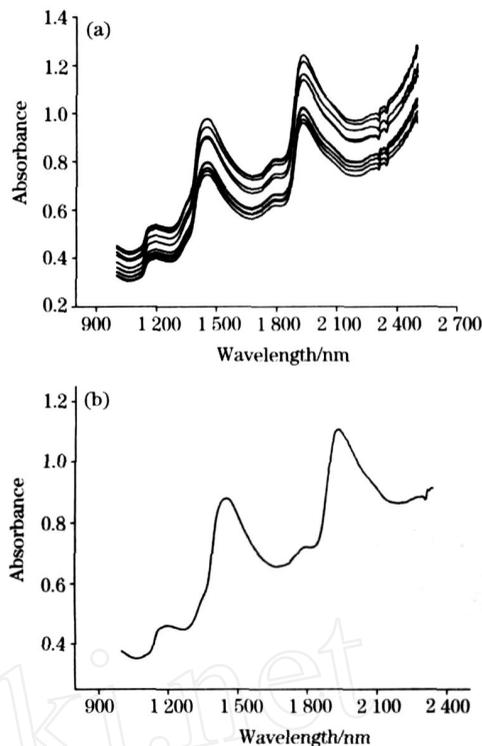


Fig 1 Spectra of apple samples (a) FTIR spectrum of 10 samples (b) average FTIR spectrum of 167 samples

Table 1 Statistic firmness values of calibration and prediction sets of apples^a

Sample set	n^b	Range	Mean	SD ^c
Calibration	126	40.96 ~ 107.60	70.63	15.26
Prediction	41	44.10 ~ 99.76	71.26	14.75

^a: Unit used $N \cdot cm^{-2}$; ^b: number of samples; ^c: stand deviation

2.3 苹果硬度的特征波长选择

2.3.1 Iterative GA-PLS

经过 5 次重复遗传算法后, 280 个原始波长点被挑选出来。对这 280 个变量每 2 个变量作为一个窗口再进行遗传算法, 5 次重复遗传算法后, 最终挑选出 90 个特征波长点。选出特征波长见图 2(a), 利用这 90 个特征波长建模效果见表 2。

2.3.2 Dynamic biPLS

研究采用 dynamic biPLS, 采用 16 到 25 个间隔数, 连续运算 3 次来选择特征波长。这 30 次中选择的频率见图 3(a)。通过计算, 共有 5 个波段的 356 个波长点被选中, 选出特征波长见图 3(b), 利用这 356 个特征波长建模效果见表 2。

2.3.3 Dynamic biPLS & GA-PLS

在图 3 所示的光谱中选出的 356 个波长点, 每 2 个变量作为一个窗口再进行遗传算法, 5 次重复遗传算法后, 最终挑选出 58 个特征波长点。选出特征波长见图 2(b), 利用这 56 个特征波长建模效果见表 2。

Table 2 PLS results of the NIR spectra obtained by different wavelength selection methods

Methods	Selected wavelengths	r	RMSEC/ ($N \cdot \text{cm}^{-2}$)	RMSEP/ ($N \cdot \text{cm}^{-2}$)
Full spectrum	1 485 wavelengths 1 000 ~ 2 234 nm	0.76	9.91	10.11
Dynamic biPLS	356 wavelengths (five regions)	0.82	8.55	9.75
Dynamic biPLS & GA-PLS	58 wavelengths (nine regions)	0.83	8.35	9.65
Iterative GA-PLS	90 wavelengths (eleven regions)	0.85	7.93	8.65

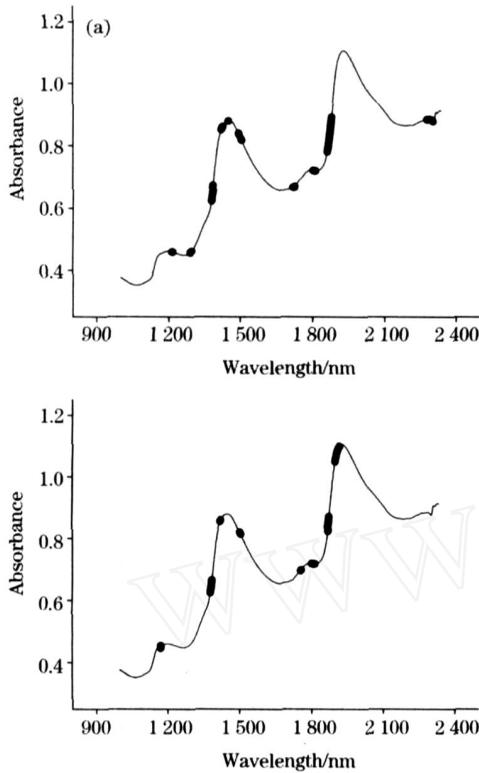


Fig 2 Spectra data set (a) location of the retained regions in the spectrum by iterative GA-PLS; (b) location of the retained regions in the spectrum by dynamic biPLS & GA-PLS

表 2 显示了原始波长下和通过各种波长选择方法处理后 PLS 建模的效果。从表中, 可以看到经过特征波长选择后模型的预测精度较原始波长下均有不同程度的提高, 其中运用 dynamic biPLS 选出 5 个特征波长后模型 RMSEP 由原始的 10.11 ($N \cdot \text{cm}^{-2}$) 下降到 9.75 ($N \cdot \text{cm}^{-2}$), r 由 0.76 提高到 0.82。通过运用 dynamic biPLS & GA-PLS 算法, 模型的 RMSEP 进一步下降到 9.65 ($N \cdot \text{cm}^{-2}$), r 也同时提高到 0.83。同时, 在运用 iterative GA-PLS 选择特征波长后, 模型预测精度达到很大提高, 模型 RMSEP 由原始的 10.11 ($N \cdot \text{cm}^{-2}$) 下降到 8.55 ($N \cdot \text{cm}^{-2}$), r 由 0.76 提高到 0.85, 同时模型的波长点数有 1 485 下降到 90, 模型的复杂度大幅下降。

2.4 苹果硬度特征波长的分析

如图 2 所示, 通过 dynamic biPLS & GA-PLS 选择的特征波长主要在 1 156 ~ 1 159 nm 和 1 900 ~ 1 920 nm 两个波段。同时, iterative GA-PLS 特征波长中有 5 个波段位于

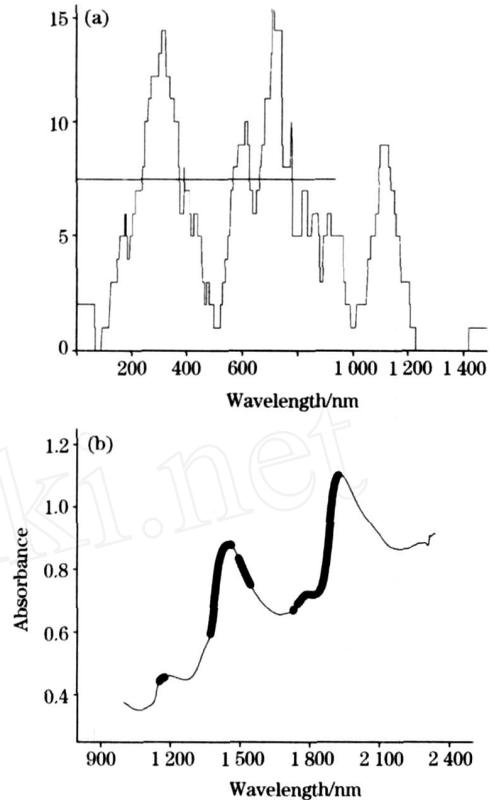


Fig 3 Spectra data set (a) the frequency of selections original wavelengths after dynamic biPLS(30 runs); (b) location of the retained regions in the spectrum by dynamic biPLS

1 380 ~ 1 505 nm 这个范围内。通过上述两种方法选出的特征波长有很多波段是一致的, 包括: 1 380 ~ 1 387, 1 420 ~ 1 425, 1 502 ~ 1 506, 1 803 ~ 1 807 和 1 870 ~ 1 880 nm。除此之外, iterative GA-PLS 还选择了 1 078 ~ 1 081 及 2 280 ~ 2 305 nm 为特征波长。

从图 2 所示的光谱图中, 可以得到, 苹果在近红外谱区的光谱图主要吸收峰位于 1 190 nm, 1 450 和 1 940 nm, 这些吸收峰均是由于水的吸收所导致。其中 1 190 nm 为 O—H 的合频吸收峰, 而 1 940 nm 为 O—H 的伸缩振动的二级倍频^[10], 由于水的吸收很强, 因此对其他成分的检测干扰很大。而在运用 dynamic biPLS & GA-PLS 和 iterative GA-PLS 算法选出的特征波长中均没有选择上述波长, 从而避免了水吸收造成的影响。同时, 通过图 2 和表 2 可以分析出, 运用这两种方法选择的特征波长有很多位于 1 450 nm 附近, 同样作为 O—H 的伸缩振动的一级倍频^[10], 1 450 nm 也是一

个明显的水的吸收峰。而据文献报道,苹果中和硬度有关的有机物主要是有果胶,纤维素,半纤维素等物质,由于 1 440 和 1 446 nm 也是 C—H 的合频吸收带^[10],而 1 450 nm 处又是 C=O 伸缩振动的三级倍频^[10],所以,虽然在此波长下存在水吸收峰的影响,上述算法还是选择了在 1 450 nm 附近的波长为苹果硬度的特征波长,说明上述算法能够在去除无关波长影响的同时,有效提取了硬度的特征波长。

由于果胶是在苹果成熟过程中一种和硬度有很大关联的物质^[11],其中在未成熟的水果中,果胶物质以原果胶形态与纤维素或半纤维素结合在一起,起到支撑和保持细胞组织形状的作用,使组织表现出坚硬状态,具有抗机械损伤的能力。进入成熟阶段的水果,其果胶物质以可溶性果胶的形态存在于细胞液中,使细胞间结合力松弛,苹果质地变软。过熟的苹果,将果胶发生去甲酯化作用生成无粘性的果胶酸,与纤维素、半纤维素分离,引起明显的胞间层解离,使组织

成疡状态,硬度急剧下降并进入细胞结构衰老期。既然近红外检测原理是检测有机物的吸收,对于硬度这个物理指标,分析其特征波长与果胶的关系就显得尤为重要了。

作为一种主要由甲氧基构成的高分子化合物,果胶的主要特征波长位于 2 250 nm 处,该处也正是 C—H 的伸缩振动的二级倍频^[10]。Boeriu 等^[12]和 Sohn 等^[13]研究表明在 2 248 nm 处的甲氧基吸收峰是果胶的特征波长。本研究通过 Iterative GA-PLS 法也选择了 2 248 nm 附近波长作为苹果的特征波长。此外,选出的苹果硬度的特征波长如还有 1 226 ~ 1 227, 1 380 ~ 1 387 和 1 862 ~ 1 880 nm,这与 Sirsomboon 等研究得到的果胶的特征波长是一致的^[14]。因此,本文采用遗传算法和间隔偏最小二乘法得到的苹果硬度的特征波长反映了果胶的吸收信息,从而很好地解释了近红外技术检测硬度这个物理量的机理。

参 考 文 献

- [1] Abbott J A. Postharvest Biology and Technology, 1999, 15: 207.
- [2] JIAO Qun-ying, WANG Shu-mao(焦群英,王书茂). Advances in Mechanics(力学进展), 1999, 29(4): 583.
- [3] Dull G, Birth G S, Smittle D A, et al. Journal of Food Science, 1989, 54(2): 393.
- [4] Lu R, Ariana D. Applied Engineering in Agriculture, 2002, 18: 585.
- [5] G ónez H A, He Y, Pereira A G. Journal of Food Engineering, 2006, 77: 313.
- [6] Leardi R. Journal of Chemometrics, 1994, 8(1): 65.
- [7] Leardi R, González A L. Chemometrics and Intelligent Laboratory Systems, 1998, 41: 195.
- [8] Norgaard L, Saudland A, Wagner J, et al. Applied Spectroscopy, 2000, 54: 413.
- [9] Kennard R W, Stone L A. Technometrics, 1969, 11: 137.
- [10] Knee M, Bartley I M. Composition and Metabolism of Cell Wall Polysaccharides in Ripening Apples. In Friend J & Rhodes M J C(Eds.), Recent Advances in the Biochemistry of Fruit and Vegetables. New York: Academic Press, 1981. 133.
- [11] Van Buren J P. Function of Pectin in Plant Tissue Structure and Firmness. In Walter R H(Ed.), The Chemistry and Technology of Pectin. New York: Academic Press, 1991. 1.
- [12] Boeriu C G, Stolle-Smits T, Van Dijk C. Journal of Near Infrared Spectroscopy, 1998, 6: A299.
- [13] Sohn M R, Cho R K. Journal of Korean Society for Horticultural Science, 2000, 41: 65.
- [14] Sirsomboon P, Tanaka M, Fujita S, et al. Journal of Food Engineering, 2007, 78: 701.

Analysis of NIR Characteristic Wavelengths for Apple Flesh Firmness Based on GA and iPLS

TU Zhen-hua¹, JI Bao-ping¹, MENG Chao-ying², ZHU Da-zhou¹, SHI Bo-lin³, QIN G Zhao-shen^{1*}

1. College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China

2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

3. Institute of Food and Agriculture Standardization, China National Institute of Standardization, Beijing 100088, China

Abstract In the present study, the fruit flesh firmness of apple was analyzed by near infrared (NIR) spectroscopy using an FT-NIR spectrometer. The sensitive spectral regions that provide the lowest prediction error were analyzed by different well-known variable selection methods, including dynamic backward interval partial least-squares (dynamic biPLS), sequential application of backward interval partial least-squares and genetic algorithm (dynamic biPLS & GA-PLS), and iterative genetic algorithm partial least-squares (iterative GA-PLS). Iterative GA-PLS, dynamic biPLS & GA-PLS led to a distinct reduction in the number of spectral data points with better predictive quality. Furthermore, the majority of selected wavelengths were content with the characteristic of the sorption bands of fruit flesh firmness. Pectin constituents, complex non-starch polysaccharides, which are related to texture change in apple, play an important role in their harvest maturity, ripening and storage. Comparing NIR characteristic wavelengths of apple flesh firmness and typical absorption bands for pectin, it was found that characteristic wavelengths of apple flesh firmness were consistent with the pectins relevant spectral regions. Therefore, the NIR characteristic wavelengths of apple firmness based on GA and iPLS reflected the chemical component of apple and the results were reasonable.

Keywords NIR; Firmness; GA; iPLS; Apple; Pectics

(Received Apr. 16, 2008; accepted Jul. 22, 2008)

* Corresponding author