

DOI:10.6041/j.issn.1000-1298.2012.09.029

银杏叶总黄酮含量近红外光谱检测的特征谱区筛选*

邹小波 黄晓玮 石吉勇 陈正伟 张德涛

(江苏大学食品与生物工程学院, 镇江 212013)

【摘要】 通过区间偏最小二乘法(iPLS)谱区筛选方法、反向区间偏最小二乘法(biPLS)谱区筛选方法和联合区间偏最小二乘法(siPLS)谱区筛选方法优化光谱特征区间,建立黄酮含量分析模型,并与波数范围为4 000 ~ 8 000 cm^{-1} 的全光谱偏最小二乘(PLS)模型进行比较。结果表明,采用siPLS谱区筛选方法将全光谱均匀划分21个子区间,选择两个子区间(7、12区间)联合时,建立的siPLS谱区筛选模型预测效果最佳,其交互验证均方根误差和预测均方根误差分别为2.950 0和3.000,校正集和预测集相关系数分别为0.938 4和0.943 7。因此采用siPLS谱区筛选方法可以有效选择光谱特征区域,提高建模预测能力,实现银杏叶总黄酮含量的快速检测。

关键词: 银杏叶 总黄酮 含量检测 近红外光谱 偏最小二乘法

中图分类号: 0657.3 文献标识码: A 文章编号: 1000-1298(2012)09-0155-05

Selection of Wavelength Regions to Determine Flavonoids Content in Ginkgo Leaves by FT – NIR Spectroscopy

Zou Xiaobo Huang Xiaowei Shi Jiyong Chen Zhengwei Zhang Detao

(College of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract

In order to improve the detecting accuracy rating and stability of total flavonoids content in ginkgo leaves by near infrared spectroscopy technique, a precision model was established by selecting efficient spectral regions combined with different partial least squares (PLS) selecting wavelength regions methods. Three improved partial least squares (PLS) methods, including interval partial least squares (iPLS) selecting wavelength regions method, backward interval partial least squares (biPLS) selecting wavelength regions method and synergy interval partial least squares (siPLS) selecting wavelength regions method were used to find the most informative ranges and build models with better predictive flavonoids content in ginkgo leaves at first. And then the models were compared with PLS model which was developed on the whole wavelength range 4 000 ~ 8 000 cm^{-1} . Results showed that the models built by the three improved PLS methods had higher predictive ability than that of PLS method. The optimal model was the one that obtained by siPLS selecting wavelength regions method and it separated the whole spectra into 21 intervals and combined two intervals including interval 7 and interval 12, the RMSECV and RMSEP were 2.950 0 and 3.000, calibration and the prediction correlation coefficient were 0.938 4 and 0.943 7. The conclusion is siPLS method can accurately and rapidly predict flavonoids content in ginkgo leaves.

Key words Ginkgo leaves, Flavonoids, Content determination, Near-infrared spectrum, Partial least squares

收稿日期: 2011-10-23 修回日期: 2011-10-31

* 国家高技术研究发展计划(863计划)资助项目(2008AA10Z208)、国家自然科学基金资助项目(60901079)、全国优秀博士基金资助项目(200968)、江苏省农业自主创新计划资助项目(CX(11)2028)和江苏大学拔尖人才启动基金资助项目

作者简介: 邹小波,教授,主要从事农产品、食品品质无损检测研究,E-mail: zou_xiaobo@ujs.edu.cn

引言

黄酮是银杏叶中的主要活性成分之一^[1-3]。银杏叶中黄酮含量的常规检测方法为高效液相色谱法(HPLC)^[4]和紫外分光光度法^[5],虽然高效液相色谱法具有重现性好、结果精准的优点,但费时、费力,消耗大量化学试剂,运行和维护的成本较高,不利于大样本的快速检测。液相色谱法适合检测银杏叶提取物制成的食品和保健品,紫外分光光度法适合检测固体银杏制品中的黄酮含量^[6],但单独使用紫外分光光度法易受标准曲线精度等问题的影响。近红外光谱常被用于农产品检测^[7-8],为了克服化学检测方法的不足,胡刚亮等采用近红外光谱法测定了银杏提取液和银杏提取物粉末中的黄酮含量,取得了较好的预测效果^[9];黄倩倩等采用近红外光谱法测定了黄芩中的黄酮含量^[10];杨玉兰等采用近红外光谱法检测了银杏叶中的黄酮含量,预测集相关系数为90.15%,但该研究中采用的样本数较少(校正集12个,预测集5个),且没有充分优选黄酮对应的特征波段^[11]。

本文采用iPLS、biPLS、siPLS谱区筛选方法对新鲜银杏叶的近红外光谱进行特征子区间优选,并建立对应的定量分析模型,讨论不同划分区间数和区间选择方法对总黄酮含量(质量比)预测模型的影响。

1 材料与方 法

1.1 实验材料与仪器设备

从江苏大学校园里收集银杏叶共120片,为了使收集的样本具有一定代表性,选择不同生长环境的不同植株进行样本采集。随机选择其中90个样本作为校正集,30个样本作为预测集。

设备包括:Antaris II型傅里叶变换近红外光谱仪(美国赛默飞世尔公司);UV-1601型紫外分光光度计;电子天平(北京赛多利斯仪器系统有限公司);离心机(长沙相仪公司)。

1.2 总黄酮含量测定

准确称取4g新鲜银杏叶加热干燥粉碎后用10mL70%乙醇进行超声萃取30min,然后置于离心机中于4000r/min离心10min,取上清液倒入50mL容量瓶并重复上述步骤两次,加入5%NaNO₂溶液2mL,摇匀,静置6min后,加入10%Al(NO₃)₃溶液2mL,摇匀,静置6min后,再加入20%NaOH试液6mL,混匀后,用70%乙醇溶液定容到刻度,静止放置15min之后,取3mL样液用分光光度计在508nm下测定吸光度,以芦丁作标准物计算样本中

的黄酮质量比。新鲜银杏叶总黄酮质量比统计结果如表1所示。

表1 银杏叶样本总黄酮质量比实测值

Tab.1 Statistics of total flavonoids content for calibration and prediction set of ginkgo leaves mg/g

模型	样品数	最小值	最大值	平均值	方差
校正集	90	2.7300	36.6800	13.0706	7.8882
预测集	30	5.9200	34.6200	18.0010	9.7657

1.3 光谱采集及预处理

选取刚采摘的新鲜银杏叶片主叶脉两边叶片的中间部位各一个点进行光谱采集,采用InGaAs检测器,波数范围10000~4000cm⁻¹,扫描次数为32次;波数间隔为3.8536cm⁻¹,每条光谱包含有1557个变量。数据采集过程中,室内湿度基本保持不变,温度保持在25℃左右^[12-13],原始光谱如图1所示。为消除基线漂移和噪声对光谱的影响,对原始光谱采用标准正交变换(SNV)进行预处理。

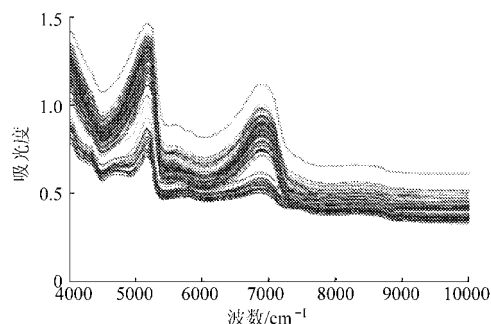


图1 银杏叶原始光谱

Fig.1 Ginkgo leaf sample primitive spectrum chart

1.4 iPLS、biPLS、siPLS谱区筛选方法优选区间的原理

iPLS谱区筛选的基本原理为:将全光谱均匀划分为 n 个子区间,对每个子区间分别建立谱区筛选模型,最优模型对应的区间为入选区间。biPLS谱区筛选的基本原理为:将全光谱均匀划分为 n 个子区间,每次剔除一个区间使得剩余区间对应的PLS模型最优,如此重复直至剩余1个子区间,全局最优PLS模型对应的区间组合为入选区间。siPLS谱区筛选的基本原理为:将全光谱均匀划分为 n 个子区间后,联合其中 m 个区间进行谱区筛选建模,建模时按排列组合的思想对 C_n^m 种组合方式分别进行考察,挑选最优模型对应的 m 个子区间组合作为入选区间,由于此方法的计算量非常大,因此一般 $m < 5$ ^[14-16]。

从上述不同谱区筛选方法的基本原理可以看出,光谱子区间数 n 是一个重要参数,决定了每个子

区间的窗口宽度,且黄酮分子式比较复杂,难以确定其特征区间的宽度,因此本文对光谱区间数 $n \in [5, 80]$ 范围内进行了优化。目前尚不能从理论上确定划分区间数,故每种方法先将整个光谱区域分别划分为 10、11、12、...、25 个子区间,以考察不同数目的子区间划分对模型性能以及最佳波长区间的影响,综合考虑 R (相关系数)值最接近 1 和 RMSECV(交互验证均方根误差)值最小两个原则,选择最佳的划分区间数。

2 结果与讨论

由图 1 可明显看出近红外光谱中 8000 cm^{-1} 以后的波数其信号强度较低,质量较差,因此在之后的建模过程中只取波数范围为 $4000 \sim 8000\text{ cm}^{-1}$ 的光谱值进行研究。

2.1 iPLS 谱区筛选模型优化特征区间

iPLS 谱区筛选模型校正集的交互验证均方根误差在全光谱划分为 15 个子区间、选择第 5 个子区间时最小,校正集的相关系数 R 为 0.934 4,但预测集相关系数较差,为 0.896 0。综合考虑 R 、RMSECV 及 RMSEP(预测均方根误差)的值,研究认为全光谱划分为 19 个子区间,选择第 7 个子区间时最佳,其 RMSECV 和 RMSEP 分别为 3.25 和 3.67,校正集和预测集相关系数分别为 0.920 0 和 0.917 5(如图 2)。对应波数范围 $5270 \sim 5480\text{ cm}^{-1}$ 。

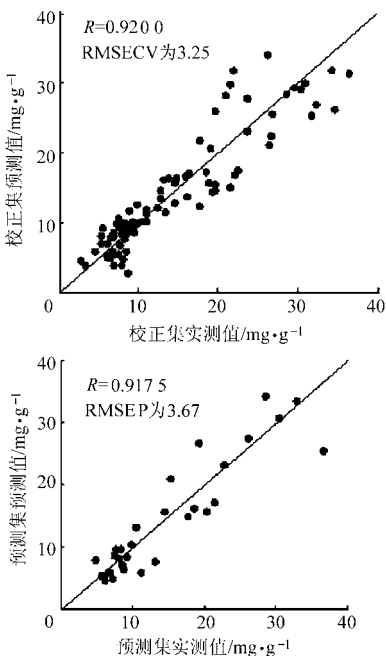


图 2 iPLS 最佳模型的样本预测值与实测值关系

Fig. 2 Reference measured versus NIR predicted by iPLS in prediction and calibration sets

2.2 biPLS 谱区筛选模型优选特征区间

biPLS 谱区筛选模型校正集的交互验证均方根

误差在全光谱划分为 25 个子区间时、5 个子区间(2、4、13、14、18)联合时最小,其 RMSECV 为 3.12,其 RMSEP 相对较差,为 4.17,校正集的相关系数为 0.930 2,预测集相关系数为 0.887 5(如图 3)。对应波数范围 $4160 \sim 4320\text{ cm}^{-1}$ 、 $4480 \sim 4640\text{ cm}^{-1}$ 、 $5910 \sim 6070\text{ cm}^{-1}$ 、 $6070 \sim 6230\text{ cm}^{-1}$ 、 $6710 \sim 6870\text{ cm}^{-1}$ 。

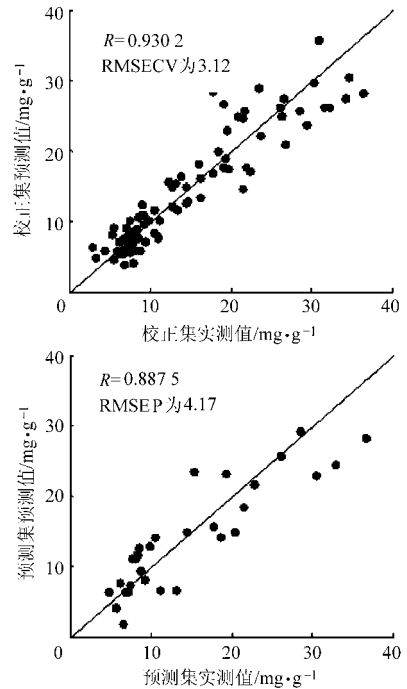


图 3 biPLS 最佳模型的样本预测值与实测值关系

Fig. 3 Reference measured versus NIR predicted by biPLS in prediction and calibration sets

2.3 siPLS 谱区筛选模型优选特征区间

siPLS 谱区筛选模型,2 个区间联合(7、12 区间联合),划分 21 个区间数时最佳,其 RMSECV 和 RMSEP 分别为 2.950 0 和 3.000,校正集和预测集相关系数分别为 0.938 4 和 0.943 7,对应波数范围 $5700 \sim 5890\text{ cm}^{-1}$ 、 $6830 \sim 7020\text{ cm}^{-1}$;3 个区间联合(7、10、16 区间联合),划分 21 个区间数时最佳,其 RMSECV 和 RMSEP 分别为 2.81 和 3.73,校正集和预测集相关系数分别为 0.944 0 和 0.907 8,对应波数范围 $5130 \sim 5320\text{ cm}^{-1}$ 、 $6080 \sim 6260\text{ cm}^{-1}$ 。综合考虑 2 个区间联合效果最佳(图 4)。

2.4 模型比较

通过对划分区间数、联合区间数、区间选取等因素的选择建立的不同谱区筛选回归模型,全光谱模型的校正集和预测集的相关系数 R 分别为 0.913 4 和 0.798 2,交互验证均方根误差和预测均方根误差分别为 3.44 和 5.37,iPLS、biPLS、siPLS 谱区筛选模型与之相比建模的效果都更理想。其中 siPLS 谱区筛选模型(联合 2 个区间)结果最好。

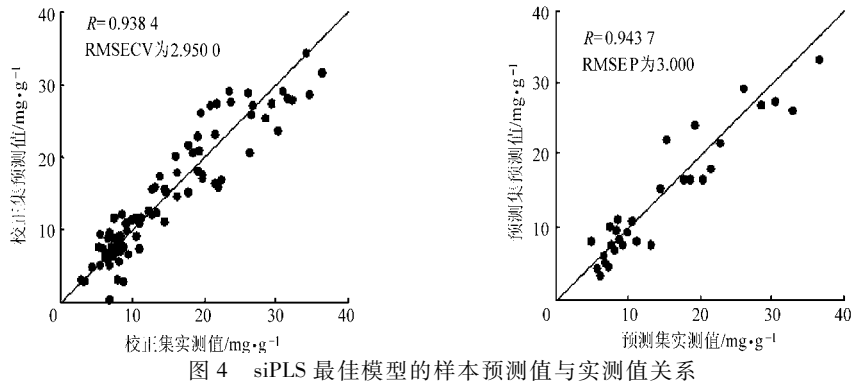


图4 siPLS最佳模型的样本预测值与实测值关系

Fig.4 Reference measured versus NIR predicted by siPLS in prediction and calibration sets

3 结束语

采用 iPLS、biPLS、siPLS 谱区筛选方法对银杏叶近红外光谱的总黄酮特征区间进行了优选,并考察了光谱子区间数对 PLS 建模的影响。将光谱均匀

分为 21 个子区间,7、12 区间联合对应的 siPLS 谱区筛选预测模型最佳,对应的 RMSECV 和 RMSEP 分别为 2.950 0 和 3.000,校正集和预测集相关系数分别为 0.938 4 和 0.943 7。研究结果表明近红外光谱法快速检测新鲜银杏叶中黄酮含量是可行的。

参 考 文 献

- 赵丽繁,庞翔. 脑出血急性期银杏叶制剂的应用[J]. 实用医技杂志,2006,13(3):70~71.
Zhao Lifan, Pang Xiang. Ginkgo biloba application to acute cerebral hemorrhage [J]. Journal of Practical Medical Techniques, 2006, 13(3):70~71. (in Chinese)
- Miao Shifeng, Yu Jinpeng, Du Zhe, et al. Supercritical fluid extraction and micronization of ginkgo flavonoids from ginkgo biloba leaves[J]. Industrial & Engineering Chemistry Research, 2010,49(11):5461~5466.
- 孙旭东,章海亮,欧阳爱国,等. 南丰蜜桔可溶性固形物近红外特征波段选择[J]. 农业机械学报,2009,40(7):129~132.
Sun Xudong, Zhang Hailiang, Ouyang Aiguo, et al. Selection of NIR characteristic wavelength bands for soluble solids content in Nanfeng mandarin fruit[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(7):129~132. (in Chinese)
- 董文庚,邓晓丽,苗凤智,等. HPLC 法测定银杏叶中黄酮的含量[J]. 理化检验:化学分册,2005,41(8):563~565.
Dong Wengeng, Deng Xiaoli, Miao Fengzhi, et al. HPLC determination of flavone in ginkgo leaves[J]. Physical Testing and Chemical Analysis: Part B Chemical Analysis, 2005, 41(8):563~565. (in Chinese)
- 杨静红,丁晓燕,周扬,等. 紫外分光光度法测定银杏制品中黄酮的研究[J]. 中国卫生检验杂志,2011,21(2):337~338.
Yang Jinghong, Ding Xiaoyan, Zhou Yang, et al. Determination of flavonoids in ginkgo biloba product by direct ultraviolet spectrophotometry[J]. Chinese Journal of Health Laboratory Technology, 2011, 21(2):337~338. (in Chinese)
- 池玉梅,居聆,邓海山,等. 分光光度测定总黄酮法的适用性[J]. 分析化学,2010,38(6):893~896.
Chi Yumei, Ju Ling, Deng Haishan, et al. Suitability of spectrophotometry for determination of total flavonoids spectrophotometry[J]. Chinese Journal of Analytical Chemistry, 2010, 38(6):893~896. (in Chinese)
- 徐惠荣,陈晓伟,应义斌. 基于多元校正法的香梨糖度可见/近红外光谱检测[J]. 农业机械学报,2010,41(12):126~129,147.
Xu Huirong, Chen Xiaowei, Ying Yibin. Multivariate approach to determinate sugar content of fragrant pears with temperature variation by Visible/NIR spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010,41(12):126~129,147. (in Chinese)
- 代芬,洪添胜,岳学军,等. 砂糖橘可溶性总糖可见/近红外光谱无损检测[J]. 农业机械学报,2011,42(4):133~138.
Dai Fen, Hong Tiansheng, Yue Xuejun, et al. Nondestructive examination of sugar content of intact 'Shatangju' with visible-near Infrared spectroscopy based on variables selection[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(4):133~138. (in Chinese)
- 胡钢亮,吕秀阳,吴建国,等. 近红外漫反射光谱法直接测定银杏叶提取物粉末中总黄酮的含量[J]. 药物分析杂志,2004,24(1):25~27.
Hu Gangliang, Lü Xiuyang, Wu Jianguo, et al. Direct determination of total flavones in powder of ginkgo biloba extract by near-infrared diffuse reflection spectroscopy[J]. Chinese Journal of Pharmaceutical Analysis, 2004, 24(1):25~27. (in

Chinese)

- 10 黄倩倩,潘瑞乐,魏建和,等. 近红外漫反射光谱法测定黄芩中总黄酮及黄芩苷的含量[J]. 光谱学与光谱分析, 2009,29(9):123~126.
Huang Qianqian, Pang Ruile, Wei Jianhe, et al. Determination of baicalin and total flavonoids in radix scutellariae by near infrared diffuse reflectance spectroscopy[J]. Spectroscopy and Spectral Analysis, 2009, 29(9):123~126. (in Chinese)
- 11 杨玉兰,吴龙奇,王建华. 用近红外光谱检测银杏叶中的黄酮含量[J]. 河南科技大学学报:农学版,2003,32(2):7~9.
Yang Yulan, Wu Longqi, Wang Jianhua. Near-infrared spectroscopy technology on ginkgo leaves flavonoids[J]. Journal of Henan University of Science and Technology: Agricultural Science, 2003, 32(2):7~9. (in Chinese)
- 12 胡钢亮,吕秀阳,罗玲,等. 近红外光谱法同时测定银杏提取液中总黄酮和总内酯含量[J]. 分析化学,2004,32(8):83~85.
Hu Gangliang, Lü Xiuyang, Luo Ling, et al. Simultaneous determination of total flavones and total lactones in ginkgo extracts by near infrared spectroscopy[J]. Chinese Journal of Analytical Chemistry, 2004, 32(8):83~85. (in Chinese)
- 13 石吉勇,邹小波,赵杰文,等. 黄瓜叶片叶绿素含量近红外光谱无损检测[J]. 农业机械学报,2011,42(5):178~182,141.
Shi Jiyong, Zou Xiaobo, Zhao Jiewen, et al. NIR spectra in non-invasive measurement of cucumber leaf chlorophylls content [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(5):178~182,141. (in Chinese)
- 14 Chen Quansheng, Jiang Pei, Zhao Jiewen. Measurement of total flavone content in snow lotus (*Saussurea involucre*) using near infrared spectroscopy combined with interval PLS and genetic algorithm [J]. Spectrochimica Acta Part A—Molecular and Biomolecular Spectroscopy, 2010, 76(1):50~55.
- 15 Chen Quansheng, Zhao Jiewen, Liu Muhua, et al. Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms [J]. Determination of Total Polyphenols Content in Green Tea Using FT-NIR, 2008, 46(3):568~573.
- 16 Cho Soohwa, Kwon Kyunghee, Chung Hoeil. Varied performance of PLS calibration using different overtone and combination bands in a near-infrared region [J]. Chemometrics and Intelligent Laboratory Systems, 2006, 82(1~2):104~108.

(上接第 154 页)

- 7 Rao Yulan, Xiang Bingren, Zhou Xiaohua, et al. Quantitative and qualitative determination of acid value of peanut oil using near-infrared spectrometry [J]. Journal of Food Engineering, 2009, 93(2):249~252.
- 8 Gülgün M D, Cabo N. Fourier transform infrared spectra data versus peroxide and anisidine values to determine oxidative stability of edible oils [J]. Food Chemistry, 2002, 77(4):503~510.
- 9 Armenta S, Garrigues S, Guardia de la M. Determination of edible oil parameters by near infrared spectrometry [J]. Analytica Chimica Acta, 2007, 596(2):330~337.
- 10 Sato T. New estimation method for fatty acid composition in oil using near infrared spectroscopy [J]. Bioscience Biotechnology and Biochemistry, 2002, 66(12):2543~2548
- 11 西北农林科技大学. 一种连续式可变光程的近红外检测流通池:中国,ZL201020110561.4[P]. 2010-11-03.
- 12 廖宜涛,樊玉霞,伍学千,等. 猪肉肌内脂肪含量的可见/近红外光谱在线检测[J]. 农业机械学报,2010,41(9):104~107.
Liao Yitao, Fan Yuxia, Wu Xueqian, et al. On-line prediction of intramuscular fat content in pork muscle with visible/near-infrared spectroscopy [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(9):104~107. (in Chinese)
- 13 Yu Xiuzhu, van de Voort F R, Sedman J. Determination of peroxide value of edible oils by FTIR spectroscopy with the use of the spectral reconstitution technique [J]. Talanta, 74(2):241~246.
- 14 陆婉珍,袁洪福,褚小立. 近红外光谱仪器[M]. 北京:化学工业出版社,2010:47~49.
- 15 杰尔·沃克曼,洛伊斯·文依. 近红外光谱解析实用指南[M]. 褚小立,许育鹏,田高友,译. 北京:化学工业出版社,2009:49~67.