

基于 CARS - PLS 的食用油脂肪酸近红外定量分析模型优化*

吴静珠¹ 徐云²

(1. 北京工商大学计算机与信息工程学院, 北京 100048; 2. 中国农业大学信息与电气工程学院, 北京 100083)

【摘要】 采用 CARS 波长变量挑选方法优化建模, 对食用油中 4 种主要脂肪酸(棕榈酸、硬脂酸、油酸和亚油酸)进行近红外定量分析。应用预测浓度残差法剔除奇异样本后, 对样品集光谱进行标准化预处理, 通过 CARS 优选出的波长变量分别建立 4 种脂肪酸的偏最小二乘法(PLS)模型。与采用 OPUS 软件自动优化建模相比, CARS 法所建模型的决定系数(R^2)、交叉校验均方根误差(RMSECV)和预测均方根误差(RMSEP)都优于后者所建模型。CARS 法有效地简化了模型, 且所挑选出的特征波长较少。

关键词: 食用油 脂肪酸 近红外光谱 CARS 变量筛选 偏最小二乘法

中图分类号: O657.33; O623.61; TS225 **文献标识码:** A **文章编号:** 1000-1298(2011)10-0162-05

NIR Quantitative Model Optimization of Fatty Acid in Edible Oil Based on CARS - PLS

Wu Jingzhu¹ Xu Yun²

(1. School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

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

Abstract

Competitive adaptive reweighted sampling (CARS) method was employed to improve the prediction accuracy of the NIR quantitative model of four kinds of fatty acid (palmitic acid, stearic acid, oleic acid and linoleic acid) in edible oil. Predict concentration residual method was employed to detect the outlier before preprocessing the spectroscopy by normalization. The key variables were selected by CARS method. The partial least squares (PLS) calibration models of four kinds of fatty acid were established respectively in the optimal conditions, and compared with the results using OPUS software. Determination coefficient (R^2), root mean square error of cross validation (RMSECV) and root mean square error of prediction (RMSEP) were used to evaluate the quality of the modes. The results showed that better prediction was obtained by CARS. The result showed that using CARS could effectively simplify the model and the less number of wavelength variables selected could be reference for developing filter spectrometer of edible oil.

Key words Edible oil, Fatty acid, Near-infrared spectroscopy, Competitive adaptive reweighted sampling, Partial least squares

引言

我国居民食用植物油主要有大豆油、菜籽油、花生油、葵花籽油、橄榄油和芝麻油等。植物油中常见的高级脂肪酸主要有棕榈酸、硬脂酸、油酸和亚油酸

等。不同植物油的脂肪酸组成有很大差异, 因此其功效和营养价值也随之不同。随着营养学研究发展, 人们对植物油中脂肪酸认识不断提高。中国营养学会也推荐膳食中必需脂肪酸摄入理想比值, 市场上也出现大量调和油, 为了评估食用植物油营养

价值,研究分析常用食用植物油中脂肪酸组成是很有必要的。

近红外分析技术是利用近红外谱区包含的物质信息,用于有机物质定性和定量分析的一种分析技术。该方法具有快速、高效、无污染、无需前处理、无损分析及多组分同时测定等优点^[1-2]。国内外已有将近红外技术应用于油脂检测的研究报道^[3-5]。提高模型预测能力始终是近红外分析中的研究热点和难点问题。有研究报道,有效地变量筛选不仅可以简化模型,更主要的是可以剔除不相关变量,建立更好的校正模型,从而加强模型预测能力与稳健性^[6]。目前,在近红外分析中,常用的波长变量选择方法主要有相关系数法、方差分析法、无信息变量消除法、遗传算法、移动窗口偏最小二乘回归法等^[6-7]。CARS (competitive adaptive reweighted sampling) 是近年来提出的一种新的变量选择方法^[8]。此算法能够在一定程度上克服变量选择中的组合爆炸问题,筛选出优化的变量子集,提高模型的预测能力和降低预测方差^[9]。

本文拟采用红外光谱技术,建立食用油中 4 种主要脂肪酸(棕榈酸、硬脂酸、油酸和亚油酸)定量分析模型,通过 CARS 挑选特征波长变量优化模型,提升脂肪酸近红外模型的预测能力。

1 CARS 变量筛选法

近红外光谱通常由大量数据点构成,建模时波长点数远大于样本数,且光谱的共线性非常严重。近红外光谱本身有效信息量很弱,有些波段与所测样品成分之间缺乏相关关系。各谱区的噪声水平也不一致,即使作相关的谱图预处理也可能导致模型不稳定,降低模型的精度。因此,波长变量筛选是很有必要的预处理手段,不但可以简化模型,并能有效提高模型的预测能力。

CARS 方法采用“适者生存”原则,将每个变量看成一个个体,通过自适应加权采样技术筛选出 PLS 模型中回归系数绝对值较大的波长,去掉权重较小的波长,并通过交互校验优选出模型交互校验均方根误差最小时所对应的波长组合^[9]。此算法中引入了指数衰减函数来控制变量的保留率,具有很高的计算效率,适用于高维数据的变量选择。

文中所用的 CARS 和 PLS 等程序均在 Matlab R2008a 环境下运行。

2 试验

2.1 试验材料

60 个植物油样品,包括花生油、大豆油、橄榄

油、芝麻油、葵花籽油、玉米胚芽油和芥花油等,均购自当地超市。

2.2 食用油脂脂肪酸测定

采用气相色谱法测定 60 个食用油样品中 4 种脂肪酸(棕榈酸、硬脂酸、油酸和亚油酸)的含量。

2.3 近红外光谱采集

采用德国 Bruker 公司 VERTEX 70 型红外光谱仪采集样品近红外光谱,使用液体光纤探头采样方式,光程 2 mm。仪器参数设定如下:波数范围 4 000 ~ 12 500 cm^{-1} ,分辨率 16 cm^{-1} ,每个样品重复扫描 32 次、采样点 1 102。

全部样品未经任何化学处理,将光纤探头插入装有样品的小瓶中逐一扫描样品,每次测量前均用石油醚清洗探头,避免样品间交叉污染。样品近红外光谱如图 1 所示。

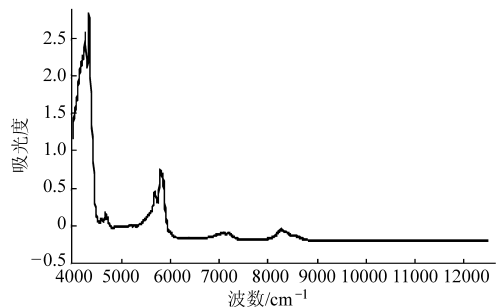


图 1 样品集近红外光谱图

Fig. 1 Near-infrared spectroscopy of samples

3 结果与分析

建立性能较好的近红外模型,进行光谱数据预处理是非常必要的。数据预处理步骤主要包括:异常样品的剔除、校正集和校验集的划分、消除光谱噪声和其干扰的影响、优化光谱范围、净化谱图信息等^[10]。

近红外校正模型的预测精度和稳健性采用主成分分数 n 、决定系数 R^2 、交叉校验标准差 (RMSECV) 和预测均方根误差 (RMSEP) 指标评价。

3.1 样品集划分

根据预测浓度残差法^[11]分别剔除 4 种脂肪酸样品中的 3 个脂肪酸异常样品后,样品集共有 57 个样品组成,样品信息如表 1 所示。校正集样品的选取直接影响所建模型的适用性和准确性。为了充分考虑校正样品集的浓度代表性,采用含量梯度法以校正集和校验集样品个数的比例为 3:1 进行划分,校正集样品 43 个,校验集样品 14 个^[12]。

3.2 软件优化建模

采用 BRUKER 公司的 OPUS 6.5 软件自动优化建模功能,采用常规的光谱预处理方法及其组合进

表1 样品集统计信息

Tab.1 Statistical information of samples

脂肪酸	最小值/%	最大值/%	标准偏差
棕榈酸	5.6	16.8	2.478
硬脂酸	1.8	5.0	0.747
油酸	22.7	77.9	14.461
亚油酸	8.6	65.9	14.848

表2 软件自动优化建模及预测结果

Tab.2 Result of calibration and validation by software OPUS

脂肪酸	预处理	n	R^2	RMSECV	RMSEP	保留波长数
棕榈酸	一阶导(窗口点数:17 pts.) + 矢量归一化, 12 489.4 ~ 7 498.3 cm^{-1} , 6 102.0 ~ 4 597.7 cm^{-1}	10	0.977	0.383	0.319	844
硬脂酸	常数偏移量消除, 7 506.0 ~ 4 597.7 cm^{-1}	13	0.961	0.150	0.143	378
油酸	常数偏移量消除, 7 506.0 ~ 5 446.3 cm^{-1}	12	0.991	1.270	0.966	268
亚油酸	最小最大归一化, 7 506.0 ~ 5 446.3 cm^{-1}	11	0.992	1.370	1.46	268

图2a为筛选过程中选出变量的变化趋势,变量数随着运行次数增加由快到慢递减。运行次数和保留的变量数之间为指数关系^[9]。

图2b为经过5折交互检验所得残差RMSECV的变化趋势图。在1~32次间残差呈现递减趋势,表明筛选过程剔除了与样本性质无关的变量,32次后开始递增则可能剔除了关键变量,从而导致残差增大。

图2c中“*”所对应点即为RMSECV最低点,图中各线表示随着运行次数增加各变量回归系数的变化趋势。

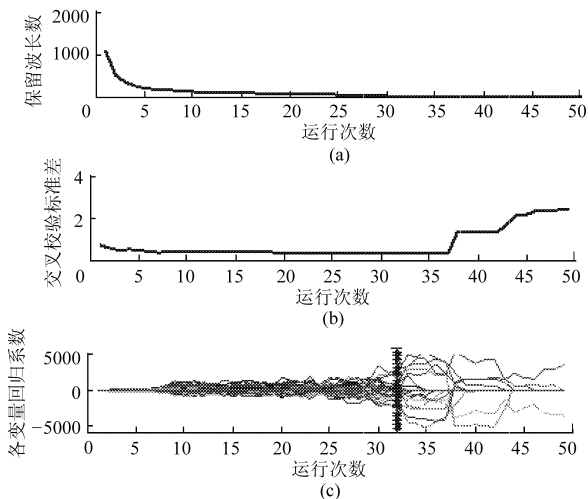


图2 棕榈酸光谱波长变量筛选图

Fig.2 Plots of wavelength variable selection on palmitic acid spectrum

(a) 保留波长数 (b) 交叉校验标准差
(c) 各波长变量回归系数

行消噪,采用软件中设定的用于近红外 NIR 的 5 个典型的谱段范围来优化谱区范围,采用偏最小二乘法 PLS 建立最优校正模型,结果如表 2 所示。

3.3 CARS-PLS 模型优化

对校正集光谱进行标准化处理后,采用 CARS 方法分别对 4 种脂肪酸波长变量进行多次筛选,以棕榈酸为例说明变量筛选过程,其筛选过程如图 2 所示。

集校验,结果如表 3 所示。由表 3 可知,经波长筛选模型质量明显提高,优选波长后建模得到的 4 种脂肪酸模型的 RMSECV 值分别是 OPUS 自动优化建模 RMSECV 值的 79.3%、80.7%、58.3%、62.7%。根据 RMSEP 指标,CARS-PLS 模型的预测能力明显优于软件自动优化所建模型。

表3 CARS-PLS 优化建模及预测结果

Tab.3 Results of calibration and validation by CARS-PLS

脂肪酸	n	R^2	RMSECV	RMSEP	保留波长数
棕榈酸	12	0.986	0.304	0.248	23
硬脂酸	12	0.975	0.121	0.110	23
油酸	13	0.997	0.740	0.868	23
亚油酸	12	0.997	0.859	0.833	20

4 种脂肪酸的分子结构中都含有亚甲基 CH_2 和羧酸 COOH ,这类含 H 基团在近红外区有比较明显的吸收峰。其中,线性脂肪族分子的亚甲基在一级倍频区的主要吸收峰约在 $5\ 800\ \text{cm}^{-1}$ 和 $5\ 680\ \text{cm}^{-1}$ 处^[13]。羧酸中的羟基 OH 伸缩振动和 $\text{C}=\text{O}$ 伸缩振动的组合频在 $5\ 290\ \text{cm}^{-1}$ 也有吸收峰。CARS 方法挑选的 4 种脂肪酸的特征波长变量如图 3 所示。从图中可以看出,CARS 法对 4 种脂肪酸挑选出的特征波长在 $5\ 000\sim 6\ 000\ \text{cm}^{-1}$ 中有公共区域,而亚甲基 CH_2 和羧酸 COOH 正是在该区域有主要吸收峰。另外,与表 2 中保留的波长数相比,CARS 法挑选的波长个数显著降低。

4 结束语

采用 CARS 分别挑选出食用油中 4 种主要脂肪

利用所选波长建立 PLS 回归模型,并通过校验

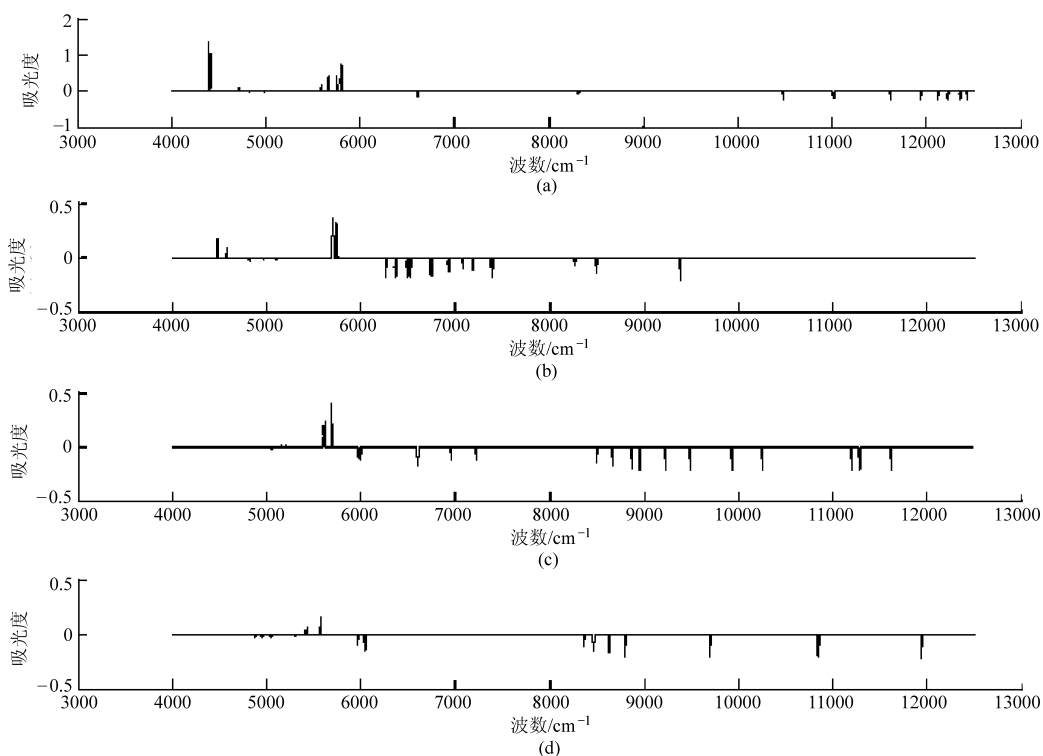


图 3 CARS 挑选的食用油脂脂肪酸近红外光谱的特征波长

Fig. 3 Wavelength variable selection on fatty acid NIR spectrum by CARS

(a) 棕榈酸 (b) 硬脂酸 (c) 油酸 (d) 亚油酸

酸(棕榈酸、硬脂酸、油酸和亚油酸)特征波长变量, 试验结果表明 CARS 法不但能有效地减少特征波长个数, 而且挑选出的波长与物质在近红外区的特征吸收峰相近, 能真实地反映出物质所含基团, 可为滤

光片式食用油近红外分析仪的波长选取作参考。试验中基于 CARS-PLS 的食用油脂脂肪酸近红外定量分析模型的优化, 大大提升脂肪酸近红外模型的预测能力。

参 考 文 献

- 1 严衍禄, 赵龙莲, 韩东海, 等. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005.
- 2 陆婉珍. 现代近红外光谱分析技术[M]. 北京: 中国石化出版社, 2007.
- 3 毕艳兰, 鲍丹青, 田原, 等. 利用傅里叶近红外技术快速测定食用植物油的过氧化值[J]. 中国油脂, 2009, 34(3): 71~74.
Bi Yanlan, Bao Danqing, Tian Yuan, et al. Rapid determination of peroxide value of edible vegetable oils by near-infrared spectroscopy[J]. China Oils and Fats, 2009, 34(3): 71~74. (in Chinese)
- 4 吴静珠, 刘翠玲, 李慧, 等. 近红外光谱技术在食用油种类鉴别及脂肪酸含量检测中的应用[J]. 北京工商大学学报: 自然科学版, 2010, 28(5): 56~59.
Wu Jingzhu, Liu Cuiling, Li Hui, et al. Application of NIR technology on identifying types and determining main fatty acid content of edible vegetable oil[J]. Journal of Beijing Technology and Business University: Natural Science Edition, 2010, 28(5): 56~59. (in Chinese)
- 5 于燕波, 臧鹏, 付元华, 等. 近红外光谱法快速测定植物油中脂肪酸含量[J]. 光谱学与光谱分析, 2008, 28(7): 1554~1558.
Yu Yanbo, Zang Peng, Fu Yuanhua, et al. The rapid analysis of fatty acids in vegetable oils by near infrared spectrum[J]. Spectroscopy and Spectral Analysis, 2008, 28(7): 1554~1558. (in Chinese)
- 6 张巧杰. 直链淀粉检测方法与技术研究[D]. 北京: 中国农业大学, 2005.
Zhang Qiaojie. Study on detection methods and technology of amylose content[D]. Beijing: China Agricultural University, 2005. (in Chinese)
- 7 Jiang J H, Berry R J, Siesler H W, et al. Wavelength interval selection in multi component spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data[J]. Analytical Chemistry, 2002, 74(14): 3555~3565.
- 8 Li Hongdong, Liang Yizeng, Xu Qingsong, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica Chimica Acta, 2009, 648(1): 77~84.

- 9 张华秀,李晓宁,范伟,等. 近红外光谱结合 CARS 变量筛选方法用于液态奶中蛋白质与脂肪含量的测定[J]. 分析测试学报,2010,29(5):430~434.
Zhang Huaxiu, Li Xiaoning, Fan Wei, et al. Determination of protein and fat in liquid milk by NIR combined with CARS variables screening method[J]. Journal of Instrumental Analysis, 2010, 29(5): 430~434. (in Chinese)
- 10 吴静珠,王一鸣,张小超,等. 基于近红外光谱的奶粉品质检测技术研究[J]. 光谱学与光谱分析,2007,27(9):1735~1738.
Wu Jingzhu, Wang Yiming, Zhang Xiaochao, et al. Study on quality detection of milk powder based on near infrared spectroscopy (NIR)[J]. Spectroscopy and Spectral Analysis, 2007, 27(9): 1735~1738. (in Chinese)
- 11 祝诗平,王一鸣,张小超,等. 近红外光谱建模异常样品剔除准则与方法[J]. 农业机械学报,2004,35(4):115~119.
Zhu Shiping, Wang Yiming, Zhang Xiaochao, et al. Outlier sample eliminating criterions and methods for building calibration model of near infrared spectroscopy analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2004, 35(4): 115~119. (in Chinese)
- 12 吴静珠,王一鸣,张小超,等. 近红外光谱分析中定标集样品挑选方法研究[J]. 农业机械学报,2006,37(4):80~82.
Wu Jingzhu, Wang Yiming, Zhang Xiaochao, et al. Study on algorithms of selection of representative samples for calibration in near infrared spectroscopy analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2006, 37(4): 80~82. (in Chinese)
- 13 Jerry Workman, Jr Lois Weyer. 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.
- 14 于修焯,杜双奎,岳田利,等. 食用油反式脂肪酸傅里叶红外光谱重组检测技术[J]. 农业机械学报,2009,40(1):114~119.
Yu Xiuzhu, Du Shuangkui, Yue Tianli, et al. FTIR measuring technique of trans fatty acids in edible oil using spectral reconstitution[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(1): 114~119. (in Chinese)

(上接第 185 页)

- 12 高大文,彭永臻,王淑莹. 控制 pH 实现短程硝化反硝化生物脱氮技术[J]. 哈尔滨工业大学学报,2005,37(7):1664~1666.
Gao Dawen, Peng Yongzhen, Wang Shuying. Nitrogen removal from wastewater via shortcut nitrification-denitrification achieved by controlling pH [J]. Journal of Harbin Institute of Technology, 2005, 37(7): 1664~1666. (in Chinese)
- 13 Suthersan S, Ganczarcezyk J. Inhibition of nitrite oxidation during nitrification: some observations [J]. Water Pollut. Res. J. Can., 1986, 21(2): 257~262.
- 14 李伟光,闫立龙,李娟,等. SBR 工艺处理高质量浓度尿素废水中短程硝化的特性研究[J]. 哈尔滨工业大学学报,2009,41(2):61~65.
Li Weiguang, Yan Lilong, Li Juan, et al. Short-cut nitrification characteristics in treating wastewater with high-concentration urea with SBR [J]. Journal of Harbin Institute of Technology, 2009, 41(2): 61~65. (in Chinese)
- 15 孙振世,柯强,陈英旭. SBR 生物脱氮机理及其影响因素[J]. 中国沼气,2001,19(2):16~19.
Sun Zhenshi, Ke Qiang, Chen Yingxu. Mechanism of biodenitrification of SBR and its effecting factors [J]. China Biogas, 2001, 19(2): 16~19. (in Chinese)
- 16 Vadivelu V M, Yuan Z, Fux C, et al. Stoichiometric and kinetic characterization of nitrobacter in mixed culture by decoupling the growth and energy generation processes [J]. Biotechnol Bioeng, 2006, 94(6): 1176~1188.