

食用植物油中反式脂肪酸含量的激光拉曼光谱检测

蒋雪松¹, 莫欣欣³, 孙通^{2,3*}, 胡栋²

1. 南京林业大学机械电子工程学院, 江苏 南京 210037
2. 浙江农林大学工程学院, 浙江 临安 311300
3. 江西农业大学工学院, 江西 南昌 330045

摘要 油脂中的反式脂肪酸(TFA)有害人们的身体健康,有必要对其含量进行监测。共收集各类食用植物油样本79个,涉及9个品种和27个品牌,分配到校正集和预测集的样本数分别为53个和26个。采用QE65000拉曼光谱仪采集79个样本的拉曼光谱,利用自适应迭代惩罚最小二乘法去除样本拉曼光谱的荧光背景;在此基础上,采用多种归一化方法对样本拉曼光谱进行处理,并对拉曼光谱的建模波数范围进行初选;再利用竞争性自适应重加权采样(CARS)方法筛选与食用植物油TFA含量相关的光谱变量,并应用偏最小二乘(PLS)回归将食用植物油TFA的特征变量光谱强度与气相色谱测定的TFA真实含量进行关联,建立食用植物油中TFA含量的定量预测模型。研究表明,多种归一化方法中,有4种归一化方法均能提高PLS定量预测模型的性能,其中Area normalization方法的效果最优;经建模波数范围初选,波数范围由 $686\sim 2301\text{ cm}^{-1}$ 缩减为 $737\sim 1787\text{ cm}^{-1}$,确定较优的建模波数范围为 $737\sim 1787\text{ cm}^{-1}$;经CARS方法筛选,共有31个光谱变量被选择,其选择的光谱变量主要分布在 $1265, 1303, 1442$ 及 1658 cm^{-1} 拉曼振动峰附近,且 974 cm^{-1} 拉曼振动峰两侧均有光谱变量被选择;此外,CARS方法的PLS建模结果优于常用的无信息变量消除及连续投影算法。由此可知,激光拉曼光谱技术结合化学计量学方法检测食用植物油中的TFA含量是可行的。归一化方法、建模波数范围初选及竞争性自适应重加权采样(CARS)方法能有效提高TFA定量预测模型的预测精度和稳定性,优化后的TFA定量预测模型的校正集及预测集的相关系数和均方根误差分别为0.949, 0.953和0.188%, 0.191%。与未优化的预测模型相比,预测均方根误差由0.361%下降为0.191%,下降幅度为47.1%;建模所用的变量数由683个下降为31个,仅占原变量数的4.54%。

关键词 拉曼光谱;反式脂肪酸;食用植物油;竞争性自适应重加权采样

中图分类号: TS227 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)12-3821-05

引言

反式脂肪酸(trans fatty acid, TFA)是一类不饱和脂肪酸,食用油脂的氢化加工、食用植物油的精炼、脱臭工艺以及高温烹调均会产生反式脂肪酸。食用植物油中的反式脂肪酸是一种不好的脂肪酸,食用过量会对人体健康产生诸多危害,如影响儿童生长发育,引起动脉硬化,诱发糖尿病、心血管疾病及神经系统疾病,增加前列腺癌、乳腺癌、结肠癌等癌症风险^[1]。而食用植物油是人们生活的必需品,因此非

常有必要对食用植物油中的反式脂肪酸含量进行监测。

激光拉曼光谱技术是近年来发展起来的现代光谱分析技术,具有分析速度快、非破坏性检测、无需样品预处理、不污染环境以及成本低等优点。目前,在农产品/食品安全检测领域,激光拉曼光谱分析技术已应用于辣椒苏丹红^[2-3]、鸡尾酒防腐剂^[4]、玉米真菌毒素^[5]、奶粉三聚氰胺^[6]、果蔬农药残留^[7-8]、牛奶抗生素^[9]及猪肉瘦肉精^[10]等的定量检测。对于反式脂肪酸的拉曼光谱检测,Stefanov等^[11]对牛奶中反式脂肪酸进行拉曼分析,确定了单不饱和脂肪酸和共轭亚油酸的C=C双键特征拉曼峰分别为 1674 和 1653

收稿日期: 2019-08-24, 修订日期: 2019-10-29

基金项目: 国家自然科学基金项目(31401278), 江西省自然科学基金项目(20181BAB204011), 浙江农林大学科研发展基金项目(2019FR033)资助

作者简介: 蒋雪松, 1979年生, 南京林业大学机械电子工程学院副教授 e-mail: xsjiang@126.com

* 通讯联系人 e-mail: suntong980@163.com

cm^{-1} 。Zhao 等^[12]采用拉曼光谱技术对乳制品中总反式脂肪酸含量进行检测。对于黄油,其总反式脂肪酸验证集模型的决定系数为 0.56~0.78。Numata 等^[13]利用拉曼光谱技术对混合溶液中的油酸和反油酸含量进行检测。总结上述文献发现,反式脂肪酸的拉曼光谱检测研究非常稀少,仅为初步的特征峰判别和定量分析,需要进一步深入研究。此外,尚未有食用植物油中反式脂肪酸的拉曼光谱检测研究报告。

研究利用激光拉曼光谱技术对食用植物油中的反式脂肪酸含量进行快速定量检测研究,采用归一化方法对样本光谱进行预处理,并对光谱建模波数范围进行初选及(competitive adaptive reweighted sampling, CARS)变量筛选,最后应用偏最小二乘(partial least squares, PLS)回归方法建立食用植物油中反式脂肪酸含量的定量预测模型。

1 实验部分

1.1 样本与试剂

试验所用的食用植物油样本共有 79 个,均从各正规大超市购买。品种包括大豆油、菜籽油、玉米油、花生油及调和油等,含有福临门、胡姬花、香满园、金龙鱼、鲁花、游天下及道道全等 27 个品牌。为保证校正集和预测集样本分配具有代表性及合理性,首先将 79 个样本按照 TFA 含量高低进行排序,再按相应比例随机分配到校正集和预测集。校正集及预测集各分配到 53 个和 26 个样本。

试验所用的主要试剂:顺式-9-十八烯酸甲酯(46902-U)、反式-9-十八烯酸甲酯(46903)、4 种亚油酸甲酯顺反异构体(47791)及 8 种亚麻酸甲酯顺反异构体(47792)标准品,均购买于美国 Sigma 公司;异辛烷(色谱纯),氢氧化钾-甲醇溶液($2 \text{ mol} \cdot \text{L}^{-1}$),硫酸氢钠(分析纯)。

1.2 仪器及光谱采集

试验所用的便携式拉曼光谱检测系统主要包括 QE65000 拉曼光谱仪(Ocean Optics Co., USA)、激光器、光纤及样品池等。激光器波长为 785 nm,功率为 460 mW。GC-2010 气相色谱仪(Shimadzu Co., Japan)配有 FID 检测器和 GC SOLUTION 数据处理工作站。色谱柱为 HP88 毛细管色谱柱($100 \text{ m} \times 0.25 \text{ mm} \times 0.25 \mu\text{m}$)。

采用移液枪将 79 个食用植物油样本依次加入 2 mL 石英进样瓶中,再将进样瓶置于样品池中进行拉曼光谱采集。光谱采集软件为 Ocean Optics Spectrasuite (Ocean Optics Co., USA)。光谱积分时间为 10 s,平均次数为 2 次。

1.3 TFA 含量测定

对于 79 个食用植物油样本,参照国家标准方法 GB 5009.257—2016 测定其 TFA 含量。每个食用植物油样本均测定两次,取测定结果的平均值作为样本 TFA 含量。气相色谱的检测条件如下:进样口及检测器温度分别设定为 240 和 300 $^{\circ}\text{C}$;色谱柱采用程序升温,初始温度为 120 $^{\circ}\text{C}$,先以 5 $^{\circ}\text{C} \cdot \text{min}^{-1}$ 升温至 170 $^{\circ}\text{C}$,保持 10 min,再以 2 $^{\circ}\text{C} \cdot \text{min}^{-1}$ 升温至 185 $^{\circ}\text{C}$,保持 8 min,然后以 3 $^{\circ}\text{C} \cdot \text{min}^{-1}$ 升温至 240 $^{\circ}\text{C}$,保持 10 min;载气为氮气,其流速及压力分别为 1.99 $\text{mL} \cdot \text{min}^{-1}$ 和 360.4 KPa;氢气及空气流速分别为 50 和 500 $\text{mL} \cdot \text{min}^{-1}$;

分流比为 20:1。

1.4 数据处理

首先,采用自适应迭代惩罚最小二乘法对样本拉曼光谱进行预处理,以扣除光谱的荧光背景。在此基础上,利用多种归一化方法(area normalization, mean normalization, maximum normalization, unit vector normalization, range normalization)对拉曼光谱进行处理,以消除外界因素及激光能量波动等的影响。

然后,利用竞争性自适应重加权采样(competitive adaptive reweighted sampling, CARS)方法筛选变量,并与连续投影算法(successive projections algorithm, SPA)及无信息变量消除(uninformative variable elimination, UVE)进行比较。本研究中,由蒙特卡罗交叉验证决定 CARS 方法的最大因子数,采样次数为 50 次,并根据交互验证均方根误差(root mean square error of cross validation, RMSECV)最小确定最优变量子集。CARS 方法的原理参见文献[14]。

最后,应用 PLS 方法建立食用植物油中 TFA 含量的定量预测模型,并利用预测集样本对模型进行验证。

2 结果与讨论

2.1 TFA 含量统计分析

表 1 为校正集、预测集及所有样本的 TFA 含量平均值、范围及标准偏差。样本最高及最低的反式脂肪酸含量分别为 2.662% 和 0.014%。此外, TFA 含量在 0~0.5%, 0.5~1%, 1~1.5%, 1.5~2% 之间的样本数分别为 30 个、29 个、11 个和 6 个, TFA 含量在 2% 以上的样本数为 3 个。

表 1 校正集、预测集及所有样本的反式脂肪酸含量平均值、范围及标准偏差

Table 1 Means, ranges, and SDs of TFA contents for calibration, prediction and all samples

参数	数据集	样本数	平均值	标准偏差	范围
TFA/%	Total	79	0.758	0.581	0.014~2.662
	Calibration	53	0.757	0.592	0.014~2.662
	Prediction	26	0.758	0.568	0.016~2.280

2.2 光谱分析及预处理

图 1(a)为所有样本的原始拉曼光谱。由图 1(a)可知,原始光谱在 1 200~1 800 cm^{-1} 波数范围存在较为明显的谱峰;在 400~1 200 cm^{-1} 波数范围,由于不同品种的样本颜色深度有差异,其光谱存在较强的荧光背景及基线漂移,导致谱峰被掩盖,需要对光谱进行预处理。图 1(b)为扣除荧光背景后的样本拉曼光谱。由图 1(b)可知,样本光谱的荧光背景及基线漂移得到有效消除,各拉曼谱峰更加清晰。其中,974 cm^{-1} 为反式双键的拉曼振动峰,1 265 及 1 658 cm^{-1} 为顺式双键的拉曼振动峰,1 303 及 1 442 cm^{-1} 分别为亚甲基和甲基的拉曼振动峰^[15]。上述拉曼振动峰对 TFA 定量检测可能均为有用的光谱信息,需要包含在建模波数范围内。此外,在 279~685 cm^{-1} 波数范围,由于受荧光背景及基线漂移影

响最大,其光谱信噪比较低,不纳入建模波数范围。因此,后续的光谱数据分析在 $686\sim 2\,301\text{ cm}^{-1}$ 波数范围内进行,该范围共有 683 个变量。

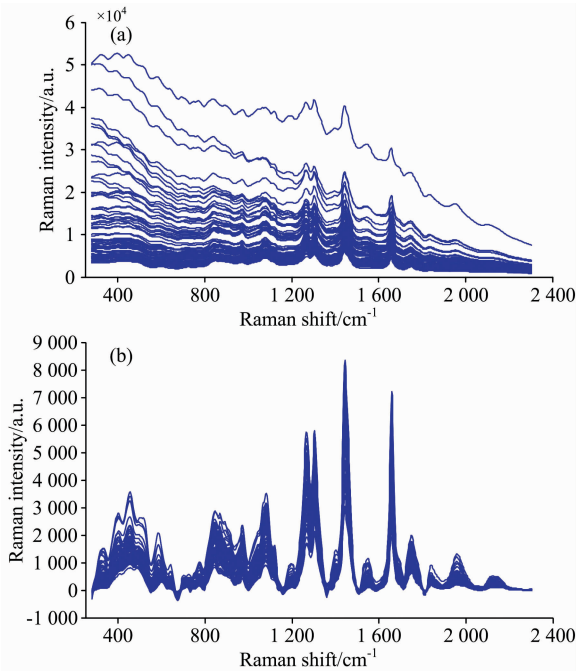


图 1 食用植物油样本的拉曼光谱

(a): 原始光谱; (b): 荧光背景扣除后的光谱

Fig. 1 Raman spectra of edible vegetable oil samples

(a): Original spectra;

(b): Spectra after fluorescence background subtraction

在扣除荧光背景的基础上,采用多种归一化方法对样本拉曼光谱进行处理,以确定较优的归一化处理方法。表 2 为不同归一化预处理方法下样本 TFA 含量的 PLS 建模结果。由表 2 可知,除 Range normalization 方法外,其他 4 种归一化方法均能在一定程度上提高 TFA 定量预测模型的性能。5 种归一化方法中,经 Area normalization 预处理后所建立的 TFA 定量预测模型性能最优,其校正集和预测集的相关系数及均方根误差分别为 0.956, 0.905 和 0.176%, 0.276%。

表 2 不同归一化预处理方法下反式脂肪酸的 PLS 建模结果

Table 2 PLS results of TFA at different normalization pretreatment methods

归一化方法	因子数	校正集		预测集	
		r	校正均方根误差/%	r	预测均方根误差/%
None	9	0.868	0.296	0.816	0.361
Area	14	0.956	0.176	0.905	0.276
Maximum	14	0.948	0.189	0.893	0.303
Mean	13	0.950	0.185	0.895	0.295
Range	12	0.930	0.219	0.797	0.389
Unit vector	12	0.934	0.212	0.832	0.350

2.3 光谱波段初选

在 Area normalization 预处理的基础上,进一步对样本拉曼光谱的建模波数范围进行初步选择。表 3 为不同波数范围下样本 TFA 含量的 PLS 建模结果。由表 3 可知,当波数范围由 $686\sim 2\,301\text{ cm}^{-1}$ 缩减为 $686\sim 1\,882\text{ cm}^{-1}$ 时, TFA 含量的校正集和预测集相关系数均有上升,而校正均方根误差 (root mean square error of calibration, RMSEC) 及预测均方根误差 (root mean square error of prediction, RMSEP) 均下降, TFA 定量预测模型性能提高,表明缩减的波数范围为无用光谱信号,对 TFA 定量预测模型有干扰作用。波数范围由 $686\sim 2\,301\text{ cm}^{-1}$ 一直缩减为 $737\sim 1\,787\text{ cm}^{-1}$ 时, TFA 定量预测模型的性能均有不同程度地提高,而波数范围由 $737\sim 1\,787\text{ cm}^{-1}$ 缩减为 $801\sim 1\,787\text{ cm}^{-1}$ 时, TFA 定量预测模型的性能下降,表明缩减的波数范围含有对 TFA 有用的光谱信号。因此,确定 $737\sim 1\,787\text{ cm}^{-1}$ 波数范围为较优的建模波数范围,该波数范围共有 349 个变量。

表 3 不同波数范围下反式脂肪酸的 PLS 建模结果

Table 3 PLS results of TFA at different wavenumber ranges

波数范围 / cm^{-1}	因子数	校正集		预测集	
		r	校正均方根误差/%	r	预测均方根误差/%
$686\sim 2\,301$	14	0.956	0.176	0.905	0.276
$686\sim 1\,882$	12	0.940	0.203	0.925	0.241
$737\sim 1\,882$	11	0.932	0.217	0.927	0.238
$737\sim 1\,787$	11	0.932	0.216	0.934	0.232
$801\sim 1\,787$	10	0.927	0.223	0.929	0.237

2.4 CARS 变量选择

在 $737\sim 1\,787\text{ cm}^{-1}$ 波数范围内,利用 CARS 变量选择方法进一步筛选与 TFA 相关的重要变量。图 2 为 TFA 的 CARS 变量选择结果。由图 2(a)可知,采样次数由 1 增加到 5 时,被选择的变量数迅速下降,而后随着采样次数的增加,

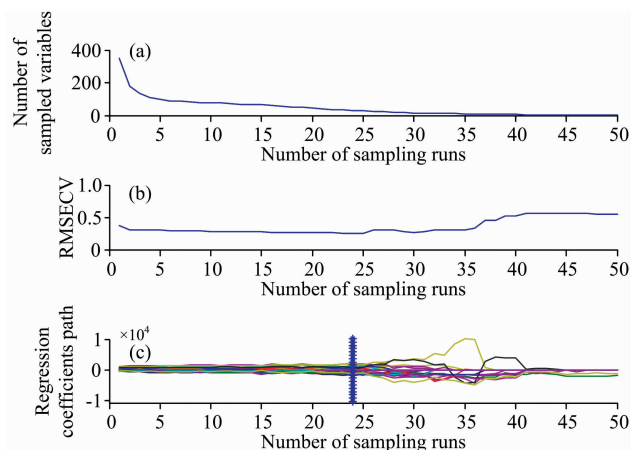


图 2 反式脂肪酸的 CRAS 变量选择结果

(a): 被选择的变量数; (b): RMSECV; (c): 回归系数变化趋势

Fig. 2 Results of CRAS variable selection of TFA

(a): Number of selected variables; (b): RMSECV;

(c): Trends of regression coefficients

被选择的变量数缓慢下降,存在变量粗选和精选两个过程。由图 2(b)可知,采样次数由 1 增加到 2 时, RMSECV 有较大下降,而采样次数由 2 增加到 24 时, RMSECV 逐渐缓慢下降,并达到最小值,表明 CARS 方法正在逐步剔除与 TFA 含量无关的变量;而后,随着采样次数的再增加, RMSECV 上下波动,但总体呈上升趋势,表明有与 TFA 含量相关的有效变量被剔除。图 2(c)为被选择变量的回归系数变化趋势,“*”对应的位置为 24 次采样,此时 RMSECV 达到最小值。根据 RMSECV 最小原则,最终共选中 31 个变量。图 3 为由 CARS 方法选择的变量的分布情况。由图 3 可知,被选择的变量主要分布在 1 265, 1 303, 1 442 及 1 658 cm^{-1} 拉曼振动峰附近, 974 cm^{-1} 拉曼振动峰虽未被选中,但其谱峰两侧均有变量被选中。

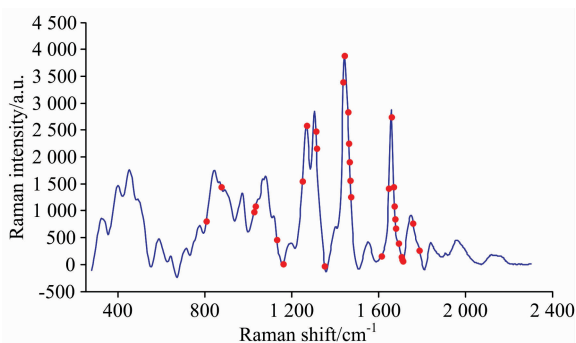


图 3 CARS 方法选择的变量的分布情况

Fig. 3 Distribution of variables selected by CARS method

2.5 模型建立及预测

对 CARS 方法选择的 31 个变量,应用 PLS 方法建立食用植物油中 TFA 含量的定量预测模型,并与常用的 UVE 及 SPA 方法的结果进行比较。表 4 为不同变量选择方法下 TFA 的 PLS 建模结果。由表 4 可知,3 种变量选择方法中, CARS 方法的结果最优, SPA 方法次之,而 UVE 方法则最差。由表 3 和表 4 可知,经 CRAS 变量选择后, TFA 的定量预测模型性能有较大提高,校正集和预测集的相关系数分别

由 0.932, 0.934 上升为 0.949, 0.953, RMSEC 及 RMSEP 则分别由 0.216%, 0.232% 下降为 0.188%, 0.191%, 且建模所用的变量数由 349 个下降为 31 个,表明 CARS 方法能有效筛选 TFA 的拉曼特征变量。

此外,由表 2、表 3 及表 4 可知,经归一化处理、波段初选及 CARS 变量选择后,对于 TFA 定量预测模型,其校正集和预测集的相关系数分别由 0.868, 0.816 上升为 0.949, 0.953, RMSEC 及 RMSEP 则分别由 0.296%, 0.361% 下降为 0.188%, 0.191%, 且建模所用的变量数由 683 个下降为 31 个,仅占原变量数的 4.54%,有效提高了定量预测模型的预测精度和稳定性。

表 4 不同变量选择方法下反式脂肪酸的 PLS 建模结果

Table 4 PLS results of TFA at different variable selection methods

方法	变量数	校正集		预测集	
		r	校正均方根误差/%	r	预测均方根误差/%
CARS	31	0.949	0.188	0.953	0.191
UVE	29	0.910	0.247	0.886	0.307
SPA	15	0.923	0.229	0.906	0.269

3 结 论

利用激光拉曼光谱技术对食用植物油中的 TFA 含量进行定量检测,采用 CARS 方法筛选有用变量,并应用 PLS 方法建立 TFA 含量的定量预测模型。研究表明,激光拉曼光谱技术检测食用植物油中的 TFA 含量是可行的,归一化处理、波段初选及 CARS 方法均有效提高了 TFA 预测模型的预测精度和稳定性;与未优化的模型相比,其校正集和预测集的相关系数分别由 0.868, 0.816 上升为 0.949, 0.953, RMSEC 及 RMSEP 则分别由 0.296%, 0.361% 下降为 0.188%, 0.191%, 且建模所用的变量数由 683 个下降为 31 个,仅占原变量数的 4.54%。

References

- [1] ZHANG Ying, DONG Quan, LU Yan-gang (张莹,董全,卢沿钢). Science and Technology of Food Industry(食品工业科技), 2012, 33(6): 448.
- [2] Ou Y, Pei L, Lai K, et al. Food Analytical Methods, 2017, 10(3): 565.
- [3] Gao F, Hu YX, Chen D, et al. Talanta, 2015, 143: 344.
- [4] FANG Xiao-qian, PENG Yan-kun, WANG Wen-xiu, et al(房晓倩,彭彦昆,王文秀,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(9): 2794.
- [5] Guo Z M, Wang M M, Wu J Z, et al. Food Chemistry, 2019, 286: 282.
- [6] Huy B T, Pham Q T, An N T T, et al. Journal of Luminescence, 2017, 188: 436.
- [7] Kumar S, Goel P, Singh J P. Sensors and Actuators B-Chemical, 2017, 241: 577.
- [8] Chen J M, Huang Y J, Kannan P, et al. Analytical Chemistry, 2016, 88(4): 2149.
- [9] Jiang Y F, Sun D W, Pu H B, et al. Talanta, 2019, 197: 151.
- [10] Cheng J, Wang S, Zhang S, et al. Sensors and Actuators B-Chemical, 2019, 279: 7.
- [11] Stefanov Colman E, Vlaeminck B, et al. Journal of Agricultural and Food Chemistry, 2011, 59(24): 12771.
- [12] Zhao M, Beattie RJ, Fearon AM, et al. International Dairy Journal, 2015, 51: 41.

- [13] Numata Y, Kobayashi H, Oonami N, et al. *Journal of Molecular Structure*, 2019, 1185: 200.
[14] Li H, Liang Y, Xu Q, et al. *Analytica Chimica Acta*, 2009, 648(1): 77.
[15] Muik B, Lendl B, Molina-Diaz A, et al. *Chemistry & Physics of Lipids*, 2005, 134(2): 173.

Determination of Trans Fatty Acids in Edible Vegetable Oil by Laser Raman Spectroscopy

JIANG Xue-song¹, MO Xin-xin³, SUN Tong^{2, 3*}, HU Dong²

1. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China

2. College of Engineering, Zhejiang A&F University, Lin'an 311300, China

3. College of Engineering, Jiangxi Agricultural University, Nanchang 330045, China

Abstract Trans fatty acids (TFA) in oils and fats are harmful to people's health, so it is necessary to monitor their content. In this research, 79 samples of edible vegetable oils were collected, involving 9 varieties and 27 brands. The number of samples that were allocated to calibration and prediction sets was 53 and 26, respectively. Raman spectra of 79 edible vegetable oil samples were collected by a QE65000 Raman spectrometer, and adaptive iteratively reweighted penalized least squares was used to remove fluorescence background of Raman spectra. Then, various normalization methods were used to process Raman spectra, and preliminary selection of modeling wavenumber range of Raman spectra was carried out. After that, competitive adaptive reweighted sampling (CARS) method was used to select TFA-related variables, and partial least squares regression was used to correlate the spectral intensity of TFA characteristic variables with the real content determined by gas chromatography to establish quantitative prediction model of TFA content in edible vegetable oils. The results indicate that among various normalization methods, four normalization methods can improve the performance of PLS quantitative prediction model, and area normalization method has the best effect. After primary selection of wavenumber range, the range of wavenumber is reduced from 686 to 2 301 cm^{-1} to 737 to 1 787 cm^{-1} , and the optimum range of wavenumber is determined to be 737 to 1 787 cm^{-1} . Thirty-one spectral variables are selected by CARS method. The selected spectral variables are mainly distributed near the Raman vibration peaks of 1 265, 1 303, 1 442 and 1 658 cm^{-1} , and the variables in the both sides of the Raman vibration peaks of 974 cm^{-1} are also selected. In addition, the PLS modeling results of CARS method were better than those of the commonly used methods such as uninformative variable elimination and successive projections algorithm. Therefore, it is feasible to detect TFA content in edible vegetable oil by laser Raman spectroscopy combined with chemometrics. Normalization method, wavenumber range selection and CARS method can effectively improve the prediction accuracy and stability of TFA quantitative prediction model. The correlation coefficients and root mean square errors of optimized TFA quantitative prediction model in calibration and prediction sets are 0.949, 0.953 and 0.188%, 0.191%, respectively. Compared with the unoptimized prediction model, the root mean square error of prediction decreases from 0.361% to 0.191%, with a decrease of 47.1%. The number of variables used in modeling decreases from 683 to 31, accounting for only 4.54% of the original variables.

Keywords Raman spectra; Trans fatty acid; Edible vegetable oil; CARS

(Received Aug. 24, 2019; accepted Oct. 29, 2019)

* Corresponding author