

# 近红外光谱结合区间偏最小二乘法 应用于花生油酸价的测定

周小华<sup>1</sup>, 张 玫<sup>1</sup>, 相秉仁<sup>2</sup>

(1. 江苏省食品药品监督检验研究院, 江苏 南京 210008; 2. 中国药科大学, 江苏 南京 210009)

**摘要:**采用近红外光谱法结合不同区间偏最小二乘波长筛选法建立花生油酸价的定量分析模型。采用酸碱滴定法测定花生油样本的酸价同时采集近红外光谱数据;采用区间偏最小二乘法(iPLS)、向后区间偏最小二乘法(BiPLS)、移动窗口偏最小二乘法(mwPLS)优选光谱特征区间;采用偏最小二乘法(PLS)对优选出来的谱段建立酸价的定量模型。结果表明,采用mwPLS选择的谱段建立的模型预测效果最佳, RMSECV和RMSEP分别为0.247 76和0.131 5,校正相关系数和预测相关系数分别为0.993 2和0.996 9。因此,近红外光谱结合移动窗口偏最小二乘法可以快速准确测定花生油的酸价。

**关键词:**近红外;区间偏最小二乘;花生油;酸价;定量分析

中图分类号:O 657.3;TS 227 文献标识码:A 文章编号:1007-7561(2017)02-0062-03

## Determination of acid value of peanut oil by near - infrared spectrometry combined with interval partial least squares

ZHOU Xiao - hua<sup>1</sup>, ZHANG Mei<sup>1</sup>, XIANG Bing - ren<sup>2</sup>

(1. Jiangsu Food and Drug Supervision & Inspection Research Institute, Nanjing Jiangsu 210008;  
2. China Pharmaceutical University, Nanjing Jiangsu 210009)

**Abstract:**Quantitative analysis model of acid value of peanut oil was performed using near - infrared spectrometry combined with different interval partial least squares (iPLS). Acid value of peanut oil were determined by acid - base titration while NIR spectra data were recorded; three improved partial least squares (PLS) methods, including interval partial least squares (iPLS), backward interval partial least - squares (BiPLS) and moving window partial least - squares (mwPLS), were used to find the most informative ranges; PLS regression models of acid value were built based on the optimal ranges. The results showed that the model by mw PLS method had the best predictive ability, the RMSECV and RMSEP were 0.247 76 and 0.131 5, and the calibration and prediction coefficient were 0.993 2 and 0.996 9. Rapidly and accurately determination of acid value of peanut oil can be achieved by NIR combined with mwPLS.

**Key words:**near - infrared; interval partial least squares (iPLS); peanut oil; acid value; quantitative analysis

食用油中游离脂肪酸含量的多少,反映了食用油质量、精炼程度和储藏品质变化,它是食用油品质评定的重要指标,常用酸价来表示。酸价以中和1 g油脂中的游离脂肪酸所需氢氧化钾的毫克数表示。国家标准中,食用油的检测是利用酸碱滴定的方法测定酸价。这种方法过程繁琐,需要消耗很多有机

试剂,不便于现场检测。近红外光谱法具有快速、无损、环保等特点<sup>[1]</sup>,近年来在食用油有关成分的检测应用方面有一定的报道<sup>[2-5]</sup>,Yulan Rao等<sup>[5]</sup>首次将近红外光谱法应用于花生油酸价的测定,结果较佳,校正相关系数为0.972 5,校正均方根误差为0.308,预测相关系数为0.937 9,预测均方根误差为0.333。本实验采用近红外光谱法结合不同区间偏最小二乘波长筛选法建立花生油中的酸价的定量分析模型,旨在利用改进偏最小二乘法可以有效提取

收稿日期:2016-10-27

作者简介:周小华,女,江苏南京,硕士研究生。

复杂体系中目标光谱信息的优点,提高花生油中酸价定量模型的预测精度及稳定性。

## 1 材料与方法

### 1.1 材料

花生油:市售,一共采集 117 个样品。

### 1.2 仪器

MPA 型傅立叶变换近红外光谱仪:德国 Bruker 公司,配有液体光纤探头和 InGaAs 检测器,以及 OPUS™ 近红外操作软件包。

### 1.3 方法

#### 1.3.1 花生油酸价测定<sup>[6]</sup>。

按照国标 GB 2716—2005 规定的酸碱滴定法对 117 个样品的酸价进行测定。

#### 1.3.2 近红外光谱采集

液体光纤漫反射测样:量取样品 15 mL 装于 20 mL 烧杯中。扫描次数:64 次,分辨率:8 cm<sup>-1</sup>,扫描范围:4 000 ~ 12 000 cm<sup>-1</sup>。每个样品重复测定 6 次,取其平均光谱。测样温度:25 °C。花生油样品的近红外原始图谱如图 1。

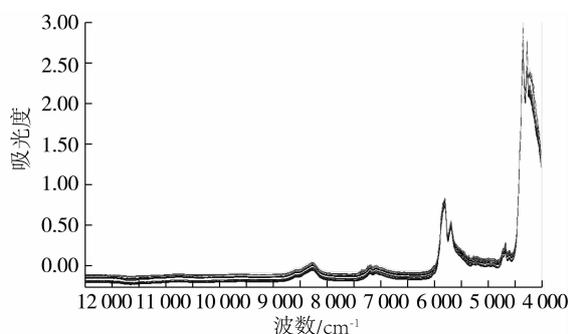


图1 花生油样品的近红外原始图谱

### 1.4 数据处理

考察了区间偏最小二乘法(iPLS)、向后区间偏最小二乘法(BiPLS)以及移动窗口偏最小二乘法(mwPLS)这三种波长筛选法对花生油中酸价检测的影响。用内部交叉验证均方根误差(RMSECV)和校正相关系数(Rc)评估校正模型的质量,用预测均方根误差(RMSEP)和预测相关系数Rp检验所建立模型的预测效果。所有的数据分析都在 MATLAB © 7.0 上实现。

## 2 结果与分析

实验选用市售花生油 117 个,根据 Kennard - Stone 算法<sup>[7]</sup>选取 78 个样品作为训练集,39 个作为预测集。

### 2.1 iPLS 模型

将整个光谱区域划分为 10、11、12、……、19、20 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。表 1 列出了不同子区

间划分的最佳局部模型的结果。由表 1 可以看出,在不同的区间划分中,最佳局部模型的子区间主要集中在 4 500 ~ 5 500 cm<sup>-1</sup> 这一光谱区域内。建立在第 18 个子区间上的模型是所有模型中预测能力最强的一个,其交互验证均方根误差(RMSECV)值为 0.222 1,因此,采用这一光谱区域(4 886.9 ~ 4 443.4 cm<sup>-1</sup>)建模,并对验证集样品进行预测,其校正相关系数(Rc)为 0.991 8,预测相关系数(Rp)和预测均方根误差(RMSEP)分别为 0.996 7 和 0.184 1。

表 1 不同区间划分时各部分的最佳模型

区间划分	产生最佳模型区间	产生最佳模型波数范围/cm <sup>-1</sup>	最佳模型的主因子数	最佳模型的 RMSECV 值
全光谱区域	全光谱区域	12 493.1 ~ 3 999.8	11	0.924 7
10	9	5 693.1 ~ 4 848.4	10	0.350 8
11	10	5 538.8 ~ 4 771.2	10	0.388 5
12	11	5 407.6 ~ 4 705.7	10	0.368 8
13	12	5 299.6 ~ 4 651.7	8	0.364 5
14	13	5 207.1 ~ 4 605.4	10	0.247 2
15	14	5 122.2 ~ 4 562.9	7	0.237 0
16	15	5 052.8 ~ 4 528.2	8	0.238 4
17	16	4 991.1 ~ 4 497.4	6	0.242 5
18	17	4 937.1 ~ 4 470.4	6	0.227 9
19	18	4 886.9 ~ 4 443.4	7	0.222 1
20	19	4 844.5 ~ 4 424.1	6	0.227 7

### 2.2 BiPLS 模型

将光谱划分为 20 个子区间进行 BiPLS 运算,从表 2 可以看出,根据最小 RMSECV 的原则,建立在 2 个子区间上的 PLS 模型为最佳模型。该 PLS 模型校正时的相关系数(Rc)和校正均方根误差(RMSECV)分别为 0.991 5 和 0.226 9,预测时的相关系数(Rp)和预测均方根误差(RMSEP)分别为 0.995 3 和 0.192 9。

表 2 向后区间偏最小二乘法(BiPLS)优化结果

模型中的区间个数	被选区间	RMSECV	模型中变量个数
20	20	1.753 4	2 203
19	1	0.247 6	2 093
18	15	0.240 4	1 982
17	13	0.239 1	1 872
16	8	0.238 4	1 762
15	14	0.237 7	1 652
14	18	0.236 8	1 542
13	5	0.234 0	1 432
12	9	0.230 6	1 322
11	10	0.229 2	1 212
10	4	0.230 1	1 102
9	7	0.230 8	992
8	3	0.237 5	882
7	11	0.233 0	771
6	2	0.235 5	661
5	16	0.246 4	550
4	12	0.247 4	440
3	6	0.242 9	330
2	17	0.226 9	220
1	19	0.227 7	110

### 2.3 mwPLS 模型

将窗口的宽度设置为 95 个波长通道跨度,窗口从左向右以每次一个波数点的速度移动,图 2 给出了所建立的一系列宽度为 95 个波长的窗口偏最小二乘模型的性能谱 RMSECV 曲线,图中的虚线表示基于全谱含有 11 个潜变量的最优全局偏最小二乘模型的交互验证均方根误差, RMSECV = 0. 853 7。图中的光谱图为样本集的均值谱,进行了比例调节,以便对应和确定最优的建模窗口,图中以椭圆示出了较优区域。并在此区域建立 PLS 模型,以含有 6 个潜变量的模型的性能最好, RMSECV = 0. 247 76。为了比较特征谱段建模与全谱建模的差异,采用 PLS 法对全谱进行了建模,结果见表 3。可见 mw-PLS 模型的预测精度最好,表 4 列出了 mwPLS 模型预测集中 39 个样品酸价的真实值与预测值。

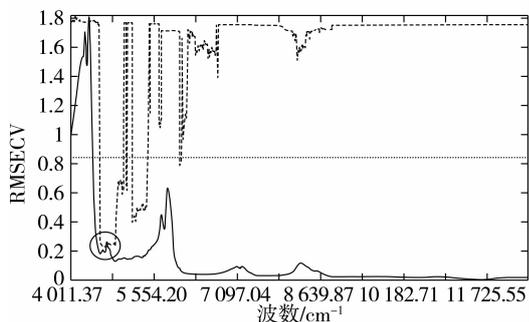


图 2 mwPLS 建模区间

表 3 不同区间偏最小二乘法模型结果比较

模型	建模区间/cm <sup>-1</sup>	RMSECV	Rc	RMSEP	Rp
全谱	12 500 ~ 4 000	0. 853 7	0. 924 7	0. 356 1	0. 879 1
iPLS	4 886 ~ 4 443	0. 222 1	0. 991 8	0. 184 1	0. 996 7
BiPLS	4 845 ~ 4 424 5 693 ~ 5 273	0. 226 9	0. 991 5	0. 192 9	0. 995 3
mwPLS	4 782 ~ 4397	0. 247 8	0. 993 2	0. 131 5	0. 996 9

表 4 mwPLS 模型预测集样品酸价预测值与真实值的比较 mgKOH/g

样品编号	真值	预测值	样品编号	真值	预测值	样品编号	真值	预测值
1	0.8	0.9	14	2.8	2.7	27	3.8	3.9
2	2.9	2.9	15	3.0	2.9	28	3.4	3.6
3	4.0	4.1	16	5.1	4.9	29	2.3	2.4
4	2.3	2.3	17	2.9	3.0	30	4.6	4.8
5	3.9	3.8	18	4.9	5.0	31	4.1	4.1
6	3.4	3.5	19	1.0	0.9	32	1.1	1.1
7	1.4	1.6	20	3.7	4.1	33	4.7	4.5
8	3.0	2.9	21	2.5	2.7	34	2.8	2.9
9	2.7	2.8	22	2.6	2.7	35	3.0	3.0
10	3.3	3.3	23	9.7	9.7	36	1.9	1.9
11	3.4	3.6	24	4.7	4.9	37	4.0	4.1
12	4.7	4.7	25	4.8	5.0	38	2.2	2.2
13	3.6	3.6	26	4.7	4.8	39	2.3	2.5

### 3 结论

采用区间偏最小二乘法及其改进算法筛选波段,3 种方法都能有效选择有用信息,从而提高模型的预测能力。全光谱模型的偏最小二乘模型精度并不高,区间偏最小二乘法处理所得的最佳模型,不论对校正集还是预测集模型的预测能力都好于全光谱模型,由于游离脂肪酸的含量小,花生油的近红外光谱图中大部分信息是与游离脂肪酸无关的干扰物质,用全谱进行建模时,引进了无关信息,不仅使得到的模型复杂化,而且降低了模型的稳定性与预测能力。用区间偏最小二乘法进行建模谱段的选择,能有效地去除无关信息,从而提高模型的稳定性和预测能力。采用的 3 种方法即常规区间偏最小二乘法、向后区间偏最小二乘法以及移动窗口偏最小二乘法所得到的模型精度均好于全光谱模型。由于 4 600 cm<sup>-1</sup>附近是顺式不饱和脂肪酸中 C = C 和 C - H 的合频振动吸收,游离脂肪酸的吸收主要在这一波段,而 5 500 cm<sup>-1</sup>附近是 O - H 伸缩振动的一级倍频吸收,主要包含有水等非目标成分的信息,因此包括这一波段的 BiPLS 模型精度比 iPLS 和 mwPLS 的模型精度低。mwPLS 模型的预测能力比 iPLS 模型的预测能力稍强 (RMSEP 较小),这可能是由于区间偏最小二乘法的区间固定,有效信息提取不完全的缘故。

近红外光谱法结合移动窗口偏最小二乘法建立的花生油酸价定量模型预测准确度比现有的文献方法高,可以实现花生油酸价的快速测定,可以用于花生油质量控制的现场检测。

#### 参考文献:

- [1] 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.
- [2] 于修焯,张静亚,李清华,等. 基于近红外光谱的食用油酸价和过氧化值自动化检测[J]. 农业机械学报,2012, 43(9): 150 - 154.
- [3] 李书国,陈辉,李雪梅,等. 食用油酸价分析检测技术研究进展[J]. 粮油食品科技,2008, 16(3): 31 - 33.
- [4] 吴静珠,石瑞杰,陈岩,等. 食用油油酸的近红外特征谱区优选[J]. 中国粮油学报,2015,30(2):118 - 121.
- [5] Yulan Rao, Bingren Xiang, Xiaohua Zhou, et al. Quantitative and qualitative determination of acid value of peanut oil using near - infrared spectrometry[J]. Journal of Food Engineering, 2009 (93): 249 - 252.
- [6] GB 2716—2005, 食用植物油卫生标准[S].
- [7] Groot P J de, Postma G J, Melssen W J, Buydens L M C. Selecting a representative training set for the classification of demolition waste using remote NIR sensing[J]. Anal. Chim. Acta. 1999, 392(1): 67 - 75. 完