

小波变换去噪用于食用油脂酸价 近红外光谱分析^①

王立琦^② 张礼勇 王铭义

(哈尔滨理工大学测控技术与通信工程学院 哈尔滨市南岗区学府路 52 号 150086)

摘要 以三级大豆油脂酸价检测为研究对象,采用 db4 小波对 36 个油脂样品的近红外光谱进行预处理,选择出最佳的小波分解层数及阈值选取方式,并对原始光谱和经小波去噪后的光谱利用偏最小二乘回归建模,以决定系数、交互验证误差均方根和相对误差均值 3 个统计量来评价回归模型性能。结果表明原始光谱决定系数 0.8668、交互验证误差均方根 0.0641、相对误差均值 3.28%;小波去噪后光谱决定系数 0.9772、交互验证误差均方根 0.0536、相对误差均值 2.26%。小波变换能在消除噪声的基础上最大限度地保留原始信号的能量和特征信息,为建立稳健可靠的预测模型打下基础。

关键词 油脂酸价;近红外光谱分析;db4 小波;去噪

中图分类号: O657.33 **文献标识码:** A **文章编号:** 1004-8138(2011)02-0911-04

1 引言

随着计算机技术、化学计量学等学科的发展,近红外光谱技术已成为现代最为流行的分析方法,在食品加工、石油化工、医学医药以及烟草工业、纺织工业等各个领域中都得到广泛应用^[1]。由于近红外光谱分析是一种二次分析方法,需要通过近红外光谱仪扫描得到的光谱与样本性质真值之间关联,利用化学计量学方法建立校正模型,再通过校正模型来实现对未知样本性质的测定。校正模型性能的好坏将直接影响预测结果的准确度与精度。因此,在近红外光谱分析研究中,如何建立稳健可靠的校正模型就成为了关键所在。

近红外光谱在获取过程中会伴随各种干扰,以致得到的光谱数据中常包含着不同程度的噪声^[2]。为了减弱随机噪声、样品背景干扰以及测样器件引起的光谱差异等因素对校正结果的影响,通常在建模前要对原始光谱数据进行一定的预处理。光谱预处理方法多种多样,如平滑、微分、多元散射校正等等,其中每种方法都有一定的适用范围,都存在一定的局限性,且一般不能单独使用,需要几种方法结合才能使预处理效果达到最佳,运用起来比较繁琐^[3],于是人们不断寻求新的光谱预处理方法,小波变换就是其中的一种,它在信号去噪中一般都能达到比较理想的效果,成为现代最流行的信号去噪方法。

本文以三级大豆油脂酸价检测为研究对象,对其近红外光谱采用小波变换预处理,通过

① 国家科技支撑计划 (2009BADB9B08);黑龙江省教育厅科学技术研究项目(11551109)

② 联系人,手机:(0)15204656615;E-mail:hsdlwq@163.com。地址:哈尔滨市松北区学海街1号哈尔滨商业大学计算机与信息工程学院,150028

作者简介:王立琦(1966—),女,哈尔滨市人,教授,硕士,主要从事油脂质量快速检测及安全控制工作。

Matlab 仿真进行深入研究, 选择出最佳的小波分解层数及阈值选取方式, 达到理想的去噪效果。

2 实验部分

2.1 样本光谱数据的采集

利用 DA7200 型二极管阵列固定光栅近红外光谱仪(瑞典波通公司), 对从油脂加工生产线上收集的 36 个有代表性的三级食用大豆油脂样本进行透射光谱扫描。仪器的光谱范围是 950—1650nm, 最大采样次数 100 次/s, 为减小光谱数据的测量误差, 采用多次测量求平均值的方式, 每个样本测量 3 次, 再以 3 次的平均值作为最终的光谱值, 36 个样本的原始光谱如图 1 所示。

由图 1 可以看出, 三级大豆油脂的近红外光谱吸收整体比较微弱, 其中, 在 1200nm 和 1400nm 附近有两个很强的吸收峰, 且吸收峰较尖锐, 符合近红外光谱吸收的特征。

2.2 样本化学值的测定

对选取的油脂样本酸价化学值的测定是采用国标法(GB 5530-85): 先利用中性乙醚-乙醇(2:1)混合溶液溶解样本油脂及其所含的脂肪酸, 再用碱标准溶液(通常为 0.1mol/L 氢氧化钾溶液)利用酸碱中和原理进行滴定, 最后根据选取样本试样的质量以及所消耗碱溶液的体积计算出酸价。计算公式:

$$\text{酸价}(\text{mg KOH/g 油脂}) = \frac{V \times C \times 56.1}{m}$$

式中: V ——滴定过程中消耗掉的氢氧化钾溶液的体积, mL; C ——氢氧化钾溶液浓度, mol/L; 56.1——氢氧化钾的摩尔质量, g/mol; m ——称取的试样质量, g。

双试验结果允许差不超过 0.2mg KOH/g 油, 求其平均数, 即为测定结果。测得样本酸价化学值范围 1.46—2.56。

3 食用油脂酸价检测近红外光谱小波阈值去噪实现

3.1 小波阈值去噪基本理论

小波变换具有多分辨率分析的特性, 能同时在时域和频域有效地表征信号的局部特征信息, 并把信号分解成低频和高频两部分, 其中低频部分的信号比较平稳, 能表征真实信号的大部分信息, 而高频部分变化较大, 通常视为噪声^[4]。小波变换去噪主要是利用正交小波的去相关性, 选取正交小波基对信号进行分解, 再选取适当的阈值对高频系数进行量化处理, 最后再利用小波分解得到的低频系数与经过阈值量化处理的高频系数进行重构即可得到去噪后的信号。

小波阈值去噪函数有硬阈值函数和软阈值函数两种^[5]。硬阈值法在均方误差意义上比软阈值优越, 但连续性差, 不具有与原始信号一致的光滑性; 软阈值估计的整体连续性较好, 但重构信号与真实信号间的相似性稍差。一般情况下, 软阈值去噪时相似性上的不足往往在误差允许范围内, 所以较硬阈值方法应用多, 因此本文选择软阈值法。

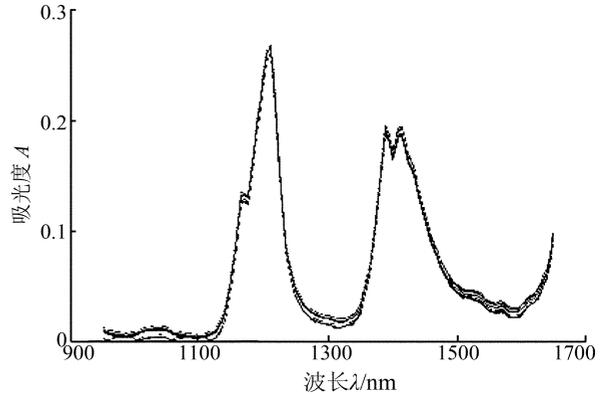


图 1 三级大豆油脂近红外光谱图

噪比取得的。有 3 种阈值选取方式: 缺省阈值、penalty 阈值和 Birge-Massart 策略确定阈值, 其中前两种是全局阈值方式, 而第 3 种是分层阈值方式^[6]。

本文选择 db4 小波基对 36 个油脂样品的近红外光谱数据进行小波分解。dbN 系列小波是著名小波学者 Ingrid Daubechies 创造的, 她发明的紧支集正交小波是小波领域中宏伟的里程碑, 使小波由理论研究走进实践, 逐步应用到各个领域。dbN 系列小波具有很好的正交性, 虽无解析的表达式, 但其扩展性和光滑性好, 可灵活权衡增加支集长度带来的边界问题, 提高信号能量的集中程度, 对信号去噪有非常好的效果。

3.2 小波分解层数的确定

任何一个信号的小波分解都不是无限制的, 一般根据信号的长度来确定分解的最大层数。在 Matlab 小波工具箱中, 往往通过语句 $N = wmaxlev(n, 'wname')$ 来实现最大分解层数的确定。本课题中对每个校正集样本的光谱数据进行采样, 采样间隔为 5nm, 从 950—1650nm 共 141 个采样点, 所以信号长度 $n = 141$, 计算出分解的最大层数是 4, 也就是说实际的小波变换去噪中使用的小波分解层数可以为 2、3 或 4。那么, 到底多少层分解能达到最佳效果呢? 为此分别对油脂样品近红外光谱数据采用不同分解层数的去噪效果进行了对比研究, 并分别选用 Brige-Massart 分层阈值、Penalty 全局阈值和缺省阈值 3 种不同阈值方式及软阈值函数对其进行处理。表 1 分别列出 db4 小波 3 种分解层数下的去噪结果统计值。

表 1 db4 小波变换去噪结果统计

	阈值函数	去噪信号在原信号中的能量	去噪信号与原信号标准差
4 层分解	Penalty 阈值	0.9991	0.0650
	Brige-Massart	0.9659	0.0650
	缺省阈值	0.9927	0.0427
3 层分解	Penalty 阈值	0.9995	0.0735
	Brige-Massart	0.9919	0.0735
	缺省阈值	0.9959	0.0374
2 层分解	Penalty 阈值	0.9998	0.0590
	Brige-Massart	0.9988	0.0262
	缺省阈值	0.9985	0.0290

由以上的统计值可以看出, 小波分解层数越少, 去噪后信号在原信号中的能量比例越高、与原始信号的标准差越小, 去噪效果越好, 由此可以得出最佳小波分解层数应为 2。

3.3 小波去噪中阈值方式的确定

因小波阈值去噪中不同阈值选取方式会带来不同去噪效果, 于是对经 db4 小波 2 层分解的 3 种阈值方式去噪后的光谱数据和未经预处理的光谱数据建立 PLS 回归校正模型, 统计结果对比如表 2 所示。

表 2 回归模型统计结果

回归模型	主成分数	PRESS	RMSECV	r^2	绝对误差均值	相对误差均值(%)
原始数据模型	3	0.1304	0.0641	0.8668	0.0541	3.28
缺省阈值去噪模型	3	0.1279	0.0538	0.9770	0.0430	2.27
分层阈值去噪模型	3	0.1278	0.0536	0.9772	0.0429	2.26
Penalty 去噪模型	3	0.1290	0.0537	0.9771	0.0429	2.27

注: RMSECV——交互验证误差均方根; PRESS——预测残差平方和。

由表 2 可以看出, 虽然去噪前后建立 PLS 回归模型时选取的主成分都为 3, 但去噪后的 PRESS 比去噪前要小, 而每个统计结果在去噪后都有所改善。在 3 种阈值选取方式去噪后的模型中, 分层阈值的统计结果优于其他两种全局阈值方法, 再次说明对近红外光谱数据预处理是十分必

要的,并确定分层阈值方式、软阈值函数时的去噪效果最好。

3.4 校正模型的验证

为了进一步验证结论的可靠性以及所建立回归模型在预测结果上是否能达到要求,选取 5 个与校正集相同条件下测定的样本作为验证集,对小波分解去噪后的全谱数据建立 PLS 回归模型,结果见表 3。

表 3 校正模型的验证结果统计

决定系数 r^2	预测集验证误差均方根 RMSEP	相对标准差 RSD(%)
0.9734	0.0484	2.44

表 3 可以看出,校正模型的验证集绝对误差与相对误差都在允许范围内,以上结果说明所建立的校正模型通过验证,性能良好,可应用于实践。

4 结论

利用小波变换对油脂样品的光谱进行预处理,可以在消除噪声的基础上最大限度地保留原始信号的能量和特征信息,为建立稳定、可靠的预测模型打下基础。

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. 1—20.
- [2] Karstang T V, Kvalheim O. Multivariate Prediction and Background Correction Using Local Modeling and Derivative Spectroscopy [J]. *Analytical Chemistry*, 1991, **63**(8): 767—772.
- [3] 彭玉华. 小波变换与工程应用[M]. 北京: 科学出版社, 1999. 1—175.
- [4] Karim G O, David J A. Noise Reduction in Multichannel Neural Recordings Using a New Array Wavelet Denoising Algorithm [J]. *Neurocomputing*, 2001, **38**(4): 1687—1693.
- [5] Donoho D L, Johnstone I M. Wavelet Shrinkage: Asymptotia [J]. *Journal of the Royal Statistical Society, Series (B)*, 1995, **7**(2): 301—369.
- [6] 刘娟花, 李福德. 基于小波变换的信号去噪研究[J]. 西安理工大学学报, 2004, **38**(1): 34—35.

Application of Wavelet Transform Denoising in Near Infrared Spectral Analysis of Edible Oil Acid Value

WANG Li-Qi ZHANG Li-Yong WANG Ming-Yi

(School of Measurement-Control Technology and Communication Engineering, Harbin University of Science and Technology, Harbin 150086, P. R. China)

Abstract The detection of third degree soybean oil acid value was used as research object, near infrared spectral datas of 36 oil samples were preprocessed by db4 wavelet, best wavelet decomposition layer and selection way of threshold were chose, then models of initial spectrum and wavelet denoised spectrum were constructed using partial least square regression, and the performances of models were evaluated by three statistics named decision coefficient r^2 , RMSECV and the mean of relative error, the r^2 , RMSECV and the mean of relative error were 0.8668, 0.0641 and 3.28%, respectively, for initial spectrum, and were 0.9772, 0.0536 and 2.26%, respectively, for wavelet denoised spectrum. Obviously, the wavelet transform can remain as much as possible energy and feature information of initial signals, at the same time eliminate the noise, which laid the foundation for constructing stable and reliable forecasting model.

Key words Oil Acid Value; Near Infrared Spectral Analysis; db4 Wavelet; Denoising