

文章编号:1672-8785(2015)01-0030-05

用 Daubechies 小波软阈值法对近红外光谱数据进行预处理

李 敏

(乐山师范学院物理与电子工程学院, 四川乐山 614000)

摘 要: 提出了一种基于 Daubechies 小波软阈值的近红外光谱预处理方法, 并用该方法处理了两类不同苹果的近红外光谱数据。该方法选用归一化相关系数 (Normalized Correlation, NC) 和峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 作为定量分析指标。与常用的矢量归一化法 (Vector Normalization, VN) 和标准正态变量校正法 (Standard normal Variate, SNV) 相比, 该方法优势明显, 既能有效去除噪声, 又能很好地保留光谱的特征细节信息, 提高了后续光谱分析过程中建模的稳健性和模型预测的精确度。

关键词: 近红外光谱; 预处理; Daubechies 小波; 软阈值

中图分类号:TN911.7

文献标志码:A

DOI: 10.3969/j.issn.1672-8785.2015.01.006

Near Infrared Spectrum Preprocessing Using Daubechies Wavelet Soft Threshold Method

LI Min

(School of Physics and Electrical Engineering, Leshan Normal University, Leshan 614000, China)

Abstract: A near infrared spectrum preprocessing method based on Daubechies wavelet threshold is proposed. It is used to process the near infrared spectra of two kinds of apples. In the method, the normalized Correlation Coefficient (NC) and Peak Signal to Noise Ratio (PSNR) are selected as a quantitative analysis index. Compared with the Vector Normalization (VN) and Standard Normal Variate (SNV), the method has obvious advantages. It not only can effectively remove noises, but also can retain the details of spectral characteristics. So, the robustness of modeling and the accuracy of the established model can be improved in the subsequent spectral analysis.

Key words: near infrared spectrum; data preprocessing; daubechies wavelet; soft threshold

0 序言

在近红外光谱分析技术中, 光谱数据的预处理是一个非常重要的环节。采集的光谱数据由于受仪器以及环境等因素的影响, 不可避免地含有噪声信息。光谱数据预处理的目的是

既要有效地去除噪声, 同时又要很好地保留光谱的特性和细节信息, 从而提高后续光谱分析建模的稳定性和模型预测的准确性^[1]。

近年来, 常用的预处理方法有 VN 法和 SNV 法。VN 法能有效去除由于测量值大小不同所

收稿日期: 2014-10-31

基金项目: 四川省教育厅重点项目 (12ZA070)

作者简介: 李敏 (1977-), 女, 四川汉源人, 副教授, 硕士研究生。主要研究方向为信号处理、模式识别、近红外光谱分析和小波分析等。

E-mail: cassie_li@163.com

导致的数据集方差,但也可能会丢失重要的方差。SNV 方法在完全消除或部分消除实验误差的同时,会引入人为误差,从而导致后续的结果错误。小波分析由于同时具有时-频局部化特性,近年来被广泛用于信号分析。本文提出一种用 Daubechies 小波软阈值方法预处理两类苹果的红外光谱的方法。定性分析和定量分析表明,该方法优于 VN 法和 SNV 法,表现出超强的去噪能力,同时又能较好地保留光谱的特征细节信息。

1 常用预处理方法简介

设 x_{ik} 为第 i 个样本的第 k 个测量数据; x'_{ik} 为处理后的数据, n 为样本总数。

1.1 矢量归一化处理

归一化处理的目的是使数据集中,使各数据矢量具有相同的长度,且一般为单位长度^[2],计算公式为

$$x'_{ik} = \frac{x_{ik}}{\sqrt{\sum_{i=1}^n x_{ik}^2}} \quad (1)$$

归一化处理能有效去除由于测量值大小不同导致的数据集的方差,但也可能会丢失重要的方差。

1.2 标准正态变量校正

标准正态变量校正也称为方差归一化,其计算公式如下:

$$x'_{i,k} = \frac{x_{i,k} - \bar{x}_k}{s_k} \quad (2)$$

$$s_k = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{i,k} - \bar{x}_k)^2} \quad (3)$$

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{i,k} \quad (4)$$

Canddfi 等^[3]的研究表明,SNV 法能有效地减少由于样本粉末粒度不同而引起的类内距离,从而提高模式识别的准确率。而 Guo Q 等^[4]的研究则表明,经 SNV 法处理后,各波长点的吸收度值之和始终保持为 0,光谱吸收峰处的差异被分散到背景差异中,也就是说 SNV 方法在完全消除或部分消除实验误差的同时,又引入人为误差,导致后续的结果错误。

2 Daubechies (dbN) 小波软阈值法

2.1 本文预处理方法

本文预处理算法流程如图 1 所示。苹果的红外光谱数据经 db4 小波三层分解后,噪声信号主要在各级高频系数中。采用软阈值方法,对各级高频系数进行阈值处理,最后由处理后的各级高频系数和最后一级低频系数重构光谱数据。处理结果有效地去除了光谱数据中的噪声信息,同时较好地保留了光谱数据的细节信息。

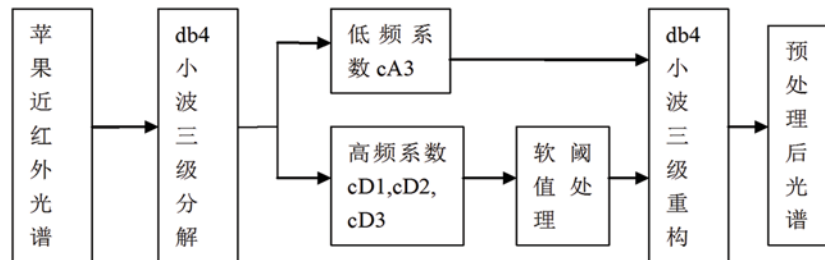


图 1 本文预处理方法的流程图

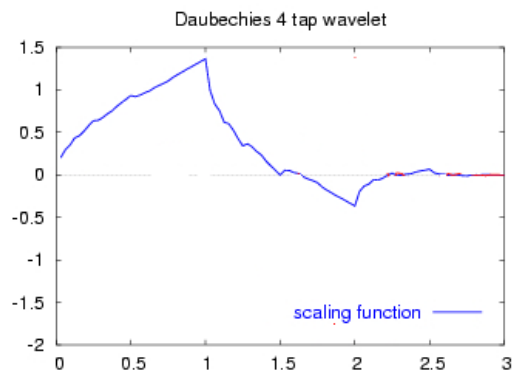
2.2 dbN 小波分解和重构

Daubechies 函数是由世界著名的小波分析学者 Inrid Daubechies 构造的小波函数。Daubechies

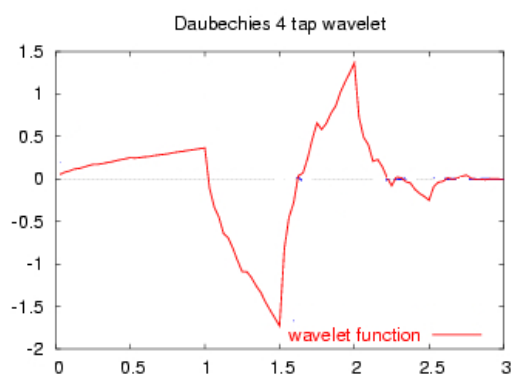
小波函数比 Haar 小波具有更有效的分析和综合能力。Daubechies 系中的小波基计为 dbN, N 为序号, $N = 1, 2, \dots, 10$ 。db4 小波的尺度函数 φ 和

小波函数 ψ 如图 2 所示。

dbN 函数是紧支撑标准正交小波, 它的出现使离散小波分析成为可能。尺度函数 φ 和小波函数 ψ 的有效支撑长度为 $2N-1$, 小波函数的消失矩阶数为 N 。dbN 小波基本不具有对称性, 其正则性随着 N 的增加而增加。正则性越强, 信号重构时越光滑稳定 [5]。



(a) db4 尺度函数 ϕ



(b) db4 小波函数 ψ

图 2 db4 小波

若信号序列 s 的长度为 N , 其一维离散小波分解最多 $\log_2 N$ 级。信号一级小波分解的过程如图 3 所示 [6], 信号 s 和低通滤波器 Lo_D 卷积得到低频系数 cA1, s 和高通滤波器 Hi_D 卷积得到高频系数 cD1。让低频系数 cA1 分别和低通滤波器、高通滤波器卷积, 得到第二级分解低频系数 cA2 和 cD2。第三级、第四级等小波分解以此类

推。最后由各级分解的高频系数和最后一级低频系数通过小波逆变换重构信号 s 。

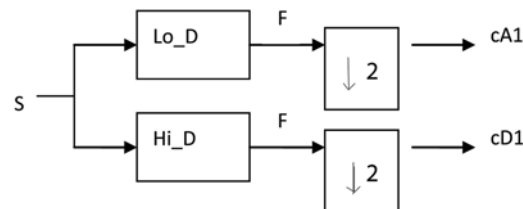


图 3 一层小波分解

2.3 软阈值处理

1995 年, Donoho [7] 提出软阈值算法。定义阈值 T_s 为

$$T_s = s\sqrt{2\ln n}, s = 3 \times \left[\frac{1}{n-1} \sum_{i=1}^n d_i^2 \right]^{1/2} \quad (5)$$

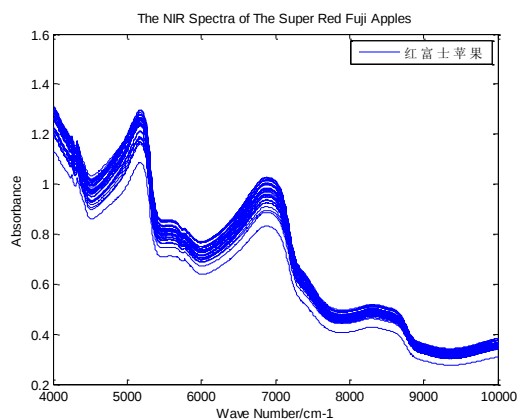
式中, s 为噪声的估计值, n 是数据的点数, d_i 是第 i 个数据。软阈值处理的依据为

$$d_{i,j}^* = \begin{cases} 0, & |d_{i,j}| < T_s \\ \text{sgn} d_{i,j} \cdot (|d_{i,j}| - T_s), & |d_{i,j}| \geq T_s \end{cases} \quad (6)$$

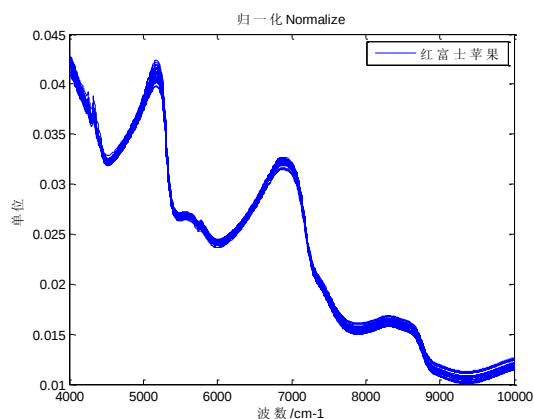
式中, $d_{i,j}$ 为高频系数, sgn 为符号函数。若采用该软阈值法进行处理, 经小波分解后, 光谱数据的高频系数既能有效消除噪声, 又能保留光谱信息中的特征信息和细节信息。

3 实验分析

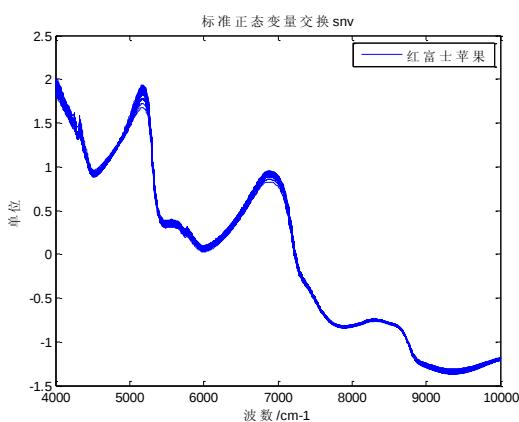
选用大小均匀、无损伤的山东产红富士苹果、甘肃产花牛苹果各 30 个组成实验样品, 样品在室温为 $20 \sim 25^\circ\text{C}$ 的实验室内存放 12 h。所用光谱采集器为赛默飞世尔 (Thermo Fisher) 公司生产的 Antaris II 近红外光谱分析仪。采集光谱时, 先把近红外光谱分析仪预热 1 h, 并采用反射积分球模式。光谱扫描的波数为 $4000 \sim 10000 \text{ cm}^{-1}$, 扫描间隔为 3.856 cm^{-1} 。为减少误差, 近红外光谱分析仪沿每个苹果的赤道轨迹扫描 3 次, 取其平均值作为最终的实验数据, 每个样本的光谱数据是一个 1557 维的数据。采集到的两类苹果的近红外光谱如图 4 (a) 和图 5 (a) 所示。



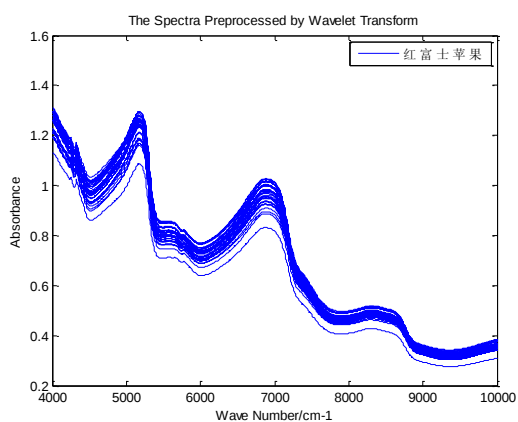
(a) 红富士苹果的近红外光谱



(b) VN 法预处理后的光谱



(c) SNV 法预处理后的光谱

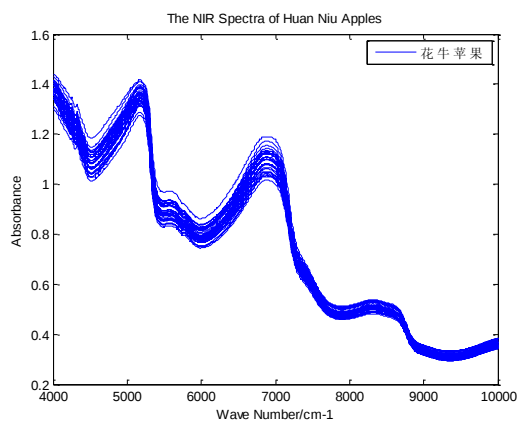


(d) 本文方法预处理后的光谱

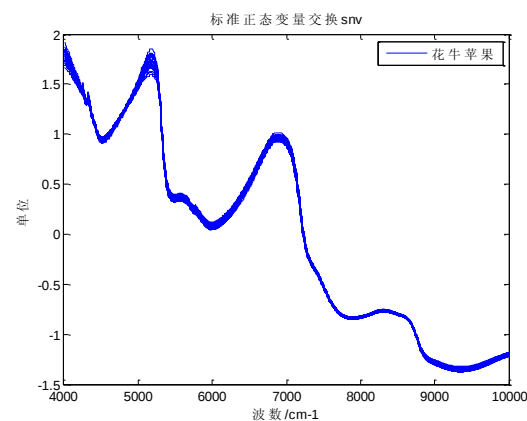
图 4 红富士苹果的近红外光谱预处理

对两类苹果的近红外光谱数据分别采用 VN 法、SNV 法和本文提出的方法进行处理。预处理结果分别如图 4 和图 5 所示。由图可知, 采用

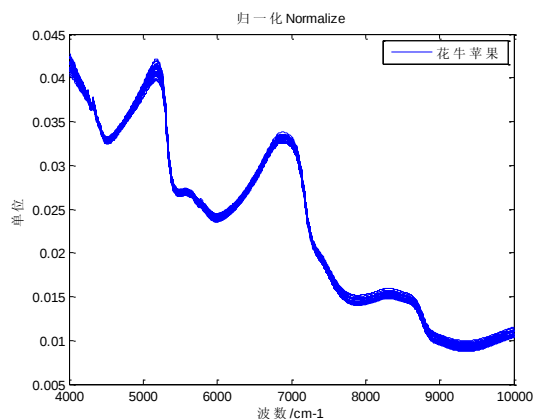
VN 法和 SNV 法处理后, 光谱图像和原始光谱图像的差异较大; 采用本文提出的方法进行预处理后, 光谱图像和原始光谱图像的相似度很高。



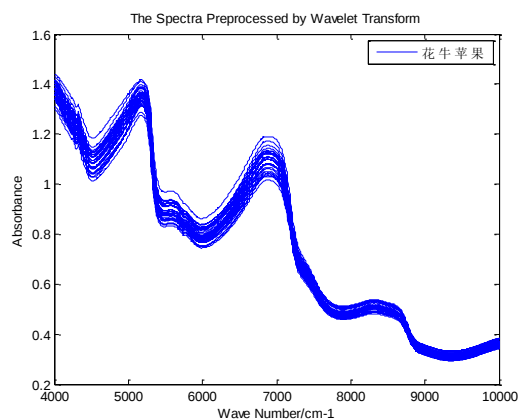
(a) 花牛苹果的近红外光谱



(b) SNV 法预处理后的光谱



(c) VN 法预处理后的光谱



(d) 本文方法预处理后的光谱

图 5 花牛苹果近红外光谱的预处理结果

采用 NC 和 $PSNR$ 对三种预处理方法的结果进行定量分析。 NC 用来评价预处理前后光谱的相似程度, NC 越大, 表明预处理后保留的光谱信号特征和细节信息越多。 $PSNR$ 用来评价预处理方法的去噪能力, $PSNR$ 值越大, 表明去噪效果越好。

表 1 列出了三种预处理方法的归一化相关系数。采用本文方法对两种不同类型苹果的光谱进行预处理后的 NC 值都是最大的, 且保持为 1。这表明与其它两种方法相比, 本文算法能更好地保留近红外光谱的细节信息, 可提高后续模型预测的准确度。

表 1 三种预处理方法的 NC 比较

苹果种类	VN 法	SNV 法	本文方法
红富士苹果	0.9992	0.3789	1.0000
花牛苹果	0.9995	0.4040	1.0000

表 2 列出了三种预处理方法的峰值信噪比。与其它两种方法相比, 采用本文方法进行预处理后, 两类苹果的峰值信噪比都是最大的, 而且

是其它两种方法的 10~14 倍左右。显然, 本文方法表现出超强的去噪能力, 能明显地提高后续建模的稳定度。

表 2 三种预处理方法的 $PSNR$ 比较

苹果种类	VN 法	SNV 法	本文方法
红富士苹果	5.0233	2.3538	79.0332
花牛苹果	4.9486	3.0544	68.0728

4 结束语

从实验结果不难看出, 采用矢量归一化法和标准正态变量矫正法对苹果近红外光谱数据进行预处理的效果非常接近, 去噪能力不强, 光谱信息损失较多。本文提出的基于 Daubechies 小波软阈值的预处理方法不仅能有效地去除光谱

信息中的噪声, 还能很好地保留其特征信息, 大大提高了后续建模的稳定性和预测的准确性, 值得推广应用。

近红外光谱分析属于弱信号分析的范畴。随着我们的预处理方法研究的深入, 更多具有针

(下转第 45 页)