

红外光谱预处理中去噪的研究

刘志宏^{①a,b} 邓波^a 周玉荣^a 庞小峰^a 张晓春^b

^a(电子科技大学生命科学与技术学院 成都市 610054)

^b(成都信息工程学院电子工程系 成都市 610041)

摘 要 红外(IR)光谱由于包含了噪声等各种外界干扰因素。应该先进行光谱预处理,以便降噪,提高分析准确度。本文采用了一种基于最优小波包基的信号去噪算法,该算法根据最小代价原理,采用不同的阈值算法对光谱的高频和低频信号进行量化处理,用量化后的系数重构得到去噪信号,从而达到较好的去噪效果。实验表明,本方法处理后光谱曲线非常光滑、噪声消除效果明显。

关键词 去噪,小波包变换,最优小波包基,红外光谱。

中图分类号:O657.33

文献标识码:A

文章编号:1004-8138(2006)04-0815-05

1 引言

滤波去噪是光谱分析数据处理中最基本的数据预处理环节之一。傅里叶变换红外(FTIR)光谱分析是由被测样品的红外特征光谱主导,包含了噪声等各种外界干扰因素,从而导致分析结果的准确度和精密度下降。所以应该先进行光谱预处理,以便降噪、减少各种干扰的影响,提高分析准确度,改善光谱分析信号的性能^[1,2]。

小波变换由于具有多分辨率分析的特性,可实现信号的平滑和去噪。目前,光谱分析中利用小波变换对信号进行平滑和去噪的方法主要有两种。第一种方法是直接利用Mallat 算法将信号分解为高频和低频信息,使频率高于截断尺度的空间向量全部置为零,然后进行信号重构即可达到信号消噪的目的^[3]。另一种方法是根据信号与噪声小波系数模极大值在小波变换域的不同表现,去除噪声所对应的极大值,然后重构原始信号,以达到滤波的目的^[4]。

这些方法的滤波结果受操作者人为因素的影响。更重要的是,仅适用于信号频率(低频)与噪声频率(高频)相差较大的情况,不适用与窄带锐峰信号(高频信号)。因此,为了得到更好的滤波效果,本文采用了一种利用小波包进行光谱信号的滤波方法—基于最优小波包基的去噪算法,该算法利用最小代价原理,采用不同的阈值对高频和低频系数进行量化处理,重构得到去噪后的信号,无论对高频、低频信号均有很好的消噪能力,实验结果令人满意。

2 去噪原理

2.1 最优小波包基

小波包基是小波包组成的不同正交基分解结果。对于所有的小波包基,信号代价熵值最小的为最优小波包基,即 $E(s) = \sum_i E(s_i)$ 最小,其中 s 代表信号, s_i 代表 s 在一个正交小波包基上的投影系

① 联系人,电话:(028)81928135; E-mail: lzhh@cuit.edu.cn

作者简介:刘志宏(1968—),女,山西省汾阳县人,电子科技大学生命科学与技术学院在读博士,从事生物医学工程专业;

庞小峰(1944—),男,四川省西充县人,博士生导师,主要研究方向为生物物理学。

收稿日期:2006-01-26;接受日期:2006-04-03

数。通常所使用的代价函数要求具有可加性^[5,6],即:

$$M(0) = 0, M(\{x_i\}) = \sum_i M(x_i)$$

根据双尺度方程:

$$\begin{aligned} U_{2n}(t) &= \frac{1}{2} \sum_k h^k u_n(2t - k) \\ U_{2n+1}(t) &= \frac{1}{2} \sum_k g^k u_n(2t - k) \end{aligned} \quad (1)$$

称由公式(1)所定义的函数族 $\{U_n(t) | n \in Z\}$, (Z 表示非负整数)为由 $U_0 = \varphi$ 所确定的小波包,由此小波包 $\{U_n(t)\}_{n \in Z}$ 包括尺度函数 $U_0(\varphi)$ 和小波函数 $U_1(\varphi)$ 在内的一定联系的函数集合,而从小波包 $\{2^{-(j-k)/2} U_n(2^{(j-k)}t - l), n = 2^k + m, l \in Z\}$ 中抽取的能组成 $L^2(R)$ 的一组正交基就称为 $L^2(R)$ 的一个小波包基。已知小波库里有很多小波包基,并且能在一定尺度下提供一种最好的分析频率成分瞬变的方法。本文以信息熵(Shannon 熵)作为代价来选取最优基,以便能最有效地表达信号的特征,从而改变光谱分析信号的性能。

为了找到一个最优的小波包基,以便能最有效地表达出信号的特性。首先给定一个序列的信息代价函数 M ,然后在所有小波包基中寻找使 M 最小的基。对于一个给定的序列来说,这样一个信息代价函数 M 应当能够测得准确描述该序列所需要的小波包分解系数的个数。若函数 M 已选定,从小波包基库中选出的某一正交基 B ,而 B_x 是 x 在基 B 下展开的系数序列,如果 $M(B_x)$ 是最小的,则 B 为“信息代价函数最小”意义下的最优基。其实质是要使小波包分解系数间彼此有较大差异而信息损失又较少,使信号中隐含的信息能集中反映在少数几个分解系数上,为光谱信号的消噪处理提供一个较好的分解和重构途径。

2.2 最优小波包基的快速搜索法

信息熵即定义序列 $x = \{x_j\}$ 的Shanon-Weaker 熵为 $M(x) = - \sum_j p_j \lg p_j$,其中 $p_j = \frac{|x_j|^2}{\|x\|^2}$,且 $p = 0$ 时, $p \lg p = 0$,由信息代价函数,采用从底向顶的搜索方法从小波包基库中选取最优基。即在小波库中的所有小波包基中使代价函数最小的基。具体算法如下^[7-9]:

(1) 根据信息代价函数,对序列进行归一化处理,确定所要选择的小波包函数以及分解层数 N ,进行 N 层小波包分解,并计算出各节点的熵值;(2) 对最低层的所有节点开始标记,将它们的信息代价作为一个初始值,称上层节点为父节点,下层节点为子节点。当父节点的信息代价比子节点低,那么就标记父节点;否则不标记。如此上推,直到顶层;(3) 检查所有的节点取最上层所标记的节点。当高层有节点被标记时,删除相应子节点的标记,从下往上对相邻两层按上述过程逐层进行比较,直至对所有层均处理完毕;(4) 被作上标记的所有节点所对应的小波包函数即是搜索到的最优小波包基。最后将最优基中的系数选取出来,按一定的顺序输出。

3 基于最优基的信号消噪

3.1 阈值的选取

在用小波进行信号的消噪的过程中,需要用阈值进行小波分解系数的量化处理,选择合适的阈值是阈值化处理的关键问题,直接决定了信号消噪的效果。对于阈值的选取,文中通过对小波包分解在不同的频带上系数,通过对Donoho硬阈值和软阈值的改进,采用不同的阈值进行计算^[8,9]。对于高频系数,采用自适应阈值选取算法。采用最小风险所对应的小波变换系数作为阈值,求出与其对应的风险值,即得到它的似然估计,进行非似然最小化,得到所选的阈值。其具体算法为:(1) 将每一层变换后的系数由小到大排列,得到一个向量 $p = [p_1, p_2, p_3 \cdots p_N]$, N 为待估计向量的长度;(2) 对应每一个元素下标 k ,取门限为待估计向量 R 元素中的最小值 r_k 为风险值,由 r_k 求出对应的 p^k ,风险算法为:

$$R \in [r_1, r_2, \cdots, r_N], r_k = [N - 2k + (N - P_k) \frac{p_k}{\sum_{i=1}^k p_i}] / N, k = 1, 2, \cdots, N. \quad (2)$$

(3) 根据公式(2), 做出对应于不同 k 值的风险曲线, 找出最小风险点及与之对应得 k 值, 使风险值最小的门限就是所要选取的门限, 其门限为: $T = P_k$, 对于低频系数, 阈值的计算是根据公式 $T = \sigma \sqrt{2 \ln N / N}$ 来计算, 对于噪声能量的计算, 可以从能量的角度出发, 各个系数所携带信号能量的大小计算其权值, 再利用这个权值和小波包系数的中位数来估计各频段的噪声能量^[9], 可根据公式:

$$\sigma = \frac{1}{0.6745 N} \sum_{i=1}^N |d_i^k| \tag{3}$$

式中 d_i^k 为第 k 层第 i 个系数($1 < k < j$) 来计算。

3.2 基于最优小波包基的信号消噪方法

通过最优小波包基分解进行信号的消噪处理. 具体方法可按以下步骤进行: (1) 选择一个小波函数并确定一个小波包分解的层数 N , 然后对信号进行 N 层小波包分解; (2) 根据前面介绍的最优小波包基搜索算法, 搜索信号的最优小波包基; (3) 采用上面介绍的阈值选择算法, 对分解的各频段采用不同的阈值选择方法, 对最优小波包基各个节点上的分解系数进行阈值量化; (4) 利用阈值量化后的最优小波包基的分解系数进行重构, 重构得到的信号就是经过最优小波包基消噪处理后的信号。

4 实验结果及分析

4.1 实验部分

测试样品为川芎(购自成都杏林医药股份有限公司), 使用的仪器设备是 Thermo Electrongon 公司的 Nexus670 傅里叶变换红外光谱仪, 红外光谱分辨率为 4cm^{-1} , 测量范围中红外 $4000\text{—}400\text{cm}^{-1}$, 增益为自动增益。最后测得的光谱如图 1 所示。用此光谱作为标准光谱。在此标准信号上, 随机加入一些噪声信号, 如图 2 所示。

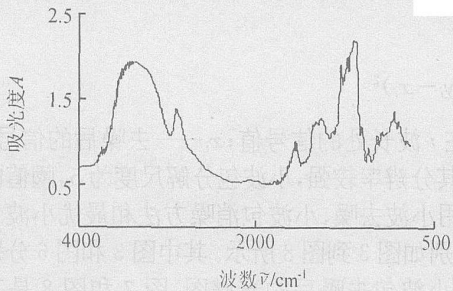


图 1 标准光谱图

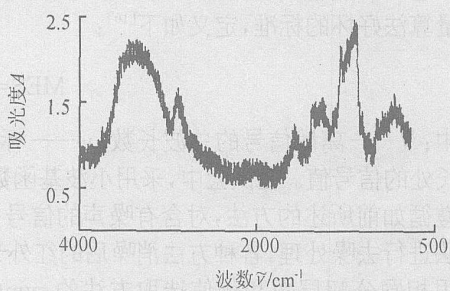


图 2 含有噪声的光谱图

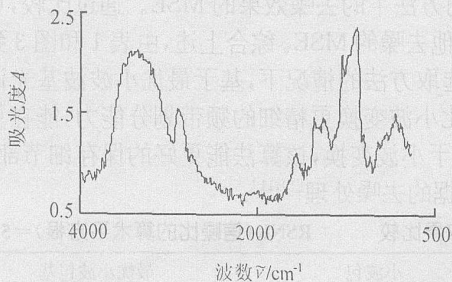


图 3 软阈值下小波消噪光谱图

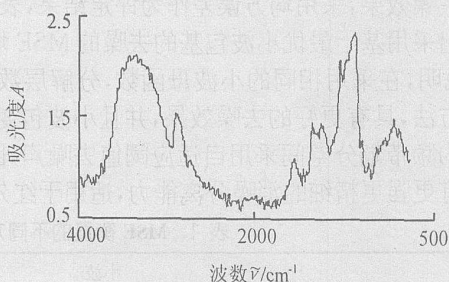


图 4 硬阈值下小波消噪光谱图

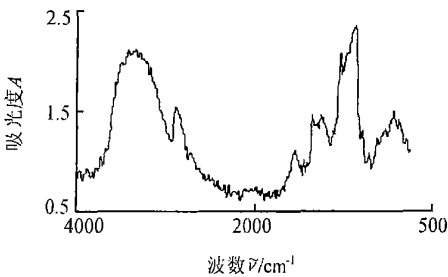


图 5 软阈值下小波包消噪光谱图

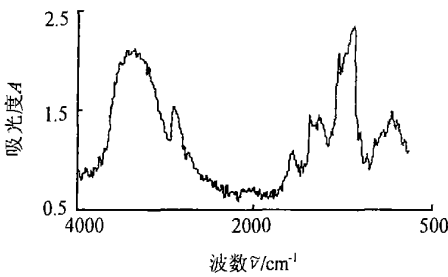


图 6 硬阈值下小波包消噪光谱图

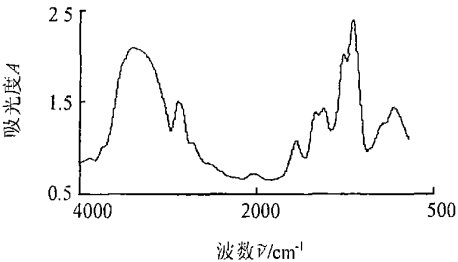


图 7 软阈值下最优基消噪光谱图

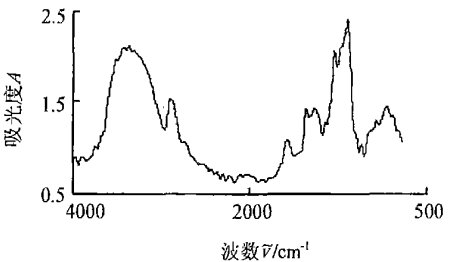


图 8 硬阈值下最优基消噪光谱图

4.2 去噪效果评估

经过小波包去噪处理后的信号与原始信号的均方误差(Mean Squared Error, MSE), 可以作为衡量算法好坏的标准, 定义如下^[10]:

$$MES = \frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2 \tag{4}$$

其中: N —— 离散信号的总波长数; y_i —— 标准信号在 i 波长处的信号值; x_i —— 去噪后的信号在 i 波长处的信号值。在试验中, 采用小波基函数 sym6, 其分辨率较强, 小波包分解尺度为 5, 阈值的选取遵循如前所述的方法, 对含有噪声的信号, 分别采用小波去噪、小波包消噪方法和最优小波包基方法进行去噪处理, 各种方法消噪后的红外光谱图分别如图 3 到图 8 所示。其中图 3 和图 6 分别是采用相同分解层数和阈值选取方法的 sym6 小波和小波包去噪后的光谱图, 图 7 和图 8 是基于 shannon 熵准则的 sym6 最优小波包去噪后的光谱图。比较它们的消噪效果, 可以看出含噪声的光谱图, 采用文中算法进行去噪处理, 较其他算法的去噪效果相比, 变得更为光滑。为了更清楚的去噪效果, 采用均方误差作为评定结果, 表 1 为不同方法下的去噪效果 MSE。通过比较, 可以看出采用基于最优小波包基的去噪的 MSE 均小于其他去噪的 MSE。综合上述, 由表 1 和图 3 到图 8 说明, 在采用相同的小波母函数, 分解层数和阈值选取方法的情况下, 基于最优小波包基光谱去噪方法, 具有更好的去噪效果, 并且小波包变换具有比小波变换更精细的频带划分能力, 能在更精细的频带划分空间采用自适应阈值去噪声, 因而相对于小波变换, 该算法能更好的保存细节部分, 具有更强更精细的消噪分离能力, 适用于红外光谱数据的去噪处理^[10, 11]。

表 1 MSE 衡量的不同方法去噪效果比较 RSNR(信噪比的算术平方根) = 5

	小波	小波包	最优小波包基
软阈值	2.24	2.04	1.1
硬阈值	2.45	2.18	1.44

5 结束语

采用基于 shannon 准则的最优小波包基的阈值处理算法, 对含有噪声的近红外光谱进行去噪处理, 在消噪阈值的选取问题上, 对于不同频段采用不同的阈值算法, 以降低原始信号与降噪后信号之间的均方误差(MSE)为目标, 取得了较好的去噪结果. 实验结果表明, 基于最优小波包基的信号消噪效果比传统小波方法和普通小波包方法都有所改进, 失真度小, 重现性好, 提高了红外光谱信号的信噪比。

参考文献

- [1] 陆婉珍, 袁洪福, 徐广通, 强冬梅. 现代近红外光谱分析技术[M]. 北京: 中国石化出版社, 2000.
- [2] 高荣强, 范世福, 严衍禄, 赵丽丽. 近红外光谱的数据预处理的研究[J]. 光谱学与光谱分析, 2004, **24**(12): 1563—1565.
- [3] Bjorn K A, Andrew M W, Michael K W *et al*. Wavelet Denoising of Infrared Spectra[J]. *Analyst*, 1997, **122**(10): 645—652.
- [4] 吴新生, 谢益民. 近红外漫反射光谱的小波变换的去噪方法[J]. 光谱实验室, 2005, **22**(3): 656—659.
- [5] 彭玉华. 小波变换与工程应用[M]. 北京: 科学出版社, 2003.
- [6] Stéphane Mallat 著. 信号处理的小波导引[M]. (第2版). 杨利华译. 北京: 机械工业出版社, 2002.
- [7] 韩璞, 张君, 董泽, 潘筱. 汽轮机振动信号的最优小波包基消噪与检测[J]. 中国动力工程学报, 2005, **25**(1): 92—96.
- [8] 戴朝华, 朱云芳, 陈维荣. 一种新的小波消噪阈值选取方法[J]. 设计与研究, 2005, **6**: 33—35.
- [9] 戎丽霞. 基于最优小波包基的信号去噪算法[J]. 德州学院学报, 2005, **21**(2): 72—75.
- [10] 刘刚, 屈梁生. 自适应阈值选择和小波消噪方法研究[J]. 信号处理, 2002, **18**(6): 509—512.
- [11] 胡昌华, 张军波, 夏军, 张伟. 基于 MATLAB 的系统分析与设计—小波分析[M]. 陕西: 西安电子科技大学出版社, 1999.

Study on Denoising in Pretreatment of Infrared Spectrum

LIU Zhi-Hong^{a, b} DENG Bo^a ZHOU Yu-Rong^a PANG Xiao-Feng^a ZHANG Xiao-Chun^b

a(School of Life Science and Technology UESTC of China, Chengdu 610054, P. R. China)

b(School of Electronic CUIT of Chengdu, Chengdu 610051, P. R. China)

Abstract IR spectroscopic data of samples are confused by a series of noise, which greatly influences the achievement of accurate analytical result. An algorithm based on best wavelet packet groups with least cost principle was used to deal with the spectrum data in high and low frequency regions by different methods. The denoised signal indicates that the algorithm can increase the accuracy of spectral analysis, smooth the spectra perfectly, and eliminate noises obviously.

Key words Denoising, Wavelet Packet Transforming, Best Wavelet Packet Groups, Infrared Spectrum.

本刊可上网查阅

由于本刊在2001—2005年被《中国核心期刊(遴选)数据库》收录, 全文上网, 因此, 读者、作者均可直接上网查阅。网址:

<http://www.periodicals.net.cn>

<http://www.wanfangdata.com.cn>

<http://gpsys.periodicals.net.cn>

<http://gpsc.chinajournal.net.cn>