

CEEMDAN 去噪在拉曼光谱中的应用研究

韩庆阳 孙 强 王晓东 李丙玉 高 群

中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033

摘要 提出一种基于自适应噪声的完备经验模态分解(CEEMDAN)拉曼光谱去噪方法。利用排列熵(PE)作为代表噪声的内在模式分量(IMF)的判据,并对其进行阈值滤波以消除噪声,得到消除噪声的拉曼光谱重构信号。采用该方法对乙醇水溶液的拉曼光谱数据进行去噪研究。结果表明,该方法有效地消除了拉曼光谱中的噪声;与经验模态分解法(EMD)和总体平均经验模态分解法(EEMD)的对比表明,CEEMDAN 去噪法不但在高信噪比的拉曼光谱信号去噪中取得良好的去噪效果,而且在低信噪比的拉曼光谱信号去噪中依然优势明显。

关键词 光谱学; 基于自适应噪声的完备经验模态分解; 排列熵; 拉曼光谱; 噪声

中图分类号 O657.37

文献标识码 A

doi: 10.3788/LOP52.113003

Application of CEEMDAN in Raman Spectroscopy Denoising

Han Qingyang Sun Qiang Wang Xiaodong Li Bingyu Gao Qun

Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun, Jilin 130033, China

Abstract A Raman spectroscopy denoising method based on complete ensemble empirical mode decomposition with adaptive noise is proposed. Permutation entropy is employed to judge intrinsic mode functions (IMFs), if they are on behalf of noise. The IMFs should be filtered based on threshold value. Then the reconstructed signal is without noise. This method is used to denoise the Raman spectra of ethanol solution. The result shows that the method can effectively remove noise from Raman spectroscopy. Compared with empirical mode decomposition and ensemble empirical mode decomposition, the method does not only achieve better performance at high signal to noise ratio (SNR), but also has obvious superiority at low SNR.

Key words spectroscopy; complete ensemble empirical mode decomposition with adaptive noise; permutation entropy; Raman spectroscopy; noise

OCIS codes 300.6330; 300.6450; 070.4340

1 引 言

拉曼光谱具有分析效率高、样品少、携带信息丰富等优点,在生物化学、高分子、矿物检测和半导体材料等领域具有广泛应用^[1-5]。拉曼光谱中的噪声有两类,一类是测量仪器本身的热噪声,另一类是外部通信系统的干扰^[6]。这些噪声影响测量得到的拉曼光谱真实信息的解析,因此须从光谱数据中剔除。拉曼光谱有效信息主要集中在高频段,所以拉曼光谱去噪应多关注高频段^[6]。

常见的拉曼光谱去噪方法有傅里叶变换和小波变换等^[7]。傅里叶变换无法从细节上识别高频噪声和高频信号;小波变换则需要选择小波基函数和分解层数,去噪效果受参数选择的影响,是非自适应的^[7-8]。李卿等^[9]提出经验模态分解(EMD)拉曼光谱去噪法,该方法自适应地将信号分解,通过阈值滤波的方法消除噪声。由于EMD分解存在模态混叠的问题,滤波效果比小波变换差。赵肖宇等^[10]提出基于总体平均经验模态分解(EEMD)的拉曼光谱去噪法,克服了EMD分解模态混叠的问题,在信噪比(SNR)大于10 dB时取得了与小波变换相似的效果,但相对小波变换,EEMD更加简便,自适应性更强。然而,EEMD分解后的内在模式分量(IMF)噪声有残留,并且每次需选择加入不同幅值的白噪声;基于自适应噪声的完备经验模态分解

收稿日期: 2015-05-15; 收到修改稿日期: 2015-06-19; 网络出版日期: 2015-10-25

基金项目: 吉林省科技厅项目(201026015, 20125092)、长吉图开发开放先导区科技创新合作专项(2011CJT0004)

作者简介: 韩庆阳(1988—),男,硕士,助理研究员,主要从事光电信号检测及信号处理等方面的研究。

E-mail: lanbohesky@126.com

(CEEMDAN)过程中通过对加入的高斯白噪声进行EMD分解获得IMF分量,代替EEMD分解过程中每次加入的高斯白噪声,实现自适应加入噪声,且残留的噪声较少^[11]。因此,提出应用CEEMDAN结合检测时间序列的随机性排列熵(PE)的方法,排列熵用于判断CEEMDAN分解得到的IMF分量是否代表噪声,对代表噪声的分量阈值滤波,最后进行信号重构,就得到消除噪声的拉曼光谱。

本文研究了基于CEEMDAN的拉曼光谱去噪方法,并将该方法应用于乙醇水溶液的拉曼光谱中,详细讨论了判断代表噪声的IMF分量的方法,并与EMD和EEMD去噪方法进行了对比分析。

2 算法与原理

与EEMD方法相同,在原始拉曼光谱信号 $x(\lambda)$ 中添加不同幅值高斯白噪声 $\omega(\lambda)$,对进行EMD分解得到的IMF分量取均值,得到的IMF分量就是第一个IMF分量,记为 $\widetilde{C}_{\text{IMF}_1}(\lambda)$;则第一个剩余分量为

$$r_1(\lambda) = x(\lambda) - \widetilde{C}_{\text{IMF}_1}(\lambda), \quad (1)$$

定义 $E_j(\cdot)$ 为对信号进行EMD分解操作后的第 j 个IMF分量,信号 $x(\lambda)$ 的第二个IMF分量获取方式为

$$\widetilde{C}_{\text{IMF}_2}(\lambda) = \frac{1}{n} \sum_{m=1}^n E_1\{r_1(\lambda) + \varepsilon_1 E_1[\omega(\lambda)]\}, \quad (2)$$

其中 $1 \leq m \leq n$, n 为信号的采样点,如此类推第 k 个剩余分量为

$$r_k(\lambda) = r_{k-1}(\lambda) - \widetilde{C}_{\text{IMF}_k}(\lambda), \quad (3)$$

则第 $k+1$ 个IMF分量为

$$\widetilde{C}_{\text{IMF}_{k+1}}(\lambda) = \frac{1}{n} \sum_{m=1}^n E_1\{r_k(\lambda) + \varepsilon_k E_k[\omega(\lambda)]\}, \quad (4)$$

分解直到剩余分量为单调函数,即不满足EMD分解条件时,分解停止。最后,拉曼光谱信号 $x(\lambda)$ 表示为

$$x(\lambda) = \sum_{k=1}^K \widetilde{C}_{\text{IMF}_k}(\lambda) + R(\lambda), \quad (5)$$

式中 λ 为拉曼光谱的波长, $\widetilde{C}_{\text{IMF}_k}(\lambda)$ ($0 \leq k \leq K$)为分解得到的IMF分量, $R(\lambda)$ 为剩余分量。这样,加入一次白噪声就能实现信号的分解。信号的排列熵代表信号的随机性大小,其值越大说明信号越随机,越小说明信号越规则,一般白噪声的排列熵大于0.9,噪声属于间歇性信号,其排列熵大于0.6^[12]。计算每个IMF分量的排列熵,排列熵计算方法参见文献[12],选取阈值,将代表噪声的IMF分量 $\widetilde{C}_{\text{IMF}_i}(\lambda) \sim \widetilde{C}_{\text{IMF}_L}(\lambda)$ 按照软阈值滤波方法进行滤波,公式为

$$\widetilde{C}'_{\text{IMF}_i}(\lambda) = \begin{cases} \text{sign}[\widetilde{C}_{\text{IMF}_i}(\lambda)] \left[|\widetilde{C}_{\text{IMF}_i}(\lambda)| - \alpha V_{\text{TH}_i} \right], & |\widetilde{C}_{\text{IMF}_i}(\lambda)| \geq V_{\text{TH}_i} \\ 0, & |\widetilde{C}_{\text{IMF}_i}(\lambda)| \leq V_{\text{TH}_i} \end{cases} \quad 0 \leq \alpha \leq 1, \quad (6)$$

其中 $1 \leq i \leq L \leq K$, 阈值 V_{TH_i} 计算公式为

$$V_{\text{TH}_i} = \sigma_i \sqrt{2 \ln n}, \quad (7)$$

式中 n 为IMF信号的长度, σ_i 为噪声的标准差,

$$\sigma_i = \frac{\text{median}(V_{\text{TH}_i})}{0.6745}. \quad (8)$$

得到滤波后的IMF分量 $\widetilde{C}'_{\text{IMF}_i}(\lambda) \sim \widetilde{C}'_{\text{IMF}_L}(\lambda)$ 信号重构:

$$x'(\lambda) = \sum_{i=1}^L \widetilde{C}'_{\text{IMF}_i}(\lambda) + \sum_{j=L+1}^K \widetilde{C}_{\text{IMF}_j}(\lambda) + R(\lambda), \quad (9)$$

式中 $1 \leq i \leq L$ 且 $L+1 \leq j \leq K$,得到的 $x'(\lambda)$ 为消除噪声的拉曼光谱信号。

该算法的步骤为:

- 1) 按照(1)~(5)式的方法对拉曼光谱信号进行CEEMDAN分解;
- 2) 计算分解后的IMF分量的PE值,选择阈值;
- 3) 对代表噪声的IMF分量按(6)式进行阈值滤波;
- 4) 按照(9)式进行信号重构,得到消除噪声后的拉曼光谱信号。

3 实 验

采用美国海洋光学(Ocean Optics)公司生产的 QE65000 拉曼光谱仪和 Laser-785 拉曼光谱激发激光器以及 RIP-RPS-785 光纤探头搭建测量系统,在常温 25 °C 的实验环境中测量由无水乙醇(分析纯)和去离子水配制的浓度(体积分数)为 65% 的乙醇水溶液的拉曼光谱数据,如图 1(a)所示。向该数据中加入 20 dB 的白噪声,得到含有噪声的乙醇水溶液拉曼光谱数据,如图 1(b)所示:

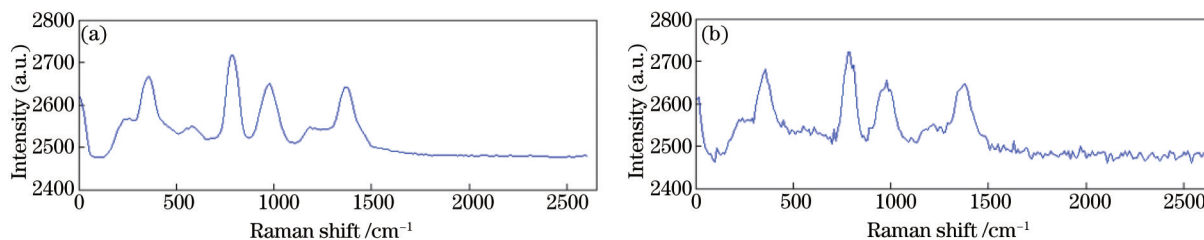


图 1 (a) 原始拉曼光谱信号; (b) 加入 20 dB 白噪声后的信号

Fig.1 (a) Original Raman spectral signal; (b) signal with 20 dB white noise

利用 CEEMDAN 分解乙醇水溶液拉曼光谱,如图 2 所示,从图中可以看出,高频噪声主要分布在 IMF1 内,IMF2 在 10 cm^{-1} 处存在拉曼光谱信息,在其他波长处更多的是高频噪声,IMF3 的主要成分为光谱信息。

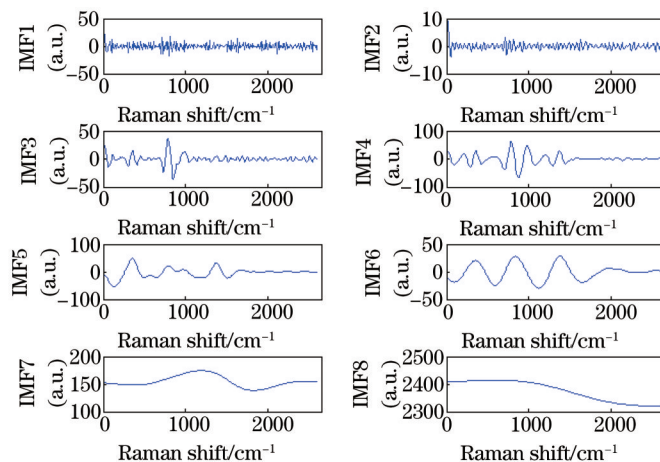


图 2 CEEMDAN 分解得到的 IMF 分量

Fig.2 IMFs after CEEMDAN decomposition

计算各 IMF 分量的 PE,PE 值越大说明该 IMF 分量越随机,反之越规则;代表高频噪声的 IMF 分量的排列熵值比代表信号的排列熵值大,故通过选取合适的阈值来判断,本文选取阈值范围为 0 ~ 0.6,则前三层代表高频噪声。

表 1 各 IMF 分量的排列熵

Table 1 PE values of IMFs

No.	PE
IMF1	0.873
IMF2	0.728
IMF3	0.728
IMF4	0.506
IMF5	0.402
IMF6	0.267
IMF7	0.179
IMF8	0.127

对前三层 IMF 分量进行阈值滤波,得到滤波后的 IMF 分量,如图 3 所示。

由图 3 可以看出,IMF1 滤波后没有剩下任何信息,证明其主要成分为光谱噪声,IMF2 分布在 0~10 cm^{-1} 的光谱信息被保留,而 10~2500 cm^{-1} 之间的噪声得以消除,IMF3 在 750~900 cm^{-1} 之间有光谱信息被保留下

来。将滤波后的IMF1~IMF3分量与IMF4~IMF8分量进行信号重构,得到的信号即为消除噪声后的拉曼光谱信号。

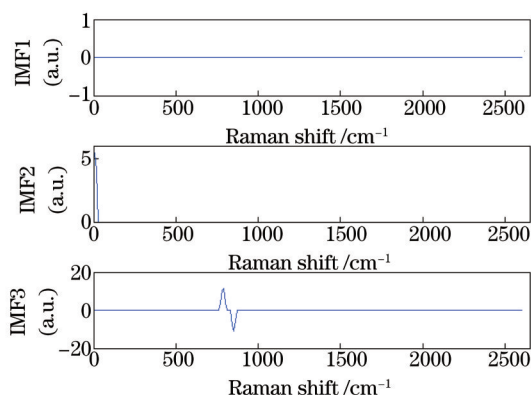


图3 滤波后的前三层IMF分量

Fig.3 First three IMFs after filtering

4 结果与讨论

4.1 实验结果

将信号重构后的信号与原始拉曼光谱信号和加入噪声的拉曼光谱进行对比,如图4所示。

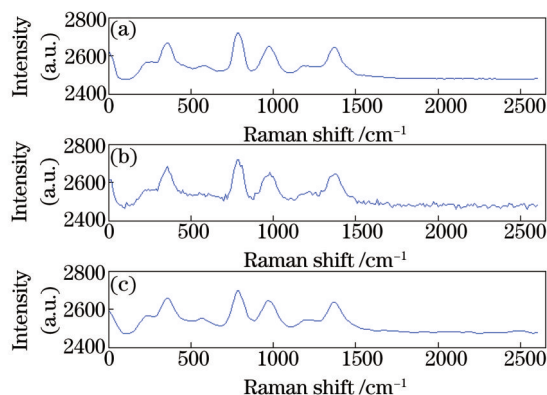


图4 (a) 原始光谱; (b) 加入噪声的拉曼光谱; (c) CEEMDAN滤波后的光谱

Fig.4 (a) Original signal; (b) signal with 20 dB white noise; (c) signal after CEEMDAN filtering

从图4(b)、(c)可以看出拉曼光谱中的噪声被消除,从图4(a)、(c)可以看出,滤波后的拉曼光谱信号与原始信号相差很小,证明滤波效果明显。为了进一步考察该方法的滤波效果,采用信号的频谱作为评价工具。

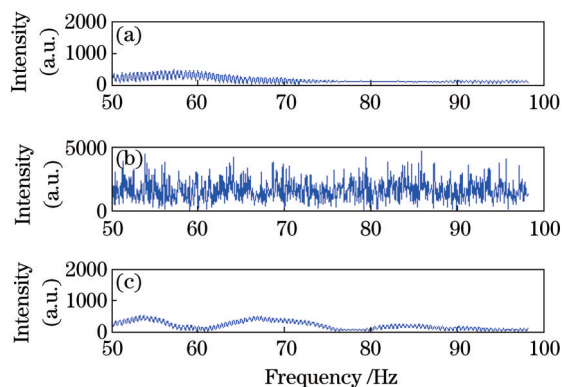


图5 (a)原始拉曼光谱的频谱; (b)加入噪声光谱的频谱; (c)滤波后光谱的频谱

Fig.5 Frequency spectra of (a) original signal; (b) signal with 20 dB white noise; (c) signal after CEEMDAN filtering

图5为50~100 Hz的原始拉曼光谱信号、加入噪声的拉曼光谱信号和滤波后的拉曼光谱信号的频谱,由图5(b)、(c)可以看出加入的噪声被滤除,由图5(a)、(c)可以看出,即使噪声被消除,但是相对原始信号还是

有差别,不能完全还原原始信号。

4.2 讨论

为了进一步直观评价该方法的滤波效果,分别与 EMD 和 EEMD 两种滤波方法的滤波结果进行对比。

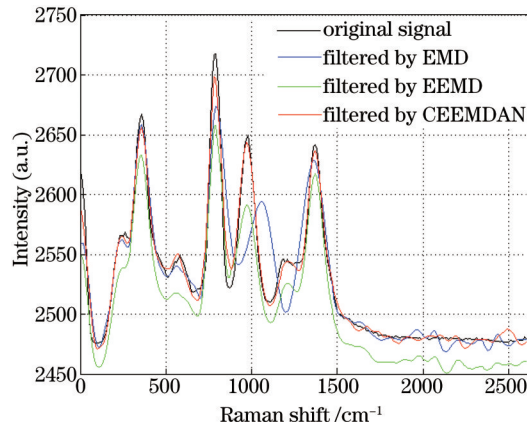


图6 原始拉曼光谱及三种方法滤波后的拉曼光谱

Fig.6 Original signal and signal after filtering by three methods

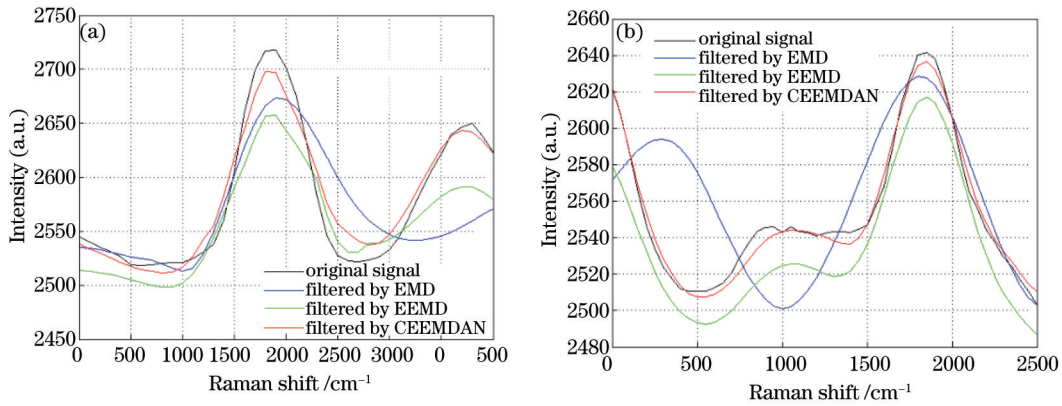


图7 原始拉曼光谱及三种方法滤波后的拉曼光谱局部放大图。(a) 原始光谱; (b) 滤波后光谱

Fig.7 Partially enlarged Raman spectra of original signal and signal after filtering by three methods.

(a) Original signal; (b) signal after filtering

图6为三种方法滤波后的拉曼光谱及原始拉曼光谱的对比,可以看出EMD的滤波效果是三种方法中最差的。图7(a)、(b)分别为光谱在0~500 cm⁻¹和0~2500 cm⁻¹波段内乙醇水溶液的拉曼光谱局部放大图,从图7(a)可以看出,EMD和EEMD滤波后的拉曼光谱特征峰偏移明显,其中,EMD滤波后光谱数据失真很大;而CEEMDAN与原始信号偏移较小;图7(b)中EMD滤波后的拉曼光谱在1000 cm⁻¹处的吸收峰消失,数据失真较大,正确率很低,EEMD滤波后的拉曼光谱特征峰偏移较CEEMDAN明显。

三种方法的滤波效果还可以通过更为客观的方法进行评价,即SNR、均方根误差(RMSE)和相关系数 r 。SNR表示拉曼光谱信号与噪声能量的大小关系, RMSE表明去噪信号与原始信号之间在数值上的差异性, r 描述原始信号和去噪信号在波形上的相似程度^[4],计算公式为

$$R_{\text{SN}} = 10 \log \left\{ \frac{\sum_n [x'(n)]^2}{\sum_n [x(n) - x'(n)]^2} \right\}, \quad (10)$$

$$f_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_n [x(n) - x'(n)]^2}, \quad (11)$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(x'_i - \bar{x}')}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (x'_i - \bar{x}')^2}}, \quad (12)$$

式中 x 为原始光谱信号, x' 为去噪后的拉曼光谱信号; $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{x}' = \frac{1}{n} \sum_{i=1}^n x'_i$ 。

表2 加入 20 dB 噪声的滤波效果

Table 2 Results of filtering 20 dB noise

	R_{SN} /dB	f_{RMSE}	r
EMD	11.115	211.362	0.7821
EEMD	17.553	188.162	0.8957
CEEMDAN	37.986	61.373	0.9389

表2为加入 20 dB 噪声时三种方法的滤波效果。CEEMDAN 滤波后的信噪比最高,比 EMD 提高了 26.871 dB,比 EEMD 提高了 20.433 dB;三种方法中 CEEMDAN 的均方误差最小,比 EMD 减小了 71.0%,比 EEMD 减小了 67.4%;CEEMDAN 的相关系数比 EMD 高 20.0%,比 EEMD 高 4.82%;由此足见,CEEMDAN 的滤波效果远好于 EMD 和 EEMD。EMD 滤波效果在三种方法中最差,原因在于分解过程中的模态混叠致使滤波后的重构信号相对原始信号失真较大。

表3 加入 1 dB 噪声的滤波效果

Table 3 Results of filtering 1 dB noise

	R_{SN} /dB	f_{RMSE}	r
EMD	5.61	348.159	0.7421
EEMD	9.55	298.040	0.8514
CEEMDAN	14.08	250.141	0.9021

表3为加入 1 dB 噪声时三种方法的去噪效果。可以看出三种方法的效果均比加入 20 dB 噪声时差,去噪后的拉曼光谱和原始光谱在数值上的差异性比加入 20 dB 噪声时大,相关性小。但是,CEEMDAN 相对于 EMD 和 EEMD 依然保持优势。

5 结 论

针对拉曼光谱的信号与噪声有交叠的特点,提出应用 CEEMDAN 与排列熵相结合的方法消除拉曼光谱的噪声,同时与 EMD 和 EEMD 去噪法进行了对比分析。实验结果表明,无论是对高信噪比还是低信噪比的拉曼光谱信号,该方法均有效地消除了噪声,提高了信噪比。该方法优于 EMD 和 EEMD。尽管基于经验模态分解的信号处理方法还存在着很多问题需要解决,但相对传统方法的优势明显,未来必将取代传统信号处理方法。

参 考 文 献

- Liu Bing, Yu Fanju, Sun Qiang, *et al.*. New approach of stray light suppression for probe system of handheld Raman spectrometer[J]. Chinese J Lasers, 2014, 41(1): 0115001.
刘 兵, 于凡菊, 孙 强, 等. 手持式拉曼光谱仪探头系统的杂光抑制新方法[J]. 中国激光, 2014, 41(1): 0115001.
- Qin Zhaojun, Lai Junzhuo, Liu Bin, *et al.*. Raman spectroscopy analysis of ethanol fermentation at various initial pH levels[J]. Chinese J Lasers, 2013, 40(2): 0215001.
覃赵军, 赖钧灼, 刘 斌, 等. 不同初始 pH 值得乙醇发酵过程拉曼光谱分析[J]. 中国激光, 2013, 40(2): 0215001.
- Ma Jing. Low-concentration detection of chlorobenzene based on laser Raman spectroscopy[J]. Chinese J Lasers, 2014, 41(2): 0215001.
马 靖. 基于激光拉曼光谱的氯苯低浓度探测[J]. 中国激光, 2014, 41(2): 0215001.
- Xiong Yang, Si Minzhen, Gao Fei, *et al.*. Study on cervical cancer oxyhemoglobin using near-infrared surface-enhanced Raman spectroscopy[J]. Chinese J Lasers, 2015, 42(1): 0115001.
熊 洋, 司民真, 高 飞, 等. 基于 NIR-SERS 光谱技术分析宫颈癌氧合血红蛋白[J]. 中国激光, 2015, 42(1): 0115001.
- Qin Zhaojun, Peng Lixin, Zhu Libo, *et al.*. Raman spectral profiles of PHB synthesis by *Cupriavidus necator* H16 at different fructose levels[J]. Chinese J Lasers, 2015, 42(3): 0315003.
覃赵军, 彭立新, 竺利波, 等. 碳源浓度影响微生物 PHB 合成代谢的单细胞拉曼光谱分析[J]. 中国激光, 2015, 42(3): 0315003.

- 6 Chu Xiaoli. Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications[M]. Beijing: Chemical Industry Press, 2011: 311-356.
褚小立. 化学计量学方法与分子光谱分析技术[M]. 北京: 化学工业出版社, 2011: 311-356.
- 7 Liu Xiaofang, Wang Hangong, Quan Gaofeng, *et al.*. Application of wavelet transform to detect the waveform of electrochemical noise[J]. Chinese Journal of Analytical Chemistry, 2001, 29(2): 161-164.
刘晓方, 王汉功, 权高峰, 等. 利用小波变换检测电化学噪声信号波形[J]. 分析化学, 2001, 29(2): 161-164.
- 8 Wang Juxiang, Li Hua, Xing Zhina, *et al.*. Application of wavelet multi-scale piecewise direct standardization on near infrared analysis calibration[J]. Chinese Journal of Analytical Chemistry, 2011, 39(6): 846-850.
王菊香, 李 华, 邢志娜, 等. 小波多尺度分段直接校正法用于近红外光谱模型传递的研究[J]. 分析化学, 2011, 39(6): 846-850.
- 9 Zhao Xiaoyu, Fang Yiming, Wang Zhigang, *et al.*. EEMD denoising adaptively in Raman spectroscopy[J]. Spectroscopy and Spectral Analysis, 2013, 33(12): 3255-3258.
赵肖宇, 方一鸣, 王志刚, 等. EEMD 自适应去噪在拉曼光谱中的应用[J]. 光谱学与光谱分析, 2013, 33(12): 3255-3258.
- 10 Li Qing, Zhang Guoping, Liu Yang. A study of Raman spectra denoising based on empirical mode decomposition[J]. Spectroscopy and Spectral Analysis, 2009, 29(1): 142-145.
李 卿, 张国平, 刘 洋. 基于 EMD 的拉曼光谱去噪方法研究[J]. 光谱学与光谱分析, 2009, 29(1): 142-145.
- 11 M A Colominas, G Schlotthauer, M E Torres. Improved complete ensemble EMD: A suitable tool for biomedical signal processing[J]. Biomedical Signal Processing & Control, 2014, 14: 19-29.
- 12 Zheng Dejin, Cheng Junsheng, Yang Yu. Modified EEMD algorithm and its applications[J]. Journal of Vibration and Shock, 2013, 32(21): 22-46.
郑德近, 程军圣, 杨 宇. 改进的 EEMD 算法及其应用研究[J]. 振动与冲击, 2013, 32(21): 22-46.

栏目编辑: 吴秀娟