

基于集合经验模态分解的拉曼光谱信号特征研究

李明^{1,2}, 赵迎^{1,2}, 崔飞鹏², 刘佳²

1. 钢铁研究总院, 北京 100081

2. 钢研纳克检测技术股份有限公司, 北京 100094

摘要 拉曼光谱信号是一种基于分子振动的散射信号, 拉曼光谱仪的激光源波长一般为纳米级, 考虑到散射频移, 拉曼光谱有效信息主要集中在较高频段。拉曼信号是典型的非平稳信号, 并且由于拉曼散射比较弱, 信号很容易被高频噪声和荧光背景干扰, 想获取较为全面的拉曼信息, 需要对信号进行处理, 小波变换对拉曼信号的分析结果取决于小波基的选择, 不同小波基处理结果有差异; 经验模态分解(EMD)方法可以自适应的分析信号, 不需要设置参数, 但存在模态混叠的问题; 集合平均经验模态分解(EEMD), 有效的解决了EMD方法中存在的模态混叠问题, 能更加清晰的将信号中的不同频率成分分离开来, 因此更加适合频率成分丰富的拉曼信号的特征分析和处理。采集了市面上常见的大豆油、花生油、玉米油和葵花籽油样本, 通过拉曼光谱仪获得了各自的拉曼光谱信号。使用集合经验模态分解对食用油拉曼光谱信号进行自适应分解和处理, 一共获得了10阶固有模态函数(IMF), 根据信号的能量分布以及幅值特性, IMF1和IMF2表征为信号中的噪声部分, IMF3-IMF7表征为拉曼特征信号部分, 最后一阶IMF10表征为荧光背景成分, IMF8和IMF9为其他物理意义的频率成分。通过对有效信号段的特征增强并重构拉曼信号, 使拉曼信号的信噪比获得了2~5倍的提升, 其中, 难以探测的酯键羰基伸缩振动位于 1745 cm^{-1} 的谱峰得到了显著的增强。最后, 将原始信号和经过特征增强的信号通过基于连续小波变换的惩罚最小二乘法进行了二次处理, 并将获得的信号进行主成分分析后, 可知: 没有增强的不同类数据样本相互有重叠, 不存在明显的类间距, 很难完整的区分类型; 基于特征增强的数据样本各自聚集, 每种类型都可以相互鉴别, 可为拉曼光谱信号处理提供一种新的途径。

关键词 信号特征增强; EEMD方法; 拉曼光谱; 食用油分析

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)01-0054-05

引言

拉曼光谱中的噪声主要来自电子元件随机热运动而造成的电子噪声, 在谱图中表现为高频频段; 外部杂散光以及样品在入射光的作用下产生的荧光或其他杂散光, 在谱图中表现为低频频段。赵肖宇等应用经验模态分解(EMD)和集合平均经验模态分解(ensemble empirical mode decomposition, EEMD)方法处理了含噪光谱信号, 获得了不错的效果, 但仅应用于信号的去噪处理。EEMD方法最大的优势就是根据时频特性将信号本身的特征成分自适应的分解为具有物理意义的特征分量, 本文利用EEMD方法对拉曼信号进行自适应时频特征分解, 得到有物理意义的高频噪声段, 拉曼有效信号段和荧光背景段, 通过增强有效信号并重构拉曼

信号获得了更好的特征谱图^[1], 该方法应用于食用油的分析, 取得了较好的效果, 为拉曼光谱信号处理提供了一种新的方法。

1 理论部分

Huang等^[2]于1998年提出一种信号时频分析方法—EMD, 该方法能自适应的将信号分解成有限个具有物理意义的分量—固有模态函数(intrinsic mode function, IMF)。EMD方法的一个重要缺陷就是模态混叠, 为了克服这个问题, Wu和Huang在对白噪声进行EMD分解深入研究的基础上, 提出了EEMD方法, 其基本思路^[3]为: 将高斯白噪声加入被分析信号, 使信号和噪声成为一个“集合”, 在EMD的分解过程中, 当被分析信号加在这些一致分布的白色背景

收稿日期: 2019-09-18, 修订日期: 2019-11-02

基金项目: 国家重点研发计划项目(重大科学仪器设备专项 2018YFF01012000)资助

作者简介: 李明, 1982年生, 钢铁研究总院高级工程师 e-mail: lmsq@163.com

上时,不同尺度的信号,会自适应的映射到合适的频率范围上;同时,由于白噪声的零均值特性,加入多次白噪声分解,将这些多次分解的结果取“平均”后,噪声最终被相互抵消而消除,即集合平均的分解结果被当做是被分析信号的 EMD 分解结果。

EEMD 方法的基本步骤^[4]为:

步骤 1: 给被分析信号 $x(t)$ 加入一组白噪声 $w(t)$ 来获得一个总体 $X(t)$

$$X(t) = x(t) + w(t) \quad (1)$$

步骤 2: 对 $X(t)$ 进行 EMD 分解,得到相应的各阶 IMF

$$X(t) = \sum_{i=1}^n c_i(t) + r(t) \quad (2)$$

步骤 3: 给被分析信号 $x(t)$ 加入一组不同的白噪声 $w_j(t)$ 来获得另一个总体 $X_j(t)$,对 $X_j(t)$ 进行 EMD 分解后,得到相应的各阶 IMF

$$X_j(t) = \sum_{i=1}^n c_{ji}(t) + r_j(t) \quad (3)$$

步骤 4: 求得各个 IMF 的均值作为最终信号的 IMF

$$c_j(t) = \frac{1}{N} \sum_{i=1}^N c_{ji}(t) \quad (4)$$

式中, N 为加入白噪声的次数,即总体的个数, ϵ 是加入白噪声的幅度, ϵ_n 是原始信号和由最终的 IMF 之和所得信号的误差。

$$\epsilon_n = \frac{\epsilon}{\sqrt{N}} \quad (5)$$

EEMD 方法较之 EMD 方法,有效的克服了模态混叠现象,能更加清晰的将信号中不同频率的成分划分开来,因此更加适合频率成分丰富的拉曼信号的特征分析和处理。

2 实验部分

2.1 食用油样本

本研究用的食用油采购市场常见的品牌,其中大豆油 5 种,花生油 5 种,玉米油 5 种,葵花籽油 5 种,如图 1 所示,所有食用油样品置于不含荧光背景的 5 mL 样品瓶中。

2.2 仪器与参数

实验仪器为美国 Bayspec 公司的 AgilityTM 拉曼光谱仪,如图 2 所示,参数设置:激发波长 785 nm,激光器功率 450 mW,积分时间 1 000 ms,扫描波数范围: 100~2 700 cm^{-1} 。

2.3 方法

实验室环境温度 25 $^{\circ}\text{C}$,每种样品分装在 2 个样品瓶中,直接测量,一共获得的 40 张拉曼光谱原始谱图,如图 3 所示,图示曲线从上到下,1—10 为大豆油拉曼谱图;11—20 为花生油拉曼谱图;21—30 为玉米油拉曼谱图;31—40 为葵花籽油拉曼谱图。

3 结果与讨论

3.1 拉曼信号的 EEMD 方法处理

选取原始拉曼谱图作为信号处理对象,经 EEMD 分解,自适应的得到了 10 个 IMF,如图 4 所示,其特征频率从高到



图 1 食用油样品

Fig. 1 Samples of edible oil



图 2 拉曼光谱仪

Fig. 2 Raman spectrometer

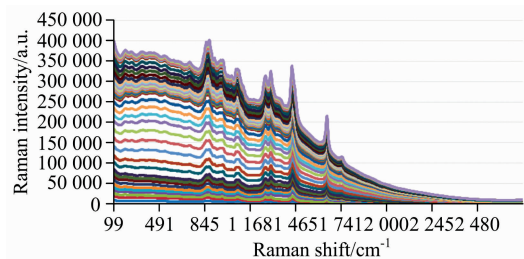


图 3 食用油原始拉曼谱图

Fig. 3 Original Raman spectra of edible oils

低排布, 其中 IMF1 频率最高, IMF10[即信号的趋势 $r(t)$] 最低。根据信号的能量分布以及幅值特性, IMF1 和 IMF2 表征为信号中的噪声部分; IMF3—IMF7 表征为拉曼特征信号部分; 最后一阶 IMF10 表征为荧光背景成分, IMF8 和 IMF9

为其他物理意义的频率成分。由此可见, EEMD 方法可以完整的将拉曼信号中各个有物理意义的特征频率成分分解出来。

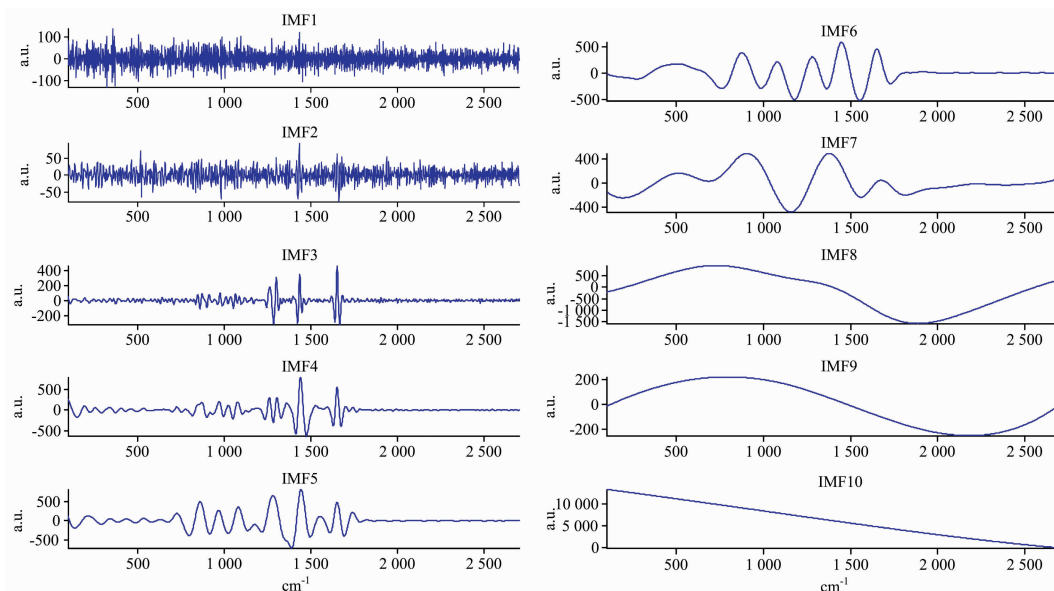


图 4 原始拉曼信号的 EEMD 分解

Fig. 4 EEMD processing of original Raman signal

3.2 拉曼信号的特征重构及增强处理

上述处理结果, IMF1 和 IMF2 为信号中的噪声部分, IMF10 为荧光频率段, 通过增强 IMF3—IMF7 段可有效增加信噪比, 而拉曼信号的噪声及荧光背景不变, 获得特征增强后的拉曼信号 $Y(t)$, 其中 k 定义为增强因子。

$$Y(t) = \text{IMF}_1 + \text{IMF}_2 + k \sum_{i=3}^7 \text{IMF}_i + \text{IMF}_8 + \text{IMF}_9 + \text{IMF}_{10} \quad (6)$$

如图 5 所示, 为特征增强后的拉曼信号和原始拉曼信号, 图中可见, 信号的高频噪声和荧光背景与原始谱图一致, 在 $k=3$ 的有效信号特征增强下获得了信噪比 2~5 倍的提升。

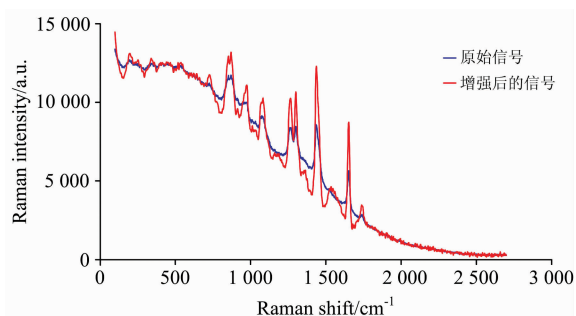


图 5 原始拉曼信号和增强后的拉曼信号对比

Fig. 5 Original Raman signal and the enhanced Raman signal

如图 6 所示, 结合基于连续小波变换的惩罚最小二乘法^[5] (continuous wavelet transform-penalized least squares,

CWT-PLS) 进行进一步处理, 可见有效特征峰均得到不同程度的增强, 其中, 食用油中难以探测的酯键碳基伸缩振动位于 1745 cm^{-1} 的谱峰得到了显著的增强。

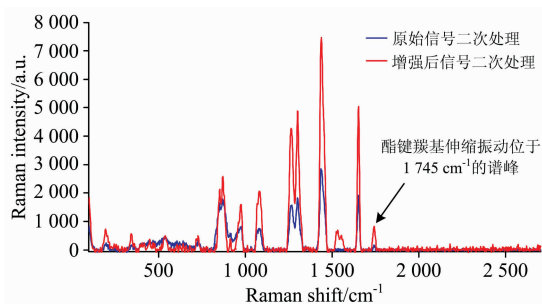


图 6 原始拉曼信号和增强后的拉曼信号二次处理后对比

Fig. 6 Original Raman signal and the enhanced Raman signal after CWT-PLS

3.3 食用油类型 PCA 分析

将试验的拉曼谱图均按上述方法进行信号特征增强后, 获得了 4 种食用油各 10 张增强后的拉曼谱图, 如图 7 所示。

上述数据仅从拉曼谱图上无法鉴别食用油样本类型, 本文结合 PCA(主成分分析)方法后进一步进行判别, 选取信噪比较高拉曼位移为 $300 \sim 1800 \text{ cm}^{-1}$ 的谱段作为 PCA 分析的主要分析数据, 图 8 显示为原始信号以及使用 EEMD 方法的特征增强后的信号经过 CWT-PLS 方法处理后的 PCA 分析结果对比。结果显示: 原始信号直接经过 CWT-PLS 方法处理后获得的结果比较分散, 不同类样本也有重叠, 不存在明显的类间距, 很难完整的区分 4 种类型的食用油; 基于

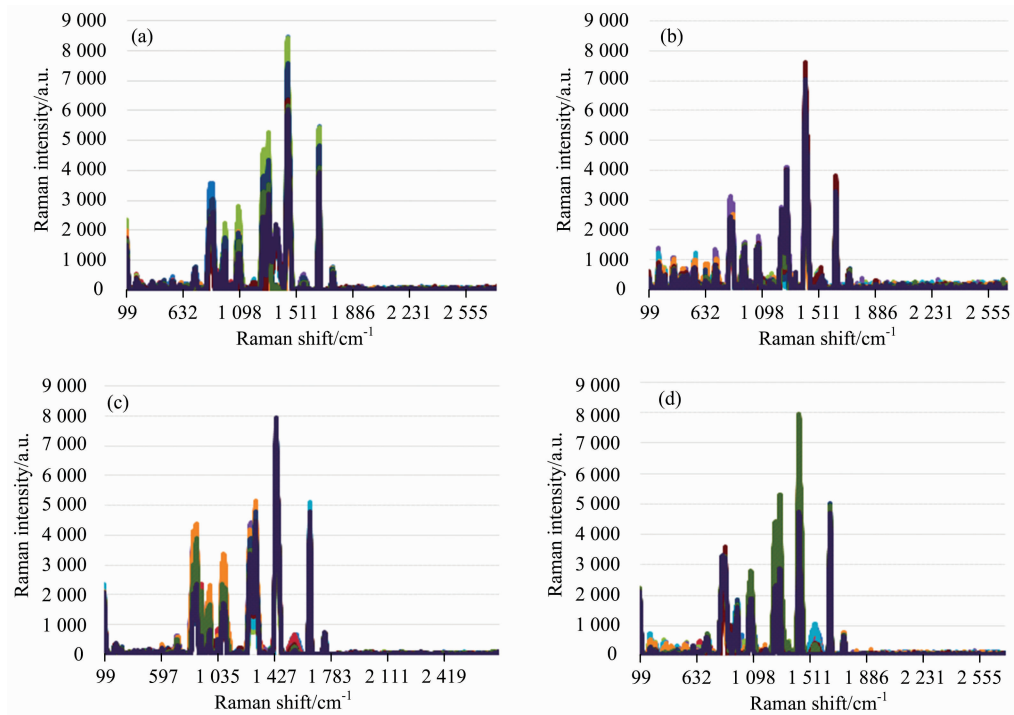


图 7 食用油增强拉曼谱图

(a): 大豆油谱图; (b): 花生油谱图; (c): 玉米油谱图; (d): 葵花籽油谱图

Fig. 7 The enhanced Raman spectra of edible oils

(a): Spectrum of Soybean oil; (b): Spectrum of peanut oil; (c): Spectrum of corn oil; (d): Spectrum of sunflower oil seed

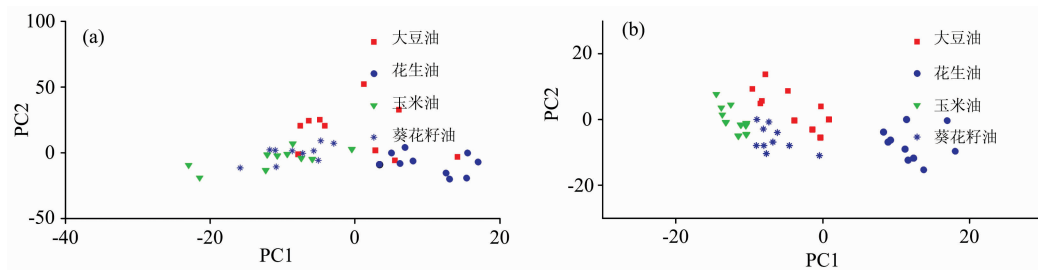


图 8 PCA 结果对比

(a): 原始信号 PCA 结果; (b): 特征增强后的信号 PCA 结果

Fig. 8 The comparison of PCA

(a): PCA results of original signals; (b): PCA results of enhanced signals

EEMD 特征增强后经过 CWT-PLS 方法处理的数据样本各自聚集, 每种食用油样本都可以相互鉴别。

4 结 论

将集合经验模态分解应用于食用油样本的拉曼光谱信号的特征提取, 增强后并重构拉曼光谱信号, 可得如下结论:

(1) 集合经验模态分解方法可自适应的将食用油拉曼光

谱信号分解成 10 阶 IMF, 其 IMF1 和 IMF2 表征为信号中的噪声部分; IMF3—IMF7 表征为拉曼特征信号的部分; 最后一阶 IMF10 表征为荧光背景部分。

(2) 基于 EEMD 的拉曼特征增强方法具有自适应的特点, 不需要预先设置其他参数, 通过应用于食用油拉曼特征增强, 获得了拉曼特征更强的信号, 可以清晰的鉴别 4 种类别的食用油, 为拉曼光谱的数据处理提供了新的方式。

References

- [1] XU Jing-yang, FANG Shao-bo, ZHOU Jing(徐静阳, 方少波, 周 婧). Acta Physica Sinica(物理学报), 2019, (6): 068701.
- [2] Huang N E, Long S R, Wu M C, et al. Proc. R. Soc. Lond., 1998, 454: 903.
- [3] Wu Zhaohua, Huang N E. Technical Report, 2005, 193: 51.
- [4] LI Ming, ZHAO Ying, CUI Fei-peng, et al(李 明, 赵 迎, 崔飞鹏, 等). Metallurgical Analysis(冶金分析), 2019, (5): 57.
- [5] Zhang Zhimin, Chen Shan, Liang Yizeng, et al. Journal of Raman Spectroscopy, 2010, 41: 659.

Research on Raman Spectral Signal Characteristics Based on Ensemble Empirical Mode Decomposition

LI Ming^{1, 2}, ZHAO Ying^{1, 2}, CUI Fei-peng², LIU Jia²

1. Central Iron and Steel Research Institute, Beijing 100081, China

2. NCS Testing Technology Co., Ltd., Beijing 100094, China

Abstract Raman spectrum signal is a kind of scattering signal based on molecular vibration. The laser source wavelength of Raman spectrometer is generally nanometer. As it is a typical non-stationary signal and considering the scattering frequency shift, the effective information of Raman spectrum is mainly concentrated in the higher frequency band. Because Raman scattering is very weak, and the signal is easily disturbed by high frequency noise and fluorescence background. In order to obtain more comprehensive Raman information, the signal needs to be processed. The results of Raman signal analysis by wavelet transform depend on the choice of wavelet bases, and the results of different wavelet bases are different; Empirical Mode Decomposition (EMD) method can analyze signals adaptively without setting parameters, but it has the problem of mode mixing. The Ensemble Empirical Mode Decomposition (EEMD) effectively solves the problem of mode mixing in EMD method, and can more clearly divide the components of different frequencies in signals, so it is more suitable for the characteristic analysis and processing of Raman signal which has rich frequency components. In this paper, Raman spectrum of soybean oil, peanut oil, corn oil and sunflower seed oil samples are collected by Raman spectrometer. Raman spectrum of edible oil are adaptively decomposed and processed by EEMD, and a total of 10 orders Intrinsic Mode Function (IMF) are obtained. According to the energy distribution and amplitude characteristics of the signal, IMF1 and IMF2 are characterized as the noise components of the signal, IMF3—IMF7 as the Raman characteristic signal components, the last order IMF10 as the fluorescence background component, and IMF8 and IMF9 as the frequency components of other physical meanings. After filtering out the high frequency noise components of IMF1 and IMF2, it obtains the Raman signal after de-noising. In addition, the signal-to-noise ratio of Raman signal is increased by 2~5 times by enhancing and reconstructing the characteristics of the effective signal component. Among them, the dynamic peak at $1\ 745\ \text{cm}^{-1}$ caused by the ester bond carbonyl stretching vibration is significantly enhanced, which is difficult to detect. Finally, the baseline of original signal and the characteristic enhancing signal are deducted by PLS method based on continuous wavelet transform. After principal component analysis, the different data samples without enhancement overlap with each other, and there is no obvious class spacing, so it is difficult to distinguish the type of samples completely. The data samples based on feature enhancement are gathered separately, and each kind of data samples is clustered obviously. Types can be identified from each other, which provides a new way for Raman spectroscopic signal processing.

Keywords Signal characteristics enhancement; EEMD method; Raman spectroscopy; Edible oil analysis

(Received Sep. 18, 2019; accepted Nov. 2, 2019)