

一种基于偏微分方程的信号增强模型

李俊¹ 李远禄^{1,2} 蒋民¹

(1. 南京信息工程大学自动化学院, 南京, 210044; 2. 江苏省大气环境与装备技术协同创新中心, 南京, 210044)

摘要: 在峰检测中, 由于重叠峰和弱峰的存在, 峰的漏检和错检率较高。为此本文在峰检测环节之前增加一个峰增强环节, 使重叠峰的分离度以及弱峰幅度均增大。增强环节中采用的方法是将经典非线性扩散与导数谱相结合, 即将导数谱增强后的信号作为经典非线性扩散的初始信号, 经过一定时间的扩散得到增强后的信号。作为效果检验, 首先对比了信号经过所提模型增强前后的效果, 之后对比了其他信号增强模型的效果, 结果表明本文所提模型有效。最后将本文模型应用于MALDI质谱峰增强。

关键词: 非线性扩散; 导数谱; 重叠峰; 信号增强; 峰检测

中图分类号: TP911.72 **文献标志码:** A

A Signal Enhancement Model Based on Partial Differential Equations

Li Jun¹, Li Yuanlu^{1,2}, Jiang Min¹

(1. School of Automation, Nanjing University of Information Science & Technology, Nanjing, 210044, China; 2. Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing, 210044, China)

Abstract: Overlapping peaks and low amplitude peaks in peak detection stage will lead to a high false detection rate. Therefore, a peak enhancement step is added before the peak detection stage to improve the resolution of overlapping peaks and increase the amplitude of the low amplitude peaks. The method in the model is to combine the classical nonlinear diffusion with the derivative spectra. In other word, the signal after the derivative spectrum enhancement is used as the initial signal of the classical nonlinear diffusion, and the enhanced signal is the result of the diffusion. As a test of the proposed mode, the performance of signal enhancement by the proposed model is compared with non-enhancement one, then compared with the performance of other signal enhancement methods. Results show that the proposed model is effective. Finally, the proposed model is applied to the enhancement of MALDI mass spectrometry.

Key words: nonlinear diffusion; derivative spectra; overlapped peak; signal enhancement; peak detection

引 言

峰检测被应用于很多领域, 如心电图、基于质谱分析的癌症诊断、水质分析以及化学混合物鉴定等^[1-2]。由于实验环境和仪器分辨水平等因素的影响, 以及谱峰自身的特点, 会存在部分弱峰, 或者部分重叠峰, 若直接对信号进行峰检测, 容易造成峰的漏检和错检。因此, 如何提高重叠峰的分辨力, 或者对弱峰进行增强是一个关键问题^[3]。

常见的信号增强方法包括Fourier去卷积方法、连续小波方法和导数谱方法。Kauppinen等人提出了Fourier去卷积法^[4];Du等人提出了连续小波的峰检测方法^[5];导数谱方法是增强信号的常用方法^[6],目前被广泛应用于水体分析、乳腺肿瘤检测和心电图检测等^[7-9]。然而导数谱方法在增强谱峰的同时也会降低信号的信噪比,因此通常将导数谱方法和平滑方法相结合使用,如与小波方法结合用于化学信号检测^[10],与Savitzky-Golay方法结合测定红外光谱^[11],与高斯平滑方法结合用于蛋白质谱峰检测^[12]等。

平滑方法较多,主要包括时域和频域两类方法。在时域中,最简单的平滑方法是滑动均值滤波,即将相邻的奇数个点求平均值代替原中心点^[13];Savitzky-Golay滤波目前被广泛使用,该方法是一种广义的滑动均值滤波方法,它将一小组连续数据点做最小二乘拟合,并将多项式拟合曲线的中心点作为输出^[14],相比于滑动均值滤波法,该方法具有更好的保峰效果;高斯滤波是对滑动均值滤波平滑窗口的改进方法,使用高斯函数作为平滑窗口^[15];类似地,还有使用kaiser窗作为平滑窗口的滤波方法^[16]。在频域中,需要先将信号转换到频域,再将高频系数抛弃,然后重构得到平滑信号。小波方法是目前最常用的频域平滑方法,该方法可以较好地保护信号的特征,然而结果与所用小波基和尺度有关,不同尺度下所得结果可能相差很大^[17-18]。

20世纪80年代,Witkin发现齐次线性扩散方程的解等价于一定尺度下初始信号与高斯函数的卷积,因此偏微分方程在信号处理中越来越受到重视^[19]。然而线性扩散滤波在处理信号时没有考虑信号的特征,在去除噪声同时也可能会模糊信号的细节。针对这个问题,Perona和Malik在1990年提出了经典非线性扩散方法,它根据原始信号的特征在信号的不同位置设置不同的扩散强度,既有好的平滑效果,又能保护信号细节特征^[20]。近年来,经典非线性扩散滤波被广泛运用到图像处理中,在保护图像纹理特征方面有很好的作用^[21-23]。

鉴于非线性扩散模型有保护信号细节特征的能力,本文将非线性扩散与导数谱方法结合,得到一种新的信号增强模型。具体步骤是先对原信号应用导数谱方法增强,再将增强后的信号作为非线性扩散模型的初始信号,经扩散后得到平滑的增强信号。

1 基于非线性扩散的导数谱增强模型

经典非线性扩散模型如下^[20]

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = g[u(x,t)] \frac{\partial^2 u(x,t)}{\partial x^2} \\ u(x,0) = f(x) & 0 < x < L \\ u(0,t) = u(L,t) = 0 & 0 < t < T \end{cases} \quad (1)$$

式中: $f(x)$ 为原始信号; $g[u(x,t)]$ 为扩散函数。

导数谱增强模型如下^[6]

$$F = f - cf^{(n)} \quad (2)$$

式中: f 为原始信号; F 为增强后的信号; n 为导数阶次,通常取偶数; c 为增强系数。

将经典非线性扩散模型与导数谱增强模型相结合,得到本文模型为

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = g[u(x,t)] \frac{\partial^2 u(x,t)}{\partial x^2} \\ u(x,0) = f(x) - cf^{(2)}(x) & 0 < x < L \end{cases} \quad (3)$$

式中

$$g[u(x,t)] = \exp \left[- \left(\frac{u(x,t)}{\lambda} \right)^2 \right] \quad (4)$$

在扩散函数中,λ是控制扩散强度的阈值。如果λ取值很大,扩散函数的值趋近于1,这时相当于高斯平滑。当选择合适的λ时,既可以保峰又可降噪。通常λ的值可根据经验给定,当然也可根据公式给定,若期待最高峰处的扩散强度为一个小值ε,那么第k次迭代的λ可设为

$$\lambda(t_k) = \frac{\max(u(x, t_{k-1}))}{\sqrt{-\ln(\epsilon)}} \quad 0 \leq \epsilon \leq 1 \quad (9)$$

当噪声水平较高时,需要在前几次迭代中取较大的ε值,通常情况下取0.9以上;之后的迭代中则可取较小的ε值,从而较好地保护峰的形状。

图1是通过式(8)得到的由两个不同高度的洛伦兹峰组成的模拟信号。对图1中的噪声信号做平滑,设较低峰处的扩散强度为0.1,则较高峰处的扩散强度ε为0,其扩散函数如图2(a)所示,平滑结果如图2(b)所示,平滑后较高峰处仍存在较大噪声;设较高峰处的扩散强度ε为0.1,其扩散函数如图2(c)所示,平滑结果如图2(d)所示,平滑后效果较好。由图2可以看出,对最高峰取不同的扩散强度ε,平滑效果是有差异的。

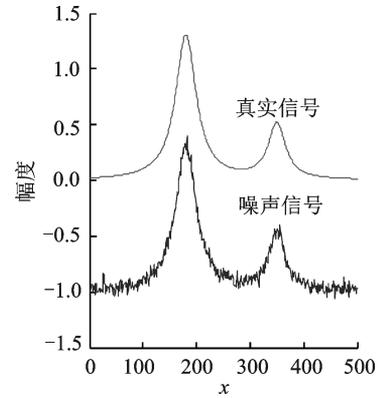


图1 模拟信号
Fig.1 Simulated signal

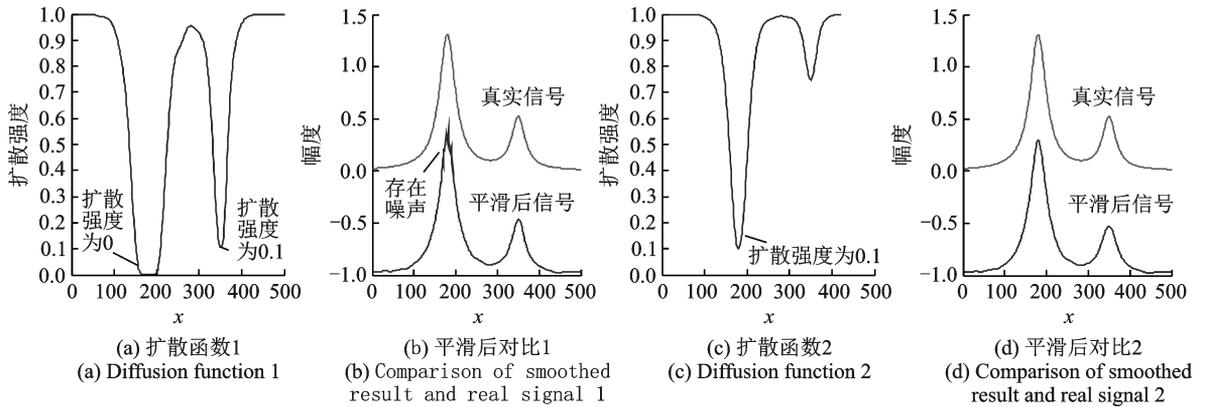


图2 不同情况下扩散函数对平滑结果的影响

Fig.2 Influence of diffusion function on smoothing results under different conditions

2.3 信号产生与实验对比

为了对比增强与未增强的效果,本文通过式(8)的洛伦兹峰模型产生模拟信号,如图3所示。图3(a, b)分别是由8个和12个洛伦兹峰组成的信号,信号中存在弱峰和重叠峰,可以看出噪声信号增强后噪声水平也会被增强。

图4是将图3中2个信号经过本文模型增强后得到的结果。为对比导数谱对峰的增强效果,需要为增强系数取一个合适值,本文对增强系数分别取0(未增强)和其他值时的增强效果进行对比。实验中两个信号取相同的平滑参数,时间步长τ取1,迭代次数取20,前5次最高峰扩散强度ε取0.99,之后取0.05。由图4可以看出本文模型具有很好的峰增强效果,能增强弱峰的幅度和提高重叠峰的分离度。

为更好地表现增强的效果,本文将图4的检测结果用表1表示。通过图4和表1的对比可以看出,只对噪声信号做平滑,其中的弱峰会和噪声容易一起被平滑掉,还容易导致重叠峰,从而影响检测结果;使用本文模型处理后,信号中的弱峰幅度被增大,如信号1中x=150处,x=310处,信号2中x=270

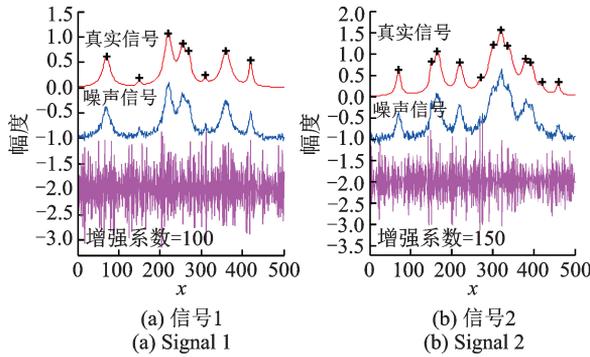


图3 重叠峰真实信号、噪声信号和增强后的噪声信号 (SNR=20 dB)

Fig.3 Real signal, noisy signal and enhanced noisy signal of overlapping peaks (SNR=20 dB)

处, $x=420$ 处的峰, 同时重叠峰的分辨率会提高, 如信号 1 中 $x=270$ 处, 信号 2 中 $x=160$ 处, $x=330$ 处, $x=380$ 处的峰。

2.4 基于非线性扩散的导数谱增强模型与其他模型的对比

导数谱方法常和小波方法(WMD)或 Savitzky-Golay 方法(SGD)结合应用于信号增强, 因此为进一步验证基于非线性扩散的导数谱增强模型(CDD)的效果, 将其与前两种模型作对比。在选取参数时尽量保证所有子峰都被检测到, 同时观察信号中其他位置的平滑效果, 两个信号增强系数分别取 100 和 150, WMD 尺度取 3, 小波基选择“sym12”, SGD 框长度取 13, 阶次取 2, 迭代次数取 10, 3 种模型对图 2 两个信号的增强结果如图 5 所示。

通过两组结果的对比, 可以发现 3 种模型都具有较好的峰增强效果。然而和真实信号相比, WMD 模型和 SGD 模型的增强结果中新增加了一些伪峰(与真实峰相比新产生的峰), 由于这些伪峰的幅度较大, 容易导致误检, 如信号 1 中 $x=200$ 处, 信号 2 中 $x=200$ 处和 $x=250$ 处等。相比于其他两种模型, 本文模型在增强峰的同时, 对峰以外的部分具有更好的去噪效果, 有效降低了峰检测中伪峰的数目。

2.5 基于非线性扩散的导数谱增强模型的应用

基于非线性扩散的导数谱增强模型可以应用于 MALDI 质谱, MALDI 质谱的检测对于发现疾病生物标志物和研究药物代谢等方面具有重要意义。本文从 Bioinformatics Toolbox 中选取一组 MALDI 模拟数据作为实验样本(<http://bioinformatics.mdanderson.org/Supplements/Datasets/Simulations/index.html>), 使用本文模型对其做处理。为方便对比, 分别取增强系数=0(未增强)和增强系数=200, 处理结果如图 6 所示, 其中竖线表示真实峰的位置。

从图 6 可以看出本文模型对质谱具有一定的增强效果, 弱峰的高度被增加, 同时从 8 200 左右的位置可以看出该模型能有效增加重叠峰的分离度。

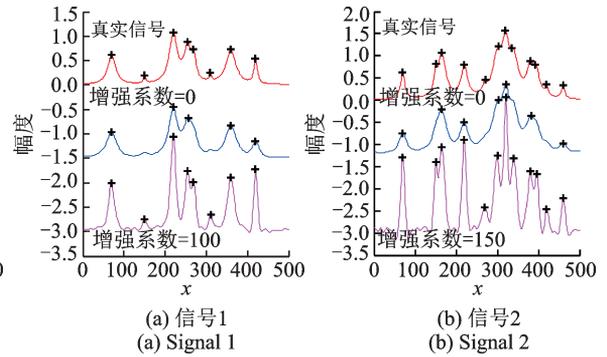


图4 基于非线性扩散的导数谱增强模型效果对比

Fig.4 Comparison for the derivative spectrum enhancement models based on nonlinear diffusion

表 1 使用基于非线性扩散的导数谱增强模型后峰检测结果对比

Tab. 1 Comparison of peak detection performance using derivative spectrum enhancement model based on nonlinear diffusion

编号	信号类别	检测到峰个数	漏检峰个数
信号 1	真实信号	8	
	增强系数=0	5	3
	增强系数=100	8	0
信号 2	真实信号	12	
	增强系数=0	7	5
	增强系数=150	12	0

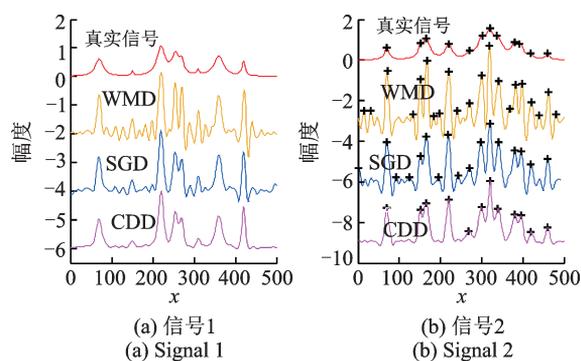


图5 不同模型信号增强结果对比

Fig.5 Comparison of enhancement results for different models

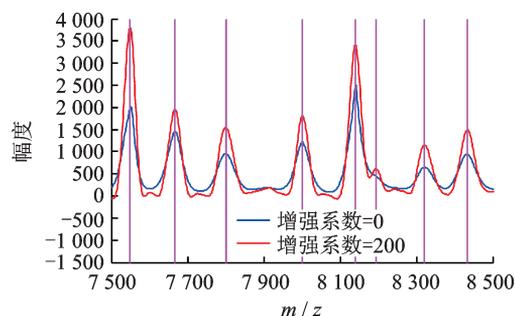


图6 应用增强与未增强质谱的处理效果对比

Fig.6 Comparison for the mass spectrometry with enhanced methods

3 结束语

本文提出了一种新的信号增强模型——基于非线性扩散的导数谱增强模型,并通过模拟信号和MALDI质谱验证所提模型的效果。结果表明所提模型能有效增强峰的特征,当信号中存在噪声,并有弱峰和重叠峰存在时,该模型能在去除噪声的同时有效增强弱峰的幅度和提高重叠峰的分离度。

参考文献:

- [1] Aebersold R, Mann M: Mass spectrometry-based proteomics[J]. Nature, 2003, 422: 198-207.
- [2] 汪鲁才,陈超华,王桃,等.水质检测中色谱峰峰峰算法的设计与实现[J].计算机工程与应用,2017,53(11): 237-240.
Wang Lucai, Chen Chao-hua, Wang Tao, et al. Design and realization of chromatography peak identification algorithm in water quality testing[J]. Computer Engineering and Applications, 2017, 53(11): 237-240.
- [3] Aqil M, Jbari A, Bourouhou A. Adaptive ECG wavelet analysis for R-peaks detection[C]//International Conference on Electrical and Information Technologies. [S.l.]: IEEE, 2016: 164-167.
- [4] Cameron D G, Moffatt D J, Mantsch H H, et al. Fourier self-deconvolution: A method for resolving intrinsically overlapped bands[J]. Applied Spectroscopy, 1981, 35(3): 271-276.
- [5] Du P, Kibbe W A, Lin S M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching[J]. Bioinformatics, 2006, 22(17): 2059-2065.
- [6] O'Haver T C, Fell A F, Smitii Q L, et al. Derivative spectroscopy and its applications in analysis[J]. Analytical Proceedings, 1982, 19: 22-46.
- [7] Goodin D G, Han L, Fraser R N, et al. Analysis of suspended solids in water using remotely sensed high resolution derivative spectra[J]. Photogrammetric Engineering & Remote Sensing, 1993, 59(4): 505-510.
- [8] Pera V E, Heffer E L, Siebold H, et al. Spatial second-derivative image processing: An application to optical mammography to enhance the detection of breast tumors[J]. Journal of Biomedical Optics, 2003, 8(3): 517-524.
- [9] Kher R, Vala D, Pawar T, et al. Implementation of derivative based QRS complex detection methods[C]//International Conference on Biomedical Engineering and Informatics. [S.l.]: IEEE, 2010: 927-931.
- [10] 王瑛,莫金垣.化学信号的近似高阶导数计算[J].分析科学学报,2004,20(5): 453-457.
Wang Ying, Mo Jinyuan. Approximate higher derivative calculation of chemical signals[J]. Journal of Analytical Science, 2004, 20(5): 453-457.
- [11] 刘茜阳,高楠,杜振辉,等.基于二阶导数谱与特征吸收窗的红外光谱定量检测方法[J].光谱学与光谱分析,2017,37(6): 1765-1770.
Liu Xiyang, Gao Nan, Du Zhenhui, et al. Quantitative detection of infrared spectra based on second derivative spectrum and

- characteristic absorption window[J]. *Spectroscopy and Spectral Analysis*, 2017, 37(6): 1765-1770.
- [12] Li Y, Pan C, Meng X, et al. A method of approximate fractional order differentiation with noise immunity[J]. *Chemometrics & Intelligent Laboratory Systems*, 2015, 144: 31-38.
- [13] Li X, Gentleman R, Lu X, et al. SELDI-TOF mass spectrometry protein data[C]//*Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. [S.l.]: Springer, 2005: 91-109.
- [14] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures[J]. *Analytical Chemistry*, 1964, 36(8): 1627-1639.
- [15] Danielsson R, Bylund D, Markides K E. Matched filtering with background suppression for improved quality of base peak chromatograms and mass spectra in liquid chromatography-mass spectrometry[J]. *Analytica Chimica Acta*, 2002, 454: 167-184.
- [16] Mantini D, Petrucci F, Pieragostino D, et al. LIMPIC: A computational method for the separation of protein MALDI-TOF-MS signals from noise[J]. *BMC Bioinformatics*, 2007, 8(1): 1-17.
- [17] 文鸿雁. 基于小波理论的变形分析模型研究[D]. 武汉: 武汉大学, 2004.
Wen Hongyan. Research on deformation analysis model based on wavelet theory[D]. Wuhan: Wuhan University, 2004.
- [18] 薛泽春, 李永刚, 李连之, 等. 应用连续墨西哥帽小波变换对弱信号提取分析研究[J]. *重庆文理学院学报(自然科学版)*, 2012(10): 79-90.
Xue Zechun, Li Yonggang, Li Lianzhi, et al. Study on the weak signal extraction analysis using continuous mexican hat wavelet transform[J]. *Journal of Chongqing University of Arts and Science(Natural Science Edition)*, 2012(10): 79-90.
- [19] Witkin A P. Scale-space filtering[C]//*Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. Karlsruhe, West Germany: Morgan Kaufmann Publishers Inc, 1983: 1019-1022.
- [20] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion, pattern analysis and machine intelligence[J]. *IEEE Transactions on*, 1990, 12(7): 629-639.
- [21] Shen S, Liu F, Anh V, et al. Detailed analysis of a conservative difference approximation for the time fractional diffusion equation[J]. *Journal of Applied Mathematics and Computing*, 2006, 22: 1-19.
- [22] Alikhanov A A. A new difference scheme for the time fractional diffusion equation[J]. *Journal of Computational Physics*, 2015, 280: 424-438.
- [23] Esmaili S, Garrappa R, A pseudo-spectral scheme for the approximate solution of a time-fractional diffusion equation[J]. *International Journal of Computer Mathematics*, 2015, 92: 980-994.

作者简介:



李俊(1993-),男,硕士研究生,研究方向:信号增强、峰检测,E-mail:2671225644@qq.com。



李远禄(1973-),男,教授,博士生导师,研究方向:信号处理、分数阶控制,E-mail:lyl_nuist@nuist.edu.cn。



蒋民(1995-),男,硕士研究生,研究方向:信号去噪,E-mail:717262287@qq.com。

(编辑:夏道家)