

一种自适应的 EMD 端点延拓方法

邵晨曦^{1,2,3,4}, 王 剑¹, 范金锋¹, 杨 明³, 王子才³

(1. 中国科学技术大学多媒体计算与通信教育部 微软重点实验室, 安徽合肥 230026;

2. 中国科学技术大学计算机科学技术系, 安徽合肥 230026; 3. 哈尔滨工业大学控制与仿

真中心, 黑龙江哈尔滨 150001; 4. 安徽省计算与通讯软件重点实验室, 安徽合肥 230027)

摘 要: 由美国国家航空航天局(NASA)的 Huang 等发明的经验模态分解(EMD)是一种先进的信号处理方法,能够有效地获得非平稳信号的时频特征,但是其利用样条曲线构造信号上下包络线的过程中存在严重的端点问题. 在研究了该问题已有方法的基础上,提出了一种基于波形匹配的自适应端点延拓方法,采用信号内部和端点处变化趋势最为相似的子波来对端点处的信号进行延拓. 该方法充分考虑了信号的内在特性以及边缘处的变化趋势,使端点处的延拓更加合理,从而使得三次样条曲线在端点处不会发生大的摆动. 实验表明该方法能够有效地抑制端点效应.

关键词: 经验模态分解; 端点效应; 自适应方法

中图分类号: TN911.7 文献标识码: A 文章编号: 0372-2112(2007)10-1944-05

A Self Adaptive Method Dealing with the End Issue of EMD

SHAO Chexi^{1,2,3,4}, WANG Jian¹, FAN Jinfeng¹, YANG Ming³, WANG Zicai³

(1. MOE-Microsoft Key Laboratory of Multimedia Computing and Communication, University of Science and Technology of China, Hefei, Anhui

230026, China; 2. Department of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230026, China;

3. Control and Simulation Center, Harbin Institute of Technology, Heilongjiang, Harbin 150001, China;

4. Anhui Province Key Laboratory of Software in Computing and Communication, Hefei, Anhui 230027, China)

Abstract: The Empirical Mode Decomposition (EMD) developed by Huang etc. of NASA is an advanced method for signal analysis. But there is an involved end issue in the course of getting two envelopes of the data using spline interpolation. In this paper, a new self adaptive method to dealing with the end issue is proposed. Generally, it extends the extrema sequence near the ends of the data by the most suited sequence in the inner data. For the chosen inner sequence has the most similarity of tendency with the ends, the data extension is reasonable. After the extension, the spline does not swing at both ends of the data. The result of experiment proved that the new method can be used to solve the end issue effectively.

Key words: EMD (empirical mode decomposition); end issue; self adaptive method

1 引言

瞬时频率是解析信号相位的导数,而对于实信号,可以通过 Hilbert 变换构造出与之相对应的解析信号,从而可以进一步得到其瞬时特征,如瞬时相位、瞬时频率等. 然而现实中的绝大部分信号都是多分量信号,对于这类信号通过 Hilbert 变换得到的结果就失去了其原有的物理意义. 1998 年,美国国家航空航天局(NASA)的 Huang 等人^[1]提出了一种新的信号处理方法——经验模态分解方法(Empirical Mode Decomposition, EMD). 该方法通过逐步的“筛选”,将一个多分量信号分解成多个单

分量信号——Huang 称之为固有模态函数(Intrinsic Mode Function, IMF)的组合. 基于这些 IMF 分量再进行 Hilbert 变换,便能够得到具有确切物理意义的瞬时特征.

EMD 方法的一个重要步骤是构造信号的上下包络线,以得到信号的瞬时平均. Huang 等提出以信号的极大值点和极小值点拟合三次样条曲线的方式来构造上下包络线. 由于信号在端点处往往并非极值点,因此三次样条曲线容易在端点处形成较大的摆动. 并且伴随着各个 IMF 的筛选过程,这种摆动会从边缘逐渐扩散到信号的内部,破坏整个数据序列,使得筛选结果严重失真. 这就是 EMD 方法的端点效应. 目前,人们已经提出了一

些抑制端点效应的方法, 包括镜像延拓法^[2]、基于神经网络的延拓法^[3]、基于多项式拟合的延拓方法^[4]等。这些方法对端点效应的抑制都有一定的效果, 却又都存在着各自的问题。镜像延拓方法在处理短数据时效果比较差, 因为它可能要截去部分数据。而对于神经网络方法而言, 它的不足在于速度太慢。至于多项式拟合方法, 则是适应性较差, 对某些数据它能得到比较好的结果, 而对另外的一些数据它的效果则不理想。

本文针对 EMD 的端点效应问题, 从信号自身的内在相似特性着手, 提出了一种基于波形匹配的自适应端点延拓方法, 通过对仿真和实际信号的分析, 验证了该方法能够有效地抑制端点效应。

2 EMD 方法和端点问题

2.1 EMD 方法简介

EMD 方法基于这样的假设: 任何信号都是由一系列幅度和相位都随时间变化的基本模式分量构成, 这种基本模式分量必须满足两个条件, 即它的零点数与极点数相等或至多相差 1 个, 以及由它的极大值和极小值确定的上下包络线关于时间轴局部对称。Huang 将这种基本模式分量定义为固有模态函数, 即 IMF。EMD 方法可以将多分量信号的各阶 IMF 一一筛选出来, 它的具体步骤如下:

设原始的信号为 $s(t)$,

(1) 确定 $s(t)$ 的所有极大值和极小值;

(2) 根据极大值和极小值作三次样条插值来构造 $s(t)$ 的上下包络线;

(3) 根据上下包络线, 计算出 $s(t)$ 的局部均值 $m_{11}(t)$, 以及 $s(t)$ 和 $m_{11}(t)$ 的差值 $h_{11} = s(t) - m_{11}(t)$;

(4) 以 h_{11} 代替原始信号 $s(t)$, 重复以上三步, 直到 $h_{1,k-1}$ 与 $h_{1,k}$ 之间的方差小于一设定值, 即认为 $h_{1,k}$ 是一个 IMF 分量, 记 $c_1 = h_{1,k}$, $r_1(t) = s(t) - c_1$, $s(t) = r_1(t)$;

(5) 重复以上四步, 直到 r_n 小于一设定值, 或者它变成了一个单调函数时, 原始信号的 EMD 分解结束, 得到 $s(t)$ 的分解式如下:

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

每一个 IMF 分量都反映了原信号不同时间尺度的内在模态特征, 并且是窄带信号, 从而使得瞬时频率具有了确切的物理意义, 于是可以合理地进行 Hilbert 谱分析。此外, 除了谱图的应用外, EMD 也可以作为良好的自适应滤波器或提取特殊应用有用的特征^[5]等。

2.2 端点问题

EMD 方法自推出以来已经成功地应用到许多非线性研究领域^[6]。但是在应用 EMD 方法时存在着较为严重的端点问题, 这个问题是在构造信号的包络线时产

生的。由于采用以极值点作为插值点的三次样条拟合的方式来构造信号的包络线, 而信号在端点处往往并非极值点, 因而, 样条曲线在数据序列的两端会出现发散现象, 并且这种发散的结果会伴随着“筛选”过程的不断进行逐渐向内“污染”整个数据序列而使得所得结果严重失真。端点问题已经成为 EMD 方法的一个瓶颈。

对于一个较长的数据序列而言, 可以根据极值点的情况不断抛弃两端的数据来保证所得包络的失真度达到最小。对于一个短数据序列来说, 这样的操作就变得完全不可行, 而只有采用数据延拓的方法, 即在数据的两端分别增加一段构造的数据, 使得两端各增加一对极大值和极小值, 然后把新增的极值点也作为插值点来构造样条曲线。如果这种延拓能够合理地延续原始信号在端点处的走向, 那么构造出来的两对虚拟的极大极小值点就能够对样条曲线在原信号端点处的走向进行较好的控制, 从而使得样条曲线在信号端部的翘起现象就会得到有效的抑制。端点延拓是解决 EMD 端点效应问题的有效手段, 是目前普遍采用的方法, 如镜像延拓方法^[2]、基于神经网络延拓方法^[3]、基于多项式拟合的延拓方法^[4]等, 与本文提出的基于波形匹配的端点延拓方法一样, 他们的本质都是在原始信号的两端分别构造出一对虚拟的极大和极小值点, 以此来抑制样条曲线在端部的翘起。数据延拓的质量直接决定了抑制的效果, 进而影响 EMD 分解的质量。

3 基于波形匹配的 EMD 端点问题的处理

3.1 波形匹配度计算

设 $s_1(t)$ 和 $s_2(t)$ 是两组长度相同的数据序列, 它们的匹配度在直观上反映为它们的波形的相似程度。设 $P_1(x_0, y_1)$ 和 $P_2(x_0, y_2)$ 分别是 $s_1(t)$ 和 $s_2(t)$ 上两点。按如下方式计算 $s_1(t)$ 和 $s_2(t)$ 相对于 P_1 、 P_2 的匹配度。首先, 对 $s_1(t)$ 进行平移操作, 使得它的 P_1 和 $s_2(t)$ 的 P_2 相重合。这个过程是两段波形在坐标轴上依照参考点的一个对齐过程, 参见图 1。

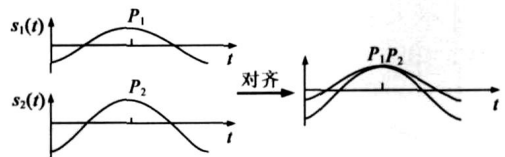


图 1 波形匹配度计算

设对齐后, 即设 $s_1(t)$ 平移之后得到的新波形为 $s'_1(t)$ 。按下面的公式计算 $s_1(t)$ 和 $s_2(t)$ 的针对参考点 P_1 (或 P_2) 的匹配度:

$$m(x, y, P_1) = \sum_{i=1}^N [s_2(i) - s'_1(i)]^2, \quad N \text{ 为序列长度}$$

(2)

对于一个信号内部的子波,可以采取同样的方式计算它们的匹配度.首先通过平移操作,将待匹配的子波依据参考点对齐,然后依据公式(2)进行计算.

3.2 基于波形匹配的端点延拓方法

对端点处的数据延拓不可能是盲目的延拓,延拓出的波形必须要符合原始信号在端点处的变化趋势,这样的延拓才是有意义的.因此,延拓的关键就是要确定原始信号在端点处的趋势.原始信号的发展趋势不仅会表现在端点处,在信号的内部必定也存在.特别是对于规律性较强的信号,这种特性更加明显.这就是说,如果在信号的内部有一段子波和边缘处的变化趋势非常相似,那么就可以用这段子波来延拓信号边缘处的那段波形.这样的延拓可以最大限度地维护原始信号的变化趋势,从而可以有效地抑制端点效应的产生.而如果在信号的内部没有任何一段子波的变化趋势和边缘处相似,这就说明信号本身的边缘处发生了异常的变化,此时参考信号内部已不再合适,数据延拓只需考虑边缘处的局部数据即可.

根据以上分析,从而可以推出一种基于波形匹配的数据延拓方法,通过采用信号内部和边缘处变化趋势最为相似的子波来对端点处数据进行延拓,这样,在信号内在规律性较强的情况下,它可以最大限度地维护信号的内在趋势,同时,它是一种自适应的方法,即在信号内在规律性较弱的情况,特别是信号边缘发生异常变化的情况下,它也能得到令人满意的结果,因为此时它只考虑边缘处的局部信息.在具体实现中,通过计算波形匹配度来量化两段波形的变化趋势,该方法的详细描述如下:

信号的延拓包括左右两端,下面以左端的延拓为例,设原始信号为 $s(t)$,

(1) 确定 $s(t)$ 最左端的两个相邻极值点,不妨设其分别为极大值点和极小值点,分别记为 P_0 和 P_1 ,从起始点到 P_1 的这段波形记为 w_0 ,设其长度为 l ;

(2) 设 E_{\max} 为 $s(t)$ 的极大值点集合,以 $E_{\max} - \{P_0\}$ 中的每一个极大值点 P_i 作为参考点,计算该段相同长度的波形 w_i 和 w_0 的匹配度 $m(w_0, w_i, P_i)$;

(3) 记

$$m(w_0, w_{i_0}, P_{i_0}) = \min \{ m(w_0, w_i, P_i), i = 1, 2, \dots \} \quad (3)$$

若 $m(w_0, w_{i_0}, P_{i_0}) < \alpha \cdot l$, 其中 α 为一常数,则取 w_{i_0} 左端包含了一个极大值和极小值的子波,作为原始 $s(t)$ 左端的延拓,延拓完毕;否则转(4);

(4) 直接指定端点处的极大和极小值:取原始信号最左端的两个相邻极大值点的均值作为左端点的极大值,取信号最左端的两个相邻极小值点的均值作为左端点的极小值.完毕

右端点处的延拓与左端点类似.该方法的图示如图 2.

常数 α 的大小可以根据实际情况加以调整,在实验中将其设定为 4. α 越大,端点的延拓越趋向于在第三步结束, α 越小,则越趋向于在第四步结束.

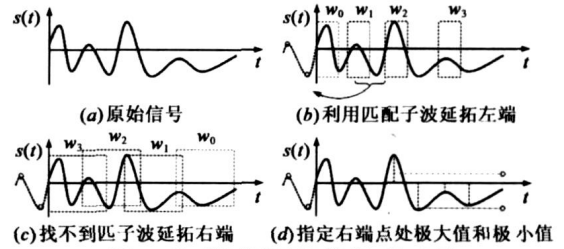


图 2 基于波形匹配的数据延拓方法

4 实验分析

4.1 仿真信号

为了便于分析比较,采用其他文献中^[3,4]经常使用的仿真信号作为实例来分析,该信号的表达式为

$$x(t) = \cos\left(\frac{\pi}{25}t\right) + \frac{3}{5}\cos\left(\frac{2\pi}{25}t\right) + \frac{1}{2}\sin\left(\frac{\pi}{100}t\right) \quad t \in [5, 95] \quad (4)$$

其波形如图 3(a) 所示,其中实线是原始信号,虚线是基于它的极值点构造的上下包络线.从图可见,该包络线在数据两端出现了巨大的失真,尤其是上包络线,发散

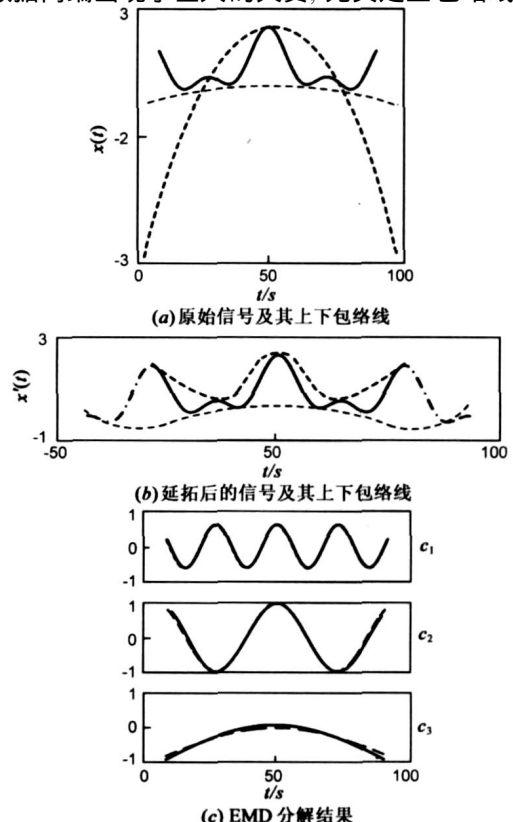


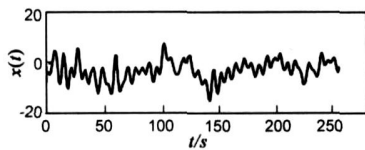
图 3 仿真信号的端点延拓与 EMD 分解

现象非常严重,由于数据序列很短,产生的端点效应会“污染”整个数据序列,使得 IMF 的分解失去意义.采用基于波形匹配的端点延拓方法对两侧端点进行延拓后得到新的数据序列 $x'(t)$,该新序列以及基于该序列的上下极值点构造的包络线如图 3(b) 所示,从图可见,该方法对 $x(t)$ 的两端的延拓有效地抑制了包络线在端点处的发散,并且,这种延拓和原始信号在端点处的变化趋势的吻合度很高,因而包络线的形式和量值都非常准确.

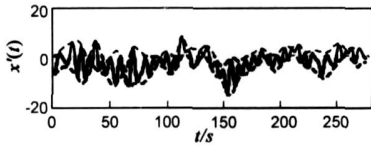
最终的 EMD 分解结果如图 3(c) 所示,其中实线为分解得到的三个 IMF 分量,虚线为信号的三个真实分量,从图可见,由于端点效应被有效抑制,EMD 分解的结果非常理想.

4.2 实际信号

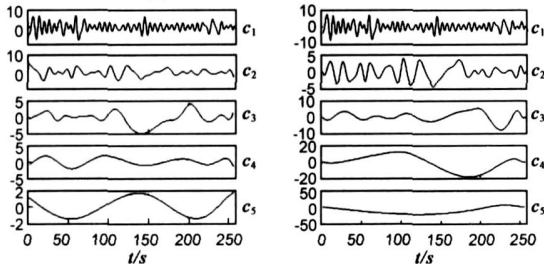
对于现实中的非平稳随机信号,该方法也能很好地抑制端点效应的产生.图 4(a) 是一段脑电波 (EEG),该数据来自于美国纽约州立大学健康中心的神经动力学实验室. EEG^[7] 是大脑头皮细胞活动所产生的一种生物电信号,它是一种典型的非平稳随机信号,下面对该段信号进行 EMD 分解,为了获得端点处的极值信息,使用波形匹配方法对其端点处数据进行延拓,图 4(b) 是延拓后得到的新序列.虚线部分是其上下包络线,从图可见,上下包络线构造的非常理想,没有出现任何的发散现象或者包络不完全现象.



(a) 原始 EEG 信号



(b) 延拓后信号及其上下包络线



(c) 经端点延拓后的 EMD 结果 (d) 端点不作任何处理的 EMD 结果

图 4 EEG 信号的 EMD 分解

最终的分解结果如图 4(c) 所示,而不针对端点问题进行任何处理得到的 EMD 分解结果如图 4(d) 所示,从图中可以直观看出,如果不对端点进行处理,会得到一些毫无意义的分量,具体表现为在高频上的部分结

果尚能接受,而到了低频分量上,由于包络线误差的不断积累,信号的发散现象非常严重,甚至出现了幅值超出原信号 2 倍多的情形.而采用了基于波形匹配的数据延拓方法对端点进行处理后,得到的分解结果非常理想,从高频到低频的得到了原始信号的五个 IMF 分量,这表明端点效应得到了有效的抑制.

为了进一步验证该方法的有效性,分别采用神经网络方法、镜像延拓方法和多项式拟合方法对该 EEG 数据进行延拓,并得到基于这三种方法的 EMD 分解结果,分别如图 5(a)、(b)、(c) 所示.

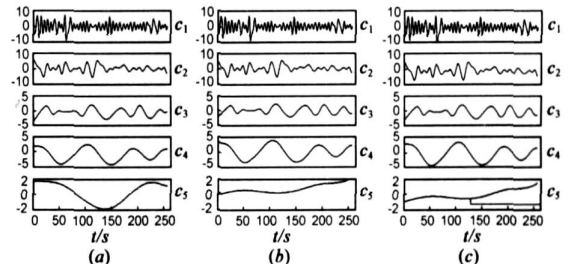


图 5 (a)神经网络法分解结果;(b)镜像延拓法分解结果;(c)多项式拟合法分解结果

从四个结果的对比来看,基于波形匹配的方法和神经网络方法的效果都比较好,它们都较好地抑制了端点效应,而镜像延拓法和多项式拟合法的效果要稍差一点,在高频分量上差异还不是很明显,由于包络线的误差不断累积,在低频分量上表现出了明显的差异.从图上可以看出,几个方法得到的前四个 IMF 分量基本相似,第五个分量 c_5 表现出了明显的不同.

神经网络方法需要通过一个学习过程不断地调整网络模型的权重向量和偏移量,这需要花费很长的时间,而镜像延拓法和多项式拟合法不具备适应性,前者常常需要对原始信号进行截断,不适用于短数据序列情形,后者则完全依赖边缘处的极值信息,延拓的信号的变化趋势完全受该信息影响,而大多数现实信号并非如此,因此数据延拓的效果不是很理想.从几个 EMD 分解的结果图对比可见,数据延拓的质量直接决定了 EMD 分解的质量,基于波形匹配的端点延拓方法最大限度地维护了原始信号在端部的变化趋势,因而,相比另外三种方法而言,它所构造出来的虚拟的极大极小值点更加合理,进而对端点效应的抑制效果更加理想.综合计算复杂度和适应性两个方面性能来考虑,基于波形匹配的方法要优于以上三种方法.

5 结论

提出了一种克服 EMD 分解中的端点效应的新方法,对于具有一定内在相似性的信号,通过计算波形匹配度,在信号内部寻找与边缘处变化趋势最为相似的子波,通过该子波对原始信号进行延拓,有效地抑制了

三次样条拟合构造上下包络线过程中的发散现象. 同时它是一种自适应方法, 即如果信号边缘处的变化趋势和内部的任何一部分都不具备相似性, 进行延拓时则仅考虑信号边缘处的变化趋势, 根据靠近端点处的两个极值的平均来近似端点处的极值, 这既保证了信号在端点处的变化趋势不会发生大的变化, 同时也可以遏制三次样条曲线在端部的翘起. 从实验结果来看, 绝大部分信号只要通过波形匹配的方式就可以得到延拓, 也就是说绝不部分信号都具备一定的内在相似性. 从仿真信号和实际信号的分析结果表明, 该方法对能够非常有效地抑制端点效应. 需要说明的是, 对于这种端点延拓方法的理论依据, 包括 EMD 方法本身的理论依据, 还需要做进一步的研究, 这是一个非常有意义的课题, 也是作者下一步要做的工作.

参考文献:

- [1] N E Huang, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non stationary time series analysis[J]. Proceeding of the Royal Society of London Series A, 1998, 454: 903- 995.
- [2] Zhao Jin Ping, et al. Mirror extending and circular spline function for empirical mode decomposition method[J]. Journal of Zhejiang University, 2001, 2(3): 247- 252.
- [3] 邓拥军, 等. EMD 方法及 Hilbert 变换中边界问题的处理 [J]. 科学通报, 2001, 46(3): 257- 263.
- [4] 刘慧婷, 等. 基于多项式拟合算法的 EMD 端点问题的处理 [J]. 计算机工程与应用, 2004, 40(16): 84- 86, 100.
- [5] 谭善文, 等. Hilbert Huang 变换的滤波特性及其应用 [J]. 重庆大学学报, 2004, 27(2): 9- 12.
- [6] Chir Hsiung Loh, et al. Application of the empirical mode decomposition hilbert spectrum method to identify near fault

ground motion characteristics and structural responses [J]. Bulletin of the Seismological Society of America, 2001, 91(5): 1339- 1357.

- [7] 王兆源, 等. 脑电信号的分析方法 [J]. 第一军医大学学报, 2000, 20(2): 189- 200.

作者简介:



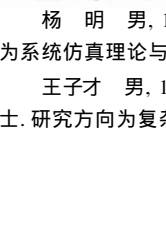
邵晨曦 男, 1954 年 7 月生于山东荣成, 副教授, 硕士. 研究方向为生物信息建模仿真理论.
E mail: cxshao@ustc.edu.cn



王 剑 男, 1983 年 7 月出生于江苏盐城, 中国科学技术大学计算机科学技术系硕士研究生. 主要研究方向为生物医学信号时频分析.
E mail: jianw3@mail.ustc.edu.cn



范金锋 男, 1978 年生于江苏江阴, 博士生. 研究方向为生物信息建模仿真



杨 明 男, 1963 年生于吉林蛟河, 博士、教授、博导. 研究方向为系统仿真理论与方法.

王子才 男, 1932 年生于山东聊城, 教授、博导, 中国工程院院士. 研究方向为复杂系统理论、系统仿真理论.