

文章编号: 2095-4980(2013)02-0277-05

基于小波包去噪和 EMD 的混合算法

汤继磊^{1a}, 黄玉清^{1b}, 潘泽友²

(1.西南科技大学 a.信息工程学院; b.国防科学与技术学院, 四川 绵阳 621000;
2.中国工程物理研究院 计算机应用研究所, 四川 绵阳 621999)

摘要: 经验模态分解(EMD)是希尔伯特-黄变换(HHT)中的关键步骤, 并伴有过冲和端点效应的产生。利用遗传算法(GA)对支持向量机(SVM)中的未知参数: 惩罚函数 C 和高斯核函数中的预设参数 σ 进行优化选取, 运用 GA-SVM 对信号进行端点延拓来处理端点效应问题并提出采用分段三次 Hermite 多项式插值进行包络线拟合; 为了机械设备早期故障频率的特征提取, 采用小波包降噪预处理, 结合改进的 Hilbert-Huang 变换进行轴承故障特征频率的提取实验; 实验表明该方法提高了故障频率提取的准确性。

关键词: 希尔伯特-黄变换; 分段三次 Hermite 多项式插值; 遗传算法-支持向量机; 小波包
中图分类号: TN957.54; TP391.5 **文献标识码:** A

Hybrid algorithm based on wavelet packet denoising and EMD

TANG Ji-lei¹, HUANG Yu-qing², PAN Ze-you³

(1a.School of Information Engineering; 1b.School of National Defence Science & Technology, Southwest University of Science and Technology, Mianyang Sichuan 621000, China;
2.Computer Application Research Institute, China Academy of Engineering Physics, Mianyang Sichuan 621999, China)

Abstract: Empirical Mode Decomposition(EMD) decomposition is a critical step in Hilbert-Huang Transform(HHT), accompanied by overshoot and endpoint effect. The Genetic Algorithm(GA) is used to optimize and select the unknown parameters including the penalty function C and default parameters σ of Gaussian kernel of Support Vector Machines(SVM). GA-SVM is applied to extend signals to deal with endpoint effect, and cubic Hermite polynomial interpolation is adopted for envelope fitting. In order to extract the early stage fault frequency features of mechanical equipment, wavelet packet noise reduction pretreatment is performed, combined with the extraction experiment of bearing fault feature frequency by using improved HHT transform. The experimental results show that the proposed method can improve the accuracy of fault frequency extraction.

Key words: Hilbert-Huang Transform(HHT); segmentation cubic Hermite polynomial interpolation; GA-SVM; wavelet packet

故障信号特征的提取是机械设备早期故障诊断的关键, 基于信号处理的特征提取方法一直是故障诊断领域的研究重点。传统的信号处理方法包括短时傅里叶变换、快速傅里叶变换、小波变换以及小波包变换等。但其各有缺陷, 基于傅里叶变换的信号处理方法受到频响函数估计所带来误差的影响; 基于小波变换的信号分析方法, 可以对信号进行降噪处理, 但其本质是一种线性变换, 对于处理非线性问题存在不足之处。针对上述问题, 1998 年美国科学家 Norden E Huang 等人提出了一种新的信号分析方法: 希尔伯特-黄变换(HHT)^[1], 此方法主要用于分析非平稳信号。他创造性地引入了经验模态分解(EMD)方法, 将信号经过 EMD 分解成本征模函数(Intrinsic Mode Function, IMF)的和, 然后再对每个 IMF 分别进行 Hilbert 变换谱分析就可以获得有意义的瞬时频率, 它无需采用信号的先验知识, 从原信号中自适应地分解得到基函数。然而毕竟 HHT 是刚刚提出的理论成果, 像其他的信号处理方法一样也存在着不足之处, 如插值问题、频率混叠现象、端点问题等, 其理论框架还需进一步完善,

目前对其的完善取得了一定的成果,文献[2]采用分段幂函数插值方法,文献[3]采用 B 样条插值方法,文献[4]采用频率偏差法,文献[5]采用自回归模型,文献[6]采用加窗函数法解决端点问题。本文采用分段三次 Hermite 多项式插值方法解决过冲现象,利用 GA-SVM 算法解决端点问题,将改进的 HHT 与小波包变换相结合,并在轴承的外圈故障检测试验中进行了验证。

1 信号的小波包降噪

1.1 基本原理

所谓小波包,简单地讲就是一个函数簇,由它们构造出 $L^2(R)$ 的规范正交基库。从此库中可以选出 $L^2(R)$ 的许多组规范的正交基,所谓的小波正交基就是其中的一组,小波包其实就是小波的推广。所以在信号降噪的思想上小波包分析和小波分析基本相同,但小波包提供了一种更为复杂、灵活的分析手段,因为小波包分析是对上一层的低频部分和低频部分同时进行分解,这样具有更加精确的局部分析能力。另外在对信号进行小波包分解时,可以根据分析信号的要求和熵标准,从中选择最好的最优基。在 Matlab 的小波工具箱中可通过 `besttree` 函数进行最优基的选择。

1.2 信号小波包降噪步骤

根据小波包具有任意多尺度分解的特性^[7],以及在时、频域具有表征信号局部信息的能力,可利用小波包对信号进行降噪处理,其步骤一般按照如下进行:

- 1) 选择小波并确定分解的层次,进行小波包分解;
- 2) 根据熵标准,确定最优小波基;
- 3) 根据每一个小波分解系数,选择合适的阈值对系数进行阈值量化;
- 4) 根据最低层的小波包分解系数和经过量化处理的系数进行小波包重构。

2 HHT 算法的改进

EMD 中关键的步骤是通过信号的极值点拟合信号的包络线,但包络的过冲/欠冲以及端点效应等问题迄今尚未给出确切的解决方法。

2.1 EMD 分解

对于每个 IMF 必须满足下列 2 个条件:

- 1) 在整个信号长度上,极值点和过零点的数目必须相等或者至多只相差 1 个;
- 2) 在任意时刻,由极大值点定义的上包络线和由极小值定义的下包络线的平均值为零,即信号的上下包络线对称于时间轴。

EMD 分解步骤:

- 1) 确认信号 $x(t)$ 所有的极值点,用极大值点拟合出 $x(t)$ 的上包络线 E_{up} ,用极小值点拟合 $x(t)$ 的下包络线 E_{down} ,则上下包络线的平均值:

$$m(t) = (E_{up} + E_{down}) / 2 \quad (1)$$

- 2) 将 $x(t)$ 减去 $m(t)$ 得:

$$h(t) = x(t) - m(t) \quad (2)$$

判断 $h(t)$ 是否满足 IMF 的条件,若满足,则 $h(t)$ 就为第 1 阶的 IMF: $c_1(t)$;若不满足,令 $h(t) = x(t)$,重复以上的操作。

- 3) 重复步骤 1)~步骤 2),直到提取所有的 $c_i(t)$ ($i=1,2,\dots,n$),最后剩余残差 $r(t)$ 代表信号的平均趋势。

所以最终 $x(t)$ 可表示为:

$$x(t) = \sum_{i=1}^n c_i(t) + r(t) \quad (3)$$

可以看出,通过极值点拟合包络线的算法影响 EMD 的全过程,因而包络线的算法是 EMD 分解的关键问题,直接影响到最终的结果。

2.2 包络线拟合

EMD 中包络线拟合算法至今没有从理论上严格地确定下来,黄等人最初所用的三次样条插值具有最佳的逼近性,但此方法拟合的曲线存在严重不足,由于三次样条插值在待拟合的 2 个邻近的插值点之间不具备单调性,所以在拟合过程中会出现过冲和欠冲的现象,本文利用分段三次 Hermite 多项式插值方法进行插值,可以有效解决算法的过冲问题。

由文献[8]可知,对于 $N+1$ 个插值点满足: $t_0 < t_1 < \dots < t_N$, 分段插值函数 $H_k(t)$ 在 2 个相邻结点构成的小区间上 $[t_i, t_{i+1}]$ 满足插值条件: $H_k(t_i) = x_i, H_k(t_{i+1}) = x_{i+1}, H_k'(t_i) = x_i', H_k'(t_{i+1}) = x_{i+1}'$ 。

利用两点 Hermite 插值结论,当 $t \in [t_i, t_{i+1}], i = 0, 1, \dots, n-1$ 时,有:

$$\tilde{H}(t) = x_i \left(1 + 2 \frac{t-t_i}{t_{i+1}-t_i}\right) \left(\frac{t-t_{i+1}}{t_i-t_{i+1}}\right)^2 + x_{i+1} \left(1 + 2 \frac{t-t_{i+1}}{t_i-t_{i+1}}\right) \left(\frac{t-t_i}{t_{i+1}-t_i}\right)^2 + x_i' (t-t_i) \left(\frac{t-t_{i+1}}{t_i-t_{i+1}}\right)^2 + x_{i+1}' (t-t_{i+1}) \left(\frac{t-t_i}{t_{i+1}-t_i}\right)^2$$

由以上可知,经验模态分解过程中构造包络线的极值点满足了三次 Hermite 插值的要求,因为很显然各极值点一阶导数为零,且在各极值点处的导数均相等,且均等于零,这就满足了三次 Hermite 插值所需的各节点的导数连续的要求。且此方法可以保证 2 个插值点间的拟合曲线是单调的,可以有效避免过冲问题的产生。如图 1、图 2 所示,分别用三次样条插值和分段三次 Hermite 多项式插值对其进行上下包络线拟合,由图可以清楚地看到由三次样条插值所拟合的上下包络线在 A,B 两点存在着明显的过冲现象,并在端点处也出现了严重的失真,由分段三次 Hermite 多项式插值拟合的上下包络线的结果可以看出,过冲问题得到了改善。

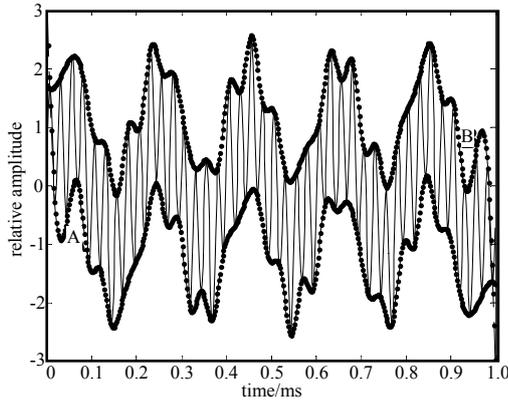


Fig.1 Cubic spline interpolation fitting
图 1 三次样条插值拟合

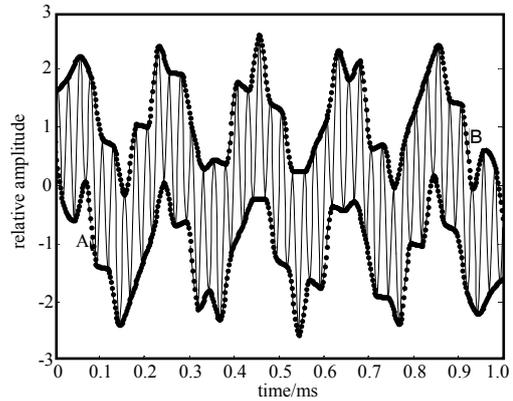


Fig.2 Segmentation cubic Hermite polynomial interpolation fitting
图 2 分段三次 Hermite 多项式插值拟合

2.3 GA-SVM 处理端点问题

支持向量机(Support Vector Machine, SVM)自 1995 年由 Vapnik 提出以来就受到理论研究和工程应用的双重重视,在很多机器学习问题上 SVM 能够获得比神经网络更好的性能^[9]。运用 SVM 处理问题仅依赖于训练数据的子集——支持向量(Support Vector, SV),便可获得同使用整个训练集相同的效果。由于遗传算法简单通用,鲁棒性强,而且求解组合优化问题具有很好的效果,因此支持向量机中惩罚函数 C ,以及高斯核函数中的预设参数 σ 的最优值可由它来优化。假设每个个体为 $g_i = (g_{i1}, g_{i2}) = (\lg C, \lg \sigma)$, 具体步骤如下:

- 1) 读入样本集 S , 且令前 N 个样本为工作集 W , 即 $W = \{s_1, s_2, \dots, s_N\}$;
 - 2) 群体初始化; 对参数进行浮点数编码, 且令当前进化代数为 0;
 - 3) 计算种群中个体的适应度值, 将适应度值最大的保留到下一代;
 - 4) 根据适应度值进行非支配排序, 选出非支配集种群;
 - 5) 将非支配集种群离散均匀化处理, 然后再进行选择、交叉、变异;
 - 6) 以种群最大进化代数为终止条件, 反复执行步骤 3)~步骤 5)。最大适应度值对应参数组合即为最优解。
- 对于给定的信号:

$$x(t) = \sin(24\pi t) + 0.5\sin(50\pi t) + 1.2\sin(162\pi t)$$

其对应的 3 个频率成分为: 12 Hz, 25 Hz, 81 Hz, 采用 GA-SVM 算法对原始信号 50 个采样点进行延拓仿真, 其中 GA 中交叉概率取 0.8, 变异率取 0.1, 进化最大代数为 20, 群体规模取 50。经过 GA 优化的参数 $\lg C$ 为 0.6, $\lg \sigma$ 为 1.2, 结果如图 3 所示, 其中“o”为原始信号波形, “*”为预测信号波形, 从图中可以看出预测信号波形与原始信号波形基本拟合。图 4 为信号经过原有的 Hilbert-Huang 变换中 EMD 分解的结果(不包含残余分量)。图 5 为本

文改进后的 EMD 分解结果(不包含残余分量),从图 4 中可以看出,从上到下,3 个分量都存在过冲和端点效应问题,而从图 5 中看到上述问题得到了有效改善。

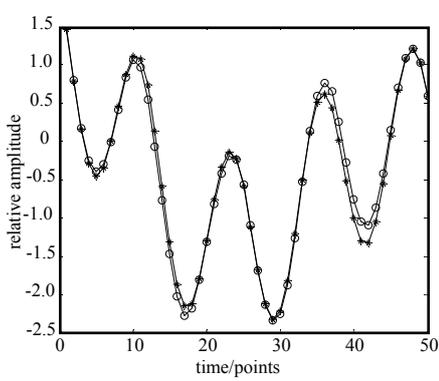


Fig.3 GA-SVM endpoint continuation
图 3 GA-SVM 端点延拓

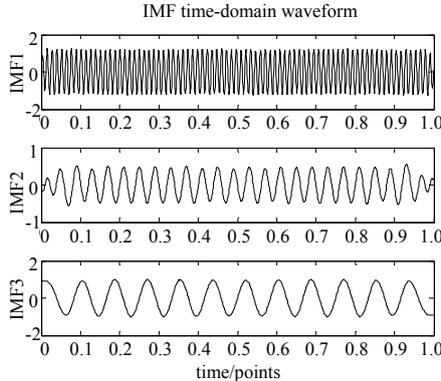


Fig.4 Traditional EMD decomposition
图 4 传统 EMD 分解

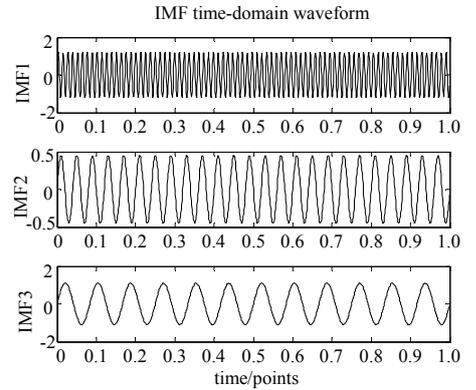


Fig.5 EMD decomposition of proposed algorithm
图 5 本文算法 EMD 分解

3 滚动轴承故障特征提取实验

实验对象采用型号为 6205-2RS,SKF 深沟球轴承,内圈直径为 25 mm,外圈直径为 52 mm,厚度为 15 mm,滚动体直径为 8 mm,节径为 39 mm,接触角为 0,滚动体个数为 9,在轴承上布置单点故障,故障直径为 4.27 mm,轴承内圈转速 1 777 r/min,采样频率 12 kHz,经过计算,轴承外圈的故障频率为 106 Hz,滚动轴承外圈出现故障时的时域波形如图 6 所示,图 7 为经过小波包降噪处理后的信号时域波形图。图 8 和图 9 是采用原有的 Hilbert-Huang 变换中 EMD 分解结果(取前 5 组分量)和对应的边际谱图。从图中可以看出故障造成的高频冲击部分出现了模态混叠的现象,从边际谱图中可以看到故障频率 106 Hz 也被其他频率所淹没。

图 10 为采样序列经过本文改进的 EMD 分解的结果,图 11 为对应的边际谱图,其中图 10 中 IMF1 分量为轴承外圈故障冲击的高频分量,具有明显的周期性。且在图 11 中可以有效提取故障频率 106 Hz。

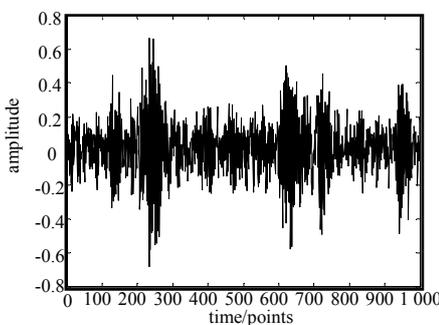


Fig.6 Time domain waveform under the fault of bearing outer ring
图 6 轴承外圈故障时域波形图

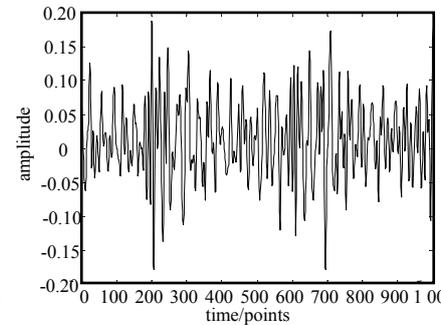


Fig.7 Domain waveform under the bearing outer ring fault after wavelet packet noise reduction process
图 7 小波包降噪处理后的轴承外圈故障时域波形

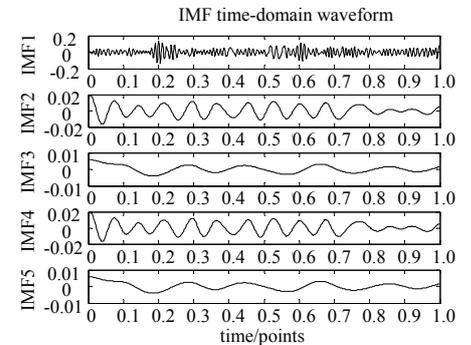


Fig.8 Decomposition results of traditional EMD
图 8 传统 EMD 分解结果

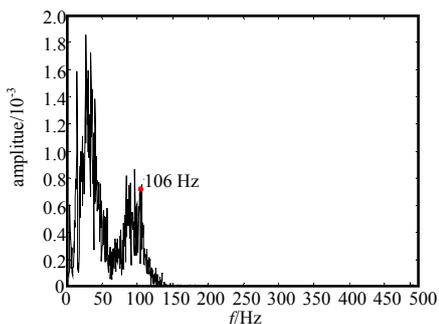


Fig.9 Corresponding marginal chart of traditional EMD decomposition
图 9 传统 EMD 分解对应边际谱图

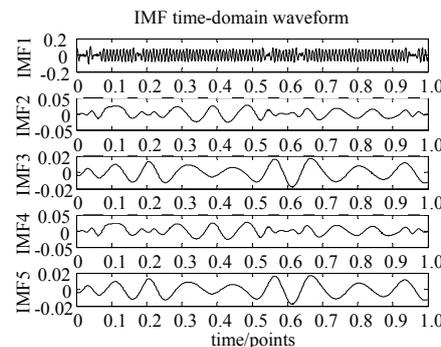


Fig.10 EMD decomposition results of the proposed algorithm
图 10 本文算法 EMD 分解结果

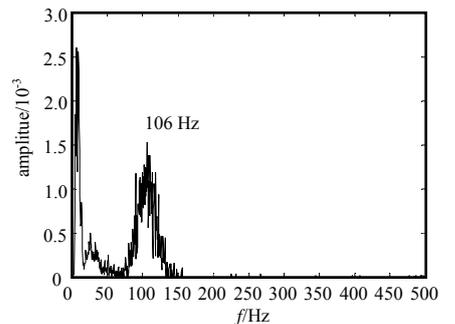


Fig.11 Corresponding marginal spectra of EMD decomposition of the proposed algorithm
图 11 本文算法 EMD 分解对应的边际谱图

4 结论

本文对原有的 EMD 算法进行了分析和探讨,针对其原有的缺陷分别提出了分段三次 Hermite 多项式插值拟合的上下包络线和运用 GA-SVM 进行端点延拓的改进方法,并与小波包变换相结合应用于轴承的外圈故障检测实验中,能准确地确定故障频率,具有很好的效果。

参考文献:

- [1] HUANG N E, SHEN Z, LONG S R, et al. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis[J]. Proceedings of the Royal Society of London Series A-Mathematical, Physical and Engineering Sciences(S1364-5021), 1998,454(1971):903-995.
- [2] 钟佑明,金涛,秦树人. 希尔伯特-黄变换中的一种新包络线算法[J]. 数据采集与处理, 2005,20(1):13-17. (ZHONG Youming, JIN Tao, QIN Shuren. A new envelope algorithm of Hilbert-Huang transform[J]. Data Acquisition and Processing, 2005,20(1):13-17.)
- [3] CHEN Q, HUANG N, RIEMENSCHNEIDER S, et al. A B-Spline Approach for Empirical Mode Decompositions[J]. Advances in Computational Mathematics(S1019-7168), 2006,24(1):171-195.
- [4] 杨德昌, Rehtanz C, 李勇, 等. 基于改进希尔伯特-黄变换算法的电力系统低频振荡分析[J]. 中国电机工程学报, 2011,31(10):102-108. (YANG Dechang, Rehtanz C, LI Yong, et al. Based on improved Hilbert-Huang transform algorithm of power system analysis of low frequency oscillation[J]. Journal of Chinese electrical engineering, 2011,31(10):102-108.)
- [5] 张郁山, 梁建文. 应用自回归模型处理 EMD 方法中的边界问题[J]. 自然科学进展, 2003,13(10):1054-1059. (ZHANG Yushan, LIANG Jianwen. Application processing EMD method regression models of boundary problem[J]. Progress in natural science, 2003,13(10):1054-1059.)
- [6] 任达千, 吴昭同, 严拱标. EMD 端点效应的评价指标及抑制端点效应的窗函数法[J]. 制造业自动化, 2007(1):21-24. (REN Daqian, WU Zhaotong, YAN Gongbiao. EMD endpoint effect evaluation index and the suppressing endpoint effect of the window function method[J]. Manufacturing automation, 2007(1):21-24.)
- [7] 韩朝晖. 利用小波包分析抑制窄带干扰[J]. 信息与电子工程, 2006,4(6):448-453. (HAN Zhaohui. Rejection of Narrowband Interference Based on Wavelet Packet Analysis[J]. Information and Electronic Engineering, 2006,4(6):448-453.)
- [8] 钟佑明. 小波包时频分析及其特性[J]. 振动、测试与诊断, 2009,29(1):51-54. (ZHONG Youming. Wavelet Packet Analysis and the Characteristics of Frequency[J]. Vibration, Testing and Diagnosis, 2009,29(1):51-54.)
- [9] Vapnik V N. Statistical Learning Theory[M]. New York:Springer-Verlag, 1998.

作者简介:



汤继磊(1986-),男,安徽省淮南市人,在读硕士研究生,研究方向故障诊断及人工智能.email:527691683@qq.com.

黄玉清(1962-),女,四川省绵阳市人,教授,主要研究方向为网络测控、智能控制.

潘泽友(1955-),男,四川省绵阳市人,研究员,主要研究方向为控制理论与控制工程、通信与信息系统.