

# EMD 端点效应抑制方法<sup>\*</sup>

郝如江, 李 非

(石家庄铁道大学机械工程学院 石家庄, 050043)

**摘要** 针对经验模态分解(empirical mode decomposition,简称 EMD)中的端点效应问题,在研究总结了现有端点效应抑制方法的基础上,提出一种新的方法——基于支持向量机(support vector machine,简称 SVM)延拓和窗函数相结合的方法,弥补了 SVM 延拓依然找不到端点以及窗函数会改变原始信号的缺点。首先,采用 SVM 对原始信号两端分别进行延拓,将延拓后的数据进行加窗处理(中间加矩形窗,延拓数据加海明窗);然后,利用 EMD 方法对加窗后的信号进行分解,得到若干个内禀模态函数(intrinsic mode function,简称 IMF);最后,将 IMF 分量的两端延拓部分去掉,以此来达到抑制端点效应的目的。以正交性为量化评价指标,对比分析了不同方法的性能,通过仿真和实验结果表明,该方法可以更好地抑制端点效应的发生。

**关键词** 经验模态分解;端点效应;支持向量机;窗函数;正交性指标

**中图分类号** TH132.41; TN911.72

## 引 言

EMD 是由 Huang 等<sup>[1]</sup>于 1998 年提出的一种自适应的信号处理算法,其分解得到的 IMF 分量具有真实的物理意义,并且因其具有较好的自适应性、正交性、完备性和 IMF 分量的调制性等优点,而被广泛地应用于处理非线性、非平稳信号。但是 EMD 分解过程中信号的上、下包络线是根据原始信号端点的局部极值点由三次样条曲线插值得到的,而信号的两端不一定存在极值点,所以样条插值在数据序列的两端就会出现发散现象,并且这种发散的结果会逐渐向内“污染”整个数据序列而使得分解结果严重失真,这就是“端点效应”问题<sup>[2]</sup>。为此,专家学者提出几种有效的方法:邓拥军等<sup>[3]</sup>提出利用神经网络来延拓原始信号序列的算法,该算法对短信号具有较好的端点效应抑制效果,但神经网络学习时间长,影响 EMD 的分解速度,不适用于实时信号的处理。Zhao 等<sup>[4]</sup>提出镜像延拓法,该算法对原始信号仅作一次延拓,在处理较长周期信号时可以得到较为理想的抑制效果,但当信号的端点无法确定是极值点时,要截去一部分数据,在处理短信号时效果欠佳。此外,还有 B 样条经验模式分解<sup>[5]</sup>、ARMA

模型<sup>[6]</sup>、窗函数<sup>[7]</sup>、多项式拟合<sup>[8]</sup>等方法,这些方法大多数都可以改善端点效应问题,但由于各种方法延拓结果因人而异,得到的结果也不尽相同,加之非平稳信号的随机性,最终结果都不可避免地存在一些问题。

为了解决 EMD 端点效应问题,笔者提出一种改进算法——基于 SVM 延拓和窗函数相结合的方法,并以仿真信号和实测齿轮裂纹故障信号作了定性和定量分析。结果表明,该方法可以有效地抑制端点效应的发生。

## 1 EMD, SVM 和窗函数法的分析

### 1.1 EMD 及其端点效应

EMD 方法对于非线性、非平稳的信号具有较好地自适应性,其分解过程描述<sup>[9]</sup>如下。

给定连续信号  $x(t)$ , 确定该信号的所有局部极值点,用三次样条曲线把这些局部极值点连接起来形成上、下包络线。若上、下包络线的平均值记为  $m$ , 求出  $x(t) - m = h$ , 则把  $h$  当作新的  $x(t)$ , 重复上述步骤直到满足 IMF 条件时,记  $c_1 = h$ 。  $c_1$  视为一个 IMF,  $r = x(t) - c_1$ ,  $r$  当作新的  $x(t)$ , 重复上述步

<sup>\*</sup> 国家自然科学基金资助项目(51375319);河北省杰出青年科学基金资助项目(E2013210113);河北省百名优秀创新人才支持计划资助项目(BR2-222)

收稿日期:2016-07-10;修回日期:2017-02-24

骤得到  $n$  个满足 IMF 条件的分量, 当  $r_n$  成为一个单调函数不能再提取出满足 IMF 条件的分量时, 循环结束, 由此得到

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

其中:  $r_n$  为残余函数, 表示信号的平均趋势。

由上述可知, EMD 分解过程中多次利用三次样条插值(根据信号的局部极值点来插值)来拟合信号的上、下包络线。由于信号的两端不一定存在极值点, 因此样条曲线在数据两端会出现发散现象, 并随着分解的过程逐渐向内部传播, 最终导致分解得到的 IMF 分量失去真实的物理意义, 从而影响后续故障特征的提取研究。

## 1.2 基于支持向量回归机的序列延拓

支持向量回归机是一种新的机器学习算法<sup>[10]</sup>, 其基本原理是通过非线性映射  $\phi$  将数据  $x$  映射到多维特征空间  $F$ , 并在这个特征空间进行线性回归, 即

$$\begin{cases} f(x) = \omega \cdot \phi(x) + b \\ \phi: \mathbf{R}^n \rightarrow F \end{cases} \quad (\omega \in F) \quad (2)$$

其中:  $b$  为阈值;  $\mathbf{R}^n$  为  $n$  维向量。

对于某一个训练集  $\mathbf{T} = \{(x_1, y_1), \dots, (x_l, y_l)\}$ , 其中,  $l$  为样本数目, 其回归方法如下。

1) 选择精度参数  $\epsilon$ 、惩罚参数  $C$ 、损失函数  $e$  和适当的核函数  $k(x_i, x_j)$ 。

2) 构造回归模型

$$f(x) = \sum_{i=1}^l (\bar{a}_i^* - \bar{a}_i) k(x_i, x) + \bar{b} \quad (3)$$

其中,  $\bar{a}_i^*$ ,  $\bar{a}_i$  由式(4)得到

$$\min \left( \frac{1}{2} \sum_{i=1}^l (a_i^* - a_i)(a_j^* - a_j) k(x_i, x_j) + \epsilon \sum_{i=1}^l (a_i^* + a_i) - \sum_{i=1}^l y_i (a_i^* - a_i) = 0 \right) \quad (4)$$

$\bar{b}$  按式(5)计算, 选择开区间  $(0, C/l)$  内的  $\bar{a}_j$  或  $\bar{a}_k^*$ , 如果选到的是  $\bar{a}_j$ , 则

$$\bar{b} = y_j - \sum_{i=1}^l (\bar{a}_i^* - \bar{a}_i)(x_i \cdot x_j) + \epsilon \quad (5)$$

若选到的是  $\bar{a}_k^*$ , 则

$$\bar{b} = y_k - \sum_{i=1}^l (\bar{a}_i^* - \bar{a}_i)(x_i \cdot x_k) - \epsilon \quad (6)$$

构造回归模型后, 便可以对信号进行前后延拓。对信号  $x(t)$  向后延拓  $m$  个点, 首先确定训练样本数  $l$ , 按一定的规则产生一个训练集, 然后构造如式(3)所示的回归模型, 利用该回归模型得到边界外的第 1 个预测值, 再将第 1 个预测值作为原始数据的新

边界点, 得到第 2 个延拓点, 以此类推, 得到第  $m$  个延拓点, 向前延拓方法与此类似, 这样便完成了对原始数据序列的延拓。

SVM 延拓可以在一定程度上改善端点效应问题, 但延拓后信号的端点仍然不确定, 这样在利用三次样条曲线插值后, 得到的上、下包络线的两端还是有可能出现发散现象, 对分解得到的 IMF 分量依然会有影响。因此, SVM 延拓只是减小了端点效应对分解结果的影响, 其端点效应依然存在。

## 2 端点效应抑制改进算法

为了解决 SVM 延拓之后端点仍然不确定以及窗函数会改变原始信号的缺点, 本研究巧妙地将两种方法结合起来, 提出一种有效抑制端点效应的新方法——基于 SVM 延拓和窗函数相结合的方法, 具体步骤如图 1 所示。首先, 采用 SVM 对原始信号两端分别延拓有限个极大值点和极小值点; 其次, 对延拓后的数据序列加窗处理, 原始信号为加窗后信号的中间部分, 这样就不会改变原来的信号特征; 然后, 对延拓加窗后的信号进行 EMD 分解; 最后, 把分解完的各个 IMF 分量的两端延拓部分去除, 中间的留下, 从而达到抑制端点效应的目的。

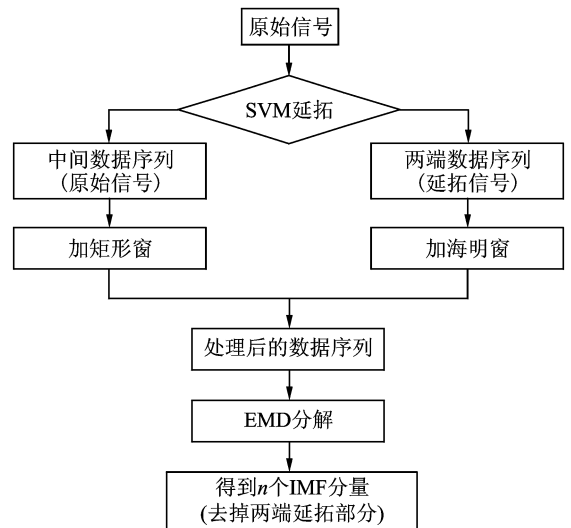


图1 基于 SVM 延拓和窗函数相结合的方法

Fig.1 The flow chart of combined SVM extension and window function

为了验证所述方法的有效性, 采用正交性指标 (index of orthogonality, 简称 IO) 对实验分解精度进行定量分析, 将式(1)中的  $r_n$  写成  $c_{n+1}$ , 那么式(1)可以写成

$$x(t) = \sum_{i=1}^{n+1} c_i \quad (7)$$

对该式两边平方得到

$$x^2(t) = \sum_{i=1}^{n+1} c_i^2(t) + 2 \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} c_i(t)c_j(t) \quad (8)$$

如果分解是正交的,那么式(5)右边的交叉项应该等于零,因此 IO 可定义为

$$x(t)_{IO} = \sum_{i=0}^T \left( \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} c_i(t)c_j(t)/x^2(t) \right) \quad (i \neq j) \quad (9)$$

其中:  $T$  为信号总长度。

文献[1]经过大量的实验指出,正交性指标越小分解精度越高,EMD 端点效应抑制效果也就越好。因此,笔者选取正交性指标作为评价 EMD 端点效应抑制效果的好坏是可行的。

### 3 实验分析

#### 3.1 仿真信号分析

选取仿真信号

$$x = \sin(20\pi t) + \cos(100\pi t) + \sin(200\pi t) + n(t)$$

如图 2(a)所示,该信号是由一段正弦、余弦和均值为 0、方差为 1 的随机噪声叠加而成的,其频率成分分别为 10, 50 和 100 Hz,采样频率为 512 Hz。对其直接进行 EMD 处理(分解得到 8 个 IMF 分量,由于篇幅有限,文中只截取了前 4 个分量),结果见图 2(b)。由图 2 可以看出,各个 IMF 分量存在严重的端点效应问题,其分解结果不能很好地反映原始信号特征。

为了解决端点效应,采用所述方法进行处理。其中,选择精度参数  $\epsilon=0$ ,惩罚参数  $C=\infty$ ,损失函数采用最常用的线性  $\epsilon$  不敏感损失函数,核函数有线性核函数、多项式函数以及高斯函数等,笔者采用最常用的线性核函数。实验中将原始信号序列分成 100 个训练组,并对信号分别向前、后各延拓 50 个点,延拓加窗后的结果见图 3(a),对应的 EMD 分解结果见图 3(b)。

与图 2(b)比较,可以看出图 3(b)中各个 IMF 分量在端点处出现的发散现象均得到了改善,表明提出的方法可以有效地抑制 EMD 分解中出现的端点效应问题,并且各个 IMF 分量也都可以反映原始信号的特征,基本符合实际情况。

为了更直观地验证所述方法的有效性,利用正交性指标对原始 EMD 分解、SVM 延拓及加窗三种

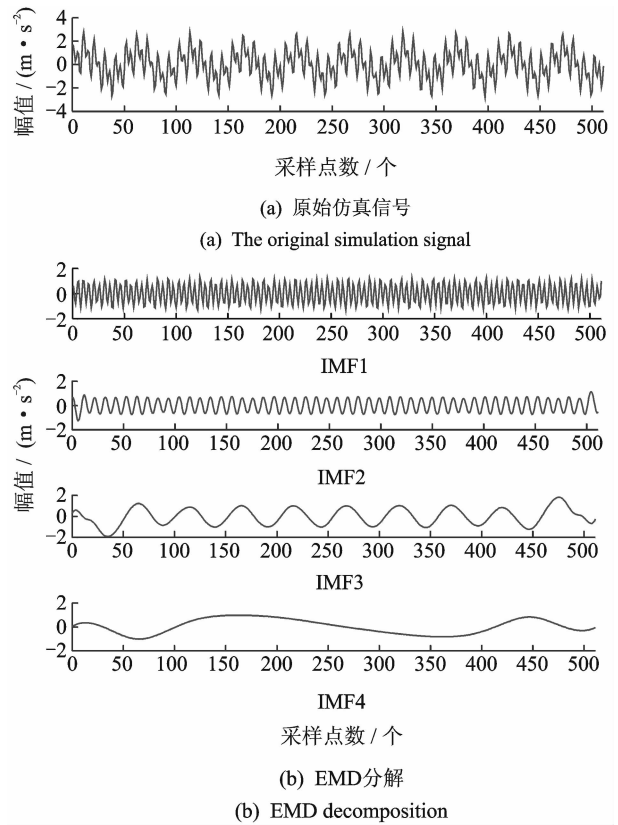


图 2 原始仿真信号及其 EMD 分解结果  
Fig. 2 The original simulation signal and the result of EMD decomposition

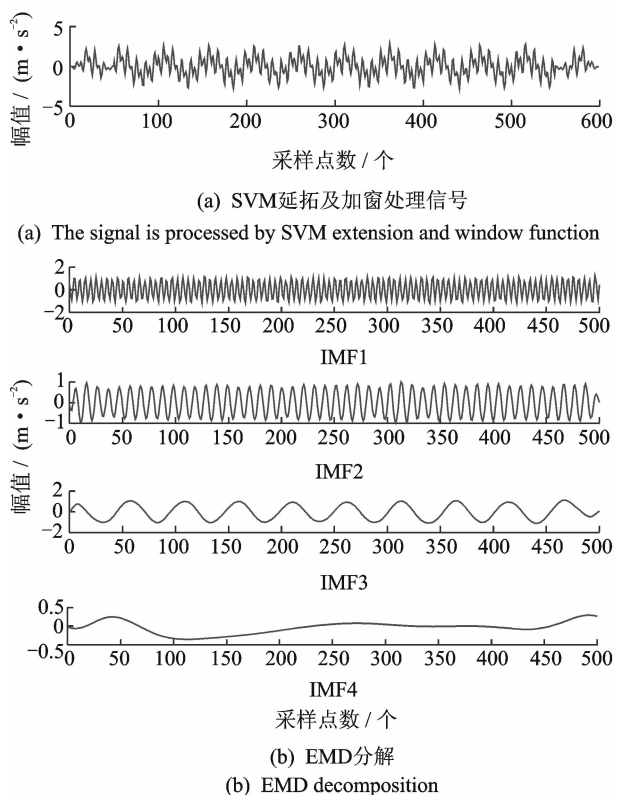


图 3 SVM 延拓及加窗处理的 EMD 结果  
Fig. 3 The results processed by SVM extension and window function

方法的性能进行定量分析,并根据式(9)得出相应的正交性指数,如表1所示。

表1 仿真信号处理结果的正交性能比较

Tab. 1 The orthogonal performance comparison of three kinds of methods for simulation signal

方法	正交性指标
原始 EMD	0.155 8
SVM 延拓	0.130 9
SVM 延拓及加窗	0.096 7

由表1可以看出,未作任何处理的 EMD 分解结果得到的正交性指标较大,SVM 延拓得到的较小,所述方法得到的更小,而正交性指标越低表明分解结果精度越高。由此可知,笔者所述方法可以更好地抑制端点效应的发生。

### 3.2 实测信号分析

为了进一步地确认该方法的有效性,笔者以实测齿轮箱齿轮裂纹故障信号来进行分析。实验采用江苏千鹏诊断系统——QPZZ-II,其输入轴转速为990 r/min,采样频率为20 kHz,采样点数为262 144,故障齿轮为3.2 mm的裂纹程度。为了方便研究截取前800个点,并采用SVM向前、向后延拓40个点。图4所示为实验平台及其传感器安装位置。

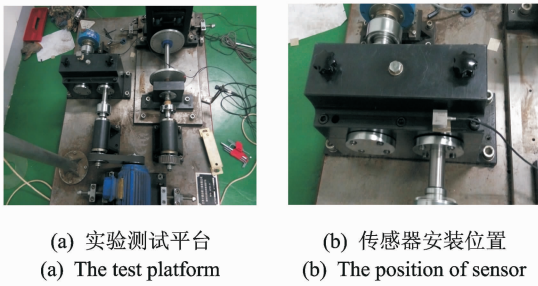


图4 实验测试平台及传感器位置

Fig. 4 Experimental test platform and sensor position

图5(a)所示为原始测试信号的时域波形,其EMD分解的结果见图5(b),由图可以看出,信号两端存在严重的端点发散现象。为了解决端点效应问题,将该数据序列按文中所述方法处理(SVM延拓数据序列的参数选择同前),其SVM延拓加窗处理结果如图6(a)所示。由图可以看出,处理后的信号与原始信号存在着显著的区别,这是在信号两端加窗处理的原因。因此,最后EMD分解得到的各个IMF分量要剔除两端延拓部分,这样就可以真实地反映原始信号特征。处理后数据的EMD分解结果

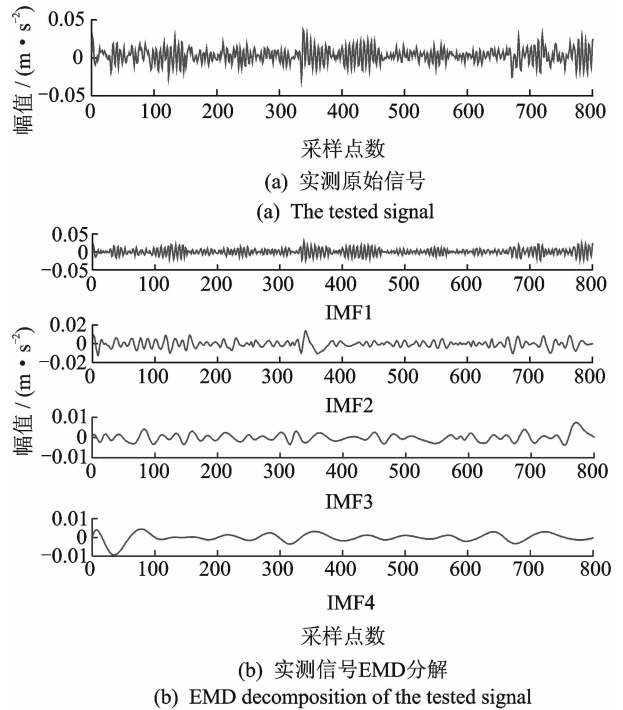


图5 实测原始信号及其EMD分解图

Fig. 5 The tested signal and the results of EMD decomposition

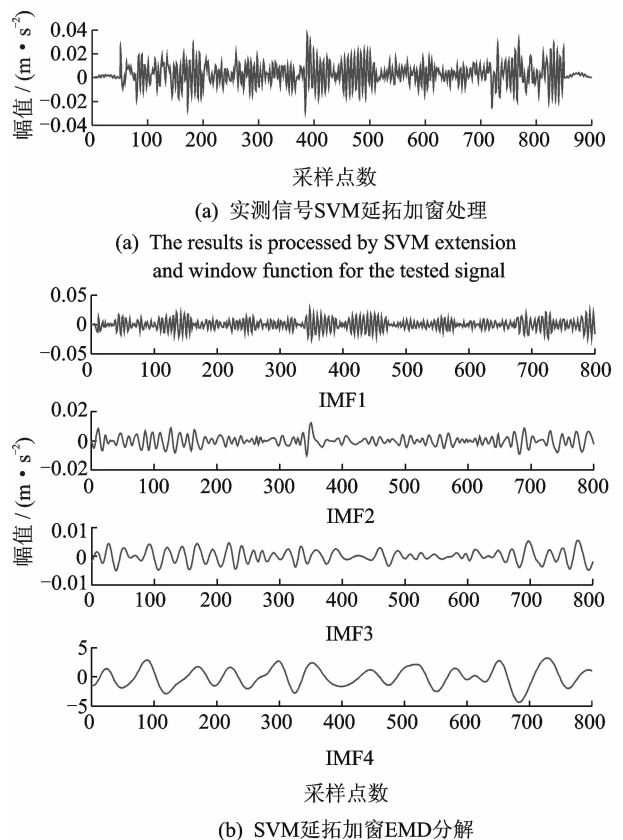


图6 实测信号SVM延拓加窗EMD处理结果

Fig. 6 The results of EMD decomposition combined with SVM extension and window function for the tested signal

见图 6(b),由图可以看出各个 IMF 分量的端点效应都得到了有效的抑制,得到的 IMF 分量具有真实的物理意义,有助于进一步地提取特征。

同样采用正交性指标进行定量分析来验证文中所提方法的有效性,分析对比结果如表 2 所示。

表 2 实测信号处理结果的正交性能比较

Tab. 2 The orthogonal performance comparison of three kinds of methods for tested signal

方 法	正交性指标
原始 EMD	0.057 8
SVM 延拓	0.047 1
SVM 延拓及加窗	0.017 1

由表 2 可知,未作任何处理的 EMD 分解结果得到的正交性指标较大,SVM 延拓得到的较小,本方法得到的最小,充分说明所述方法的有效性。并且由文献[2]可知 SVM 延拓对确定性、平稳随机和非平稳随机信号都有很好的延拓效果,且不依赖于参数的选择,所以本研究所述方法可以很好地应用于大量的连续信号,为进一步地特征提取提供了帮助。

## 4 结束语

由于在 EMD 分解过程中多次利用了三次样条插值,从而导致信号分解出现失真现象,即端点效应,并且这种现象会严重地影响整个数据序列。为此,笔者提出一种有效抑制端点效应的方法——基于 SVM 延拓和窗函数相结合的方法,该方法弥补了 SVM 延拓之后依然找不到端点以及窗函数会改变原始信号的缺点,仿真信号和实测信号均验证了该方法的有效性。以正交性指标做定量分析,与原始 EMD 分解和 SVM 延拓方法相比,该方法可以更好地抑制端点效应,得到的 IMF 分量更加真实地反映了原始信号的特征,为后续齿轮故障诊断中精确地提取故障特征频率提供了帮助。

## 参 考 文 献

- [1] Huang N E, Shen Z, Long S R. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. The Royal Society, 1998, 454: 903-995.
- [2] 于德介,程军圣,杨宇. 机械故障诊断的 Hilbert-Huang 变换方法[M]. 北京:科学出版社, 2006: 52-55.
- [3] 邓拥军,王伟,钱成春,等. EMD 方法及 Hilbert 变换中边界问题的处理[J]. 科学通报, 2007, 46(3): 257-263.  
Deng Yongjun, Wang Wei, Qian Chengchun, et al.

- Boundary-processing-technique in EMD method and Hilbert transform[J]. Science Bulletin, 2007, 46(3): 257-263. (in Chinese)
- [4] Zhao Jinping. Improvement of the mirror extending in empirical mode decomposition method and the technology for eliminating frequency mixing[J]. High Technology Letters, 2002, 8(3): 40-47.
- [5] 邓蕾,胡小林,李锋,等. 基于支持向量机的 S-EMD 端点效应消除方法[J]. 振动、测试与诊断, 2011, 31(3): 344-347.  
Deng Lei, Hu Xiaolin, Li Feng, et al. Support vector machines-based method for restraining end effects of B-spline empirical mode decomposition[J]. Journal of Vibration, Measurement & Diagnosis, 2011, 31(3): 344-347. (in Chinese)
- [6] 杨建文,贾民平. 希尔伯特-黄谱的端点效应分析及处理方法研究[J]. 振动工程学报, 2006, 19(2): 283-288.  
Yang Jianwen, Jia Minping. Study on processing method and analysis of end problem of Hilbert-Huang spectrum[J]. Journal of Vibration Engineering, 2006, 19(2): 283-288. (in Chinese)
- [7] 徐力彬,宋余庆,刘毅. 基于镜像延拓和窗函数的端点效应抑制方法[J]. 计算机工程, 2015, 41(4): 112-116.  
Xu Libin, Song Yuqing, Liu Yi. Restraining method for end effect based on mirror extension and window function[J]. Computer Engineering, 2015, 41(4): 112-116. (in Chinese)
- [8] 刘慧婷,张旻,程家兴. 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.  
Liu Huiting, Zhang Yu, Cheng Jiaying. Dealing with the end issue of EMD based on polynomial fitting algorithm [J]. Computer Engineering & Applications, 2004, 40(16): 84-86. (in Chinese)
- [9] 张志刚,石晓辉,施全,等. 基于改进 EMD 和谱峭度法滚动轴承故障特征提取[J]. 振动、测试与诊断, 2013, 33(3): 478-482.  
Zhang Zhigang, Shi Xiaohui, Shi Quan, et al. Fault feature extraction of rolling element bearing based on improved EMD and spectral kurtosis[J]. Journal of Vibration, Measurement & Diagnosis, 2013, 33(3): 478-482. (in Chinese)
- [10] Thorsten B J. Making large-scale SVM learning practical[J]. Technical Reports, 1998, 8(3): 499-526.



**第一作者简介:**郝如江,男,1972 年 2 月生,博士、教授。主要研究方向为机车车辆状态监测与故障诊断。曾发表《滚动轴承故障信号的多尺度形态学分析》(《机械工程学报》2008 年第 44 卷第 11 期)等论文。

E-mail: haorj@stdu.edu.cn

