

文章编号:1004-9037(2012)02-0196-06

一种免疫算法与 SVR 的 Hilbert-Huang 边界优化

姚 莉¹ 李磊民² 黄玉清¹

(1. 西南科技大学信息工程学院, 绵阳, 621000; 2. 西南科技大学国防科技学院, 绵阳, 621000)

摘要: Hilbert-Huang 变换(Hilbert-Huang transform, HHT)在对信号进行经验模态分解(Empirical mode decomposition, EMD)和对各内禀模态函数(Intrinsic mode function, IMF)进行 Hilbert 变换时都会出现边界问题。为了克服该问题,本文提出了基于离散均匀免疫算法(Discrete uniform immune algorithm, DUIA)和支持向量回归(Support vector regression, SVR)的 HHT 边界优化方法。该方法采用 DUIA 优化 SVR 的参数,并利用 SVR 对数据延拓,以有效分析 HHT 边界问题。通过对正弦叠加信号和实际信号的仿真分析表明:所提出的算法可有效解决 HHT 中存在的边界问题,且其效果优于 SVR 的数据延拓方法。

关键词: 希尔伯特-黄变换; 遗传算法; 支持向量回归机

中图分类号: TN911.7

文献标识码: A

Hilbert-Huang Boundary Optimization of Immune Algorithm and SVR

Yao Li¹, Li Leimin², Huang Yuqing¹

(1. School of Information Engineering, Southwest University of Science and Technology,
Mianyang, 621000, China;

2. School of National Defence Science & Technology, Southwest University of Science and Technology,
Mianyang, 621000, China)

Abstract: The Hilbert-Huang transform (HHT) boundary problem appears when the signal is decomposed by the empirical mode decomposition (EMD) method as well as the intrinsic mode function (IMF) in Hilbert transform. Therefore, the HHT boundary optimization method based on discrete uniform immune algorithm (DUIA) and support vector regression (SVR) is proposed to overcome the problem. To effectively analyze the boundary problem of HHT, the scheme can use DUIA to optimize parameters of SVR, and then predict the signal by the trained optimal SVR model. For the sine superposition and practical signals, the corresponding simulation results demonstrate that the proposed algorithm can effectively solve the boundary problem of HHT, and its performance is better than the prediction method by SVR.

Key words: Hilbert-Huang transform (HHT); genetic algorithm; support vector machine (SVM)

引 言

Hilbert-Huang 变换(Hilbert-Huang transform, HHT)是近年来发展起来的一种新的时间序列信号分析方法。HHT 的创新点是引入了基于信号局部特征的内禀模态函数(Intrinsic mode function, IMF),以获得具有物理意义的瞬时频率。目前 HHT 已广泛用于信号处理和故障诊断等领域的研究^[1]。

虽然 HHT 非常适合处理非平稳信号,但在应用该变换时存在边界问题。首先是进行经验模态分解(Empirical mode decomposition, EMD)时,在数据的两端会产生发散现象,并且这种发散会在 EMD“筛选”的过程中向内扩散,逐渐“污染”整个数据序列,使得 HHT 结果严重失真;其次是在对各个 IMF 分量进行 Hilbert 变换时,由于 Hilbert 变换算法中的加窗效应,信号的两端也会出现严重的边界问题。边界问题直接导致 Hilbert 谱不能反映原始信号的真实特征^[2-3]。

目前,提出的一些边界优化方法,要有采用AR模型对数据延拓^[4]、神经网络对数据进行延拓^[5]、边界波形匹配预测法^[6]和极值点延拓法^[7],这些方法对边界问题都有一定的优化效果,但同时也存在着缺陷。其中,神经网络具有局部极小点、过学习以及结构和类型的选择过分依赖于经验等固有的缺陷,而SVR具有较强的理论依据和较好的泛化性能,且克服了神经网络的上述缺陷,目前已经广泛用于故障诊断、回归估计和数据分类等领域^[8]。当前采用SVR数据延拓方法来解决HHT边界问题的研究仅是在参数选定情况下的信号进行分析得到较满意的结果,对于如何选择参数及不同参数对延拓效果的影响没有做进一步研究^[9]。基于此,本文考虑SVR模型参数及训练样本长度的多目标优化选择。通过DUIA优化SVR参数,然后采用建好的SVR模型对原始数据的延拓来实现HHT边界优化。

1 Hilbert-Huang 变换

HHT主要由两部分组成,即EMD和Hilbert变换。EMD通过多次的移动过程进行筛选,一方面消除模态波形的叠加,另一方面使波形轮廓更加对称,使得每一个IMF具有如下两个特征:(1)极值点(极大值或极小值)数目与跨零点数目相等或者最多相差一个;(2)由局部极大值构成的上包络和由局部极小值构成的下包络的平均值为零。

上述的IMF的两个特征也是EMD分解结束的收敛准则。由于EMD的自适应性,因此其分解快速有效,同时EMD是基于信号的局部变化特性,从而可用于非线性和非平稳过程分析。

原始信号 $X(t)$ 可表示为所有的IMF c_i 及余量 r_n 之和,即

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

任意信号经EMD分解得到的IMF数目通常有限,对所有的IMF进行Hilbert变换得到Hilbert谱,记 $x_i(t) = c_i(t)$, $x_i(t)$ 的Hilbert变换为

$$y_i(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x_i(\tau)}{t - \tau} d\tau \quad (2)$$

根据Hilbert变换的定义, $x_i(t)$ 和 $y_i(t)$ 为复共轭对,由其构成解析函数 $z_i(t)$ 为

$$z_i(t) = x_i(t) + iy_i(t) = a_i(t)e^{i\theta_i(t)} \quad (3)$$

幅值函数

$$a_i(t) = |x_i(t) + iy_i(t)| \quad (4)$$

相位函数

$$\theta_i(t) = \arctan \frac{y_i(t)}{x_i(t)} \quad (5)$$

瞬时频率函数

$$\omega_i(t) = \frac{d\theta_i(t)}{dt} \quad (6)$$

其中幅值函数表示信号每个采样点的瞬时幅度(能量),相位函数表示信号每个采样点的瞬时频率,这里的振幅 $a_i(t)$ 和频率 $\omega_i(t)$ 均是时间的函数。在振幅-频率-时间的三维空间中,振幅在频率-时间平面上的分布称为Hilbert振幅谱($H(t, \omega)$),简称Hilbert幅谱。

2 算法模型实现

本文提出的模型算法实现框图如图1所示。图1中,信号输入模块负责各项参数的初始化及原始信号的输入与采集,如种群规模、进化代数、采样点数等,并将处理后的信息输出,以配合后面离散均匀免疫算法(Discrete uniform immune algorithm, DUIA)模块的应用。SVR预测能力分析模块获得的信息(如SVR预测数据的误差)作为DUIA模块优化参数组合的适应度信息,DUIA模块根据该信息迭代优化出使SVR具有很好预测能力的最优解,即建立了最优的SVR模型。SVR模型训练完成后,原始波形延拓模块采用已建立好的SVR模型对原始数据进行预测延拓,并将信息输入到HHT模块。

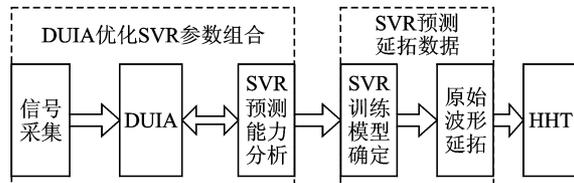


图1 模型算法实现

2.1 DUIA 算法描述及性能分析

本文提出的DUIA算法经过参数及种群的初始化,并计算个体的拥挤度,选出非支配集种群作为活性抗体种群,然后对选出的活性抗体种群进行离散均匀化处理,通过选择、交叉、变异并逐步迭代,以实现种群多样性和快速收敛性。

将DUIA算法与当前性能最好的非支配邻域免疫算法(Nondominated neighbor immune algorithm, NNIA)^[10]对比,并根据ZDT系列测试函数来验证DUIA算法的优越性。NNIA模拟了免疫响应中多样性抗体共生、少数抗体激活的现象,通过一种基于非支配邻域的个体选择方法,只选择少数相对孤立的非支配个体作为活性抗体,根据活性抗体的拥挤度进行比例克隆复制,对克隆后的抗体群

采用了有别于现有的遗传算法的重组及变异操作,对于 ZDT 问题得到了较其他多目标遗传算法都好的性能^[11]。

ZDT 是一系列具有代表性的测试函数,本文选取 ZDT₂, ZDT₃ 测试。其中 ZDT₂ 有一个凹的最优 Pareto 前沿,较易收敛;ZDT₃ 具有离散型,其 Pareto 最优前沿是由几个非连续的凹凸区域组成。

在实验时,两种算法选用相同的参数:进化代数 40,种群规模 100,交叉概率 1,变异概率为 0.03,且都采用实数编码。每个算法对各个测试函数独立运行一次。图 2,3 为实验结果,图中菱形线代表 DUIA 的结果,实线代表 Pareto 前沿,星形线代表 NNIA 的结果。

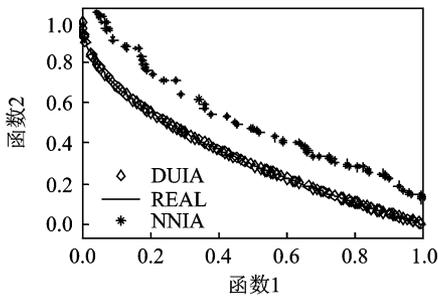


图 2 ZDT₂ 函数

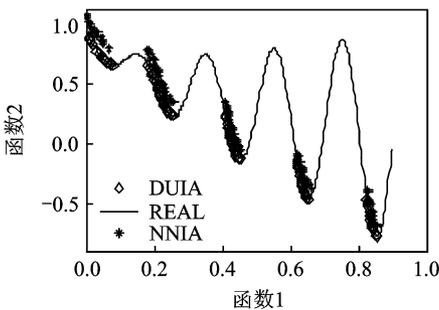


图 3 ZDT₃ 函数

由仿真运行结果可以看出,采用 DUIA 得到的 Pareto 曲线分布更加均匀,接近标准前沿,且解的精度较高;NNIA 算法对 Pareto 前沿的接近度为 40%~60%左右,且种群分布不均匀,只是反应了曲线的明显走势。仿真结果证明,在参数一致、种群数目相等的情况下,DUIA 的性能明显优于 NNIA。

2.2 DUIA 优化 SVR 参数实现

本文以建立最优的 SVR 模型为目标,并通过实数编码,用 SVR 预测数据误差作为指标并实现参数优化选择,算法实现步骤为:

(1)优化参数进行随机初始化:本文选取的以多项式核为例,对核参数 c , p 及 SVR 训练步长 l

进行浮点数编码;

(2)计算种群中个体的适应度值。种群中每个个体即为 SVR 参数组合,利用训练样本对 SVR 进行模型训练,以 SVR 对训练样本的预测误差为个体适应度;

(3)根据适应度值进行非支配排序,选出非支配种群;

(4)将非支配种群离散均匀化处理,然后再进行选择、交叉、变异;

(5)以种群最大进化代数为终止条件,反复执行步骤(2~4)。最小适应度值对应的参数组合即为最优解。

2.3 支持向量回归预测

SVR 算法是通过核函数升维后,在高维空间中构造线性决策函数来实现线性回归。本文在经过多次实验验证后,选取多项式核函数进行回归预测。

采用 SVR 对数据序列的延拓分两步进行,第 1 步是预测学习的过程;第 2 步是延拓的过程。对于原始数据序列首先要按照一定的规则产生一个学习样本矩阵和与之对应的目标向量。假设数据序列的采样时间间隔是 t ,采样点数为 N ,得到的数据序列为 $x(0), x(t), \dots, x(it), \dots, x((N-1)t)$ 。该数据序列也是对未来值进行预测时的输入序列。由 $n(n \leq N-1)$ 时刻的前 m 个值预测第 n 个值的问题可表示为寻找如下的对应关系 f 的问题

$$x_n = f(x_{n-1}, x_{n-2}, \dots, x_{n-m}) \quad (7)$$

式中 x_i 是 $x(it)$ 的简写, $i=1, 2, \dots$ 。构造如下的训练样本对: $(x_1, x_2, \dots, x_m; x_{m+1}), (x_2, x_3, \dots, x_{m+1}; x_{m+2})$, 并依此类推,由 k 个训练样本构造 $k-m$ 个训练样本对。用构造的训练样本训练 SVR 模型,同时以 SVR 对数据的预测误差为指标,利用 DUIA 进行 SVR 参数组合优化,建立最优的 SVR 模型。当模型训练完成后,接着进行数据的延拓。对未来值的预测采用式(7)的关系,产生如下的结果 $(x_{n-m+1}, \dots, x_{n-1}, x_n; x_{n+1}^*), (x_{n-m+2}, \dots, x_{n-1}, x_n, x_{n+1}^*; x_{n+2}^*)$, 后续各步的预测依次类推,并由此形成对原始数据序列的延拓。其中, x_i 为第 i 点的真实值, x_i^* 为第 i 点的预测值, $i=1, 2, \dots$ 。

3 仿真分析

3.1 正弦叠加信号仿真

针对非线性、非平稳正弦信号进行仿真实验,分析了本文提出的边界优化的性能。仿真信号为

$$x(t) = \sin(6\pi \times t) + 0.4\sin(22\pi \times t) + 0.9\sin(40\pi \times t) \quad (8)$$

该仿真信号为正弦叠加信号,3 个频率分别为 3,11,20 Hz,幅值分别为 1,0.4,0.9 的正弦信号叠加。仿真实验中选取采样频率 500 Hz,采样长度 500 点。

图 4 为原始时间序列,图 5 为采用本文算法对原始信号 50 个采样点的延拓仿真,从图 5 中可以看出预测信号波形与原始信号波形基本拟合。图 6(a,b)为希尔伯特谱,图 7(a,b)为 Hilbert 三维谱图,图 8(a,b)为 EMD 分解结果。3 组图形所含信息相同。此处,图 6(a),7(a)和图 8(a)为采用 SVR 对数据序列进行延拓后进行 HHT 得到的效果图,SVR 核函数及参数采用文献[9]中的线性核函数,精度参数 0,惩罚参数 ∞ ,训练样本点数 100^[9]。由图 6(a)可知,采用参数固定的 SVR 延拓数据后进行 HHT,信号的两端点处仍然存在频率发散现象,并且向内扩散,越靠近边界,发散现象越严重,只有信号的中间部分准确反映了信号的特征。图 6(b)采用本文提出的算法进行数据延拓,并用 HHT 正确分解出 3 个正弦信号,且基于数据延拓后的 Hilbert 谱消除了边界问题,由图 7(a)可知,采用参数固定的 SVR 延拓数据后进行 HHT 得到的三维频谱信息存在较多虚假成分,不能清晰地分辨出

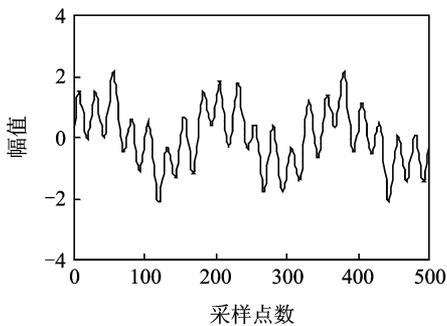


图 4 原始信号时间序列图

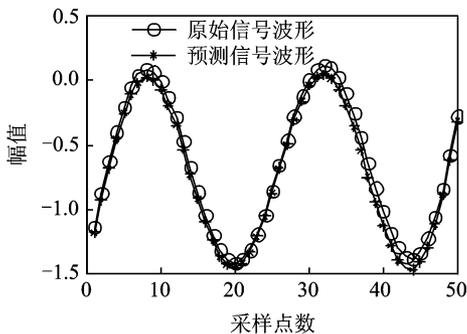
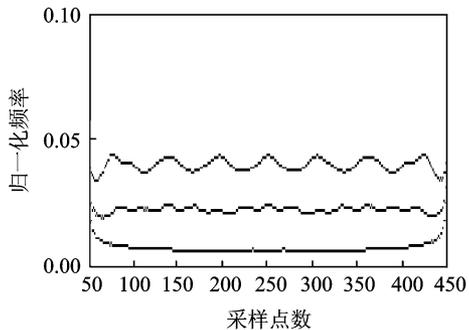
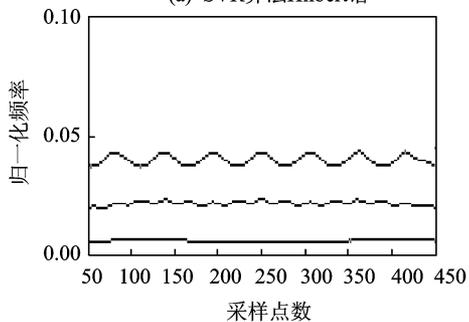


图 5 本文算法延拓仿真

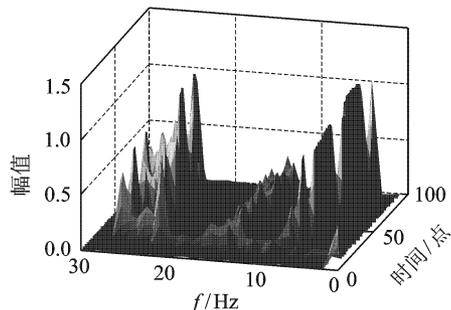


(a) SVR算法Hilbert谱

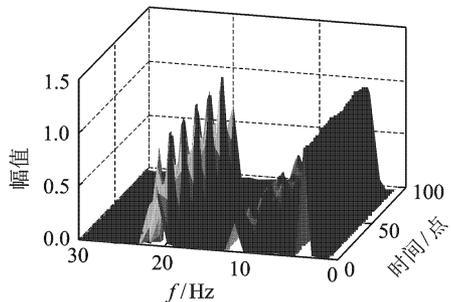


(b) 本文算法Hilbert谱

图 6 两种算法的 Hilbert 谱



(a) SVR算法Hilbert 三维谱图



(b) 本文算法Hilbert三维谱图

图 7 两种算法的 Hilbert 三维谱图

频谱信息,而图 7(b)表明,采用本文算法可以有效分辨出原始时间序列的频谱特征。图 8(a,b)从上到下分别为第 1~3 个分量,省略了余项。图 8(a)的 3 个分量在波形的两端都出现了较为严重的边界问题,而图 8(b)采用本文算法,很好地解决了边界问题。

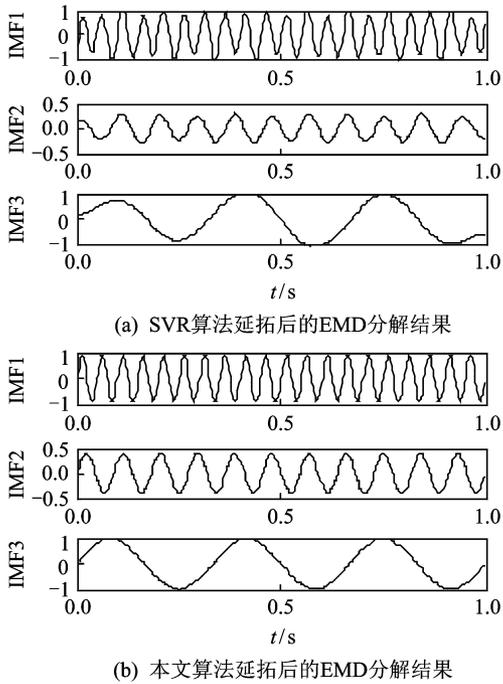


图 8 两种算法延拓后的 EMD 分解结果

3.2 振动信号分析

为验证本文算法,采用了 Loparo 实验室测试数据^[12]。该实验室实验设备由功率为 1.5 kW 的电动机、扭矩传感器/译码器、测振仪和电器控制装置组成。由测试轴承支撑着电机轴承,以每秒 12 000 样本点和 48 000 样本点的采样速率,分别在测试轴承上收集直径为 0.017 78, 0.035 56, 0.050 8, 0.071 12, 0.101 6 cm 的单点故障数据,以获得驱动端轴承数据。本文所用数据采自电机驱动端轴承、型号为 6205-2RS JEM SKF 滚动轴承振动信号,电机负载 0 kW、轴承内圈转速 1 797 r/min、采样频率 12 kHz、每组数据长度 500,采集滚动轴承振动信号 20 组。

分别对 20 组振动信号进行仿真后取平均效果,利用 3.2 节方式构造 SVR 训练样本,并使用同样的初始化参数进行离散均匀免疫算法的迭代优化,可得 20 个最优值,以误差为标准按照升序排列。本文取前 5 个个体的仿真结果如表 1 所示。

由表 1 可得该信号的 SVR 的最优参数为: $l=54$, $c=23.522$, $p=0.002\ 716$ 。

表 1 离散均匀免疫算法的训练仿真结果

参数	1	2	3	4	5
l	54	35	36	36	72
c	23.522	1.595	1.175	1.175	23.484
$P/10^{-3}$	1.359	12.324	14.439	14.434	1.361
误差 $e/10^{-3}$	2.716	2.722	2.729	2.730	2.784

确定 SVR 模型后,训练模型算法,并对数据向前向后延拓,向前延拓与向后延拓方法相同,本文只对向后的 50 个延拓点进行比较。图 9 为测得的电机滚动轴承振动加速度信号的时域波形。其中前 450 个点用于训练模型,后 50 个点用于预测。图 10,11 分别为采用本文算法预测和 SVR 算法(参数与 3.1 节相同)预测后与原始数据的比对。由 3.1 节可知,在得到良好的信号延拓后,本文算法能有效地进行信号特征分析,解决 HHT 边界问题,因此该部分重点验证算法的延拓性能,其进一步的特征分析不再赘述。

由图 10,11 知,基于实际信号的复杂性,本文算法对于实际信号的预测能很好地反映曲线走势,拟合度大于 80%,而 SVR 预测信号的结果没有很好地反映曲线走势,对信号的拟合度小于 30%,本文算法的误差相对于 SVR 预测误差仍然减小了约 1/3。表 2 对这两种算法的比对表明,本文算法的训

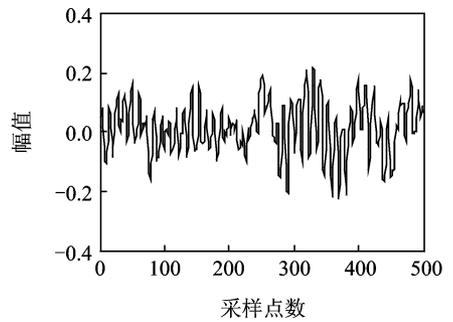


图 9 原始图形

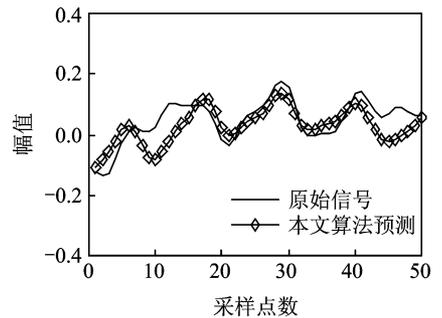


图 10 本文算法预测与原始波形比较

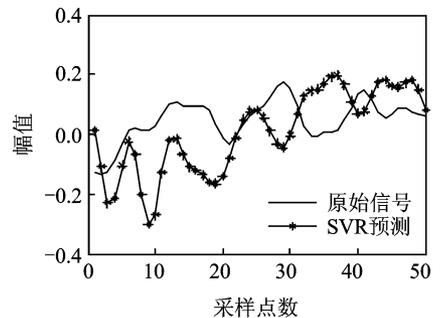


图 11 SVR 预测与原始波形比较

表2 本文算法与SVR性能比较(滚动轴承信号分析)

方法	训练样本长度	预测值与真实值均方误差
本文算法	36	0.002 3
SVR	100	0.020 2

训练样本点数比SVR减少了50%,均方误差也减小了20%。由此可知,对实际中的振动信号,采用本文提出的方法延拓数据同样能取得较好效果。

4 结束语

针对HHT的边界问题,本文提出了一种新的优化方法,对目前SVR波形延拓法进行了改进,克服了参数固定对波形延拓效果的影响问题。仿真结果表明,本文提出的算法对正弦叠加信号和实际信号均取得了较好的延拓效果,实现了HHT的边界优化,并且效果明显优于SVR方法。在实际应用过程中,实际信号的复杂性等情况使得波形预测不准确,会对后续HHT信号特征提取的准确度造成负面影响,因此算法还需进一步改进。

参考文献:

- [1] 王维珍. Hilbert-Huang变换在旋转机械故障诊断中的应用研究[D]. 河北:华北电力大学,2006.
Wang Weizhen. Studying and application of Hilbert-Huang transform in the fault diagnosis of rotating machinery[D]. Hebei: North China Electric Power University, 2006.
- [2] 杨宇. 基于EMD和支持向量机的旋转机械故障诊断方法研究[D]. 湖南:湖南大学,2005.
Yang Yu. Fault diagnosis of rotating machinery based on EMD and support vector machine[D]. Hunan: Hunan University, 2005.
- [3] 陈娟. Hilbert-Huang变换及其在信号处理中的应用[D]. 大连:大连理工大学,2006.
Chen Juan. Hilbert-Huang transform and its application in signal processing[D]. Dalian: Dalian University of Technology, 2006.
- [4] 胡劲松,杨世锡. EMD方法基于AR模型预测的数据延拓与应用[J]. 振动、测试与诊断,2007,27(2):116-120.
Hu Jinsong, Yang Shixi. AR model prediction-based EMD method and its application to data extension[J]. Journal of Vibration, Measurement & Diagnosis, 2007,27(2):116-120.
- [5] 瞿伟廉,程磊. 应用径向基函数神经网络处理EMD方法中的边界问题[J]. 华中科技大学学报,2006,29(4):1-4.

- Qu Weilian, Cheng Lei. Disposal of boundary problem in EMD method applying RBF neural network[J]. Journal of Huazhong University of Science and Technology, 2006,29(4):1-4.
- [6] 邵晨曦,王剑,范金锋,等. 一种自适应的EMD端点延拓方法[J]. 电子学报,2007,35(10):1944-1948.
Shao Chenxi, Wang Jian, Fan Jinfeng, et al. A self-adaptive method dealing with the end issue of EMD[J]. Acta Electronica Sinica, 2007, 35(10):1944-1948.
- [7] Huang N E, Chern C C, Huang K, et al. A new spectral representation of earthquake data: Hilbert spectral analysis of station TUC129[J]. Bulletin of the Seismological Society of America, 1999,91(5):1310-1338.
- [8] 陶卿,姚穗,范劲松,等. 一种新的机器学习算法:Support Vector Machines[J]. 模式识别与人工智能,2000,13(3):285-289.
Tao Qing, Yao Sui, Fan Jinsong, et al. A new kind of machine learning algorithm: support vector machines[J]. Pattern Recognition and Artificial Intelligence, 2000,13(3):285-289.
- [9] 程军圣,于德介,杨宇. 基于支持向量回归机的Hilbert-Huang变换端点效应问题的处理方法[J]. 机械工程学报,2006,42(4):64-65.
Cheng Junsheng, Yu Dejie, Yang Yu. Process method for end effects of Hilbert-Huang transform based on support vector regression machine[J]. Chinese Journal of Mechanical Engineering, 2006, 42(4):64-65.
- [10] Gong Maoguo, Jiao Licheng, Du Haifeng, et al. Multi objective immune algorithm with nondominated neighbor based selection[J]. MIT Press Journals, 2008(7):225-255.
- [11] 公茂果,焦李成,杨咚咚,等. 进化多目标优化算法研究[J]. 软件学报,2009,20(2):271-289.
Gong Maoguo, Jiao Licheng, Yang Dongdong, et al. Research on evolutionary multi-objective optimization algorithms[J]. Journal of Software, 2009, 20(2):271-289.
- [12] Laparo K A. Bearing data center. Case western reserve university [EB/OL]. [2008-6-20]. <http://www.eecs.cwru.edu/laboratory/learning>.

作者简介:姚莉(1984-),女,硕士研究生,研究方向:数字信号处理、智能控制,E-mail: yaoliwb@163.com;李磊民(1960-),男,教授,研究方向:信号处理及硬件实现、智能控制、图像处理;黄玉清(1962-),女,教授,研究方向:信号处理、图像处理、智能控制。