

# 采用支持向量回归抑制噪声的经验模态分解方法

宋 剑,邱晓晖

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

**摘 要:**在实际信号分解中,经验模态分解(EMD)是对噪声敏感的,往往会分离出一些虚假的本征模函数,对信号的分析产生一定影响。为了提高 EMD 分解的正确率,减少其出现虚假本征模函数的情况,文中提出了一种基于支持向量回归(SVR)的去噪方法。先对一次 EMD 分解结果进行 SVR 逐层滤波并且对信号进行重组,然后利用 EMD 方法对重组信号进行二次分解。实验表明,二次分解结果已经非常接近于理想的分解结果,不会出现虚假 IMF。这种分解方法对噪声不敏感,能有效提高 EMD 方法对噪声的容忍度。

**关键词:**信号处理;经验模态分解;支持向量回归;噪声抑制

中图分类号:TN911.7

文献标识码:A

文章编号:1673-629X(2014)11-0122-05

doi:10.3969/j.issn.1673-629X.2014.11.031

## Empirical Mode Decomposition Method Using Support Vector Regression to Suppress Noise

SONG Jian, QIU Xiao-hui

(School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** Empirical Mode Decomposition (EMD) is sensitive to noise in actual signal decomposition. False intrinsic mode functions tend to exist in decomposition results, leading to negative effects to signal analysis. To improve the accuracy of EMD and reduce the condition of existing the false intrinsic mode function, in this paper, a new de-noising method based on Support Vector Regression (SVR). Firstly, decompose the signal with EMD, filtering every IMF by SVR and recombining the regression results. Then decompose the recombined signal with EMD once more time. Experimental results show that the secondary decomposition result is very close to ideal situation and no false IMF is appeared in it. This method is not sensitive to noise, which can effectively improve the tolerance of EMD to noise.

**Key words:** signal processing; EMD; support vector regression; noise suppression

## 0 引言

1998年,N. E. Huang等人提出了经验模态分解(EMD)方法,该方法具有完全自适应的特点,无需预先设定基函数,即可对信号进行分解,从而克服了小波变换中如何选取小波基的困难;同时,分解的停止条件也是完全基于信号本身的,无需设定分解层数,对于平稳或非平稳信号,都有非常好的效果。

EMD方法将信号分解为若干本征模函数(Intrinsic Mode Function, IMF),在文献[1]中,Huang等人明确提出了IMF的特征,但是由于其难以用数学公式表示的特点,一般不用作IMF的筛选判定停止条件。目前主流的停止条件有,利用SD的判断条件、基于零极点判断条件以及基于幅值比的判断条件<sup>[2]</sup>。

当信号中频率成分间隔较大时,上述方法都可以取得很理想的结果。然而实际信号通常都会含有频率分布很广的噪声,如果不采用合适的判断条件,就会出现过分解或者分频率混叠的现象<sup>[3-4]</sup>,原本应该出现的IMF淹没在噪声中,或者出现多个虚假的IMF。这样的分解结果对于信号分析来说是没有意义的。

为了消除这样的影响,可以对信号进行预处理,滤除其中的噪声。目前有很多种方法滤除噪声,例如Fourier滤波、FIR滤波、频域分析、小波阈值去噪等等。

收稿日期:2013-11-26

修回日期:2014-03-05

网络出版时间:2014-07-28

基金项目:江苏省自然科学基金(BK2011789);东南大学毫米波国家重点实验室开放课题(K201318)

作者简介:宋 剑(1989-),男,硕士研究生,研究方向为信号与信息处理、电磁兼容;邱晓晖,教授,通信作者,研究方向为信号与信息处理、电磁兼容等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140728.1222.017.html>

对于非平稳噪声,小波阈值去噪法通常具有较好的效果,但是在噪声特性未知的情况下,如何选取小波基函数,以及确定分解层数,成为了不得不面对的问题。亦有文献[5]提出一种基于 EMD 的硬去噪方式,即将信号经过 EMD 分解后直接去除第一层分解结果,这样固然可以去除大部分噪声,但是信号中部分高频成分也有可能随之损失。

文中提出了一种基于支持向量回归(SVR)的新型 EMD 分解方法,支持向量机具有训练速度快、分类效果好、泛化能力强的优点<sup>[6]</sup>。将各层本征模函数作为输入训练样本,采用 SVM 的回归功能对各层的本征模函数进行回归处理,滤除大部分的噪声,然后对信号进行重组,再次利用 EMD 进行分解,从而得到最终的分解结果。在无需特别设定停止判断条件的情况下,有效提升了 EMD 方法对噪声的容忍度。

## 1 EMD&SVR

### 1.1 EMD

经验模态分解是一种基于信号本身的自适应信号分解方法,EMD 将信号分解为若干本征模函数,每个 IMF 满足如下条件:

- (1)在整个时间范围内,局部极值点和过零点的数目必须相等或至多相差 1;
- (2)任意时刻点,其上包络线和下包络线的均值为零。

每个本征模函数都满足以上两个条件,可以得到具有实际意义的瞬时频率。

大多数需要分析的信号都不是本征模函数,因此,必须通过对信号的不断“筛选”,得到一系列组成信号的 IMF,直至余项单调或低于阈值。分解过程概括如下:

- (1)初始化  $X(t) = s(t)$ ,  $i = 1$ ;
- (2)采用三次样条函数,获取  $X(t)$  的上下包络线  $up(t)$  和  $low(t)$ ,并求得其包络均值  $m(t) = \frac{1}{2}(up(t) + low(t))$ ,  $c(t) = X(t) - m(t)$ ;
- (3)若  $c(t)$  不满足停止条件,  $X(t) = X(t) - c(t)$ ,跳至步骤 2;
- (4)若  $c(t)$  满足停止条件,则  $c(t)$  为一个本征模函数,  $imf_i(t) = c(t)$ ,  $i = i + 1$ ,  $X(t) = X(t) - c(t)$ ;
- (5)若  $X(t)$  非单调,则跳至步骤 2;
- (6)若  $X(t)$  单调,则  $r(t) = X(t)$ ,分解结束。

通过上述分解过程,信号  $s(t)$  可分解为如下形式:若干本征模函数  $imf_i(t)$  以及一个余项  $r(t)$ :

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

其中,  $imf_i(t)$  为第  $i$  次分解出的本征模函数;  $r(t)$  为余项。

### 1.2 支持向量机

支持向量机(Support Vector Machine, SVM)是一种以统计学习理论(Statistical Learning Theory, SLT)为基础的监督式学习的方法,于 1995 年由 V. Vapnik 等人首先提出<sup>[7]</sup>。SVM 本质上是一种最小化结构风险的分类器,由线性分类器发展而来。一个二元线性分类问题描述如下<sup>[8]</sup>:

在一个平面上,存在  $C_1, C_2$  两类点,采用二元取值的  $y_i$  进行标记。因此,该点集可记为  $D_i = (x_i, y_i)$ 。需要找到一个超平面(又称分类函数)  $g(x) = \omega x + b$ ,使得  $C_1, C_2$  两类点分别分布在超平面  $g(x)$  两侧。

在实际分类中,通常会以 +1 和 -1 作为两类点的标签,即:

$$g(x) = \text{sign}(\omega x + b) = \begin{cases} -1 & x \in C_1 \\ +1 & x \in C_2 \end{cases}$$

定义一个样本点到某一超平面的间隔为:  $\sigma_i = y_i(\omega x_i + b)$ 。

经过归一化处理后,间隔可以写成:  $\sigma_i = \frac{|\text{sign}(\omega x_i + b)|}{\|\omega\|}$ 。因此,要使得误分次数最少,显然间隔要最大化,因此在样本可分的条件下,分类问题就转化为一个最优化问题:

$$\text{Minimize } \frac{1}{2} \|\omega\|^2 \quad (2)$$

$$\text{s. t. } y_i[(\omega \cdot x_i) + b] - 1 \geq 0, i = 1, 2, \dots, l$$

但是一旦样本点中存在一个错误点,样本就会因为这个错误点而变得线性不可分,出现所谓“少数点控制多数点”的问题。另外错误点可能出现在数据分布的任何区域内,而转化为二类分类问题会带来新的数据不平衡问题<sup>[9]</sup>。

以人类的常识判断,  $N$  个点因为其中一个不合理的点而变得不可分,这是不合理的,更加可能的情况是,这个不合理的点是“噪声”。但是上述最优化问题中并不具备这样的排错能力,仍然会因为自身“硬分类”的特性而判断这个样本点群不可分。

为了增强分类器的容错能力,引入松弛变量  $\xi_i$  和惩罚因子  $C$ ,原最优化问题变为:

$$\begin{aligned} \text{Minimize } & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s. t. } & y_i[(\omega \cdot x_i) + b] - 1 + \xi_i \geq 0, \\ & i = 1, 2, \dots, l \end{aligned} \quad (3)$$

其中,惩罚因子  $C$  需要预先设置。 $C$  的大小代表对离群点损失的重视程度,  $C$  越大,分类器对错误点的容忍程度越低。

引入拉格朗日算子  $\alpha_i \geq 0$  以及核函数<sup>[10]</sup>  $K(x, y)$   
 $= \varphi(x) \cdot \varphi(y)$ , 原优化问题转化为:

$$\begin{aligned} \text{Maximize } L(\omega, b, \alpha) = & \\ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) & \quad (4) \\ \text{s. t } & \begin{cases} \alpha_i \geq 0 \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases} \end{aligned}$$

分类符号函数表示为:  $f(x) = \text{sign}(\sum_{i=1}^l \alpha_i y_i K(x, x_i) + b)$ 。

### 1.3 支持向量回归(SVR)的实现

M. Pontil 等在文献[11]中指出,通过变量代换,分类是可以推广到回归情况中的,即分类是一种特殊的回归。设给定的样本集为:  $(x_i, y_i), x_i \in R^n, i = 1, 2, \dots, l$ , 对分类情况,  $y_i \in (-1, +1)$ ; 对于回归情况,  $y_i \in R^n$ 。

回归问题是解下面的二次规划:

$$\begin{aligned} \text{Minimize } & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t } & y_i - [(\omega \cdot x_i) + b] \leq \varepsilon + \xi_i, i = 1, 2, \dots, l \\ & [(\omega \cdot x_i) + b] - y_i \leq \varepsilon + \xi_i^*, i = 1, 2, \dots, l \\ & \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (5)$$

文献[12]经过变量代换后,给出如下结论:若优化问题(4)存在一个最优解  $(\omega, b)$ , 则存在一个  $a \in (0, 1)$ , 使得  $\forall \varepsilon \in (a, 1)$  时,回归问题(5)的最优解是  $(1 - \varepsilon)(\omega, b)$ 。

因此,回归问题和分类问题本质上是一样的,回归问题完全可以通过分类问题的方法求解。

## 2 基于 SVR 的去噪技术在 EMD 中的应用

### 2.1 SVR 在去噪方面的应用

任何被噪声污染的信号都可以由原始信号  $s(n)$  和噪声  $r(n)$  表示为  $x(n) = s(n) + r(n)$ , 事实上,无法在仅仅知道观测信号  $x(n)$  的情况下就准确给出  $s(n)$  和  $r(n)$ , 因为不知道  $s(n)$  或者  $r(n)$  的先验知识。

SVR 不需要输入信息的任何先验信息即可对信息进行回归处理。那么可以将去噪过程看作在已知观测信号  $x(n)$  的条件下,对  $s(n)$  进行回归估计的过程。即将  $x(n)$  作为 SVR 的输入,通过解最优化问题(5),得到最优解  $y_i$ , 则此时  $y_i$  为估计结果  $s(n)$ ,  $s(n)$  相对于原始信号  $s(n)$  的准确程度反映了 SVR 去噪能力的强弱。

SVR 是一种具有良好泛化能力和鲁棒性的机器学习策略,这种基于 SVM 分类的学习方法在多维变量

估计方面应用极为广泛<sup>[13-14]</sup>。并且,无论是对高斯噪声还是非高斯噪声,都有着优于维纳滤波器和多层感知滤波器的良好效果<sup>[15]</sup>。

如图 1 所示,输入信号  $s(t) = x(t) + n(t) = \text{sinc}(t) + 0.1 * \text{normrnd}(0, 1, 1, 1001)$ , 其中,  $t = -5:0.01:5$ ,  $\text{normrnd}(0, 1, 1, 1001)$  表示均值为 0, 方差为 1 的高斯白噪声序列,序列长度为 1 001。对两种信号进行加性混合,得到输入信号。

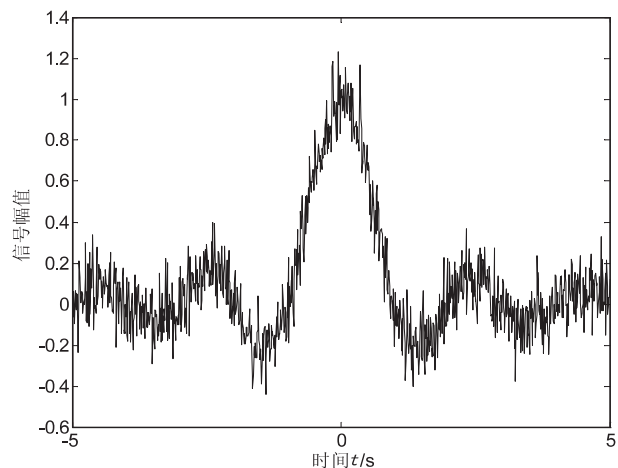


图 1 被噪声严重污染的信号

将输入信号作为最小二乘支持向量机的训练样本,采用基于 PSO 的 LS-SVM 进行学习<sup>[16]</sup>,最终回归结果如图 2 所示。

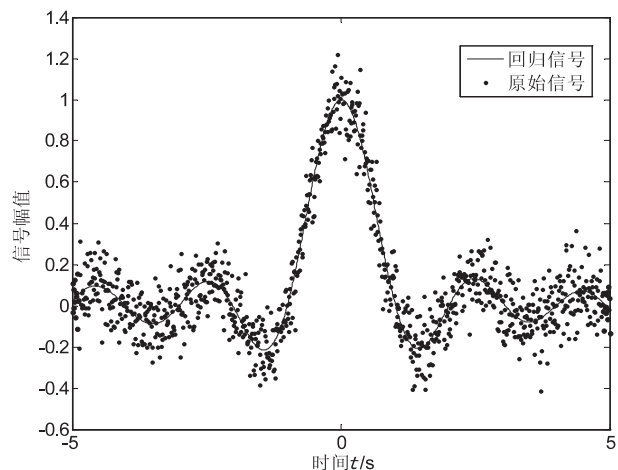


图 2 经过 LS-SVM 回归后的信号

定义信噪比  $\text{SNR} = 10 * \log(\frac{x^2(t)}{n^2(t)})$ , 那么,在 SVR

滤波前,输入信号的信噪比为 0.361 1 dB。经过 SVR 滤波后,可以看到回归结果和原始输入信号几乎一致,信噪比为 18.351 2 dB,提升了 17.990 1 dB,效果非常显著。

如图 3 所示,回归结果和原始输入样本的拟合度很高,回归效果比较理想。

### 2.2 二次分解—虚假 IMF 的消除

在传统的 EMD 分解中,采用三次样条函数对信号

进行包络线的获取,但是,这种基于极点的估计方法,对噪声极为敏感。一旦混入噪声,分解出的本征模函数就会产生严重的畸变,而且还有可能产生虚假的 IMF,对后续的分析产生干扰。

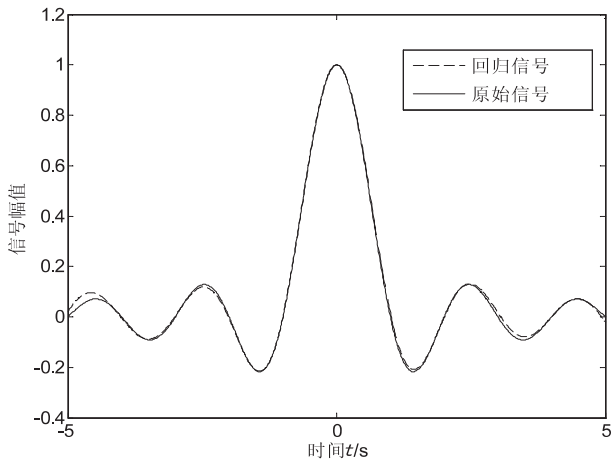


图 3  回归结果与原始信号的对比

为了使得在信噪比较小的情况下依旧能够分解出正确的本征模函数,需要对初次 EMD 分解结果进行逐层滤波去噪,然后对重组的信号进行二次 EMD 分解,过程如下:

- (1)对原信号  $x(t)$  进行 EMD 分解,得到  $imf_i(t)$  ;
- (2)对  $imf_i(t)$  进行 SVM 回归,得到回归后的  $imf_i'(t)$  和剩余分量  $Re(t)$  ;
- (3)将  $imf_i'(t)$  重组,得到  $\hat{x}(t)$  ;
- (4)对  $\hat{x}(t)$  进行二次 EMD 分解,得到  $imf_i''(t)$  ,分解结束。

对一次 EMD 分解结果进行逐层滤波去噪后,可以有效地滤除其中大部分的噪声成分,将回归结果进行重组后,原始信号中的大部分噪声成分已经被滤除,此时,对信噪比较高的重组信号进行二次分解,剩余噪声对分解过程将不会产生大的影响。因此,EMD 分解结果将不再受初始信噪比的影响,变得对噪声不敏感。

3  仿真结果分析

以信号  $inData = 4 * \sin 20 \pi t + 5 * \sin 200 \pi t + 10 * t$  为例,理论上应用 EMD 应该分解出 2 个正弦分量和一个直流分量。如图 4 所示,在没有噪声干扰的情况下,EMD 分解的效果非常好。

如图 5 所示,虽然 IMF1,IMF3,IMF4 被正确地分解出来,但是却错误地分解出了一个虚假的 IMF2,进一步加大了噪声功率。分解出的 IMF 已然发生了严重的畸变,变得无法识别,不具有实际的意义。

因此,采用二次分解的方式,首先对源信号进行一次 EMD 分解,并对分解结果进行 SVR 回归处理,将 SVR 回归结果整合重组后,进行 EMD 二次分解,对噪

声权值 0.1 和 0.2 两种情况进行观察和对比。

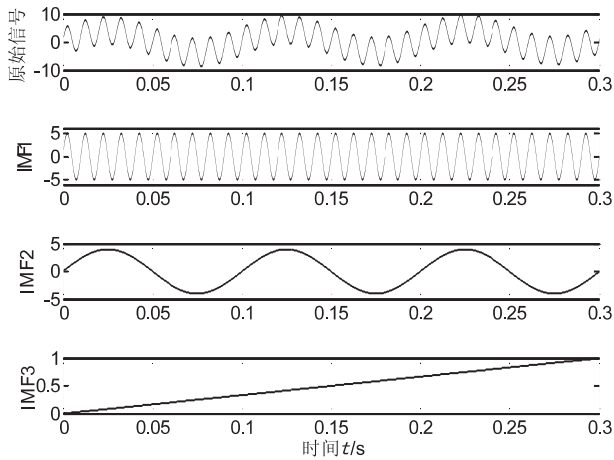
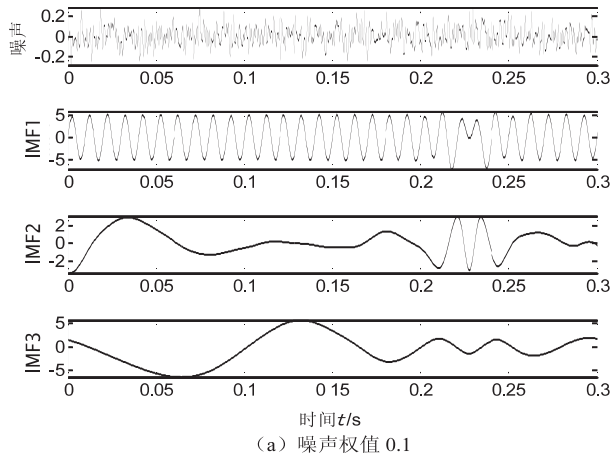
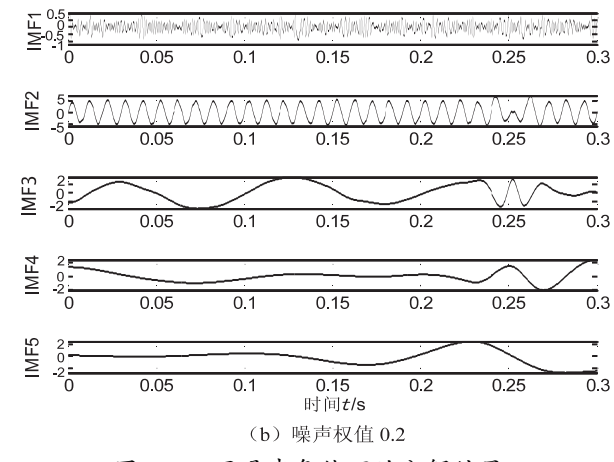


图 4  无噪声情况下 EMD 分解的效果



(a) 噪声权值 0.1



(b) 噪声权值 0.2

图 5  不同噪声条件下的分解结果

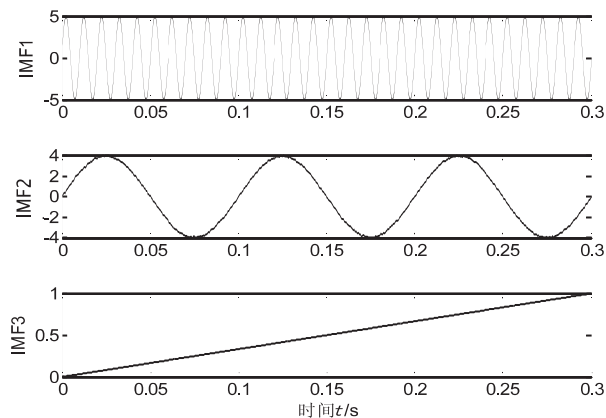
从图 6 中可以看到,在噪声权值为 0.1 时,原先出现的虚假“IMF2”已经消失,可以得到正确的分解结果;将噪声权值增加到 0.2 时,经过两次 EMD 分解,依旧可以得到应有的 3 个 IMF 分量,而不会出现一次 EMD 分解时,多个虚假错误 IMF 的情况。

4  结束语

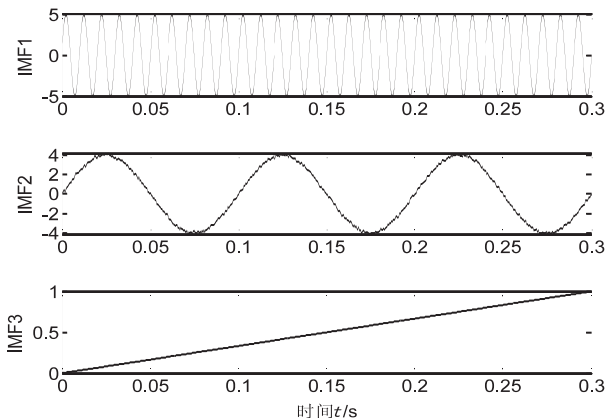
EMD 是一种优秀的自适应分解方法,但是在噪声干扰较为严重时,会出现误分解的情况。文中介绍了



一种基于 SVR 的去噪方法,并将之应用到 EMD 分解中,通过一次去噪,二次分解的方法,可以显著改善 EMD 分解的性能,使之在信噪比较低的情况下,依旧可以分解出正确的 IMF 分量,对数据采集和分析,都有重要意义。



(a) 噪声权值 0.1



(b) 噪声权值 0.2

图 6 不同噪声条件下二次 EMD 分解结果

## 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, 1998, 454(1971): 903-995.
- [2] Rilling G, Flandrin P, Gonçalves P. On empirical mode decomposition and its algorithms [C]//Proc of IEEE - EURASIP workshop on nonlinear signal and image processing. [s. l.]:

[s. n.], 2003.

- [3] 胥保春, 袁慎芳. IMF 筛选停止条件的分析及新的停止条件[J]. 振动、测试与诊断, 2011, 31(3): 348-353.
- [4] 许学明. 改进的希尔伯特—黄变换及其在信号时频分析中的应用[D]. 南京: 南京信息工程大学, 2012.
- [5] Huang C, Guo J, Yu X, et al. The study of interferogram de-noising method based on empirical mode decomposition[J]. International Journal of Computer Science Issues, 2013, 10(1): 750-756.
- [6] 张战成, 王士同, 邓赵红, 等. 支持向量机的一种快速分类算法[J]. 电子与信息学报, 2011, 33(9): 2181-2186.
- [7] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [8] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.
- [9] 田江, 顾宏. 孤立点一类支持向量机算法研究[J]. 电子与信息学报, 2010, 32(6): 1284-1288.
- [10] Smola A J, Schölkopf B. A tutorial on support vector regression [J]. Statistics and Computing, 2004, 14(3): 199-222.
- [11] Pontil M, Rifkin R, Evgeniou T. From regression to classification in support vector machines[R]. Cambridge: MIT Artificial Intelligence Lab, 1998.
- [12] 孙德山. 支持向量机分类与回归方法研究[D]. 长沙: 中南大学, 2004.
- [13] Hsiao C C, Su S F, Chuang C C. A rough-based robust support vector regression network for function approximation [C]//Proc of IEEE international conference on fuzzy systems. Taipei, Taiwan: IEEE, 2011: 2814-2818.
- [14] 于波, 邵高平, 孙红胜, 等. 直扩系统中基于 SVM 的干扰自动分类识别方法[J]. 信号处理, 2010, 26(10): 1539-1543.
- [15] Zhang Jian, Peng Qicong, Shao Huaizong, et al. Nonlinear noise filtering with support vector regression [C]//Proc of sixth international conference on intelligent systems design and applications. [s. l.]: IEEE, 2006: 172-176.
- [16] Wang L X, Wang L N, Li G F, et al. Application research of support vector machine based on particle swarm optimization in runoff forecasting[J]. Applied Mechanics and Materials, 2012, 226: 2303-2307.

(上接第 121 页)

- 质量的新方法[J]. 计算机研究与发展, 2004, 41(1): 98-103.
- [11] Rosenberg J. The Session Initiation Protocol (SIP) update method[S]. RFC 3311, 2002.
- [12] Ikeda H, Deng Hui, Niu Zhisheng, et al. Context-aware quality of service control in session based IP networks [C]//Proc of

2004 and the 5th international symposium on multi-dimensional mobile communications. [s. l.]: IEEE, 2004: 814-818.

- [13] 叶婷, 杜旭, 潘鹏, 等. 支持 QoS 的 SIP 代理服务器方案的设计与实现[J]. 计算机工程, 2006, 32(1): 139-141.
- [14] 3GPP TS 23. 207 (V6. 6. 0). End-to-end QoS concept and architecture[S]. 2005.

# 采用支持向量回归抑制噪声的经验模态分解方法

作者：[宋剑](#)，[邱晓晖](#)，[SONG Jian](#)，[QIU Xiao-hui](#)

作者单位：[南京邮电大学 通信与信息工程学院, 江苏 南京, 210003](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2014(11)

引用本文格式：[宋剑](#).[邱晓晖](#).[SONG Jian](#).[QIU Xiao-hui](#) 采用支持向量回归抑制噪声的经验模态分解方法[期刊论文]  
]-[计算机技术与发展](#) 2014(11)