

基于 EMD 的太阳黑子时间序列组合预测模型

王 曦, 毕贵红, 唐京瑞

(昆明理工大学电力工程学院, 昆明 650500)

摘 要: 针对太阳黑子的复杂性, 利用经验模态分解(EMD)方法, 将太阳活动在各时间尺度上的变化分量分解为平稳的固有模态函数(IMF)分量及余项。观察各分量的频谱, 根据低频 IMF 分量和高频 IMF 分量的特点, 分别采用自回归滑动平均模型和神经网络方法进行预测。通过各分量的预测值, 重构出原始信号的预测序列。仿真结果表明, 该模型具有较高的预测精度。

关键词: 太阳黑子数; 经验模态分解方法; 自回归滑动平均模型; 反向传播

Composite Forecasting Model of Sunspot Time Sequences Based on EMD

WANG Xi, BI Gui-hong, TANG Jing-rui

(College of Electric Power Engineering, Kunming University of Science and Technology, Kunming 650500, China)

【Abstract】 According to the complexity of sunspots, this paper uses Empirical Mode Decomposition(EMD) method, the solar activity contains all of the time scale changes separated into the inherent weight smooth Intrinsic Mode Function(IMF) and remainders. It observes each component of the spectrum, based on the characteristics of the low frequency IMF component selection Auto-regressive Moving Average(ARMA) model predicted the average directly, and the high frequency IMF using neural network forecast. Through the various components of the primary signal reconstruction predicts a prediction sequence, and increases the prediction accuracy. Simulation results show that the model has higher forecast accuracy.

【Key words】 sunspot number; Empirical Mode Decomposition(EMD) method; Auto-regressive Moving Average(ARMA) model; Back Propagation(BP)

DOI: 10.3969/j.issn.1000-3428.2011.24.059

1 概述

太阳黑子数是一个表征太阳总的活动水平的指数。在实际应用中, 只有它的月均值、月滑值和年均值才有较大的实用意义。太阳黑子数的演化具有高度的复杂性, 太阳黑子数据时间序列被认为是非线性、非平稳和混沌的, 被广泛应用于评价非线性时间序列模型的有效性。长期以来, 国内外学者对太阳黑子时间序列的性质和预测的理论和方法做了大量研究, 提出了各种各样的研究和预测方法。目前, 利用长期记录所积累的历史数据构造时间序列通过回归分析或人工神经网络方法对太阳黑子数进行预测是空间环境预报的一个重要内容。文献[1]用径向基函数神经网络方法预报太阳黑子数平滑月均值, 取得了较好的预测精度, 但是随着预报实效的延长预报误差被逐渐放大。文献[2]将自回归求和滑动平均(Auto-regressive Integrated Moving Average, ARIMA)与人工神经网络组合模型用于太阳黑子序列预测。利用 ARIMA 模型良好的线性拟合能力和人工神经网络强大的非线性关系映射能力, 把时间序列看成由线性自相关结构和非线性结构两部分组成, 采用 ARIMA 模型对太阳黑子序列的线性部分进行预测, 用人工神经网络模型对其非线性残差部分进行预测。结果表明: 组合模型的预测准确性高于各自单独使用时的准确性。文献[3]采用 Fourier 方法分析过太阳黑子相对数的周期特征。Fourier 方法的缺点是它有频率分辨率但没有时间分辨率。现在小波分析越来越多地被用于分析信号的周期特征, 因为它可以得到时间序列周期和频率特征的局部精细结构。文献[4]为了更深刻地认识太阳黑子相对数的周期规律特征,

采用小波分析方法分析过太阳黑子相对数周期的几个重要特性。但是小波分析不具有自适应性的特点, 一旦基本小波基被选定, 就必须用它来分析所有待分析数据, 小波变换实质上是窗口可调的 Fourier 变换, 其小波窗内的信号必须是平稳的, 因此, 没有从根本上摆脱 Fourier 变换的固有局限。综上所述, 本文提出一种基于 EMD 的太阳黑子时间序列组合预测模型。

2 相关知识

HHT(Hilbert Huang Transform)是全新的信号处理方法, 它由经验模态分解(Empirical Mode Decomposition, EMD)方法及 Hilbert 变换(HT)两部分组成。EMD 算法的目的是将性能不好的信号分解为一组性能较好的具有瞬时频率的固有模态函数(Intrinsic Mode Function, IMF)和具有单调性或一个极值点的余项。该方法分解过程基于数据信号局部特征, 不需要预先设定参数, 是自适应的。近年来, HHT 方法已经被成功应用于太阳黑子数据时间序列的研究。文献[5]利用 EMD 方法提取太阳活动周期成分, 将太阳活动包含的各个时间尺度上的变化分量很好地分离出来, 它们可以很好地得到整个太阳活动的变化规律和特征, 有助于太阳活动建模和预测。

基金项目: 云南省自然科学基金资助项目(2009CD028); 昆明理工大学科学研究基金资助项目(201001)

作者简介: 王 曦(1987—), 男, 硕士研究生, 主研方向: 信号处理, 模式识别; 毕贵红(通讯作者), 教授、博士; 唐京瑞, 硕士研究生

收稿日期: 2011-05-11 **E-mail:** wangxilich@126.com

文献[6]利用 EMD 方法将震动脉冲信号分解并提取第一模态、第二模态的能量和平均值进行训练, 结果也表明了 EMD 分解后提取的特征能有效地浓缩信号信息。

因此, 根据 EMD 方法将太阳活动包含的各个时间尺度上的变化分量很好地分离成平稳的 IMF 分量及余项, 通过各分量的频谱观察, 针对低频 IMF 分量规律性及周期性强, 高频分量相对较弱的特点, 对低频 IMF 分量选择自回归滑动平均(Auto-regressive Moving Average, ARMA)预测模型直接进行预测, 高频 IMF 采用神经网络预测方法, 然后用各分量的预测值重构出原始信号的预测序列, 从而提高预测精度。

3 基于 EMD 的组合预测模型

3.1 太阳黑子时间序列的 EMD 分解

EMD 分解源于 HHT 方法, HHT 是一种新的数据或者信号处理方法, 可以处理非线性非平稳信号。文献[7]认为任何信号都是由基本信号——固有模态信号或固有模态函数组成, IMF 相互叠加就形成复合信号。基于这样的思想, 提出的 HHT 分 2 个步骤, 即: (1)EMD。把复杂的信号用经验模式分解方法分解成若干阶 IMF。(2)HT。对 IMF 进行 Hilbert 变换, 得到每一个 IMF 随时间变化的瞬时频率和振幅, 求得时间-频率-振幅的三维谱分布, 称为 Hilbert 谱。在 Hilbert 谱的基础上, 通过积分可以求边际谱。

具体方法如下:

(1)确定信号所有的局部极值点, 然后样条函数拟合所有的局部极大值点和极小值点, 形成上、下包络线 $u(t)$ 和 $v(t)$ 。

(2)计算上、下包络线的平均值 $m=(u(t)+v(t))/2$ 。

(3)进一步计算 $h=x(t)-m$, $x(t)$ 为初始信号。

(4)如果 h 符合固有模态函数的特点, 则 h 是第 1 个被分解出来的 IMF 分量。否则, 把 h 作为原始数据, 重复上述步骤, 直到得到一个固有模态函数。此时, 将 h 记作 c 。

(5)将 c 从 $x(t)$ 分离出来, $r=x(t)-c$ 。再将 c 作为原始数据重复上面步骤, 得到第 2 个满足 IMF 条件的分量, 循环 n 次, 直到 r 只有一个极值点的时候, 循环停止, 此时的 r 作为残余分量。

这里使用 Matlab 环境, 编写相应 EMD 分解程序, 在控制分解停止时采用筛选停止准则, 即设:

$$SD = \frac{\sum_{t=0}^T |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^T h_{k-1}^2(t)}$$

其中, SD 为判断标准, 当 SD 在 0.2~0.3 时, 筛选停止。

(6)端点处理。EMD 分解采用极值法, 但是数据最初的点和最后一个点, 很难说是否是极值点, 而方法中直接将这 2 个点作为极值点处理, 所以, 必然会对分解效果产生影响, 甚至会污染到整个数据。采用镜像延拓法^[8], 得到 IMF 分量再截去后面延拓的数据, 即为原始数据的 EMD 分解, 数据长度与原始数据相同。

3.2 ARMA 模型

ARMA 模型是研究时间序列的重要方法, 由自回归(AR)模型与滑动平均(MA)模型为基础“混合”构成。ARMA 是序列建模方法, 其建模的基本思想是对非平稳的时间序列通过差分使其成为平稳序列, 再用以 p 、 q 为参数的 ARMA 模型对该平稳序列建模, 之后经过反变换得到原序列。以 p 、 q 为参数的 ARMA 模型预测方程可以表示为:

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

其中, y_t 为样本值; $\varphi_i(i=1, 2, \dots, p)$ 和 $\theta_j(j=1, 2, \dots, p)$ 为模型参数; ε_t 为服从独立正态分布 $N(0, \sigma^2)$ 的白噪声序列。 p 、 q 为模型的阶数, 通过对不同的 p 、 q 组合测试可以优化模型预测结果, 找到最合适的模型参数。

3.3 BP 神经网络

BP 神经网络是一种多层前馈神经网络, 该网络的主要特点是信号前向传递, 误差反向传播。在前向传递中, 输入信号从输入层经隐含层逐层处理, 直至输出层。每一层的神经元状态只影响下一层神经元状态。如果输出层得不到期望的输出, 则转入反向传播, 根据预测误差调整网络权值和阈值, 从而使 BP 神经网络预测输出不断逼近期望输出。BP 神经网络的拓扑结构如图 1 所示。

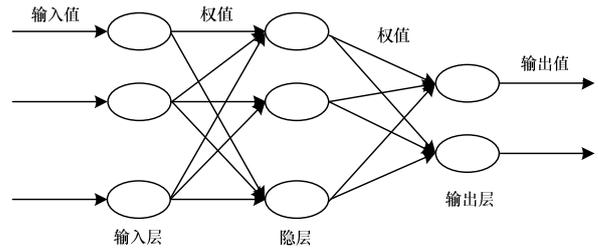


图 1 BP 神经网络的拓扑结构

当输入节点数为 n 、输出节点数为 m 时, BP 神经网络表达从 n 个自变量到 m 个因变量的函数映射关系。

BP 神经网络预测前首先要训练网络, 通过训练使网络具有联想记忆和预测能力。BP 神经网络的训练过程包括以下步骤: (1)网络初始化。根据系统输入输出序列 (X, Y) 确定网络输入层节点数 n 、隐含层节点数 1, 输出层节点数 m , 初始化输入层、隐含层和输出层神经元之间的连接权值初始化隐含层阈值、输出层阈值, 给定学习速率和神经元激励函数。(2)隐含层输出计算。根据输入向量 X 、输入层和隐含层间连接权值以及隐含层阈值, 计算隐含层输出。(3)输出层计算, 并计算误差。(4)根据网络预测误差更新网络连接权值和节点阈值。这里利用 BP 神经网络的非线性函数拟合算法对数据进行处理。由于复杂的非线性系统的状态方程比较复杂, 因此难以用数学方法准确建模。在这种情况下可以建立 BP 神经网络表达这些非线性系统。

3.4 组合预测模型

EMD 方法从本质上讲是将时间序列分解成有限个具有特征振荡尺度的分量, 尽管仍然保持不同程度的非平稳性, 但是它们之间的相互影响却被隔离了, 有利于对分解出来的信号根据其特征采取不同的预测方法。然后用各分量的预测值重构出原始信号的预测序列, 从而提高预测精度, 见图 2。

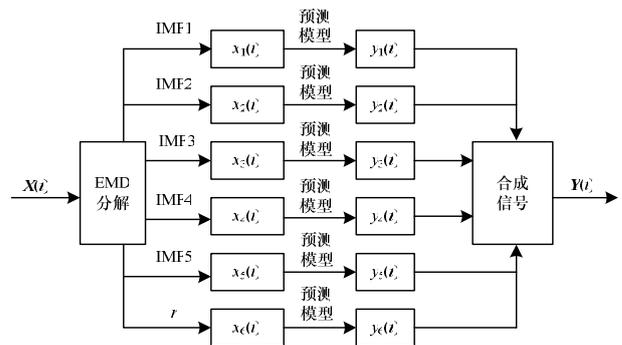


图 2 基于 EMD 的组合模型结构

基于 EMD 的组合模型描述如下: (1)应用 HHT 算法, 通

过 EMD 对太阳黑子数据序列进行分解, 得到若干个频率由高到低的 IMF 分量和余项。(2)提取规律。对各个 IMF 分量通过 Hilbert 求频谱, 通过观察其瞬时频率值、频带宽度、频率波动情况掌握规律。(3)低频分量处理。将频率值较低、频率变化范围较窄的 IMF 分量和余项定位低频分量, 对它的预测较简单, 选择一种合适的预测方法即可。(4)高频分量处理。将频率值较高、频率变化波动性较大、频带宽的分量定为高频分量。它除了包含高频的周期信息外, 绝大部分为太阳黑子数据序列中的随机变化量和非线性分量, 对于这些分量采用一般的预测方法通常效果不佳, 这里将采用神经网络组合模型的预测方法。(5)预测结果综合。将各个 IMF 预测值相加得到最终的预测结果。

4 数据仿真与分析

首先对所有的数据进行 EMD 分解, 得到的分量和原始数据进行比较, 频率相对平滑利于预测。然后, 分别使用 ARMA 模型和 BP 神经网络模型对 IMF 分量进行计算并预测, 这里使用计算的数据为 1959 年-2008 年月平滑太阳黑子数, 对比数据为 2009 年月平滑太阳黑子数。导入 1959 年-2008 年月平滑太阳黑子数, 经过 EMD 分解得到 5 个 IMF 分量和 1 个余项 r , 如图 3、图 4 所示。

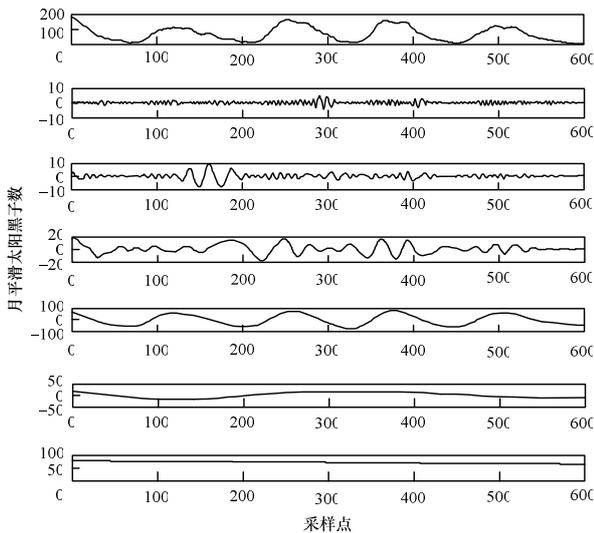


图 3 1959 年-2008 年月平滑太阳黑子数的 EMD 结果

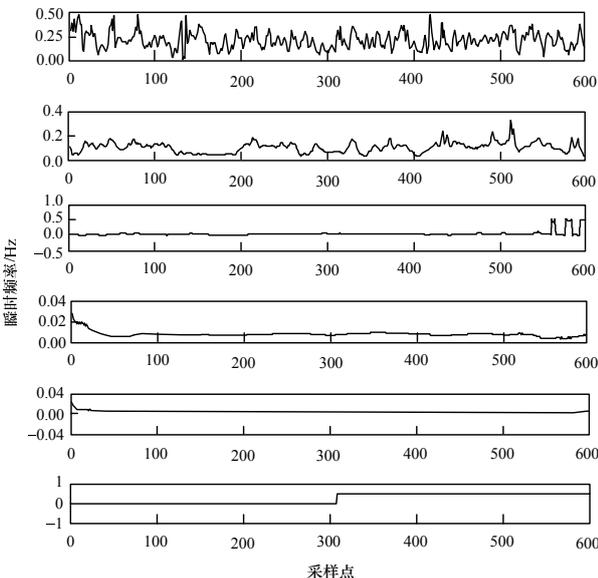


图 4 各模态分量的瞬时频率

在图 3 中, 从上到下分别为原始信号、模态 IMF1~IMF5 和剩余分量; 在图 4 中从上到下分别为模态 IMF1~IMF5 和剩余分量。从分解结果可以看出, 分解后的 IMF 分量比较平稳, 频率带窄, 瞬时频率较低, 基本平稳。对分解后的 IMF 预测可以达到一个比较高的预测结果。IMF1、IMF2 相对其他分量还是比较复杂, 所以对 IMF1 和 IMF2 使用 BP 神经网络模型做预测分析, 对剩下的各分量使用 ARMA 模型做预测分析, 以此比较各个模型对不同分量的预测精度。

(1)直接使用 ARMA 模型。选择使用 1959 年-2008 年月平滑太阳黑子数进行建模, 经过对多个 AR 阶数和 MA 阶数 (10 以内) 的实验比较, ARMA[10, 7] 这组阶数是误差 (绝对误差) 最小的, 均方误差为 0.518 2。预测曲线与实测曲线的比较见图 5。

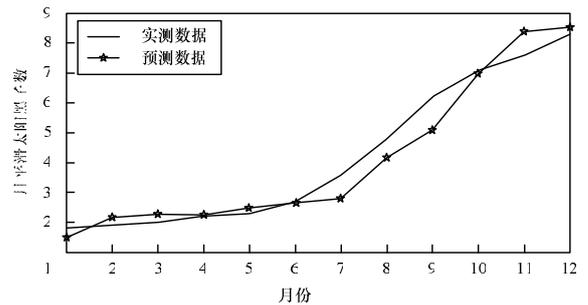


图 5 直接使用 ARMA 模型的预测结果

(2)直接使用 BP 神经网络模型。设置计算步数为 1 000、精确目标为 0.01、学习速率为 0.1, 使用 1959 年-2008 年的数据进行网络训练, 预测后的均方误差为 0.512 8。图 6 为实际数据和预测数据的比较。

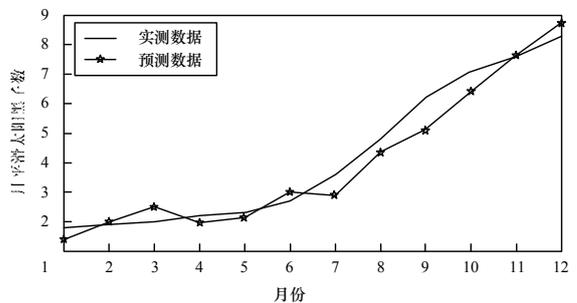


图 6 直接使用 BP 神经网络模型的预测结果

(3)EMD 结合 ARMA 模型。使用上面 1959 年-2008 年月平滑太阳黑子数的 EMD 分解结果, 得到 IMF 分量然后进行 ARMA 模型建模并预测。设定参数 ARMA[10, 7] 进行预测。

首先是使用 1959 年-2008 年的数据, 从 IMF1~IMF5 以及剩余分量的均方误差分别为 0.150 8、0.095 4、0.022 0、0.005 2、0、0。最后拟合数据的均方误差为 0.152 9。预测结果见图 7。

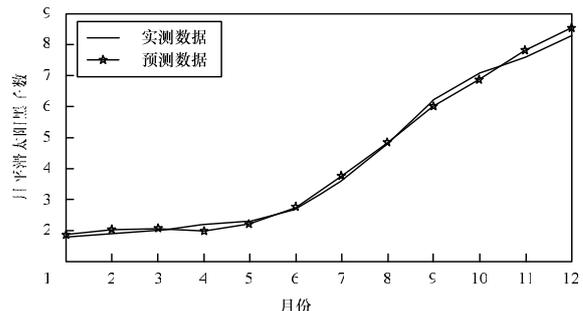


图 7 EMD 结合 ARMA 模型的预测结果

(4)EMD 结合 BP 神经网络模型。使用 1959 年-2008 年月平滑太阳黑子数的 EMD 分解结果, 然后对各 IMF 分量进行 BP 神经网络建模并预测, 采用计算步数为 1 000、误差目标为 0.01、计算速率仍然为 0.1, 从 IMF1~IMF5 以及剩余分量得到的均方误差分别为 0.034 3、0.047 7、0.018 3、0.074 4、0.068 0、0.050 3, 最后拟合数据的均方误差为 0.139 3。EMD 分解后各分量用 BP 神经网络模型的预测结果如图 8 所示。

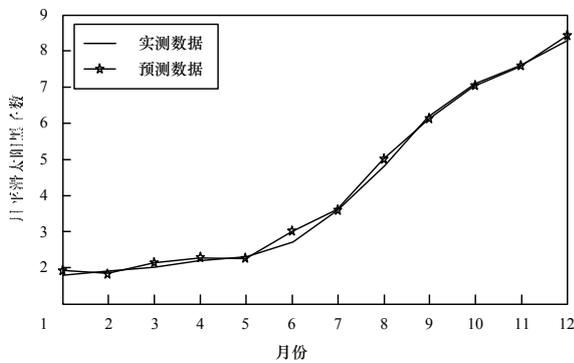


图 8 EMD 结合 BP 神经网络模型的预测结果

(5)从瞬时频率图可以看出 IMF1、IMF2 分量信号的频率变化较快、信号较复杂, 可以考虑使用 BP 神经网络模型进行预测, 其余分量频率相对简单平滑, 使用 ARMA 模型进行预测, 然后将预测数据结合得到最终的预测数据。预测后得到每个 IMF 分量的均方误差为 0.035 3、0.047 4、0.022 0、0.005 2、0、0, 最终数据的均方误差为 0.073 3。预测结果如图 9 所示。

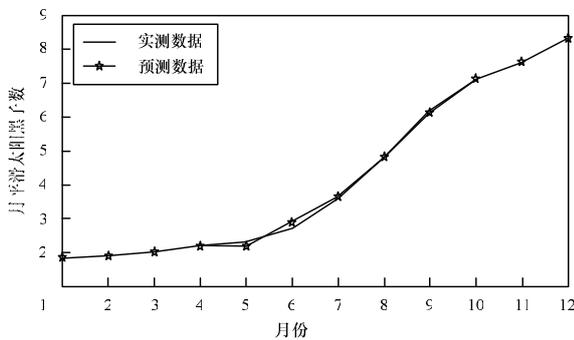


图 9 EMD 结合 BP 神经网络与 ARMA 模型的预测结果

文中使用这 5 种方法的均方误差如表 1 所示。

表 1 5 种方法的均方误差比较

模型方法	均方误差
直接使用 ARMA 模型的方法	0.518 2
直接使用 BP 神经网络模型的方法	0.512 8
EMD 结合 ARMA 模型的方法	0.152 9
EMD 结合 BP 神经网络模型的方法	0.139 3
EMD 结合 BP 神经网络与 ARMA 模型的方法	0.073 3

从对比结果可以看出, EMD 结合 ARMA 模型或者 BP 神经网络模型比单独使用这 2 种模型准确, 预测误差更小。

5 结束语

太阳黑子序列是典型的复杂时间序列, 对太阳黑子的预测有很重要的实际意义。本文实验结果表明, 使用 EMD 方法能够明显地提取出较平稳的信号, 这些信号包含了不同的信息内容, ARMA 模型对其中的较平滑数据的预测较好, BP 神经网络一般用于非线性的复杂数据, 所以, 在 EMD 分解的各分量中, 选择合适的模型处理不同的分量, 取得很好的效果, 多种预测模型的结合会进一步提高预测准确性。

参考文献

- [1] 赵海娟, 王家龙, 宗位国, 等. 用径向基函数神经网络方法预报太阳黑子数平滑月均值[J]. 地球物理学报, 2008, 51(1): 31-36.
- [2] Zhang G P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model[J]. Neurocomputing, 2003, 50(1): 159-171.
- [3] Romanov Y S, Zgonyaiko N S. The Periodicity of Solar Activity Cycles[J]. Solar Physics, 1994, 152: 31-34.
- [4] 乐贵明, 郭汉伟. 太阳黑子群周期的小波分析[J]. 自然科学进展, 2004, 14(4): 597-600.
- [5] 李 强, 吴 健, 许正文, 等. 利用 EMD 方法提取太阳活动周期成分[J]. 空间科学学报, 2007, 27(1): 1-6.
- [6] 杨永超, 汪同庆. 基于 EMD 和 SVDD 的铸钢支座故障诊断[J]. 计算机工程, 2009, 35(22): 236-238.
- [7] Huang N E, Shen Zheng. 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, 1998, 454(1971): 903-995.
- [8] 王 凯, 安 钢, 樊新海, 等. 基于支持向量机和镜像延拓的 Hilbert-Huang 变换及其在滚动轴承故障诊断中的应用[J]. 测量与仪器, 2009, (12): 40-43.

编辑 陆燕菲

(上接第 175 页)

参考文献

- [1] Kennedy J, Eberhart R C. Particle Swarm Optimization[C]//Proc. of IEEE International Conference on Neural Networks. Piscataway, USA: IEEE Service Center, 1995: 1942-1948.
- [2] 倪庆剑, 张志政, 王蓁蓁, 等. 一种基于可变多簇结构的动态概率粒子群优化算法[J]. 软件学报, 2009, 20(2): 339-349.
- [3] 李 睿, 李伟娟, 李 明. 基于加权量子粒子群的分类器设计[J]. 计算机工程, 2010, 36(7): 203-204, 207.
- [4] Robinson J, Rahmat-Samii Y. Particle Swarm Optimization in Electromagnetics[J]. IEEE Trans. on Antennas Propagation, 2004, 52(2): 397-407.

- [5] Xu Shenheng, Rahmat-Samii Y. Boundary Conditions in Particle Swarm Optimization Revisited[J]. IEEE Trans. on Antennas and Propagation, 2007, 55(3): 760-765.
- [6] Shi Yuhui, Eberhart R C. A Modified Swarm Optimizer[C]//Proc. of IEEE International Conference on Evolutionary Computation. Anchorage, USA: IEEE Press, 1998: 69-73.
- [7] Eberhart R C, Shi Yuhui. Particle Swarm Optimization: Developments, Applications, and Resources[C]//Proc. of IEEE International Conference on Evolutionary Computation. [S. l.]: IEEE Press, 2001: 81-86.

编辑 张正兴