

基于 EMD 和 RVM—AR 的航空发动机 磨损故障预测模型

范 庚, 马登武

(海军航空工程学院 兵器科学与技术系, 山东 烟台 264001)

摘要: 针对滑油光谱数据的非线性非平稳特性及现有预测方法的不足, 提出了一种基于经验模态分解 (EMD) 和相关向量机—自回归模型 (RVM) 的预测方法; 首先, 利用 EMD 算法将序列分解为若干固有模态函数 (IMF) 和残余函数 (RF), 分离原始数据中的趋势项和随机项; 然后, 建立各 IMF 分量的 AR 模型和 RF 分量的 RVM 模型, 以留一交叉验证误差作为适应度函数, 采用量子粒子群优化算法进化搜索 RVM 的最优核参数和嵌入维数; 最后, 将各 AR 模型和 RVM 模型的预测值叠加作为最终预测值; 实验结果表明: 相比 AR 和 SVM 方法, 该方法具有较高的预测精度, Fe 元素的平均相对预测误差分别提高了 15% 和 6%, Si 元素的平均相对预测误差分别提高了 49% 和 23%。

关键词: 故障预测; 滑油光谱分析; 经验模态分解; 相关向量机; 自回归模型; 量子粒子群优化; 航空发动机

Aero—Engine Wear Faults Prediction Based on EMD and RVM—AR

Fan Geng, Ma Dengwu

(Department of Ordnance Science and Technology, Naval Aeronautical and Astronautical University, Yantai 264001, China)

Abstract: To deal with the non—linearity and non—stationarity of oil spectrum data and overcome the deficiencies of current prediction methods, a prediction method based on empirical mode decomposition (EMD) and RVM—AR is proposed. Firstly, the original data were decomposed into several intrinsic mode functions (IMF) and a residual function (RF) by the EMD algorithm. Secondly, the AR models of IMFs and the RVM model of RF were established, and the kernel parameter and the embedding dimension were optimized by quantum—behaved particle swarm optimization algorithm, which used cross—validation error as the fitness function. Finally, the prediction results of AR models and the RVM model were added as the final prediction values. Experimental results show that the proposed method has better performance in prediction accuracy than current methods and can predict the trend of the lubrication metal content accurately.

Key words: fault prognostic; oil spectral analysis; empirical mode decomposition; relevance vector machine; quantum—behaved particle swarm optimization; aero—engine

0 引言

故障预测与健康管理 (PHM) 技术是实现航空发动机高可靠性、高安全性、高可用率和低耗费的关键使能技术, 对于军用维修保障体制变革具有重大的技术引导作用^[1-2]。航空发动机滑油中的金属元素含量值能较准确地反映出发动机的状态信息, 利用油液光谱分析技术对滑油中各种金属含量值及其趋势进行分析是实现发动机磨损故障预测的有效手段^[3-4]。文献^[4]运用自回归 (AR) 模型对采集的航空发动机滑油光谱数据进行时序建模和预测分析。AR 模型的特点是在平稳时间序列建模时具有较好的效果, 但在处理非平稳时间序列时会产生较大的误差。文献^[5-7]建立了滑油光谱数据的支持向量机 (SVM) 预测模型。SVM 利用核函数将输入空间通过非线性变换映射到高维空间, 将非线性问题转化为线性问题并能得到全局最优解, 但 SVM 也存在一些缺陷, 如参数设置复杂核函数受 Mercer 条件限制等。RVM 是 Tipping^[8-9]提出的一种与

SVM 相似的稀疏概率模型, 它利用 Bayesian 方法进行推理, 模型结构简单。与 SVM 相比, RVM 不仅得到更高的准确率, 而且在模型稀疏性、参数设置、核函数选择等方面还具备了 SVM 无法达到的优势^[10]。

针对油液光谱数据的非线性非平稳特性以及现有单一预测方法的不足, 提出了一种基于经验模态分解 (EMD) 和 RVM—AR 的预测方法, 并将通过预测实例说明该方法的有效性和优越性。

1 经验模态分解

1.1 传统 EMD 算法

EMD^[11]是 1998 年由美国工程院院士 Huang 等人提出的一种自适应信号处理方法。该方法吸取了小波变换多分辨的优势, 同时克服了小波变换中小波基选取与分解尺度确定的难题, 更适于非线性、非平稳信号的分析。令 x 为原始时间序列, 传统 EMD 算法的步骤为:

- 1) 确定 x 所有局部极值点, 并采用三次样条函数拟合形成上包络线和下包络线。记上下包络线的均值为 m_1 , 计算 $h_1 = x - m_1$;
- 2) 令 $x = h_1$, 重复步骤 1) 直到 h_1 满足 IMF 的两个条

收稿日期:2012-12-16; 修回日期:2013-01-30。

作者简介:范 庚(1985—),男,博士研究生,主要从事航空装备故障诊断与预测方向的研究。

件, 此时记 $c_1 = h_1$, 则 c_1 为 x 的第 1 个 IMF 分量;

3) 将 $r_1 = x - c_1$ 作为原始数据, 重复步骤 1) 和 2), 得到 x 的第 2 个 IMF 分量 c_2 , 依此继续分解直至 r_n 足够小或为单调函数, 得到 x 的 n 个 IMF 分量 $c_1 \sim c_n$;

通过上述分解, 原始序列 x 可表示为

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

式中: $c_1 \sim c_n$ 分别代表从信号从高到低不同频段的成分, r_n 代表信号的平均趋势。

1.1 EMD 端点效应的镜像延拓抑制方法

EMD 算法的关键步骤之一是利用三次样条函数拟合形成上、下包络线, 它将直接影响到分解得到的所有 IMF 分量的精度。然而由于无法确定信号两边端点位置的极值点, 导致得到的包络线在端点附近出现严重的失真, 出现端点效应, 最终导致得到一组不可靠的 IMF 分量和残余分量。为了抑制传统 EMD 算法的端点效应, 国内外学者提出了很多处理方法, 工程上常用的有线性外延法、多项式拟合法、镜像延拓法、AR 预测法和 ANN 预测法等。文献 [12] 利用准周期信号和随机信号对它们的端点效应抑制效果进行了测试, 指出镜像延拓法是目前相对最优和应当首选的 EMD 端点效应抑制方法。因此, 本文采用基于镜像延拓的 EMD 算法对油液光谱数据进行分解。

镜像延拓法^[13]的基本思想是: 根据信号的分布特性, 将镜子放在具有对称性的极值位置, 通过镜像法把镜内信号映射成一个周期性的环形信号。由于不存在端点, 从根本上避免了 EMD 的端点效应。

2 相关向量机

2.1 稀疏 Bayesian 回归原理

假设给定输入-输出样本对 $\{x_n, t_n\}_{n=1}^N$ 作为训练数据集, 式中: $x_n \in \{R^d\}$, $t_n \in \{R\}$, 则与 SVM 类似, RVM 的模型输出可以表示为:

$$y(x, w) = \sum_{i=1}^N w_i K(x, x_i) + w_0 \quad (2)$$

式中: $K(x, x_i)$ 为核函数; w_i 为模型的权值; N 为样本数。

假定目标是独立的, 且来自带噪声的模型:

$$t_n = y(x_n, w) + \varepsilon_n \quad (3)$$

式中: ε_n 为噪声, 服从均值为 0, 方差为 σ^2 的 Gaussian 分布, 即 $p(t_i | x_i) = N(t_i | y(x_i, w), \sigma^2)$ 。方便起见, 引入超参数 $\beta = \frac{1}{\sigma^2}$, 则相应的训练样本集的似然函数为:

$$p(t | w, \beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \exp\left\{-\frac{\beta}{2} \|t - \Phi w\|^2\right\} \quad (4)$$

式中: $t = [t_1, t_2, \dots, t_N]^T$; $\Phi \in R^{N \times (N+1)}$ 是设计矩阵, 定义为 $\Phi = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_N)]^T$, 基函数向量 $\varphi(x_n) = [1 \ K(x_n, x_1) \ \dots \ K(x_n, x_N)]^T$ 。

采用稀疏 Bayesian 方法对权值 w 赋予先验的条件概率分布:

$$p(w | \alpha) = \prod_{i=1}^N N(w_i | 0, \alpha_i^{-1}) \quad (5)$$

根据 Bayesian 公式, 利用样本似然函数式 (4) 和 w 先验

分布式 (5) 可以得到 w 的后验分布计算公式:

$$p(w | t, \alpha, \beta) = \frac{p(t | w, \beta) p(w | \alpha)}{p(t | \alpha, \beta)} \quad (6)$$

由于 $p(t | w, \beta)$ 和 $p(w | \alpha)$ 均为 Gaussian 分布, 而 $p(t | \alpha, \beta)$ 中不含 w , 可看作归一化系数。则 w 的后验分布可进一步表示为:

$$p(w | \alpha, \beta) = N(w | \mu, \Sigma) \quad (7)$$

后验协方差和均值分别为:

$$\Sigma = (\beta \Phi^T \Phi + A)^{-1} \quad (8)$$

$$\mu = \beta \Sigma \Phi^T t \quad (9)$$

式中: $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ 。超参数 α 和 β 直接影响 w 的后验分布, 需对其优化以得到 w 的最大后验分布。通过最大化边缘似然函数 $p(t | \alpha, \beta)$ 实现超参数的优化。将 $p(t | \alpha, \beta)$ 取负对数得到目标函数, 将目标函数对超参数 α 和 β 求偏导并令导数式为 0, 得到以下迭代公式:

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i^2} \quad (10)$$

$$(\sigma^2)^{\text{new}} = \frac{\|t - \Phi \mu\|}{N - \sum_i \gamma_i} \quad (11)$$

式中: μ_i 为第 i 个后验平均权; $\gamma_i \equiv 1 - \alpha_i \Sigma_{ii}$, Σ_{ii} 为后验权协方差矩阵的第 i 个对角元素; N 为样本数据的个数。公式 (8) ~ (11) 依次进行迭代计算直到所有参数都收敛或达到最大训练次数时为止。

若给定输入值 x_* , 则相应输出的概率分布为

$$p(t_* | t, \alpha_{MP}, \sigma_{MP}^2) = \int p(t_* | w, \sigma_{MP}^2) p(w | t, \alpha_{MP}, \sigma_{MP}^2) w \quad (12)$$

式 (12) 服从正态分布, 即

$$p(t_* | t, \alpha_{MP}, \sigma_{MP}^2) = N(t_* | y_*, \sigma_*^2) \quad (13)$$

预测均值和方差分别为:

$$y_* = \mu^T \varphi(x_*) \quad (14)$$

$$\sigma_*^2 = \sigma_{MP}^2 + \varphi(x_*)^T \Sigma \varphi(x_*) \quad (15)$$

则 RVM 新观测值的预测输出为 $y_* = (x_*; \mu)$ 。

2.2 时间序列 RVM 预测模型

对于时间序列 $\{x_1, x_2, \dots, x_N\}$, 利用 t 时刻以前的值预测 t 时刻的值, 需要建立 $\{x_{t-1}, x_{t-2}, \dots, x_{t-l}\}$ 与 x_t 之间一一映射的关系 $f: R^l \rightarrow R$, 其中 l 为嵌入维数。因此, RVM 回归模型的训练样本集为:

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_{N-l} \\ x_2 & x_3 & \dots & x_{N-l+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_l & x_{l+1} & \dots & x_{N-1} \end{bmatrix} = [x_1, x_2, \dots, x_{N-l}],$$

$$Y = [x_{l+1}, x_{l+2}, \dots, x_N] = [t_1, t_2, \dots, t_{N-l}].$$

根据式 (2) 建立 RVM 回归预测模型如下:

$$y(x) = \sum_{i=1}^{N-l} w_i K(x, x_i) + w_0 \quad (16)$$

$K(\cdot, \cdot)$ 选为 Gaussian 核函数, 则:

$$K(x_i, x_j) = \exp(-||x_i - x_j||^2 / 2\gamma^2) \quad (17)$$

2.3 模型参数的自适应优化

RVM 模型参数的选择将直接影响 RVM 模型的预测性能。本文建立的时间序列 RVR 回归模型有两个参数: 嵌入维数 l

和 Gaussian 核参数 γ 。由于凭借主观经验确定模型参数的人为方法缺乏科学依据且不具备推广性，因此，为了实现模型参数的自适应优化选择，本文引入量子粒子群优化（QPSO）算法对建立模型的参数进行优化。

传统的 PSO 算法将群体中的每个个体（可行解）看成 D 维搜索空间中的一个没有体积的粒子，在搜索空间中以一定的速度飞行，其速度根据目前所找到的最优位置和整个粒子群所找到的最优位置进行动态调整。QPSO 算法中粒子只有位置向量，没有速度向量，由于粒子满足聚集态的性质与 PSO 完全不同，使粒子能在整个可行解空间中搜索全局最优解。算法具体描述如下^[14]：

$$Z(t+1) = p(t) \pm \eta \cdot |m_{\text{best}} - Z(t)| \cdot \ln(1/u) \quad (18)$$

$$p(t) = \varphi \cdot p_i(t) + (1 - \varphi) \cdot p_g(t) \quad (19)$$

$$m_{\text{best}} = \left(\frac{1}{M} \sum_{i=1}^M p_{i,1}(t), \dots, \frac{1}{M} \sum_{i=1}^M p_{i,D}(t) \right) \quad (20)$$

式中： M 为种群中粒子的数目； D 为粒子维数； u 和 φ 是在 $[0, 1]$ 上均匀分布的随机数； m_{best} 是种群中所有粒子的平均最好位置点； p_i 和 p_g 分别表示粒子 i 所经历的最好位置和种群中所有粒子所经历的最好位置； η 称为收缩扩张系数，是 QPSO 惟一的参数。本文采用通过距离自适应调节的方法来控制 η 的取值：

$$\eta = 0.5 + 0.5 \times \frac{i_{\text{max}} - i_c}{i_{\text{max}}} \quad (21)$$

式中： i_{max} 为算法的最大迭代次数； i_c 为当前的迭代次数。

好的推广能力估计方法是实现 RVM 模型参数选择的基础。鉴于交叉检验法原理简单，实现方便，而且寻优结果比较接近实际最优值，本文以留一交叉验证^[15]得到的平均误差估计 RVM 的推广能力。留一交叉验证法的基本思想是：逐一从训练集中去掉一个样本，利用剩余的样本训练 RVM 回归模型，并利用训练好的模型对去掉的样本进行预测；最后取所有相对预测误差的平均值作为推广能力的估计。利用 QPSO 算法实现 RVM 模型参数 (l, γ) 自适应进化搜索的步骤如下：

1) 置当前迭代次数 $i_c = 1$ ，设定种群中粒子数目 M ，算法的最大迭代次数 i_{max} ，嵌入维数 l 和 Gaussian 核参数 γ 的取值区间，并在取值区间中随机初始化粒子位置 $Z_n(i_c) = [l_n(i_c), \gamma_n(i_c)]^T$ ， $n = 1, 2, \dots, M$ 。

2) 根据 2.2 节方法建立训练样本集 X_n 。

3) 根据本节中留一交叉验证法计算得到交叉验证误差 e_n 。

4) 将每个粒子的当前位置作为粒子的个体最优值 p_n ，通过比较 e_n 确定种群的全局最优值 p_g 。

5) 根据公式 (18) ~ (21) 更新粒子的位置。

6) 将每个粒子的适应度值 e_n 与其经历的最好位置 p_n 进行比较，若优则将其作为当前最好位置，更新 p_n ，否则保留原值；

7) 将更新后每个粒子的 p_n 与全局极值 p_g 比较，若优则将 p_g 设置为最好粒子的当前位置，否则保留原值；

8) 判断是否满足终止条件，如果不满足，则设定 $i_c = i_c + 1$ 转至步骤 3)。

3 基于 EMD 和 RVM-AR 的滑油金属含量预测模型

基于 EMD 和 RVM-AR 的航空发动机滑油金属含量预测模型的基本思路是：首先，利用 EMD 算法将序列分解为若干固有模态函数（IMF）和残余函数（RF），分离原始数据中的趋势项和随机项；然后，建立各 IMF 分量的 AR 模型和 RF 分量的 RVM 模型，AR 模型参数采用最终预测误差（FPE）准则确定，以交叉验证误差作为优化准则，采用量子粒子群优化（QPSO）算法进化搜索 RVM 的最优核参数和嵌入维数；最后，将各 AR 模型和 RVM 模型的预测值叠加作为最终预测值。

4 预测实例

以文献 [16] 中某型航空发动机的 16 份油液光谱分析数据（如图 2 所示）为例进行预测实验。

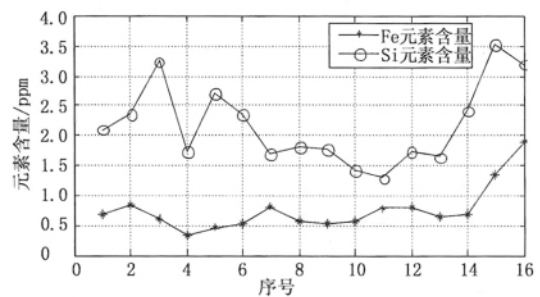


图 1 油液光谱分析数据

利用 EMD 算法分别对 Fe 元素和 Si 元素含量数据进行分解，分解结果如图 2 和图 3 所示。由图可以看出，RF 分量较为准确地表现出了原始数据的变化趋势。这说明采用镜像延拓法能够有效地抑制 EMD 存在的端点效应，经其分解得到的分量可以较好地反映原始数据中的真实成分。为将本文方法与已有方法的性能进行对比，分别建立油液光谱的 AR 模型和 SVM 模型对后 4 个数据进行一步预测。各模型参数的设置如下：AR 模型参数通过最终预测误差（FPE）准则选取；SVM 的核函数选为 Gaussian 核函数，模型参数包括核参数 γ 、惩罚系数 C 、不敏感损失函数 ϵ 和嵌入维数 l ，参考本文 2.3 节的方法进行自适应优化选择， M 设为 30， i_{max} 设为 100， γ 的取值区间为 $[0.1, 10]$ ， C 的取值区间为 $[1, 1000]$ ， ϵ 的取值区间为 $[0.0001, 0.5]$ ， l 的取值区间为 $[1, 5]$ ；RVM 模型根据本文 2.3 节的方法进行自适应优化选择， M 设为 30， i_{max} 设为 100， γ 的取值区间为 $[1, 10]$ ， l 的取值区间为 $[1, 5]$ 。各模型的预测结果如表 1 和表 2 所示。

表 1 Fe 元素含量预测结果

序号	真实值 /ppm	AR		SVM		EMD+RVM-AR	
		预测值 /ppm	相对误差/%	预测值 /ppm	相对误差/%	预测值 /ppm	相对误差/%
13	0.64	0.71	10.94	0.63	1.56	0.62	3.13
14	0.68	0.65	4.41	0.63	7.35	0.88	29.41
15	1.35	0.70	48.15	0.81	40.00	1.18	12.59
16	1.91	0.91	52.36	1.36	28.80	1.73	9.42
最大相对误差/%		52.36		40.00		29.41	
平均相对误差/%		28.96		19.43		13.64	

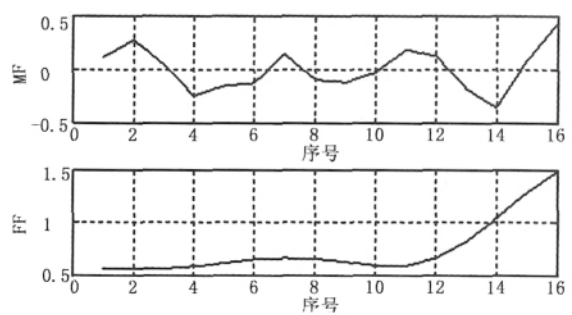


图 2 Fe 元素数据 EMD 分解结果

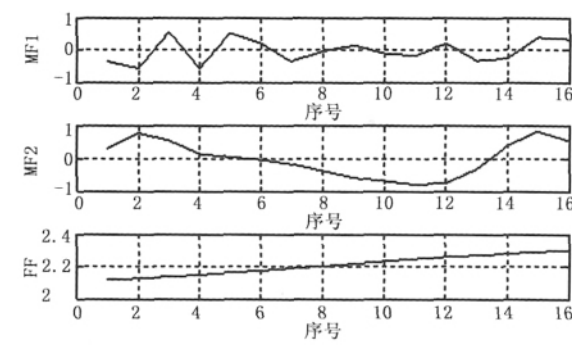


图 3 Si 元素数据 EMD 分解结果

表 2 Si 元素含量预测结果

序号	真实值 /ppm	AR		SVM		EMD+RVM-AR	
		预测值 /ppm	相对误差 /%	预测值 /ppm	相对误差 /%	预测值 /ppm	相对误差 /%
13	1.64	1.15	29.88	1.46	10.98	1.60	2.44
14	2.42	1.00	58.68	1.60	33.88	2.35	2.89
15	3.51	1.22	65.24	2.45	30.20	3.59	2.28
16	3.19	1.26	60.50	4.37	36.99	3.56	11.60
最大相对误差/%		65.24		36.99		11.60	
平均相对误差/%		53.57		28.01		4.80	

根据预测结果可知, AR 模型作为一种传统的时间序列分析方法, 在用于非线性非平稳时间序列预测时具有较大的预测误差, 本例中 Fe 元素和 Si 元素含量预测最大相对误差分别高达 52.36% 和 65.24%, 平均相对误差分别高达 28.96% 和 53.57%。SVM 模型的预测精度较 AR 模型有较大提高, 但预测误差仍然较大。相对于 AR 模型和 SVM 模型, 本文方法的预测性能显著提高, 其中 Si 元素含量预测的平均相对误差仅为 4.8%, 大大低于 AR 模型的 53.57% 和 SVM 模型的 28.01%。本文方法取得了优异的预测性能, 原因主要有两方面: 一方面, 采用 EMD 算法分离原始润滑油光谱数据的随机项和趋势项, 能够有效降低数据的复杂度, 并通过子序列单独建模分散了预测风险; 另一方面, RVM 作为机器学习发展的最新成果, 具有强大的非线性处理能力。

5 结论

1) 本文建立的基于 EMD 和 RVM-AR 模型的航空发动

机润滑油金属含量预测模型, 结合了 EMD 多分辨的优势和 RVM 强大地非线性处理能力, 具有优异的预测性能, 能够实现润滑油金属含量变化趋势的准确预测。

2) 采用基于 QPSO 的智能优化算法实现了模型参数的自适应选择, 克服了人为设置 RVM 模型参数缺乏科学依据和推广性不强的缺点。

参考文献:

[1] 姜彩虹, 孙志岩, 王 曦. 航空发动机预测健康管理系统的关
键技术 [J]. 航空动力学报, 2009, 24 (11): 2589-2594.

[2] 尉询楷, 冯 悦, 刘 芳, 等. 军用航空发动机 PHM 发展策略及关
键技术 [J]. 航空动力学报, 2011, 26 (9): 2107-2115.

[3] 任国全, 张培林, 张英堂. 装备油液智能监控原理 [M]. 北京: 国
防工业出版社, 2006.

[4] Gan M L, Yang Z. Time-series modeling and fault forecast study
on spectral analysis of lubricating oil [J]. Transactions of Nanjing
University of Aeronautics & Astronautics, 2001, 18 (1): 86-90.

[5] 尉询楷, 李应红, 王 硕, 等. 基于支持向量机的航空发动机滑油
监控分析 [J]. 航空动力学报, 2004, 19 (3): 392-397.

[6] 李本威, 张 赞, 孙 涛. 基于免疫粒子群算法的滑油屑末支持向
量机预测模型设计 [J]. 航空动力学报, 2009, 24 (7):
1639-1643.

[7] 邓 森, 杨军锋, 杨 朴, 等. 基于混沌粒子群优化 LS-SVM 的
发动机磨损状态监测研究 [J]. 计算机测量与控制, 2011, 19 (8):
1853-1856.

[8] Tipping M E. The relevance vector machine [A] // Advances in
Neural Information Processing Systems [C]. Cambridge, USA:
MIT Press, 2000: 652-658.

[9] Tipping M E. Sparse Bayesian learning and the relevance vector ma-
chine [J]. Journal of Machine Learning Research, 2001, 1 (3):
211-244.

[10] 段 青, 赵建国, 马 艳. 优化组合核函数相关向量机电力负荷预
测模型 [J]. 电机与控制学报, 2010, 14 (6): 33-38.

[11] 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]. Proceedings of the Royal Society London,
1998, 454 (1): 903-995.

[12] 胡维平, 莫家玲, 龚英姬, 等. 经验模态分解中多种边界处理方
法的比较研究 [J]. 电子与信息学报, 2007, 29 (6): 1394-1398.

[13] 黄大吉, 赵进平, 苏纪兰. 希尔伯特-黄变换的端点延拓 [J]. 海
洋学报, 2003, 25 (1): 1-11.

[14] Sun J, Feng B, Xu W B. Particle Swarm Optimization with Parti-
cles Having Quantum Behavior [A] // Proceedings of IEEE 2004
Congress on Cybernetics and Intelligent Systems [C]. Portland,
USA, 2004: 325-331.

[15] 董春曦, 杨绍全, 饶 鲜, 等. 支持向量机推广能力估计方法比较
[J]. 电路与系统学报, 2004, 9 (4): 86-91.

[16] 周新立, 钟新辉, 李 艳. 基于灰色时序的航空发动机磨损故障预
测模型 [J]. 长沙航空职业技术学院学报, 2004, 4 (1): 20-21.