

基于相关向量 EMD 和 GMDH 重构的 故障率预测方法^{*}

徐廷学¹, 逯程¹, 王虹², 韩旭³

(1. 海军航空大学岸防兵学院 烟台, 264001) (2. 中央军委联合参谋部第 55 研究所 北京, 100094)
(3. 海军航空大学航空基础学院 烟台, 264001)

摘要 针对故障率数据的非线性非平稳特性及现有预测方法精度不足的问题,提出了一种基于相关向量经验模态分解(relevance vector empirical mode decomposition,简称 RVEMD)和数据处理组合法(grouped method of data handling,简称 GMDH)重构的预测方法。首先,通过 RVEMD 分离故障率时间序列的波动项和趋势项,分解产生序列的固有模态函数(intrinsic mode function,简称 IMF)和残余函数(residual function,简称 RF),通过相关向量机(relevance vector machine,简称 RVM)预测端点局部极值的方法抑制传统经验模态分解(empirical mode decomposition,简称 EMD)存在的端点效应,进一步利用 RVM 回归生成序列的上下包络,替代了常规三次样条插值法;其次,建立各 IMF 分量的 RVM 预测模型和 RF 分量的灰色预测模型,其中对标准 RVM 回归模型进行了改进,通过构建一种方差高斯核函数(variance Gauss kernel function,简称 VGKF)来提高核函数的全局性能和泛化能力,利用 H-Q 准则对训练空间预测嵌入维数进行优化,避免了主观选取的盲目性,同时构建了一种基于背景值优化的灰色预测模型;最后,通过 GMDH 算法产生的最优智能组合模型得到最终的预测结果。仿真实例结果表明,相比常规 EMD 分解后叠加预测法和其他单一模型预测法,该方法具有更加优异的预测性能,能够对故障率的变化趋势进行准确预测。

关键词 相关向量经验模态分解; 数据处理组合法; 端点效应; 方差高斯核函数; 灰色预测模型

中图分类号 TJ760; TH86

引 言

随着现代工业技术的不断发展,武器装备在自动化与信息化水平得到快速提升的同时,其组成结构也日趋复杂,装备使用过程中所出现的各种故障问题使得后勤维修保障面临极大的挑战。实现装备故障率的准确预测不仅对备件的贮存策略具有重要指导意义,还能够为视情维修(condition-based maintenance,简称 CBM)决策提供有效支持,使装备满足规定的战备完好率要求^[1-2]。

目前,针对故障率的预测方法主要分为两类,即基于模型和数据驱动的方法^[3]。由于装备组成结构复杂,部件关系模糊,加上众多不确定的未知因素相互影响,很难对故障率分布函数进行精确建模,若建模不当则会带来较大的预测误差,这也大大限制了

该类方法的应用^[4-6]。对于故障率数据这种具有一定波动的非线性非平稳时间序列,灰色模型预测法^[7-8]和自回归滑动平均模型(auto-regressive and moving average model,简称 ARMA 模型)预测法^[9-10]的预测效果也并不理想。由于不需要对系统模型有足够深入的了解,基于数据驱动的方法以其较强的灵活性日益成为了研究的热门方向。当前,相关学者对基于人工神经网络^[11-12](artificial neural network,简称 ANN)和支持向量机^[13-14](support vector machine,简称 SVM)的预测方法已进行了广泛研究,但它们也存在一些固有缺陷,如 ANN 在优化过程中易陷入局部最优值,且性能受到训练样本量的限制;SVM 模型参数设置复杂,稀疏性有限,它们都并非解决此类问题的最佳方案。因此,对于像装备故障率预测这一复杂的小样本建模问题,使用单一模型难以对其进行准确预测。

^{*} 国家自然科学基金资助项目(51605487);山东省自然科学基金资助项目(ZR2016FQ03);中国博士后科学基金资助项目(2016M592965)

收稿日期:2017-02-09;修回日期:2017-09-08

EMD^[15]是一种基于信号局部特征的自适应信号分解方法,该方法在保留小波变换多分辨优势的基础上,还解决了小波变换中分解尺度确定与小波基选取的问题,因此更加适于对故障率这类非线性非平稳时间序列进行分析。但是由于无法判断断点处的极值情况,常规EMD被端点效应所困扰,不仅信号的包络线会在端点处发散,还会随着分解过程的进行向内传播,造成分解后的信号失真,严重影响分解精度。

RVM^[16-17]是一种基于稀疏Bayes学习理论的机器学习新算法。与SVM相比,其核函数不受Mercer条件的限制,选取更加灵活;参数设置简单,避免了复杂的附加参数;解的稀疏性也远高于SVM,具有较好的泛化能力。RVM已在解决非线性回归估计等问题中逐渐展现出了超越经典机器学习算法的独特优势,所以笔者结合装备保障实际,针对现有单一预测方法精度不足的问题,提出了一种基于RVEMD和GMDH重构的预测方法。首先,通过RVEMD分离故障率时间序列的波动项和趋势项;其次,建立了各固有模态函数分量的改进RVM预测模型和残余函数分量的背景值优化灰色预测模型;最后,通过GMDH算法产生的最优智能组合模型得到最终预测结果。该方法的有效性及其优越性将在故障率预测实例中得到验证。

1 相关向量EMD分解

通过对复杂信号进行自适应多尺度分解,EMD算法逐步展现了其在非线性非平稳信号分析领域所特有的优势。EMD是将时间序列中不同尺度的波动或趋势逐级分解为一系列具有不同特征尺度的IMF分量。由于每个IMF分量具有一定的物理意义,分解结果突显了信号的局部特征,进一步分析即可准确把握原始信号的特征信息。

IMF必须具备下面两个条件^[15]:a.在整个时间范围内,局部极值点和过零点数目相同或至多相差一个;b.在任意时刻点,其上下包络线关于时间轴局部对称。EMD分解后的IMF分量实质上包含了从高频到低频的不同频率成分,残余函数分量则代表了原始序列的平均趋势。

传统EMD在利用三次样条插值法拟合上下包络线的过程中,将序列的两个端点同时默认为极大值和极小值点,这种不当的处理方式会导致包络线在端点处产生发散现象,并随着“筛选”过程的不断迭代而向序列内扩散,最终造成分解结果严重失真,

这就是EMD的端点效应。特别是像装备故障率时间序列这类相对长度较短的数据集,端点对EMD分解质量的影响更为显著,此时分解得到的各分量也失去了其实际的物理意义。

针对EMD的端点问题,基于信号外推预测的抑制方法目前已有不少研究成果,包括AR回归预测^[18]、神经网络预测^[19]、灰色预测^[20]和支持向量机预测^[21]。该类方法在EMD分解前即对信号两端进行多步预测以确定原信号端点的极值情况,作为后续采用三次样条插值拟合判断的依据。然而当面对时间序列的EMD分解预测问题时,这类处理方法并不适用,因为上述方法要通过对外推来判断端点的极值情况,必然会产生一些时间序列外的数据,这也就失去了EMD分解后再进行预测的意义。笔者针对时间序列预测的问题,基于RVM这一机器学习领域的新算法,提出了一种利用RVM回归生成信号上、下包络线的新方法——RVEMD,若 \mathbf{X} 为包含 n 个数据的时间序列,其基本过程如下:

1) 找出序列中局部极大值点 $\mathbf{X}_{\max} = [x_{\max}(1) \quad x_{\max}(2) \quad \cdots \quad x_{\max}(p)]^T$,并确定其对应时刻 $\mathbf{T}_{\max} = [t_{\max}(1) \quad t_{\max}(2) \quad \cdots \quad t_{\max}(p)]^T$;同理确定局部最小值点 $\mathbf{X}_{\min} = [x_{\min}(1) \quad x_{\min}(2) \quad \cdots \quad x_{\min}(q)]^T$ 和其对应时刻 $\mathbf{T}_{\min} = [t_{\min}(1) \quad t_{\min}(2) \quad \cdots \quad t_{\min}(q)]^T$;

2) 分别构建RVM训练集 $\{t_{\max}(k), x_{\max}(k)\}_{k=1}^p$ 和 $\{t_{\min}(k), x_{\min}(k)\}_{k=1}^q$ 进行学习,其中 $t_{\max}(k)$, $t_{\min}(k)$ 为输入变量, $x_{\max}(k)$, $x_{\min}(k)$ 为输出值,利用训练完成后的RVM分别对输入集 $[1 \quad 2 \quad \cdots \quad n]^T$ 进行预测,生成原序列 \mathbf{X} 的上包络线 \mathbf{U}_1 和下包络线 \mathbf{L}_1 ;

3) 令 $\mathbf{H}_1 = \mathbf{X} - \mathbf{M}_1$,其中 \mathbf{M}_1 为均值包络线,对 \mathbf{H}_1 重复步骤1和2直至其满足IMF条件,记 $\mathbf{C}_1 = \mathbf{H}_1$ 为 \mathbf{X} 的第1个IMF分量;

4) 将 $\mathbf{R}_1 = \mathbf{X} - \mathbf{C}_1$ 作为原始序列,重复上述步骤得到 \mathbf{X} 第2~ m 个IMF分量与一个余量(RF) \mathbf{R} 。

针对传统EMD的端点效应抑制问题,RVEMD采用了一种新的包络线生成方式,不仅避免了信号端点的极值判断,而且利用RVM良好的非线性预测能力获得了边界极值点。与此同时,由于跳过了分解前的外推预测步骤,该方法对故障率的时间序列分解更为适用,分解后的IMF波动分量可以采取RVM进一步预测,而像RF这一波动不大的趋势分量,笔者采用针对小样本的GM(1,1)模型进行预测。

2 改进 RVM 预测模型

2.1 高斯核函数的改进

RVM 是一种新的非线性稀疏 Bayes 学习理论,其良好的泛化性能和较少的相关向量使其在预测领域逐步成为研究的新热点。通过在 SVM 基础上把基于高斯过程中的 Bayesian 推理应用到核理论上,RVM 算法在先验参数的结构下利用自相关判定理论(automatic relevance determination,简称 ARD)移除不相关的样本点,从而实现了模型的稀疏化^[22]。虽然与 SVM 相比,RVM 的核函数由于稀疏、较少的超参数而计算量大减,且脱离了 Mercer 条件的限制,但是预先确定的核函数性能依然会直接影响到 RVM 回归预测的精度。

目前常用的核函数有高斯核函数、Laplace 核函数以及 Cauchy 核函数等。高斯基核函数是目前应用最广泛的一种核函数,其标准形式定义为 $K_{\text{std}}(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2b^2}\right\}$ ($i=1,2,\dots,N$)。其中:参数 b 为核宽度; N 为训练样本数。

作为一种基于距离方程的核函数,高斯核函数的测试点在由源空间映射到高维特征空间后往往会变得稀疏。只有当核函数不仅能在测试点保持较快的衰减,而且在无限远处也能保持一定的衰减时,上述情况才能得到改变。然而高斯核函数只满足前者,所以它是一种典型的局部核函数。近些年,很多学者通过构造高斯核函数与全局性核函数组成的混合核来提高核函数的性能^[23-24],但在这个过程中又引入了多个参数变量,大大增加了 RVM 的计算复杂度。笔者在高斯核函数的基础上,通过构造一种改进 VGKF 来提高 RVM 的性能。

可以观察到在高斯核函数的标准形式里,所有的特征元素都利用相同的尺度因子来衡量,这一特性也使得具有大数值尺度范围的特征要更占优势,这一局限性也严重影响了其在实用中的性能。为克服这一问题,使各特征元素进行零均值归一化并将其标为单位标准差形式

$$\bar{x}_i = (x_i - \mu_i) / \sigma_i \quad (1)$$

其中: μ_i 和 σ_i 为特征向量的均值与标准差。

将式(1)代入标准形式,并将得到的结果改写成核函数的形式,就产生了一个新的核函数,即

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2\sigma_i^2 b^2}\right\} \quad (i=1,2,\dots,N) \quad (2)$$

由于 σ_i^2 为第 i 个向量的方差,将式(2)称为方差高斯核函数(VGKF),这样特征向量就通过不同的尺度因子进行了刻画。进一步分析,若要使函数在远离样本点时仍保持一个缓慢的衰减速度,考虑将 Euclid 距离方程放到指数函数的分母上去,这样就得到改进后的 VGKF 表达式

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{2\sigma_i^2 b^2}{\|\mathbf{x}-\mathbf{x}_i\|^2}\right\} \quad (i=1,2,\dots,N) \quad (3)$$

这样构造的核函数不仅满足前文提到的两个条件,而且成功避免了未知参数的引入,另外在模型数据的输入前也不需要数据的归一化过程。文献[25]指出,调整唯一的带宽参数 b 对提高核函数的性能效果并不明显,可以视为正态分布中的标准差 σ 。

2.2 训练样本空间维数的优化

实现时间序列 $\{x_1, x_2, \dots, x_n\}$ 预测的关键在于寻找到滑动时间窗口 $X_t = \{x_{t-1}, x_{t-2}, \dots, x_{t-m}\}$ 与输出 $Y_t = \{X_t\}$ 之间的准确映射 $f: R^m \rightarrow R, t=m+1, m+2, \dots, n$,其中 m 为预测嵌入维数。因此,参数 m 的确定对于预测建模的准确性至关重要。为了克服 m 选取的主观盲目性,笔者基于混沌时序的局域预测法^[26-27]中邻近点个数选取的思想,对 RVM 的训练样本空间最优嵌入维数进行了优化,该方法适合小数据量的情况,具有计算简单和精度高的特点。

根据局域法小样本训练^[28-30]的思想,目标点的预测值可以通过邻近相点的运动趋势而进行推断,即通过映射来近似当前矢量 $\mathbf{X}(n)$ 与未来值 \hat{x}_{n+T} 的演化值,而这一映射的近似过程是依赖于 $\mathbf{X}(n)$ 的 d 个邻近点 $\mathbf{X}(n_i)$ ($i=1,2,\dots,d$) 而实现的。在本研究中,考虑利用邻近点训练 RVM,提出一种 RVM 训练样本空间维数优化的方法。把 $\mathbf{X}(n_i)$ ($i=1,2,\dots,d$) 作为输入样本, x_{n_i+1} ($i=1,2,\dots,d$) 作为输出值进行 RVM 的回归预测训练,进而得到权重系数和相关向量。这样就通过 RVM 回归模型代替了局域线性模型,充分利用了 RVM 良好的逼近性能使其训练后得出了一个最优函数来拟合 $f(\cdot)$,计算便可得到后续数据的预测值 \hat{x}_{n+T} 。

在邻近点个数的确定过程中,笔者基于 Hannan-Quinn 准则^[31],过程如下。

首先对 d 给定一个宽泛的范围 $d \in [d_{\min}, d_{\max}]$,依次计算每个 d 值下的 HQ 准则值

$$C(d) = \ln\sigma(d)^2 + (d+1) \frac{H \ln(\ln N)}{N} \quad (4)$$

$$\sigma^2(d) = \frac{1}{S} \sum_{j=1}^S (x_j - x'_j)^2 \quad (5)$$

其中: H 为表征权重的常数, 一般取 $H > 2$; N 为预测样本个数; σ^2 为拟合方差, x_j 为样本数据点; \bar{x} 为样本均值; x'_j 为样本预测值。

式(4)取得最小值时所对应的 d_{opt} 值即为最佳邻近点个数, 此时模型在理论上取得精度和复杂度的最佳平衡。

这一过程通过明确 $m = d_{\text{opt}}$ 将训练样本的空间维数进行了优化, 克服了以往仅凭主观经验和 FPE 准则^[32]进行回归训练的不足, 具有一定参考性和实用价值。

3 GM(1,1)的背景值修正模型

由于故障率变化趋势含有一些复杂的不确定及未知因素, 而灰色模型对“贫信息、少数据”的小波动序列有较好的预测效果, 所以笔者采用 GM(1,1)模型^[33]对 RVEMD 分解后的 RF 项进行预测。

通过分析常规 GM(1,1)模型建模过程可知, 决定模型预测精度的关键在于参数 v 和 u 的估计是否准确, 而参数的最小二乘求解过程又与 $z^{(1)}(k)$ 的构造形式密不可分。因此, 构建精确的背景值 $z^{(1)}(k)$ 计算公式对于提高 GM(1,1)模型的预测精度, 避免模型从连续到离散所产生的跳跃误差都具有重要意义。

将模型的白化微分方程两边在区间 $[k-1, k]$ 上同时进行积分得

$$x^{(0)}(k) + v \int_{k-1}^k x^{(1)}(t) dt = u \quad (6)$$

参数 v, u 在此形式下所求得的误差最小。

进一步从几何角度进行分析, 传统背景值 $z^{(1)}(k)$ 表示为 $x^{(1)}(k-1)$ 与 $x^{(1)}(k)$ 的平均值, 而背景值的准确表达式 $z_0(k) = \int_{k-1}^k x^{(1)}(t) dt$ 应为 $x^{(1)}(t)$ 与横坐标所围曲边梯形的面积。因此, 背景值选取是否准确直接影响到参数 v, u 的估计值, 这也从几何意义上解释了传统背景值构造方法使得 GM(1,1)模型对高速增长序列进行预测会产生较大系统误差的原因。

本节从推导背景值与发展系数 v 之间数量关系的角度出发, 通过最小二乘法对二次参数进行估计, 还原得到原始参数估计值, 提供一种背景值优化新思路。

3.1 背景值与发展系数的数量关系

根据中值定理可知, 存在 $\lambda \in [0, 1]$, 使得 $z_0(k)$ 满足

$$z_0(k) = \lambda x^{(1)}(k-1) + (1-\lambda)x^{(1)}(k) \quad (7)$$

在 GM(1,1)模型中, 通过指数形式对 1-AGO 序列 $X^{(1)}$ 进行近似, 所以设 $x^{(1)}(k) = ae^{-v(k-1)} + b$, 则式(7)为

$$z_0(k) = \frac{a}{e^{v(k-2)}} \left(\frac{1-\lambda + \lambda e^v}{e^v} \right) + b \quad (8)$$

根据背景值的几何意义有

$$z_0(k) = \int_{k-1}^k x^{(1)}(t) dt = \frac{a}{e^{v(k-2)}} \left(\frac{1}{v} - \frac{1}{ve^v} \right) + b \quad (9)$$

由于式(8)和式(9)相等, 化简得到背景值 $z_0(k)$ 中的权重

$$\lambda = \left(\frac{1}{v} - \frac{1}{e^v - 1} \right) \quad (10)$$

3.2 拟合误差最小化的 GM(1,1)模型求解

将新的背景值构造公式代入灰色微分方程, 模型转化为

$$x^{(0)}(k) + v[\lambda x^{(1)}(k-1) + (1-\lambda)x^{(1)}(k)] = u \quad (11)$$

化简得到

$$(1 - v\lambda)x^{(0)}(k) + vx^{(1)}(k) = u \quad (12)$$

首项系数归一化, 进而得到

$$x^{(0)}(k) + \frac{v}{(1-v\lambda)}x^{(1)}(k) = \frac{u}{(1-v\lambda)} \quad (13)$$

令 $v^* = \frac{v}{1-v\lambda}$, $u^* = \frac{u}{1-v\lambda}$, 则式(13)变为

$$x^{(0)}(k) + v^*x^{(1)}(k) = u^* \quad (14)$$

该式称为 GM(1,1)模型的变换式。利用最小二乘法对新参数向量 $Q^* = (v^*, u^*)^T$ 进行估计, 得到

$$Q^* = (\hat{v}^*, \hat{u}^*) = (C^T C)^{-1} C D \quad (15)$$

其中

$$C = \begin{bmatrix} -x^{(1)}(2) & 1 \\ -x^{(1)}(3) & 1 \\ \vdots & 1 \\ -x^{(1)}(n) & 1 \end{bmatrix} \quad (16)$$

$$D = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T \quad (17)$$

根据式(10)的推导及 v, u, λ 与二次参数 \hat{v}^*, \hat{u}^* 的关系, 可得方程组

$$\begin{cases} \lambda = \frac{1}{v} - \frac{1}{e^v - 1} \\ \hat{v}^* = \frac{v}{1 - v\lambda} \\ \hat{u}^* = \frac{u}{1 - v\lambda} \end{cases} \quad (18)$$

解方程组得

$$\begin{cases} v = \ln(\hat{v}^* + 1) \\ \lambda = \frac{1}{\ln(\hat{v}^* + 1)} - \frac{1}{\hat{v}^*} \\ u = \frac{\hat{u}^* \ln(\hat{v}^* + 1)}{\hat{v}^*} \end{cases} \quad (19)$$

将还原得到的发展系数 v 与灰色作用量 u 重新代入 GM(1,1)模型中的白化微分方程

$$\frac{dx^{(1)}}{dt} + vx^{(1)} = u \quad (20)$$

方程的离散解为

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{u}{v})e^{-vk} + \frac{u}{v} \quad (21)$$

3.3 新陈代谢

将预测值 $\hat{x}^{(0)}(k+1)$ 添加到原始序列中,同时删除 $x^{(0)}(1)$,构造新序列

$$\{x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(k+1)\} \quad (22)$$

取 $x^{(1)}(2) = x^{(0)}(2)$,依据新模型的步骤实现下一时刻的动态预测更新。

4 基于相关向量 EMD 与 GMDH 重构的故障率预测方法

4.1 GMDH 算法

利用所提的 RVEMD 方法将故障率序列分解为趋势分量(残余项)和各波动分量(各阶 IMF 项)后,依据各分量序列自身特点进行预测,很自然地会将分量的预测值相加得到最终的预测结果。但是根据 EMD 的完备性与正交性特点,这一处理方式很有可能会造成某些信息的丢失,从而影响预测的精度。考虑到各分量背后所蕴含的物理意义,高频到低频分量可能代表了不规则因素和重大事件等因素对故障率序列的影响,实际过程中这些因素的影响程度并不完全相同,所以需要挖掘出隐含在 IMF 分量和余项中的隐含信息,以确定各因素的重要程度,构建基于客观权重的最优智能组合模型,实现提高预测精度的目的。

作为自组织数据挖掘建模的核心,GMDH 算法是一种新兴的数据挖掘理论^[34],目前已在经济预测

领域中得到了广泛应用^[35]。GMDH 算法实际上是一种数据分组处理方法,它基于遗传与进化的思想,按照一定的准则从一系列待选模型中挑选一个最优复杂度模型。该算法在建模过程中以“主动”方式自主选择,不存在主观因素,在小数据情况下,是一种有效的非线性预测模型组合方法^[36]。其基本步骤如下。

1) 样本数据分组。在预测问题中,将 N 个样本数据集分为训练集 A 、检测集 B 和预测集 C ,其中 $N_\omega = N_A + N_B + N_C$, $\omega = A \cup B \cup C$ 。

2) 选择参考函数。通过一个参考函数建立输出与输入变量的一般关系。

3) 确定一个目标函数。选择一个具有外补充性质的准则作为每层模型的筛选准则。

4) 产生第 1 层局部模型。选取某传递函数 $y_k = f_k(v_i, v_j)$ 作为第 1 层模型,其中 v_i, v_j 为目标函数的子项模型,并在训练集 A 上估计 y_k 参数。根据外准则在检测集 B 上进行筛选作为下一层局部模型的输入变量。

5) 重复上一步,得到更多的局部模型,从而生成最优复杂度结构模型

4.2 基于 RVEMD 与 GMDH 重构的故障率预测模型

若 \mathbf{X} 为包含 n 个数据的原始故障率时间序列,则基于 GM-RVEMD 与 GMDH 重构的预测方法主要计算步骤如下。

1) 利用 RVEMD 对 \mathbf{X} 进行分解,得到 m 个 IMF 分量 $\mathbf{C}_i = [c_i(1) \ c_i(2) \ \dots \ c_i(n)]^T$ ($i=1, 2, \dots, m$) 和一个 RF 分量 $\mathbf{R} = [r(1) \ r(2) \ \dots \ r(n)]^T$ 。

2) 利用 2.2 节方法确定嵌入维数 d ,构建 IMF 分量 \mathbf{C}_i 的 RVM 训练样本集 $P_i = \{\mathbf{I}_i(j), O_i(j)\}_{j=1}^d$,其中输入向量 $\mathbf{I}_i(j) = [c_i(j) \ c_i(j+1) \ \dots \ c_i(j+d-1)]^T$,模型输出 $O_i(j) = c_i(j+d)$ 。

3) 以相对误差平方和最小作为外准则,在样本集上利用 GMDH 算法对每个 IMF 分量和 RF 分量进行最优复杂度的筛选,得到最优组合模型

$$x(k) = \sum_{i=1}^m \omega_i c_i(k) + r(k) + a \quad (23)$$

其中: ω_i 为 c_i 的权重值; a 为常数项。

4) 利用 P_i ($i=1, 2, \dots, m$) 对 RVM 进行训练,分别将 $\mathbf{P}_i \mathbf{I}_i = [c_i(n-d+1) \ c_i(n-d+2) \ \dots \ c_i(n)]^T$ 作为输入变量代入模型,得到 $n+1$ 时刻的

IMF 分量预测值 $\hat{c}_i(n+1), i=1, 2, \dots, m$ 。利用第3节的灰色背景值修正模型得到余项 RF 分量的预测值 $\hat{r}(n+1)$ 。

5) 将 $\hat{c}_i(n+1)$ 和 $\hat{r}(n+1)$ 代入式(23), 将得到的 $\hat{x}(n+1) = \sum_{i=1}^m \omega_i \hat{c}_i(n+1) + \hat{r}(n+1)$ 作为 $x(n+1)$ 的预测值。

该算法流程图如图1所示。

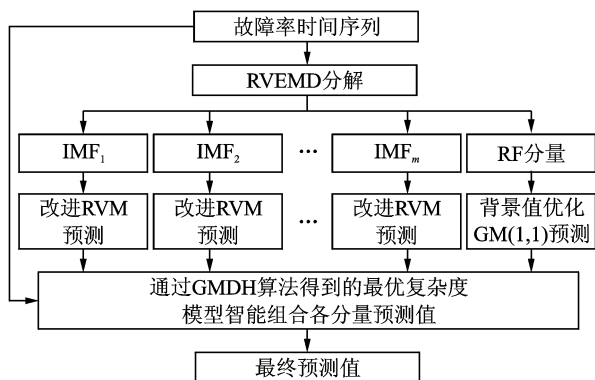


图1 算法流程图

Fig. 1 Flow chart of algorithm

5 实例分析

5.1 导弹故障率预测

以某部队处于贮存状态的整批导弹为研究对象, 一般情况下为满足一定的战备完好率要求, 导弹会定期进行通电检测, 每次间隔为半年。笔者以2007年~2016年共20个故障率数据为样本对本研究所提预测方法进行检验, 如图2所示。

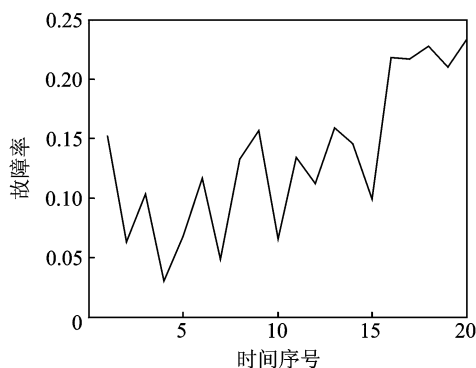


图2 导弹故障率时间序列

Fig. 2 Missile failure rate time series

EMD 作为非平稳非线性信号分析领域的有力工具, 利用其进行时间序列分解时必须保证故障率数据满足非线性非平稳特性, 在此做简要分析。

首先, 对故障率数据的非平稳特性进行分析。由图2可见, 故障率数据整体上呈缓慢上升的变化趋势, 并非围绕一个水平的中心趋势, 所以从直观上可以初步判定其非平稳特性。下面从平稳性的定义出发, 一个平稳的时间序列 \mathbf{X}_t 须满足以下3个条件:

- 1) 对任意时间 t , 均值 $E(\mathbf{X}_t) = \mu$ 为常数;
- 2) 对任意时间 t , 方差 $\text{Var}(\mathbf{X}_t) = \sigma^2$ 为常数;
- 3) 对任意时间 t 和 s , 协方差 $\text{Cov}(\mathbf{X}_t, \mathbf{X}_s) = \gamma_{t-s}$, 只与时间间隔 $t-s$ 有关。

故障率数据的平稳性可以采用单位根检验法, 利用 Matlab 软件中的 `adftest` 函数对故障率数据进行计算, 得到函数返回值为 0, 证明了故障率数据的非平稳特性。

接下来分析故障率数据的非线性特性。BDS 检验^[37] 是 Broock 提出的一种检验时间序列非线性特性的有效方法, 在 5% 显著性水平下, 若检验统计量的绝对值超过 1.96, 则拒绝原独立同分布的假设, 认为该序列存在非线性相关特征。通过 Matlab 软件对故障率数据进行 BDS 检验, 嵌入维数设置为 4, 距离 δ 为 1.5 σ , 检验统计量为 15.1361, 强烈地拒绝了原假设, 也证明了故障率数据的非线性特性。

下面利用本研究提出的 RVEMD 与 EMD 对故障率时域波形进行分解, 结果如图3、图4所示。取样本的前 18 个故障率数据进行建模分析, 后 2 个数据作为预测结果检验。由图3可以看出, 分解数据包含 2 个 IMF 分量与 1 个 RF 余量, 余量准确反映了故障率数据稳中有升的变化趋势。图4所示的传统 EMD 分解结果中包含了 3 个 IMF 分量和 1 个 RF 余量, 复杂度较高。同时由于端点效应的影响, $\text{IMF}_1 \sim \text{IMF}_3$ 的两端数据都存在幅值减小变为零的现象, 误差的积累导致 RF 余量表现为先减小后增大的趋势, 这与工程实践中该批导弹的实际故障率变化趋势不符。因此, 从图3和图4的对比中可以看出, 经 RVEMD 分解的故障率数据更具规律性, 得到的波动项和趋势项对反映原始数据的真实物理成分也更具代表意义, 以此为基础进行下一步的预测建模分析。

利用 GMDH 算法对 RVEMD 分解后的分量数据进行智能组合, 得到的最优复杂度模型为

$$x(k) = 0.865c_1(k) + 1.312c_2(k) + r(k) \quad (24)$$

由式(24)可以看出, 经 RVEMD 分解后的各 IMF 变量都参与了模型的构建, 并未产生没有意义

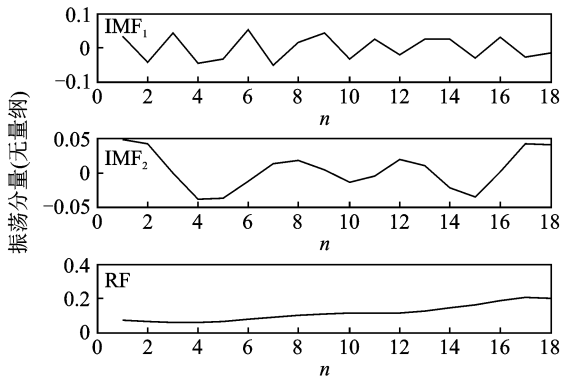


图 3 故障率数据 RVEMD 分解结果

Fig. 3 RVEMD results of failure rate data

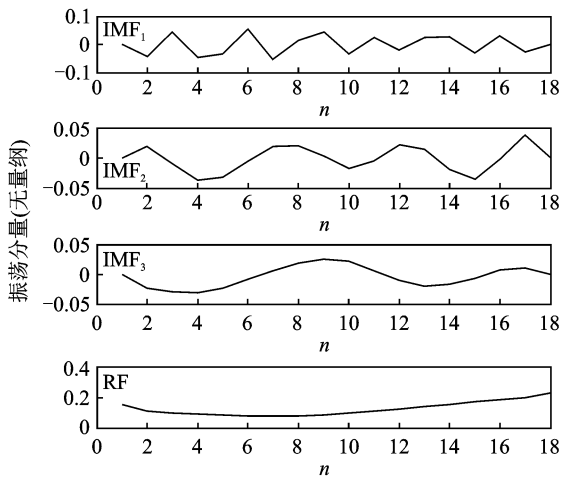


图 4 故障率数据 EMD 分解结果

Fig. 4 EMD results of failure rate data

的虚假分量,这也从侧面说明了 RVEMD 分解的有效性。进一步分析,第 1 个 IMF 分量 c_1 的权系数仅为 0.865,说明该 IMF 分量对最终结果的贡献率较小,可能代表一些随机因素对故障率造成的波动;第 2 个 IMF 分量 c_2 的权系数达到了 1.312,说明该分量可能代表一些影响故障率的重要因素,如部队规律性的维修活动等。因此,通过该模型可以明确各分量信息的重要程度,再利用本研究提出的改进 RVM 和灰色模型分别对分解得到的 IMF 分量和 RF 分量进行预测,对预测分析具有重要意义。

为将本研究方法与已有方法进行对比分析,表 1 和表 2 分别为本研究方法与各单一模型在 RVEMD 与 EMD 下的预测性能比较结果。其中,平均绝对百分比误差(MAPE)为预测性能评价指标,LS-SVM 选取 Gaussian 核函数,核参数和正则化参数通过留一交叉验证法确定,嵌入维数为 4。

由表 1 可以看出,由于通过 RVM 预测局部极值的方法对 EMD 的端点效应进行了抑制,故障率数据的各 IMF 分量与 RF 分量的分解失真程度得到了有效降低,因此联合 RVEMD 得到的预测结果要明显优于常规 EMD 分解的预测结果。与此同时,通过加入 GMDH 算法对 RVEMD 分解后的各分量进行智能组合,预测精度较单纯的分解后叠加法又有了显著提高。这是由于 GMDH 算法生成的最优复杂度模型可以有效挖掘各分量因素间的隐含信息,体现了分量的物理特征。

表 1 预测结果对比

Tab. 1 The results of prediction

序号	真实值	RVEMD+GMDH+GM-RVM		RVEMD+GM-RVM		EMD+GM-RVM	
		预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%
19	0.210 5	0.223 7	6.2	0.227 6	8.1	0.190 2	9.6
20	0.233 3	0.229 8	1.5	0.220 3	5.6	0.214 7	7.9
MAPE/%		3.85		6.85		8.75	

表 2 单一模型预测结果

Tab. 2 The results of prediction based on single model

序号	真实值	RVM		LS-SVM		GM(1,1)	
		预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%
19	0.210 5	0.188 6	10.4	0.236 8	12.5	0.242 6	15.2
20	0.233 3	0.208 5	10.6	0.260 2	11.5	0.277 4	18.9
MAPE/%		10.50		12.00		17.05	

进一步比较表 1 与表 2 可以看出,通过单一模型对故障率进行预测效果并不理想,尤其是针对故障率数据这类波动较大的时间序列,常规 GM(1,1)

模型并不适用。RVM 和 LS-SVM 的预测精度又取决于模型的具体参数,在未对参数进行优化的常规模型框架下得到的预测精度也较为一般。但是,作

为故障率数据主要成分的趋势项分量,笔者通过背景值优化后的改进 GM(1,1)模型对较为平稳的趋势项 RF 分量进行预测,对提高组合后预测结果的精度有重要影响。利用 RVM 优秀的非线性处理能力对波动项 IMF 分量进行预测,改进的核函数也进一步提高了预测精度。根据 GMDH 算法对各分量进行了智能融合,预测结果有了明显改善,而且这种子序列单独建模预测的方法可以有效规避预测风险。

5.2 Boeing 飞机故障率预测

选取故障率预测领域内的经典算例,即 Boeing757~700 飞机 1996 年 9 月至 1998 年 8 月两年内的 24 个故障率数据^[38](见图 5),继续检验本研究所提方法。

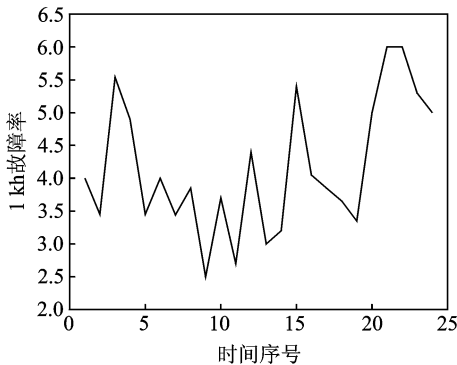


图 5 飞机故障率时间序列

Fig. 5 Aircraft failure rate time series

首先依然是通过 RVEMD 与 EMD 对前 20 个数据进行分解,结果如图 6 和图 7 所示,最后 4 个故障率数据作为预测结果检验。由图 5 可以看出,故障率时间序列大致呈先减小后增大的趋势。进一步对比图 6 和图 7 的分解结果可以看出,RVEMD 分解得到的趋势项 RF 分量与故障率时序的真实变化趋势更加吻合,呈经典“浴盆曲线”形状,而且克服了常规 EMD 的端点效应,同时复杂程度也大幅降低,更适用于以此基础进行预测建模。

利用 GMDH 算法得到的最优权重模型为

$$x(k) = 0.958c_1(k) + r(k) + 0.193 \quad (25)$$

由式(25)可以看出,RVEMD 分解得到的两个分量都被选进了模型中,证明了分解的有效性。另外,模型中还出现了一个小的常数项,它代表了一些未知的影响故障率的不确定性因素。通过 GMDH 算法挖掘出有效信息后,下一步分别对各分量进行预测。

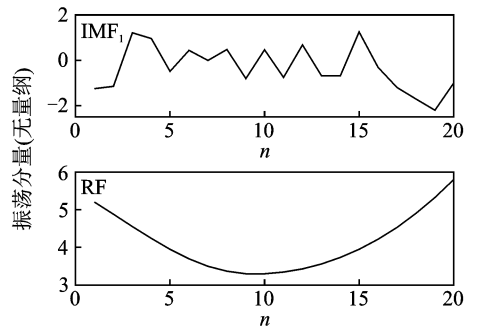


图 6 故障率数据 RVEMD 分解结果

Fig. 6 RVEMD results of failure rate data

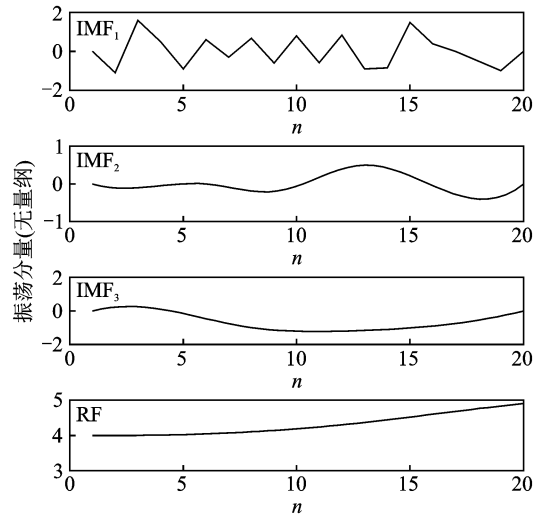


图 7 故障率数据 EMD 分解结果

Fig. 7 EMD results of failure rate data

RVEMD 分解得到的余项 RF 分量是故障率数据的主要成分,在利用背景值优化的 GM(1,1)模型进行预测时,不断加入新的预测信息向后预测,以提高精度;波动项 IMF 分量依然采用改进 RVM 算法;最后将各分量预测结果代入式(25)得到故障率最终预测值。为将本研究方法与已有方法进行对比分析,表 3 和表 4 列出了本方法与各单一模型在 RVEMD 与 EMD 下的预测性能比较结果,各参数设置如上节所示。

从表 3 和表 4 的结果可以看出,在与 RVEMD 联合模式下,故障率的预测精度较单一预测方法有较大的改善。由于故障率数据的波动特性,独立 GM(1,1)模型的预测效果很差,其 MAPE 甚至达到了 19.3%,这也说明了笔者提出的 RVEMD 方法对故障率数据分解的有效性,不仅降低了时间序列的复杂程度,还增强了其规律性,适于以此为基础进行建模预测。最终的预测结果由各具体模型来决定,由于 RVEMD 对端点效应的抑制,使分解得到

表 3 预测结果对比
Tab. 3 The results of prediction

序号	真实值	RVEMD+GMDH+GM-RVM		RVEMD+GM-RVM		EMD+GM-RVM	
		预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%
21	6.0	5.78	3.7	6.33	5.5	6.47	7.8
22	6.0	6.37	6.2	5.54	9.0	6.42	7.0
23	5.3	5.44	2.6	5.76	8.7	5.78	14.7
24	5.0	5.33	6.6	5.56	11.0	5.64	12.8
MAPE/%		4.78		8.55		10.58	

表 4 单一模型预测结果
Tab. 4 The results of prediction based on single model

序号	真实值	RVM		LS-SVM		GM(1,1)	
		预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%
21	6.0	5.34	11.0	6.73	12.2	6.89	14.8
22	6.0	5.08	15.3	6.82	13.6	7.04	17.3
23	5.3	5.96	12.5	6.04	14.0	6.52	23.0
24	5.0	5.67	13.4	5.84	16.8	6.07	21.4
MAPE/%		13.05		14.15		19.13	

的 IMF 分量和 RF 分量较常规 EMD 算法失真极小,以此为基础采用优化的 RVM 和 GM(1,1)模型进行预测,精度必然有所提高。利用 GMDH 算法挖掘有效信息并智能组合各分量预测值后,最终得到的预测结果精度有了进一步的改善。

6 结束语

针对传统 EMD 中存在的端点效应,笔者提出了一种基于 RVM 预测局部均值的经验模态分解新方法,在利用 EMD 多分辨优势的基础上有效克服了时间序列分解的失真现象,显著提高了基于 RVEMD 的预测模型质量。通过引入 GMDH 算法得到了一个最优复杂度模型,明确了各分解因素间的重要程度,使分量的物理特征得以体现。最后,对各子序列单独进行预测建模,通过改进的 RVM 模型和背景值优化后的 GM(1,1)模型分别对 IMF 分量和 RF 分量进行预测。通过故障率的预测实例表明,基于 RVEMD 分解与 GMDH 重构的方法可以很好地反映故障率的变化规律,且具有较高的预测精度,为解决故障率的预测问题提供了一个新的思路。

参 考 文 献

[1] 李瑞莹,康锐. 基于 ARMA 模型的故障率预测方法研究[J]. 系统工程与电子技术, 2008,30(8):1588-

1591.
Li Ruiying, Kang Rui. Research on fault rate prediction method based on ARMA model [J]. Systems Engineering and Electronics, 2008,30(8):1588-1591. (in Chinese)
[2] 李生彪. 基于 ARMA 模型的故障率时间序列预测[J]. 自动化与仪器仪表, 2015(12):218-219.
Li Shengbiao. Prediction of failure rate time series based on ARMA model [J]. Automation and Instrumentation, 2015(12):218-219. (in Chinese)
[3] Sikorska J Z, Hodkiewicz M, Ma Lin. Prognostic modelling options for remaining useful life estimation by industry[J]. Mechanical Systems and Signal Processing, 2011,25(5):1803-1836.
[4] Vichare N M, Pecht M G. Prognostics and health management of electronics[J]. IEEE Transactions on Components and Packaging Technologies, 2006, 29(1):222-229.
[5] Lorton A, Fouladirad M, Grall A. A methodology for probabilistic model-based prognosis [J]. European Journal of Operational Research, 2013, 225(3): 443-454.
[6] 胡昌华,张琪,乔玉坤. 强跟踪粒子滤波算法及其在故障预报中的应用[J]. 自动化学报, 2008,34(12):1522-1528.
Hu Changhua, Zhang Qi, Qiao Yukun. Strong tracking particle filter algorithm and its application in fault prediction [J]. Journal of Automation, 2008,34(12): 1522-1528. (in Chinese)

- [7] Deng Xiujian, Luo Qiang, Zhao Yiyang, et al. Avionics equipment failure prediction based on genetic programming and grey model[C]//Seventh International Conference on Electronics and Information Engineering. Nanjing:International Society for Optics and Photonics, 2017.
- [8] 邵延君, 潘宏侠, 马春茂, 等. 基于新陈代谢组合模型的装备故障预测[J]. 振动、测试与诊断, 2015, 35(2):359-362.
Shao Yanjun, Pan Hongxia, Ma Chunmao, et al. Equipment fault prediction based on metabolic combination model [J]. Journal of Vibration, Measurement & Diagnosis, 2015, 35(2):359-362. (in Chinese)
- [9] Pham H T, Yang B S. Estimation and forecasting of machine health condition using ARMA/GARCH model[J]. Mechanical Systems and Signal Processing, 2010, 24(2):546-558.
- [10] 丛飞云, 陈进, 董广明. 基于 AR 模型的 Kolmogorov-Smirnov 检验性能退化及预测研究[J]. 振动与冲击, 2012, 31(10):79-82.
Cong Yunfei, Chen Jin, Dong Guangming. Research on performance degradation and prediction of Kolmogorov-Smirnov test based on AR model [J]. Vibration and Shock, 2012, 31(10):79-82. (in Chinese)
- [11] Yuan Lifan, He Yigang, Huang Jiaoying, et al. A new neural-network-based fault diagnosis approach for analog circuits by using kurtosis and entropy as a pre-processor[J]. IEEE Transactions on Instrumentation and Measurement, 2010, 59(3):586-595.
- [12] Wu Bairong, Tian Zhigang, Chen Mingyuan. Condition-based maintenance optimization using neural network-based health condition prediction [J]. Quality and Reliability Engineering International, 2013, 29(8): 1151-1163.
- [13] Fronza I, Sillitti A, Succi G, et al. Failure prediction based on log files using random indexing and support vector machines[J]. Journal of Systems and Software, 2013, 86(1):2-11.
- [14] 姜媛媛, 王友仁, 罗慧, 等. 电力电子电路故障评估新指标及基于 LSSVM 的预测新方法[J]. 电工技术学报, 2012, 27(12):43-50.
Jiang Yuanyuan, Wang Youren, Luo Hui, et al. New indices for fault assessment of power electronic circuits and a new prediction method based on LSSVM[J]. Transactions of China Electrotechnical Society, 2012, 27(12):43-50. (in Chinese)
- [15] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis[J]. Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences, 1998, 454(1971):903-995.
- [16] Tipping M E. Sparse Bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research, 2001(1):211-244.
- [17] Samui P, Dixon B. Application of support vector machine and relevance vector machine to determine evaporative losses in reservoirs[J]. Hydrological Processes, 2012, 26(9):1361-1369.
- [18] 郭明威, 倪世宏, 朱家海, 等. 振动信号中的 HHT/EMD 端点延拓方法研究[J]. 振动与冲击, 2012, 31(8):62-66.
Guo Mingwei, Ni Shihong, Zhu Jiahai, et al. Study of HHT/EMD endpoint extension method in vibration signals [J]. Vibration and Shock, 2012, 31(8):62-66. (in Chinese)
- [19] 孟宗, 闫晓丽, 王赛. 基于 HMM 校正与神经网络延拓的 EMD 端点效应抑制方法[J]. 中国机械工程, 2015, 26(14):1920-1925.
Meng Zong, Yan Xiaoli, Wang Sai. EMD endpoint suppression method based on HMM correction and neural network extension [J]. China Mechanical Engineering, 2015, 26(14):1920-1925. (in Chinese)
- [20] Zong Yongtao, Shen Yanxia, Ji Zhicheng, et al. Research on restraining the end effect of EMD based on grey prediction model[C]//Control and Decision Conference (2014 CCDC), the 26th Chinese. Changsha: IEEE, 2014:214-217.
- [21] 旷欢, 王如龙, 张锦, 等. 基于 SVM 的 EMD 端点效应抑制方法研究[J]. 计算机工程与应用, 2015(11): 196-200.
Kuang Huan, Wang Rulong, Zhang Jin, et al. Study of EMD endpoint effect suppression method based on SVM [J]. Computer Engineering and Applications, 2015(11):196-200. (in Chinese)
- [22] 周建宝. 基于 RVM 的锂离子电池剩余寿命预测方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2013.
- [23] Miao Shuwei, Xie Kaigui, Yang Hejun, et al. A mixture kernel density model for wind speed probability distribution estimation [J]. Energy Conversion and Management, 2016, 126:1066-1083.
- [24] Liu Kenghao, Lin Yenyu, Chen Chusong. Linear spectral mixture analysis via multiple-kernel learning for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(4):2254-2269.
- [25] Remaki L, Cheriet M. KCS-new kernel family with compact support in scale space: formulation and impact[J]. IEEE Transactions on Image Processing,

- 2000,9(6):970-981.
- [26] Li Hengchao, Zhang Jiashu. Local prediction of chaotic time series based on support vector machine[J]. Chinese Physics Letters, 2005,22(11):2776-2779.
- [27] 杜杰,曹一家,刘志坚,等.混沌时间序列的局域高阶Volterra滤波器多步预测模型[J].物理学报,2009,58(9):5997-6005.
Du Jie, Cao Yijia, Liu Zhijian, et al. A multi-step prediction model of local high order Volterra filters with chaotic time series [J]. Acta Physica Sinica, 2009,58(9):5997-6005. (in Chinese)
- [28] Meng Qingfang, Chen Yuehui, Peng Yuhua. Small-time scale network traffic prediction based on a local support vector machine regression model[J]. Chinese Physics B, 2009,18(6):2194-2199.
- [29] Bing Qichun, Gong Bowen, Yang Zhaosheng, et al. Short-term traffic flow local prediction based on combined kernel function relevance vector machine model [J]. Mathematical Problems in Engineering, 2015, 154703:1-9.
- [30] Meng Qingfang. Nonlinear prediction of small scale network traffic based on local relevance vector machine regression model[J]. Acta Physica Sinica, 2013, 62(15):1-7.
- [31] Hannan E J, Quinn B G. The determination of the order of an autoregression[J]. Journal of the Royal Statistical Society, Series B :Methodological, 1979, 41(2):190-195.
- [32] 孟庆芳,彭玉华,曲怀敬,等.基于信息准则的局域预测法邻近点的选取方法[J].物理学报,2008,57(3):1423-1430.
- Meng Qingfang, Peng Yuhua, Qu Huaijing. Selection method of adjacent points based on information criterion of local prediction method[J]. Acta Physica Sinica, 2008,57(3):1423-1430. (in Chinese)
- [33] Liu Sifeng, Zeng Bo, Liu Jiefang, et al. Four basic models of GM(1,1) and their suitable sequences[J]. Grey Systems: Theory and Application, 2015, 5(2): 141-156.
- [34] Mueller J A, Lemke F. Self-organising data mining: extracting knowledge from data [M]. Berlin: Libri Books, 2000:45.
- [35] 贺昌政.自组织数据挖掘与经济预测[M].北京:科学出版社,2005:127.
- [36] 王磊. HHT与AC算法在金融数据分析中的应用研究[D].杭州:浙江大学,2008.
- [37] Broock W A, Scheinkman J A, Dechert W D, et al. A test for independence based on the correlation dimension[J]. Econometric Reviews, 1996,15(3):197-235.
- [38] Seymour M. Reliability prediction methods an overview[J]. The Journal of RAC, 1999(3):8-11.



第一作者简介:徐廷学,男,1962年3月生,博士、教授。主要研究方向为装备综合保障。曾发表《EM算法在Wiener过程随机参数的超参数估计中的应用》(《系统工程与电子技术》2015年第37卷第3期)等论文。

E-mail: xtx-yt@163.com

(第38卷卷终)