

基于 EMD 和 LS-SVM 的非平稳振动信号趋势预测

王洪波,朱启兵

WANG Hong-bo,ZHU Qi-bing

江南大学 通信与控制工程学院,江苏 无锡 214122

School of Communication and Control Engineering,Jiangnan University,Wuxi,Jiangsu 214122,China

WANG Hong-bo,ZHU Qi-bing.Trend prediction of non-stationary vibration signals based on Empirical Mode Decomposition and Least Square Support Vector Machine.Computer Engineering and Applications,2008,44(16):157-159.

Abstract: The trend forecasting of vibration signals is an important content of condition monitoring and fault diagnosis.The old method of identification of machinery system is not practicable because the non-linear and non-stationary character is becoming more and more prominent.A prediction modelling method based on Empirical Mode Decomposition (EMD) and Least Square Support Vector Machine (LS-SVM) is proposed.Firstly,the trend time series is adaptively decomposed into a series of stationary Intrinsic Mode Functions (IMF) in different scale space using EMD.Then the right parameter and kernel functions are chosen to build different LS-SVM respectively to each and every IMF.Finally,these forecasting results of each IMF are combined to obtain final forecasting result.The simulation results show that the hybrid method has faster speed,higher precision and greater generalization ability than that of the single LS-SVM method.

Key words: trend prediction;non-linear and non-stationary time series;Least Squares Support Vector Machine(LS-SVM);Empirical Mode Decomposition(EMD)

摘要: 振动信号的趋势预测是设备状态监测与故障诊断中的一个重要内容。随着运行设备的非线性、非平稳特点越来越明显,传统的数学建模预报方法已不能满足设备的复杂化和现代化要求。提出了一种基于经验模式分解 EMD(Empirical Mode Decomposition)和最小二乘支持向量机 LS-SVM(Least Square Support Vector Machine)的新模型。首先,运用 EMD 将趋势时间序列自适应地分解成一系列不同尺度的本征模式分量 IMF(intrinsic mode function);其次,对每个本征模式分量,采用合适的核函数和超参数构造不同的 LS-SVM 进行预测;最后对各分量的预测值进行拟合得到最终的预测值。仿真实验表明,此方法与单一的 LS-SVM 预测法相比,具有较高的精度和较强的推广能力。

关键词: 趋势预测;非线性、非平稳时间序列;最小二乘支持向量机(LS-SVM);经验模式分解(EMD)

DOI:10.3778/j.issn.1002-8331.2008.16.048 **文章编号:**1002-8331(2008)16-0157-03 **文献标识码:**A **中图分类号:**TP39

1 前言

趋势分析主要是在振动信号监测的基础上,估计设备故障的传播、发展趋势,对设备的优劣趋势做出预测。技术人员根据预测的结果,合理地安排设备的维修,可以充分发挥设备的工作潜力,并对设备的使用寿命做出估计。因此^[1],趋势预测及寿命估计作为故障诊断中一个重要环节倍受重视,是合理安排生产的主要依据之一。近年来,神经网络方法^[2]成功应用于非平稳时间序列的预测中,但是神经网络存在着网络结构难以确定、容易陷入局部极小值、收敛速度慢、推广能力较差等缺点。与神经网络相比,支持向量机结构简单、推广能力强,对于解决非线性以及高维模式识别问题表现出了很大的优越性,文献[3,4]将 SVM 应用于机械设备状态的趋势预测中。但这些方法都是在时间序列是平稳信号的假设下建模的,并不能有效应用与机械设备的状态评估和预测。经验模式分 EMD(Empirical Mode

Decomposition)是一种处理非线性、非平稳信号的新方法,它将非平稳信号按不同尺度的波动或趋势逐级分解成若干个本征模式分量(IMF),对信号作了平稳化处理,减少了信号间的特征信息的干涉或耦合。

本文尝试将 EMD 和最小二乘支持向量机 LS-SVM(Least Square-Support Vector Machine)相结合,对非平稳时间序列进行预测。首先运用 EMD 将非平稳的时间序列分解成具有不同特征尺度的本征式分量 IMF(Intrinsic Mode Function),然后根据 IMF 的变化规律,利用不同的 LS-SVM 建立不同的 IMF 分量预测模型。最后,拟合所有分量的预测值组合得到最终的预测值。

2 经验模式分解

经验模态分解方法 EMD 和相应的 Hilbert 变换是由Huang

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.60674026)。

作者简介:王洪波(1982-),男,硕士生,主要研究方向:现代信号处理及应用;朱启兵(1973-),男,博士,副教授,主要研究方向:小波、神经网络、故障诊断及容错控制。

收稿日期:2007-09-11 **修回日期:**2007-11-26

等提出的一种适用于分析和处理非线性、非平稳随机信号的新方法^[5]。该方法从本质上讲是对一个信号进行平稳化处理,其结果是将信号中不同尺度的波动或趋势逐级分解开来,产生一系列具有不同特征尺度的数据序列,每一个序列代表一个本征模式函数 IMF(Intrinsic Mode Function)。Huang 认为,时间序列通过经验模式分解最终可得到:

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

式中 c_i 就是从原始信号中获得的基本模式分量, m 为趋势项。即原始信号被分解为 n 个基本模式分量和一个趋势项。 $c_1 \rightarrow c_n$ 就是原数据序列经分离后得到的不同尺度信息,包含了时间序列从高频到低频的不同频率成分,每个本征模式分量均是平稳时间序列, c_1 中应包含原始信号中最短的周期分量,即信号尺度最小的部分; r_n 中包含信号中较大尺度的部分,是一个非震荡的单调序列,是一个相对缓变的信息,最终可能是一个单调函数,也可能是一个恒量。

3 LS-SVM 回归估计原理

LS-SVM 是标准 SVM 的一种扩展,定义了与标准 SVM 不同的损失函数,并将其不等式约束改为等式约束^[9]。给定 l 个样本数据,其值表示为: (x_k, y_k) , 式中 $x_k \in R^n$ 的 n 维向量, $y_k \in R$ 为相对应的输出变量。

优化问题为:

$$\min \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{i=1}^l \xi_i^2,$$

$$\text{st. } y_i = \omega^T \varphi(x_i + b + \xi_i), i=1, 2, \dots, l)$$

正实数 γ 是调节常数,它能够在训练误差和模型复杂度之间取一个折衷以便使所求的函数具有较好的泛化能力,并且 γ 值越大,模型的回归误差越小。

引入 Lagrangian 函数:

$$L(\omega, b, \xi, a) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l a_i [\omega^T \varphi(x_i) + b + \xi_i - y_i]$$

式中: $a_i (i=1, 2, \dots, l)$ 是 Lagrangian 乘子。最优的 a 和 b 可以通过 KKT 条件获得:

$$\frac{\partial L}{\partial \omega} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial \xi} = 0, \frac{\partial L}{\partial a} = 0$$

可得:

$$\omega = \sum_{i=1}^l a_i \varphi(x_i), \sum_{i=1}^l a_i = 0, a_i = \gamma \xi_i$$

$$\omega^T \varphi(x_i) + b + \xi_i - y_i = 0$$

定义核函数 $K(x, x_i) = \psi(x) \psi(x_i)$ 代替非线性映射,根据上式可得求解的优化问题转化为求解线性方程:

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & K(x_1, x_1) + \frac{1}{\gamma} & \dots & K(x_1, x_l) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_l, x_1) & \dots & K(x_l, x_l) + \frac{1}{\lambda} \end{bmatrix} \begin{bmatrix} b \\ a_1 \\ \vdots \\ a_l \end{bmatrix} = \begin{bmatrix} 1 \\ y_1 \\ \vdots \\ y_l \end{bmatrix}$$

最后得最小二乘支持向量机的估计函数为:

$$f(x) = \sum_{i=1}^l a_i K(x, x_i) + b$$

4 时间序列的构造

设时间序列 $\{a(1), a(2), \dots, a(N)\}$, 时间序列预测即是根据时间序列的历史数据 $\{a(t), a(t-1), \dots, a(t-m+1)\}$ 预测未来 $t+k$ 时刻的值 $a(t+k)$, 即寻找 $a(t+k)$ 与历史数据 $\{a(t), a(t-1), \dots, a(t-m+1)\}$ 之间的关系。当 $k=1$ 时,称为一步预测;当 $k>1$ 时,称为多步预测。参数 m 称为嵌入维数,选取尚无严格意义上的理论依据,本文通过对比方法得到,并且考虑一步预测。假设时间序列 $\{a(1), a(2), \dots, a(N)\}$ 的前 n 个数据用于模型的建立,其余 $N-n$ 个数据用于验证模型的推广性能。形成下面的矩阵作为预测模型的输入输出:

$$(x_i, y_i) (i=1, 2, \dots, n)$$

$$X = \begin{bmatrix} a(1) & a(2) & \dots & a(m) \\ a(2) & a(3) & \dots & a(m+1) \\ \vdots & \vdots & \ddots & \vdots \\ a(n-m) & a(n-m+1) & \dots & a(n-1) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-m} \end{bmatrix}$$

$$Y = \begin{bmatrix} a(m+1) \\ a(m+2) \\ \vdots \\ a(n) \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-m} \end{bmatrix}$$

即找出输入 X 和输出 Y 之间的函数 f , 使得 $y_i = f(x_i)$, 其中 $(i=1, 2, \dots, n)$ 。这样将时间序列建模与预测问题转换成样本数据 (x_i, y_i) 的函数回归估计。

5 建模与应用实例仿真

建立预测模型的过程见图 1。

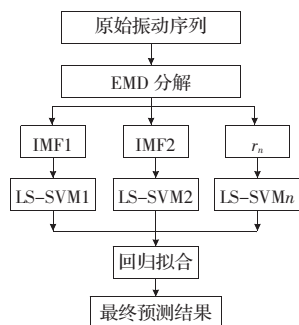


图1 EMD 和 LS-SVM 混合预测模型

某化工厂空气压缩机连续几个月的轴振动测量数据趋势序列见图 2, 共 500 个数据,可以看出此序列有明显的是非平稳特性。

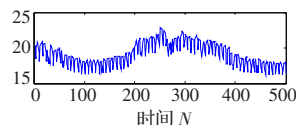


图2 原始趋势序列

取前 400 个数据进行建模(将预测值与后面的 100 个真实数据进行比较), 预测误差采用 (Pe) 相对误差和平均绝对百分误差 $(MA Pe)$:

$$Pe = \left| \frac{r_i - f_i}{r_i} \right| \times 100\%$$

$$MA Pe = \frac{1}{N} \sum_{i=1}^N \left| \frac{r_i - f_i}{r_i} \right|$$

其中, r_i 为真实值, f_i 为预测值, N 为预测值的个数。经过 EMD

分解后的到 5 个本征模式分量和 1 个剩余分量,如图 3 所示。

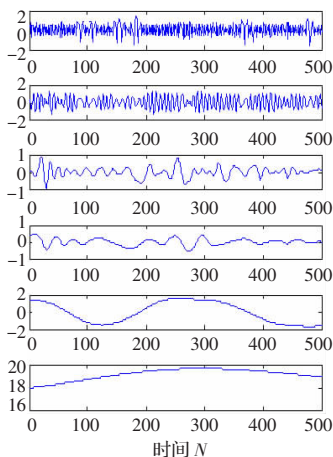


图 3 原始序列的 EMD 分量

本文使用 LS-SVM 软件包对两种方法分别进行预测并进行比较:一是基于 EMD 和 LS-SVM 的预测方法;二是不使用 EMD 直接使用 LS-SVM 的预测方法。核函数的选取和参数的选取是非常重要的,经过反复试验比较,对基本模式分量和趋势项共同选取 RBF 核函数 $k(x, y) = \exp(-\|x - y\|^2 / \sigma^2)$ 的效果更佳。需要确定的参数有核函数 σ^2 和调节常数 γ ,利用交叉验证的网格搜索法^[7],分别找到对于模式分量和趋势项算法性能最好的参数,然后对各个分量和趋势项进行预测,最后利用线性回归模型对各预测分量组合即得到最终预测值,预测相对误差如图 4,点线是 LS-SVM 的预测误差曲线,平均绝对百分误差(MAPE)为 1.44;直线是 EMD 和 LS-SVM 的预测误差曲线,平均绝对百分误差(MAPE)为 0.89。

对比分析两种预测方法的相对误差,新方法的预测精度高,而且波动性最小。这是因为 EMD 将非平稳的原始序列分解为一系列平稳的具有一定规律的单一分量,跟原始序列相比,这些分量更易预测;其次,LS-SVM 不是以经验风险最小化为原则,而是建立在结构风险最小化原理基础上,通过调整常数,使误差尽可能小的同时,使回归函数尽可能平滑,故它具有更强的推广能力。

(上接 153 页)

4.4 实验分析

从表 4 中可以看出,ZW、ZZ、DE、PO、DZ<de>、DZ 这 6 种结构 PP 较低,而 SB<bu>、DE、JY、AD、ZW 这 5 种结构 PR 较低。表 5 中,词与词的匹配较理想,问题主要在词与短语、短语与短语的匹配。当然,由于下层的匹配错误会传递给上层结构(如表 4 中的 ZW 多为上层结构,表 5 中的短语与短语也一般处于上层),因此上层的匹配错误未必完全和自身相关。

5 结语

本文实现的这种基于 GFM 的句法分析算法只是一个初步成果,实验结果并不十分理想,尚有不少问题亟待解决。笔者认为,进一步提高分析的正确率和召回率主要有以下 4 点工作要做:(1)为消歧补充必要的词义知识和词语搭配知识,避免目前消歧中只关注语法功能概率而不顾语义的偏颇;(2)区分出同

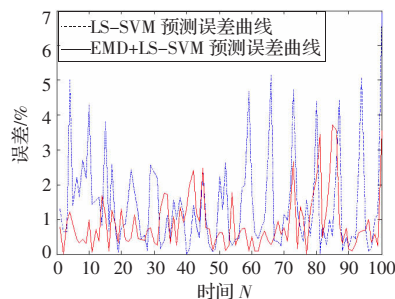


图 4 预测相对误差

6 结束语

将振动非平稳序列用 EMD 平稳化处理,能得到一系列的不同频率的平稳分量,这些分量更能准确地反映原序列的特征信息。根据这些分量选择不同的 LS-SVM 参数,最后通过回归预测得到最终预测结果。仿真试验结果表明,此方法精度高,是一种行之有效的预测方法。

参考文献:

- [1] 杨叔子,吴雅,王治藩,等.时间序列分析的工程应用[M].武汉:华中理工大学出版社,1991.
- [2] 裴树毅.基于神经网络的预测理论及其应用[D].西安:西安交通大学机械工程学院,1994.
- [3] 杨俊燕,张伏云,赵荣珍.支持向量机在机械设备振动信号趋势预测中的应用[J].西安交通大学学报,2005(9).
- [4] Thissen U,Brakel R V,Weijer A P D,et al.Using support vector machines for time series prediction[J].Chemometrics and Intelligent Laboratory Systems,2003,69(12):35-49.
- [5] 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].Proc R Soc London A,1998,454.
- [6] Suykens J A K,Vandewalle J.Least Squares Support Vector Machines[J].Neural Processing Letters,1999,9(3):293-300.
- [7] Momma M,Bennett K P.A pattern search method for model selection of support vector regression[C]//Proceedings of the SIAM International Conference on Data Mining.Philadelphia:SIAM,2002:261-274.

形词与多义词的每个词形的不同语法功能集;(3)解决数据稀疏问题,包括非词典词的语法功能集和词典词未统计到的语法功能;(4)深入调查短语语法功能的分布特点和匹配规律。

参考文献:

- [1] 陈小荷.从自动句法分析角度看汉语词类问题[J].语言教学与研究,1999(3):63-72.
- [2] 徐艳华.现代汉语实词语法功能考察及词类体系重构[D].南京师范大学,2006.
- [3] 周强.汉语句法树库标注体系[J].中文信息学报,2004,18(3):1-8.
- [4] 孟遥.四种基本统计句法分析模型在汉语句法分析中的性能比较[J].中文信息学报,2003,17(3):1-8.
- [5] 詹卫东.面向中文信息处理的现代汉语短语结构规则研究[D].北京大学,1999.
- [6] 陈小荷.现代汉语自动分析——Visual C++实现[M].北京:北京语言文化大学出版社,2000.