

基于 EMD 和组合模型的太阳黑子时间序列预测

曾庆山, 张晓楠

(郑州大学 电气工程学院, 河南 郑州 450001)

摘要: 太阳黑子是非线性、非平稳、多尺度变化的时间序列, 且观测结果大多存在噪声的干扰. 针对太阳黑子时间序列预测的复杂性, 首先将原始数据序列通过小波去噪进行预处理, 然后将去噪后的信号通过 EMD 分解产生若干个从高频到低频的 IMF 分量和余项. 针对低频分量变化缓慢和高频分量波动性较大的特点, 分别采用 RBF 神经网络模型和 SVM 模型进行预测, 最后将各个分量的预测结果相叠加得到最终预测值. 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; EMD 分解; 组合模型; 预测

中图分类号: TP391

文献标志码: A

doi:10.3969/j.issn.1671-6833.2014.03.019

0 引言

太阳黑子是表示太阳活动强弱的一项重要指标, 它是典型的复杂时间序列, 地磁变化、大气运动、气候异常、海洋活动等都和太阳黑子数的变化有着不同程度的关系. 对太阳黑子活动进行有效的预测以此来分析地球环境的变化有着十分重要的价值. 赵海娟等^[1]采用 RBF 神经网络来预测太阳黑子数平滑月均值, 但是预测误差会随着预报时效的延长逐渐放大; 姚淑萍等^[2]将小波分解与重构引入到太阳黑子月平均值这一典型的混沌时间序列预测中, 但小波分解存在选择小波基的困难且不具备自适应的信号分解特性; 张海望^[3]首先将太阳黑子年平均值序列进行小波去噪, 然后用 BP 神经网络进行训练预测, 但存在预测模型单一的问题.

基于此, 笔者提出基于 EMD 的太阳黑子时间序列组合预测模型方法, 将 EMD 分解后的高频分量和低频分量采用不同的预测方法进行预测以此来提高预测的整体精度.

1 EMD 分解

Hilbert-Huang 变换是 N.E.Huang 等人于 1998 年提出的新型的、具有自适应性的时频分析方法, 特别适用于非线性、非平稳信号的分析处理, 它由

经验模态分解 (EMD) 和 Hilbert 变换两部分组成^[4].

EMD 分解是把复杂的信号分解成有限个内禀模态函数 (IMF) 之和. 每个 IMF 分量必须满足以下两个条件: ①极值数和过零点的个数必须相等或最多只能相差一个; ②在任意时刻, 由局部极大值点和局部极小值点分别形成的上、下包络线的平均值为零, 即上包络线和下包络线相对于时间轴局部对称.

对任意一个信号 $X(t)$ 的 EMD 分解的过程如下:

(1) 确定信号所有局部极大值点和极小值点, 采用三次样条曲线形成上、下包络线, 并计算上、下包络线的平均值记为 m_1 .

$$h_1(t) = X(t) - m_1(t). \quad (1)$$

(2) 如果 $h_1(t)$ 满足 IMF 的条件, 那么 $h_1(t)$ 就是 $X(t)$ 的第一个 IMF 分量, 如果 $h_1(t)$ 不满足 IMF 的条件还需对 $h_1(t)$ 重复上述处理过程. 经过 k 次筛分后将产生第 1 个 IMF 分量 $c_1(t)$, 即

$$c_1(t) = h_{1k}(t) = h_{1(k-1)} - m_{1k}(t). \quad (2)$$

(3) 将 $c_1(t)$ 从 $X(t)$ 中分离出来, 得到

$$r_1(t) = X(t) - c_1(t). \quad (3)$$

将 $r_1(t)$ 进行上述筛分处理可以得到第 2 个 IMF 分量 $c_2(t)$, 如此重复循环 n 次最终得到 $X(t)$ 的 n 个 IMF 分量和残余分量 $r_n(t)$, 即

收稿日期: 2013-11-23; 修订日期: 2013-12-27

基金项目: 河南省科技攻关计划项目 (122102210102)

作者简介: 曾庆山 (1963-), 男, 湖北武汉人, 郑州大学教授, 博士, 主要研究方向为复杂系统建模、图像信号处理, E-mail: qszeng@126.com.

$$X(t) = \sum_{j=1}^n c_j(t) + r_n(t). \tag{4}$$

EMD 分解能够将信号按照从高频到低频的顺序依次分解出来,无需选择基函数,避免了小波分解需要选择基函数的困扰,是一种真正的自适应分析方法.

2 预测模型

2.1 RBF 神经网络模型

RBF 神经网络是一种性能良好的前向神经网络,它训练简洁、学习收敛速度快,并且可以避免局部最优. RBF 神经网络模型结构如图 1 所示,它由输入层、隐含层和输出层三部分组成. 这里由前 7 年的太阳黑子数据预测下一年的数据,共 206 组样本,其中输入样本数据 $X_i = [x(t)x(t + 1) \cdots x(t + 6)]$,输出样本数据 $Y_i = x(t + 7), t = 1, 2, \cdots, 206$. 合理选择扩展系数的大小很重要,这里采用逐一筛选的方法来确定使训练样本误差达到最小的扩展系数值.

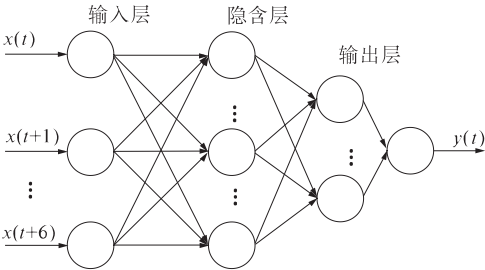


图 1 RBF 神经网络结构图

Fig.1 The structure of RBF neural network

2.2 SVM 模型

SVM 是一种基于统计学习理论的新型机器学习方法,该方法实现了结构风险最小化并且具有较好的泛化性能,较好地解决了小样本、过学习、高维数、局部最小等问题,目前已成功应用于模式识别、函数逼近、时间序列预测等领域^[5].

设数据集 $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)\}$, x_i 为输入值, y_i 为对应的预测值, $i = 1, 2, \cdots, l$. SVM 算法的基本思想是利用一个非线性映射 ϕ , 把数据映射到一个高维的特征空间里,免去了在高维空间计算复杂的点积运算^[6],即

$$f(x) = \omega \times \phi(x) + b. \tag{5}$$

式中: ω 是自回归系数,即权重矢量; b 是误差值.

通过最小化正则风险函数获得 ω 和 b 值.

$$R(C) = C \frac{1}{n} \sum_{i=1}^n L_{\varepsilon}(y) + \frac{1}{2} \|\omega\|^2. \tag{6}$$

$$L_{\varepsilon}(y) = \begin{cases} |f(x) - y| - \varepsilon, & |f(x) - y| \geq \varepsilon; \\ 0, & |f(x) - y| \leq \varepsilon. \end{cases} \tag{7}$$

式中: $\|\omega\|^2$ 为结构风险; ε 为不敏感损失函数; C 为惩罚参数.

引入非负松弛变量 ξ 和 ξ^* , 则式(5) 的最优化问题变为

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*). \tag{8}$$

约束条件为

$$\begin{cases} y_i - \omega\phi(x) - b \leq \varepsilon + \xi_i; \\ \omega\phi(x) + b - y_i \leq \varepsilon + \xi_i^*; \\ \xi_i, \xi_i^* \geq 0. \end{cases} \tag{9}$$

引入拉格朗日乘子 α 和 α^* , 将式(8) 优化问题转换为对偶问题:

$$\min \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j). \tag{10}$$

其中, $K(x_i, x_j)$ 为核函数. 常用的核函数有线性核、多项式核、径向基(RBF) 核等.

径向基核函数的形式为

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2). \tag{11}$$

这里利用交叉验证法选择惩罚参数 C 和核函数参数 σ 的值; 损失函数 ε 的值设置为 0.01; 核函数选择回归泛化能力强、处理非线性序列效果较好的径向基核函数.

3 太阳黑子年平均数序列预测

太阳黑子年平均数据来自 Solar Influences Data Analysis Center(SIDC) 的官方网站, 选择 1800 ~ 1949 年的太阳黑子数据作为训练样本的输入, 1950 ~ 2012 年的数据作为测试样本的输入. 预测步骤如下:

- (1) 将原始数据序列通过小波去噪进行预处理;
- (2) 将去噪后的信号通过 EMD 分解产生若干个从高频到低频的 IMF 分量和余项;
- (3) 对低频分量和余项采用 RBF 神经网络进行预测;
- (4) 对频率值较大、频率变化波动性较大的分量定为高频分量,采用 SVM 进行预测;
- (5) 将各分量的预测结果相叠加得到最终预测值.

3.1 数据预处理

由于数据在采集过程中不可避免的会存在误差和噪声,为了减少噪声干扰,提高 EMD 分解的

准确性,这里采用小波消噪的方法对原始数据进行预处理^[7]. 经过多次对比试验,小波基函数选取 sym8 小波,分解层数为 5 层,选择软阈值函数去噪. 原始信号和去噪后的信号如图 2 和图 3 所示. 从图 2,3 可以看出:原始数据序列在细节处有锯齿一样的噪声出现,经过小波去噪后曲线光滑了许多,说明噪声得到了有效的去除.

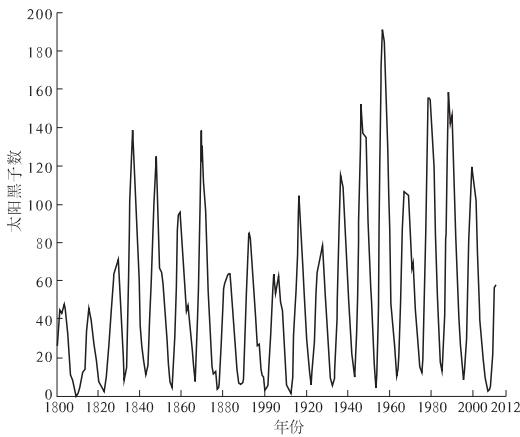


图 2 原始数据序列
Fig. 2 The original data sequence

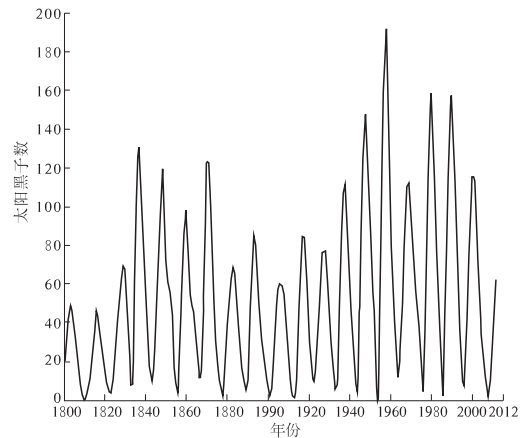


图 3 去噪后数据序列
Fig. 3 The denoised signal

3.2 数据预测与分析

对去噪后的信号进行 EMD 分解,得到 4 个 IMF 分量和一个余项 (RES), 如图 4 所示. 可以看到 IMF1、IMF2 频带较宽、突变较大,采用 SVM 模型进行预测, 这里使用 libsvm-mat 工具箱中的 SVMcgForRegress 函数进行参数寻优,获取最佳惩罚函数 C 和核函数参数 σ . IMF3、IMF4 和余项变化比较平稳,直接采用 RBF 模型即可^[8]. 各个分量的预测值与真实值如图 5 所示.

将各个分量的预测结果相叠加得到最终预测值,如图 6 所示. 这里通过平均绝对百分比误差

(MAPE) 和均方根误差 (RMSE) 这两个评价指标来衡量预测的准确性.

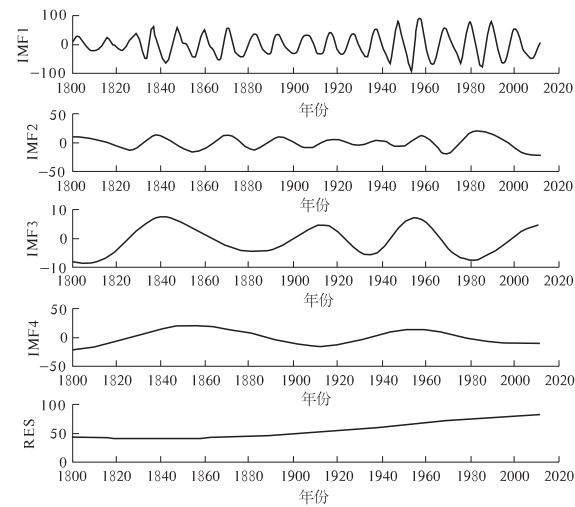


图 4 EMD 分解图
Fig. 4 EMD decomposition results

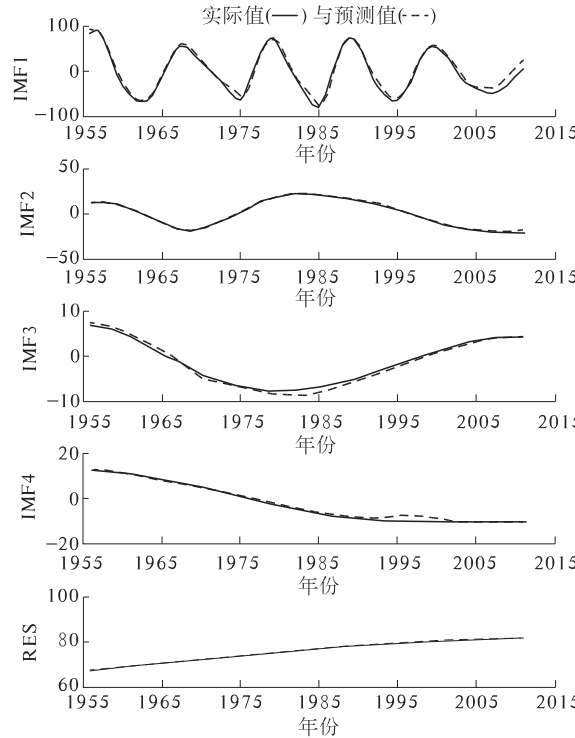


图 5 IMF 分量实际值与预测值
Fig. 5 The actual value and the predicted value of IMFs

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{x(t) - \bar{x}(t)}{x(t)} \right| \times 100\% . \tag{12}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (x(t) - \bar{x}(t))^2} . \tag{13}$$

式中: $x(t)$ 和 $\bar{x}(t)$ 分别为 t 时刻的实际值和预测值.

对同一组太阳黑子数据分别采用 RBF 模型、

SVM 模型、EMD 结合 RBF 模型以及 EMD 结合 SVM 模型分别进行预测. 表 1 中列出了 5 种预测模型方法的 MAPE 和 RMSE. 从对比结果可以看出,基于 EMD 和组合模型的预测方法预测精度要明显高于单一预测方法.

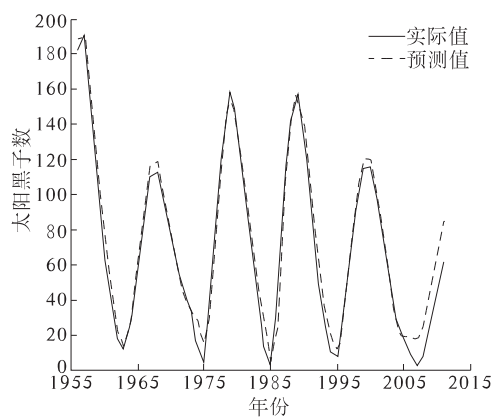


图 6 最终预测值
Fig. 6 Final prediction results

表 1 5 种模型方法预测误差对比

Tab. 1 Error contrast of five forecasting methods		
模型方法	MAPE/%	RMSE
RBF	23.62	0.539 4
SVM	19.28	0.532 1
EMD + RBF	13.56	0.213 7
EMD + SVM	10.37	0.176 5
EMD + SVM + RBF	4.02	0.087 7

4 结论

笔者将 EMD 方法与组合神经网络模型相结合应用于太阳黑子时间序列预测中,优势互补进

行建模. 首先,利用 EMD 自适应分解的特性将时间序列中不同尺度的波动或趋势分离开来,然后,针对高频分量采用 SVM 预测模型进行预测,低频分量采用 RBF 预测模型. 通过和单一预测模型进行对比,结果表明组合预测模型可进一步提高预测的整体精度.

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):31-35.

[2] 姚淑萍,胡昌振,郑链. 基于小波分解与重构的混沌时间序列预测[J]. 弹箭与制导学报,2006,26(2):1133-1135.

[3] 张海望. 基于小波消噪与 BP 神经网络的太阳黑子时间序列预测[D]. 哈尔滨:东北师范大学电气学院,2009.

[4] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and nonstationary time series analysis[J]. Proc R Soc Lond,1998(454):903-995.

[5] 陈海英. 基于支持向量机的上证指数预测和分析[J]. 计算机仿真,2013,30(1):297-300.

[6] 王杰,陈锴鹏. 基于决策函数及 PSO 优化的 SVM 预测控制应用研究[J]. 郑州大学学报:工学版,2013,34(2):678-681.

[7] 朱瑜,王海洋. 基于小波去噪和 EMD 的齿轮箱故障诊断研究[J]. 煤矿机械,2012,33(4):279-280.

[8] 白玮莉,刘志刚,彭权威,等. 基于 HHT 和神经网络组合的负荷预测模型研究[J]. 电力系统保护与控制,2009,37(19):31-35.

Sunspots Time-series Prediction Based on EMD and Combination Model

ZENG Qing-shan, ZHANG Xiao-nan

(School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Sunspots are non-linear, non-stationary, multi-scale changes time-series, and the observations were often interfered by noise. According to the complexity of sunspots time-series prediction, first of all, this paper preprocessed the original data through wavelet de-noising method, then the denoised signal was decomposed into several IMF components and remainder by EMD. In view of the characteristics of the low frequency and high frequency components, RBF neural network model and SVM model were used to predict them respectively, the final predicted value would be got by adding each component's result at last. The simulation results show that this model has higher prediction accuracy.

Key words: sunspot; combination model; EMD decomposition; forecasting