第38卷 第24期
Vol.38 No.24

计 算 机 工 程
Computer Engineering

2012年12月
December 2012

• 人工智能及识别技术 • 文章编号：1000—3428(2012)24—0152—04 文献标识码：A 中图分类号：TP391

基于集合经验模式分解的火灾时间序列预测

张 烨[1]，田 雯[1]，刘盛鹏[2]

(1. 南昌大学电子信息工程系，南昌 330031；2. 公安部上海消防研究所，上海 200438)

摘 要：采用集合经验模式分解(EEMD)和多变量相空间重构技术，结合非线性支持向量回归(SVR)模型，提出一种火灾次数时间序列组合预测方法。根据EEMD将非平稳的火灾时间序列分解为一系列不同尺度的固有模态分量，利用多变量相空间重构技术对分解的各个分量进行相空间重构，构建其训练数据，对重构的训练数据建立各分量的非线性支持向量回归预测模型，使用SVR集成预测方法对火灾时间序列进行预测。仿真结果表明，与单变量相空间重构方法以及SVR方法相比，该方法具有较高的预测精度。

关键词：火灾时间序列；集合经验模式分解；相空间重构；支持向量回归；非平稳

Fire Time Series Forecasting Based on Ensemble Empirical Mode Decomposition

ZHANG Ye[1], TIAN Wen[1], LIU Sheng-peng[2]

(1. Department of Electronic Information Engineering, Nanchang University, Nanchang 330031, China;
2. Shanghai Fire Research Institute, Ministry of Public Security, Shanghai 200438, China)

【Abstract】Based on a combination of Ensemble Empirical Mode Decomposition(EEMD) and multivariate phase space reconstruction, a new combined forecasting model is proposed for fire time series by using Support Vector Regression(SVR). The fire time series is decomposed into a series of Intrinsic Mode Function(IMF) in different scale space by using EEMD. The phase space of IMF is reconstructed by using of multivariate phase-space reconstruction. Based on nonlinear SVR, a prediction model is developed for each intrinsic mode functions, and these forecasting results of each IMF are combined with SVR again to obtain final forecasting result. Experimental results show that this method is more accurate than single variable phase space reconstruction method and SVR method.

【Key words】fire time series; Ensemble Empirical Mode Decomposition(EEMD); phase space reconstruction; Support Vector Regression(SVR); non-stationary

DOI: 10.3969/j.issn.1000-3428.2012.24.036

1 概述

火灾发生的次数与许多时变因素有关，如气候、人口密度、经济发展水平等，尽管某一次火灾的发生具有随机性，但对一个区域、一段时间内火灾的发生具有一定的规律性。因此，可以通过对火灾起数时间序列的分析，建立火灾起数预测模型。目前，对于火灾时间序列的预测主要有2种方法：

(1)采用灰度理论分时段对火灾发生次数进行预测，以及对火灾发生趋势进行分析预测[1-2]。灰色理论计算简单，且所需样本数较少，但灰色预测算法只能对系统作出简单的平滑模拟，并不能反映系统中的“突变”现象，如果数据序列随机性较大，该方法的预测误差较大[3]。

(2)基于人工智能方法，如采用人工神经网络和支持向量机回归，该方法有较强的学习和逼近能力，通过反复的迭代和学习可用来预测复杂的非线性系统的输出[4]。但是，采用人工神经网络方法需要足够多的训练数据来建立预测模型，而火灾数据一般不多[5]。尽管采用支持向量回归模型所需要的训练数据不多，但是必须选择合适的维数才能得到较高好的预测效果[6]。

为解决上述问题，本文提出一种避免人为设定嵌入维数，适合具有非线性、非平稳特点的火灾时间序列预测方法。首先采用集合经验模式分解(Ensemble Empirical Mode Decomposition, EEMD)，将非平稳火灾时间序列进行不同尺度的分解，减少非平稳特性对预测的影响[7]。针对不同尺度的分量，采用多变量相空间重构技术来估计嵌入维数，重建系统的高维相空间，恢复并刻画出火灾时间序列内在变化规律，构建训练数据。利用所构建的训练数据，采用非线性回归预测模型建立火灾次数的组合预测模型。

基金项目：国家自然科学基金资助项目(61162014, 61141007)；公安部应用创新计划基金资助项目(2009YYCXSHXF148)

作者简介：张 烨(1965－)，男，副教授、博士，主研方向：信号处理，智能信息系统；田 雯，硕士研究生；刘盛鹏，助理研究员、博士

收稿日期：2012-02-27 **修回日期：**2012-04-30 **E-mail：**zhye901@126.com

对我国1950年-2008年火灾数据，采用上述方法建立火灾数的预测模型进行仿真。

2 预测模型

对火灾起数的预测一般是采用非线性回归模型[8]，即利用过去若干年火灾次数来预测未来的次数，其数学模型如下：

$$\hat{x}_{t+p}=f(x_t,x_{t-1},\cdots,x_{t-m+1}) \tag{1}$$

其中，x_t为第t年火灾的次数；m是模型的阶，也称为嵌入维数；当$p=1$时，称为一步预测；当$p>1$时，称为多步预测，本文考虑一步预测；$f(\cdot)$为非线性函数。

上述模型实际上是利用时间延迟法，重构火灾次数时间序列的相空间或状态空间。也就是将观测到的一维时间序列重构到高维空间，通过重构得到多维观测变量，向量$\boldsymbol{X}_t=[x_t,x_{t-1},\cdots,x_{t-m+1}]^{\mathrm{T}}$。信号嵌入维数$m$是相空间重构和非线性回归模型的主要参数，通过对火灾起数序列非线性动力学研究，通常可采用非线性动力学方法来估计参数m。由于原始火灾时间序列呈现出非平稳的变化特点，并不适合直接用来作为预测模型的输入，因此为了减小时间序列的非平稳性对模型的影响，采用集合经验模式分解将非平稳时间序列化为若干个近似于平稳时间序列进行处理，然后，利用多变量相空间来估计嵌入维数，将重构数据作为训练数据来训练非线性支持向量回归模型，以构建预测模型。

2.1 时间序列的平稳化处理

为了减小火灾时间序列非平稳特性对预测精度的影响，由于火灾时间序列数据的非平稳性，因此将其直接作为训练数据用来训练预测模型会增加预测难度，预测精度也不高。为此，先采用EEMD将原始火灾时间序列分解为一系列不同尺度的固有模态，以减少非平稳特性对预测的影响。

经验模式分解(Empirical Mode Decomposition, EMD)是将信号中真实存在的不同尺度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固有模式分量(Intrinsic Mode Function, IMF)[9]。分解的IMF符合2个条件：(1)局部最大值和局部最小值定义的包络均值必须为0。(2)极值点的数量和过零点的数量必须相等，或最多相差不多于一个。其分解方法如下：

(1)已知信号$y(t)$，寻找$y(t)$的所有局部极大值点和所有极小值点。

(2)拟合极值点求得到上包络线$y_{\min i}(t)$和下包络线$y_{\max i}(t)$。

(3)根据上下包络线，计算出$y(t)$的局部均值：

$$m_{11}(t)=\frac{[y_{\min i}(t)+y_{\max i}(t)]}{2}$$

以及$y(t)$和$m_{11}(t)$的差为$h_{11}(t)$。

(4)将$h_{11}(t)$看作新的$y(t)$，重复以上3步，直到$h_{11}(t)$满足IMF的条件，$h_{11}(t)$就是第1个从$y(t)$中获得的IMF_1，记$c_1=h_{11}(t)$。

(5)将$r_1(t)=y(t)-h_{11}(t)$看作新的$y(t)$，重复上述步骤，即可依次得到$IMF_2,IMF_3,\cdots$，直到剩余分量r_n满足给定的终止条件时筛选结束。

经验模式分解的最终结果可表示为：

$$y(t)=\sum_{i=1}^{n}c_i(t)+r_n(t) \tag{2}$$

其中，$c_i(t)$就是从原始信号中获得的基本模式分量，包含了时间序列从高频到低频的不同频率成分；$r_n(t)$剩余分量，反映了原始信号的变化趋势。

当观测信号包含噪声时，上述EMD方法会出现模式混叠现象，为了克服模式混叠问题，在EMD的基础上文献[7]提出了EEMD方法。EEMD采用了噪声辅助分析技术，也就是在信号中加入高斯白噪声，使信号和噪声组成一个总体。而且由于零均值噪声的特点，加入的次数越多，将这些多次分解结果取平均后，噪声最终将被抵消。总体平均的结果即被当作真实信号。

EEMD实现的具体方法如下[7]：

(1)在目标数据上加入白噪声序列。

(2)将加入白噪声的序列分解为IMF。

(3)每次加入不同的白噪声序列，反复重复步骤(1)和步骤(2)。

(4)把分解得到的各个IMF的均值作为最终的结果。

图1为我国1950年-2008年火灾发生次数。

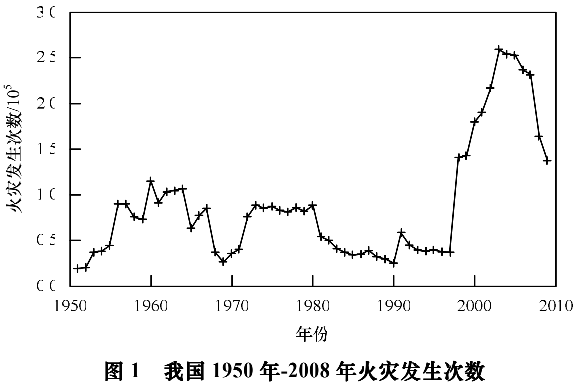

图1 我国1950年-2008年火灾发生次数

对火灾次数时间序列进行EEMD分解，加入零均值，方差为0.25的高斯白噪声，经过分解得到4个固有模式分量和一个剩余分量，如图2所示。

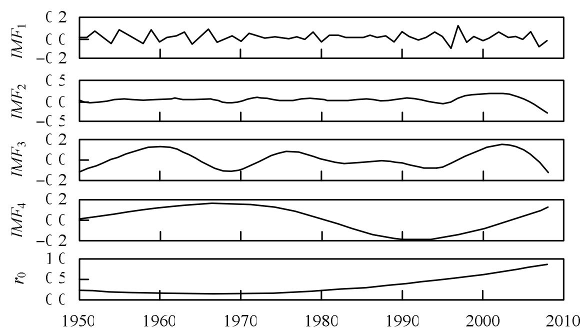

图2 火灾次数的EEMD分解结果

从图 2 可以看出，由上至下分别是 $IMF_1 \sim IMF_4$，依次由高频到低频排列，最后一个是剩余分量 r_0。其中，只有 IMF_1 分量的波动性最大，但相对原始数据其波动范围不大，对预测精度影响相对也小，而其余分量表现出较强的平稳性，对其预测精度也较高。因此，可以通过对集合经验模式分解分解的各个分量的预测来得到火灾起数的预测，由于减小了火灾起数非平稳性的影响，可以提高火灾起数的预测精度。

2.2 多变量时间序列的相空间重构

为了估计各个模式分量，必须将各个分量从一维空间重构到高维空间。由于火灾起数通过集合经验模式分解分解得到多个不同尺度的固有模式分量(C_i)分量和剩余分量 r_n，必须采用多变量相空间重构技术，将各个分量从一维空间重构到高维空间，反映出火灾起数时间序列变化的内在规律，重构过程如下：

假设已知 M 个分量的时间序列向量 $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_M$，其中，$\boldsymbol{X}_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,n}), i = 1,2,\cdots,M$，重构的相空间向量 $\boldsymbol{V}_n$ 为[10]：

$$\begin{aligned} \boldsymbol{V}_n = (&x_{1,n}, x_{1,n-T_1}, \cdots, x_{1,n-(m_1-1)T_1}, x_{2,n}, x_{2,n-T_2}, \cdots, \\ &x_{1,n-(m_2-1)T_2}, \cdots, x_{M,n}, x_{M,n-T_M}, \cdots, x_{1,n-(m_M-1)T_M}) \\ n &= J_0, J_0+1, \cdots, N \\ J_0 &= \max_{1 \leqslant i \leqslant M}(m_i - 1)T_i + 1 \end{aligned} \tag{3}$$

其中，T_i、$m_i (i = 1,2,\cdots,M)$ 分别为时间延迟和嵌入维数，通常采用自相关函数法[11]和 G-P 算法[12]分别确定时间延迟和相空间嵌入维数。

根据嵌入理论，存在一个映射 $F: R^d \to R^d$，F 是 m 维空间上的光滑函数($m = \sum_{i=1}^{M} m_i$)，F 可使得 $\boldsymbol{V}_{n+1} = F(\boldsymbol{V}_n)$ (假设 m 或 m_i 均充分大)，也可写作：

$$x_{i,n+1} = F_i(\boldsymbol{V}_n), i = 1,2,\cdots,M \tag{4}$$

其中，F 与 $F_i (i = 1,2,\cdots,M)$ 均未知，需要先通过拟合逼近其函数形式，然后再用 F 来进行预测，即 $\boldsymbol{V}_{n+1} = F(\boldsymbol{V}_n)$ 或 $\hat{x}_{i,n+1} = F_i(\boldsymbol{V}_n), (i = 1,2,\cdots,M)$，这里，$\boldsymbol{V}_{n+1}, \hat{x}_{i,n+1}$ 为预测值。将嵌入向量 $\boldsymbol{V}_n$ 对应的时间序列下一步的值 $x_{i,n+1}$ 组成训练数据对集 $\left\{(\boldsymbol{V}_n, d_i)\right\}_{i=1}^{M}$ 来训练预测模型，其中，$d_i = x_{i,n+1}$，M 为 IMF 分量个数。

2.3 预测模型的构建

根据上述方法确定的训练数据 $\left\{(\boldsymbol{V}_n, d_i)\right\}_{i=1}^{M}$，对每个模式分量采用非线性支持向量回归来构建各个分量的预测模型，即求出 F 函数。如对第 k 个基本模式分量的预测如下：

给定训练样本 $Y_k = \{(\boldsymbol{V}_n, d_k)\}$，寻找一组 Lagrange 乘子 $\{\alpha_i\}_{i=1}^{T}$ 和 $\{\alpha'_i\}_{i=1}^{T}$ 使其最大化目标函数：

$$\begin{aligned} Q(\alpha_i, \alpha'_i) = &\sum_{i=1}^{T} d_{t_i}(\alpha_i - \alpha'_i) - \varepsilon \sum_{i=1}^{T}(\alpha_i + \alpha'_i) - \\ &\frac{1}{2}\sum_{i=1}^{T}\sum_{j=1}^{T}(\alpha_i - \alpha'_i)(\alpha_j - \alpha'_j)K(\boldsymbol{V}_i, \boldsymbol{V}_j) \end{aligned} \tag{5}$$

满足约束条件：

$$\sum_{i=1}^{T}(\alpha_i - \alpha'_i) = 0,\ 0 \leqslant \alpha_i, \alpha'_j \leqslant C,\ i = 1,2,\cdots,T$$

其中，C、ε 为常数；$K(\boldsymbol{V}_i, \boldsymbol{V}_j)$ 为核函数。获得最优的 α_i 和 α'_i 的值后，假设 d 的估计为 l，则：

$$l = f(\boldsymbol{V}) = \sum_{i=1}^{T}(\alpha_i - \alpha'_i)K(\boldsymbol{V}_i, \boldsymbol{V}_j) \tag{6}$$

核函数 $K(\boldsymbol{V}_i, \boldsymbol{V}_j)$ 要求满足 Mercer 定理，在这个要求内，它的选择有一定的自由度。本文采用径向基函数：

$$K(\boldsymbol{V}_i, \boldsymbol{V}_j) = \exp\left(-\frac{\left\|\boldsymbol{V}_i - \boldsymbol{V}_j\right\|^2}{2\sigma^2}\right),\ i = 1,2,\cdots,T \tag{7}$$

在得到各个分量的预测模型后，将各个分量的预测结果作为集成预测模型的输入。为了得到集成支持向量回归(Support Vector Regression, SVR)预测模型，将各个分量的预测结果与其相对应时刻的火灾起数构成训练数据，采用上述建立分量预测模型的方法构建 SVR 集成预测模型，得到组合预测模型，如图 3 所示。

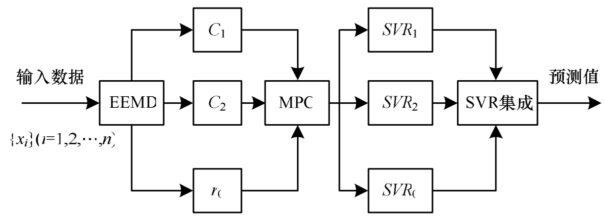

图 3 组合预测模型

在图 3 中，EEMD 为模式分解单元；C_i 为分解得到的第 i 个 IMF；r_0 为剩余分量；MPC 为 M 个分量的多变量相空间重构单元；SVR_i 为第 i 个分量的支持向量回归预测模型单元，采用 SVR 作为预测集成单元。

3 仿真实验

本文采用我国 1950 年-2008 年火灾发生起数作为实验数据，如图 1 所示。在实验中，采用相对误差(P_e)为评估指标，即：

$$P_e = \frac{|x_i - \hat{x}_i|}{x_i} \times 100\% \tag{8}$$

其中，$\hat{x}_i$ 表示预测值；x_i 表示真实值。

首先将火灾起数时间序列进行 EEMD 分解，即加入零均值，方差为 0.25 的高斯白噪声，分解得到 4 个固有模式分量和一个剩余分量，如图 2 所示。利用相空间重构技术估计出各个分量的嵌入维数分别为：2，2，2，2，1，最佳延迟分别为：2，3，5，12，19。将 1950 年-2007 年的火灾起数数据作为训练数据，2008 年的数据作为预测数据，利用重构后的时间序列构建各个预测模型的训练数据。选取径向基函数作为核函数，取 $C = 1000$，$\varepsilon = 0.001$，$\sigma = 0.26$，结果如图 4 所示，其中，2008 年的预测值约为 138 070，实际值为 136 835，相对误差约为 0.902 5%，可见本文提出的预测模型有较高的预测精度。

为进一步验证本文方法的有效性，并与以下 2 种预测方法进行比较：(1)SVR 方法，选取嵌入维数为 5，将火灾起数时间序列重构得到训练数据，利用支持向量回归构建

预测模型，训练时取 $C=1000$，$\sigma=0.5$，$\varepsilon=0.01$。(2)利用相空间重构技术，将一维火灾起数时间序列重构到高维空间得到训练数据，同样利用支持向量回归建立预测模型，在实验中，火灾起数时间序列嵌入维数的估计是 3，时间延迟为 8，训练时取 $C=1000$，$\sigma=0.5$，$\varepsilon=0.01$。

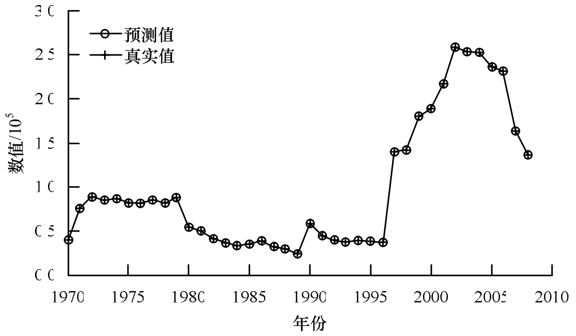

图 4　基于 EEMD 和相空间重构的预测值与真实值

在比较实验中，同样取 1950 年-2007 年的火灾起数数据作为训练数据，将 2008 年的数据作为预测数据，对 SVR 预测模型建模。由于在 EEMD 分解时加入了随机噪声，使得每次 EEMD 分解得到的 IMF 分量也有所不同，估计出的嵌入维数和延迟时间也不同，因此通过训练数据学习得到的 SVR 模型核的个数也不同。所以，为评估预测模型的估计精度，将 2008 年的数据作为预测数据，在条件不变的情况下，将上述 3 种方法重复做 10 次预测，相对误差如表 1 所示。

表 1　3 种方法对火灾时间序列的相对误差　　(%)

实验次数	SVR 方法	单变量相空间重构	本文方法
1	3.041 2	1.786 3	0.902 5
2	3.041 2	1.786 3	1.507 1
3	3.041 2	1.786 3	0.381 3
4	3.041 2	1.786 3	1.483 1
5	3.041 2	1.786 3	0.308 7
6	3.041 2	1.786 3	0.744 8
7	3.041 2	1.786 3	1.159 1
8	3.041 2	1.786 3	0.879 7
9	3.041 2	1.786 3	1.543 4
10	3.041 2	1.786 3	0.601 8
平均误差	3.041 2	1.786 3	0.950 6
标准差	4 161.426 0	2 444.284 0	1 431.381 0

由表 1 可知，本文方法优于单变量相空间重构以及 SVR 方法，提高了预测精度。通过训练 SVR 方法和单变量相空间重构方法得到的 SVR 核的个数分别为 45 和 40，而本文方法的 SVR 模型核个数的变化范围为 43~51。

4　结束语

本文采用基于非线性支持向量回归的组合预测模型，提出一种火灾次数时间序列组合预测方法。采用集合经验模式对火灾次数的原始序列进行分解，得到多个不同尺度的固有模式分量和剩余分量，使用多变量相空间重构，得到火灾次数时间序列预测的训练数据，结合非线性支持向量回归模型，利用得到训练数据，构建火灾时间序列的组合预测模型。仿真结果表明，该方法能减小火灾次数非平稳性对预测精度的影响，具有较高的预测精度。

参考文献

[1] 郑双忠. 基于灰色系统理论的城市火灾预测分析[J]. 数学的实践与认识, 2005, 35(1): 72-76.

[2] 姜学鹏, 徐志胜. 我国火灾起数的灰色拓扑预测[J]. 中国公共安全: 学术版, 2006, (2): 58-61.

[3] 谢正文. 火灾预测的改进 GM(1, 1)模型[J]. 中国计量学院学报, 2007, 18(3): 241-244.

[4] Lean Y, Wang Shouyang, Kinkeung L. Forecasting Crude Oil Price with an EMD-based Neural Network Ensemble Learning Paradigm[J]. Energy Economics, 2008, 30(5): 2623-2635.

[5] Urbina M B, Mendez A F. Time Series Forecasting Through Polynomial Artificial Neural Networks and Genetic Programming[C]//Proc. of IEEE International Joint Conference on Neural Networks. Hong Kong, China: [s. n.], 2008.

[6] Chen Chia-Chuang, Shun Feng-Su, Jin Tsong-Jeng, et al. Robust Support Vector Regression Networks for Function Approximation with Outliers[J]. IEEE Transactions on Neural Networks, 2002, 13(6): 1322-1330.

[7] Wu Zhaohua, Norden E H. Ensemble Empirical Mode Decomposition: A Noise-assisted Data Analysis Method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41.

[8] 刘盛鹏, 张 烨. 基于相空间重构和独立分量分析的火灾预测[J]. 自然灾害学报, 2011, 20(3): 47-50.

[9] Norden E H, Shen Zheng, Steven R L, et al. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis[J]. Proceeding of the Royal Society of London Series: A Mathematical Physical and Engineering Sciences, 1998, 454(1971): 903-905.

[10] 王海燕, 盛昭瀚, 张 进, 等. 多变量时间序列复杂系统的相空间重构[J]. 东南大学学报: 自然科学版, 2003, 33(1): 115-118.

[11] 吕金虎, 陆君安, 陈士华. 混沌时间序列分析及其应用[M]. 武汉: 武汉大学出版社, 2002.

[12] Wang Shuohe, Hao Ruilin, Chang Yujian, et al. Research of Short-term Load Forecasting Based on Combined Grey Neural Network and Phase Space Reconstruction[C]//Proc. of International Conference on Machine Learning and Cybernetics. Baoding, China: [s. n.], 2009.

编辑　刘　冰