

时间序列非线性模型的研究¹

马 竟 王锦山 韦 岗

(华南理工大学电子与通信工程系 广州 510641)

摘 要 本文提出了几种时间序列非线性模型, 对这些模型的结构和训练方法作了分析和讨论, 并针对几类常见的时间序列, 给出了实验结果.

关键词 时间序列分析, 非线性模型, 非线性预测

中图分类号 TN911.7, TN-052

1 引 言

传统的时间序列分析方法, 主要是线性方法^[1]. 但许多时间序列隐含大量的动态特征和具有高度的非线性, 因而其非线性分析有更大理论和实践价值. 随着计算能力的提高, 时间序列的非线性分析已成为可能.

神经网络是一种典型的非线性模型. 利用多层前馈时延神经网络 (TDNN) 来进行时间序列的建模和分析, 已取得了一些有价值的实用成果^[2]. 回归神经网络通过在隐层神经元引入“时延反馈”环节, 能更好地用于时间序列分析. 本文将一种回归神经网络用于时间序列的非线性模型的分析中.

ARMA 模型作为一种经典的线性模型, 已能较好地反映时间序列的线性特征. 为反映时间序列的非线性特征, 人们提出了许多基于 ARMA 模型的非线性建模和预测方法^[3]. 本文在此方面也有所尝试.

模型的训练方法对时间序列建模也至关重要. 本文结合诸多改进算法, 提出一种“合成训练算法”, 提高了学习效率和建模的准确性.

2 时间序列和预测模型

时间序列 $\{x_t\}$ 实质上是一随机变量 X 在某一时间阶段的历程, 即某一序列的个别观测值 x_t 的集合. 其预测模型可以广义地用某种映射 ϕ 表示为

$$\hat{x}_{t+k-i} = \phi(x_t, x_{t-1}, \dots, x_{t-i}), \quad (1)$$

其中 \hat{x}_{t+k-i} 为对 x_{t+k-i} 的预测. 如果构造 ϕ 是线性映射, 那么 (1) 式称为线性时间序列模型; 当构造 ϕ 是非线性映射时, 那么 (1) 式即为非线性模型. 构造 ϕ 实际上是随机变量 X 的函数.

理论上已证明^[4], 任意函数在某一条件下均可表示为某一区间的正交函数之和, 即函数的正交展开或正交变换. 然而, 另一类非正交展开也被广泛使用, 如人们熟知的泰勒级数展开, 若取前几项, 有 $f_n(x) = \sum_{i=0}^n a_i x^i$, 其中函数系 $X = (1, x, x^2, \dots, x^n)$. X 不满足

¹ 1997-04-22 收到, 1998-01-15 定稿

国家自然科学基金 (69602002, 69772027)、国家教委博士点基金 (96056112) 和广东省自然科学基金资助 (960227, 963037) 项目

正交关系, 但确满足线性独立关系. 另外亦可证明, 任一线性独立函数系都可以通过施密茨正交变换, 使之正交化.

神经网络常用的 Sigmoid 激励函数 $f(x) = 1/(1 + e^{-x})$, 即是满足此类条件的非正交函数系. 而三角函数、指数函数也是满足此类条件的正交函数系. 利用这些函数的组合就能近似表示一切函数.

3 时间序列非线性模型

3.1 回归神经网络模型

首先我们引入回归神经网络模型. 设有如图 1 所示的三层网络, 包括输入, 隐层和输出层. 为使系统的动态特性得以体现, 我们在隐层节点处引入反馈环节, 形成一局部自回归神经网络. 假设网络输入为时间序列, 而输出则为预测信号. 这样, 该网络构成一通用时间序列预测模型.

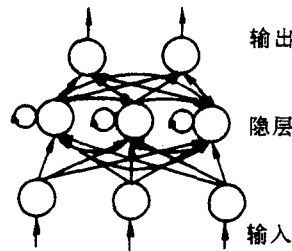


图 1 局部回归神经网络非线性时间序列预测模型

3.2 基于 ARMA 模型的非线性时序模型

传统的线性 ARMA 模型可表示为

$$\hat{x}_n = \sum_{i=1}^p a_i x_{n-i} + \sum_{j=1}^q b_j e_{n-j} = f(x_{n-1}, x_{n-2}, \dots, x_{n-p}, e_{n-1}, e_{n-2}, \dots, e_{n-q}), \quad (2)$$

其中 \hat{x}_n 表示第 n 时刻预测值, x_n 表示该时刻真实值, e_n 表示预测误差, $e_n = x_n - \hat{x}_n$, a, b 为预测系数. 将此模型用于语音信号预测、机械工程信号分析中已取得比较好的效果. 但它在描述生物医学信号等含有较强非线性和动态特征的时间序列时性能则不够好. 一种改进方案就是利用非线性函数系的组合来逼近真实模型. 本文提出四种方案, 以期得到较完备的时序非线性模型.

3.2.1 以三角函数为函数系的非线性 ARMA 模型

$$\begin{aligned} \hat{x}_n = & f(x_{n-1}, x_{n-2}, \dots, x_{n-p}, e_{n-1}, e_{n-2}, \dots, e_{n-q}) + \mu_1 \sin(\omega_1 x_{n-1} + v_1) \\ & + \mu_2 \sin(\omega_2 x_{n-2} + v_2) + \mu_3 \sin(\omega_3 x_{n-1} + \omega_4 x_{n-2} + v_3) \end{aligned} \quad (3)$$

或

$$\begin{aligned} \hat{x}_n = & f(x_{n-1}, x_{n-2}, \dots, x_{n-p}, e_{n-1}, e_{n-2}, \dots, e_{n-q}) \\ & + \mu_1 \sin\left(\sum_{i=1}^{r_1} \omega_{1i} x_{n-i} + v_1\right) + \mu_2 \sin\left(\sum_{i=1}^{r_2} \omega_{2i} x_{n-i} + v_2\right), \end{aligned} \quad (4)$$

这里, 三角函数充当非线性神经元的角色, 并根据三角函数的特性作了上述权值参数的选择.

3.2.2 以指数函数为函数系的非线性 ARMA 模型

$$\hat{x}_n = f(x_{n-1}, x_{n-2}, \dots, x_{n-p}, e_{n-1}, e_{n-2}, \dots, e_{n-q}) + c_1 e^{-d_1 (\sum_{i=1}^{r_1} s_{1i} x_{n-i} - v_1)^2} + c_2 e^{-d_2 (\sum_{i=1}^{r_2} s_{2i} x_{n-i} - v_2)^2} + h \quad (5)$$

或

$$\hat{x}_n = f(x_{n-1}, x_{n-2}, \dots, x_{n-p}, e_{n-1}, e_{n-2}, \dots, e_{n-q}) + b x_{n-1} e^{-a x_{n-1}^2} + d x_{n-2} e^{-c x_{n-2}^2}. \quad (6)$$

指数函数的动态特性比三角函数好, 但模型学习过程不容易稳定, 同时训练稍复杂.

3.2.3 以指数函数和三角函数为函数系的非线性 ARMA 模型

$$\hat{x}_n = f(x_{n-1}, x_{n-2}, \dots, x_{n-p}, e_{n-1}, e_{n-2}, \dots, e_{n-q}) + c e^{-d (\sum_{i=1}^r s_i x_{n-i} - v)^2} + \mu \sin \left(\sum_{k=1}^l \omega_k x_{n-k} + z \right). \quad (7)$$

它结合了三角函数和指数函数的动态特性, 但训练过程仍很复杂.

3.2.4 基于混合多项式的非线性 ARMA 时序模型

$$\hat{x}_n = f(x_{n-1}, x_{n-2}, \dots, x_{n-p}, e_{n-1}, e_{n-2}, \dots, e_{n-q}) + \sum_{k=1}^{r_1} c_k x_{n-1}^{k+2} + \sum_{k=1}^{r_2} d_k e_{n-1}^{k+2}. \quad (8)$$

完备的非线性项应表示为 $\varphi(x_{n-1}^{k+1}, x_{n-2}^{k+1}, \dots, x_{n-p}^{k+1}, e_{n-1}^{l+1}, e_{n-2}^{l+1}, \dots, e_{n-q}^{l+1})$, ($k = 1, 2, \dots, r_1; l = 1, 2, \dots, r_2$), 这样复杂的结构对实际应用不利, 因此本文采用简化了的非线性项.

3.3 基于多元多项式的非线性时序模型

$$\hat{x}_n = \sum_{j=1}^{q-1} c_j \left(\sum_{i=0}^{p-1} d_{ji} x_{n-i-1} \right)^{j+1} + \sum_{j=1}^{q-1} b_j \left(\sum_{i=0}^{p-1} a_{ji} x_{n-i-1} \right)^{j+1} + \sum_{i=0}^{p-1} a_{0i} x_{n-i-1}. \quad (9)$$

这是一种完全的非线性模型, 它适于描述部分混沌序列.

4 模型训练方法

对非线性模型参数辨识问题, 到目前为止, 还未找到一种最佳方案. 传统的方法是最速下降法. 但它有两个突出的问题: 第一, 学习速度慢; 第二, 容易陷于局部极小. 因而人们提出了许多改进算法. 如二阶梯度法^[5]、改进的梯度搜索算法^[6]和对训练数据进行预处理^[7]等. 摆脱局部极小通常是对权系数加入一个扰动, 即模拟退火 (SA) 算法. Wessels 和 Barnard^[8]则认为: 合理选取权系数的初始值, 能减少模型训练时陷入局部最小的可能. 设第

k 层网络节点数是 N , 则连接第 k 层和第 $k+1$ 层权系数初始值应为 $(-3/\sqrt{N+1}, 3/\sqrt{N+1})$ 之间的随机数。

综合上述算法, 本文提出一种“合成训练算法”, 即在最初训练过程中, 采用最速下降法, 当下降速度放慢时, 采用二阶梯度法, 在接近于收敛处, 则采用改变步长和搜索方向的方法。同时用最佳初始权值系数和 SA 算法使搜索过程远离局部最小。

4.1 二阶梯度法用于时间序列非线性模型的学习算法

二阶梯度法主要涉及对二阶梯度的求解。以混合多项式的非线性 ARMA 时序模型为例, 我们来简单推导一下该算法中的二阶梯度。

定义系统误差目标函数为

$$J(w) = \frac{1}{2} \sum_t \sum_k e_k^2(t). \quad (10)$$

设权参数 $w = (a_i, b_j, c_k, d_k)^T$, t 时刻的预测误差为 $e_k(t)$, k 为输出维数。对一维时间序列, $k = 1$, 因此在不失一般性的情况下可假设

$$J(w) = \left(\frac{1}{2N}\right) \sum_{n=1}^N e_n^2. \quad (11)$$

预测误差则为

$$e_n = x_n - \sum_{i=1}^p a_i x_{n-i} - \sum_{j=1}^q b_j x_{n-j} - \sum_{k=1}^{r_1} c_k x_{n-1}^{k+1} - \sum_{k=1}^{r_2} d_k e_{n-1}^{k+1}. \quad (12)$$

w 可按下式迭代

$$w^{(n+1)} = w^{(n)} - \frac{\partial J / \partial w}{\xi + \partial^2 J / \partial w^2}, \quad (13)$$

其中 ξ 为一常数 (其值很小), 以防止二阶偏导数为 0 时, 系统溢出。由 (11) 式可得

$$\frac{\partial J}{\partial w} = \frac{1}{N} \sum_{n=1}^N \frac{\partial e_n}{\partial w}, \quad \frac{\partial^2 J}{\partial w^2} = \frac{1}{N} \sum_{n=1}^N \left[\left(\frac{\partial e_n}{\partial w} \right)^2 + e_n \frac{\partial^2 e_n}{\partial w^2} \right]. \quad (14)$$

因而由 (12) 式分别对各权值系数求二阶偏导, 可得

$$\frac{\partial^2 e_n}{\partial a_i^2} = - \sum_{j=1}^q b_j \frac{\partial^2 e_{n-j}}{\partial a_i^2} - \sum_{k=1}^{r_2} d_k (k+1) \left[k e_{n-1}^k \left(\frac{\partial e_{n-1}}{\partial a_i} \right)^2 + e_{n-1}^k \frac{\partial^2 e_{n-1}}{\partial a_i^2} \right], \quad (15)$$

$$\frac{\partial^2 e_n}{\partial b_j^2} = -2 \frac{\partial e_{n-j}}{\partial b_j} - \sum_{i=1}^q b_i \frac{\partial^2 e_{n-j}}{\partial b_j^2} - \sum_{k=1}^{r_2} d_k (k+1) \left[k e_{n-1}^k \left(\frac{\partial e_{n-1}}{\partial b_j} \right)^2 + e_{n-1}^k \frac{\partial^2 e_{n-1}}{\partial b_j^2} \right], \quad (16)$$

$$\frac{\partial^2 e_n}{\partial c_k^2} = - \sum_{j=1}^q b_j \frac{\partial^2 e_{n-1}}{\partial c_k^2}, \quad (17)$$

$$\begin{aligned} \frac{\partial^2 e_n}{\partial d_k^2} = & - \sum_{j=1}^q b_j \frac{\partial^2 e_{n-j}}{\partial d_k^2} - 2(k+1) e_{n-1}^k \frac{\partial e_{n-1}}{\partial d_k} \\ & - \sum_{i=1}^{r_2} d_i (i+1) \left[i e_{n-1}^{i-1} \left(\frac{\partial e_{n-1}}{\partial d_k} \right)^2 + e_{n-1}^i \frac{\partial^2 e_{n-1}}{\partial d_k^2} \right]. \end{aligned} \quad (18)$$

初始值一般设为 0，由以上公式可递归计算出二阶梯度。

4.2 改变步长和搜索方向的搜索算法

最速下降算法在接近于收敛处，收敛速度变慢，该收敛点也很可能是一个局部极小点。因此，改变搜索方向和步长就是很自然的事情。本文的策略如下：

- 步骤 1 确定起始点权值系数 w_0 和初始步长 λ ；
- 步骤 2 求出目标函数在 w_0 处的梯度向量 ∇w_0 ，取 $S_0 = -\nabla w_0$ ；
- 步骤 3 从 w_0 出发，沿 S_0 前进 λ 达到 w_1 ，求出 ∇w_1 ，取 $S_1 = -\nabla w_1$ ；
- 步骤 4 令 $M = 0$ ，选取 K ，求出 w_1 的目标函数 $J(w_1)$ ，设搜索目标 $TAR = J(w_1)$ ；
- 步骤 5 $M = M + K$ ， $S = S_0 + M \times S$ ，由 W_0 前进 λ 到 w_2 ；
- 步骤 6 如 $J(w_2) < J(w_1)$ ，则 $w_1 = w_2$ ， $K = 2K$ ，返回步骤 5，如 $J(w_2) > J(w_1)$ 且 $|K| < \text{某一门限 } \xi_1$ 或 $|M| > \text{某一门限 } \xi_2$ ，转至步骤 7，否则 $K = -K/3$ ，返回步骤 5；
- 步骤 7 设 $D = (TAR - J(w_1))/J(w_0)$ ，如 $D > \text{某一门限 } \xi_3$ ，则 $S_1 = -\nabla w_1$ ，返回步骤 4，如 $D < \xi_3$ ，则转至步骤 8；
- 步骤 8 如 $J(w_1) < J(w_2)$ ，则 $w_0 = w_1$ ，返回步骤 2，如 $\lambda > \text{某一门限 } \xi_4$ ，则 $\lambda = \lambda/2$ ，否则如 $\lambda < \xi_4$ ，则停止运算。

本算法为一从 w_0 开始起步的递推搜索法。在确定步长 λ 后，无论方向如何，搜索目标函数最终将落在以 w_0 为中心， λ 为半径的一超球面上。最优方向肯定指向目标函数在该超球面上最小点处。由于此超球面为一无边界的有限闭集，当搜索目标函数是一有下界的连续可微函数时，这一最小点必存在且为一极小点。

5 非线性模型应用于时间序列预测

5.1 时间序列的非线性预测

语音信号、生物医学信号等序列被认为含有非线性和时变特征，因而用非线性模型进行处理有一定的实用价值。本文对这类典型的时间序列进行了预测分析。

为分析方便，本文将输入取为信号采样 $x(t-1), x(t-2), \dots, x(t-p)$ ，而实际输出亦取信号采样值 $x(t)$ ，即用时刻 t 以前 p 个时刻的样本值来预测 t 时刻的采样，以预测相对误差作为评判标准。这里的相对误差是指“一帧”信号总的相对误差，即：
$$\epsilon = \frac{\sum_{t=1}^N (\hat{x}(t) - x(t))^2}{\sum_{t=1}^N x^2(t)}$$
，其中 ϵ 为相对误差， $\hat{x}(t)$ 为对真实值 $x(t)$ 的预测值， N 为帧长。

5.2 实验结果分析

本文用多段信号进行了实验。图 2 为一段典型语音波形，用前述几种非线性模型对其进行分析，其结果见图 3(a) - 3(d)。为公平比较，所有模型输入阶数相等，模型权值参数的总数也相等。

图 3(a) 是用回归和前馈神经网络模型分别对该段信号作预测分析得出的结果，图中横坐标表示帧长，纵坐标代表预测误差（以下同）。从图中可看出：回归网络是比前馈网络更好的非线性模型，特别是对辅音信号，它所得预测误差比前馈网络模型的预测误差减小了将近 20%。

本文学习算法与传统梯度下降算法比较结果见图 3(b)，该图中横坐标表示学习次数。所采用的信号为初始语音帧。显然本文算法能保持一定的收敛速度直至寻到最优。

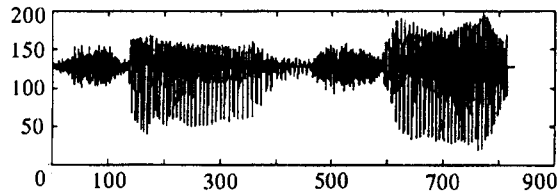


图 2 一段典型语音波形

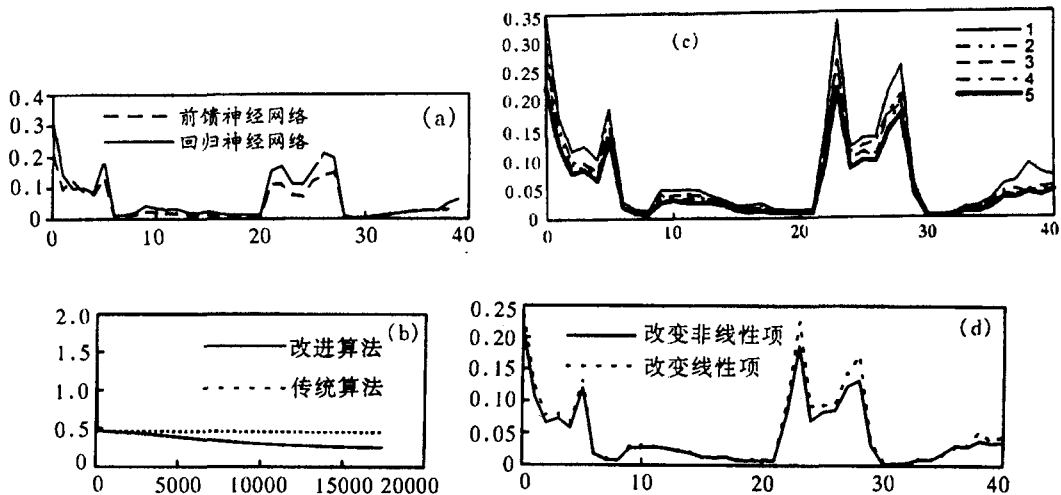


图 3

几种非线性 ARMA 模型对该段信号分析的结果见图 3(c)。图中 1 - 5 各自代表 ARMA 模型和 3.2.1 节到 3.2.4 节所描述的 4 种模型。从实验结果来看, 非线性 ARMA 模型预测效果都比线性 ARMA 模型好, 其中以混合多项式非线性模型结果最好。这在一定程度上表明了语音信号的非线性成分存在的方式为: $\varphi(x_{n-1}^{k+1}, x_{n-2}^{k+1}, \dots, x_{n-p}^{k+1}, e_{n-1}^{l+1}, e_{n-2}^{l+1}, \dots, e_{n-q}^{l+1})$, 但其训练过程太复杂。

在总参数数目一定的条件下, 调整内部权参数所得结果见图 3(d)(以混合多项式 ARMA 模型为例, 分别改变线性和非线性参数的数目)。不难看出, 改变非线性项对预测辅音的影响较明显。

对图 4 心电图信号分析的结果见图 5。我们可以有同样的结论。

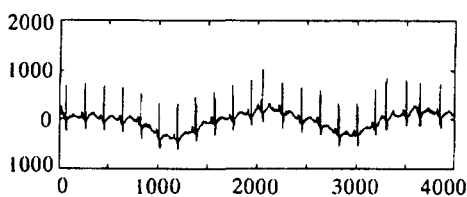


图 4 一段心电图波形

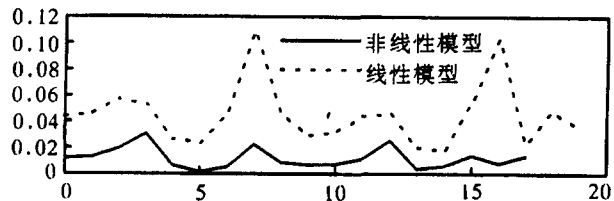


图 5 心电图信号预测结果

6 结束语

本文在传统的线性时序模型的基础上提出了非线性时序模型, 并对其作了实验研究. 结合诸多改进算法, 本文提出了一种“合成学习算法”. 对几种典型时间序列的预测分析表明, 本文的时序非线性模型具有较好的预测能力.

参 考 文 献

- [1] Pandit S M, Wu S M. Time Series and System Analysis with Applications. New York: John Wiley, 1983, Chapter I.
- [2] 韦 岗, 贺前华. 神经网络模型、学习及应用. 北京: 电子工业出版社, 1994, 191-236.
- [3] Auestad B, Tjostheim D. Identification of nonlinear time and order determination. *Biometrika*, 1990, 77(4): 669-687.
- [4] Cybenko G. Approximations by superpositions of sigmoidal function. *Mathematics of Control, Signals and Systems*, 1989, (2): 303-314.
- [5] Funahashi K. On the approximate realization of continuous mapping by neural networks. *Neural Networks*, 1989, 2(2): 183-192.
- [6] Hush D R, Sales J M. Improving the learning rate of back-propagation with the gradient reuse algorithm. *IEEE International Conference on Neural Networks*, San Diego, CA, U.S.A., July, 24-27, 1988, Vol.1, 441-449.
- [7] Stornetta W S, Huberman B A. An improved three layer, back-propagation algorithm. *IEEE First International Conference on Neural Networks*. San Diego, CA, U.S.A.: June, 21-24, 1987, Vol.2, 637-643.
- [8] Wessels L A, Barnard E, Avoiding false local minima by proper initialization of connections. *IEEE Trans. on Neural Networks*, 1992, NN-3(6): 899-905.

RESEARCHES ON THE NON-LINEAR MODELS OF TIME SERIES

Ma Ni Wang Jinshan Wei Gang

(Dept. of Electron. and Comm. Eng., South China University of Technology, Guangzhou 510641)

Abstract This paper presents some kinds of non-linear models for time series. The structure and training methods for each model have been analysed and studied. The experimental results for some common time series are given.

Key words Time series analysis, Non-linear model, Non-linear prediction

马 霓: 男, 1971 年生, 广州金鹏电子信息机器有限公司博士后, 现从事信号处理、神经网络及语音处理方面的研究工作.

王锦山: 男, 1972 年生, 博士生, 现主要从事通信网络技术方面的研究工作.

韦 岗: 男, 1963 年生, 教授, 博士生导师, 系副主任, 主要研究兴趣包括信号处理、神经网络、回波理论、图象处理、模式识别及多媒体技术等.