

文章编号: 1005—0523(2006)04—0141—03

# 基于混合自回归模型和神经网络的时间序列预测

武俊峰, 江其保

(东南大学 数学系, 江苏 南京 210096)

**摘要:** 在求出混合自回归时间序列模型的成分个数的基础上, 应用 BP 神经网络对时间序列进行了预测, 并对模型进行了数值模拟表明该预测模型的具有较高的精确度和广泛的应用前景.

**关 键 词:** 混沌时间序列; BP 神经网络; 混合自回归时间序列模型

**中图分类号:** O21

**文献标识码:** A

## 1 混合自回归模型

在时间序列分析中, 线性回归模型是最常用的一种模型. 在假定模型的噪声为正态分布时, 时间序列的边缘分布和条件分布都服从正态分布. 所以, 线性回归模型有着很好的性质. 但是, 在实际中, 很多数据的并非服从这样的分布. 比如, 在股票市场上, 股票价格有时会周期性地某些时间段内变化波动特别大. 对于这样的数据, 用通常的线性模型拟合效果就会很差. 为了解决这样的问题, 人们提出了很多非线性回归模型来拟合数据. Wang and Li(2000)<sup>[1]</sup>提出的混合自回归模型(MAR)就是一种非线性时间序列模型.

模型的定义如下:

$$F(x_t | F_{t-1}) =$$

$$\sum_{k=1}^K \alpha_k \Phi\left(\frac{x_t - \phi_{k0} - \phi_{k1}x_{t-1} - \dots - \phi_{kp_k}x_{t-p_k}}{\sigma_k}\right)$$

记作  $MAR(K; p_1, p_2, \dots, p_K)$ . 这里  $F(x_t | F_{t-1})$  是  $F_t$  在已知过去信息情况下的条件分布,  $F_t$  是到达时刻  $t$  时刻获得的信息,  $\alpha_1 + \dots + \alpha_K = 1$ ,  $\alpha_k > 0$ ,  $k = 1, \dots, K$ . 之所以把这个模型称作混合自回归模型(MAR), 是因为它实际上是  $K$  个自回归模型的混

合. 也就是有  $K$  个自回归时间序列模型

$$x_t = \phi_{k0} + \phi_{k1}x_{t-1} + \dots + \phi_{kp_k}x_{t-p_k} + \epsilon_{kt}, \epsilon_{kt} \sim N(0, \sigma^2), k=1, \dots, K$$

以概率  $\alpha_k$  来自于第  $k$  个模型.

针对未知成分个数和回归阶数的混合模型, Mu and Jiang<sup>[2]</sup>在贝叶斯框架下利用 DBMCMC 方法<sup>[3]</sup>确定了成份个数和回归阶数, 但要使其有很好的预测性需要有较多的计算. 鉴于神经网络强大的非线性映射能力, 完全可以通过它来学习时间序列, 然后进行预测.

## 2 时间序列的重构

对于时间序列  $x_1, x_2, \dots, x_n, \dots$  可以适当选取嵌入维数  $m$  和时间延滞重构相空间.

$$X(t_i) = [x(t_i), x(t_i + \tau), x(t_i + 2\tau), \dots, x(t_i + (m-1)\tau)], i=1, 2, \dots$$

其中  $\tau$  可以取为时间序列的时间间隔.  $m$  的选取可以由自回归模型中的成分个数确定. 因为混合自回归模型的成分个数描述了组成数据的空间维数. 如果时间序列为混沌时间序列,  $m$  的选取可以通过 G-P 算法<sup>[4]</sup>求得. 是否为混沌时间序列可以通

收稿日期: 2005—10—08

作者简介: 武俊峰(1981—)男, 吉林人, 东南大学数学系研究生.

过 Lyapunov 指数判别,若最大的 Lyapunov 指数则时间序列为混沌的.

G-P 算法是由 Grassberger 和 Procaccia 提出的,主要步骤如下:

1) 对于时间序列  $x_1, x_2, \cdots, x_n, \cdots$ 先给出一个较小的值  $m_0$ , 对应一个重构的相空间.

2) 计算关联函数

$$C(r) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i,j=1}^N \theta(r - |X(t_i) - X(t_j)|)$$

其中  $|X(t_i) - X(t_j)|$  表示相点  $X(t_i)$  和  $X(t_j)$  之间的距离,  $\theta(z)$  是 Heaviside 函数, 当  $z > 0$  时为 1,  $z < 0$  时为 0,  $C(r)$  是一个累积分布函数, 表示相空间中两点之间距离小于  $r$  的概率.

3) 对于  $r$  的某个适当范围, 吸引子的维数  $d$  与累积分布函数  $C(r)$  应满足对数线性关系, 即  $d(m) = 1/n(C(r))/1/n(r)$  从而由拟和求出对应于  $m_0$  的关联维数估计值  $d(m_0)$ .

4) 增加嵌入维数  $m_1 > m_0$ , 重复上面的步骤, 直到相应的维数估计值  $d(m)$  不再随  $m$  的增长而在一定误差范围内不变为止. 此时  $d$  为吸引子关联维数, 取  $m$  为大于等于  $2d+1$  的整数.

3 基于混合自回归模型和神经网络的时间序列预测的具体步骤

1) 建立网络, 根据时间序列计算出嵌入维数  $m$ , 用  $m$  作为神经网络的输入个数, 建立 BP 网络

2) 学习阶段, 将重构的时间序列的前  $N-1$  个数值序列顺序输入, 通过 BP 算法学习网络的连接权值和阈值. 直到误差控制在允许的范围内.

3) 预测阶段, 将第  $N$  个数值输入到神经网络中, 此时的输出就是该时间序列的预测值.

4 数值实验

假设时间序列来自于以下的混合自回归模型, 参数如表 1 所示. 首先通过 DBMCMC 模拟确定混合自回归时间序列的成分个数, 具体方法参见文献 [2]. 可求出成分的个数为 4, 恰好为真实模型的成分个数. 于是  $m=4$ , 取时间延迟, 建立神经网络, 进行学习. 并将真实值与预测值比较, 真实值为 3.214, 一步预测值为 3.0176, 预测相对误差为 6.11%, 真实值为 17.296, 二步预测值为 17.163, 预测相对误差为 0.77%, 平均相对误差为 0.344%, 结果如图 1 所示, 可以看出基于混合自回归模型和神经

网络的时间序列预测方法具有较好的预测结果.

表 1 模型的真实参数

| K  | $\alpha_1$  | $\alpha_2$  | $\alpha_3$  | $\alpha_4$  | $\sigma_1$           | $\sigma_2$  | $\sigma_3$  | $\sigma_4$  | $p_1$ |
|----|-------------|-------------|-------------|-------------|----------------------|-------------|-------------|-------------|-------|
| 4  | 0.37350     | 0.21090     | 0.20930     | 0.20630     | 0.83950              | 0.58020     | 0.4207      | 0.810       | 7     |
| p2 | p3          | p4          | $\phi_{10}$ | $\phi_{11}$ | $\phi_{12}\phi_{13}$ | $\phi_{14}$ | $\phi_{15}$ | $\phi_{16}$ |       |
| 4  | 2           | 2           | 0.1167      | -0.1982     | 0.06330              | 0.3010      | 1.7441      | 0.1203      | 1697  |
|    | $\phi_{20}$ | $\phi_{21}$ | $\phi_{22}$ | $\phi_{23}$ | $\phi_{30}$          | $\phi_{31}$ | $\phi_{40}$ | $\phi_{41}$ |       |
|    | 0.3782      | -0.2188     | 0.3838      | 1.445       | -0.1614              | 0.2228      | 0.0449      | 1.0162      |       |

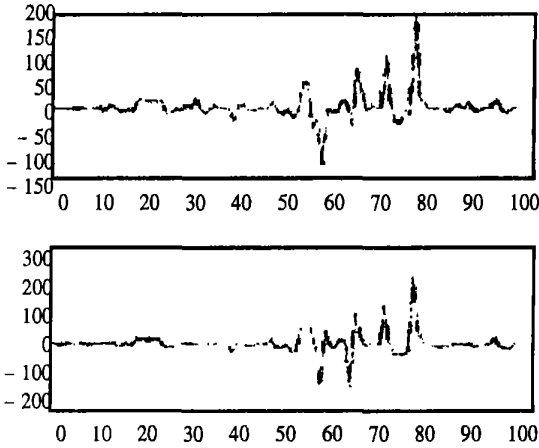


图 1 真实值与预测值的比较

5 总结

基于混合自回归模型和神经网络的时间序列预测, 当然还有许多问题值得深入研究, 应该注意的是这种方法具有较强的推广性, 在许多领域有着广阔的应用前景.

参考文献:

[1] Wang, C. S. and W. K. Li. (2000). On a mixture autoregressive model, J. R. Statist. Soc. B 62, 95~115.

[2] 穆建勇, 江其保. 未知成分个数和回归阶数的混合自回归模型的贝叶斯分析.

[3] Stephens, M. (2000) Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. Ann. Statist. 28, 40-74

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

[5] M. Casdagli. Nonlinear prediction of chaotic time series, Physica D, 1989, 35, 335~356.

[6] J. L. Chern, J. Y. Ko, J. S. Lih, H. T. Su and R. R. Hsu. Recognizing hidden frequencies in a chaotic time series, Phys. Lett. A, 1998, 238, 134~140

[7] Hwang J, Lay S, Maechler M. Regression modeling in back-propagation and projection pursuit learning. IEEE Trans Neural Networks, 1994, 5(3): 342-353

[8] Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711-732.

# Time Series Prediction Based on Mixture Autoregressive Model and BP Neural Network

WU Jun-feng, JIANG Qi-bao

(Department of Mathematics, Southeast University, Nanjing 210096, China)

**Abstract :** On the basis of components number estimation in mixture autoregressive model of time series, this paper use BP neural network method to forecast the time series, and example calculation manifest the high accuracy of model and wide prospect in practical forecasting.

**Key words :** chaotic time series; BP neural networks; mixture autoregressive model of time series