

文章编号: 1001-0920(2001)03-333-04

基于神经网络的时间序列鲁棒预测

何丕廉, 侯越先, 常虹, 孙学军

(天津大学 计算机系, 天津 300072)

摘要: 为了实现神经网络预测模型的鲁棒预测, 提出一种基于非线性偏自相关的一般化预测模型辨识方法。该方法通过考察待预测时间序列的当前序列值对各阶历史序列值的不可约自依赖, 确定神经网络这类非线性自回归模型的自回归阶数。实验表明, 该方法可有效地提高神经网络预测模型的鲁棒预测能力。

关键词: 鲁棒预测; 模型辨识; 非线性偏自相关; 神经网络

中图分类号: TP 13

文献标识码: A

Robust Time Series Prediction of Neural Network

HE Pi-lian, HOU Yue-xian, CHANG Hong, SUN Xue-jun

(Computer Science Department, Tianjin University, Tianjin 300072, China)

Abstract: To solve the problem of the robust prediction of neural networks, the paper proposed a universal method of nonlinear model identification. The method is based on the nonlinear partial autocorrelation. It could determine the order of nonlinear auto-regression by investigating the irreducible dependence between current quantities of time series and high order historical quantities. Computer simulations perfectly supported our idea.

Key words: robust prediction; nonlinear partial autocorrelation; model identification; neural network

1 引言

相对于传统的线性模型 (ARMA)^[1], 神经网络作为一种通用的非线性预测模型^[2,3], 具有诸多优势: 适当拓扑结构的 MLP, 可任意逼近紧集上的单值连续函数; MLP 和多种反馈网络 (如 TLRN^[4], RNN^[5], NARX^[6]), 分别具有实现非线性自回归 (NAR) 和非线性自回归滑动平均 (NARMA) 的能力; 参数空间为有理数域的反馈网络 (RNN 和 NARX), 具有图灵等价的计算能力^[5,6], 由丘奇-图灵假设^[7], 可认为反馈网络能力的界限即为算法能

力的界限。综上, 神经网络在理论上具备了优于线性模型的更一般的表达能力, 因而能更准确地模拟现实世界中的复杂现象。

神经网络在表达能力方面的优势是以牺牲其可解析性为代价的。由于在模型中引入非线性而增加的解析困难, 难以象对线性模型那样发展出一套基于概率统计的模型辨识技术^[3]。而在实际应用中, 经常需要在没有或缺乏先验知识辅助实现模型辨识的情况下, 完成系统辨识或预测的任务。例如, 可能需要通过某种股票价格的历史时间序列来预测股价的走势, 也可能需要通过某个生态系统中特定种群个

收稿日期: 2000-05-19; 修回日期: 2000-09-14

基金项目: 国家自然科学基金项目 (69783004); 天津市自然科学基金项目 (993800111)

作者简介: 何丕廉 (1943—), 男, 天津人, 教授, 从事人工智能、CSCW 等研究; 侯越先 (1973—), 男, 天津人, 博士生, 从事非线性时间序列分析、神经网络研究。

体数的历史时间序列来预测其未来的发展趋势。在类似的例子中,研究者所掌握的全部资料可能仅是一个一维(或多维)时间序列,由于缺乏模型辨识的先验知识,或不得不退而采用传统的线性模型,依靠盲目的试错和不完备的经验来进行神经网络模型的辨识。显然,前者不能有效地应用于非线性时间序列;后者无法保证对最优模型的选择。

本文利用非线性偏自相关^[8]实现了一种神经网络预测子的辨识和鲁棒预测方法。对仿真时间序列和真实世界时间序列的预测实验表明,该方法通过对合理的网络模型及模型规模的选择,实现了预测算法的鲁棒化。

2 非线性偏自相关及时间序列的不可约自依赖性

本文利用非线性偏自相关来确定神经网络预测子的非线性自回归阶数以及网络的类别(前馈或反馈网络)。这种考察仅涉及时间序列本身,不需要通过对基于不同备选模型的预测效果做出某种概率意义上的估计来实现辨识。模型的独立性使其得以回避非线性的解析困难。

按照动力系统的观点,可赋予非线性自回归模型一个物理解释。一般地,可假设时间序列 $X_t(t=1, 2, \dots)$,由微分方程组

$$\dot{x} = f(x), \quad f \in C^1 \quad (1)$$

的某一维可观测参量通过离散采样获得。根据分形延迟嵌入普遍性定理^[9],只要采样的时间延迟单位 T 合适(T 合适的概率为1),并且 p 大于由微分方程组所确定的动力系统的吸引子的盒计数维数^[10]的两倍,则由式(1)确定的流的吸引子与利用延迟坐标嵌入算法生成^[11]的重构空间之间存在一个微分同胚^[9]。于是可证 X_t 与其历史序列值之间存在如下函数关系^[11]

$$X_t = g(X_{t-1}, X_{t-2}, \dots, X_{t-p}) \quad (2)$$

由此,NAR模型简单地利用有限的历史信息来拟合 g ,而NARMA模型则利用滑动平均项来进一步提高对 g 的拟合精度。

非线性自回归模型自回归阶数的辨识等价于确定使 g 成为函数的最小的 p ,这个最小的 p 对应于最简约而完备的非线性自回归阶数。传统上,对 p 的估计首先要对产生时间序列背景动力系统的吸引子的某种维数(如盒维数,相关维数等)进行估计,然后利用估计结果和Taken嵌入定理^[12]来猜测合适的

p 值。但对吸引子维数的估计往往难以精确实现^[13,14],即使这种估计得以完成,Taken嵌入定理也只能大致确定合适的 p 值范围,并且这个范围随着吸引子盒维数的增加而成正比增加。由于上述困难和不确定性,这种方法在实际中往往难以奏效。

为此,我们利用时间序列的非线性偏自相关^[8]来确定 p 的数值,该指标是对非线性自相关(互信息^[15])的发展。非线性自相关反映了时间序列的自依赖性,而非线性偏自相关则反映了时间序列的不可约自依赖性,这里的不可约是相对预测而言,即无助于提高预测精度。例如,为预测由一阶递归函数 $X_t = f(X_{t-1})$ 生成的时间序列的当前值 X_t ,仅需了解 X_{t-1} 即可。任何高阶的历史序列值与当前值之间的函数依赖关系,对于预测而言均为冗余信息。

非线性偏自相关的定义如下^[8]

$$\begin{aligned} I[X_{t+1} | X_{t-(q-1)}, X_{t-(q-1)-1}, \dots, X_{t-(q-1)-(p-1)}] = \\ H(X_{t+1}) + H[X_{t-(q-1)}, X_{t-(q-1)-1}, \dots, X_{t-(q-1)-(p-1)}] - \\ H[X_{t+1}, X_{t-(q-1)}, X_{t-(q-1)-1}, \dots, X_{t-(q-1)-(p-1)}] \\ I(p, q), \quad p, q \geq 1 \end{aligned}$$

则 k 阶非线性偏自相关

$$\rho_{Gp}(k) = I(k, 1) - I(k-1, 1) \quad (3)$$

这里 $H(\cdot)$ 是香农的信息熵,可以证明 $\rho_{Gp}(k)$ 具有如下属性^[11]:

1) 正定性: $\rho_{Gp}(k) \geq 0$;

2) 渐近收敛性: 若 $X_{t+1}(t=p, p+1, \dots)$ 与其 p 阶历史值之间存在函数关系

$$X_{t+1} = g(X_t, X_{t-1}, \dots, X_{t-(p-1)})$$

则

$$\begin{aligned} \rho_{Gp}(k+1) = I(k+1, 1) - I(k, 1) = \\ I[X_{t+1} | X_t, X_{t-1}, \dots, X_{t-p}] - \\ I[X_{t+1} | X_t, X_{t-1}, \dots, X_{t-(p-1)}] = 0 \end{aligned}$$

3 非线性偏自相关的计算

非线性偏自相关的直接算法对于空间资源的要求相当高。例如若每一维有8个可能的取值,则为了计算10阶的非线性偏自相关,直接的算法须构造 2^{30} 个元素的多维数组。这里的计算复杂性在于由实验数据估计 k 维联合分布 p 。为此,可根据实验数据的散点图进行递归划分求解。

图1以计算2维信息熵 $H[X_{t+1}, X_t]$ 为例,给出了该递归算法的程序流程图。实验表明,相对于直接的计算,递归算法不仅能解决计算空间的问题,而且可有效地提高算法的时间效率。

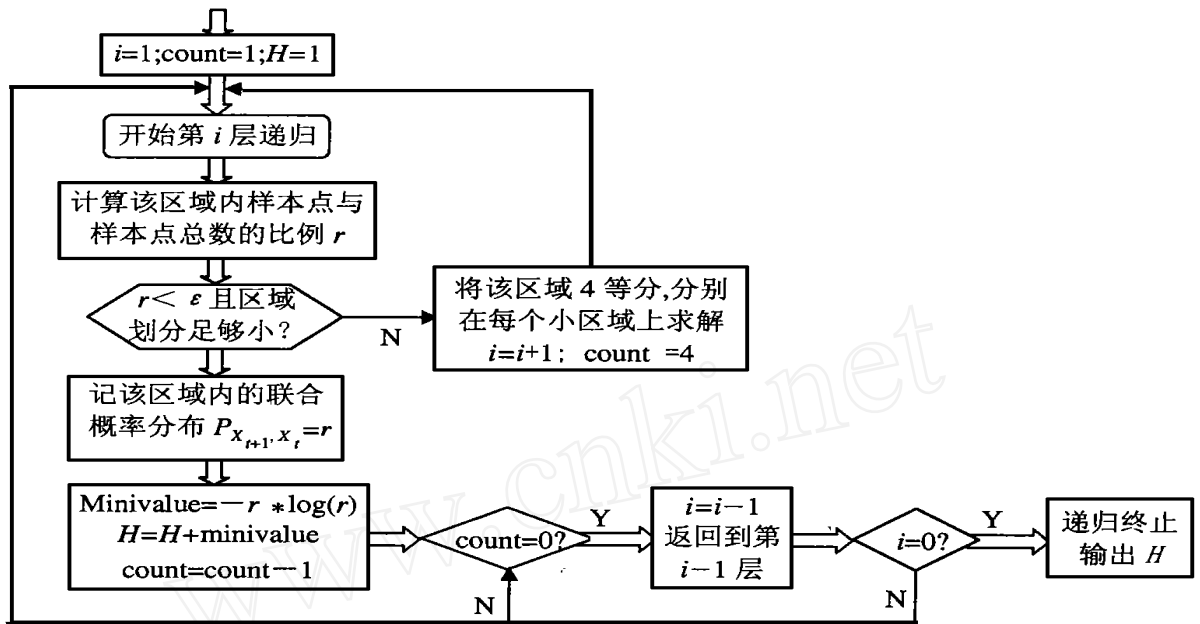


图 1 计算 $H[X_{t+1}, X_t]$ 的程序流程图

4 神经网络的鲁棒时间序列预测

经验表明, 相对于现象本身, 描述现象的模型过于简单可能无法充分反映现象的复杂性, 过于复杂又可能导入不属于现象本身的复杂性, 这两种情况都会导致模型描述现象准确程度的下降。具体到神经网络模型的辨识, 我们希望由其实现的NAR或NARMA模型的自回归阶数 p , 是使函数关系(2)成立的最小的 p 。过大的 p 不仅增加了神经网络参数空间的局部极优点, 而且可能降低网络的收敛性, 更重要的是将导致“过分拟合”现象, 从而降低模型对有噪声的测试数据的预测精度^[4]。

我们选用 3 组数值仿真样本数据, 前 2 组数值仿真数据由离散映象获得, 其迭代公式分别是

$$\text{Logistic: } X_{t+1} = 4X_t(1 - X_t)$$

$$\text{Triangle2: } X_{t+1} = 3(\sin(X_t) + \cos(X_{t-1}))$$

第 3 组数值仿真数据由 Lorenz 方程组^[10]生成。其数值解法采用欧拉方法, 步长取为 0.01。对 X 参量进行采样, 每 4 次迭代采样 1 个样本数据, 采样时滤去了迭代初期迁移态下生成的样本点。样本数据的学习集和预测集各包括 200 个样本点, 并进行规范化, 使其取值于 $[-1, 1]$ 之间。

图 2 是各级仿真数据的最初几阶非线性偏自相关曲线, 其中, —— 为 Lorenz 样本数据, - - 为 Logistic 样本数据, 为 Triangle 2 样本数据。对于 Logistic 和 Triangle 2 样本数据, 由迭代公式可知

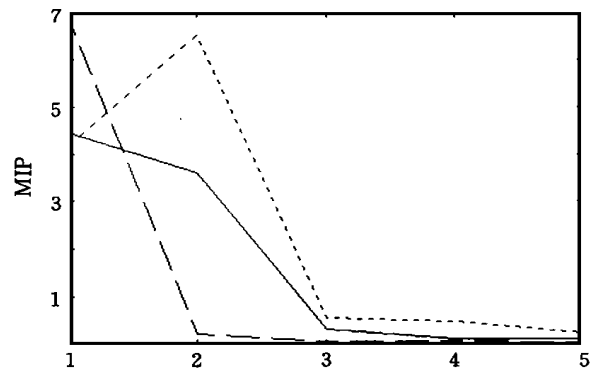


图 2 各样本数据的非线性偏自相关

其顺序样本值之间分别具有 1 阶和 2 阶不可约自依赖性, 这与图 2 的计算结果相吻合。对于 Lorenz 样本数据, 其背景动力系统的吸引子盒维数为 $2.06^{[10]}$, 非线性偏自相关反映了拓扑等价于重构背景吸引子所需的嵌入维(实验表明, 分形延迟嵌入普遍性定理中的重构条件常可弱化为“ p 等于吸引子盒维数”^[16])。表 1 给出了利用不同拓扑结构的 MLP 对各级样本数据的预测结果, 各项数值均为预测集的最小平方误差(SSE)。其中, MLP1 表示输入层单元数

表 1 实验结果

	Logistic	Triangle 2	Lorenz
MLP1	0.58	13.7	15.8
MLP2	0.69	2.86	1.41
MLP3	0.69	3.03	1.39

为1的MLP,其余类同。

综合考察上述图表,易见非线性偏自相关作为时间序列不可约自依赖性度量指标的合理性,以及这种自依赖性用于模型辨识的有效性。例如由表1,对于Logistic这样的1阶不可约自依赖样本数据,利用只有1个输入层单元的MLP1实现的NAR(1)即可进行有效的预测,增加模型的阶数并不能明显地提高预测效果。由图2,该时间序列1阶以上的非线性偏自相关都近似为零,这从理论上说明了1阶模型的充分性。类似地,对于Triangle 2和Lorenz这样的2阶不可约自依赖的样本数据,表1的实验结果表明自回归模型的必要阶数为2。

需要特别指出的是,由复杂真实世界系统生成时间序列(如股票价格时间序列)的非线性偏自相关有时是缓慢衰减的,这种现象称为一般长相关^[8]。而MLP只有短程记忆能力,无法有效预测一般长相关的时间序列。为了改进预测效果,增加反馈网络是必要的。这类模型是利用输出层单元反馈、隐含层单元反馈或输入卷积(无限脉冲滤波器 IIR)^[4]来延长模型对高阶历史的记忆能力。由于计算机开销的限制,记忆能力的提高总是以牺牲记忆解析度^[4]为代价的。尽管如此,反馈网络在某些情况下仍可对历史的不可约自依赖较长的时间序列进行预测,因此它不失为一种处理一般长相关时间序列的有效方法。

5 结 语

本文针对非线性预测模型的辨识问题,提出了基于非线性偏自相关的辨识技术。实验表明该方法可以实现非线性自回归模型自回归阶数的辨识,有效地提高了神经网络预测模型的鲁棒预测能力。

参考文献:

- [1] G Box, George E P. Time series analysis: Forecasting and control[M]. Revised Edn. San Francisco: Holden-day, 1976
- [2] J T Connor, R D Martin, L E Atlas. Recurrent neural networks and robust time series prediction [J]. IEEE Trans on Neural Networks, 1994, 5(2): 240-254
- [3] A G Parlos, K T Chong, A F Atiya. Application of the recurrent multilayer perceptron in modeling complex process dynamics [J]. IEEE Trans on Neural Networks, 1994, 5(2): 255-266
- [4] J C Principe. Neural systems: Fundamentals through simulations [M]. Chichester: John Wiley & Sons, 2000
- [5] H T Siegelmann, E D Sontag. On the computational power neural networks [J]. J of Comput Syst Sci, 1995, 50(1): 132-150
- [6] H T Siegelmann, B G Horne, C L Giles. Computational capabilities of recurrent NARX neural networks [J]. IEEE Trans on Syst, Man and Cyb, 1997, 27(2): 208-215
- [7] John E Hopcroft. Introduction to automata theory, languages and computation [M]. Addison-Wesley Publishing Company, 1979
- [8] 侯越先, 何丕廉. 基于预测复杂性的神经网络预测子辨识 [J]. 信息与控制, 2001, 30(1): 16-20.
- [9] T Sauer, J A Yorke, C S G Devaney. Embedology [J]. J of Stat Phys, 1991, 65(3/4): 579-616
- [10] D Kaplan, L Glass. Understanding nonlinear dynamics [M]. New York: Springer-Verlag, 1995
- [11] 侯越先. 预测复杂性及预测模型辨识 [D]. 天津: 天津大学计算机系, 2000.
- [12] T Shinbrot, C Grebogi, E Ott *et al*. Using small perturbations to control chaos [J]. Nature, 1993, 363(6428): 411-417.
- [13] H Abarbanel. Analysis of observed chaotic data [J]. New York: Springer, 1996
- [14] E N Lorenz. Dimension of weather and climate attractors. Nature, 1991, 353(19): 241-244
- [15] A M Fraser, H L Swinney. Independent coordinates for strange attractors from mutual information [J]. Phys Review A, 1986, 33(2): 1134-1140
- [16] F Takens. Detecting strange attractors in turbulence: Lecture notes in mathematics [M]. New York: Springer-Verlag, 1981.