文章编号: 1000-6788(2001)05-0106-04

混沌时间序列建模及预测

孙海云, 曹庆杰
(山东大学数理与系统科学学院, 山东 济南 250061)

摘要: 讨论了混沌时间序列的建模及预测方法, 给出了各重要参数的选取算法, 并应用于实例, 与传统的时间序列预测方法相比较, 取得了精度更高的预测结果, 从而为一类非线性时间序列提供了从数据采集识别到建模预测的完整技术.

关键词: 混沌; 时间序列; 相空间重构
中图分类号: O322 **文献标识码**: A

The Modeling and Forecasting of Chaotic Time Series

SUN Hai-yun, CAO Qing-jie
(School of Mathematics and System Science, Shandong University of Technology, Ji'nan 250061, China)

Abstract: We present a forecasting technique for a kind of nonlinear time series. In the analysis of chaotic time series, a good technique is to reconstruct an image of the original dynamical system using delay coordinate. It can get better forecasting result than conventional methods.

Keywords: chaotic; time series; phase space reconstruction

1 前言

人们对时间序列的分析研究已有数十年的历史了, 但主要集中在线性方法的研究上. 近年来, 对非线性系统尤其是混沌背景下产生的时间序列的分析越来越受到人们的重视. 混沌现象介于确定关系和随机关系之间, 是对现有确定模式的推广, 是自然界客观存在的一类重要的形式. 一方面, 在一个确定性系统中, 混沌现象使得对初始条件非常敏感, 一个小小的扰动变化就会被放大, 产生意想不到的结果, 这使混沌运动产生了长期不可预测的特性; 另一方面, 混沌蕴含着有序, 它不同于无从控制的随机运动, 轨迹发散但逃逸不出奇异吸引子的约束, 这使得做短期预测是可行的, 且比利用传统的线性预测模型所获得的结果要好. 对于如太阳黑子数目, 股票行情等一些看似随机的现象的建模及预测有着重要的理论和实际意义.

混沌时间序列预测的基础是状态空间的重构理论, 即把具有混沌特性的时间序列重建为一种低阶非线性动力学系统. 通过相空间重构, 可以找出隐藏区在混沌吸引子的演化规律, 使现有的数据纳入某种可描述的框架之下, 从而为时间序列的研究提供了一种崭新的方法和思路. 相空间重构是非线性时间序列分析的重要步骤, 重构的质量直接影响到模型的建立和预测.

2 相空间重构的理论基础及方法

Takens 指出, 系统中的任一分量的演化都是由与之相互作用着的其它分量所决定, 因此, 这些相关分量的信息就隐含在任一分量的发展过程之中. Packard[1]等人提出的时间延迟的思想, 可重构出观测到的动力学系统的相空间. 我们以 Lorenz 吸引子为例, 看一下他的原图与 x 分量重构图(图 1).

° 收稿日期: 1999-09-08
资助项目: 国家自然科学基金(19872041); 山东省自然科学基金(Y98A 73016)

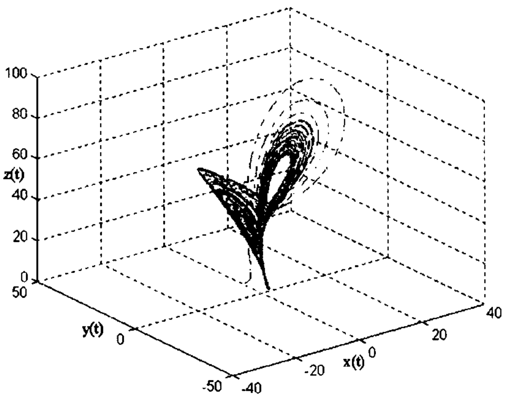

图 1　Lorenz 吸引子

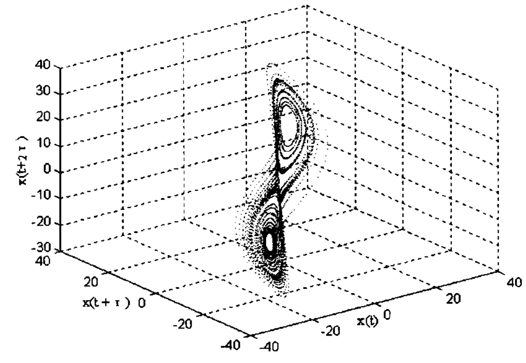

图 2　重构的 Lorenz 吸引子

由图 1 在已知 Lorenz 数学模型的基础上可知该系统的动力特性: 吸引子有两个焦点, 轨道绕两个焦点随机旋转, 轨道具有稳定的动力特性．图 2 为 $\tau = 0.2$sec 的重构图, 尽管嵌入变换使得吸引子形状、大小发生了变化, 但吸引子许多根本的动力特性并没有改变．所以, 这种方法确实可以从系统的一个变量恢复和研究整个系统的动力特性．这对于不能直接测量深层的自变量而仅仅知道一组数据序列的研究人员来说, 也有了研究系统的动力行为的可能．

我们选择一组在实验室测得的由远红外线激光器在混沌状态下产生的单变量激光数据, 选择这组数据的原因是: 其物理背景已知, 是由低自由度的简单动力系统产生的类似随机的复杂动力行为, 而且噪声低, 不需再经降噪处理．假设我们事先和不知道产生这组数据的动力系统特征, 而仅仅想从获得的一列数据中分析、重构和预报原来的动力系统模型．

在具体实施相空间重构过程中, 如何正确确定延迟时间 τ 和嵌入维数 d 是相空间重构成功的关键所在．

2.1　选择延迟时间 τ

当 τ 选择过小时, $x(t)$ 和 $x(t+\tau)$ 在数值上彼此接近, 因此不能相互独立．而当 τ 过大时, 就混沌吸引子而言, 由于蝴蝶效应的影响, $x(t)$ 和 $x(t+\tau)$ 相互之间的关系就变成随机的了．因此我们需要一种方法来选择恰当的 τ, 使得 $x(t)$ 和 $x(t+\tau)$ 之间既可相互独立, 又不至于在统计意义上完全无关．

通常取使 $x(t)$ 和 $x(t+\tau)$ 的自关联函数首次通过零点的 τ 为延迟时间, 因为此时是使 $x(t)$ 和 $x(t+\tau)$ 线性无关的最小值．自关联函数的优点是计算简单, 但它只是描述变量间线性相关程度的一种方法, 并不适用于所用情况[2]．而互信息函数可将非线性关系也考虑在内, 这种方法的根据是我们可从事件 b_j 在 B 中发生的概率中得到多少关于 a_i 在 A 集中发生概率的信息．由采农信息理论, 事件 a_i, b_j 之间的关系可用互信息熵 I_{AB} 来表示

$$I_{AB} = \sum_{ij} P_{AB}(a_i, b_j)\log_2\left[\frac{P_{AB}(a_i, b_j)}{P_A(a_i)P_B(b_j)}\right]$$

把 A 看作是由 $x(t_0 + i\tau_s)$ 组成的集合, B 是由 $x(t_0 + i\tau_s + \tau)$ 组成的集合, 则上式变为:

$$I(\tau) = \sum_i P[x(t_0 + i\tau_s), x(t_0 + i\tau_s + \tau)] \times \log_2\left\{\frac{P[x(t_0 + i\tau_s), x(t_0 + i\tau_s + \tau)]}{P[x(t_0 + i\tau_s)]P[x(t_0 + i\tau_s + \tau)]}\right\}$$

在实际计算中, 通常采用划分网格的方法, 将变量 a_i 和 b_j 组成的样本空间划分为若干‘网格’或‘盒子’, 然后通过统计各盒中的点数来求出其概率值．一般选取互信息函数第一极小值点时的 τ 为延迟时间．对所选激光数据, 计算的延迟时间 $\tau = 1$．

互信息函数的计算方法过于复杂, 无法避开大量的计算和复杂的空间划分要求．通过大量计算和对已知系统的数值实验, 我们认为, 当取 $\tau = T/4$ 为延迟时间时, 可接近最佳重构．时间域采样定理表明, 若

$x(t)$为单值、频带宽度有限的时间函数, 则当采用间隔 $\tau \le T/2$ 时, 即可精确的复现 $x(t)$. 混沌吸引子虽无周期而言, 但其具有半稳定的周期轨道[3], 寻找合适的相点 $x_i(t)$, 依次计算它与 $x_{i+1}(t)$, $x_{i+2}(t)$, $\cdots$ 的距离, 直到找到一个 $x_k(t)$, 使得 $x_i(t)$ 与 $x_k(t)$ 的距离 $\rho(x_k(t) - x_j(t)) < \epsilon$, 从 $x_i(t)$ 到 $x_k(t)$ 的轨道就是一个周期轨道, 我们可以将从 $x_i(t)$ 到 $x_k(t)$ 所用的平均时间当作周期 T. 取延迟时间 $\tau = T/4$ 是在不过分减少信息损失和不过分增加数据量之间做出的合理选择.

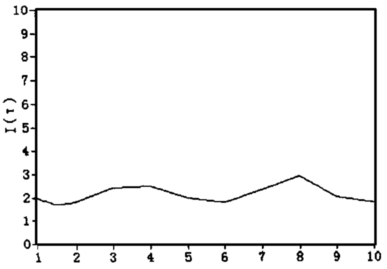

图 3 激光数据互信息函数图

2.2 嵌入维数的选择

设原始系统的吸引子维数为 d_0, 嵌入维数为 d. 在 Takens 的嵌入定理中, $d \ge 2d_0$ 仅仅只是充分条件. 在实际应用中, d 并非越大越好, 如果嵌入维数过大, 就需要更多的观测值, 更大的计算量. 在有噪声存在的非线性系统中, 维数大了, 就要花费不必要的时间来观测充满噪声的信息.

寻找一个嵌入维数为 d 的相空间, 由于投影到低维空间内, 所以会出现一些轨道的交叉点; 另外, 当 d 不是很大时, 在原始相空间中离的较远的点在重构的相空间中有可能离的很近, 因而产生了'伪邻近点'. 为了确定这些邻近点, 需要鉴别两个邻近的状态是因为动力系统行为还是因为投影到低维空间中产生的. 当逐步增加嵌入维数 d 时, 就可消除伪邻近点, 从而确定出嵌入维数.

假定嵌入维数为 d, 延迟时间为 $k\tau$, 则重构向量 $y_n = [x_n, x_{n+k}, x_{n+2k}, \cdots, x_{n+(d-1)k}]^{\mathrm{T}}$ 的邻近点由向量 $\hat{y}_n = [\hat{x}_n, \hat{x}_{n+k}, \hat{x}_{n+2k}, \cdots, \hat{x}_{n+(d-1)k}]^{\mathrm{T}}$ 确定. 根据欧氏空间理论, y_n 与 $\hat{y}_n$ 之间的距离为:

$$R_n^2(d) = \sum_{i=1}^{d} (\hat{x}_{n+(i-1)k} - x_{n+(i-1)k})^2$$

如果 $R_n(d+1)$ 与 $R_n(d)$ 相比大很多, 则可推断出 y_n 与 $\hat{y}_n$ 为伪邻近点. 在计算时, 根据实际情况选择临界值 R_T (一般 $10 \le R_T \le 50$), 看其是否满足下列不等式:

$$\frac{\hat{x}_{n+kd} - x_{n+kd}}{R_n(d)} > R_T$$

由此来确定 y_n 与 $\hat{y}_n$ 是否伪邻近点.

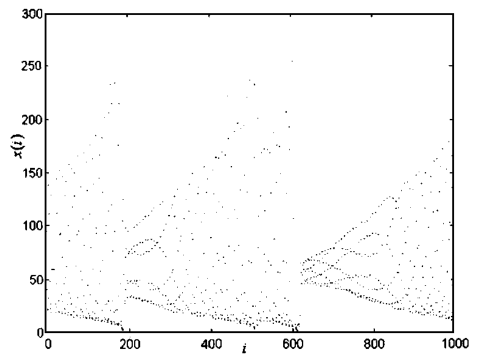

图 4 激光数据原始序列图

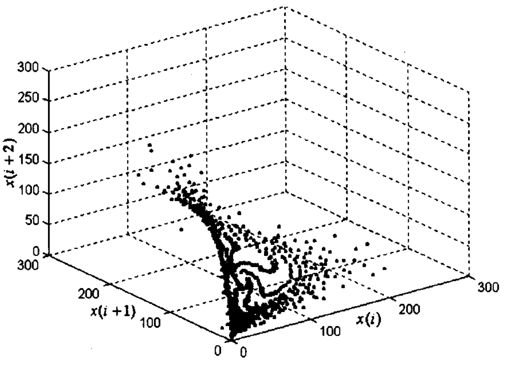

图 5 激光数据重构图

通过上述方法, 计算激光数据的重构参数可得: $\tau = 1$, $d = 3$, 重构图如图 5 所示.

3 由最大 Lyapunov 指数判断时间序列的类型

轨道的收敛率或发散率称为 Lyapunov 指数, 它是研究混沌的一个重要参数. 最大 Lyapunov 指数大

于 0, 就可判定该系统为混沌的, 存在混沌吸引. 利用相空间重构技术可从离散时间序列中得到与原系统吸引子相同的 Lyapunov 指数谱. 对重构的三维相空间利用 Wolf[4]提出的方法计算所采集的 1000 年激光数据的最大 Lyapunov 指数, 可得 $\lambda \approx 0.15 (> 0)$, 因此可判定该时间序列为混沌时间序列.

4　预测

对于平稳的时间序列来讲, 利用传统的 AR、MA、ARMA 等模型通常可获得较好的预报结果. 而对混沌时间序列而言, 即使模型对数据匹配的很好, 有时也无法做出准确的预测, 未来趋势会在性质上与原有时间序列趋势发生根本不同的变化. 因此, 对混沌时间序列的预测研究我们需另找出路.

混沌时间序列预测的基础是状态空间的重构理论[5]. 首先按上述方法重构 d 维空间, 然后建立当前值 X_t 与预测值 X_t' 的关系式, 我们希望找到合适的预测算子 f_T, 使得 $X_t' = f(X_t)$, 其中 X_t 为 d 维向量. 对混沌时间序列, 通常采用基于邻近点状态的局部预测法. 局域预测方法就是要在 X_t 的 k 个邻近似和一个线性映射. 假设任何邻近点 $X_{[t]}$ 与它的未来状态点 $X_{[t]}'$ 有下面的线性关系:

$$X_{[t]}' = f(X_{[i]}) \approx A X_{[i]} + b \qquad i = 1, 2, \cdots, k \tag{1}$$

式中: A 为元素 $a_{ij}(i, j = 1, 2, \cdots, d)$ 的常量矩阵; b 为元素 $b_j(j = 1, 2, \cdots, d)$ 的常向量, 再确定矩阵 A 和元素 b 以后, 可以把关系式 $X_t' \approx A X_t + b$ 作为预测函数, 要预测的值为

$$x_{t+d-1}' \approx a_{d_1} x_t + a_{d_2} x_{t+1} + \cdots a_{d_d} x_{t+d-1} + b_d$$

为了得到 X_{t+d-1}', 只要确定系数 $a_{d_i}(i = 1, 2, \cdots, d)$ 和 b_d 即可. 建立局部预测算子, 还有一种更为简单直接的方法即零阶预测, 是利用相空间中当前状态的邻近状态点的后续值作为当前状态的预测值. 零阶近似虽然可以很容易的提高到上述线性近似, 但除非 f 的第一次预测就是非常精确的, 否则预测精度还不如直接法好.

在预测过程中, 我们发现 X_t 的邻近点即满足 $\| X(t) - X(t') \| \le \epsilon$ 条件的点会有很多个, 但并非每个点都可作为邻近状态点进行预测, 最近的点也不一定是最好的预测点. 我们还应计算所选的邻近状态点的变化趋势是否与当前点的变化趋势相一致, 即是否满足:

$$(\overrightarrow{(X(t-1), X(t))}, \overrightarrow{(X(t'-1), X(t'))}) \le \alpha$$

利用直接法我们对已有的第 800- 810 激光数据进行预测, 并将其与采用最小最终预报误差准则[6]而建立的自回归模型得到的预测值及真实值进行比较, 结果见图 6. 由图可知, 对于混沌时间序列, 采用上述的分析建模方法比传统的自归模型所得到的预测值误差小、精度高, 且能更好的反映出时间序列的变化趋势.

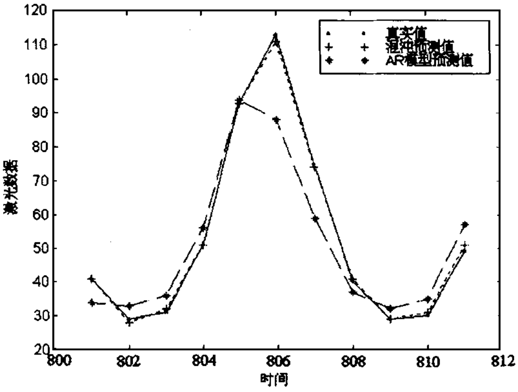

图 6　预测图

5　结束语

不论时间序列是线性还是非线性, 我们研究它的目的是相同的, 即: 利用得到的数据探究有用的模型, 使之可以分析、重构和预测原来的模型. 对非线性时间序列的分析步骤如下:

1) 区分混沌与噪声, 降噪.

2) 进行相空间重构, 其中 $\{x(t_i)\}$ 为观测到的单变量数据, τ 为延迟时间, d 为嵌入维数

$$X(t) = (x(t), x(t-\tau), \cdots, x(t-(d-1)\tau))^{\mathrm{T}}$$

3) 通过对重构后相空间的 Lyapunov 指数及分形维数等的计算判断原系统的类型.

4) 建立模型, 预测

(下转第 113 页)

$\sqrt{\hat{x}^{(0)}(k)} = -0.5011e^{-0.1893(k-1)} + 2.6177$, $k = 1, 2, \cdots, n$. $x^{(0)}$ 的预测值为 $\hat{x}^{(0)} = (4.48, 4.835, 5.189, 5.4248, 5.653, 5.8462)$.

表 2　精度检验

序号 (k)	1	2	3	4	5	6
原始值 $x^{(0)}$	4.48	4.85	5.2	5.446	5.671	5.889
预测值 $\hat{x}^{(0)}$	4.48	4.843	5.189	5.425	5.653	5.846
参差百分比 %	0	0.15	0.23	0.39	0.32	0.89

而传统建模方法得到的预测值为

$$\overline{x}^{(0)} = (4.48, 4.578, 5.1783, 5.4062, 5.667, 5.999)$$

我们从平均误差百分比, 误差平方和两个方面对两种方法进行比较, 见表 3.

表 3　两种建模方法精度比较

比较内容 / 模　型	平均误差(%)	误差平方和
中心逼近 GM (1, 1) 模型	0.33	0.00372453
传统 GM (1, 1) 模型	1.485	0.8605493

显然中心逼近式灰色 GM (1, 1) 模型的精度远远高于传统灰色 GM (1, 1) 模型, 且可调整 m 值, 使新模型更加提高精度.

参考文献:

[1]　邓聚龙. 灰色系统理论教程[M]. 武汉: 华中理工大学出版社, 1990.
[2]　熊岗, 陈章潮. 灰色预测模型的缺陷及改进方法[J]. 系统工程, 1992(6): 32-26.
[3]　刘希强, 王树泽, 宋中民. 灰色关联空间引论[M]. 贵阳: 贵州人民出版社, 1993.

(上接第 109 页)

延迟时间 τ 和嵌入维数 d 的选择是相空间重构的两个重要参数. 利用采样定理选择的延迟时间 τ, 方法简单易行. 重构效果较佳. 当经相空间重构而判定 $\{y_i\}$ 存在混沌吸引子后, 传统 AR 模型的预测值可信度不高, 而采用基于混沌吸引子的时间序列局部预测方法, 可获得较好的预测结果. 通过本文的方法对各种时间序列进行分析能有效地描述和预测混沌现象.

参考文献:

[1]　Packard N H, Crutchfield J P, Farmer J D, Shaw, R S. Geometry from a time series[J]. Phys Rev Lett, 1989, 45(9): 712-716.
[2]　Abarbanel H D I. Analysis of Observed Chaotic Data[M]. Springer, New York, 1996.
[3]　钟晓旭. 混沌吸引子中周期轨道的仿真研究[J]. 暨南大学学报, 1998, 19(1): 88-92.
[4]　Wolf A, Swift J B, Swinney H, Vastano J. Determining Lyapunov exponents from a time series[J]. Physica D, 1985, 16: 285-317.
[5]　Casdagli M. Nonlinear prediction of chaotic time series[J]. Physica D 35, 1989, 335.
[6]　杨位钦, 顾岚. 时间序列分析与动态数据建模[M]. 北京: 北京工业学院出版社, (1986).