

混沌时间序列全局预测新方法——连分式法^{*}

张 森^{1)†} 肖先赐¹⁾

¹⁾ (电子科技大学电子工程学院, 成都 610054)

²⁾ (宜宾学院数学系, 四川 644007)

(2005 年 3 月 24 日收到, 2005 年 4 月 25 日收到修改稿)

拓展了多项式逼近理论, 利用连分式法建立了混沌时间序列非线性全局预测模型, 此模型替代混沌序列的动力学方程, 实现对其动力学特性分析, 达到预测目的. 理论分析和仿真实验表明, 连分式法能够有效预测一些混沌序列, 该方法预测精度高, 并且能得到显式的预测表达式.

关键词: 混沌时间序列, 全局预测, 连分式

PACC: 0545

1. 引 言

近年来, 对非线性系统, 尤其是混沌背景下产生的时间序列分析越来越受到人们的重视. 混沌现象一种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种重要形式.

现有的混沌时间序列预测方法主要分为两类: 局域法和全局法. Farmer 和 Sidorowich 于 1987 年提出了局域预测法^[1], 这类方法特点是计算量小, 简单易行, 缺点是不能预测历史数据中没有的新点; 在大多数情形下, 还需要采用全局预测法建立全局预测模型, 以便能够预测新的数据点, 目前人们常采用多项式模型、神经网络作为工具实现全局预测. 全局预测方法主要有以多项式模型为基础的直接构造非线性模型^[2], 由 ARMA 演变的混沌模型^[3], Volterra 非线性模型^[4], 正交多项式模型^[5]; 在神经网络方面, 径向基神经网络^[6], 小波网络^[7], 混合网络^[8]. 虽然全局预测法比较多, 但是仍需要解决两个主要问题: 1) 提高方法的简便性, 尽可能减少预测模型的未知参数; 2) 尽可能减少建模所需要的样本点, 若需要样本太多, 实际情况可能不满足.

本文拓展了多项式逼近理论^[9], 引入渐进连分式, 构造非线性全局预测模型.

连分式是一个古老的数学分支, 随着科学技术

的迅速发展, 它的应用范围不断扩大, 特别是以连分式为工具的数值逼近方法已经引起人们的关注^[10]. 但是, 用连分式方法用于对混沌系统实现预测, 还未见报道, 作者在这方面作了一些研究, 得到了比较好的效果.

2. 本文预测原理

假设某一混沌系统, 可以由下面的映射来描述其动力学特性:

$$x_{n+1} = f(x_n). \quad (1)$$

但是实际上, 往往不能事先确知函数 $f(x)$ 的表达式, 只能通过观测得到部分数据点集 $\{x_i | i = 1, 2, \dots, N\}$.

对于混沌系统的重构, Packard 等^[11]首先提出了相空间重构的思想, 然后 Takens 嵌入定理^[12]和 Sauer 嵌入定理^[13]解决了这个问题.

Takens 嵌入定理说明, 通过满足一定条件的延迟坐标映射变换, 可以得到与系统真实动力学拓扑等价的动力学特性, 而且重构后的轨迹不会相交. 因此, 嵌入实际上是对一个观测时间序列的建模, 这个模型不是系统真实的动力学, 但是与真实动力学在拓扑或微分意义下等价. 从这个意义上讲, 延迟重构就是动力学系统的一个等价变换, 这个重构过程没有预测功能, 只能为进一步的预测奠定基础.

^{*} 国防预研基金(批准号 51435050101DZ0203)资助的课题.

[†] E-mail: zsenzen@163.com

本文同样根据相空间重构的思想,利用有限的观测数据构造一个高精度的连分式有理函数,使之能很好地逼近混沌系统动力学系统函数,从而达到预测目的。

引理(Weierstrass 定理)^[9] 设 $f(x)$ 是区间 $[a, b]$ 上的连续函数,则对于任意的正数 ε ,总存在多项式 $S(x)$,使得 $|f(x) - S(x)| < \varepsilon, \forall x \in [a, b]$ 。

定理 1 设 $f(x)$ 是区间 $[a, b]$ 上的连续函数,则对于任意的正数 ε ,总存在有理分式函数 $R(x) = \frac{P(x)}{Q(x)}$,使得 $|f(x) - R(x)| < \varepsilon, \forall x \in [a, b]$ 。

证明 设 $g(x) = f(x)Q(x)$ 则 $g(x)$ 为 $[a, b]$ 上的连续函数,这里 $Q(x)$ 为在 $[a, b]$ 上没有零点的多项式函数,由于 $Q(x)$ 为在 $[a, b]$ 连续则必然有界,则有 $0 < m < |Q(x)| < M$ 。

据引理,对于任意的正数 ε ,总存在多项式 $P(x)$,使得

$$|g(x) - P(x)| < \varepsilon, \forall x \in [a, b], \quad (2)$$

即

$$|f(x)Q(x) - P(x)| < \varepsilon, \forall x \in [a, b], \quad (3)$$

$$\left| f(x) - \frac{P(x)}{Q(x)} \right| \times |Q(x)| < \varepsilon, \forall x \in [a, b], \quad (4)$$

所以

$$|f(x) - R(x)| < \varepsilon/m, \forall x \in [a, b], m > 0. \quad (5)$$

由 ε 的任意性,本定理得证。

本定理从理论上肯定了闭区间上的连续函数总可以用有理分式以任意的精度来逼近。通过用 Bernstein 法证明 Weierstrass 定理的过程,不难发现,如果精度要求高,则用来逼近的 Bernstein 多项式的次数一般也很高,这往往不是很实用。在使用有理分式逼近时,有时根据需要对有理分式进行截断处理使其满足达到一定的精度即可。

3. 预测方法

本文用非线性手段构造有理渐进连分式,实际上总可以化成一个有理分式函数 $R(x) = \frac{P(x)}{Q(x)}$ 。这里 $P(x)$ 与 $Q(x)$ 分别为 m, n 阶互质的多项式,所以从本质上看,构造连分式的方法类似于构造多项式模型方法。

但不同的是,由于 $P(x)$ 与 $Q(x)$ 为互质的多项

式, $R(x)$ 的展开式是一个无穷阶的多项式,相当于构造一个无穷阶数逼近的多项式,因而能够提高模型的预测精度。

以下是方法的具体实现。

3.1. 构造数据点列

根据观测数据集 $D_N = \{x_i | i = 1, 2, \dots, N\}$, 构造一个二维数组

$$\{(x_i, y_i) | y_i = x_{i+1}, i = 1, 2, \dots, N-1\}. \quad (6)$$

在(6)式中,把 x_i 按照从小到大排列,得到一新数组

$$\{(x'_i, y'_i) | y'_i = x'_{i+1}, i = 1, 2, \dots, N-1\}. \quad (7)$$

然后把区间 $[x'_i, x'_{N-1}]$ 平均分成 K 个子区间,每个子区间至多选取一个 x' ,再加上 x_1 与 x_{N-1} ,这样便得到新数组

$$\{(x'_i, y'_i) | y'_i = x'_{i+1}, i = 1, 2, \dots\}. \quad (8)$$

(8)式就是所要构造的数据,其作用在于:

1)由一维观测数据构造二维数组,为了找 x_i 与 y_i 之间的函数关系。

2)由于 x'_i 严格递增,并且各个数字不会过于接近,方便连分式的构造,避免构造连分式时出现分母趋于零的情形。

3)让构造的数据点体现出观测数据的大小变化范围,使构造的连分式能够比较全面准确模拟系统的动力学方程。

3.2. 构造连分式预测函数

为书写方便,常把连分式表示成

$$b_0 + a_1 / (b_1 + a_2 / (b_2 + a_3 / (b_3 + \dots))). \quad (9)$$

N 阶渐进连分式记为

$$b_0 + a_1 / (b_1 + a_2 / (b_2 + a_3 / (b_3 + \dots + a_n / b_n))). \quad (10)$$

由于连分式有很好的递推算法,本文使用反差商法构造连分式。

假设给定的数据集为 $\{(x_i, y_i) | i = 1, 2, 3, \dots, N, x_i \text{ 互异}\}$ 这里 $y_i = f(x_i), i = 1, 2, 3, \dots, N$, 其中 N 为观测数据点。

定义反差商

$$d(x_i, x_j) = \frac{x_j - x_i}{y_j - y_i}, i \neq j, i, j = 1, 2, 3, \dots, N, \quad (11)$$

$$d(x_i, x_j, x_u) = \frac{x_u - x_j}{d(x_i, x_u) - d(x_i, x_j)},$$

$$u > j, u = 2, 3, \dots, n, \quad (12)$$

$$d(x_i, \dots, x_u, x_k, x_l) = \frac{x_l - x_k}{d(x_i, \dots, x_u, x_l) - d(x_i, \dots, x_u, x_k)}, \quad l > k > u. \quad (13)$$

根据这 N 组数据点, 利用反差商可以构造渐进连分式 $R(x)$, 其构造阶数为 $N - 1$.

$$L(x) = d(x_1) + (x - x_1)(d(x_1, x_2) + (x - x_2)(d(x_1, x_2, x_3) + \dots + (x - x_{N-1})(d(x_1, x_2, \dots, x_N))))). \quad (14)$$

3.3. 反差商-连分式算法

- 1) 计算反差商的第 i 阶.
- 2) 计算 i 阶渐进连分式 $L_i(x)$.
- 3) 检验点 (x_{k+i+j}, y_{k+i+j}) 是否在 $y = L_i(x)$, $j = 1, 2, \dots, N - k - i$.

记 $h_i = \sum_{j=1}^{N-k-i} g_i(j)$, 其中

$$g_i(j) = \begin{cases} 1 & \text{当点 } (x_{k+i+j}, y_{k+i+j}) \text{ 在 } y = L_i(x) \text{ 上,} \\ 0 & \text{其他,} \end{cases}$$

若 $h_i \geq (N - k - i)/2$, 则转第 5 步, 否则转第 4 步.

4) 若 $(x_{k+i+j_0}, y_{k+i+j_0})$ 是 (x_{k+i}, y_{k+i}) 之后第一个既不在 $y = L_{i-1}(x)$ 上, 也不在 $y = L_i(x)$ 上的点, 则调换点 $(x_{k+i+j_0}, y_{k+i+j_0})$ 与 (x_{k+i+1}, y_{k+i+1}) 的次序, 转第 1 步重新计算.

5) 若 $h_i \geq N - k - i$, 且 $L_i(x)$ 的最简有理分式是正则的, 则 $L_i(x)$ 为所求的渐进连分式.

根据上述算法容易得到如下定理.

定理 2 对于给定的观测数据 D_N , 利用反差法构造的渐进连分式 $L_i(x)$ 是唯一的.

证明 由观测数据集 D_N 构造的有理渐进连分式 $L(x)$, 总可以转化成一个有理分式函数 $R_{m,n}(x) = \frac{P(x)}{Q(x)}$, m, n 分别为其分子、分母的多项式的最高阶数, 且 $m + n = N - 1$; 当 N 为偶数时, $m = n$; 当 N 为奇数时, $m = n + 1$.

不妨设 N 为偶数, 现假设有两个不同的连分式分别对应着两个的有理分式函数 $R_1(x) = P_1(x)/Q_1(x)$ 和 $R_2(x) = P_2(x)/Q_2(x)$ 都满足插值条件, 则有

$$P_1(x_i)Q_1(x_i) - P_2(x_i)Q_2(x_i) = y_i \quad (i = 1, 2, \dots, N), \quad (15)$$

即

$$P_1(x_i)Q_2(x_i) - P_2(x_i)Q_1(x_i) = 0 \quad (i = 1, 2, \dots, N). \quad (16)$$

由于 $P_1(x), Q_1(x), P_2(x), Q_2(x)$ 都是次数不超过 m 的多项式, 所以次数不超过 $2m$ 的多项式 $P_1(x)Q_2(x) - P_2(x)Q_1(x)$ 有 $N = 2m + 1$ 个不同的零点, 因此它是零多项式, 即

$$P_1(x)Q_2(x) \equiv P_2(x)Q_1(x). \quad (17)$$

由此可知, 两个有理分式函数是等价的. 又因为 $L_i(x)$ 的有理分式是最简的, 所以认为 $R_1(x)$ 与 $R_2(x)$ 是同一个函数.

4. 仿真试验 :对经典混沌时间序列的预测

4.1. 预测连分式阶数的确定

确定方法 :一般把误差均方根小于某阈值(具体试验根据要求的精度而定, 例如 0.01, 0.05 等)的最低阶预测连分式作为有效预测的连分式.

因为在实际应用时, 由于对混沌时间序列没有先验知识, 因此在对连分式的阶数的确定是试探性的. 我们的方法是逐渐增加连分式的阶数, 预测连分式的阶数越高, 预测精度越高, 预测效果越理想, 这从 Ulam 和 Tent 混沌映射试验(图 1, 图 2)容易看出. 因此在这里, 当误差均方根小于 0.01 时, 我们就认为达到了预测效果, 就能确定预测的连分式的阶数.

图 1 是用 Ulam 混沌时间序列映射做了 7 次试验, 数据样本点为 500, 前 300 个点作为观测集, 分别构造出 2, 3, 4, 5, 6, 7, 8 阶渐进连分式函数 $f(x)$, 把后 200 点作为预测集, 图 1(a) 是各阶连分式在预测后 200 个点时的误差绝对值对数分布图; 从图 1(b) 可以看出, 3 阶连分式的误差均方根对数小于 -2 (即误差均方根小于 0.01), 因此我们确定 3 阶连分式就能实现有效预测.

类似地, 图 2 是用 Tent 混沌时间序列映射做了 7 次试验, 这里我们确定 8 阶连分式能达到比较高的预测精度, 即认为能实现有效预测.

4.2. logistic 混沌时间序列试验

这里以 logistic 映射的一个特例(Ulam 映射混沌映射)作介绍.

Ulam 混沌映射

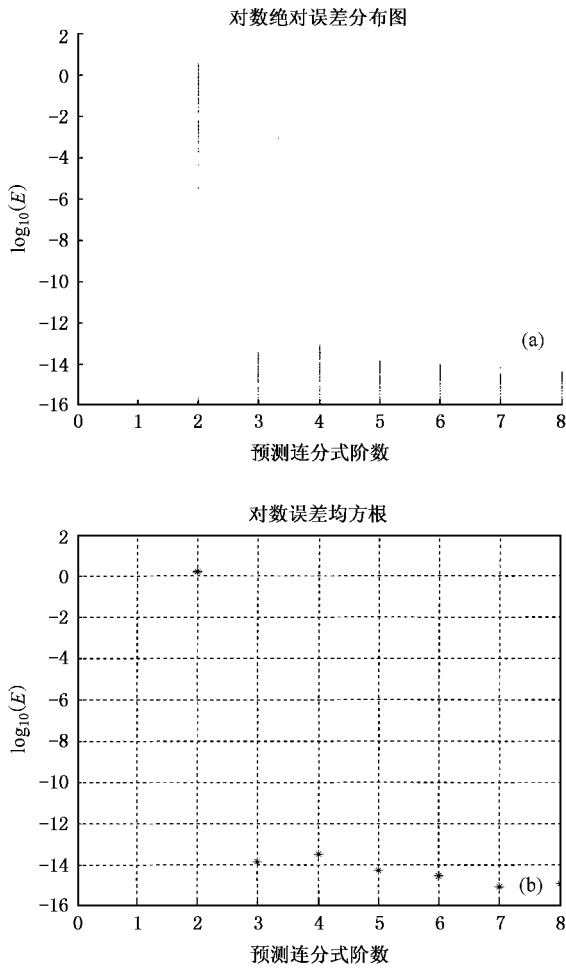


图 1 (a)Ulam 混沌映射试验对数绝对误差分布图;(b)对数误差均方根图

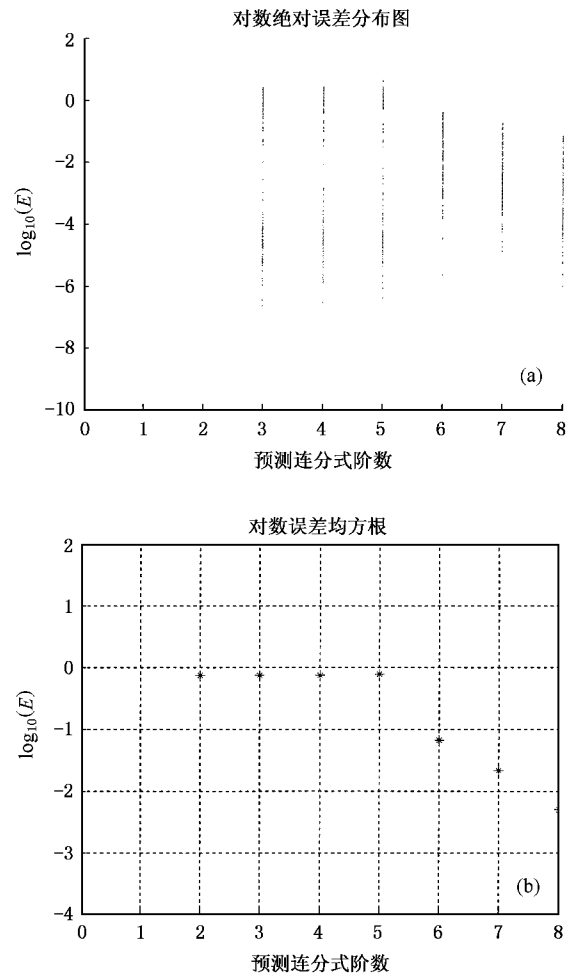


图 2 (a)Tent 混沌映射试验对数绝对误差分布图;(b)对数误差均方根图

$$x_{n+1} = 1 - 2x_n^2 \quad x_n \in [-1, 1]. \quad (18)$$

观测数据样本为 500 点,从前 300 个点中,找出 4 组数据观测值,构造出 3 阶渐进连分式函数 $f(x)$,

$$f(x) = -4503573039590813/4503599627370496 + (x + 2251796490210335/2251799813685248)(5121244468927363/18014398509481984 + (x + 3417298016567721/4503599627370496)(10141189834225365348450078668481/2883004335241789318858434150400 - 18014398509481984/5121244468927325x)). \quad (19)$$

预测效果见图 3.

实验结果分析:预测误差均方根为

$$10^{-13.8680}.$$

4.3. Tent 混沌时间序列试验

Tent 混沌映射

把后 200 点作为预测集.

构造的预测连分式为

$$x_{n+1} = \begin{cases} x(n)/a & x(n) \leq a, \\ (1-x(n))/a & x(n) > a, \end{cases} \quad \text{这里 } a = 0.4. \quad (20)$$

观测数据样本为 500 点,从前 300 个点中,找出 9 组数据观测值,构造出 8 阶渐进连分式函数 $f(x)$,把后 200 点作为预测集.

构造的 8 阶渐进连分式逼近函数 $f(x)$ 为

$$f(x) = 29/2000 + (x - 29/5000) (1405/3512 + (x - 1463/10000) (- 6232436526506409/4398046511104 + (x - 227/1000) (- 370370971923283/18446744073709551616 + (x - 607/2000) (3785533388673923/4398046511104 + (x - 1163/2500) (5467263131973579/9223372036854775808 + (x - 211/400) (4899592084047755/8796093022208 + (x - 6839/10000) (772300845976891759774186864095659/6476052483293055278002093424640000 - 1125899906842624/2300756024202519x))))))))) . \tag{21}$$

预测效果见图 4.

试验结果分析 预测误差均方根 0.0049.

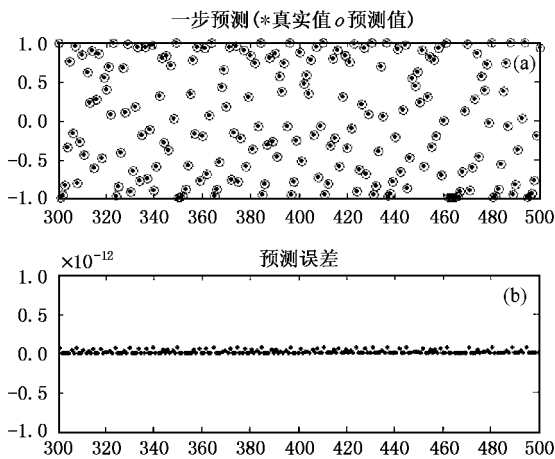


图3 Ulam混沌映射预测

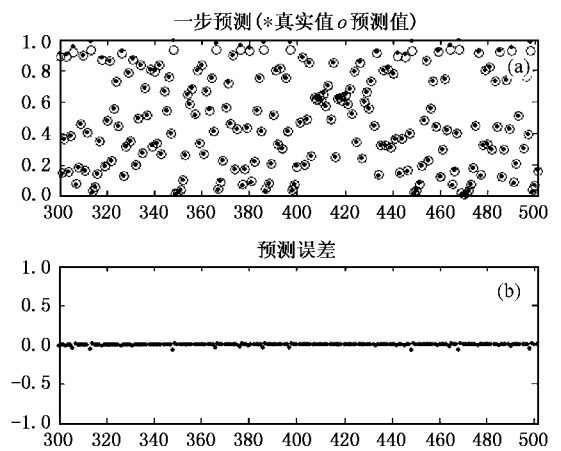


图4 Tent混沌映射预测

5. 应用 单跳频电台实测跳频码的预测

某次实验实测采样得到 162 个跳频码频率(数据略),考虑到数据分布在 65MHz 至 70MHz 之间,于

是首先利用(22)式,进行归一化处理

$$x'(n) = x(n) / \max\{x(n)\}. \tag{22}$$

在前 120 个点中,取 9 组数据观测值,构造出 8 阶渐进连分式函数 $f(x)$,并预测后 42 个点,预测连分式为

$$f(x) = 65163/69736 + (x - 65085/69736) (741/1492 + (x - 1431/1516) (2430539814382875/562949953421312 + (x - 66089/69736) (- 5336940341317727/9223372036854775808 + (x - 66655/69736) (6369660810953221/18014398509481984 + (x - 67337/69736) (- 2495379533718975/1152921504606846976 + (x - 8483/8717) (- 2791692999741229/562949953421312 + (x - 67913/69736) (- 14713818098028269992757461119863878609/8396714281732726595432477140900118528 + 9007199254740992/5012955392540379x))))))))) . \tag{23}$$

预测效果见图 5.

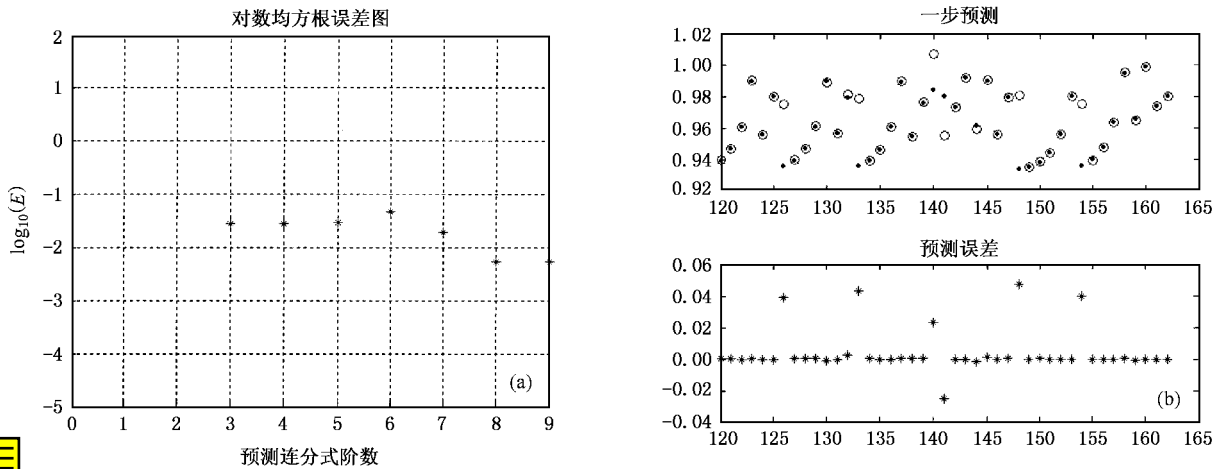


图5 跳频码预测 (a) 预测阶数与对数误差均方根 (b) 预测效果图

实验结果分析 : 将真实值 $x(n)$ 介于预测区间 $[x_p(n) - 1\% x_p(n), x_p(n) + 1\% x_p(n)]$ 的预测当作有效预测, 则有效预测达 85.71%, 预测误差均方根 0.0054.

6. 结 论

本文着重讨论了利用连分式法预测混沌时间序列, 得到了许多有用的结论:

1. 所需样本容量小 : 只需要小样本数据, 并选取的部分观测数据点, 基本能达到预测目的.
2. 计算量小 : 此方法不需要对混沌系统的相空间进行重构, 是直接利用部分观测的数据点, 即可重构构造动力学系统, 因而避免复杂的数据运算和处理.

3. 预测效果良好、预测精度高、误差小.

4. 一般地, 构造连分式的阶数越高, 则预测精度越高. 经过试验分析, 只要保证一定的预测效率和均方根误差某个阈值时, 就能确定预测连分式, 并认为达到了预测效果.

5. 本预测方法, 能够给出未知的动力学方程的近似表达式, 这优于其他许多方法^[14-20].

6. 本文在单跳频电台的跳频码预测方面提供了一种新的方法和思路.

本文方法对于连续系统有很好的预测效果. 然而, 对于有些离散系统, 则把它们近似看成连续系统来处理(本文中的试验中, logistic 时间序列比 Tent 时间序列预测效果要好); 另外对于复杂的高维混沌系统, 所涉及的方法也更加复杂, 这是作者正在进行的工作.

[1] Farmer J D and Sidorowich J J 1987 *Phys. Rev. Lett.* **59** 845
 [2] Casdagli M 1989 *Physica D* **35** 335
 [3] Ma J H, Chen Y S and Liu Z R 1999 *App. Math. & Mech.* **20** 1128 (in Chinese) [马军海、陈予恕、刘曾荣 1999 应用数学与力学 **20** 1128]
 [4] Zhang J S and Xiao X C 2000 *Acta. Phys. Sin.* **49** 403 (in Chinese) [张家树、肖先赐 2000 物理学报 **49** 403]
 [5] Jian X C and Zheng J L 2002 *Acta. Elec. Sin.* **30** 76 (in Chinese) [简相超、郑君里 2002 电子学报 **30** 76]
 [6] He X D and Lapedes A 1993 *Physica D* **70** 289
 [7] Cao L Y, Hong Y G, Fang H P and He G W 1995 *Physica D* **85** 225
 [8] Oh S K, Kim D W and Pedrycz W 2002 *International Journal of Uncertainty Fuzziness and Knowledge-based Systems* **10** 257

[9] Feng J H, Che G M and Nie Y F 2004 *Numerical Analysis Principle* (Science Press) (in Chinese) [封建湖、车刚明、聂玉峰 2004 数值分析原理 (科学出版社)]
 [10] Wang R H and Zhu G Q 2004 *Rational-function approximation & application* (Science Press) (in Chinese) [王仁宏、朱功勤 2004 有理函数逼近及其应用 (科学出版社)]
 [11] Packard N H, Farmer J D, Crutchfield et al 1980 *Phys. Rev. Lett.* **45** 712
 [12] Takens F 1981 *Dynamical Systems and Turbulence*, etc. D. Rand and L. S. Young, (Berlin, Springer) 366
 [13] Sauer T, Yorke J A and Casdagli M 1991 *Dynamical Systems and Turbulence. J. Stat. Phys* **65** 579
 [14] Yuan J 2000 *Acta. Phys. Sin.* **49** 641 (in Chinese) [袁 坚 2000 物理学报 **49** 641]

- [15] Zhang J S and Xiao X C 2000 *Acta. Phys. Sin.* **49** 1221(in Chinese] 张家树、肖先赐 2000 物理学报 **49** 1221]
- [16] Gan J C and Xiao X C 2003 *Acta. Phys. Sin.* **52** 1102(in Chinese] 甘建超、肖先赐 2003 物理学报 **52** 1102]
- [17] Gan J C and Xiao X C 2003 *Acta. Phys. Sin.* **52** 2995(in Chinese] 甘建超、肖先赐 2003 物理学报 **52** 2995]
- [18] Gan J C and Xiao X C 2004 *Chin. Phys.* **13** 312
- [19] Zhang J S and Xiao X C 2005 *Chin. Phys.* **14** 49
- [20] Zhang J S and Xiao X C 2000 *Chin. Phys.* **9** 408

A new method of global prediction for chaotic time series based on continued fractions ^{*}

Zhang Sen^{1,2)} Xiao Xian-Ci¹⁾

¹⁾(*College of Electronic Engineering , University of Electronic Science and Technology of China , Chengdu 610054 , China*)

²⁾(*Department of Mathematics , Yibin University , Yibin 644007 , China*)

(Received 24 March 2005 ; revised manuscript received 25 April 2005)

Abstract

A non-linear predictable global model for chaotic time series is built based on continued fractions after developing the method of polynomial approximants in theory. We can analyze the characteristics and realize the prediction by substituting the model for the equation of chaotic time series. The results of theoretical analysis and computer simulation have proved this method to be practically feasible for predicting a chaotic series accurately and give its explicit expression.

Keywords : chaotic time series , global prediction , continued fractions

PACC : 0545

* Project supported by the National Defense Foundation of China (Grant No. 51435050101DZ0203).