第 52 卷 第 8 期 2003 年 8 月
1000-3290/2003/52(08)/1871-06

物 理 学 报
ACTA PHYSICA SINICA

Vol.52,No.8,August,2003
©2003 Chin. Phys. Soc.

一种混沌伪随机序列复杂度分析法*

蔡觉平[1)] 李 赞[2)] 宋文涛[1)]

[1)](上海交通大学电子工程系,上海 200030)

[2)](西安电子科技大学综合业务网国家重点实验室,西安 710071)

(2002 年 9 月 30 日收到;2002 年 12 月 13 日收到修改稿)

分析了已有的序列线性复杂度分析方法,提出了用近似熵算法计算混沌运动的测度熵,作为衡量混沌伪随机序列复杂度的标准.理论研究表明,利用较短的观察序列,该方法能够准确地反映混沌系统和混沌伪随机序列复杂度的大小,可以作为判断利用混沌系统产生的伪随机序列的复杂度准则.实验结果表明该方法的有效性和理论结果的正确性.

关键词:混沌,伪随机序列,熵

PACC:0545

1. 引 言

随着混沌研究的深入,混沌在扩频通信领域中的潜在价值引起了广泛关注.由于混沌序列具有宽带、类噪声、对初始状态十分敏感等特性,因此可以取代传统的伪随机序列,应用于保密度要求高的军用和商用扩频通信系统[1—5].

混沌伪随机序列是混沌迭代产生的序列经过量化和判决得到的,为了区分两种序列,本文把由混沌迭代产生序列称为混沌序列,把量化和判决后的多进制序列称为混沌伪随机序列.目前已经提出了基于 Chebyshev 法[6]、Logistic 法[7]、耦合映象格子法[8]等用于直接扩频和跳频系统的混沌伪随机序列族,理论和实验证明它们具有较强的随机性和良好的相关性.

有限长度序列的复杂度是指它与随机序列的相似程度,是对利用序列的部分恢复出整体的难易程度的度量.因此它是衡量保密通信系统中扩频序列抗干扰和截获能力的重要指标.目前在扩频序列设计领域,主要采用线性复杂度分析算法(Berlekamp-Massey)[9]判断序列复杂度的大小.但是,研究发现对于混沌伪随机序列,该算法不能够有效地判断复杂度的大小,因此需要一种新的复杂度标准.

本文在混沌系统研究的基础上,提出利用混沌运动产生信息量的大小来度量混沌伪随机序列的复杂度,用近似熵(approximate entropy, ApEn)[10,11]作为判断复杂度大小的准则.理论和实验结果表明,该方法可以用较短的观察序列,有效地判断混沌伪随机序列的复杂度.

2. Berlekamp-Massey 算法的局限性

首先观察 Logistic 映射和耦合映象格子的混沌序列相空间结构(图 1).其中图 1(a)采用 Logistic 混沌映射

$$x_{n+1} = 4x_n(1 - x_n),$$
$$0 \leqslant x_n \leqslant 1 \qquad (n = 0,1,2,\cdots). \tag{1}$$

图 1(b)是空间维数为 6,耦合系数 $\varepsilon = 0.99$ 的耦合映象格子映射,其表达式为

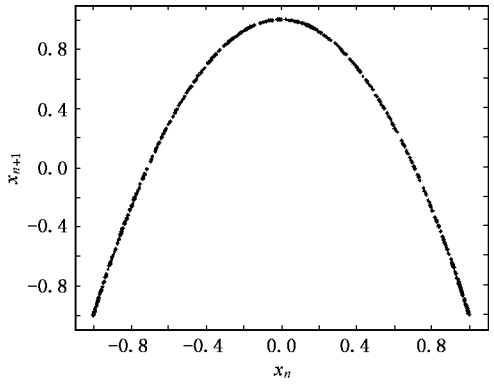
$$y_{n+1}(1) = (1-\varepsilon)f(y_n(1)) + \varepsilon g_n,$$
$$y_{n+1}(i) = (1-\varepsilon)f(y_n(i)) + \varepsilon f(y_n(i+1)),$$
$$i = 2,\cdots,6 \qquad (n = 0,1,2\cdots).$$
$$g_n = f(y_n(2)),$$
$$0 \leqslant y_n \leqslant 1 \qquad (n = 0,1,2,\cdots), \tag{2}$$

式中,边缘条件满足 $y_n(7) = y_n(1)$,映射子函数 $f(x) = 4x(1-x)$,选择耦合信号 g_n 作为伪随机序列的生成函数.

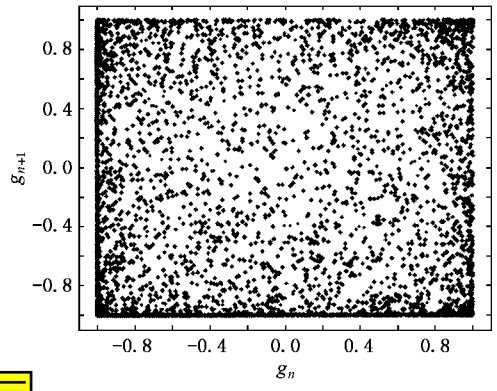
从图 1 可以看到 Logistic 映射的相空间结构是一种简单的单峰结构.与 Logistic 映射相比,图 1(b)是'混乱'的,这表明耦合映象格子生成的序列具有

* 国家重点基础研究发展规划(批准号:513060103)和国家自然科学基金(批准号:60072028)资助的课题.

更高的复杂度.



(a)



(b)

图 1 混沌序列的相空间结构 (a) Logistic 映射,(b)耦合映象格子

目前,衡量伪随机序列复杂度的最常用的方法是 Berlekamp-Massey 线性复杂度计算法[9].现有的混沌伪随机序列都采用这种方法衡量序列的复杂程度[7,12].定义判决公式(3),对(1),(2)式产生的序列进行量化,可以得到 $K = 2^n$ 进制伪随机序列 $\{\sigma_c(f^n(x))\}_{n=0}^{\infty}$.

$$\sigma_c(x) = j$$

$$\sin^2\left(\frac{j\pi}{2K}\right) < x \leqslant \sin^2\left(\frac{(j+1)\pi}{2K}\right)$$

$$(j = 0,1,2,\cdots,K-1). \qquad (3)$$

表 1 给出了两种映射所产生的 8 进制伪随机序列的线性复杂度计算结果.

表 1 混沌伪随机序列的线性复杂度

序列长度	200	400	600	800	1000
Logistic 映射	100	199	299	400	498
耦合信号 $\varepsilon = 0.99$	100	200	300	398	499

根据表 1 的结果,两种映射所产生序列的线性复杂度近似相同,因此线性复杂度不能够有效地区分混沌伪随机序列的复杂度.事实上,线性复杂度表征的是能够产生该序列的最短线性反馈寄存器长度,而混沌伪随机序列是通过混沌系统的迭代,并由其演化轨迹得到的.因此,用线性复杂度来衡量混沌伪随机序列的复杂度是不合适的.

3. 复杂度分析的 ApEn 算法

为了区分不同混沌伪随机序列的复杂程度,必须从混沌系统本身特性入手.混沌系统的相邻轨道是以指数速度分离的,在一段时间内可以区分不同的轨道数目 M 越多,那么复杂度就越大.对于混沌运动,M 随时间呈指数增长,即

$$M \propto e^{Kt}, \qquad (4)$$

式中常数 K 是测度熵,它反映了混沌运动信息产生速率.对于规则运动,$K = 0$;对于纯随机运动,$K = +\infty$;对于混沌运动,$0 < K < +\infty$.K 越大,随机性越强,被恢复的可能性越小,复杂度也就越高.

对于测度熵的一种有效方法是 K-S 熵,但是这种方法需要很大的观察样本序列,而且计算量巨大.文献[11]讨论了一种有效计算测度熵的 ApEn 方法.理论和实验证明:ApEn 可以通过较短的观察序列,有效计算混沌序列的 K-S 熵,从而确定随机序列的随机性大小.当序列是均匀分布的纯随机序列时,在相同统计条件下的 ApEn 最大.我们采用这种方法来估计混沌序列的复杂度.

对于一个长度为 N 的序列样本空间 $[u(1), u(2),\cdots,u(N)]$,定义 m 维向量组

$$x(1),x(2),\cdots,x(i),\cdots x(N-m+1) \in R^m,$$

其中

$$x(i) = [u(i),u(i+1),\cdots,u(i+m-1)]$$

$$1 \leqslant i \leqslant N-m+1.$$

定义以下几个中间量:m 维向量 $x(i)$ 和 $x(j)$ 的最大距离,

$$d[x(i),x(j)] = \max_{k=1,2,\cdots,m}(|u(i+k-1) - u(j+k-1)|). \qquad (5)$$

满足与第 i 个 m 维向量 $x(i)$ 的最大距离小于 r 的向量数,

$$C_i^m(r) = (\text{满足 } d[x(i),x(j)] \leqslant r \text{ 的 } j \text{ 的个数})/(N-m+1). \qquad (6)$$

根据 $C_i^m(r)$,定义 $\Phi^m(r)$,

$$\Phi^m(r) = (N-m+1)^{-1}\sum_{i=1}^{N-m+1}\ln C_i^m(r). \quad (7)$$

那么，ApEn 可以被定义为

$$\mathrm{ApEn}(m,r,N) = \phi^m(r) - \phi^{m+1}(r). \quad (8)$$

向量维数 m 的最大值由观察空间的长度 N 确定，当 m 越大时，ApEn 越接近测度熵. 距离参数 r 决定了该算法的分辨率，r 越小 ApEn 的分辨率越高.

混沌伪随机序列是由混沌迭代产生的序列经过量化和判决得到的. 所以存在着两种需要度量的复杂度性能指标 混沌系统的复杂度和混沌伪随机序列的复杂度.

对于混沌系统的复杂度，可以通过对其迭代映射所产生的混沌序列直接用(8)式计算 ApEn，它反映的是混沌运动的复杂程度.

混沌伪随机序列的复杂度也可以采用(8)式计算. 但是由于伪随机序列的取值是离散的，是一种特殊情况，所以计算方法有所不同. 因为 ApEn 的计算要求较小的 r，所以可以选择离散序列集的最小距离作为 r 的取值. 对于混沌伪随机序列选择 $r=0$，那么相应的(6)式可以改写为

$$C_i^m = (\text{满足 } x(i) = x(j) \text{ 的 } j \text{ 的个数})/(N-m+1), \quad (9)$$

混沌伪随机序列 ApEn 的计算公式与(8)式相同.

当观察区间 $N\to\infty$ 时，可以得到两个重要的结论.

结论 1　对于任意的 m，K 进制序列 $X_N = [x(1),x(2),\cdots,x(i),\cdots,x(N)]$ 满足

$$0 \leqslant \lim_{N\to\infty}\mathrm{ApEn}(m,N) \leqslant \ln K.$$

证　令 $\omega(y)\ y\in R^m$，是 R^m 向量 $[x_j,x_{j+1},\cdots,x_{j+m-1}]$ 的分布概率，

$$\begin{aligned}
&\lim_{N\to\infty}\mathrm{ApEn}(m,N)\\
&=\lim_{N\to\infty}\Big[(N-m+1)^{-1}\sum_{i=1}^{N-m+1}C_i^m - (N-m)^{-1}\sum_{i=1}^{N-m}C_i^{m+1}\Big]\\
&=-E[\ln(C_1^{m+1}/C_1^m)]\\
&=-E[\ln P(x_{j+m}=x_{m+1}\parallel x_{j+k-1}=x_k,(k=1,2,\cdots,m))]\\
&=-\sum_x\sum_y P(x,y)\ln P(x\parallel y)\\
&=-\sum_y\omega(y)\Big[\sum_x P(x\parallel y)\ln P(x\parallel y)\Big]\\
&\leqslant-\sum_y\omega(y)\Big[\sum_x P(x\parallel y))\ln\sum_y\omega(y)P(x\parallel y)\Big]\\
&=-\sum_y\omega(y)\Big[\sum_x P(x\parallel y)\ln P(x)\Big]\\
&=-\sum_x\sum_y\omega(y)P(x\parallel y)\ln P(x)\\
&=-\sum_x P(x)\ln P(x). \quad (10)
\end{aligned}$$

(10)式与离散集熵的结论是一致的. 对于 K 进制序列，当 $P(p_k)=\frac{1}{K}$，$(k=1,2,\cdots,K)$ 时，(10)式得到最大测度熵.

$$\lim_{N\to\infty}\mathrm{ApEn}(m,N) \leqslant \sum_x P(x)\ln P(x) \leqslant \ln K. \quad (11)$$

因为 $C_1^{m+1}\leqslant C_1^m$，所以可以得到

$$\lim_{N\to\infty}\mathrm{ApEn}(m,N) = -E(\ln(C_1^{m+1}/C_1^m)) \geqslant 0. \quad (12)$$

证毕.

结论 2　如果离散伪随机序列是在离散空间 X 中的齐次马氏链，$\pi(x)$ 是 x 的平稳分布，p_{xy} 是一步转移概率，那么对于任意的 m，

$$\lim_{N\to\infty}\mathrm{ApEn}(m,N) = -\sum_{x\in X}\sum_{x\in Y}\pi(x)p_{xy}\ln(p_{xy}).$$

证

$$\begin{aligned}
&\lim_{N\to\infty}\mathrm{ApEn}(m,N)\\
&=-E(\ln(C_1^{m+1}/C_1^m))\\
&=-E[\ln P(x_{j+m}=x_{1+m}\parallel x_{j+k-1}=x_k),k=1,2,\cdots,m)].
\end{aligned}$$

因为序列的产生是齐次马氏链，所以

$$\begin{aligned}
&\lim_{N\to\infty}\mathrm{ApEn}(m,N)\\
&=-E[\ln P(x_{j+m}=x_{1+m}\parallel x_{j+m-1}=x_m)]\\
&=-\sum_{x\in X}\sum_{y\in Y}P(x_{j+m}=y,x_{j+m-1}=x)\\
&\quad\times\ln[P(x_{j+m}=y,x_{j+m-1}=x)/P(x_{j+m-1}=x)]\\
&=-\sum_{x\in X}\sum_{y\in Y}\pi(x)P_{xy}\ln(p_{xy}).
\end{aligned}$$

证毕.

4. 参数 m,r,N 的选择

ApEn 是对有限长度的序列进行统计，计算序列的条件分布概率. 当 $N\to\infty$，$m\to\infty$，$r\to 0$ 时，(8)式就是 Eckmann-Ruelle 提出的 K-S 熵的计算方法[13].

$$\text{E-R 熵} = \lim_{r\to 0}\lim_{m\to\infty}\lim_{N\to\infty}[\phi^m(r)-\phi^{m+1}(r)]. \quad (13)$$

利用(13)式计算复杂度需要很长的观察序列,并且计算量巨大.为了通过有限长的部分序列确定复杂度,首先要确定参数 m,r 和 N,以便有效地反映真实结果.

参数 m 是距离向量的维数[14],(8)式实际是计算在已知 m 个样本的情况下,产生第 $m+1$ 个样本的条件概率.文献[15]对二进制序列的 m 的选取,进行了详细的讨论.Ornstein 和 Weiss 的研究表明:对于长度为 N 的 K 进制随机序列,为了测定长度为 m 的序列块出现的概率,所需观察序列长度的增长大于取值空间 K 的指数的增长,因此最小的观察区间长度满足 $K^{2^m} < N$.也就是

$$m_{\max} = \max(m : K^{2^m} < N). \tag{14}$$

以耦合映象格子产生的 8 进制伪随机序列为例,图 2 是 $N=300$($m_{\max}=1$)和 $N=5000$($m_{\max}=2$)时的仿真结果.结果表明 $m \geqslant 2$,$N=300$ 时的 ApEn 与 $N=5000$ 的结果有较大的差异,证明了(14)式的正确性.

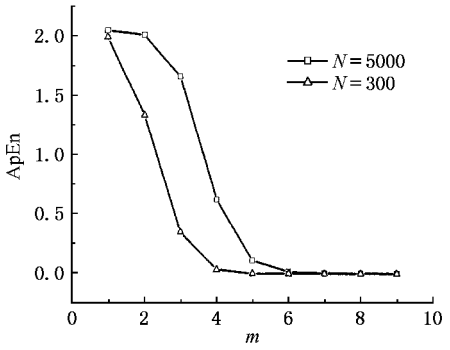


图 2 耦合映象格子产生的 8 进制伪随机序列 ApEn

在混沌运动系统的复杂度计算中,参数 r 是测量的距离大小,它决定了测量的精度.举例说明,根据(10)式,在区间(0,1)上服从均匀分布的独立随机变量 x 的最大熵为 $\ln(1/r)$,当 r 增大时 ApEn 将会减小,所以当 $r \to 0$ 时,ApEn 更接近真实情况.同时,参数 m 的选取不仅取决于 N,而且取决于 r.对于(0,1)上独立均匀分布的随机变量,观察序列的长度满足 $(1/r)^{2^m} \leqslant N$.图 3 是耦合映象格子产生的序列,在 $N=5000$,$m=2,3,4$ 时的仿真结果.r 较小时,当 m 增大,ApEn 有较大下降,证明了 m 的选取与 r 有关.同时,当 r 增大时,不同的 m 对应的曲线趋于一致,证明了参数 r 的选取影响了 ApEn 的计算精确度.实验验证,当 $N=5000$ 时,选取 $m=2$,$r<0.1$ 的 ApEn 计算结果较为准确.

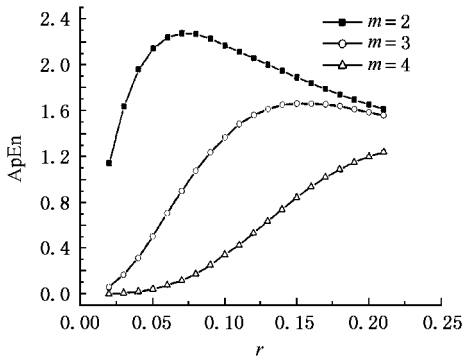


图 3 耦合映象格子混沌序列 ApEn

5 仿真及分析

图 4 是对 Logistic 映射和耦合映象格子($\varepsilon=0.99$)的混沌序列的 ApEn 计算结果,其中 $N=5000$,$m=2$.

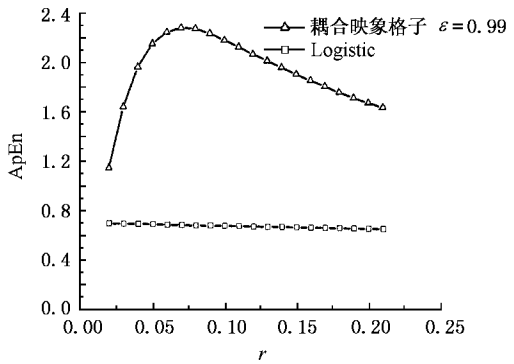


图 4 混沌序列的 ApEn 复杂度

从图中可以看到,耦合映象格子混沌序列的 ApEn 远大于 Logistic 映射的混沌序列,所以 ApEn 可以有效地判断图 1 所示混沌序列的复杂度.

对于 Logistic 映射,ApEn 的稳定值在 0.69 左右,而 Logistic 映射的 Lyapunov 指数为 0.69.测度熵与系统的 Lyapunov 指数服从 Pesin 公式,

$$K \leqslant \sum_i \lambda_i^+ .$$

这从另一个方面证明,该算法可以准确计算混沌系统的测度熵.

表 2 是对混沌序列进行判决所得到的 8 进制伪随机序列的 ApEn 计算结果,其中观察序列长度为 $N=10000$.

表 2　8 进制混沌伪随机扩频序列的 ApEn

m	1	2	3	4
Logistic 映射	0.694	0.692	0.692	0.691
耦合信号 $\varepsilon = 0.99$	2.05	2.03	1.85	1.38

从表 2 可以看到，混沌伪随机序列的 ApEn 也反映了图 1 所示混沌序列的相空间结构的复杂度，因此可以判定其复杂度取决于混沌运动的复杂度. 同时，混沌伪随机序列的 ApEn 不超过第 3 节中结论 1 所确定的理论最大值.

ApEn 是对序列样本的统计得到的，因此其应用不局限于混沌序列的复杂度分析，还可以应用于随机序列或确定性随机序列的分析.

6　结　　论

指出了 Berlekamp-Massey 算法对混沌伪随机序列复杂度分析的局限性. 利用测度熵是对混沌运动的确定随机性的度量性质，提出了用 ApEn 作为度量混沌伪随机序列复杂度分析的准则，证明了混沌伪随机序列 ApEn 的两个结论. 利用随机统计学原理，分析了对于混沌序列和混沌伪随机序列 ApEn 计算中参数的选取，并用耦合映象格子为例，验证了理论的正确性. 最后仿真并分析 Logistic 映射和耦合映象格子的混沌伪随机序列的 ApEn 复杂度. 理论和试验结果表明，该方法可以有效地判断混沌系统和混沌伪随机序列的复杂度.

[1] Parlitz U, Ergezinger S 1994 *Phys. Lett.* A **188** 149

[2] Abarbanel H D I, Linsay P S 1993 *IEEE Trans. CAS – II* **40** 643

[3] Liu J B, Ye C F, Zhang S J 1999 *Acta Phys. Sin.* **49** 20(in Chinese) [刘剑波、叶春飞、张树京 1999 物理学报 **49** 20]

[4] Yan T, Shao H H 2002 *Acta Phys. Sin.* **51** 742(in Chinese) [杨涛、邵惠鹤 2002 物理学报 **51** 742]

[5] Li J F, Li N 2002 *Chin. Phys.* **11** 1124

[6] Kohda T, Tsuneda A, 1994 *IEEE ISSSTA* '94 391

[7] Ling C, Sun S G 1998 *IEEE Trans. Commun.* **46** 1433

[8] Xiao J H, Hu G, Qu Z L 1996 *Phys. Rev. Lett.* **77** 4162

[9] Massey J L 1969 *IEEE Trans. Inform. Theory* **IT – 15** 122

[10] Jorge A *et al* 2000 *Physica* A **276** 425

[11] Pincus S, Kalman R E 1997 *Proc. Nat. Acad. Sci. USA* **94** 3513

[12] Chen Y, Ling C 2001 *Acta Electron. Sin.* **29** 868(in Chinese) [陈勇、凌　聪 2001 电子学报 **29** 868]

[13] Eckmann J P, Ruelle D 1985 *Rev. Mod. Phys.* **57** 617

[14] Grassberger P, Procaccia I 1983 *Phys.* D **9** 189

[15] Ornstein D S, Weiss B 1990 *Ann. Prob.* **18** 905

Analysis on the chaotic pseudo-random sequence complexity*

Cai Jue-Ping[1)] Li Zan[2)] Song Wen-Tao[1)]

[1)] (*Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai* 200030, *China*)

[2)] (*State Key Laboratory of ISN, Xidian University, Xi'an* 710071, *China*)

(Received 30 September 2002; revised manuscript received 13 December 2002)

Abstract

In this paper, the conventional pseudo-random sequence linear complexity is discussed, and a new criterion is proposed, based on the approximate entropy. It is proved that the criterion is able to distinguish different complex chaos and chaotic pseudo-random sequences with short observed sequence. Simulations indicate that the method is effective, and the corresponding theories are proved right.

Keywords: chaos, pseudo-random sequence, entropy

PACC: 0545

* Project supported by the State Key Development Program for Basic Research of China (Grant No. 513060103) and the National Natural Science Foundation of China (Grant No. 60072028).