基于排列熵算法的混沌伪随机序列复杂性分析

孙克辉,谈国强,盛利元
SUN Ke-hui,TAN Guo-qiang,SHENG Li-yuan

中南大学 物理科学与技术学院,长沙 410083
School of Physics Science and Technology,Central South University,Changsha 410083,China

SUN Ke-hui,TAN Guo-qiang,SHENG Li-yuan.Analysis of chaotic pseudo-random sequence complexity based on permutation entropy.Computer Engineering and Applications,2008,44(3):47-49.

Abstract: In this paper,permutation entropy algorithm is used to analyze the complexity of the chaotic series and chaotic pseudo-random series.The series complexity with different chaotic system parameters is discussed.The study results show that the series after being measured hold the traits of original series,and the complexity of TD-ERCS system is more great and stable than that of other discrete chaotic systems.So TD-ERCS is a new safe chaotic system which can be used for chaotic cryptology.
Key words: chaos;pseudo-random sequence;complexity;permutation entropy algorithm;TD-ERCS

摘 要:运用排列熵算法分析了离散混沌系统产生的混沌序列和混沌伪随机序列的复杂性,讨论了混沌系统参数对序列复杂性的影响情况。研究表明:多次粗粒化后得到的混沌伪随机序列保持了原有混沌序列的复杂性特点;与 Logistic 系统和 Henon 系统相比,TD-ERCS 系统产生的混沌伪随机序列的复杂性大且相对稳定,是一个极具密码学应用价值的安全混沌系统。
关键词:混沌;伪随机序列;复杂性;排列熵算法;TD-ERCS
文章编号:1002-8331(2008)03-0047-03 **文献标识码**:A **中图分类号**:TP13

近年来,人们试图用 Logistic 系统、Henon 系统和其它新型离散混沌系统产生的混沌序列用于混沌扩频通信、密码学等应用领域[1,2]。由于只有由复杂度高的混沌序列构成的加密系统才有高的安全性,因此,对混沌系统的复杂性分析成为了重要的研究课题。混沌系统复杂性研究,一方面可以为系统的安全性提供理论依据,另一方面可以为寻找新型的更适合密码学应用的混沌系统提供判别依据。

目前,复杂性分析理论不是很完善,复杂性也没有统一、严格的数学定义,但这并不影响人们对复杂性问题的研究。本文研究的复杂性是针对系统的行为复杂性的度量,即系统产生的序列与随机序列的相似程度。对复杂性的度量,目前有 Kolmogorov 算法[3]、Lip-Ziv 算法[4]、近似熵算法(Approximate Entropy,ApEn)[5]、排列熵算法(Permutation Entropy,PE)[6]等。Kolmogorov 算法仅提出了复杂性的概念,后由 Lip-Ziv 算法具体实现,但是所需的计算量非常大。近似熵算法存在计算结果受参数影响较大、计算结果对高维混沌系统不可靠等缺点。在这些算法中,只有排列熵算法不仅能有效地刻画系统的复杂性,而且算法简单、计算量较少、计算结果可靠,已广泛用于天气预测等领域[7]。

2004 年,盛利元等人提出了基于切延迟的椭圆反射腔离散混沌系统(Tangent Delay—Ellipse Reflecting Cavity Map System,TD-ERCS)[8]。该系统是一类新的全域性离散混沌系统,具有零相关特性和稳定的概率分布等特点,文献[9]设计了基于该系统的混沌伪随机序列发生器。但该系统的复杂性尚未得到证实。本文采用排列熵算法对量化前后的 TD-ERCS 系统、Tent 系统和 Henon 系统的复杂性进行对比分析。

1 离散混沌系统与混沌伪随机序列

1.1 经典离散混沌系统

目前,常用离散混沌系统主要有 Logistic 映射和 Henon 映射。Logistic 系统的映射方程为:

$$x_{n+1}=\mu' x_n(1-x_n) \tag{1}$$

其中,变量 $x_n\in(0,1)$,μ' 为系统参数,当 $3.569\,9\cdots<\mu'\leqslant 4$ 时,系统处于混沌态。

Henon 系统的映射方程为:

$$\begin{cases}x_{n+1}=1-ax_n+y_n\\ y_{n+1}=bx_n\end{cases} \tag{2}$$

其中,$a=1.4$,$b=0.3$,变量 $x_n\in(-1.5,1.5)$,它是一个二维离散混沌映射系统。

1.2 TD-ERCS 离散混沌系统

2004 年,盛利元等人提出了一类新的离散混沌系统(TD-ERCS)[8],其映射关系为:

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.60672041);湖南省自然科学基金(the Natural Science Foundation of Hunan Province of China under Grant No.04JJ3077)。
作者简介:孙克辉(1968-),男,博士后,主要从事混沌理论及其应用研究;谈国强(1982-),男,研究生,主要从事混沌系统的复杂性算法研究;盛利元(1956-),男,教授,主要从事混沌密码学与信息安全方面的研究。

$$\begin{cases} x_n=-\dfrac{2k_{n-1}y_{n-1}+x_{n-1}(\mu^2-k_{n-1}^2)}{\mu^2+k_{n-1}^2} \\ k_n=\dfrac{2k_n^{'}-k_{n-1}+k_{n-1}k_n^{'2}}{1+2k_{n-1}k_n^{'}-k_n^{'2}},n=1,2,3\cdots \\ k_{n-m}^{'}=-\dfrac{x_{n-m}}{y_{n-m}}\mu^2,m\leqslant n \\ y_n=k_{n-1}(x_n-x_{n-1})+y_{n-1} \end{cases} \tag{3}$$

其中,系统参数 $\mu\in(0,1]$,$|x_n|\leqslant 1$,$|y_n|\leqslant 1$;m 为整数,代表切线延迟;$k_{n-m}^{'}$ 为延迟 m 后椭圆切线的斜率;k_0 可由入射角 α 确定。显然,给定系统参数值 μ 和 m、初值 x_0 和 α,就可以求出 y_0、k_0 和 $k_0^{'}$,从而可以得到一组混沌序列$\{x_n,k_n\}$,具体的物理模型请参考文献[8]。当 $m=0$ 时,系统为 ERCS 系统;当切延迟 $m\geqslant 1$ 时,该系统称为 TD-ERCS 系统,此时,系统处于混沌状态。

1.3 混沌伪随机序列

由混沌系统方程迭代产生的序列经过量化和判决得到的序列称为混沌伪随机序列。为了区分混沌序列和混沌伪随机序列,本文把由混沌系统直接迭代产生的序列称为混沌序列,把经量化和判决后的多进制序列称为混沌伪随机序列。目前,多数文献采用的混沌序列粗粒化一般都是二次粗粒化方法,即给定一个混沌序列$\{x_n\}$,$n=1,2,\cdots,N$,先求出序列的平均值$\bar{x}$,然后定义判别式:

$$x_i=\begin{cases}1 & x_i>\bar{x} \\ 0 & x_i<\bar{x}\end{cases} \quad i=1,2,\cdots,N \tag{4}$$

即可将序列量生成 0、1 符号序列。由于只考虑了大于平均值和小于平均值两种情况,二次粗粒化的方法很可能会丢失原混沌动力系统的一些有用信息。采用多次粗粒化量化方法,其判决公式为[10]:

$$\sigma_c(x)=j$$

$$\sin^2\left(\frac{j\pi}{2K}\right)<x\leqslant\sin^2\left[\frac{(j+1)\pi}{2K}\right] \quad i=0,1,2,\cdots,K-1 \tag{5}$$

对序列$\{x_n\}$进行判决,即可以得到 $K=2^n$ 进制的混沌伪随机序列$\{\sigma_c(x_n)\}$。对于 Logisitic 系统产生的序列,其值域为(0,1),可以直接用判决公式(5),而对于 TD-ERCS 系统产生的序列$\{x_n\}\in(-1,1)$和 Henon 系统产生的序列$\{x_n\}\in(-1.5,1.5)$,则必须先对其做线性变化。对 TD-ERCS 系统,令 $x_n^{'}=\frac{1}{2}x_n+\frac{1}{2}$;对 Henon 系统,令 $x_n^{'}=\frac{1}{3}x_n+\frac{1}{2}$,使其值域变为(0,1)。由于此变换过程只有压缩和平移,故不影响原混沌序列的性质。

2 基于排列熵算法的混沌随机序列的复杂性分析

2.1 排列熵算法描述

排列熵算法是对时间序列的复杂性的一种度量计算方法。对于混沌序列,该算法描述如下:

(1)给定由系统方程迭代得到长度为的离散时间序列为$\{x_i,i=1,2,\cdots,N\}$,对$\{x_i\}$进行相空间重构,得到重构后的序列:

$$X(i)=[x(i),x(i+\tau),\cdots,x(i+(p-1)\tau)] \quad 1\leqslant i\leqslant N-p+1 \tag{6}$$

式中 p 和 τ 分别为嵌入维数和延迟时间。这里使用最大重叠情形,令 $\tau=1$,即将每个子序列向后移动一个数据点得到下一个子序列。

(2)将 $X(i)$的第 p 个重构分量$[x(i),x(i+\tau),\cdots,x(i+(p-1)\tau)]$按照升序重新进行排列,得到:

$$[x(i+(j_1-1)\tau)\leqslant x(i+(j_2-1)\tau)\leqslant\cdots\leqslant x(i+(j_p-1)\tau)]$$

$$1\leqslant j\leqslant N-p+1 \tag{7}$$

若存在序列某两个值的 $x(i)$相等,就按照 j 值的大小来进行排序。所以,任意一个向量 $\boldsymbol{X}_i$ 都可以得到一组符号序列:

$$A(g)=[j_1,j_2,\cdots,j_p] \quad 1\leqslant g\leqslant N-p+1 \tag{8}$$

(3)p 个不同的符号$[j_1,j_2,\cdots,j_p]$一共有 $p!$ 种不同的排列,也就是一共有 $p!$ 种不同的符号序列,符号序列 $A(g)$是其中的一种。将所有的排列相同的符号序列 $A(g)$归为一组,在 $N-p+1$组序列中一共有组不同的符号序列,每组序列的个数分别为 Num_1,$Num_2,\cdots,Num_k$,计算每一种符号序列出现的概率 $P_1,P_2,\cdots,P_k$。

$$P_k=\frac{Num_k}{N-p+1} \tag{9}$$

(4)时间序列$\{x_i\}$,$i=1,2,\cdots,n$ 的 k 种不同符号序列的 PE 就可以按照 Shannon 信息熵的形式定义为:

$$H(p)=-\sum_{i=1}^{k}P_k\ln P_k \tag{10}$$

(5)理论上,当 $P_k=1/p!$ 时,$H(p)$就达到了最大值 $\ln(p!)$,根据文献[6]的讨论,实际当中,$H(p)\leqslant\ln(N-p+1)$。为了方便,通常将 $H(p)$用 $\ln(N-p+1)$进行标准化处理,即:

$$0\leqslant h(p)=\frac{H(p)}{\ln(N-p+1)}\leqslant 1 \tag{11}$$

对于混沌伪随机序列的排列熵复杂性的计算,大体步骤与混沌序列的复杂性计算相似。不同的是,在第(1)步当中,应该加入对混沌序列的量化,使$\{x_i\}$,$i=1,2,\cdots,n$ 变成混沌伪随机序列。这里,仍然采用多进制量化方法对原序列进行量化。然后,对混沌伪随机序列进行重构。在第(2)步中,无需对重构后的序列进行排序,因为混沌伪随机序列本身已有一定的大小关系,故只需要统计出数目相同的序列个数 Num_k,直接进入第(3)步。

显然,$h(p)$变化体现了原序列的随机性。$h(p)$越小,原序列越规则,序列的复杂性越小;$h(p)$越大,序列越接近随机,序列复杂性越大。文献[6]中讨论了序列长短 N,以及 p 选取时计算的有效性。N 的选取不能太小,否则失去了其统计学的意义,一般 $1\,000\leqslant N\leqslant 10\,000$;$p$ 的取值范围一般为 $3\leqslant p\leqslant 15$。与其他算法相比,该算法具有概念清晰、计算快捷等特点。

2.2 排列熵算法流程

在 Matlab6.5 平台上,设计了排列熵算法的计算程序,应用排列熵算法计算混沌序列的复杂性的流程如图1 所示。

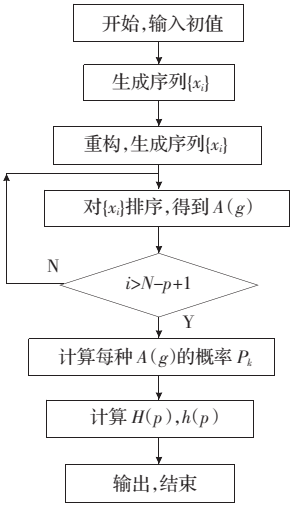

图1 排列熵算法流程图

该计算过程中，实际上是统计排序相同的重构序列的过程。在 Matlab 中,要用到两个重要的函数 perms( )和 sort( ),函数 perms( )是将所有的 p 个符号的排列都列出,结果为其返回值。而函数 sort( )则可将数组的自动按升序排列,返回值为数组的位置序号。所以,在计算中,可以很方便地利用这两个函数,将所有排列顺序一致的数组 $X(j)$找出,从而计算出概率 P_k。对于混沌伪随机序列的复杂性计算,只需按照 2.1 节中讨论的略做改动即可。

2.3　计算结果与讨论

运用排列熵算法,分析 Logistic 系统(μ'=4.0),Henon 系统(a=1.4,b=0.3)和不同参数下的 TD-ERCS 系统(μ=0.712 3)产生的混沌序列和混沌伪随机序列复杂性。选取的参数 p=5,序列长度 N=5 000,计算结果如表 1 所示。

表 1　混沌序列和混沌伪随机序列的排列熵

	Logistic	Henon	TD-ERCS(m=1)	TD-ERCS(m=2)	TD-ERCS(m=3)
混沌序列 PE	0.382 6	0.344 9	0.508 8	0.557 7	0.557 6
混沌伪随机序列 PE	0.567 8	0.552 4	0.704 3	0.975 7	0.984 1

表 1 中，混沌伪随机序列排列熵比混沌序列排列熵大,这是因为对于量化前的混沌序列,排列熵是计算序列的大小关系的排列的量度,而混沌伪随机序列是计算量化后的伪随机数的确定排列,因此,多次粗粒化量化后的序列更加适应排列熵算法,量化后的计算结果比原序列计算结果大,量化后的序列更体现了原系统的复杂程度。由表 1 的数据可知,TD-ERCS 系统的复杂性高于 Logistic 系统和 Henon 系统。另外,对于 m=2,m=3，的 TD-ERCS 系统要比 m=1 的 TD-ERCS 系统复杂性大;随着切延迟 m 的增加,其复杂性变化趋于稳定。

2.4　混沌系统参数变化对序列复杂性的影响

为了更好体现混沌序列和混沌伪随机序列的复杂性随参数变化情况，给出了 Logistic 伪随机序列和 TD-ERCS 伪随机序列的复杂性随参数变化的情况，结果如图 2~图 4 所示。同样,排列熵算法中的参数取 p=5,序列长度 N=5 000。

由图 2~图 4 可见,Logistic 系统的复杂性随参数变化比较大,而 TD-ERCS 系统的复杂性随参数变化不大,除 m=1 外,其他切延迟的排列熵值为水平直线,复杂性稳定;随着参数 μ 的增大,TD-ERCS 系统的复杂性也趋于稳定。因此,TD-ERCS 系统是一个复杂性稳定的离散混沌系统,这为该系统在密码学中的应用提供了理论依据。

此外,图 2 与熟知的 Logistic 系统的 Lyapunov 指数变化图相似,图 4 与文献[8]得到的 TD-ERCS 系统的 Lyapunov 指数“大雁图”相近，由此说明，离散混沌系统排列熵计算结果与 Lyapunov 指数变化图形状相似。在物理意义上讲,排列熵是计算序列产生新随机序列的量度，而 Lyapunov 指数是计算序列的空间发散程度,两者的物理意义有相似之处,但是在数学上这两种计算之间是否存在某种必然的联系,这有待进一步的研究。如果其中存在某种必然,那么排列熵也可以和 Lyapunov 指数一样,作为判断混沌系统的依据,而且,排列熵的概念简单,物理概念清晰,计算简单,使用方便。

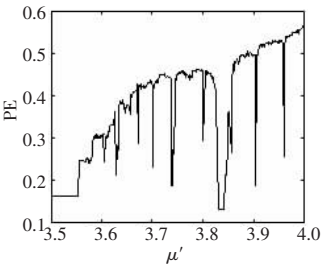

图 2　Logistic 伪随机序列的 PE 随 μ' 变化图

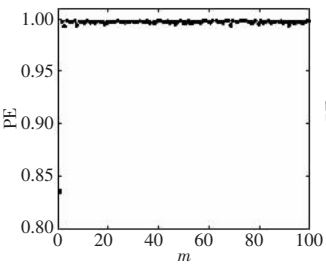

图 3　TD-ERCS 伪随机序列的 PE 随 m 变化图

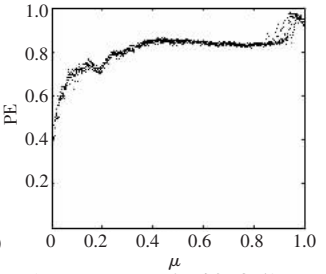

图 4　TD-ERCS 伪随机序列的 PE 随 μ 变化图

3　结论

本文运用排列熵算法对比分析了 Logistic 系统、Henon 系统和 TD-ERCS 系统的混沌序列和混沌伪随机序列的复杂性,讨论了混沌系统参数变化对复杂性的影响。通过实验分析,证明了 TD-ERCS 系统的复杂性比 Logistic 系统和 Henon 系统大,为 TD-ERCS 的密码学应用提供了复杂性实验证明。

(收稿日期:2007 年 8 月)

参考文献:

[1] 陈果,廖晓峰.一种新的基于混沌映射的分组加密方法[J].计算机工程与应用,2005,41(24):44-46.

[2] 刘建夏.一种混沌伪随机序列的设计及其应用[J].计算机工程,2005,31(18):150-152.

[3] Li M,Vitanyi P M B.Kolmogorov complexity and its applications[M].[S.l.]:Elsevier Science Publishers,1990:187-192.

[4] Lempel A,Ziv J.On the complexity of finite sequences[J].IEEE Trans,1976,22:75-79.

[5] Pincus S M.Approximate entropy as a measure of system complexity[J].Mathematics,1991,88:2297-2301.

[6] Bandt C,Pompe B.Permutation entropy:a natural complexity measure for time series[J].Phys Rev Lett,2002,88:1741-1743.

[7] 侯威,封国林,董文杰,等.利用排列熵测近 40 年华北地区气温突变的研究[J].物理学报,2006,55(5):2663-2669.

[8] 盛利元,孙克辉,李传兵.基于切延迟的椭圆反射腔离散混沌系统及其性能研究[J].物理学报,2004,53(9):2871-2876.

[9] 盛利元,曹莉玲,孙克辉,等.基于 TD-ERCS 混沌系统的伪随机数发生器及其统计特性分析[J].物理学报,2005,54(9):4031-4037.

[10] 蔡觉平,李赞,宋文涛.一种混沌伪随机序列复杂度分析法[J].物理学报,2004,52(8):1871-1876.