混沌二进制序列的伪随机性和复杂性分析

刘年生[1], 郭东辉[2]
LIU Nian-sheng[1], GUO Dong-hui[2]

1.集美大学 计算机工程学院, 福建 厦门 361021
2.厦门大学 电子工程系, 福建 厦门 361005
1.School of Computer Engineering, Jimei University, Xiamen, Fujian 361021, China
2.Department of Electronic Engineering, Xiamen University, Xiamen, Fujian 361005, China
E-mail: nsliu@jmu.edu.cn

LIU Nian-sheng, GUO Dong-hui.Analysis of pseudorandomness and complexity of chaotic binary sequences.Computer Engineering and Applications, 2008, 44(2): 16-19.

Abstract: The pseudorandomness and complexity of binary sequences generated by typical Lorenz chaotic system and Chebyshev map are analyzed and discussed.The binary sequences are obtained from the chaotic real-valued sequences generated by chaotic systems by using T.Kohda binary quantification algorithm.The statistical test, correlation function, spectral analysis, Lempel-Ziv complexity and approximate entropy are regarded as quantitative measures to characterize the pseudorandomness and complexity of binary sequences.The experimental results show the finite binary sequences generated by chaotic system approach the random sequences of Lempel-Ziv level.They are of good properties in the pseudorandomness, complexity and nonperiodicity.However, their pseudorandomness and complexity do not enhance with the sequence length increased, but degrade in the criterion of approximate entropy.Furthermore, the results of data statistics analysis show that the Lorenz system is better than Chebyshev map as the source of pseudorandomness.
Key words: chaotic system; finite binary sequence; pseudorandomness of sequence; complexity of sequence

摘 要: 分析和讨论了由经典的 Lorenz 混沌系统和 Chebyshev 映射所生成的二进制序列的伪随机性和复杂性, 采用 T.Kohda 混沌二进制量化算法, 将混沌系统所产生的实数序列转换为相应的二进制序列; 从统计检验、自相关性、频谱、Lempel-Ziv 复杂度和近似熵等多方面对序列的伪随机性和复杂性进行定量分析。统计分析结果表明对由混沌系统所产生的有限二进制序列逼近 Lempel-Ziv 意义的随机序列, 它具有较高的伪随机性、复杂性和非周期性, 但是序列的伪随机性和复杂性并不随序列长度的增加而提高, 在近似熵评价指标中呈显出降低的趋势。同时, 作为伪随机源, Lorenz 混沌系统略比 Chebyshev 映射好。
关键词: 混沌系统; 有限二进制序列; 序列伪随机性; 序列复杂性
文章编号: 1002-8331(2008)02-0016-04 **文献标识码**: A **中图分类号**: TP301.5; TP309.2

混沌系统是一类重要的伪随机序列的生成源, 近年来取得了一些研究成果[1-3]; 与真随机源(如电路噪声、混沌电路和振荡电路等)相比, 它的生成方法简单, 且不存在确定采样频率的问题[4]。与线性反馈移位寄存器方法相比, 混沌系统在序列的周期性、相关性和复杂性等方面具有优势[5], 近年来, 它被广泛应用于保密通信领域[6-8]。但是, 将混沌系统所生成的二进制序列被用于数字加密通信时, 序列的伪随机性和复杂性常常被忽略了[9-11], 没有进行定量地分析和讨论。而序列的伪随机性和复杂性是影响它应用于所设计的加密算法安全性的主要因素之一, 如何从统计分析的角度定量分析和评价混沌二进制序列的伪随机性和复杂性成为密码学家所关注的重要问题。

Martin-Löf Per 在 1966 年从统计学的角度定义了随机序列[12], Kmuth D.E.通过归纳提出了 6 种伪随机性测试方法, Ismet Bayraktaroglu 在此基础上又增加了 3 种[13]。而对于序列的复杂性, Kolmogorov A.N. 于 1965 年首次把复杂性测度引入随机性算法理论, 提出了算法复杂性概念[14]; Lempel A.和 Ziv J.在有限序列的复杂性测度中引入新模式概念, 提出了一种可计算的复杂性测度的方法[15]; 随后, Kasper F.和 Schuster H.G.在此基础上进行改进, 提出了一类完整的随机序列复杂性测度的算法[16]; Pincus S.从信息熵的角度提出了用近似熵作为序列复杂性的测度[17], 可从较短的观察序列有效计算混沌序列的 K-S 熵; Evans S.从序列相关性的角度改进了二值序列的近似熵计算表达式[18]。

目前 Chebyshev 和 Lorenz 混沌系统均有被应用于设计加

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60076015); 福建省自然科学基金(the Natural Science Foundation of Fujian Province of China under Grant No.A0640009); 福建省青年创新项目(No.2005J034); 福建省教育厅科技项目(No.JA05293); 集美大学优秀青年骨干教师基金(No.2006B003)。
作者简介: 刘年生(1967-), 男, 博士, 副教授, 主研方向为网络安全与人工智能; 郭东辉(1967-), 男, 博士, 教授, 博导。

密算法[19,20],为此本文希望通过对这两种混沌系统的伪随机性和复杂性分析,给人们应用设计加密算法提供参考。首先介绍所研究的混沌系统模型、混沌实数序列转化为二进制序列的量化算法及其相应的性能评价方法与指标;然后编程仿真,进行数据采集和统计处理,得到实验结果,并对实验结果进行分析和讨论;最后是结论。

1 混沌系统模型

混沌系统有很多类型,本文选择 Chebyshev 映射作为一维离散的混沌系统的代表,而 Lorenz 混沌系统作为高维连续的混沌系统的代表。

1.1 Chebyshev 映射

Chebyshev 映射是经典的混沌系统之一[21],它的数学表达式为:

$$X_{n+1}=\cos(P\cos^{-1}(X_n)) \quad -1\le X\le 1 \tag{1}$$

其中,P 是 Chebyshev 映射的阶数,一般为大于或等于 2 的正整数。它几乎对所有的初值它都可以产生混沌实数序列,其不变概率分布密度[21]为:

$$\rho(x)=\frac{1}{\pi\sqrt{1-x^2}} \tag{2}$$

Chebyshev 映射具有如下主要性质[22]:(1) 对于初始条件的敏感,且在 P≥2 时 Chebyshev 映射具有正的 Lyapunov 指数值;(2) 在 P≥2 时 Chebyshev 映射具有混合性和遍历性;(3) 不同阶数 Chebyshev 映射所产生的实数序列之间是正交的;(4) Chebyshev 实数序列的自相关函数是 δ函数;(5) Chebyshev 实数序列之间是正交的。

1.2 Lorenz 混沌系统

Lorenz 混沌系统是经典的多维连续的混沌系统[23],它的方程为:

$$\begin{aligned}\dot{x}&=\sigma(y-x)\\ \dot{y}&=-xz+\gamma x-y\\ \dot{z}&=xy-bz\end{aligned} \tag{3}$$

当设定参数 σ=10、b=2 和 γ=20 时,时间间隔为 dt=0.01,状态初始值为 x=0.2、y=0.3 和 z=0.2,其三维的 Lyapunov 指数[24]均大于 0,系统运动呈超混沌态,可获得状态值 x、y 或 z 的混沌实数序列,本文实验中不妨取状态值 x 的混沌实数序列。

2 量化算法

由于混沌系统为非线性动力系统,为了保证所得到的二进制序列用于密码学的安全性,混沌实数序列的二进制转化一般应优先考虑采用 T.Kohda 等人提出的实数量化算法[21,24],比直接采用符号函数量化算法安全[25]。该量化算法定义一个阈函数为:

$$\sigma_\vartheta(x)=\begin{cases}0 & x<\vartheta\\ 1 & x\ge\vartheta\end{cases} \tag{4}$$

这样,任何一个实数绝对值可以表示为二进制数:

$$|x|=0.c^1(x)c^2(x)\cdots c^m(x)\quad c^i(x)\in\{0,1\} \tag{5}$$

$$c^i(x)=\sigma_{\frac{1}{2}}\left(2^{i-1}|x|-\sum_{p=1}^{i-1}2^{i-p-1}c^p(x)\right) \tag{6}$$

其中,m为二进制数长度。对于一个无限精度的实数 x,它可以表示成为无限长度的二进制序列值,即 m→∞。

3 评价方法与指标

3.1 统计检验

从统计学的角度评价和检验任何序列的随机性,而不管这个序列是怎样产生的;它的计算公式如下[26]:

$$x^2=\sum\frac{(f_0-f_e)^2}{f_e} \tag{7}$$

其中,f_0:观察实际的次数,f_e:期望次数,用 x^2 分布检验是否存在显著性差异;如果由 0 和 1 所组成的各种子块是等分布的,就意味着序列具有较高的伪随机性[27],当子块的长度为 1 代表频数检验,而为 2 和 3 时为块频检验。

3.2 序列的自相关性

自相关函数是描述随机序列 x(t) 在任意两个不同位置 t_1,t_2,的取值之间的相关程度;相关性函数 Cor(m) 的定义如下[28]:

$$Cor(m)=\lim_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N}x_{i+m}\vartheta_i \tag{8}$$

其中:

$$\vartheta_i=f^i(x_0)-\bar{x} \tag{9}$$

$$\bar{x}=\lim_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N}f^i(x_0) \tag{10}$$

m、N 分别为序列的偏移值和长度,$f^i(x_0)=x(i)$。

3.3 频谱分析

对二进制序列进行频谱分析主要是为了检测序列是否存在中心频率,如果产生的频谱有中心频率,则说明这种方法所生成的序列具有明显的周期性,不是一个理想的随机序列。序列 x(n) 的频谱计算公式为[29]:

$$x(n)=\sum_{n=1}^{N}x(n)^{(-j*2*\pi*(k-1)*(n-1)/N)} \quad 0\le k\le N \tag{11}$$

其中:N 为序列 x(n) 的长度,即样本数,k 为谐波的级数。

3.4 Lempel-Ziv 复杂度

Lempel-Ziv 复杂度[15,16]的计算过程如下:对于一个字符串 $S(S_1,S_2,\cdots,S_n)$ 后再加一个字符串 $Q(q_1,q_2,\cdots,q_{in})$ 得到一个字符串 SQ,令 SQv 是 SQ 减去最后一个字符所得字符串,再判断 Q 是否是 SQv 的一个子串,如果 Q 是 SQv 的一个子串,把这个字符加到后面,继续增长 Q,再判断。如果 Q 不是 SQv 的一个子串,则用‘.’把前后分开,下一步把‘.’前的所有字符看成 S,重新构造 Q,重复以上过程直到结束。

序列的 Lempel-Ziv 复杂性定义为由‘.’界定的 S 的子串数目,几乎所有的 c(n) 都趋向于一定值。即:

$$\lim_{n\to\infty}c(n)=b(n)=\frac{n}{\log_2 n} \tag{12}$$

所以 c(n) 是随机序列的渐近行为。用 b(n) 来对 c(n) 进行归一化,即:

$$D(n)=\frac{c(n)}{b(n)} \tag{13}$$

因此,可以用这归一化的 D(n) 来测度给定序列的复杂性变化。对完全随机的序列 D(n) 值趋向于 1,而周期性序列的 D(n) 趋向于 0,其余情况介于两者之间。相对复杂度 D(n) 反应了一个给定序列与随机序列的接近程度,某序列的 D(n) 趋向于 1,则

表明这个序列趋近随机序列。

3.5 近似熵

近似熵[18]是基于信息熵来评估二进制序列复杂性的计算方法。对于一个 L bit 的二进制序列 c 需要记为双极的二进制序列 E, 即:

$$E=\{E(i)=2*c^{i}(x)-1\},\ 0\leq i<L \tag{14}$$

其中, $c^{i}(x)\in\{0,1\}$。这样, 二进制序列的自相关函数 R 可定义为:

$$R=\{r(i)\},\ 0\leq i<L \tag{15}$$

其中: $r(i)=\sum_{j=0}^{L-i-1}E(j)E(i+j)$。而非负功率谱密度 Φ_i 可通过 R 的 Fourier 变换计算出来, 即:

$$\Phi=abs(fft(R)) \tag{16}$$

其中 $\Phi=\{\Phi_i\}$, $0\leq i<L$, abs 表示实数绝对值或复数模。这样, 混沌系统的复杂性就可以近似熵的方法用量化后的二进制序列 c 的信息熵 Ψ 来定量表示, 即:

$$\Psi=\frac{1}{F}\sum_{i=0}^{L-1}\Phi_i\log\Phi_i \tag{17}$$

其中, F 为归一因子。

4 实验结果与分析

Chebyshev 映射和 Lorenz 混沌系统所生成二进制序列的统计检验结果如表 1 所示, 统计处理的置信水平 α=0.95, 从表 1 中可以看到, 随着序列长度的增加和子块长度的增加, 所产生的二进制序列的通过率呈下降趋势; 并且, Lorenz 混沌系统所生成二进制序列的统计检验结果比 Chebyshev 映射的好, 特别是在长序列时。

表 1 Chebyshev 和 Lorenz 混沌系统所生成二进制序列的统计检验结果

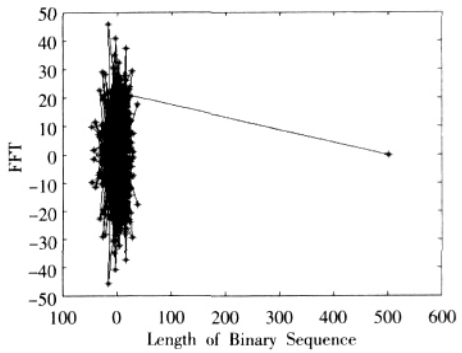
序列长度/bit	序列数	子块长度/bit	通过率/%	
			Chebyshev	Lorenz
100	1 000	1	82.2±2.1	83.6±2.1
100	1 000	2	70.5±2.2	73.7±2.2
100	1 000	3	39.4±2.5	45.6±2.1
1 000	1 000	1	70.1±2.2	84.3±2.2
1 000	1 000	2	53.0±2.0	71.4±2.1
1 000	1 000	3	28.1±2.1	28.3±2.0
10 000	100	1	11.1±1.1	67.5±2.2
10 000	100	2	8.0±1.0	39.1±2.0
10 000	100	3	2.1±1.1	21.1±2.1

Chebyshev 映射和 Lorenz 混沌系统所生成二进制序列的自相关性结果如表 2 所示, 从表 2 中, 可以看到所得到的二进制序列呈现出良好的自相关性, 并且 m 值对序列的自相关性影响甚微。

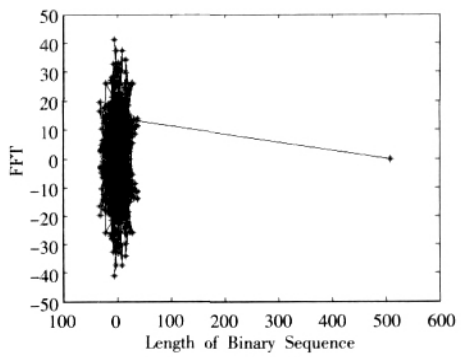
表 2 Chebyshev 和 Lorenz 混沌系统所产生的二进制序列的自相关性结果 (α=0.95)

系统类型	序列长度	序列数	m 值	自相关性
Chebyshev 映射	100	1 000	5	0.025±0.016
	1 000	1 000	5	0.008±0.006
	100	1 000	10	0.025±0.020
	1000	1 000	10	0.008±0.006
Lorenz 混沌系统	100	1 000	5	0.023±0.017
	1 000	1 000	5	0.008±0.005
	100	1 000	10	0.022±0.019
	1 000	1 000	10	0.007±0.005

对 Chebyshev 映射和 Lorenz 混沌系统所生成二进制序列的频谱, 随机选取序列长度分别为 100 bit、1 000 bit 和 10 000 bit 各 1 000 个样本进行分析, 所有结果均类似于图 1 所示; 未见明显的中心频率。



(a)Chebyshev 映射



(b)Lorenz 混沌系统

图 1 Chebyshev 映射和 Lorenz 混沌系统所生成二进制序列的频谱

Chebyshev 映射和 Lorenz 混沌系统所生成二进制序列的 Lempel-Ziv 复杂度如表 3 所示, 它们的复杂度均超过 0.9, 逼近 Lempel-Ziv 意义下的随机序列, 且与序列的长度无关; 这表明随着序列长度的增加, 不断有新的子串出现, 保证二进制序列具有较高的复杂性和为随机性。同时, 从表 3 中, Lorenz 混沌系统所生成二进制序列的 Lempel-Ziv 复杂度均值略高于 Chebyshev 映射的, 但无显著性差异。

表 3 Chebyshev 和 Lorenz 混沌系统所产生的二进制序列的 Lempel-Ziv 复杂度 (α=0.95)

系统类型	序列长度	序列数	d(n)值	Lempel-Ziv 复杂度
Chebyshev 映射	100	1 000	14.76±1.22	0.981±0.081
	1 000	1 000	91.62±2.64	0.913±0.026
	10 000	1 000	718.50±5.49	0.955±0.008
Lorenz 混沌系统	100	1 000	15.20±1.02	1.010±0.068
	1 000	1 000	91.69±2.20	0.914±0.022
	10 000	1 000	720.16±5.13	0.957±0.007

由 Chebyshev 映射和 Lorenz 混沌系统所生成二进制序列的近似熵如表 4 所示, 随着序列的长度的增加, 其近似熵呈现下降的趋势, 这表明二进制序列越长, 其可压缩性越高, 并逐渐呈现出一些有序性的特性; 造成这一现象的主要原因是有限精度效应[30,31], 在经典混沌系统可通过增加扰动和收缩等操作减少有限精度效应的影响。同时, 从表 4 中, Lorenz 混沌系统所生成二进制序列的近似熵均值略高于 Chebyshev 映射的, 但无显

著性差异。

表 4 Chebyshev 和 Lorenz 混沌系统所生成二进制序列的近似熵（α=0.95）

系统类型	序列长度	序列数	近似熵
Chebyshev 映射	100	1 000	0.249±0.071
	1 000	1 000	0.164±0.035
	10 000	1 000	0.078±0.026
Lorenz 混沌系统	100	1 000	0.258±0.070
	1 000	1 000	0.175±0.032
	10 000	1 000	0.117±0.023

5 结论

本文统计分析和讨论了 Chebyshev 和 Lorenz 两种的混沌系统所产生的二进制序列的随机性和复杂性、将混沌系统所产生的实数序列用 T.Kohda 混沌二进制量化算法转换为相应的二进制序列；从统计检验，自相关性、频谱、Lempel-Ziv 复杂度和近似熵等多方面对序列的伪随机性和复杂性进行定量分析。统计分析结果表明对由混沌系统所产生的有限二进制序列逼近 Lempel-Ziv 意义下的随机序列，它具有较高的伪随机性、复杂性和非周期性，但是序列的伪随机性和复杂性并不随序列长度的增加而提高，在近似熵评价指标中呈显出降低的趋势；在有限精度效应的制约下，随着迭代次数的增加，序列变长，逐渐呈现出有序性和周期性。为了减少有限精度效应的影响，在混沌系统实现时应增加扰动，例如用 m 序列作为扰动序列；同时，作为伪随机源，Lorenz 混沌系统与 Chebyshev 映射相比在性能上略好，应优先考虑使用。（收稿日期：2007 年 9 月）

参考文献：

[1] Kinsner W.Characterizing chaos through lyapunov metrics[J].IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, 2006, 36(2): 141-151.

[2] Anders Johansson J.Analysis of formal randomness in a Chaotic random number generator[C]//Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems, Michigan, USA, Aug 8-11 2000, 2: 724-725.

[3] Delgado-Restituto M, Rodriguez-Vazquez A.Mixed-signal map-configurable integrated Chaos generator for Chaotic communications[J]. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 2001, 48(12): 1462-1474.

[4] Wang L, Wang F, Wang Z.Novel Chaos-based pseudorandom number generator[J].Acta Physica Sinica, 2006, 55(8): 3964-3968.

[5] Thomas S, Werner J P, Hartmut B, et al.Stochastic modeling of experimental Chaotic time series[J].Physical Review Letters, 2007, 98(4).

[6] Papadimitriou S, Bezerianos A, Bountis T.Secure communication with Chaotic systems of difference equations [J].IEEE Transactions on Computers, 1997, 46(1): 27-38.

[7] Kocarev L.Chaos-based cryptography: a birief overview[J].IEEE Transactions on Circuits and Systems, 2001, 1(3): 6-21.

[8] Dachselt F, Schwarz W.Chaos and cryptography[J].IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 2001, 48(12): 1498-1509.

[9] 权安静，蒋国平，左涛，等.基于超混沌序列的分组密码算法及其应用[J].南京邮电学院学报，2005，25(4)：80-84.

[10] 徐全生，李震，杜旭强.一种基于混沌序列的图像加密算法[J].小型微型计算机系统，2006，27(9)：1754-1756.

[11] Chiaraluce F, Ciccarelli L, Gambi E, et al.A new chaotic algorithm for video encryption[J].IEEE Transactions on Consumer Electronics, 2002, 48(4): 838-844.

[12] Martin-Löf Per.The definition of random sequences[J].Information and Control, 1966, 9(6): 602-619.

[13] Bayraktaroglu I, Orailoglu A.Selecting a PRPG: randomness, primitiveness, or sheer Luck?[C]//Asian Test Symposium, 2001: 373-378.

[14] Kolmogorov A N.Three approaches to the quantitative definition of information[J].Problem of Information Transmission, 1965, 1(1): 1-7.

[15] Lempel A, Ziv J.On the complexity of finite sequence[J].IEEE Transactions on Information Theory, 1976, 22(1): 75-81.

[16] Kasper F, Schuster H G.Easily calculable measure of system complexity of spatio-temporal patterns[J].Phys Rev A, 1987, 36(7): 842-848.

[17] Pincus S.Approximate Entropy (ApEn) as a complexity measure[J]. Chaos, 1995, 5: 100-117.

[18] Evans S, Bush S F, Hershey J.Information assurance through kolmogorov complexity[C]//DARPA Information Survivability Conference and Exposition, 2001: 322-330.

[19] Liu Nian-sheng, Guo Dong-hui, Wu Bo-xi, et al.A new images hiding scheme based on Chaotic sequences[J].Wuhan University Journal of Natured Science, 2005, 10(1): 303-307.

[20] Gonzales O A, Han G, de Gyvez J P, et al.Lorenz-based Chaotic cryptosystem: a monolithic implementation[J].IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 2000, 48(8): 1243-1247.

[21] Tohru Kohda, Akio Tsuneda, Tetsuya Sakae.Chaotic binary sequences by chebyshev maps and their correction properties[C]// IEEE Second International Symposium on Spread Spectrum Techniques and Applications, Yokohama, Japan, Dec 1992: 63-66.

[22] Tohru Kohda, Akio Tsuneda.Pseudonoise sequences by Chaotic nonlinear maps and their correlation properties[J].IEICE Trans, 1993, E77-B(8): 855-862.

[23] Lorenz E N, Deterministic nonperiodic flow [J].J Atmospheric Sci, 1993, 71(1): 130-141.

[24] 刘年生，郭东辉.混沌序列复杂性分析及其数值仿真的精度问题[J].集美大学学报：自然科学版，2005，10(3)：210-215.

[25] 张申如，王庭昌.混沌二进制序列构成的安全性研究[J].信息安全与通信保密，1995，4：42-46.

[26] Rukhin A, Soto J, Nechvatal J, et al.A statistical test suite for the validation of random number generators and pseudo random number generators for cryptographic applications[EB/OL].[2001-05-15]. http://csrc.nist.gov/rng/SP800-22b.pdf.

[27] 宋一中.等分布度与伪随机性检验[J].数理统计与管理，1994，13(5)：57-60.

[28] Schuster H G.Deterministic Chaos: an introduction[M].Weinheim: Physik-Verlag, 1984.

[29] 陈兆斗，张志刚.高维离散 Fourier 变换的一种快速算法[J].自然科学进展，1999，9(9)：780-782.

[30] 王云峰，沈海斌，严晓浪.输出-密文混和反馈混沌流密码的设计[J].浙江大学学报，2006，40(11)：1972-1975.

[31] 吴芝路，任广辉，赵楠，等.混沌扩频序列有限精度研究[J].哈尔滨商业大学学报，2006，22(1)：42-45.