

微熵率算法分析及实证研究

黄奕^{1,2}, 谢维波^{1,2}

(1. 华侨大学 计算机科学与技术学院, 福建 厦门 361021;

2. 华侨大学 厦门软件园嵌入式技术开放实验室, 福建 厦门 361008)

摘要: 利用微熵率法进行相空间重构,并以 Henon map 混沌特性的理论结果为依据,分别选取两组初值以验证 Henon 序列混沌特征的稳定性,实证研究微熵率算法的各个环节.实验结果表明:利用微熵率法重构相空间能够有效地捕捉混沌序列的特性.

关键词: 微熵率法;替代数据;Henon map;混沌特性;实证研究

中图分类号: TP 311

文献标志码: A

在时间序列的分析中,决定序列的可观测因素很多,且相互作用的动力学方程往往是非线性的.20 世纪 80 年代以来,由于 Takens 对 Whitney 早期在拓扑学方面工作的发展,使得深入分析时间序列的背景和动力学机制成为可能.在确定性的基础上对序列动力学因素的分析,目前广泛采用的是相空间重构法.微熵率法是 Gautam 等提出的一个基于样本时间序列及其替代数据的相空间重构方法^[1].熵率是指随机源(1 个随机过程)随时间变化的平均不确定性;1 个随机过程的熵率是该过程平均每产生 1 个随机变量其不确定度大小的度量.微熵率法中的替代数据方法为 iAAFT. iAAFT 是一种性能稳定的替代数据产生方法,能很好匹配原始数据的傅里叶幅度谱和概率密度分布,在数据的非线性检验中被广泛采用.本文以 Henon map 混沌系统为数据源,以 Henon map 混沌特性的理论为依据,实证研究微熵率算法的各个环节.

1 Henon map 混沌系统

1976 年,Michel Henon 给出了 Henon map 混沌系统^[2-3].此后,许多学者就 Henon map 混沌系统的全局结构进行数值与仿真研究,并将 Henon map 系统扩展到更高维^[4].

1.1 Henon 混沌序列

Henon 混沌序列为

$$x_k = 1 - ax_{k-d}^2 + bx_{k-2d}. \quad (1)$$

式(1)中: a, b 为系统参数; d 为系统延迟.当 $a=1.4, b=0.3$ 时,系统具有混沌特性.经过足够多次的迭代,Henon 序列呈现“混沌”现象.

Henon 混沌序列($d=1$),如图 1 所示.图 1 中:对应初值为($x_0=0.4, x_1=0.6$)和($x_0=0.4+10^{-8}, x_1=0.6$)的 Henon 混沌序列利用 Matlab 仿真的情况(实验均采用 Matlab 完成).

大约在 $0 < k < 40$ 的范围内, x_k 序列基本一致;当 $k > 40$ 之后,初始值微小扰动(10^{-8})的两组 x_k 序列呈现急剧的差异,体现了混沌序列“初值敏感性”的特征.

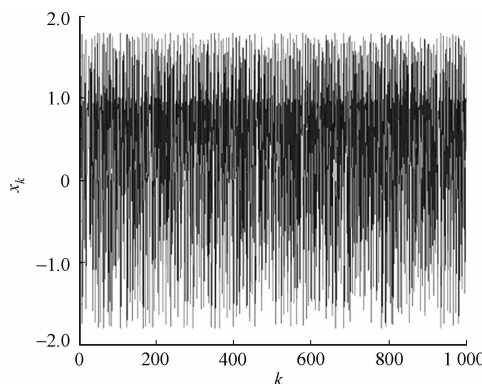


图 1 Henon 混沌序列

Fig. 1 Henon chaotic sequence

1.2 Henon map 奇异吸引子

奇异吸引子是混沌运动的主要特征. 令

$$y_k = x_{k-d}, \tag{2}$$

由式(1),(2)可得

$$x_k = 1 - ax_{k-d}^2 + by_{k-d}. \tag{3}$$

式(2),(3)构成了 Henon map 混沌系统^[5]. 点集 (x_k, y_k) 组成一条不封闭的曲线, 即 Henon map 混沌系统的奇异吸引子相图.

图 1 两种序列的奇异吸引子相图, 如图 2 所示. 由图 2 可知: 两种序列几乎“完全重合”. 尽管图 1 两种序列存在急剧的差异, 点集 (x_k, y_k) 却呈现“相同的轨迹”. 显然, 由图 1 可知: 对应点出现的“顺序”是不同的, 但是轨迹却几乎“完全重合”. 奇异吸引子给出的“确定性轨迹”, 体现了 Henon map 混沌系统的稳定性, 奠定了混沌系统发展规律的研究基础.

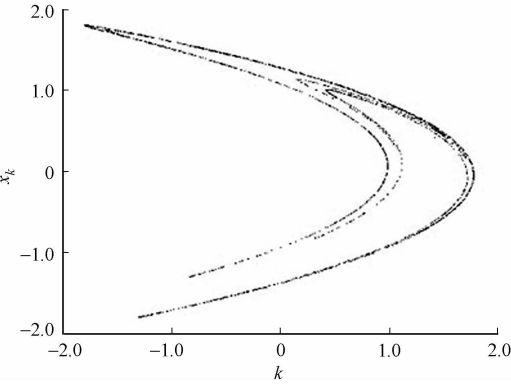


图 2 Henon map 混沌系统奇异吸引子相图
Fig. 2 Henon map chaotic system attractor

2 替代数据

为实现时间序列分析的统计学检验, 时间序列需要足够多样本. 样本的获取, 可通过构造产生时间序列的系统, 或者直接构造时间序列本身. 后者产生时间序列样本的方法, 称为替代数据法^[6]. 替代数据能够尽可能精确地复制原始数据的性质(包括时间概率分布和自相关函数), 但同时又是尽量随机的.

2.1 AAFT 生成算法

振幅调节傅里叶变换(AAFT)有以下 4 个步骤^[7].

- 步骤 1 原始数据 $\{x_n^0\}$, 排列序号 $\{\text{rank}_n^0\}$, 当 x_n^0 是 $\{x_n^0\}$ 中第 k 小时, rank_n^0 .
- 步骤 2 高斯白噪声 $\{g_0\}$, 按 $\{\text{rank}_n^0\}$ 重排得 $\{g'_n\}$.
- 步骤 3 对 $\{g'_n\}$ 傅里叶变换和相位随机化处理, 得序列 $\{s'_n\}$.
- 步骤 4 求 $\{s'_n\}$ 的排列序号 $\{\text{rank}_n^s\}$, 原始数据 $\{x_n^0\}$ 按 $\{\text{rank}_n^s\}$ 重排得替代数据 $\{s_n\}$.

替代数据 $\{s_n\}$ 是原始数据 $\{x_n^0\}$ 的重排, 保证了替代数据与原始数据有相同的时间概率分布. 因此, 也有一样的均值方差等一、二阶统计量. 替代数据 $\{s_n\}$ 的随机性体现在高斯白噪声 $\{g_n\}$, 及其傅里叶变换的相位随机化处理.

然而, 原始数据功率谱密度(自相关函数)的性质被改变了, 因为发生在步骤 2 和步骤 4 的两个重排在严格意义上不是彼此的逆操作. 这种差异导致替代数据的功率谱密度在原始数据的基础上被白化. 二者的差异程度取决于原始数据的时间概率分布与高斯分布的相似程度, 简而言之, AAFT 算法适用于类高斯分布的时间序列.

2.2 iAAFT 生成算法

为了解决 AAFT 替代数据的功率谱白化问题, 1996 年 Schreiber 提出了 AAFT 迭代生成算法(iAAFT). iAAFT 是一种性能稳定的替代数据产生方法, 能很好匹配原始数据的傅里叶频谱和概率密度分布, 在数据的统计学检验中被广泛采用. 具体有以下 4 个步骤^[8].

- 步骤 1 原始数据 $\{x_n^0\}$, 排列序号 $\{\text{rank}_n^0\}$, 傅里叶变换的幅度值 $|X_k|$, 计算 AAFT 替代数据 $\{s_n\}$.
 - 步骤 2 记 $\{s_n\}$ 傅里叶变换 $S_k = |S_k| \exp(j\varphi(k))$, 保持相位不变, 幅度值用 $|X_k|$ 代替, 得 $S'_k = |X_k| \exp(j\varphi(k))$.
 - 步骤 3 对 S'_k 进行傅里叶反变换得 $\{s'_n\}$, 再按 $\{\text{rank}_n^0\}$ 重排得 $\{s''_n\}$.
 - 步骤 4 重复步骤 2, 3, 直至所得数据 $\{s''_n\}$ 和原始数据有相近的功率谱密度.
- 确保 $\{s''_n\}$ 和原始数据的统计性质(时间概率分布、功率谱密度)趋于一致, 该算法有两个基本假设.
- 1) 步骤 2 对傅里叶变换幅度值的矫正, 造成时间概率分布的扭曲都比上一次迭代的小.
 - 2) 步骤 3 的重排, 造成功率谱密度的扭曲都比上一次迭代的小.

Schreiber 等^[7]证明:对于非线性自相关过程,替代数据的功率谱将会逐渐趋近原数据的功率谱.

3 微熵率法确定时间序列最佳嵌入参数

3.1 微熵率法

时间序列可观测性的决定因素很多,其相互作用的动力学方程往往是非线性的,甚至是混沌的.同时,计算的复杂性、有限的测量精度,以及可能存在的本质上的非确定性等多方面困难,严重制约着人们对时间序列内在机制的理解.20 世纪 80 年代,Takens^[9]对 Whitney 早期在拓扑学方面工作的发展,为深入分析时间序列的背景和动力学机制奠定基础.在确定性的基础上,对序列动力学因素的分析,目前广泛采用的是延迟坐标状态空间重构法(相空间重构法)^[10].一般来说,非线性系统的相空间可能维数很高,甚至无穷,在大多数情况下维数并不知道.

对于给定的时间序列,相空间重构法表明存在一个最优的嵌入维数 m 和时延 τ .如果 τ 太小,为了使 $m \cdot \tau$ 覆盖(大于)“捕捉信号的动力学性质所需的最小时间距”, m 将变得相当大;相反,如果 τ 大于最佳值,模型的性质将变得太离散,导致捕捉不到信号的动力学性质.Gautama 等提出基于样本时间序列及其替代数据的微熵率方法,用于确定相空间的最佳嵌入维数 m 和时延 τ .

3.2 微熵率法详细步骤

微熵率法有以下 4 点详细步骤^[1].

步骤 1 原始数据 $\{x(k) : k=1,2,\dots,N\}$, 计算其 N_s 组替代数据,记为

$$\{x_{s,i}(k) : k=1,2,\dots,N\} : i=1,2,\dots,N_s.$$

步骤 2 嵌入维数 m 和时延 τ 的相空间.原始数据的相空间为

$$\begin{bmatrix} x(1) & x(2) & \cdots & x(N-(m-1)\tau) \\ x(1+\tau) & x(2+\tau) & \cdots & x(N-(m-2)\tau) \\ \vdots & \vdots & \vdots & \vdots \\ x(1+(m-1)\tau) & x(2+(m-1)\tau) & \cdots & x(N) \end{bmatrix}.$$

相空间的状态向量记为

$$\mathbf{X}(k) = [x(k) \quad x(k+\tau) \quad \cdots \quad x(k+(m-1)\tau)]^T, \quad k=1,2,\dots,N-(m-1)\tau.$$

替代数据的相空间(共 N_s 个),即

$$\begin{bmatrix} x_{s,i}(1) & x_{s,i}(2) & \cdots & x_{s,i}(N-(m-1)\tau) \\ x_{s,i}(1+\tau) & x_{s,i}(2+\tau) & \cdots & x_{s,i}(N-(m-2)\tau) \\ \vdots & \vdots & \vdots & \vdots \\ x_{s,i}(1+(m-1)\tau) & x_{s,i}(2+(m-1)\tau) & \cdots & x_{s,i}(N) \end{bmatrix}.$$

相空间的状态向量记为

$$\mathbf{X}_{s,i}(k) = [x_{s,i}(k) \quad x_{s,i}(k+\tau) \quad \cdots \quad x_{s,i}(k+(m-1)\tau)]^T, \quad k=1,2,\dots,N-(m-1)\tau.$$

其中: $\mathbf{X}(k)$ 和 $\mathbf{X}_{s,i}(k)$ 又称为延迟矢量,延迟矢量的个数为 $M=N-(m-1)\tau$.

步骤 3 基于原始数据及其替代数据,确定原始数据 $\{x(k) : k=1,2,\dots,N\}$ 的熵率为

$$R_{\text{ent}}(m, \tau) = I(m, \tau) + \frac{m}{M} \ln M.$$

其中: $I(m, \tau) = H(x, m, \tau) / (H(x_{s,i}, m, \tau))_i$. 由原始数据构成的相空间,其熵为

$$H(x, m, \tau) = \sum_{j=1}^M \ln(M\rho_j) + \ln 2 + C_E = \sum_{j=1}^{N-(m-1)\tau} \ln(\rho_j) + M \ln M + \ln 2 + C_E.$$

其中:欧拉常数 $C_E = 0.5772$; ρ_j 是第 j 个延迟矢量与其最近邻点的欧氏距离,即

$$\rho_j = \min\{\|X(j) - X(p)\|_2, p=1,2,\dots,M, p \neq j\}.$$

相应地有

$$H(x_{s,i}, m, \tau) = \sum_{j=1}^M \ln(M\rho_{s,i,j}) + \ln 2 + C_E = \sum_{j=1}^{N-(m-1)\tau} \ln \rho_{s,i,j} + M \ln M + \ln 2 + C_E,$$

$$\rho_{s,i,j} = \min\{\|X_{s,i}(j) - X_{s,i}(p)\|_2, p=1,2,\dots,M, p \neq j\}.$$

由 N_s 组替代数据构成的 N_s 个相空间,其平均熵为 $(H(x_{s,i},m,\tau))_i = \frac{1}{N_s} \sum_{i=1}^{N_s} H(x_{s,i},m,\tau)$. 可见,熵率体现为原始数据相空间的熵与 N_s 个替代数据相空间平均熵之比.

步骤 4 计算 $R_{\text{ent}}(m,\tau)$ 的最小值,相应的 m 和 τ 即最佳嵌入维数 m_{opt} 和时延 τ_{opt} .

4 数值实验

4.1 实验数据的选取

Henon 序列混沌特性的理论结果:当 $a=1.4, b=0.3$ 时,式(1)具有混沌特性,其参数 d 为时间延迟,对应最佳嵌入时延 τ_{opt} .

不当的初值选择会造成式(1)的发散,分别选取两组初值以验证 Henon 序列混沌特征的稳定性的. 经过足够多次的迭代, Henon 序列才会进入“稳定的”混沌状态^[11]. 取 Henon 序列 10 000 个数据后的 500 个作为数据源,才得出“稳定的”结果.

4.2 实验结果

Henon 序列的时延 $d=4$,初值取 $(x_0=0.2, x_1=0.6, x_2=0.4, x_3=0.3, x_4=0.1, x_5=0.5, x_6=0.8, x_7=0.7)$ 和 $(x_0=0.2+10^{-8}, x_1=0.6, x_2=0.4, x_3=0.3, x_4=0.1, x_5=0.5, x_6=0.8, x_7=0.7)$,生成两组上述的数据源($N=500$),应用微熵率法分别求最佳嵌入维数 m_{opt} 和时延 τ_{opt} . 其中: N_s 的选取,以 $(H(x_{s,i},m,\tau))_i$ 趋于稳定为准,文中取 $N_s=10$.

图 3,表 1,2 对应 $d=4$ 的 Henon 序列. 图 3 和表 1 对应的初值取 $(x_0=0.2, x_1=0.6, x_2=0.4, x_3=0.3, x_4=0.1, x_5=0.5, x_6=0.8, x_7=0.7)$ 的结果. 表 2 对应的初值取 $(x_0=0.2+10^{-8}, x_1=0.6, x_2=0.4, x_3=0.3, x_4=0.1, x_5=0.5, x_6=0.8, x_7=0.7)$ 的结果. 从表 1,2 可以看出: $m=3, \tau=4$ 时,熵率 $R_{\text{ent}}(m,\tau)$ 均取得最小值,即最佳嵌入维数 $m_{\text{opt}}=3$ 和时延 $\tau_{\text{opt}}=4$ (对应 $d=4$),实证了不同的初值下 Henon 序列混沌特征的稳定性的.

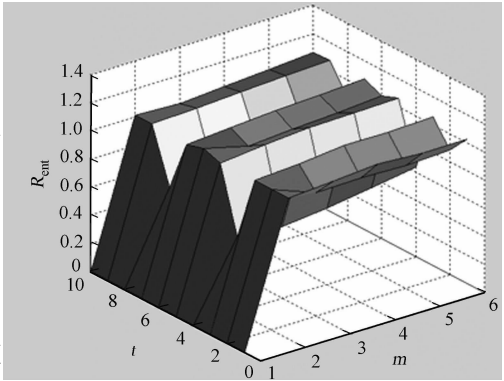


图 3 $R_{\text{ent}}(m,\tau)$ 三维示意图
Fig. 3 3D-diagram of $R_{\text{ent}}(m,\tau)$

表 1 第一组初值对应的 R_{ent}
Tab. 1 R_{ent} of the first set initial value

$m \backslash \tau$	1	2	3	4	5	6	7
2	1.039 4	1.045 9	1.036 9	0.515 5	1.022 4	1.028 3	1.021 1
3	1.037 9	0.924 5	1.041 4	0.490 3	1.026 5	0.986 5	1.042 4
4	1.051 6	0.931 3	1.053 2	0.521 6	1.053 5	1.016 0	1.054 0
5	1.032 6	0.920 1	1.054 0	0.563 4	1.060 8	1.028 2	1.064 7
6	1.035 9	0.929 1	1.065 6	0.613 0	1.076 1	1.043 3	1.079 2
7	1.044 8	0.944 8	1.078 8	0.663 8	1.088 3	1.063 1	1.091 1
8	1.056 3	0.966 2	1.091 6	0.716 6	1.105 5	1.082 1	1.106 7

表 2 第二组初值对应的 R_{ent}
Tab. 2 R_{ent} of the second set initial value

$m \backslash \tau$	1	2	3	4	5	6	7
2	1.009 3	1.028 6	1.017 6	0.468 4	1.005 5	1.010 0	1.021 7
3	1.041 1	0.915 7	1.040 8	0.455 0	1.042 0	0.983 4	1.034 5
4	1.045 1	0.920 1	1.054 0	0.492 7	1.052 2	0.999 4	1.044 6
5	1.030 6	0.909 0	1.051 7	0.534 5	1.055 1	1.009 5	1.060 5
6	1.031 9	0.916 5	1.062 2	0.580 7	1.066 2	1.031 0	1.074 2
7	1.040 8	0.929 3	1.076 2	0.630 2	1.082 8	1.050 3	1.090 1
8	1.052 5	0.949 8	1.089 9	0.681 5	1.099 4	1.070 6	1.106 7

对应 Henon 序列的不同时延 d ,表 3 给出基于微熵率的辨识结果(初值取 $[0,1]$ 的随机数),包括 $m_{\text{opt}}, \tau_{\text{opt}}, R_{\text{ent}}(m_{\text{opt}}, \tau_{\text{opt}})$. 由表 3 可知: m_{opt} 的值恒为 3, τ_{opt} 的值始终与 d 保持一致;有效地验证了 Henon

序列混沌特征的稳定性,这种稳定性奠定了混沌系统的研究基础.

表 3 不同时延的微熵率法辨识
Tab. 3 Differential entropy of different time delay

d	m_{opt}	τ_{opt}	$R_{\text{ent}}(m_{\text{opt}}, \tau_{\text{opt}})$	d	m_{opt}	τ_{opt}	$R_{\text{ent}}(m_{\text{opt}}, \tau_{\text{opt}})$
1	3	1	0.463 1	4	3	4	0.496 8
2	3	2	0.460 7	5	3	5	0.488 3
3	3	3	0.494 1	6	3	6	0.523 5

参考文献：

[1] GAUTAMA T,MANDIC D P,VANHULLE M M. A differential entropy based method for determing the optimal embedding parameters of signal[C]//Processing of the Int Conf on a Coustics, Speech and Signal Processing. Hong Kong:[s. n.],2003;29-32.

[2] 王光义,郑艳,刘敬彪. 一个超混沌 Lorenz 吸引子及其电路实现[J]. 物理学报,2007,56(6):3113-3120.

[3] GALLAS J A C. Structure of the parameter space of the Hénon map[J]. Physical Review Letters,1993,70(18): 2714.

[4] HENON M. A two dimensional mapping with stranger attractor[J]. Commun Math Phys,1976,50(1):69-77.

[5] BENEDICKS M,CARLESON L. The dynamics of the Henon map[J]. Annals of Mathematics. 1991,163(3):749-841.

[6] THEILER J,EUBANK S,LONGTIN A,et al. Testing for nonlinearity in time series: The method of surrogate data [J]. Physica D: Nonlinear Phenomena,1992,58(1):77-94.

[7] SCHREIBER T,SCHMITZ A. Improved surrogate data for nonlinearity tests[J]. Physical Review Letters,1996,77 (4):635.

[8] KEYLOCK C J. Constrained surrogate time series with preservation of the mean and variance structure[J]. Physical Review E,2006,73(3):036707.

[9] 王海燕,盛昭瀚. 混沌时间序列相空间重构参数的选取方法[J]. 东南大学学报:自然科学版,2000,30(5):113-117.

[10] TAKENS F. Detecting strange atractors in turbulence[J]. Lecture Notes in Mathematics,1982,898(12):198-366.

[11] GRASSBERGER P,KANTZ H,MOENIG U. On the symbolic dynamics of the Hénon map[J]. Journal of Physics A:Mathematical and General,1989,22(24):5217.

Analysis and Empirical Research of the Entropy Ratio Method

HUANG Yi^{1,2}, XIE Wei-bo^{1,2}

(1. College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China;
2. Open Laboratory of Embedded Technology, Xiamen Software Park, Huaqiao University, Xiamen 361021, China)

Abstract: Using differential entropy ratio method to reconstruct phase space and based on the theoretical results of the chaotic characteristic of Henon map, two groups of initial value were selected to verify the stability of the chaotic characteristics of Henon sequence. In this process, all aspects of differential entropy ratio algorithm was demonstrated either. The experimental results showed that using differential entropy ratio method to reconstruct phase space can effectively capture the characteristic of the the chaotic sequences.

Keywords: differential entropy ratio method; surrogate data; Henon map; chaotic characteristics; empirical research

(责任编辑: 钱筠 英文审校: 吴逢铁)