

基于符号分析的极大联合熵延迟时间求取方法*

张淑清[†] 李新新 张立国 胡永涛 李亮

(燕山大学电气工程学院, 河北省测试计量技术及仪器重点实验室, 秦皇岛 066004)

(2012年11月30日收到; 2013年2月4日收到修改稿)

本文通过符号分析法求取联合熵的极大值点, 进而得到相空间重构的最佳延迟时间, 通过对几个典型的混沌系统进行数值仿真试验, 结果表明, 该方法简化了计算, 提高了效率, 能够准确快速地获得最佳延迟时间, 从而有效地重构原系统的相空间, 为混沌信号识别提供一种快速有效的途径.

关键词: 相空间重构, 延迟时间, 联合熵, 符号分析

PACS: 05.45.Rq, 05.40.-a, 02.50.Fz

DOI: 10.7498/aps.62.110506

1 引言

混沌自从被确认是长期可控制的和短期可预测以后, 其研究课题引起了大家广泛的兴趣. 然而相空间重构作为混沌时间序列处理中的重要课题, 其最佳延迟时间的选取至关重要. 对于时间序列 $X(n)$, 过小的延迟时间 τ 将导致信息的冗余, 使得重构的坐标之间不可分辨, 过大的 τ 将使延迟坐标之间毫不相关, 不能代表真实的动力系统^[1], 因此选取最佳延迟时间需要使 $X(n)$ 和 $X(n+\tau)$ 最大限度相互独立却又不是完全无关.

确定相空间重构最佳延迟时间的方法有互信息函数法、自相关函数法和平均位移法等. 其中互信息函数方法是估计重构相空间延迟时间的一种有效方法, 它在相空间重构中有很广泛的应用. Shaw 首先提出互信息第一次达到最小时滞时作为相空间重构的最佳延迟时间^[2], Faster 给出了互信息计算的递归算法. 但其提出的划分网格计算方法过于复杂, 不易于编程实现及推广使用. 国内学者提出了相应的改进算法并获得很好的应用, 这在一定程度上激发了人们对准确快速地选取相空间重构中最佳延迟时间的兴趣^[3]. 文献 [4] 提出了确定延迟时间变量的联合熵之第一极大准则, 相对互信息方法减少了计算量, 但在求取联合熵时采用划分

网格的方法, 计算相对复杂而且误差较大; 文献 [5, 6] 提出了改进的网格标记法, 该方法与符号分析法相比计算仍然相对复杂; 文献 [7] 提出的符号分析法求取互信息函数, 从而确定出最佳延迟时间, 可避免复杂的网格划分和标记, 但需同时求取 $X(n)$ 和 $X(n+\tau)$ 的各自信息熵及其联合熵. 本文提出了基于符号分析的极大联合熵延迟时间求取方法, 较好地解决了上述方法中所存在的各种问题, 可以快速有效地提取混沌动力系统中有用定量信息, 重构系统的相空间.

2 最佳延迟时间的极大联合熵准则

考虑时间序列 $X(n)$ 及其延迟时间序列 $X_\tau(n) = X(n+\tau)$, $n = 1, 2, \dots, N$. 根据互信息函数的递推公式^[8], 两组序列的互信息可表示为

$$\begin{aligned} I(X, X_\tau) &= H(X) - H(X|X_\tau) \\ &= H(X) + H(X_\tau) - H(X, X_\tau) \\ &= I(X_\tau, X), \end{aligned} \quad (1)$$

其中 X 代表时间序列 $X(n)$, X_τ 代表其延迟序列 $X(n+\tau)$, $H(X)$ 是孤立的 X 的不定性, $H(X|X_\tau)$ 是已知 X_τ 的 X 的不定性, 所以 X_τ 的已知减少了 X 的不定性, 则 $I(X, X_\tau)$ 的第一极小值处的 τ 即为

* 国家自然科学基金(批准号: 61077071, 51075349)和河北省自然科学基金(批准号: F2011203207)资助的课题.

[†] 通讯作者. E-mail: zhshq-yd@163.com

最佳延迟时间. 并且为了计算 $I(X, X_\tau)$, Fraser 和 Swinney 提出了复杂的划分网格的方法^[2].

互信息 $I(X, X_\tau)$ 表征 X 和 X_τ 的相关程度, 相空间重构最佳延迟时间的确定是在保证重构坐标之间最大限度地相互独立的基础上提出来的. $I(X, X_\tau)$ 取极小值时, X 和 X_τ 的相关度也极小, 即 X 和 X_τ 的联合整体不确定度达到极大.

另一方面, 为了保证重构坐标之间最大限度地相互独立, 应使 X 和 X_τ 联合整体的不确定性达到最大. 由信息理论可知, 联合熵 $H(X, X_\tau)$ 是 X 和 X_τ 的联合整体的不确定性的度量, $H(X, X_\tau)$ 越大, 则 X 和 X_τ 联合整体的不确定性也越大. 由此推断, 联合熵 $H(X, X_\tau)$ 的第一个极大值点即为相空间重构的最佳延迟时间点^[4].

同时, 混沌吸引子在其不变集具有的遍历特性表现在其状态变量 $X(n)$ 上, 即在稳态情况下, 对任意的延迟时间 τ , $X(n)$ 和 $X(n+\tau)$ 都具有相同的概率分布和统计特性, 则 $H(X)$ 和 $H(X_\tau)$ 的值接近于常数, 根据互信息函数 (1) 式可得

$$\begin{aligned} H(X, X_\tau) &= H(X) + H(X_\tau) - I(X, X_\tau) \\ &= \text{Const} - I(X, X_\tau). \end{aligned} \quad (2)$$

可见, $H(X, X_\tau)$ 和 $I(X, X_\tau)$ 呈近似相反的变化规律. 互信息的极小值点即为联合熵的极大值点在理论上得到验证.

因此通过复杂的划分网格或者进行网格标记求取互信息第一极小值点^[6]从而确定相空间重构最佳延迟时间可以转换为求取联合熵 $H(X, X_\tau)$ 的第一极大值点. 联合熵的计算公式^[9]可以表示为

$$\begin{aligned} H(X, X_\tau) &= - \sum_{i=1}^N \sum_{j=1}^N P(x_i, x_{\tau_j}) \log_2 P(x_i, x_{\tau_j}). \end{aligned} \quad (3)$$

由此可见, 求取不同延迟时间下的联合熵, 找出联合熵的第一极大值点所对应的 τ , 即为所求的最佳延迟时间点.

3 符号分析法求取极大联合熵

符号分析法实际上是时间序列的符号化, 该方法是通过在几个可能值上将其时间序列离散化, 从而将千变万化的数据序列转换为几个数值或特殊符号的符号序列, 该过程称为粗粒化过程, 这一过程能够从动力系统中有效快速地捕获有用定量信息, 并且能够降低噪声的影响. 将混沌时间序列通

过粗粒化方法转化为符号序列^[10-14], 则其轨道演化的信息就通过符号序列达到了编码.

时间序列符号化有二进制符号化规则、角区间符号化规则和概率统计符号化规则等, 二进制划分^[13,15]是最简单的一种划分规则, 只需给定一个阈值 P_0 , 时间序列中大于此阈值 P_0 的取 1, 否则取 0. 阈值 P_0 的选取有均值, 零值等. 则时间序列 $X(n)$ 最终被转换成符号序列 $\{S(n)\}$, 所得符号序列表征了两种数据元模式, 即

$$S(n) = \begin{cases} 1, & X(n) > P_0, \\ 0, & X(n) \leq P_0, \end{cases}$$

如图 1 所示.

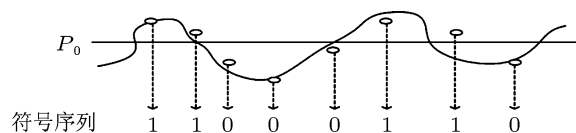


图 1 二进制符号化规则

二进制分割虽然简便, 但是由于其划分比较粗糙, 容易导致原始信息的遗失. 为了解决这个问题, 本文通过对相空间的分割从而使时间序列 $X(n)$ 实现粗粒化过程^[10-14]. 以符号数取 d 为例, 划分如下:

$$S_i = \begin{cases} 0, & X_{C_0} \leq X_n < X_{C_1}, \\ 1, & X_{C_1} \leq X_n < X_{C_2}, \\ 2, & X_{C_2} \leq X_n < X_{C_3}, \\ \vdots & \\ d-1, & X_{C_{d-1}} \leq X_n < X_{C_d}, \end{cases} \quad (4)$$

其中 X_n 为时间序列 $X(n)$ 在 n 时刻的幅值大小, $n = 1, 2, \dots, N$. $X_{C_0}, X_{C_1}, \dots, X_{C_d}$ 为阈值即临界点, S_i 为符号序列的符号集, 并用相应的整数 i 来代替, $i = 0, 1, \dots, d-1$. 则时间序列 $\{X(1), X(2), X(3), \dots, X(N)\}$ 可以转换成整数集序列 $\{S(1), S(2), S(3), \dots, S(N)\}$.

时间序列被转化成长整数集序列 $\{S(1), S(2), \dots, S(N)\}$ 后, 要提取有用信息, 需选择标准分割长度 L . 将长整数集序列分割成长度为 L 的短序列, L 个连续符号被编码为十进制数, 形成了新的整数集序列 $L_x(L, k) = \sum_{j=1}^L d^{L-j} S(j+k-1)$, k 表示短序列从长整数集序列 S_i 的第 k 个元素 $S(k)$ 开始^[10].

按照上式进行标记和辨识^[10-12], 则离散的符号替代连续数据的原始时间序列, 如图 2 所示.

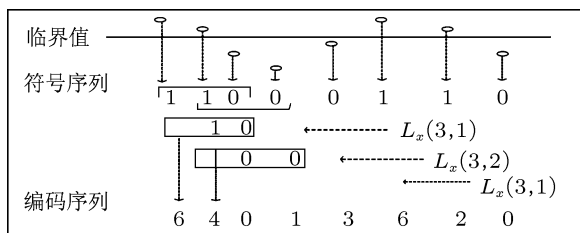


图 2 符号化时间序列图

符号数 d 的个数可以通过使符号熵最大化来寻找^[11,12], 为了方便起见, 将混沌序列的临界点 $(d+1)$ 的个数置为 11, 即 $d=10$. 且分割长度 L 取 2.

对时间序列 $X(n)$ 及其延迟时间序列 $X(n+\tau)$, $n=1, 2, \dots, N$, 根据上面所介绍的粗粒化符号方法将其编码成能捕获用定量信息的特殊的十进

制数序列 $L_x(n)$ 和 $L_{x\tau}(n)$, 则各个特殊十进制数出现的频率为时间序列分析的指标, 即为联合概率 $P(L_x, L_{x\tau})$, 则符号分析法求取的联合熵^[16](3) 式可以改写为

$$H(L_x, L_{x\tau}) = - \sum_{i=1}^N \sum_{j=1}^N P(L_{x_i}, L_{x_{\tau_j}}) \log_2 P(L_{x_i}, L_{x_{\tau_j}}). \quad (5)$$

4 数值仿真试验

为了验证该方法求取最佳延迟时间的准确性和优越性, 我们分别对 Lorenz, Rossler 和 Duffing 三种常见的混沌时间序列求取最佳延迟时间并画出其吸引子的重构.

Lorenz 系统的数值仿真实验如图 3 所示, Lorenz 系统^[17] 方程

$$x = \sigma(y-x), \quad y = rx - y - xz, \quad z = -bz + xy, \quad (6)$$

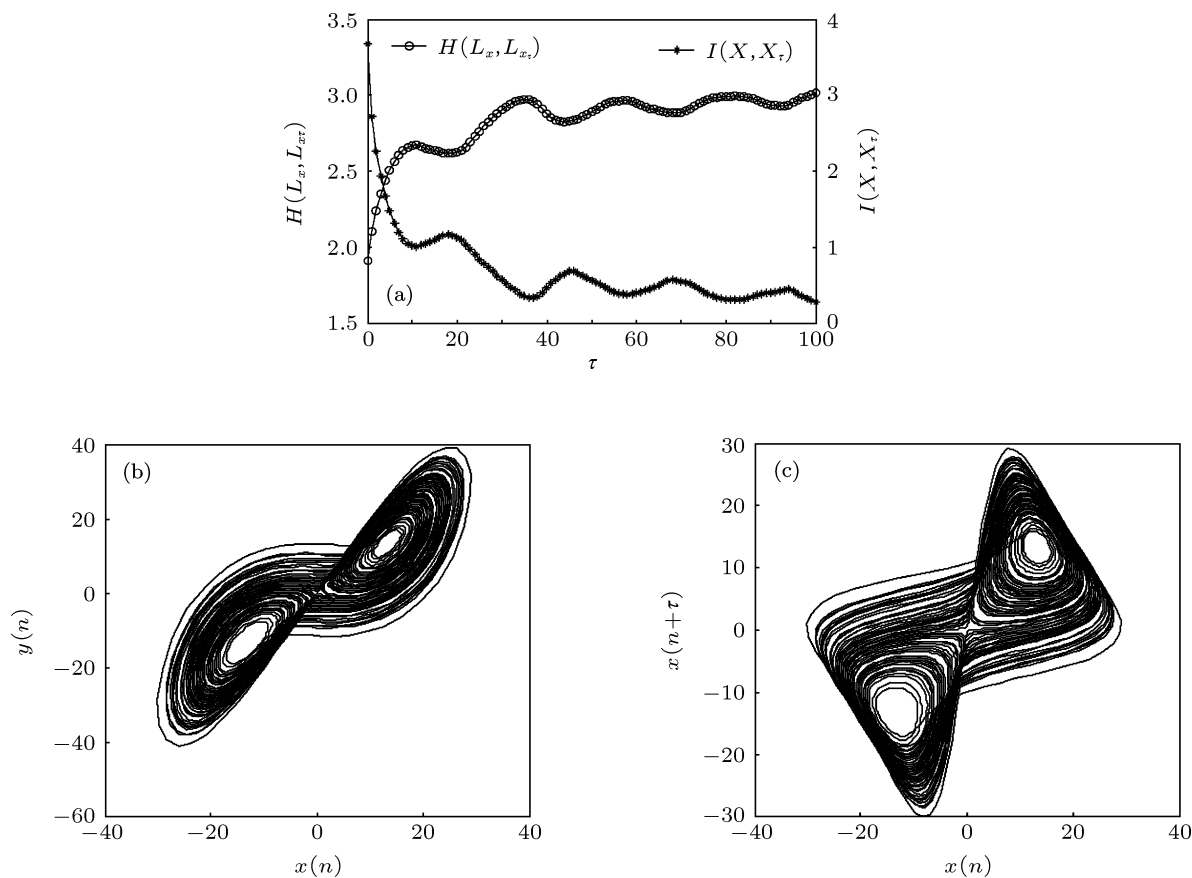


图 3 Lorenz 系统 ($\sigma = 16, b = 4, r = 45.92$) 的数值仿真实验 (a) 本文方法和互信息函数法求取最佳延迟时间的对比图; (b) Lorenz 吸引子在 x - y 平面上的投影图; (c) $\tau = 11$ 时的 Lorenz 的重构吸引子图

其中取 $\sigma = 16, b = 4, r = 45.92$, 此系统为混沌系统, 用 Runge-Kutta 法求解方程 (6), 步长 $h = 0.01$, 取变量 x 为研究对象, 去除前 8000 个点, 得到一个 7000 个点的时间序列. 通过符号分析法求取联合熵 $H(L_x, L_{x_\tau})$ 的极大值点和通过网格划分求取互信息 $I(X, X_\tau)$ 的极小值点的对比图形如图 3(a) 所示, 从图 3(a) 可以看出, $H(L_x, L_{x_\tau})$ 和 $I(X, X_\tau)$ 的变化规律近似相反, 并且都是在 $\tau = 11$ 时取得第一个所对应的极值点, 故可得由联合熵第一极大值点法求得的相空间重构最佳延迟时间达到了与互信息法相同的效果; 试验过程中也大大地简化了其计算量. 图 3(b) 是 Lorenz 吸引子在 x - y 平面的投影图; 图 3(c) 是 $\tau = 11$ 时的重构吸引子图. 从图中可以看出 $\tau = 11$ 时重构图能够很好的反映出 Lorenz 系统的双圈拓扑结构. 从而验证了符号分析法求取极大联合熵进而确定最佳延迟时间的有效性和优越性.

Rosler 系统的数值仿真实验如图 4 所示, Rosler 系统^[18] 方程

$$\dot{x} = -(y+z), \quad \dot{y} = x+dy, \quad \dot{z} = e+z(x-f), \quad (7)$$

其中取 $d = 0.2, e = 0.2, f = 0.5$, 此时 Rosler 系统为混沌系统, 用 Runge-Kutta 法求解方程 (7), 步长 $h = 0.05$, 取变量 x 为研究对象, 去除前 50000 个点, 得到一个 8000 个点的时间序列. 图 4(a) 是 Rosler 系统处于混沌状态时的联合熵 $H(L_x, L_{x_\tau})$ 和互信息 $I(X, X_\tau)$ 的变化规律图, 两者的变化成近似相反的规律, 并且在互信息取得极小值点处, 联合熵处于极大值点, 即 $\tau = 23$; 图 4(b) 是 Rosler 吸引子在 x - y 平面的投影图; 图 4(c) 是 $\tau = 23$ 时的重构吸引子图. 比较重构前后吸引子在 x - y 平面投影, 发现重构吸引子很好地保持 Rosler 吸引子的单圈结构. 从而证明了符号分析法求取联合熵极大值点即为最佳延迟时间的可行性.

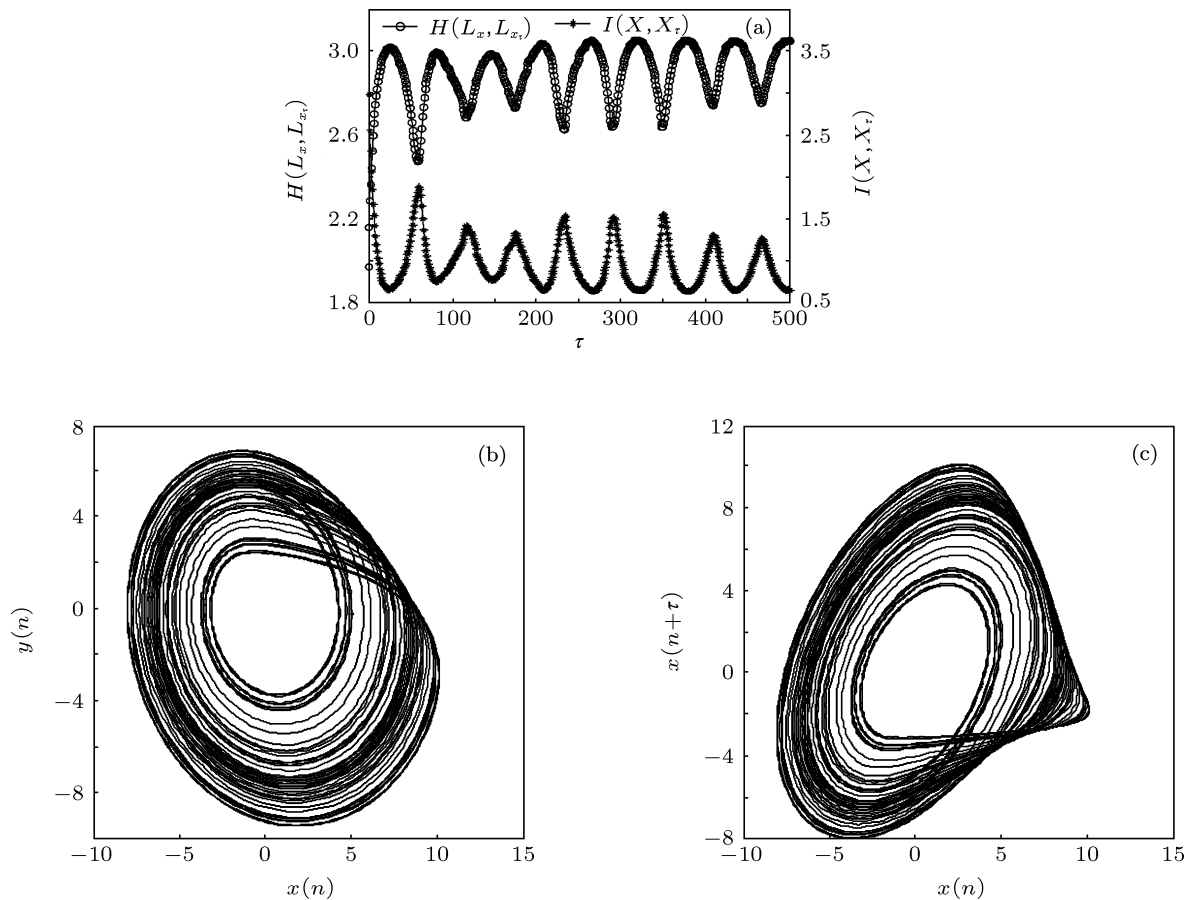


图 4 Rosler 系统 ($d = 0.2, e = 0.2, f = 0.5$) 的数值仿真实验 (a) 本文方法和互信息函数法求取最佳延迟时间的对比图; (b) Rosler 吸引子在 x - y 平面上的投影图; (c) $\tau = 23$ 时的 Rosler 的重构吸引子图

Duffing 系统的数值仿真实验如图 5 所示, Duffing 系统^[1] 方程

$$\begin{aligned} \dot{x} &= y, \\ \dot{y} &= -\sigma + ax(1-x^2) + F \cos(t), \end{aligned} \quad (8)$$

其中取 $\sigma = 0.06$, $a = 0.7$, $F = 7.5$, 此时 Duffing 系统为混沌系统, 用 Runge-Kutta 法求解方程 (8), 步长 $h = 0.05$, 取变量 x 为研究对象, 舍去开始暂态

点, 得到一个 7000 个点的时间序列. 从图 5(a) 中能够明显看出 Duffing 系统此时的联合熵 $H(L_x, L_{x_\tau})$ 和互信息 $I(X, X_\tau)$ 的近似相反的变化规律, 且由符号分析法求得的联合熵 $H(L_x, L_{x_\tau})$ 的极大值点为 $\tau = 11$, 与互信息函数法得到的最佳延迟时间 τ 相同. 图 5(b) 是 Duffing 吸引子在 x - y 平面的投影图; 图 5(c) 是 $\tau = 11$ 时的重构吸引子图. 比较重构前后吸引子在 x - y 平面投影发现重构效果良好.

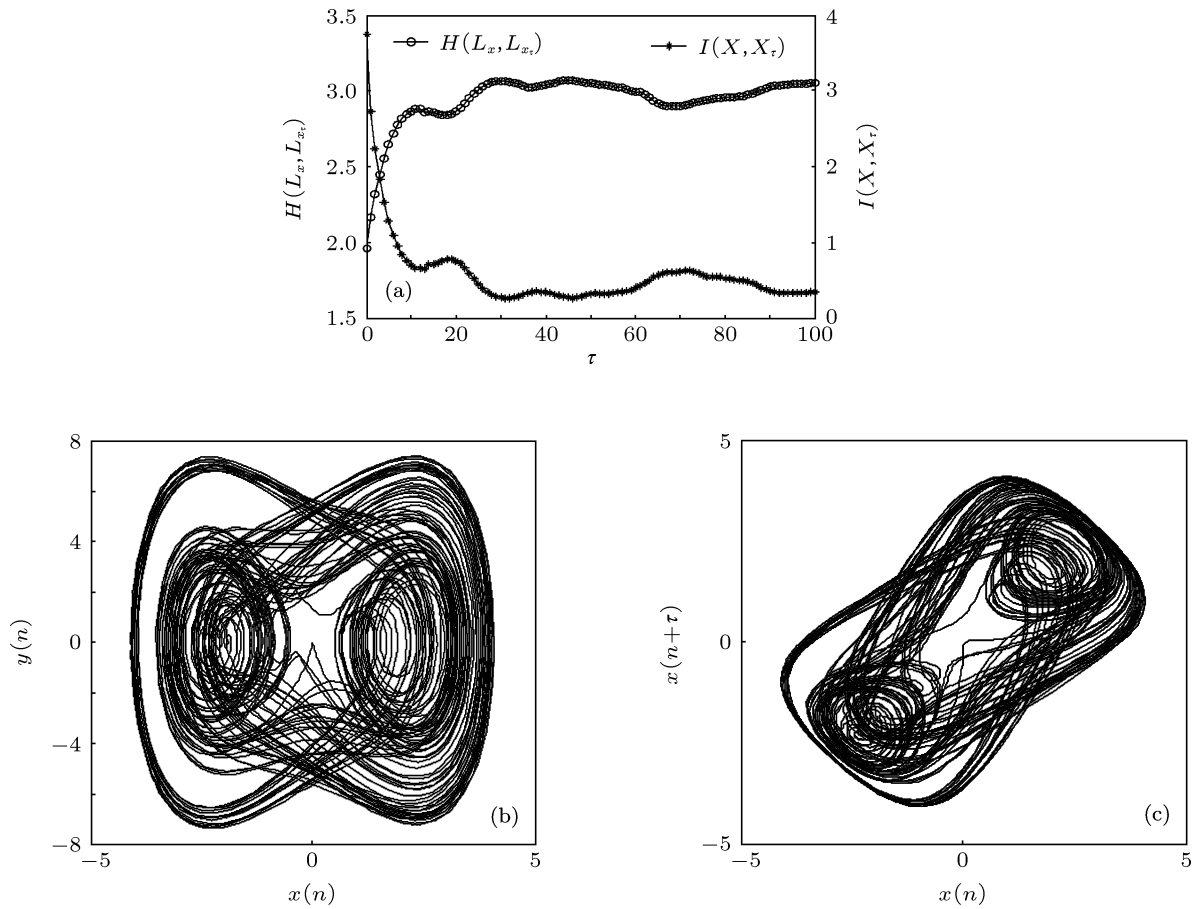


图 5 Duffing 系统 ($\sigma = 0.06$, $a = 0.7$, $F = 7.5$) 的数值仿真实验 (a) 本文方法和互信息函数法求取最佳延迟时间的对比图; (b) Duffing 吸引子在 x - y 平面上的投影图; (c) $\tau = 11$ 时的 Duffing 系统的重构吸引子图

5 结论

采用符号分析与极大联合熵结合求取相空间重构的最佳延迟时间, 不仅避免了计算互信息时的复杂的网格标记, 而且无需计算 $X(n)$ 和 $X(n + \tau)$ 的

各自的概率分布及信息熵. 该方法能达到与互信息方法计算最佳延迟时间同样的效果, 而且极大地简化了计算, 是求取最佳延迟时间的改进算法. 理论分析与仿真实验证明, 该方法物理意义明显, 实际效果较好, 能够更多地保留原系统的动力学特征.

- [1] Jiang W L, Zhang S Q, Wang Y Q 2005 *Chaos and Wavelet Based Fault Information Diagnosis* (Beijing: National Defence Industry Press) p46 (in Chinese) [姜万录, 张淑清, 王益群 2005 基于混沌和小波的故障信息诊断 (北京: 国防工业出版社) 第 46 页]
- [2] Fraser A M, Swinney H L 1986 *Phys. Rev. A* **33** 1134
- [3] Zhang S Q, Zhao Y C, Jia J, Zhang L G, Shanguan H L 2010 *Chin. Phys. B* **19** 060514
- [4] Tian Y C 1997 *Systems Engineering Theory and Practice* **05** 48 (in Chinese) [田玉楚 1997 系统工程理论与实践 **05** 48]
- [5] Lu X Q, Cao B, Zeng M 2006 *Chinese J. Comput Phys.* **23** 184
- [6] Zhang J, Fan Y Y, Li H M, Sun H Y, Jia M 2011 *Chinese J. Comput Phys.* **28** 469 (in Chinese) [张菁, 樊养余, 李慧敏, 孙恒义, 贾蒙 2011 计算物理 **28** 469]
- [7] Xiao F H, Yan G R, Han Y H 2005 *Acta Phys. Sin.* **54** 0550 (in Chinese) [肖方红, 阎桂荣, 韩宇航 2005 物理学报 **54** 0550]
- [8] Zhang S Q, Jia J, Gao M, Han X 2010 *Acta Phys. Sin.* **59** 1576 (in Chinese) [张淑清, 贾健, 高敏, 韩叙 2010 物理学报 **59** 1576]
- [9] Zhu X K, Jia Y H 2005 *Journal of Geomatics* **05** 3 (in Chinese) [祝晓坤, 贾永红 2005 测绘信息与工程 **05** 3]
- [10] Xiao F H, Yan G R, Han Y H 2004 *Acta Phys. Sin.* **53** 2877 (in Chinese) [肖方红, 阎桂荣, 韩宇航 2004 物理学报 **53** 2877]
- [11] Lehrman M, Rechester A B 1997 *Phys. Rev. Lett.* **78** 54
- [12] Rechester A B and White R B 1991 *Phys. Lett. A* **158** 51
- [13] Zhang Y 2004 *J. Xiangtan Min. Inst.* **19** 75 (in Chinese) [张雨 2004 湘潭矿业学院学报 **19** 75]
- [14] Azad R K, Rao J S, Ramaswamy R 2002 *Chao Solitons and Fractals* **14** 633
- [15] Xiang K, Jiang J P 2007 *PR&AI* **20** 154 (in Chinese) [向旭, 蒋静坪 2007 模式识别与人工智能 **20** 154]
- [16] Shi Y S, Jiang Y, Song Y X 2011 *Journal of Aerospace Power* **26** 670 (in Chinese) [史永胜, 姜颖, 宋云雪 2011 航空动力学报 **26** 670]
- [17] Yu W B 2008 *Calculate Experiment and Analysis on Chaos* (Beijing: Science Press) p33 (in Chinese) [于万波 2008 混沌的计算实验与分析 (北京: 科学出版社) 第 33 页]
- [18] Liu Z H 2006 *Fundamentals and Applications of Chaotic Dynamics* (Beijing: Higher Education Press) p71 (in Chinese) [刘宗华 2006 混沌动力学基础及其应用 (北京: 高等教育出版社) 第 71 页]

Delay time obtaining method using the maximum joint entropy on the basis of symbolic analysis*

Zhang Shu-Qing[†] Li Xin-Xin Zhang Li-Guo Hu Yong-Tao Li Liang

(Institute of Electronic engineering of Yanshan University, Qinhuangdao 066004, China)

(Received 30 November 2012; revised manuscript received 4 February 2013)

Abstract

In this paper, the local maximum of joint entropy was computed using the symbolic analysis method so as to determine the appropriate delay time of the phase space reconstruction. The numerical experiments for three typical chaotic systems show that the present method could reduce computation, increase the efficiency, and also could obtain the optimum delay time accurately and rapidly. And it could reconstruct the original phase space from the time series effectively. Thus it provides a fast and effective way to identify the chaotic signal.

Keywords: phase space reconstruction, delay time, joint entropy, symbolic analysis

PACS: 05.45.Rq, 05.40.-a, 02.50.Fz

DOI: 10.7498/aps.62.110506

* Project supported by the National Natural Science Foundation of China (Grant Nos. 61077071, 51075349), and the Natural Science Foundation of Hebei Province, China (Grant No. F2011203207).

[†] Corresponding author. E-mail: zhshq-yd@163.com