

文章编号: 1000-5471(2008)06-0019-05

# 脑电时间序列的非线性重构<sup>①</sup>

朱家富<sup>1,2</sup>, 何 为<sup>1</sup>

1. 重庆大学 电气工程学院, 重庆 400030; 2. 重庆文理学院 电子电气工程学院, 重庆 永川 402160

**摘要:** 脑电信号的非线性特征量的提取以其非线性重构为基础, 对脑电信号的进行非线性重构的关键是正确选择重构参数时间延迟和嵌入维数. 确定重构参数的方法很多, 并各有其优缺点. C-C 方法是一种可以同时确定最佳时间延迟和最佳嵌入维数的新方法, 以 Lorenz 模型数据和脑电时间序列为计算对象, 对这种方法进行了验证与比较.

**关键词:** 相空间重构; 时间延迟; 嵌入维数; 脑电信号; Lorenz 模型

**中图分类号:** TN911.7

**文献标识码:** A

脑电信号的分析与处理是近几年发展起来的热门研究课题. 自从 1924 年 H. Berger 首次完整地从小脑皮层上记录脑电信号以来, 人们尝试用各种各样的方法来分析脑电信号<sup>[1]</sup>, 并在许多领域取得了令人满意的成果. 由于脑电的复杂性与混沌特性, 脑电信号的非线性特性已经得到了证实, 学术界也尝试用非线性动力学方法来分析脑电信号, 并利用诸如分数维值、Lyapunov 指数、哥氏熵、复杂度等非线性动力学参数来刻画脑电信号<sup>[2]</sup>. 计算这些动力学参数的前提是对其进行非线性重构, 而如何选择正确的重构参数, 最大限度地防止在对脑电信号重构的过程中信息量的丢失, 这有是对脑电信号进行有效的非线性重构的关键.

## 1 非线性相空间重构理论

对单非线性序列进行非线性重构的目的是希望隐含在序列中的各种信息, 能够通过其高维重构显露出来, 并保持原有信息的连续性. 非线性相空间重构的理论基础是由 Packard 和 Takens 建立的嵌入理论<sup>[3]</sup>. 假设  $\{x_k : k = 1, \dots, N\}$  是观测得到的某一时间序列, 将其嵌入到  $m$  维欧氏空间  $R_m$  中, 得到一个点(或向量)集  $J(m)$ , 其元素记作

$$X_n(m, t, J) = (x_{1+(n-1)J}, x_{1+(n-1)J+t}, \dots, x_{1+(n-1)J+(m-1)t}) \quad n = 1, \dots, N_m \quad (1)$$

式中:  $m$  为嵌入维数,  $t$  是时间延迟,  $J$  是采样间隔数,  $N_m$  是重构点(或重构相量)的个数, 它们之间的关系满足

$$N_m = [((N-1) - (m-1)t)/J + 1] \quad (2)$$

若信号的采样频率为  $F_s$ , 则延迟时间  $\tau_d$  (delay time)(单位: s)与时间延迟  $t$  之间的关系为

$$\tau_d = t \times T_s = t/F_s \quad (3)$$

采样间隔数  $J$  的大小与信号的重采样有关, 重采样频率  $f_s$ 、原始信号采样频率  $F_s$  以及采样间隔数  $J$  之间的关系满足  $f_s = F_s/J$ , 一般情况下取  $J = 1$ , 此时  $f_s = F_s$ , 表明没对原始信号没有实现重采样.

① 收稿日期: 2008-04-18

基金项目: 重庆市教委科技资助项目(KJ081209); 重庆文理学院研究资助项目(Y2007WX44).

作者简介: 朱家富(1969-), 男, 重庆万州人, 副教授, 主要从事生物医学信号处理的理论与方法的研究.

## 2 相空间重构参数的确定

在对原始数据进行非线性重构的时候,最影响重构效果的两个参数是嵌入维数  $m$  和时间延迟  $t$ ,同时,实践证明这两个参数的选取也是非常困难的.相空间重构时的嵌入维数和时间延迟选取的原则是保证重构时间序列能最大限度地重现原时间序列的非线性特征,保持原动力学系统的信息不会丢失,而这时的嵌入维数和时间延迟称为最佳嵌入维数和最佳时间延迟.关于最佳嵌入维数和最佳时间延迟的确定,现在有两种相反的观点:一种观点认为时间延迟和嵌入维数是两个互不相干的量,可以独立地确定最佳时间延迟和最佳嵌入维数,这种观点来源于 Taken 的研究,因为根据他的研究认为,时间延迟和嵌入维数在理论上是互相独立的;另一种观点认为时间延迟  $t$  和嵌入维数  $m$  是互相依赖的,最佳时间延迟和最佳嵌入维数的确定必须作为一个整体来考虑.

### 2.1 最佳时间延迟的确定

时间延迟是一个正整数,选择最佳的时间延迟是相空间重构的一个重要步骤之一,其值既不能太小也不能太大.由(1)式可以看出,若时间延迟  $t$  的取值太小,任意两个相邻的延迟坐标  $X_i$  和  $X_{i+1}$  在数值上过于接近,从统计的角度来看,两点的彼此相关性过大,因此不能作为两个独立的坐标;相反,若时间延迟  $t$  的取值太大,任意两个相邻的延迟坐标  $X_i$  和  $X_{i+1}$  彼此相互独立,相关性太小,若要求能反映原动力学系统的整体信息,则需要更多的原始数据,这会在实际计算中增加更多的计算时间.综合而言,最佳时间延迟选择的原则是使得重构系统在统计意义下为独立时的尽可能小的值.

当把时间延迟作为一个独立的重构参数来考虑时,最佳时间延迟  $\tau$  的确定主要遵循以下两条原则:一是序列相关性原则.信号的重构将使得重构信号  $X_i$  的各元素之间的相关性较原始信号而言有所减小,同时要保证重构信号  $X_i$  中包含原动力学系统的所有信息,或者信息的丢失量很少.基于这种原则下的方法主要有自相关法、互信息量法、高阶相关法等<sup>[4]</sup>.二是相空间扩展原则.即重构相空间轨迹应从相空间的主对角线尽可能地扩展,但同时又要保证不出现折叠.基于这一原则的主要方法有填充因子法、摆动量法、平均位移法、SVF 法等.新的研究发现,可以在这两个原则之间找到另外一种兼顾方法,如象复自相关法和去偏复自相关法等<sup>[5]</sup>,它们具有很强的理论依据,其计算的复杂度不大,对数据长度的依赖性不强,具有很强的抗干扰能力,具有很强的实际操作性,在实际中应用较广.

### 2.2 最佳嵌入维数的确定

最佳嵌入维数的确定是相空间重构的又一个重要问题.嵌入维数过大,将会不必要的计算工作量,而且系统动力学行为由低维描述时,多余的相空间维数将有放大原时间序列噪声和舍入误差的效应;若嵌入维数过小,则重构吸引子不能完全拆开,重构的相空间不能反映原动力学系统的动力学特征.

根据 Takens 的理论,嵌入维数  $m$  和分数维值  $D$  应该满足关系  $m \geq 2D + 1$ ,但实际上这种关系是有缺陷的,因为关系式的得来是假设在无噪声时间序列的基础之上的,而实际中的时间序列都或多或少含有噪声,而且实验证明该方法估计出来的最佳嵌入维数要比需要的大.1985年, J P Eckman 证明对与任何的时间序列,相空间重构时的嵌入维数可以在  $D \leq m \leq 2D + 1$  间取值.但实际中这两种方法都是不可行的,因为在对相空间进行重构之前,吸引子的分数维值是并不知道的,其后效性使得这两个结果失去了应用的前提.

确定最佳嵌入维数  $m$  的方法目前有很多种,最常见的是伪邻点法,其基本思想是:在进行吸引子重构时,重构相空间实际上是原相空间的一个低维投影,若重构维数太低,则可能原来本身并不相交的相轨迹有可能相交,原来并不相邻的点在重构相空间中成为最近的邻点,即伪邻点.伪邻点法就是使重构维数从低到高变化,直到不出现伪邻点,此时的最小嵌入维数就是最佳嵌入维数.另外一个常见的方法是系统特征量饱和法,其基本思想来源于系统的特征量随着嵌入维数的增加而波动的特点,增大嵌入维数,计算每一个嵌入维数下的系统特征量,若特征量对嵌入维数出现饱和,即特征量不随嵌入维数的增加而变化,此时的最小嵌入维数即为最佳嵌入维数.这两种方法的共同点都是必须计算多个嵌入维数下的参考量来对比分析,从而找到最佳的嵌入维数,因此计算工作量都较大.

### 2.3 C-C 算法

C-C 算法是一种能够同时确定最佳时间延迟  $\tau$  和最佳嵌入维数  $m$  的改进型算法, 它是由 H S Kim, R Eykholt 和 J D Salas 于 1999 年提出的<sup>[6]</sup>, 该方法应用关联积分函数  $C$  的变化规律, 从统计的角度可以同时确定最佳时间延迟和最佳嵌入维数的算法. C-C 算法是在大量的统计规律中得到的结论, 其无雄厚的理论基础, 并没有经过严格的理论推导与证明, 但实践已经证明其可行性, 在实际的计算中可以发挥其独特的优点.

对于如(1)式的重构序列, 其关联积分定义为

$$C(m, N, r, t) = \frac{2}{N_m(N_m - 1)} \sum_{i=1}^{N_m} \sum_{j=i+1}^{N_m} \theta(r - \|X_i - X_j\|) \quad (4)$$

式中:  $m$  是嵌入维数,  $N$  原时间序列点数,  $r$  是参考半径,  $t$  是时间延迟,  $N_m$  是相空间重构点数(由(2)式确定),  $\theta$  是 Heaviside 函数( $x < 0$  时,  $\theta(x) = 0$ ;  $x \geq 0$  时,  $\theta(x) = 1$ ),  $\|\cdot\|$  表示时间向量的上确界范数.

将原时间序列  $\{x_k : k = 1 \cdots N\}$  分解成  $t$  ( $t$  为一般自然数) 个不相交的时间子序列:

$$\{x_i, x_{t+i}, x_{2t+i}, \cdots, x_{(l-1)t+i}\} \quad i = 1, 2, \cdots, t \quad (5)$$

式中:  $l$  为子序列的数据长度, 其值  $l = \lceil N/t \rceil$ , 对于较大的原始数据  $N$  而言, 定义子序列的  $S(m, r, t)$  和  $\Delta S(m, t)$  函数为

$$S(m, r, t) = \frac{1}{t} \sum_{s=1}^t [C_s(m, r, t) - C_s^m(1, r, t)] \quad (6)$$

$$\Delta S(m, t) = \max\{S(m, r_j, t)\} - \min\{S(m, r_j, t)\} \quad (7)$$

$S(m, r, t)$  和  $\Delta S(m, t)$  函数中隐含着原序列的相关性信息. Brock 根据统计得出  $2 \leq m \leq 5$ ,  $\sigma/2 \leq r \leq 2\sigma$  ( $\sigma$  是序列的标准差),  $N \geq 500$  时  $S(m, r, t)$  就能代表序列的相关性, 而对于研究序列的非线性重构问题时, 取  $N = 3\,000$  比较适宜. 根据 Brock 的统计结果, 取  $m = 2, 3, 4, 5$ ,  $r_j = j\sigma/2$ ,  $j = 1, 2, 3, 4$ , 计算

$$S(t) = \frac{1}{16} \sum_{m=2}^5 \sum_{j=1}^4 S(m, r_j, t) \quad (8)$$

$$\Delta S(t) = \frac{1}{4} \sum_{m=2}^5 \Delta S(m, t) \quad (9)$$

$$S_{\text{cor}}(t) = \Delta S(t) + |S(t)| \quad (10)$$

取不同的  $t$  进行计算,  $\Delta S(t)$  的第一个极小值对应着的时间延迟  $t$  就为最佳时间延迟  $\tau$ , 其延迟时间为  $\tau_d = \tau T_s$ , 而  $S_{\text{cor}}(t)$  的最小值就对应着时间窗口  $\tau_w = (m-1)\tau_d$ , 则最佳嵌入维数为  $m = (\tau_w/\tau_d) + 1$ . 这些计算参数的选择与结论的给出都是在统计的基础上得出的, 无可靠的理论依据与理论推导, 只有在实际应用中才能检验其正确性.

## 3 实验验证与对比结果

为了验证 C-C 方法的正确性, 分别以 Lorenz 模型产生的时间序列和实验脑电信号为例, 通过计算式(8)-(10)式来确定最佳嵌入维数  $m$  和最佳时间延迟  $\tau$ .

### 3.1 Lorenz 模型的实验仿真

一个很著名的非线性时间系统例子就是 Lorenz 模型, 其方程为<sup>[7]</sup>

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = rx - y - xz \\ \dot{z} = -bz + xy \end{cases} \quad (11)$$

取系数  $\sigma = 10$ ,  $r = 28$ ,  $b = \frac{8}{3}$ , 初始值取(1, 2, 3), 步长取 0.01, 利用四阶 Runge-Kutta 公式通过 5 000 次迭代后, 取计算所得的数值为实验用非线性时间序列.

以 Lorenz 模型的  $x$  轴坐标时间序列为计算对象, 数据长度取  $N = 3\,000$ , 时间延迟取  $t = 1 \sim 200 T_s$ ,

分别计算  $S(t)$ ,  $\Delta S(t)$  和  $S_{cor}(t)$  的值, 其随  $t$  的变化关系曲线如图 1 所示. 从图 1(b)可以看出,  $\Delta S(t)$  的第一个极小值出现在  $\tau_d = 15T_s = 0.15$  s 处, 因此为  $\tau = 15$ . 而从图 1(c)看出,  $S_{cor}(t)$  的最小值出现在  $\tau_w = 140T_s = 0.140$  s, 由此计算出最佳嵌入维数  $m = 10$ .

对于如式(11)所示的 Lorenz 模型, 取不同的系数  $\sigma, r$  和  $b$ , 初始值仍取  $(1, 2, 3)$ , 步长取 0.01, 生成其非线性时间序列. 利用本方法计算其序列的最佳嵌入维数  $m$  和最佳时间延迟  $\tau$ , 计算结果与文献[8]和文献[9]采用互信息法计算的结果相似.

### 3.2 脑电时间序列的重构参数

脑电信号的采集按照国际标准导联 10~20 系统在受试者头皮上安置电极, 采用的是十六导联连接方式: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6. 参考电极 A1 和 A2 分别置于左右耳垂. 利用脑电信号放大器同时采集 16 个通道的 EEG 波形, 采样频率是 1 000 Hz, 采样精度 12 位 A/D, 带宽 0.5~100 Hz.

与 Lorenz 模型计算方法相似, 在对原始脑电信号作包含去高频再噪声和选段等基本的预处理后, 数据长度取  $N = 3 000$ , 时间延迟取  $t = 1 \sim 200 T_s$ , 同样分别计算  $S(t)$ ,  $\Delta S(t)$  和  $S_{cor}(t)$  的值, 其随  $t$  的变化关系曲线见图 2. 得到最佳时间延迟为  $\tau = 25$ 、最佳嵌入维数  $m = 6$ .

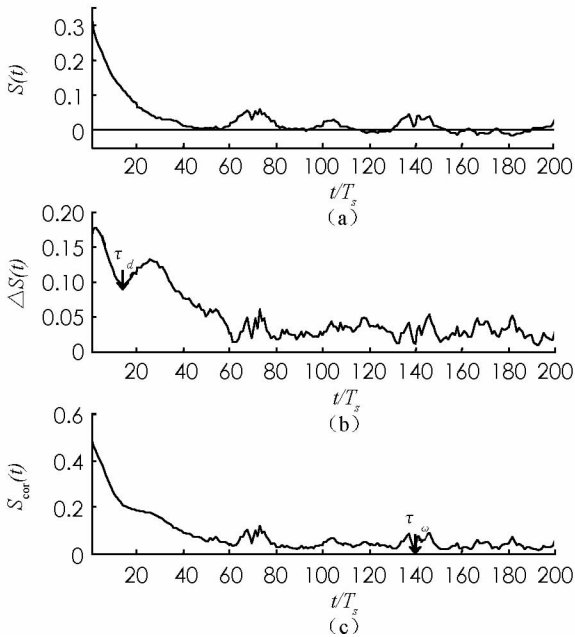


图 1 Lorenz 系统的  $S(t)$ 、 $\Delta S(t)$  和  $S_{cor}(t)$  随时间延迟  $t$  的变化关系曲线

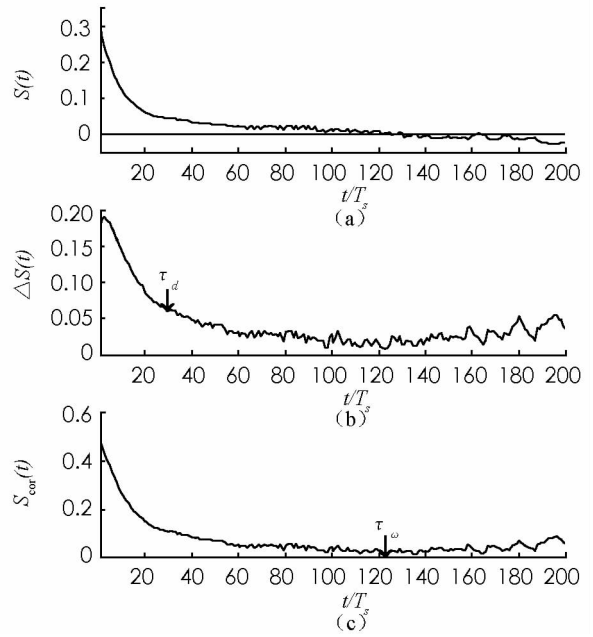


图 2 脑电时间序列的  $S(t)$ 、 $\Delta S(t)$  和  $S_{cor}(t)$  随时间延迟  $t$  的变化关系曲线

在参考文献[10]中, 针对大鼠癫痫实验数据, 采用伪邻点法和 Cao 法相结合的方法确定最佳嵌入维  $m = 7$  和最佳时间延迟为  $\tau = 25$ , 而在文献[11]中, 在计算海马脑电信号的分数维值, 讨论了嵌入维数  $m = 2 \sim 8$  的值. 尽管所采用的方法不同, 但最佳嵌入维和最佳时间延迟的确定值与本方法所确定的值是相符的.

## 4 结论与讨论

对诸如脑电时间序列等非线性时间序列描述的前提是重构相空间, 进行非线性重构的目的是希望隐含在序列中的各种信息, 能够通过其高维重构显露出来. 对非线性时间序列重构的原则是保持原有信息的连续性, 在重构过程中不能有太多信息的丢失. 重构相空间最重要的两个参数是延迟时间和嵌入维数, 确定这两个参数的方法很多, 而 C-C 算法的最大优点是计算量小、对小数据组可靠, 而且其效果和互信息方法是一致的.

C-C 算法通过计算不同  $t$  的  $\Delta S(t)$  来确定最佳时间延迟  $\tau$ , 而由  $S_{cor}(t)$  的来确定时间窗口  $\tau_w = (m - 1)\tau_d$ , 并由式  $m = (\tau_w/\tau_d) + 1$  计算最佳嵌入维数. 它能同时确定重构空间的最佳延迟时间和最佳嵌入维

数, 并能很好地保证它们的一致性.

通过以 Lorenz 模型所产生的时间序列和实验测得的脑电数据为对象, 利用 C-C 算法计算了它们的最佳延迟时间和最佳嵌入维数, 计算结果与其他方法计算结果保持一致. 在实际应用中, 由于非线性系统本身的特点, 最佳延迟时间和最佳嵌入维数的选择是有一定范围的, 在这个范围内选定的延迟时间和嵌入维数所得到结果均满足相空间重构要求.

#### 参考文献:

- [1] 陈 中, 赵联文. 信号奇异性检测中小波分析的应用 [J]. 重庆师范大学学报(自然科学版), 2004, 21(2): 15 - 18.
- [2] 吕金虎, 陆君安, 陈士华. 混沌时间序列分析及其应用 [M]. 武汉: 武汉大学出版社, 2002.
- [3] Holger Kantz, Thomas Schreiber. Nonlinear Time Series Analysis [M]. Beijing: Hsinghua University Press and Cambridge University Press, 2001.
- [4] Liebert W, Schuster H G. Proper choice of the time delay for analysis of chaotic time series [J]. Phys Lett A, 1988, 142: 107 - 111.
- [5] 王礼刚, 左源瑞, 李盛瑜. 一种基于改进型遗传算法的关联规则提取算法及其应用 [J]. 重庆师范大学学报(自然科学版), 2006, 23(2): 42 - 45.
- [6] Kim H S, Eykholt R, Salas J D. Nonlinear dynamics, delay time, and embedding windows [J]. Physica D, 1999, 127: 48 - 60.
- [7] Lorenz E N. Deterministic nonperiodic flow [J]. Atoms Sci, 1963, 20(1): 130 - 141.
- [8] Martinerie J M, Albano A M, Mees A I, et al. Mutual information, strange attractors, and the optimal estimation of dimension [J]. Physics Review A, 1992, 45: 7058 - 7064.
- [9] 肖方红, 阎桂荣, 韩宇航. 混沌时序相空间重构参数确定的信息论方法 [J]. 物理学报, 2005, 54(2): 550 - 556.
- [10] 谢 勇, 徐健学, 杨红军, 等. 皮层脑电时间序列的相空间重构及非线性特征量的提取 [J]. 物理学报, 2002, 51(2): 205 - 214.
- [11] Taikang Ning. Dimensional Complexity of the Hippocampal EEG [J]. IEEE Bioengineering Conference, 2000(8 - 9): 27 - 28.

## The Nonlinear Reconstruction of EEG Signal

ZHU Jia-fu<sup>1,2</sup>, HE Wei<sup>1</sup>

1. College of Electrical Engineering, Chongqing University, Chongqing 400030, China;

2. School of Electronic and Electrical Engineering, Chongqing University of Arts and Sciences, Yongchuan Chongqing, 402160

**Abstract:** For EEG time series, gaining the nonlinear characteristic parameters are based on the reconstruction of the time series, and it is the key to determine the best delay time and the best embedding dimension for the time series reconstruction. There are many restrictive methods to determine the reconstruction parameters. In this paper, the C-C algorithm is introduced which can determine the best delay time and the best embedding dimension synchronously, and the method is validated and compared according to calculation results of EEG signals and Lorenz time series.

**Key words:** phase-space reconstruction; time delay; embedding dimension; EEG signal; Lorenz model