

生物实验数据的某些非线性分析方法

练 军 陈丹梅 郭爱克 陈润生

(中国科学院生物物理研究所神经生物学研究室, 北京 100101)

摘要 简要介绍常用的非线性动力学参量, 结合生物学实验数据的特点, 给出几种最新的分析方法.

关键词 稳定性, 噪声, 维数, 混沌

非线性动力学作为物理学在本世纪的第三次革命, 它的影响与相对论和量子力学一样广泛, 如今它在化学、生物学、地质学, 社会科学等领域已作为一种新的分析工具而使用. 可以说非线性科学将是 21 世纪的研究热点, 它在学术上高度融合交叉的特点决定了它的发展必然是各个领域的科学家共同努力的结果.

非线性动力学在生物学中应用最成功的例子之一是对脑电的研究, 用关联维数刻划了思维状态的转化和健康状况的差异. 此后, 混沌这个新名词在研究生命现象的文献中频频出现, 心跳、脑电位, 生化反应中展现了混沌; 血管的枝形结构, 蛋白质的分子链和表面都具有分形特征. 然而, 由于生物学实验的特点和近几年来非线性分析方法的完善, 早先的一些工作不得不进行修正. 就目前的研究水平, 我们可以表明生物系统的某些特征, 但真正能建立确定性混沌的可能性是比较小的.

本文第一部分简要介绍常用的非线性动力学参量, 第二部分叙述生物学实验数据的特点, 第三部分给出几种最新的分析方法, 第四部分是本文的总结.

1 常用的非线性动力学参量

1.1 关联维数 (correlation dimension, D_2)

维数用来刻划系统的几何结构, 着眼于系统的静态特征, 维数的具体形式有很多种: 如容量维、豪斯道夫维数、信息维数, 关联维数. 关联维数易于从实验数据计算^[1], 是处理生物学实验时间序列最常用的维数.

计算关联维数的第一步是时间序列的重构^[2]. 这是一个具有开创性的思想, 乍一看来, 单变量的时间序列似乎只能提供十分有限的信息, 其实时间序列包含着远为丰富的信息, 它蕴藏了参与动态的全部其他变量的痕迹, 并使人们得以验证体系潜在的某些与任何模型化无关的重要特征. 设原始时间序列: $\{S_i\}$, $i = 1, \dots, N$, 时间延迟取为 τ , 嵌入维数取为 m 后, 重构向量 $X_i = (S_i, S_{i+\tau}, \dots, S_{i+(m-1)\tau})$, $i \leq N - (m-1)\tau$, 第二步是计算关联积分:

$$C_2(\epsilon) = \frac{2}{(N-m)(N-m-1)} \sum_{i=1}^{N-(m-1)\tau} \sum_{j=i+1}^{N-(m-1)\tau} H(\epsilon - |X_i - X_j|) \quad (1)$$

H 是亥维赛 (Heaviside) 阶梯函数. 关联维数 D_2 定义为:

$$D_2 = \lim_{N \rightarrow \infty, \epsilon \rightarrow 0} \frac{\partial \lg C_2(\epsilon)}{\partial \lg \epsilon} \quad (2)$$

图 1 是从癫痫病人的脑电图计算出的关联积分和关联维数^[3], 计算 D_2 的时间消耗正比于 N^2 , 近年来很多工作都围绕着如何提高运算效率展开, 文献 [4] 有这方面的介绍.

1.2 李雅普诺夫指数与科尔摩诺夫熵

这两个参量是用来描述系统演化过程的动态特征. 李雅普诺夫指数 (Lyapunov exponents) 反映的是系统对初始条件的敏感性, 它的定义非常简单, 具体应用时却有很多困

难, Wolf^[5]提出了实用的算法, 以后的各种方法多是基于这种方法的改进. 图 2 是从癫痫病人的脑电图计算出的最大李雅普诺夫指数^[3].

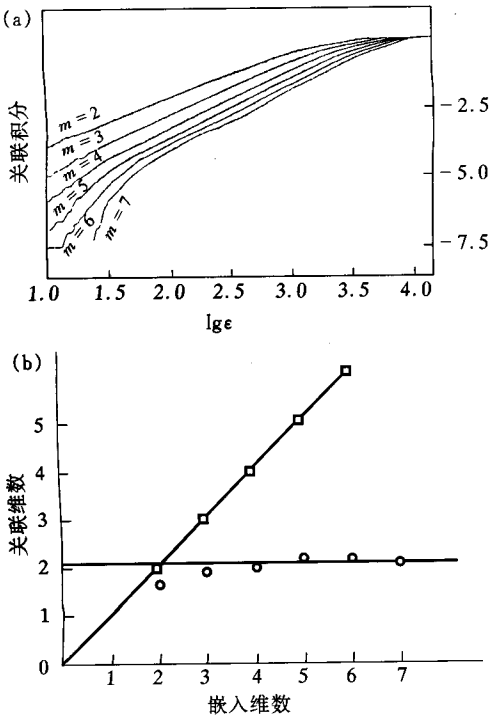


图 1 从癫痫病人的脑电图计算出的关联积分和关联维数

(a) 嵌入维数 m 取 2 至 7 维的关联积分; (b) 癫痫病人的脑电图关联维数(○)与白噪声关联维数(□)的对比.

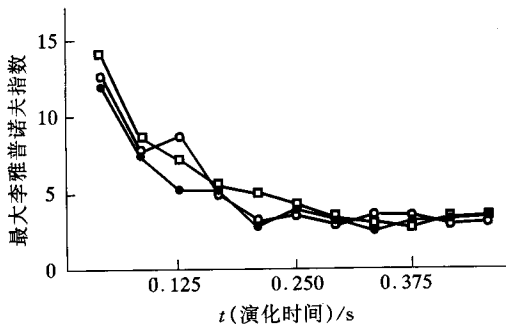


图 2 从癫痫病人的脑电图计算出的最大李雅普诺夫指数

吸引子的最大李雅普诺夫指数. □—□、○—○、●—●: 分别表示不同的空间展开度.

entropy, K 熵) 是用于标志在相空间里混沌运动的最重要的度量, 它正比于时间进程中动力系统状态信息丧失的平均速率.

1.3 混沌程度的刻划

功率谱、自相关函数、彭加勒映象可以定性判断混沌的有无, 定量刻划的主要工具就是这部分介绍的三个参量. 混沌出现的标志是: 关联维数大于 2 且是分数; 李雅普诺夫指数至少有一个为正; K 熵介于零和无穷大之间, K 越大混沌程度越严重.

2 从实验数据重构动力学的困难

当物理学家已经能在微分方程的数值实验中控制混沌时, 现有的方法在处理实验数据时却显得苍白无力, 生物学的实验数据尤其复杂, 这表现在如下三个方面.

2.1 稳定性

时间序列的稳定性定义为均值、方差及自相关函数在整个采样过程中保持不变. 稳定性是大多数分析方法能够应用的首要条件. 生物学实验由于外界环境的扰动及长时程过渡效应, 稳定性条件很容易受到破坏. 解决的方法之一是尽量选取较短的时间序列, 这时稳定性条件近似成立, 当然数据量的减少也会限制维数的计算, 维数与数据量的关系可见下文. 另一种方法是用去除趋势或对序列作一阶微分来纠正数据的整体漂移.

2.2 噪声

噪声信号的抑制已经成为用非线性参量描述实验数据的一个主要困难^[7]. 滤波是一个非常精细的过程, 原始信号经滤波变换后有可能变得更为复杂, 而某些噪声经滤波后至少有一个为正的李雅普诺夫指数, 出现了与低维确定性混沌类似的性质, 这导致了以前的一部分文献把噪声当成了混沌报道.

2.3 有限的的数据量及复杂系统的高维

从一个长度为 N 的时间序列中所能估计出的最大维数^[8]为 $D_{\max} \approx 2\lg N - 2$, 而估计李雅普诺夫指数所需的数据量是估计维数所需的平方. 时间序列的延伸对于生物学数据将是

不适宜的, 因为此时非稳定性因素将使得出的非线性参量失去意义, 即使能够增加数据量 (如 10 倍), 对于高维吸引子的维数估计也无显著帮助, 因为点间距的减少使判断维数大小的标度区几乎不变。

3 非线性分析方法的新进展

第二部分介绍的几种方法在分析微分方程及简单的易于精确控制外部条件的实验中是比较有效的. 对于高维数的, 稳定性差的复杂系统, 如一个生物系统, 应用上述方法就会遇到很多困难. 如果不能获得足够长的精确的实验数据就无法计算维数及李氏指数, 保证实验数据的稳定性和高信噪比是实验和处理数据过程中的难点. 下面介绍的几种方法有助于克服第三部分所述的几种困难。

3.1 点相关维数

点相关维数 (point correlation dimension, PD_2) 是由 Skinner^[9] 近年来发展的. 与常用的关联维数 D_2 不同点在于公式 (1) 中不对 i 求和而简单地取当前的 i 矢量为参考矢量. 在计算距离 $|x_i - x_j|$ 时, 摒弃对于维数的线性标度没有贡献的项. 点维数只搜寻处于稳定亚周期中的点, 对于不稳定性因素较不敏感, PD_2 可以探知和抑制脉冲型噪声, 对于通过相同数据获取系统的白噪声也有分辨能力。

3.2 替代数据

随机系统也可能产生和混沌类似的维数和李雅普诺夫指数. 单独地给出维数和李雅普诺夫指数已经不能足以表明混沌的出现. 替代数据 (surrogate data) 方法^[10]是把原始数据加以适当变化, 然后用相同的算法计算维数等参量, 和从原始数据计算出的结果作对比, 利用假设检验来判断系统展现的混沌是不是由噪声引发的. 替代数据的方法主要有三种: 随机位相 (phase randomized), 高斯标度 (Gauss scaled) 和傅立叶混和法 (Fouriur shuffled). 它们的原理是类似的, 以随机位相为例: 将原始数据作傅氏变换, 位相作随机化处理, 再作反傅氏变化, 这时得到的数据就是替代数据,

它们与原始数据有相同的功率谱、均值、方差和自相关函数. 图 3 是一个掺杂噪声的吸引子的关联维数和其替代数据的关联维数, 它们之间的显著差异表明了确定性因素的存在。

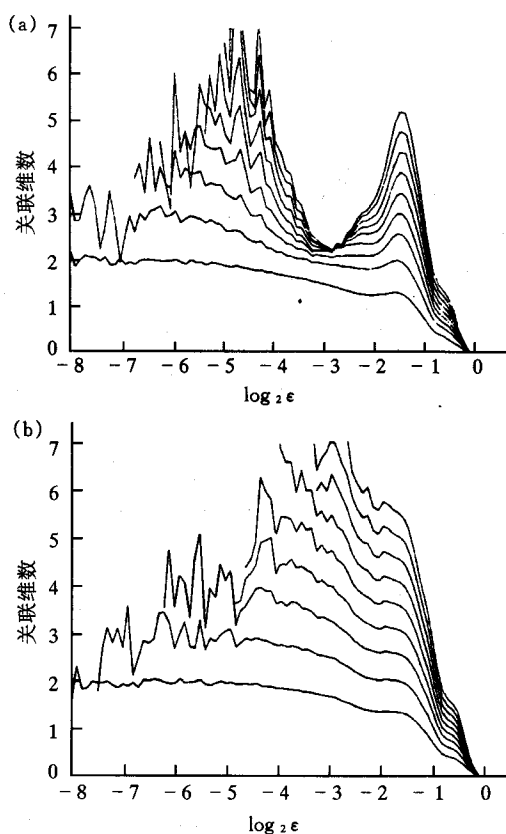


图 3 一个掺杂噪声的吸引子的关联维数 (a) 和其替代数据的关联维数 (b)

3.3 可预测性

检验可预测性目的是区分确定性混沌和随机系统^[11]. 在非线形领域目前常用的方法有四种: 例外事件法 (exceptional events), 局域流 (local flow), 局域耗散 (local dispersion) 和非线性预测 (nonlinear prediction) 方法. 第一种方法特别适用于高维系统, 因为在高维系统中偶然事件可能是确定性的唯一指示, 后三种的特点是从实验数据的分布中提取中心趋势. 非线性预测用在高维和低信噪比的生物学实验数据更为有效, 其原理是对于嵌入空间中的一点 X_0 , 选取水平标度 H , 取 $k = 0.02N$

个最近邻点, 计算 $\langle v \rangle = \frac{1}{k} \sum_{j=1}^k X_{j+H}$, 以 X_{0+H} 为例, 定义预测误差: $\epsilon_{\text{trans}} = |X_{0+H} - \langle v \rangle|$, 时间序列的平均值的预测误差为 $\epsilon_{m\text{trans}} = |X_{0+H} - \text{mean}(x)|$, 归一化后的 NPE 作为这种方法的判断参量, RMS 表示均方根.

$$NPE = \frac{RMS(\epsilon_{\text{trans}})}{RMS(\epsilon_{m\text{trans}})} \quad (3)$$

4 混沌生物系统给我们的启示

生物系统作为异常复杂的系统长期以来只被当作随机系统来研究, 而当用一些简单的非线性参量就能描述系统的特征时, 无论是生物学家还是物理学家对此都表示了极大兴趣. 当我们能够证明一个系统是确定性混沌系统, 我们能够知道^[12]: a. 系统是非线性确定性系统; b. 系统能够被有限的几个自由度刻画; c. 从原理上, 系统能够被简单的模型所描述; d. 长期行为的不可预测性; e. 利用混沌初值敏感性的特征可控制系统状态.

然而, 从生物学实验数据中真正提取出确定性混沌特征的例子并不多. 一方面是生物学实验中噪声比较大; 另一方面生物系统往往是高维的, 难以用现有只对低维问题有效的非线性方法来分析. 当我们不能在所研究的系统中发现混沌时, 目前较好的方法是用替代数据或预测性的方法来说明系统和一般随机系统的差别, 这多少有些令人沮丧, 但就象当初只把生物系统当作随机系统一样, 先入为主地去从实验数据中提取混沌特征也是不可取的. 反之, 当我们发现了混沌倒是应该更仔细地去验证, 以防被处理过程中的失误所迷惑.

参 考 文 献

- 1 Grassberger P. Generalize dimensions of strange attractors. *Phys Lett A*, 1985, **107**: 101~105
- 2 Takens F. Detecting strange attractors in turbulence. In:

Lecture notes in Mathematics, **898**. Berlin: New York Springer, 1981. 366~370

- 3 Babloyantz A. Low-dimensional chaos in an instance of epilepsy. *Proc Natl Acad Sci USA*, 1986, **83**: 3513~3517
- 4 Schreiber T. Efficient neighbor searching in nonlinear time series analysis. *Int J Bifurcation and Chaos*, 1995, **5** (2): 349~358
- 5 Wolf A, Swift J B, Swinney H L *et al.* Determining Lyapunov exponents from a time series. *Physica D*, 1985, **16** (2): 285~317
- 6 Glass L, Kaplan D. Time series analysis of complex dynamics in physiology and medicine. *Medical Progress through Technology*, 1993, **19**: 115~128
- 7 Kostelich E J. Noise reduction in chaotic time-series data: a survey of common methods. *Phys Rev E*, 1993, **48** (3): 1752~1763
- 8 Eckmann J P, Ruelle D. Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. *Physica D*, 1992, **56** (1): 185~187
- 9 Skinner J E. Reduction in the correlation dimension of heart-beat intervals precedes imminent ventricular fibrillation in human subjects. *Am Heart J*, 1993, **125**: 731~737
- 10 Theiler J, Eubank S, Longtin A. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 1992, **58** (1): 77~94
- 11 Chang T, Sauer T, Schiff S J. Test for nonlinearity in short stationary time series. *Chaos*, 1995, **5** (1): 118~126
- 12 Skinner J E. Low-dimensional chaos in biological systems. *Biotechnology*, 1993, **12**: 596~600

Some Nonlinear Methods in the Biological Experimental Data Analysis. LIAN Jun, CHEN Danmei, GUO Aike, CHEN Runsheng (Department of Neurobiology, Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China).

Abstract Some common nonlinear dynamic parameters are introduced, then on the basis of the characteristics of the biological experimental data, the newly developed analysis methods are illustrated.

Key words stability, noise, dimension, chaos