

# 非线性时间序列复杂性测度的稳定性<sup>①</sup>

罗爱萍

西昌学院 经济管理系, 西昌 四川 615000

**摘要:** 复杂性测度是刻画非线性时间序列最重要的动力学指标, 复杂度的计算涉及时间序列长度、复杂度阶次等参数的选择, 合适地选择这些计算参数是保证计算结果真实与稳定的前提. 以 Lorenz 方程所产生的时间序列和临床脑电时间序列为例, 在数值计算的基础上, 讨论了参数选择对计算结果的影响, 并给出了参数选择的参考.

**关键词:** 非线性时间序列; 复杂性测度; 稳定性; Lorenz 模型; 脑电信号

**中图分类号:** TN911.7

**文献标识码:** A

近年来, 随着非线性科学的发展, 对非线性时间序列的分析得到了广泛的应用, 特别是在对各类非线性动力学系统进行描述的时候, 利用根据系统得到的非线性时间序列的分析来表征动力学系统. 常见的参数有分维数、Lyapunov 指数、近似熵、复杂度等<sup>[1]</sup>.

非线性时间序列的复杂度是一种定量评价序列复杂性的方法, 其算法简单, 具有较强的抗干扰能力, 计算结果可以直接从时间序列的波形得到解释, 与非线性动力学的其它参数相比, 计算时只需要较少的数据, 对随机性的和确定性的信号都适用, 特别适用于对生理信号的分析<sup>[2]</sup>. 然而, 在计算时间序列的复杂度时, 涉及到对时间序列多个计算参数的选择, 合理的选择计算参数是正确利用复杂度来描述非线性动力学系统的前提.

## 1 复杂性测度

### 1.1 复杂性测度定义

复杂度的最初定义是由 Kolmogorov 于 1965 年提出的<sup>[3]</sup>, 表征为能够产生某一(0,1)序列所需的最短程序的比特数, 后来由 Lempel 和 Ziv 等人给出了实现这种定义复杂度的具体算法<sup>[4-5]</sup>, 并称为 Kolmogorov 复杂度或算法复杂度, 本文用  $C_k$  来表示, 其大小体现了时间序列的信息含量<sup>[6]</sup>.

Kolmogorov 复杂度的基本算法描述为: 以时间序列的均值为参考, 大于均值取“1”, 小于均值取“0”, 重构出一新的相同数据长度的(0,1)序列; 计算(0,1)序列中禁止字或禁止字符串的个数, 被称为绝对复杂度  $C(n)$ ; 由于任何序列的绝对复杂度  $C(n)$  都趋近于一个定值

$$b(n) = \frac{n}{\log_2 n} \quad (1)$$

而相对复杂度, 也称归一化复杂度  $C_k$ , 其定义为

$$C_k = \frac{c(n)}{b(n)} = c(n) \frac{\log_2 n}{n} \quad (2)$$

$C_k$  是一个与序列长度  $N$  有关的物理量, 可以看出完全随机序列的  $C_k$  趋近于 1, 而规律的或周期的序列的  $C_k$  趋近于 0. 我国著名学者徐京华等人于 1994 年提出了新的复杂度  $C_1, C_2$  定义<sup>[7]</sup>, 它是描述非线性

① 收稿日期: 2007-10-29

作者简介: 罗爱萍(1968-), 四川西昌人, 讲师, 主要从事计算机应用技术领域的科研与教学工作.

时间序列的又一个复杂性测度. 本研究中以 Kolmogorov 复杂度  $C_k$  为复杂性测度研究指标.

## 1.2 高阶复杂性测度

在计算时间序列的复杂度的时候, 首先是将原序列以序列的均值为参考字符化, 得到一个与原序列相同长度的  $(0,1)$  字符序列, 这一过程的实质是将原序列颗粒化, 而这一过度粗略化行为丢失了原序列中所包含的幅度信息, 使得在新序列中以高幅波起主要作用, 而淹没了原时间序列中的低幅波, 忽略了其中的细微变化信息. 为弥补这一缺陷, 引入高阶复杂度的概念, 以减小粗略化对原序列信息的丢失<sup>[8]</sup>. 其方法是通过将原序列与序列均值的差值的绝对值来重构新的序列, 如原序列为  $x_i (i=1, 2, \dots, N, N$  为数据长度), 则重构的新时间序列为

$$y_i = \left| x_i - \frac{1}{N} \sum_{j=1}^N x_j \right| \quad i = 1, 2, \dots, N \quad (3)$$

然后将新序列  $y_i$  字符化, 计算所得的复杂度称为原序列  $x_i$  的二阶复杂度. 多次重复序列的重构过程, 可以计算得到原序列的任意高阶复杂度.

## 1.3 复杂性测度的计算

$C_k$  复杂度定义为  $(0, 1)$  序列中禁止字或禁止字符串的个数, 其算法描述为: 对于一给定的字符串序列  $S = \{s_1, s_2, s_3, \dots, s_n\}$  而言, 另外两个字符串  $P = \{s_1, s_2, \dots, s_m\}$  和  $Q = \{s_{m+1}, \dots, s_r\}$  均为序列  $S$  的子字符串, 而字符串  $PQ = \{s_1, s_2, \dots, s_m, s_{m+1}, \dots, s_r\}$  为字符串  $P$  和  $Q$  的串联字符串,  $PQ_v$  为字符串  $PQ$  中去掉最后一个字符后形成的新的字符串, 判断  $Q$  是否属于  $PQ_v$  中已有的字句. 如果  $Q$  是  $PQ_v$  的子字符串, 就把下一个字符串  $s_{m+1}$  添加到  $Q$  的末尾, 并重复上述步骤, 这一过程被称为“复制”; 若  $Q$  不是  $PQ_v$  的子字符串, 则把  $Q$  添加到  $P$  的末尾, 并用分隔号“ $\cdot$ ”隔开, 把下一个字符  $s_{m+1}$  重新赋给  $Q$ , 再重复上述步骤, 这一过程被称为“插入”. 重复整个步骤使  $m$  的取值从 1 直至  $n$ , 这样  $S$  就被分隔符“ $\cdot$ ”分离成了许多子字符串段, 而子字符串段的数目就为绝对复杂度  $c(n)$  的值.

为能更清楚地描述复杂度的计算过程, 以序列  $S = \{0, 1, 1, 0, 0, 1, 1, 0, 1\}$  为例子来演示复杂度的计算步骤.

(1)  $s_1 = 0$ , 第一个字符总是插入  $\rightarrow 0 \cdot$

(2)  $s_2 = 1$ , 此时  $P = s_1 = 0, Q = s_2 = 1, PQ = 01, PQ_v = 0, Q$  不是  $PQ_v$  的子字符串,  $s_2$  再次被插入  $\rightarrow 0 \cdot 1 \cdot$

(3)  $s_3 = 1$ , 此时,  $P = s_1 s_2 = 01, Q = s_3 = 1, PQ = 011, PQ_v = 01, Q$  是  $PQ_v$  的子字符串,  $s_3$  被复制  $\rightarrow 0 \cdot 1 \cdot 1$

(4)  $s_4 = 0, P = s_1 s_2 = 01, Q = s_3 s_4 = 10, PQ = 0110, PQ_v = 011, Q$  不是  $PQ_v$  的子字符串,  $s_4$  被插入  $\rightarrow 0 \cdot 1 \cdot 10 \cdot$

(5)  $s_5 = 0, P = s_1 s_2 s_3 s_4 = 0110, Q = s_5 = 0, PQ = 01100, PQ_v = 0110, Q$  是  $PQ_v$  的子字符串,  $s_5$  被复制  $\rightarrow 0 \cdot 1 \cdot 10 \cdot 0$

(6)  $s_6 = 1, P = s_1 s_2 s_3 s_4 = 0110, Q = s_5 s_6 = 01, PQ = 011001, PQ_v = 01100, Q$  是  $PQ_v$  的子字符串,  $s_6$  再次被复制  $\rightarrow 0 \cdot 1 \cdot 10 \cdot 01$

(7) 重复上述过程, 直到序列  $S$  的最后一个字符  $s_{10} = 1$ , 最终  $S$  被分成

$S = 0 \cdot 1 \cdot 10 \cdot 011101 \cdot$

由于序列最后被分成了 4 段, 所以原序列  $S$  的绝对复杂度  $c(n) = 4$ , 根据式(2) 计算可得序列  $S$  的相对复杂度  $C_k$  为 0.98.

## 2 复杂性测度计算参数的选择

对于任何非线性时间序列, 在计算其复杂度的过程中, 主要涉及 2 个参数的选择: 一个是序列的长度, 另一个是复杂度的阶次, 这 2 个参数都将影响复杂度的计算结果, 影响了对非线性时间序列复杂度分析的稳定性.

## 2.1 实验数据的采集

本研究的实验数据从两个方面采集,一是利用典型的非线性系统——Lorenz 模型数值生成,二是临床采集得到的脑电信号,它们都是典型的非线性时间序列.

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = rx - y - xz \\ \dot{z} = -bz + xy \end{cases} \quad (4)$$

Lorenz 模型如式(4) 方程描述所示, 实验中取系数  $\sigma = 10$ ,  $r = 28$ ,  $b = 8/3$ , 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  $x, y, z$  的值为实验用非线性时间序列. 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, 本研究采用的是十六导联连接方式, 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. 图 1 分别为 EEG 时间序列和 Lorenz 模型时间序列的一个样本举例波形.

## 2.2 复杂度随数据长度的变化

根据(2)式的定义可以看出,  $C_k$  是一个与序列长度  $N$  有关的物理量, 选择合适的样本数据长度  $N$  对复杂度的计算是至关重要的. 如图 2 为不同样本复杂度随数据长度的变化关系曲线, 其中图(a) 为 Lorenz 模型  $x, y, z$  3 个坐标时间序列的复杂度随数据长度的变化关系, (b) 图为临床 5 个样本的脑电时间序列的复杂度随数据长度的变化关系. 可以看出, 对任何一类的任意一个样本而言, 当数据长度小于 1 000 点时, 复杂度的计算结果不稳定, 只要数据点在 1 000 点以上, 复杂度的计算就相对稳定的. 根据 Kolmogorov 复杂度定义可知, 序列数据长度越长计算结果越真实, 但要受计算量的影响, 实际计算复杂度的时候, 数据长度在 2 000 ~ 3 000 点为宜.

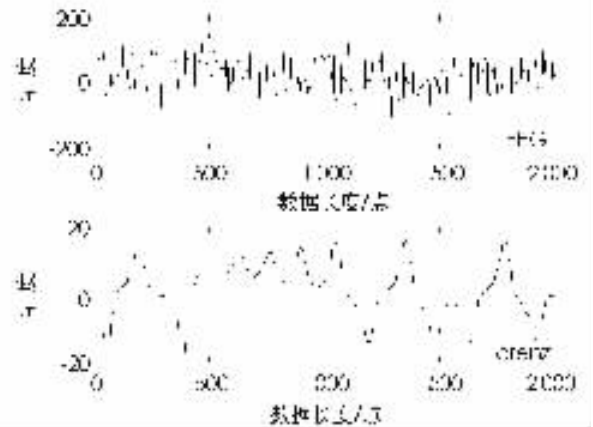


图 1 实验样本数据波形

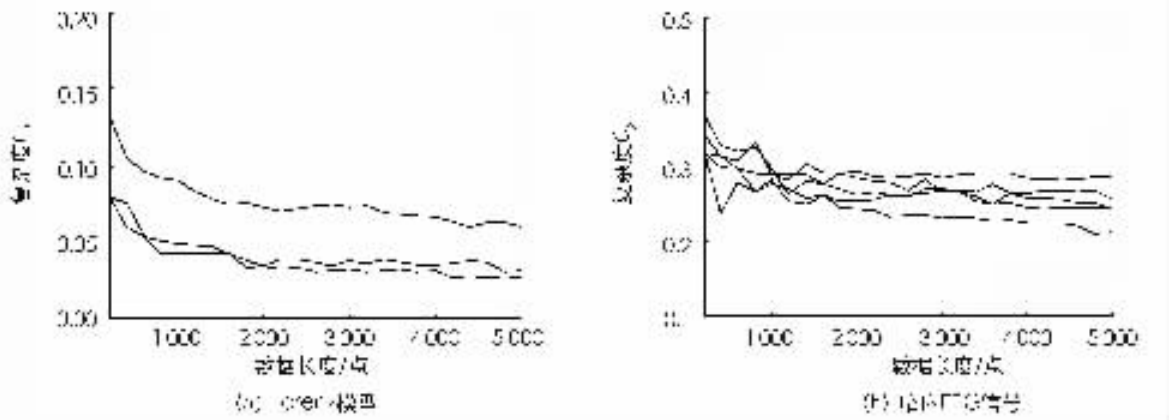


图 2 不同样本复杂度随数据长度的变化关系

## 2.3 复杂度随阶次的变化

复杂度计算的一个关键环节就是将原时间序列字符化, 这一方法将会丢失序列中低幅波所包含的信息, 高阶复杂度的引入可以很大程度地避免这一问题, 关于高阶复杂度的定义如前所述. 但在实际的计算复杂度时, 选择合适的阶次又成为了一个新的问题, 最佳的阶次可以最大程度地反映原时间序列的复杂性测度. 图 3 为不同样本复杂度随阶次的变化规律曲线, 其中(a) 图为 Lorenz 模型  $x, y, z$  3 个坐标时间序列的复杂度随阶次的变化关系, (b) 图为临床 5 个样本的脑电时间序列的复杂度随阶次的变化关系. 从图 3 可以看出, 过高的阶次会造成计算结果的振荡, 而在 1 ~ 6 阶的情况下, 复杂度的稳定性较好, 具体的阶次的确定必须在实验数据的支持下反复验证来确定.

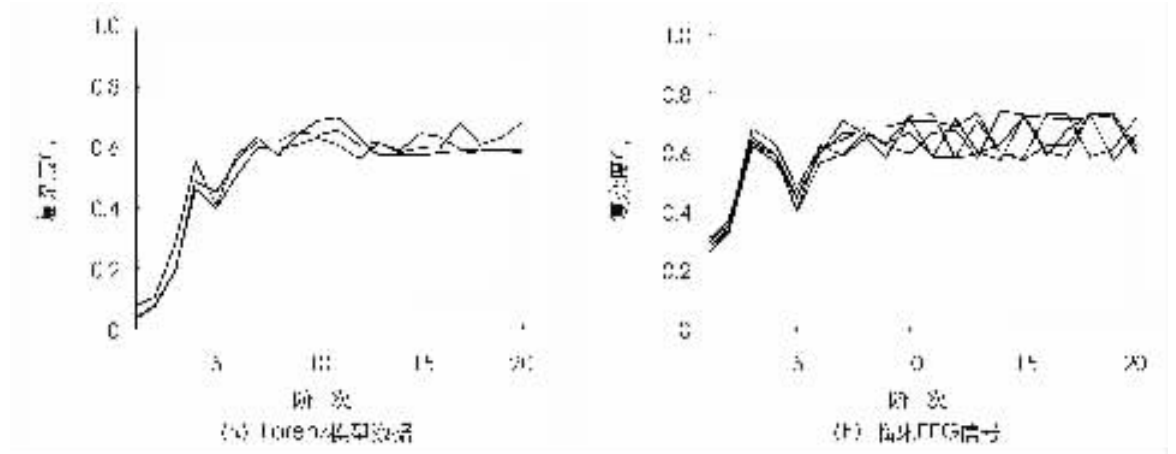


图 3 不同样本复杂度随阶次的变化关系

### 3 结 论

研究与实验计算表明,时间序列的复杂度随序列的长度  $N$  的增大而相对减小、而随着阶次的增大而增大.从理论上讲,时间序列数据长度越长计算结果越真实,但计算复杂度的工作量会大幅度提高,实际计算复杂度的时候,需要合适地选择数据的长度.高阶复杂度存在一定的限度,超过一定的阶次限度后,复杂度的数值将会出现振荡,失去了原有的稳定性.合理地选择高阶复杂度的阶次,以增加时间序列复杂度的区分度,同时保持原有的稳定性,可以使得在对时间序列的复杂度分析的时候能获得更理想的效果.

#### 参考文献:

- [1] 侯 威,封国林,董文杰.基于复杂度分析 logistic 映射和 Lorenz 模型的研究 [J].物理学报,2005,54(8):3940-3946.
- [2] Shaw Fuzen, Chen Rueifeng, Tsao Henwai, Algorithmic complexity as an index of cortical function in awake and pentobarbital-anesthetized rats [J]. Journal of Neuroscience Methods, 1999, 93(2): 101.
- [3] Kolmogorov AN. Three approaches to the quantitative definition of information [J]. Inform Tuams, 1965, 1: 3-11.
- [4] Lempel A, Ziv J. On complexity of finite sequences [J]. IEEE Trans Inform Theor, 1976, 22: 75-88.
- [5] Kasper K, Schuster HG. Easily calculable measure for complexity of sparial-temporal, pattens [J]. Physics Review(A), 1987, 36: 843-848.
- [6] 陈仲永,伍文凯,童勤业,等.基于复杂性测度的帕金森病人 EEG 分析 [J].生物医学工程杂志,1999,16(2):218-221.
- [7] Xu JH, Liu ZR, Liu R. The Measures of Sequence Complexity for EEG Studies [J]. Chaos Soliton And Fractals, 1994, 4: 2111-2119.
- [8] 孟 欣,邹 睿,邵 颀,等.简化 EEG 为符号序列及对其复杂度的研究 [J].山东生物医学工程,1995,14(2):9-13.

## Study on the Stabilization of Complexity Measure of Nonlinear Time Sequence

LUO Ai-ping

Department of Economic Management, Xichang College, Xichang Sichuan 615000, China

**Abstract:** The complexity measure is an important dynamic index to describe nonlinear time sequence. The proper length of sequence and order of complexity must be determined to ensure the validity and stability of results before calculating the complexity measures. In this paper, Lorenz model and clinical EEG signal are used as samples. Based on results of complexity measure in different data length and order, the influences of different parameters to measure are discussed and a suggestion of proper parameter is represented.

**Key words:** nonlinear time sequence; complexity measure; stabilization; Lorenz model; EEG signal

责任编辑 潘春燕