

文章编号: 1000-0887(2005) 09_1083_08

C_0 复杂度的数学基础*

沈恩华¹, 蔡志杰², 顾凡及¹

- (1. 复旦大学 生命科学学院 脑科学研究院脑科学研究中心, 上海 200433;
2. 复旦大学 数学科学学院 非线性科学研究中心, 上海 200433)

(戴世强推荐)

摘要: 对于许多同时具有强烈非线性和非平稳性的连续生物医学信号来说, 计算其复杂度往往要求: 1) 在数据长度比较短的情况下也可以得出比较鲁棒的估计值; 2) 无需对原始信号作像二值化这样的过分的粗粒化。我们以前所提出的 C_0 复杂度就是这样的一种度量, 但是这种度量缺乏严格的数学基础, 因而影响到它的应用。提出了一种改进形式, 并严格证明了它的重要性质, 从而表明这个量在一定条件下可以作为时间序列随机程度的指标, 因而在随机性复杂度的意义下也可作为复杂性的一个定量指标。由于这个量有计算速度快的优点, 因此特别适合于一些需要大量计算复杂度的场合, 例如计算长时间过程中滑动窗口中复杂度的动态变化。

关键词: 复杂性度量; 随机性复杂度; C_0 复杂度

中图分类号: TN911.72; TN911.73 文献标识码: A

引 言

有许多生物医学信号(例如脑电信号), 由于其强烈的非线性, 使得传统的线性分析手段不能得出理想的结果。而像 Liapunov 指数和经典的各种分维数这样的混沌动力学指标由于一般说来需要长的数据序列才能得出鲁棒的估计值, 又和这些信号强烈的非平稳性发生矛盾。因此找出一种适用于表征像脑电这样的同时具有强烈的非线性和非平稳性的时间序列的复杂度指标就成为一个迫切需要加以解决的问题。上世纪 90 年代初, 徐京华等人^[1]提出用 Lempel-Ziv 复杂度 C_{LZ} 作为指标来表征脑的一些不同的功能状态, 取得了一些有希望的结果, 较好地解决了上述矛盾。但是生物医学信号往往是连续信号, 而对 C_{LZ} 有定义的时间序列的元素只能取自有限集合, 因此他们在计算时先对原始信号进行粗粒化, 把原来的连续信号转换成某种二值信号, 例如把所有其值大于平均值的采样值都用 1 来代替, 而把所有其值小于等于平均值的采样值都用 0 来代替。这种过分的粗粒化潜藏着根本改变信号的动力学特性的危险, 例如一个原来是混沌的信号在经过这种过分粗粒化之后有可能变成一个周期信号。我们^[2,3]曾经以 Logistic 映射

* 收稿日期: 2003_07_18; 修订日期: 2005_05_20

基金项目: 国家自然科学基金资助项目(70271065; 10201008)

作者简介: 沈恩华(1977—), 男, 上海人, 博士;

蔡志杰(1968—), 男, 上海人, 副教授, 博士(联系人. Tel: + 86_21_65642469; Fax: + 86_21_65642342; E_mail: zhijiecai@163.com)•

$$x_{n+1} = 1 - \mu x_n^2, \quad \mu \in [0, 2]$$

所产生的时间序列为例, 说明当参数 μ 的值落在大致为 (1.4, 1.5) 的区间内时, 由此产生的时间序列已经落入混沌区, 应该有较高的复杂度值。然而如果采用把超过平均值的值置为 1, 反之置为 0 的方法进行粗粒化, 那么经粗粒化以后所得的序列的复杂度在此区间内为 0。究其原因, 这是由于原始序列在均值上下来回跳动, 但是其跳动的幅度是不规则的, 表现出某种混沌特性; 而在粗粒化之后, 跳动幅度被粗粒化成了 1 和 0, 于是粗粒化之后的序列就变成了 0、1 相间的周期序列, 其复杂度自然也就变为 0 了。显然, 这种潜在的危险性使得我们在用 C_{LZ} 来表征量值作连续变化的时间序列时必须十分小心。因此, 对于这种信号来说, 一个比较理想的复杂度指标应满足: 1) 在数据长度比较短的情况下也可以得出比较鲁棒的估计值; 2) 对连续信号也有定义, 从而无需对原始信号作像二值化这样的过分的粗粒化。

我们曾经提出过一种满足这两个要求的 C_0 复杂度^[3], 其定义如下: 对原始序列作快速 Fourier 变换(FFT), 求其幅值谱的平均值, 保留幅值大于平均值的频谱成分而将其余的置为 0, 从而得到一个新的频谱, 对此进行 Fourier 逆变换, 这样得到的时间序列与原始信号之差环绕其均值而成的面积和原始信号环绕它本身均值得到的面积之比, 就是我们所谓的 C_0 复杂度。然而, C_0 复杂度的一个严重缺陷是缺乏严格的数学基础, 这不能不大大降低了由此得出的结论的可信性。本文中, 我们将提出 C_0 复杂度的一种修正形式, 严格证明了它的一些主要性质。为了简单起见, 我们把这种修正后的指标仍记为 C_0 , 并把它试用于脑电分析, 从而说明这可能是适用于像脑电这样的非平稳非线性信号的复杂度指标。

1 C_0 复杂度的定义

记 $\{f(k), k = 0, 1, 2, \dots, N-1\}$ 是一个长度为 N 的时间序列, 则

$$F_N(j) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) e^{-2\pi i k j / N}, \quad j = 0, 1, 2, \dots, N-1$$

构成相应的 Fourier 变换序列, 其中 $i = \sqrt{-1}$ 是虚数单位。为了书写方便, 记 $W_N = e^{2\pi i / N}$, 于是 $F_N(j)$ 可写为

$$F_N(j) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) W_N^{-kj}, \quad j = 0, 1, 2, \dots, N-1$$

设 $\{F_N(j), j = 0, 1, 2, \dots, N-1\}$ 的均方值为

$$G_N = \frac{1}{N} \sum_{j=0}^{N-1} |F_N(j)|^2,$$

记

$$F_N(j) = \begin{cases} F_N(j), & \text{当 } |F_N(j)|^2 > G_N \text{ 时,} \\ 0, & \text{当 } |F_N(j)|^2 \leq G_N \text{ 时,} \end{cases}$$

对 $\{F_N(j), j = 0, 1, 2, \dots, N-1\}$ 作 Fourier 逆变换

$$f(k) = \sum_{j=0}^{N-1} F_N(j) W_N^{kj}, \quad k = 0, 1, 2, \dots, N-1,$$

定义 C_0 复杂度为

$$C_0 = \sum_{k=0}^{N-1} |f(k) - \bar{f}|^2 \sqrt{\left(\sum_{k=0}^{N-1} |f(k)|^2 \right)}.$$

2 C_0 复杂度的主要性质

性质 1 若 $\{f(k)\}$ 是常数序列, 则 $C_0 = 0$.

证 记 $f(k) \equiv c (k = 0, 1, 2, \dots, N-1)$, 则

$$F_N(j) = c\delta(j) = \begin{cases} c, & j = 0, \\ 0, & j \neq 0, \end{cases}$$

$$G_N = \frac{c^2}{N} < |F_N(0)|^2,$$

故

$$F_N(j) = \begin{cases} F_N(0) = c, & j = 0, \\ 0, & j \neq 0, \end{cases}$$

于是 $f(k) \equiv c$. 由 C_0 的定义可得 $C_0 = 0$.

注 1 若 $f(k) \equiv 0 (k = 0, 1, 2, \dots, N-1)$, 定义 $C_0 = 0$.

性质 2 若 $\{f(k)\}$ 是周期 T 序列, 即 $f(k+T) \equiv f(k)$, 则 $\lim_N C_0 = 0$.

证 记 $M = N/T$ (即 $N = TM$) 表示序列中有 M 个周期, 则

$$F_N(j) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) W_N^{-kj} = \frac{1}{N} \sum_{m=0}^{M-1} \sum_{t=0}^{T-1} f(t) W_N^{-tj} W_N^{-mj} = \frac{1}{N} \left[\sum_{t=0}^{T-1} f(t) W_N^{-tj} \right] \left[\sum_{m=0}^{M-1} W_N^{-mj} \right].$$

记 $j = pM + q (p = 0, 1, 2, \dots, T-1; q = 0, 1, 2, \dots, M-1)$, 并注意到 $W_N^M = e^{2\pi i} = 1$, 有

$$F_N(pM + q) = \frac{1}{N} \left[\sum_{t=0}^{T-1} f(t) W_N^{-t(pM+q)} \right] \left[\sum_{m=0}^{M-1} W_N^{-m(pM+q)} \right] = \frac{1}{N} \left[\sum_{t=0}^{T-1} f(t) W_N^{-t(pM+q)} \right] M\delta(q).$$

于是, 当 $q \neq 0$ 时, $F_N(pM + q) = 0$. 当 $q = 0$ 时, 由于 $W_N^M = e^{(2\pi i/N)M} = e^{2\pi i/T} = W_T$, 有

$$F_N(pM) = \frac{1}{T} \sum_{t=0}^{T-1} f(t) W_N^{-tpM} = \frac{1}{T} \sum_{t=0}^{T-1} f(t) W_T^{-tp} = F_T(p), \quad p = 0, 1, \dots, T-1,$$

故 $\{F_N(j)\}$ 除 T 个元素外均为 0, 而这 T 个元素的值恰好是一个周期的 Fourier 变换值, 与周期个数 M 无关, 即与序列长度 N 无关, 因此

$$G_N = \frac{1}{N} \sum_{p=0}^{T-1} |F_T(p)|^2 \rightarrow 0 \quad (N \rightarrow \infty).$$

所以当 N 充分大时, $F_N(j) = F_T(j)$, 即有 $f(k) = f(k)$, 于是

$$\lim_N C_0 = \lim_N \left\{ \sum_{k=0}^{N-1} |f(k) - f(k)|^2 \sqrt{\sum_{k=0}^{N-1} |f(k)|^2} \right\} = 0.$$

性质 3 对任一时间序列 $\{f(k)\}$, 成立 $0 \leq C_0 \leq 1$.

证 记 x^* 表示复数 x 的共轭复数, 则

$$\begin{aligned} \sum_{k=0}^{N-1} |f(k)|^2 &= \sum_{k=0}^{N-1} f(k)f^*(k) = \sum_{k=0}^{N-1} \left[\sum_{j=0}^{N-1} F_N(j) W_N^{jk} \right] \left[\sum_{l=0}^{N-1} F_N^*(l) W_N^{-lk} \right] = \\ &= \sum_{k=0}^{N-1} \left[\sum_{j=0}^{N-1} \sum_{l \neq j} F_N(j) F_N^*(l) W_N^{(j-l)k} + \sum_{j=0}^{N-1} F_N(j) F_N^*(j) \right] = N \sum_{j=0}^{N-1} |F_N(j)|^2. \end{aligned}$$

由 C_0 的定义,

$$C_0 = \sum_{k=0}^{N-1} |f(k) - f(k)|^2 \sqrt{\left(\sum_{k=0}^{N-1} |f(k)|^2 \right)} = \\ N \sum_{j=0}^{N-1} |F_N(j) - F_N(j)|^2 \sqrt{\left(N \sum_{j=0}^{N-1} |F_N(j)|^2 \right)} = \\ \sum_{j=0}^{N-1} |F_N(j) - F_N(j)|^2 \sqrt{\left(\sum_{j=0}^{N-1} |F_N(j)|^2 \right)}.$$

显然, $C_0 \geq 0$, 而 $|F_N(j) - F_N(j)| \leq |F_N(j)|$ ($j = 0, 1, 2, \dots, N-1$), 故 $C_0 \leq 1$.

注2 由性质3的证明过程可知, 在实际计算时, 只要计算原始序列的 Fourier 变换, 不需要作逆变换. 这样可以节省约一半的计算时间.

注3 若定义

$$\tilde{F}_N(j) = \begin{cases} 0, & \text{当 } |F_N(j)|^2 > G_N \text{ 时,} \\ F_N(j), & \text{当 } |F_N(j)|^2 \leq G_N \text{ 时,} \end{cases}$$

则

$$C_0 = \sum_{j=0}^{N-1} |\tilde{F}_N(j)|^2 \sqrt{\left(\sum_{j=0}^{N-1} |F_N(j)|^2 \right)}.$$

性质4 设 $\{f(k)\}$ 是一个随机时间序列, 服从独立同分布, 且有有限的4阶矩. 记其均值为 $Ef = \mu$, 方差为 $Df = \sigma^2$, 则当 $N \rightarrow \infty$ 时, C_0 以概率1收敛于 $\sigma^2 / (\mu^2 + \sigma^2)$. 特别地, 当 $\mu = 0$ 时, C_0 以概率1收敛于1.

证 由于

$$|F_N(j)|^2 = F_N(j) F_N^*(j) = \frac{1}{N^2} \left[\sum_{k=0}^{N-1} f(k) W_N^{kj} \right] \left[\sum_{l=0}^{N-1} f(l) W_N^{lj} \right]^* = \\ \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} f(k) f(l) W_N^{(l-k)j},$$

其数学期望为

$$E |F_N(j)|^2 = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E[f(k) f(l)] W_N^{(l-k)j}.$$

由于 $f(k)$ 与 $f(l)$ 独立同分布, 因此

$$\text{当 } k \neq l \text{ 时, } E[f(k) f(l)] = Ef(k) Ef(l) = \mu^2, \\ \text{当 } k = l \text{ 时, } E[f(k) f(l)] = Ef^2(k) = \mu^2 + \sigma^2,$$

于是

$$E |F_N(j)|^2 = \frac{1}{N^2} \sum_{k=0}^{N-1} Ef^2(k) + \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l \neq k} E[f(k) f(l)] W_N^{(l-k)j} = \\ \frac{\mu^2 + \sigma^2}{N} + \frac{\mu^2}{N^2} \sum_{k=0}^{N-1} \sum_{l \neq k} W_N^{(l-k)j}.$$

当 $j = 0$ 时,

$$E |F_N(0)|^2 = \frac{\mu^2 + \sigma^2}{N} + \frac{\mu^2}{N^2} N(N-1) = \mu^2 + \frac{\sigma^2}{N}.$$

当 $j \neq 0$ 时,

$$E |F_N(j)|^2 = \frac{\mu^2 + \sigma^2}{N} + \frac{\mu^2}{N^2} \left[\sum_{k=0}^{N-1} \sum_{l=0}^{N-1} W_N^{(l-k)j} - 1 \right] = \frac{\mu^2 + \sigma^2}{N} - \frac{\mu^2}{N} = \frac{\sigma^2}{N}.$$

于是

$$EG_N = E\left[\frac{1}{N} \sum_{j=0}^{N-1} |F_N(j)|^2\right] = \frac{1}{N} \left[\left(\mu^2 + \frac{\sigma^2}{N}\right) + \frac{\sigma^2}{N}(N-1) \right] = \frac{\mu^2 + \sigma^2}{N}.$$

利用 Chebyshev 不等式可以证明(见引理 1): 对任意给定的 $\varepsilon > 0$ 成立

$$\lim_{N \rightarrow \infty} P(| |F_N(j)|^2 - E |F_N(j)|^2 | > \varepsilon) = 0,$$

$$\lim_{N \rightarrow \infty} P(|G_N - EG_N| > \varepsilon) = 0,$$

故当 N 充分大时, 以概率 1 成立 $|F_N(0)|^2 > G_N$, $|F_N(j)|^2 < G_N$ ($j \neq 0$), 于是当 N 充分大时, 以概率 1 成立

$$C_0 = \sum_{j=0}^{N-1} |F_N(j) - F_N(j)|^2 \left\{ \sum_{j=0}^{N-1} |F_N(j)|^2 \right\} = \sum_{j=0}^{N-1} |F_N(j)|^2 \left\{ \sum_{j=0}^{N-1} |F_N(j)|^2 \right\} = \frac{\sigma^2}{N}(N-1) \left\{ \mu^2 + \sigma^2 \right\} \rightarrow \frac{\sigma^2}{\mu^2 + \sigma^2}.$$

在性质 4 的证明过程中, 需要以下引理:

引理 1 在性质 4 的条件下, 对任意给定的 $\varepsilon > 0$ 成立

$$\lim_{N \rightarrow \infty} P(| |F_N(j)|^2 - E |F_N(j)|^2 | > \varepsilon) = 0,$$

$$\lim_{N \rightarrow \infty} P(|G_N - EG_N| > \varepsilon) = 0.$$

证 我们只证第 1 式, 第 2 式类似可证.

记 $X = |F_N(j)|^2 = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} f(k)f(l)W_N^{(l-k)j}$. 在性质 4 的证明中我们已经得到

$$EX = \begin{cases} \mu^2 + \frac{\sigma^2}{N}, & j = 0, \\ \frac{\sigma^2}{N}, & j \neq 0. \end{cases}$$

又

$$EX^2 = E |F_N(j)|^4 = \frac{1}{N^4} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} E[f(k)f(l)f(p)f(q)] W_N^{(l-k+q-p)j}.$$

1) 当 $j = 0$ 时,

$$EX^2 = \frac{1}{N^4} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} E[f(k)f(l)f(p)f(q)] = \frac{1}{N^4} [P_N^4 \mu^4 + P_N^4 C_4^2 P_{N-1}^2 \mu^2 (\mu^2 + \sigma^2) + C_4^2 C_N^2 (\mu^2 + \sigma^2)^2 + C_4^2 P_N^2 \mu E f^3(k) + N E f^4(k)].$$

由于

$$\mu_3 = E(f - Ef)^3 = E[f^3 - 3Ef \cdot f^2 + 3(Ef)^2 f - (Ef)^3] = Ef^3 - \mu^3 - 3\mu\sigma^2,$$

$$\mu_4 = E(f - Ef)^4 = E[f^4 - 4Ef \cdot f^3 + 6(Ef)^2 f^2 - 4(Ef)^3 f + (Ef)^4] = Ef^4 - 4\mu\mu_3 - \mu^4 - 6\mu^2\sigma^2,$$

故

$$Ef^3 = \mu_3 + \mu^3 + 3\mu\sigma^2, \quad Ef^4 = \mu_4 + 4\mu\mu_3 + \mu^4 + 6\mu^2\sigma^2,$$

于是

$$EX^2 = \frac{1}{N^4} [N(N-1)(N-2)(N-3)\mu^4 + 6N(N-1)(N-2)\mu^2(\mu^2 + \sigma^2) +$$

$$3N(N-1)(\mu^2 + \sigma^2)^2 + 4N(N-1)\mu(\mu_3 + \mu^3 + 3\mu\sigma^2) + N(\mu_4 + 4\mu\mu_3 + \mu^4 + 6\mu^2\sigma^2) = \mu^4 + \frac{6}{N}\mu^2\sigma^2 + \frac{3(N-1)}{N^3}\sigma^4 + \frac{4}{N^2}\mu\mu_3 + \frac{1}{N^3}\mu_4$$

因此

$$DX = EX^2 - (EX)^2 = \frac{4}{N}\mu^2\sigma^2 + \frac{2N-3}{N^3}\sigma^4 + \frac{4}{N^2}\mu\mu_3 + \frac{1}{N^3}\mu_4 = O\left(\frac{1}{N}\right).$$

2) 当 $j \neq 0$ 时,

$$EX^2 = \frac{1}{N^4} \left\{ \sum_{k \neq l \neq p \neq q} E[f(k)f(l)f(p)f(q)] W_N^{(l-k+q-p)j} + \sum_{k \neq l} E[f(k)f(l)f(p)f(q)] W_N^{(l-k+q-p)j} \right\} = I + II,$$

对 II, 由于 f 的 1~4 阶矩均为有限, 即 $|\mu_1|, |\mu_2|, |\mu_3|, |\mu_4| \leq M$, 所以

$$|II| \leq \frac{1}{N^4} \sum_{k \neq l} |E[f(k)f(l)f(p)f(q)]| |W_N^{(l-k+q-p)j}| \leq \frac{M}{N^4} \sum_{k \neq l} 1 = \frac{M}{N^4} (6N(N-1)(N-2) + 3N(N-1) + 4N(N-1) + N) = O\left(\frac{1}{N}\right).$$

而对 I 有

$$\begin{aligned} |I| &= \frac{\mu^4}{N^4} \sum_{k=0}^{N-1} \sum_{l \neq k} \sum_{p \neq k, l} \sum_{q=0}^{N-1} [W_N^{(l-k+q-p)j} - W_N^{(l-p)j} - W_N^{(2l-k-p)j} - W_N^{(l-k)j}] = \\ &= \frac{\mu^4}{N^4} \sum_{k=0}^{N-1} \sum_{l \neq k} \sum_{p=0}^{N-1} [-W_N^{(l-p)j} - W_N^{(2l-k-p)j} + \\ &W_N^{(l-k)j} + W_N^{(2l-2k)j} + 1 + W_N^{(l-k)j} - (N-2)W_N^{(l-k)j}] = \\ &= \frac{\mu^4}{N^4} \sum_{k=0}^{N-1} \left[\sum_{l=0}^{N-1} (W_N^{(2l-2k)j} - (N-4)W_N^{(l-k)j}) - 1 + (N-4) + (N-1) \right] = \\ &= \frac{\mu^4}{N^4} \sum_{k=0}^{N-1} (2N-6) = \frac{\mu^4}{N^3} (2N-6) = O\left(\frac{1}{N^2}\right). \end{aligned}$$

于是

$$EX^2 \leq O\left(\frac{1}{N}\right), \quad DX = EX^2 - (EX)^2 \leq O\left(\frac{1}{N}\right).$$

由此可知, 对任意 $j = 0, 1, 2, \dots, N-1$ 均成立

$$DX \leq O\left(\frac{1}{N}\right),$$

由 Chebyshev 不等式, 对任意给定的 $\varepsilon > 0$, 有

$$P(|X - EX| > \varepsilon) \leq \frac{DX}{\varepsilon^2} \leq O\left(\frac{1}{N\varepsilon^2}\right),$$

故

$$\lim_{N \rightarrow \infty} P(|X - EX| > \varepsilon) = 0,$$

即第 1 式成立.

3 C_0 复杂度在脑电分析中应用举例

为了检验经过改进之后的 C_0 复杂度是否适用于像脑电这样的既具有高度的非线性又具

有高度的非平稳性的信号分析,我们以人们经常用来进行脑电复杂度分析的睁眼休息,闭眼休息,浅睡和深睡这4种典型状态下的脑电信号作为例子,计算其 C_0 复杂度。实验方法是标准的,已有报道^[4]。由于脑电的极度非平稳性,在计算脑电的 C_0 复杂度时,每次我们都只取窗口宽度为 0.5 s 的数据(由于采样率为 1 024 Hz,故每次计算可得 512 个数据点),在时间轴上每次滑动 0.5 s 重复计算,可以观察到即使在同一种状态下,脑电的 C_0 复杂度也是随时间变化的——非平稳性!图 1 是对两名被试者在脑电 10—20 系统的 Fp1、Fp2、C3、C4、T3、T4、O1、O2 分别记录到的各约 10 s 的脑电,将 0.5 s 的窗口每次滑动 0.5 s 所计算得的 C_0 复杂度均值和方差。由图可见,除了 O1 之外,所有各导的脑电的 C_0 复杂度都在睁眼休息时最高,而在深睡时最低,闭眼休息和浅睡则依次介于两者之间。这和我们的直观是一致的,和前人用其它复杂度所得到的结果也是一致的^[4]。

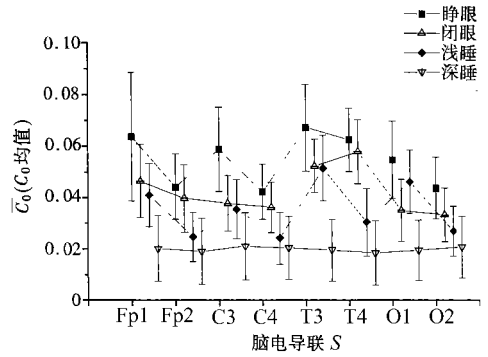


图 1 在睁眼休息,闭眼休息,浅睡和深睡 4 种不同状态下,记录所得的脑电 C_0 复杂度的平均值和方差

4 讨 论

关于时间序列的复杂度,现在有着种种不同的定义。最近 Rapp 和 Schmah^[5]把这些不同定义下的复杂度作了种种分类。按照他们的一种分类标准,可以把现有的复杂度分为两大类:“随机性复杂度”(randomness finding complexity)和“规则复杂度”(rule finding complexity)。前一类实际上是把序列接近随机的程度作为复杂性的度量^[6,7]。而按照我们上面所证明的性质, C_0 是一个介于 0 和 1 之间的数,对于常数序列和周期序列来说,其值趋于 0,而对于满足一定条件的随机序列来说则以概率 1 收敛于 1。因此,把 C_0 作为随机程度的一种度量看来是合理的,这是一种“随机性复杂度”。我们最近指出^[8],对于随机性复杂度来说,其一阶复杂度度量随机性,而其高阶复杂度则度量非平稳性。因此, C_0 复杂度也可望度量复杂性的这些不同方面。

满足我们在引言中所提出的两个要求的复杂度,除了我们在本文中所提出的 C_0 复杂度之外,近似熵(ApEn)^[7]是人们常用的一种复杂度指标。和近似熵比较起来, C_0 复杂度由于其计算主要是 FFT,而且由注 2 可知,实际上为了计算 C_0 复杂度只要计算一次 FFT 就可以了,因而有计算量小的优点。近年来,人们对于某些信号在长时间过程中所表现出来的复杂度的动态变化过程产生了越来越大的兴趣。人们通过计算滑动窗口中的复杂度来检测这种变化^{[3],[9]},以解决诸如癫痫预报这样的实际问题^[9]。显然,对于这种问题来说,由于问题牵涉到要计算大量滑动窗口中的复杂度,在实际应用中还希望能进行实时计算,因此提高计算速度有着十分重要的意义,可以预期 C_0 复杂度在这类问题中将有其优势。

致谢 本文中所用的脑电信号均由中国科学院上海生物化学研究所徐京华教授和北京中医药大学杨秦飞教授提供,特此致谢。

[参 考 文 献]

- [1] 吴祥宝, 徐京华. 复杂性和脑功能[J]. 生物物理学报, 1991, 7(1): 103—106.
- [2] 孟欣, 沈恩华, 陈芳, 等. 脑电图复杂度分析中的粗粒化问题 I: 过分粗粒化和三种复杂度的比较[J]. 生物物理学报, 2000, 16(4): 701—706.
- [3] Chen F, Xu J, Gu F, et al. Dynamic process of information transmission complexity in human brains [J]. Biological Cybernetics, 2000, 83(4): 355—366.
- [4] 杨斯环, 杨秦飞, 石继明, 等. 不同生理状态时脑电时间序列的三种复杂度计算比较[J]. 生物物理学报, 1996, 12(3): 437—440.
- [5] Rapp P E, Schmah T I. Dynamical analysis in clinical practice[A]. In: Lehnertz K, Arnhold J, Grassberger P, et al, Eds. Chaos in Brains [C]. Singapore: World Scientific, 2000, 52—62.
- [6] Lempel A, Ziv J. On complexity of finite sequences[J]. IEEE Transactions on Information Theory, 1976, IT22(1): 75—81.
- [7] Pincus S M. Approximate entropy as a measure of system complexity[J]. Proceedings of the National Academy of Sciences of the United States of America, 1991, 88(6): 2297—2301.
- [8] Gu F, Shen E, Meng X, et al. Higher order complexity of time series[J]. The International Journal of Bifurcation and Chaos, 2004, 14(8): 2979—2990.
- [9] Lehnertz K, Elger C E. Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity[J]. Physical Review Letters, 1998, 80(22): 5019—5022.

Mathematical Foundation of C_0 Complexity

SHEN En_hua¹, CAI Zhi_jie², GU Fan_ji¹

(1. School of Life Science, Research Center for Brain Science, Institute of Brain Science, Fudan University, Shanghai 200433, P. R. China;

2. School of Mathematical Sciences, Research Center for Nonlinear Science, Fudan University, Shanghai 200433, P. R. China)

Abstract: For many continuous bio_medical signals with both strong nonlinearity and non_stationarity, two criterions were proposed for their complexity estimation: 1) Only short data set is enough for robust estimation; 2) No over_coarse graining preprocessing, such as transferring the original signal into a binary time series, is needed. C_0 complexity measure proposed by us previously is one of such measures. However, it lacked solid mathematical foundation and thus its use was limited. A modified version of this measure is proposed, and some important properties are proved rigorously. According to these properties, this measure can be considered as an index of randomness of time series in some senses, and thus also a quantitative index of complexity under the meaning of randomness finding complexity. Compared with other similar measures, this measure seems more suitable for estimating a large quantity of complexity measures for a given task, such as studying the dynamic variation of such measures in sliding windows of a long process, owing to its fast speed for estimation.

Key words: complexity measure; randomness finding complexity; C_0 complexity