

分区复杂性在时间序列分析中的应用

李凌云¹, 童勤业²

(1. 浙江大学生物医学工程系, 杭州 310027; 2. 浙江大学非线性理论及应用研究中心, 杭州 310027)

摘 要: 本文对 KC 复杂性, C_1 、 C_2 复杂性等常见的时间序列分析中所用的方法进行了比较, 在此基础上提出一种新算法——分区复杂性, 同时将其应用在 EEG 时间序列分析、处理中, 以检验它的有效性.

关键词: 时间序列; 复杂性算法; 分区复杂性

中图分类号: TN919.5 **文献标识码:** A **文章编号:** 0372-2112 (2000) 12-0097-02

Time Series Analysis Using Partition Complexity Measure

LI Lingyun¹, TONG Qinye²

(1. Dept of Biomedical Engineering, Zhejiang University, Hangzhou 310037, China)

2. Research Center for Nonlinear Theory and Application, Zhejiang University, Hangzhou 310027, China)

Abstract: In this paper we discuss several important complexity algorithms for time series, such as Kolmogorov Complexity (KC) and C_1 , C_2 complexity. Had analyzed these methods, we bring forward a new complexity measure: partition complexity. The new arithmetic overcomes the disadvantage of those older methods. It sufficiently reflects its advantage through clinic tests.

Key words: time series; complexity measures; partition complexity

1 引言

复杂性测度作为一个重要的非线性指标在非线性的时间序列的分析处理中得到了越来越广泛的应用. 国内苏州大学的谢惠民教授在理论上比较系统全面的论述复杂性^[7], 徐京华先生^[4]首先将 KC 用于研究人脑 EEG 信息传输的复杂性, 同时提出了 C_1 、 C_2 两种复杂性测度. 将 KC、 C_1 、 C_2 等复杂性应用在 EEG 时间序列的分析处理中, 也得到了一定的结果^[6]. 但复杂性的定义并不统一, 不同定义的复杂性测度只适用于一定的范围, 存在各自的局限性^[1,2]. 对 KC 复杂性, C_1 、 C_2 复杂性等常见的时间序列分析中所用的方法进行了比较, 分析了它们的优缺点, 在此基础上提出了一种新的复杂性算法——分区复杂性, 并在临床脑电中检验了它的有效性.

2 分区复杂性的提出

2.1 目前复杂性算法存在的问题

复杂性测度计算中对状态空间进行粗粒化划分, 必然会导致某些细节信息的丢失. 以 KC 复杂性, C_1 、 C_2 复杂性的计算为例.

KC 复杂性是把时间序列粗粒化为符号序列, 然后用这一符号序列的复杂性来代替时间序列的复杂性, 它反映了一个时间序列随其长度的增长出现新模式的速率. C_1 、 C_2 复杂性也是针对粗粒化后的符号序列, 从符号动力学禁止字和允许字的基础上提出的.

例如, KC 是用二个符号表示的序列, 可理解为时间序列

$\{x_0, x_1, \dots, x_{n-1}, \dots\}$ 是由二个符号的映射系统产生的序列, 如图 1 所示. 可是实际情况并非如此, 一般情况下它应有多临界点如图 2 所示. 如果只考虑二个符号, 那么只能区分 X 的左右, 这就把 X_1 、 X_2 、 X_3 ……这些细节都去掉了. 而从符号动力学^[5]的观点来看, X_1 、 X_2 、 X_3 ……这些细节都是十分本质的东西. 针对这一点, 本文将状态空间细分, 由原来的两个状态变为 m 个状态. 并对算法复杂性做了改进. 因为是按区操作的, 故称为分区算法复杂性.

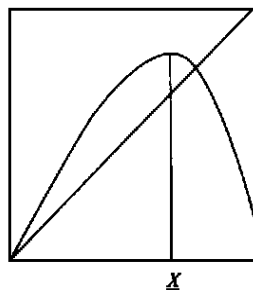


图 1

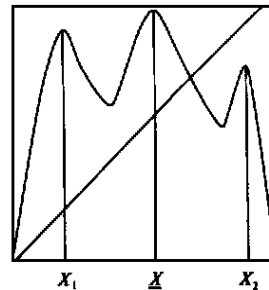


图 2

问题是不知道所研究的动力系统究竟有多少临界点, 因为象图 2, 若把它作为四个符号的动力系统来分析却是多余的. 为此提出了加权修正方案, 在同类试验中用大量数据来确定权数, 相当于来确定临界点的多少. 只有调整权数后的复杂性算法对特定对象有强说服力.

2.2 分区复杂性

分区复杂性是将状态空间细分, 考虑到更普遍的是状态空间的不均匀性和不同质性, 对状态空间划分时采用平均值分区. 方法如下:

时间序列 $\{X_0, X_1, \dots, X_{n-1}\}$ 首先求的这个序列的平均值 \bar{X} , 再把这个序列重构, 对大于平均值 \bar{X} 的 X_i , 令它对应的符号 $S_i^{(0)}$ 为 1, 记这些 X_i 的全体组成的集合为 C_1 ; 不大于平均值 \bar{X} 的 X_i , 令它对应的符号 $S_i^{(0)}$ 为 0, 记这些 X_i 的全体组成的集合为 C_0 . 这样就从时间序列 $\{X_0, X_1, \dots, X_{n-1}\}$ 得到一个 $(0, 1)$ 序列 $S^{(0)} = \{s_0^{(0)}, s_1^{(0)}, \dots, s_{n-1}^{(0)}\}$. 然后以同样方法将 C_1 、 C_0 符号化, 得到序列 $S^{(1)} = \{s_0^{(1)}, s_1^{(1)}, \dots, s_{n-1}^{(1)}\}; \dots$

通过这种分区方法最终可得到 2^m 个分区区间, 时间序列 $\{X_0, X_1, \dots, X_{n-1}\}$ 经过 m 次符号化后就得到 m 个 $(0, 1)$ 符号序列 $S^{(i)} = \{s_0^{(i)}, s_1^{(i)}, \dots, s_{n-1}^{(i)}\}$ 与之对应 $i = 0, 1, \dots, m-1$, 将 m 个 $(0, 1)$ 符号序列 $\{S_0^{(i)}, S_1^{(i)}, \dots, S_{n-1}^{(i)}\}$ 按行排列, 就组成了一个 $m \times n$ 阶矩阵, 记为 S .

记 $S^{(i)}$ 序列的复杂度为 $C(i)$, $i = 0, 1, \dots, m-1$, i 表示位值. 则总的复杂度 C 为:

$$C = \sum_{i=0}^{m-1} q_i C(i), \text{ 其中 } \sum_{i=0}^{m-1} q_i = 1 \quad (1)$$

本文使用的方法是分别计算这 m 个 $(0, 1)$ 序列的 KC 复杂性、 C_1 、 C_2 复杂性, 或涨落复杂性^[3], 最后通过加权平均的方法得到整个序列 X 的复杂度指标. 对大量同类试验的数据, 通过最优化方法调整不同位复杂性的权值, 就可以得到最佳指标.

通过研究这 m 个 $(0, 1)$ 符号序列的特性, 就可以得到原始序列 X 的特性, 即可通过符号动力学理论来研究时间序列 X 的性质. 这显然比只使用单个 $(0, 1)$ 序列要详细、精确.

3 用复杂性来分析精神分裂症病人的 EEG 时间序列

3.1 临床应用和结果

本文所用的脑电数据由澳大利亚墨尔本健康中心提供, 共有 400 人, 200 人已经由医生确诊为精神病人, 200 人为正常人, 每人有两种状态: 睁眼, 闭眼. 采样频率为 250 Hz. 记录导联采用国际脑电图标准, 记录了 32 道 EEG 数据, 分析计算时, 这里仅用了 8 道 ($fp1, fp2, c3, c3, t3, t4, o1, o2$). 结果如下:

(1) 使用以前的分区方法的处理结果:

表 1

		费歇判别		马氏距离	
		病人	正常人	病人	正常人
C_1	睁眼	81.0	82.2	73.6	88.0
	闭眼	72.5	74.6	76.3	85.2
C_2	睁眼	83.1	82.3	72.4	84.1
	闭眼	74.6	72.4	78.7	82.4
KC	睁眼	83.0	78.2	75.3	87.0
	闭眼	76.0	71.1	76.6	82.5

(2) 使用分区复杂性优化后的处理结果:

表 2

		费歇判别		马氏距离	
		病人	正常人	病人	正常人
C_1	睁眼	83.6	85.2	76.6	92.4
	闭眼	76.7	80.0	80.7	89.6
C_2	睁眼	86.4	88.2	76.1	88.0
	闭眼	79.3	80.9	80.3	86.0
KC	睁眼	86.1	82.7	79.9	92.1
	闭眼	81.8	74.5	80.5	88.4

3.2 结果和讨论

(1) 分区复杂性当权数只有 q_0 , 其它权数为零时, 这种算法就是 KC 复杂性, 或 C_1 、 C_2 复杂性. 当 $q_0 = q_1, \dots, q_{m-1}$, 采用了信息熵的算法, 这就是涨落复杂性. 针对大量同类实验数据进行权寻优, 使结果达到最佳, 这就是说这种算法效果大于或等于 KC 等复杂性.

(2) 分区复杂性克服了原有算法对状态空间进行划分时所出现的缺陷, 考虑了多临界点的映射, 通过调整权数可更大限度的提取时间序列中有用的信息, 从而提高对不同状态的区分率(1式). 我们将其应用在临床 EEG 中进行检验, 从表 1 和表 2 中可看出其判别率相对于原有算法提高了 3%~8%.

(3) 涨落复杂性虽然在一定程度上可克服粗粒化的符号方法所带来的缺陷, 但它相当于等位权的分区复杂性算法, 这种均一化方法, 在一定程度上没有突出时间序列中起关键性作用的变化, 分区复杂性可通过加大这一部分的位权来更好的反映出时间序列的本质.

用分区复杂性在临床 EEG 中已取得一定成果, 这不但为 EEG 分析提供了一个有力工具, 同时此法也可以用于其它非线性时间序列处理. 由于只是对分区复杂性算法作了一些初步尝试, 还有许多进一步完善的地方, 譬如权优化算法还需进一步探讨.

作者简介:



李凌云 1971 年 7 月出生. 1993 年毕业于湖南大学电气工程系本科. 毕业后在湖南大学电气系检测技术及自动化仪器仪表专业任教. 1997 年考入浙江大学生物医学工程系攻读硕士.



董勤业 1939 年 5 月 12 日出生. 浙江大学生物医学工程系教授, 仪器科学与工程系教授, 博士生导师. 浙江大学非线性科学理论与应用研究中心副主任. 主要从事非线性动力学的研究, 近年来完成了“心脏细胞非线性动力学研究”, “非线性动力学在临床脑电分析中的应用”, “EEG 所构成的奇怪吸引子模式识别”, “非线性技术在信息科学中的应用”等多项国家自然科学基金资助项目.

(下转第 93 页)

5 结论

本文提出了一种新的盲波束形成方法估计分布式目标波达方向. 与其它分布式目标 DOA 估计方法相比, 首先, 本文利用空间信号分布函数的共轭对称性, 提出的是一种一维 DOA 搜索估计方法, 避免了多维参数谱峰搜索或求解多维参数优化问题的迭代过程; 其次, 本文仅利用空间信号分布函数的共轭对称性信息, 不要求已知分布式目标空间信号分布的具体函数形式, 因此, 在空间信号的分布函数形式不同的分布式目标同时存在的情况下, 也能得到分布式目标 DOA 估计. 仿真实验结果表明, 这种高分辨的分布式目标 DOA 估计方法的估计性能对信号源的分布特性是不敏感的.

参考文献:

- [1] S. Valaee, B. Champagne, and P. Kabal. Parametric localization of distributed sources [J]. IEEE Trans. on Signal Processing, 1995, 43(9): 2144– 2153.
- [2] T. P. Jantti, The influence of the extended sources on the theoretical performance of MUSIC and ESPRIT method [C]: Narrow-band sources, Proceeding ICASSP, San Francisco, CA, 1992, 3: 11-429 11-432.

- [3] Y. Meng, K. Max, and Q. Wu. Direction finding for point and spread sources [C]: VEG MUSIC and its performance, Proceeding ICSP, Atlanta, 1996, 5: 2908– 2911.
- [4] Y. Meng, K. M. Wong, Q. Wu. Estimation of the direction of arrival of spread source in sensor array processing [C], Proceeding ICSP, Beijing, China, 1993, 10: 430– 434.
- [5] Y. Meng, P. Stoica, K. M. Wong. Estimation of the direction of arrival of spatially dispersed signals in array processing [J], IEE Proceeding F, 1996, 2, 143(1): 1– 9.
- [6] Y. U. Lee, J. Choi, I. Song, and S. R. Lee. Distributed source modeling and direction of arrival estimation techniques [J]. IEEE Trans. on Signal Processing, 1997, 45(4): 960– 969.
- [7] J. F. Cardoso, A. Souloumiac. Blind beamforming for non Gaussian signals [J], IEE Proceeding F, 1993, 140(6): 362– 370.

作者简介:

万 群 1971 年生, 1993 年在南京大学本科毕业, 1996 年在成都电子科技大学获硕士学位, 现为电子科技大学电子工程学院博士生, 从事阵列信号处理、稀疏目标及分布式目标波达方向估计的研究.

杨万麟 1945 年生, 电子科技大学电子工程学院教授, 博士生导师, 长期从事微波成像、信号处理、雷达系统和目标识别的科研、教学工作.

(上接第 98 页)

参考文献:

- [1] Peter Grassberger, Thomas Schreiber, and Carsten Schaffrath, Nonlinear time sequence analysis [J]. International Journal of Bifurcation and Chaos, 1991, 1(3): 521– 547.
- [2] W. R. Witt. A comparative classification of complexity measure [M]. Chaos Soliton & Fractals, 1994, 4: 133– 173.
- [3] J. E. Bates, H. K. Shepard. Information fluctuation as a measure of complexity [M]. University of New Hampshire, Durham (1991).

- [4] Jinghua Xu, Zengrong Liu. Information transmission in human cerebral cortex [M]. Physical D 106(1997) 363– 374.
- [5] Hao bar lin. Elementary Symbolic Dynamics and Chaos in Dissipative Systems [M]. World Scientific, 1989.
- [6] Qinye Tong, June Kong et al. A Note On Analyzing Schizophrenic EEG with Complexity Measure [M]. .. Chaos, Solitons & Fractals 7 (1996) 371– 375.
- [7] 谢惠民. 复杂性与动力系统(非线性科学丛书) [M]. 上海科技教育出版社, 1994 年第一版, 1995 年第二版.