

一种识别混沌时间序列动力学异同性的方法

李春贵¹⁾ 裴留庆²⁾

¹⁾ 武汉工业学院基础部, 武汉 430023)

²⁾ 北京师范大学电子系, 北京 100875)

(2002 年 11 月 19 日收到 2002 年 12 月 30 日收到修改稿)

定义了两个新的统计量, 分别记为 Q 和 R , Q 称为动力学自相关因子指数, R 称为动力学互相关因子指数. 用 Q 和 R 对不同时间序列的动力学“距离”或同一时间序列中不同区段(窗口)的动力学“距离”进行估算, 可以鉴别这些时间序列的动力学属性或它们内在的动力学结构层次以及复杂性. 一些典型实例检验表明, 这种方法是有效的.

关键词: 混沌时间序列, 动力学自相关因子指数, 动力学互相关因子指数

PACC: 0545

1. 引言——问题的提出

对于时间序列 $\{x(t_i)\}_{i=1}^N$, 在下文的讨论中, 我们将用时间延迟方法在嵌入空间上重构其动力学轨线^[1]. 重构嵌入向量表示为

$$x(t_i) = \{x(t_i), x(t_i + \tau), \dots, x(t_i + (m-1)\tau)\}.$$

其中, 时间延迟 $\tau = a \cdot \Delta t$, Δt 是采样时间, a 为延迟参数, m 是嵌入空间的维数. 集 X 的自关联和^[2]为

$$C_{XX}(\epsilon) = P(\|x(t_i) - x(t_j)\| < \epsilon) \\ = \frac{2}{(N-m)(N-m+1)} \\ \times \sum_{i=1}^{N-m} \sum_{j=i+1}^{N-m+1} \Theta(\epsilon - \|x(t_i) - x(t_j)\|),$$

表示在嵌入空间中重构距离 ϵ 内找到向量 $x(t_i)$ 的邻近点的概率. 其中 Θ 是 Heaviside 阶跃函数.

为了对集 X 和 Y 进行有效的比较, 还需计算交互关联和^[3]. 称 Grassberger 和 Procaccia 的关联和为自关联和, 集 X 和 Y 的交互关联和为

$$C_{XY}(\epsilon) = P(\|x(t_i) - y(t_j)\| < \epsilon) \\ = \frac{2}{(N-m)(N-m+1)} \\ \times \sum_{i=1}^{N-m} \sum_{j=i+1}^{N-m+1} \Theta(\epsilon - \|x(t_i) - y(t_j)\|),$$

其中 $C_{xy}(\epsilon)$ 关于 X 和 Y 是对称的. 交互关联和表示在 x 的 ϵ 邻域内找到点 y 的概率.

混沌信号的时间序列分析研究是目前非线性科学的前沿课题之一^[4]. 在描述混沌信号时, 关联和一

直受到广泛的关注. 自关联和及交互关联和具有一定的区分潜在动力学的的能力^[5], 但他们远不能作为识别混沌时间序列之间相近性的最重要的标志^[2,6-8]. 能否找到对不同的动力学有较高区分能力和比较能力的检验量? 甚至适用于较短的时间序列和小尺度下的情形?

2. Q 和 R

设 $x(i)$ 和 $x(j)$ 是时间序列 $x(n)$ (n 是离散的) 上的任意两点. 那么, 当 $\|x(i) - x(j)\| \leq \epsilon$ 时, $\|x(i+1) - x(j+1)\| \leq \epsilon$ 的概率为 $S_m = C_{XX}^{m+1}(\epsilon) / C_{XX}^m(\epsilon)$. 它比关联和具有更强的预见性, 可用于混沌时间序列动力学异同性的识别^[9,10]. 同理, 对于两个时间序列集 $\{x_i\}, \{y_i\}$, 比值 $C_{YY}^m(\epsilon) / C_{XX}^m(\epsilon)$ 也具有识别混沌时间序列动力学异同性的作用.

我们定义一个新的量

$$Q_{XY} = \lim_{\epsilon \rightarrow 0} \left| \ln \frac{C_{YY}(\epsilon)}{C_{XX}(\epsilon)} \right|,$$

Q 称作“动力学自相关因子指数”, Q 代表的意义是: 如果 Q 是统计上足够小的, 那么集 X 和集 Y 具有相近的动力学, 反之则不具有相近的动力学. 即 Q 能够反应轨线两部分是否具有相近的动力学. 也就是说, 它能起到直接测量两混沌时间序列之间的“距离”的作用.

Q 能够反应两轨线动力学的相近性, 但相近并不等于完全相同. 二者间的区别与联系就如同三角形的相似和全等之间的关系. 对混沌时间序列的识

别,停留在相近上是不够的,对于那些动力学相近的混沌时间序列,有必要进一步判断二者是否等价.为此,我们定义另一个新的量

$$R_{XY} = \lim_{\epsilon \rightarrow 0} \left| \ln \frac{C_{XY}(\epsilon)}{C_{XX}(\epsilon) \cdot C_{YY}(\epsilon)} \right|,$$

称作“动力学互相关因子指数”, R 代表的意义是:当 Q 是统计上足够小的前提下,如果 R 也是统计上足够小的,那么集 X 和集 Y 具有完全相同的动力学;反之,如果 R 不是统计上足够小的,则集 X 和集 Y 的动力学仅仅相近.

3. Q 和 R 作用的进一步讨论

当两个混沌时间序列源于完全相同的动力学时,则 Q 和 R 均必为零;反过来,如果 Q 和 R 均为零,说明两个混沌时间序列源于完全相同的动力学.下面用实例来说明这个问题.

Hénon 映射:

$$\begin{cases} x_{n+1} = y_n + 1 - a \cdot x_n^2 \\ y_{n+1} = b \cdot x_n \end{cases} \quad (a = 1.4, b = 0.3).$$

首先,由该映射产生 50000 个点(去掉暂态之后)的时间序列(为了计算方便起见,在分析之前将时间序列归一化,归一化后的时间序列的均值为 0,方差为 1.下文所分析数据均予以归一化,不再另行说明);然后从这个时间序列中挑出最初的 500 个点和最后的 500 个点组成两个较短的子时间序列,而这两个混沌时间序列就源于完全相同的动力学.最后计算这两个子时间序列的 Q 和 R 值.选取嵌入维数 $m = 2, \epsilon = 0.01$.结果发现,两个子时间序列的 Q 值和 R 值均等于零.这正是期望的结果.同时由以上实例也可看出,用 Q 和 R 识别两个混沌时间序列时,较短的数据也能够产生稳定的结果.

那么,两个混沌时间序列具有或源于完全不同的动力学时,情况又怎样呢?我们计算了两个完全不同的时间序列的 Q 和 R 值.其中第一个上面的标准 Hénon 映射的 x 分量,第二个时间序列源于 Baker 映射的 x 分量(其中 $\alpha + \beta = 1, \lambda_a + \lambda_b \leq 1, \alpha = 0.4, \beta = 0.6, \lambda_a = 0.6, \lambda_b = 0.4$).

Baker 映射:

$$x_{n+1} = \begin{cases} \lambda_a \cdot x_n, & \text{如果 } y_n < \alpha, \\ (1 - \lambda_b) + \lambda_b \cdot x_n, & \text{如果 } y_n > \alpha, \end{cases}$$

$$y_{n+1} = \begin{cases} y_n / \alpha, & \text{如果 } y_n < \alpha, \\ (y_n - \alpha) / \beta, & \text{如果 } y_n > \alpha. \end{cases}$$

每个序列都选 5000 个点,嵌入维数 $m = 2, \epsilon = 0.005$.计算结果是: $Q = 2.108560, R = 5.016684$. Q 和 R 都不等于零,这正是所预期的结果,它意味着两个混沌时间序列具有或源于完全不同的动力学.

那么, Q 和 R 在区分混沌时间序列时各自所起的作用又有什么不同呢?为了说明这个问题,我们计算标准的 Hénon 映射($a = 1.4, b = 0.3$)与参数稍作变化的 Hénon 映射($a = 1.43, b = 0.29$)二者间的 Q 和 R .将这两个映射的轨线画在同一个图上(图 1),它们的轨线稍有不同(两者近乎重叠),而关联维也稍有不同,前者为 $D_x \cong 1.2$,后者为 $D_y \cong 1.15$.

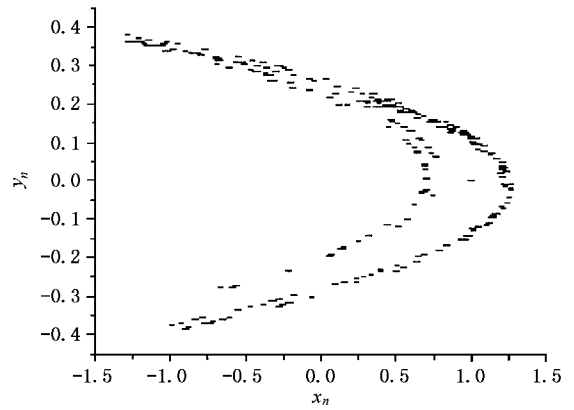


图 1 两个长度均为 50000 个点的 Hénon 映射($a = 1.4, b = 0.3$ 和 $a = 1.43, b = 0.29$)的轨线图

为了进行比较分析,我们分别从标准的 Hénon 映射与参数稍作变化的 Hénon 映射去掉前 500 暂态点后各取 50000 个点作为两个原始的时间序列(取点过程中,两个映射的初值均为 $X_0 = 0.001, Y_0 = 0.001$).然后从两个原始的时间序列中分别取出长度为 3500 的区间各两个,共组成 4 个窗口,即每个窗口含 3500 个点,其中窗口 1,2 是由标准的 Hénon 映射产生的,窗口 3,4 是由参数稍作变化后的 Hénon 映射产生的.嵌入维数取 $m = 2, \epsilon = 0.01$.表 1 列出了各个窗口之间的 Q, R 值.

表 1 各窗口之间的 Q, R 值

窗口	Q	R
1—2	0.011875	0.027623
1—3	0.054156	6.412272
1—4	0.003378	6.383505
2—3	0.066031	6.406335
2—4	0.008496	6.377568
3—4	0.057534	0.007782

从理论上说,由于 1, 2 两个窗口由标准的 Hénon 映射产生,它们具有相同的动力学,因而 Q_{12} 和 R_{12} 应该为 0. 但实际上,由于统计误差的原因使得 Q_{12} 和 R_{12} 只能很接近于 0 而不能严格等于 0. 同样的道理, Q_{34} 和 R_{34} 也只能是很接近于 0. (该例中, 误差 $e \approx 2\epsilon = 0.02$).

1) $Q = 0$ 时, $R = 0$ 表示两序列源于完全相同的动力学.

$R \neq 0$, 表示两序列源于相近的动力学.

2) $Q \neq 0$ 时, 无论 R 是否为 0, 都表示两序列源于完全不同的动力学.

4. 具有相近动力学属性的窗口族

一个复杂的时间序列可能是由一些有相互耦合的动力学共同作用产生的, 找出时间序列中具有相近动力学的区段(窗口)很有意义. 而且, 为了分析的需要, 可将几个具有相近的动力学的窗口组合在一起, 从而达到改进统计性能的目的.

把一个时间序列分割成 J 个窗口. 如果两个窗口之间的 Q 为零, 那么, 这两个窗口源于相近的动力学. 再把动力学相近的那些窗口放在一起, 就可将 J 个窗口的时间序列分成 K 组(或族). 同一组(或族)内的窗口可认为是由相近的动力学产生的, 而不同族中的窗口可认为是由不同的动力学产生的.

将动力学窗口分成族的程序和方法:

1) 设窗口族的个数为 m , m 的初值为 0; 没有组成窗口族的窗口数用 n_u 表示.

2) 计算任意两窗口之间的 Q_{ij} 和 R_{ij} ($i \neq j$).

3) 将窗口族的个数 m 增加 1 (也就是令 $m \rightarrow m + 1$).

4) 对某个确定的 i , 从 $\{Q_{ij}, j$ 为除 i 之外的 J 中的所有其他的窗口} 中找出那些满足 $Q_{ij} = 0$ 的窗口 $\{W_j, j$ 为除 i 之外的 J 中的所有其他的窗口}, 这些窗口就组成了一个族 G_m (注意第一个族中的第一个窗口就是整个的窗口序列中的第一个窗口).

5) 依次从剩余的未组成族的窗口中, 选出一个窗口, 重复 3)—4) 的过程, 直到没有未组成族的窗口为止.

6) 将上述结果中那些 $Q_{ij} \approx 0$ ($i \neq j$) 的窗口分别归属于 K 组(族), 从而可以鉴别相应时间序列中不

同区段的动力学异同性, 为理解其动力学复杂性和结构层次提供依据.

5. 模拟数据检验

本节将前面的方法应用于由计算机模拟产生的简单系统的数据中, 用一个具体的例子来检验 R 对不同时间系列的区分作用. 本例中所选用的产生时间序列的映射具有相同的自相关函数(和功率谱)^[10].

为了说明动力学互相关因子指数 R_{ij} 作为测量轨线之间距离的一种度规, 能够将不同的任意时间序列区分开来, 我们计算了如下三个映射之间的 R_{ij} :

$$x(t) = \eta(t); p(\eta) = \theta(\eta)\theta(1 - \eta) \quad (\text{uniform IID}), \quad (a)$$

$$x(t) = 2x(t-1)\theta(1/2 - x(t-1)) + \chi(1 - x(t-1))\theta(x(t-1) - 1/2) \quad (\text{帐篷映射}), \quad (b)$$

$$x(t) = 2x(t-1) \llbracket \text{mod } 1 \rrbracket, \quad (c)$$

其中 θ 是 Heaviside 阶跃函数, η 是随机数. 每个序列各产生 20000 个点, 共计 60000 个数据点. 所有这些映射在 $[0, 1]$ 上的边际概率 $p(x)$ 都为常数(三个动力学之间无暂态), 自相关函数是 δ 函数

$$x(t), x(t') = \frac{1}{4} + \frac{1}{12}\delta(t - t'),$$

因而线性分析不能将它们区分开. 因为这些动力学系统并不复杂以至于可用很多方法将它们区分开来. 此处引用的目的是以它们为例说明在实际中如何应用 R_{ij} 区分动力学过程.

在分析中, 将整个时间序列(60000 个点)分成 24 个窗口, 其中 1—8 由 (a) 式产生, 9—16 由 (b) 式产生, 17—24 由 (c) 式产生. 对时间序列进行两步延迟重构 ($x(t), x(t-1)$). 计算自关联和时, $\epsilon = 0.0625\sigma$, σ 为标准差(将时间序列归一化后, σ 为 1). 嵌入维数 $m = 2$. 为了比较方便同时也不失一般性, 我们选择 1, 9, 17 作为参考窗口, 然后计算其他所有窗口与它们之间的距离 $\{R_{1i}\}, \{R_{9i}\}, \{R_{17i}\}$. 将三个数据子集之间的 R 与 i ($i = 1, 2, \dots, 24$) 之间的关系分别画在图 2 和图 3 ($m = 3$) 中. 由图 2 和图 3 可看出, 以动力学相关因子指数为度规能将三个影射区分开来.

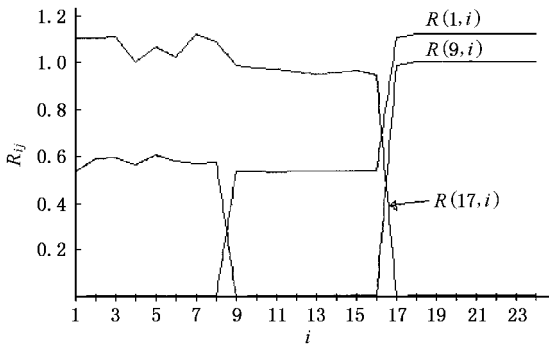


图2 嵌入维数 $m=2$,以 1,9 和 17 为参考窗口,动力学互相关因子指数 $R_{ij}(j=1,9,17; i=1,2,\dots,24)$

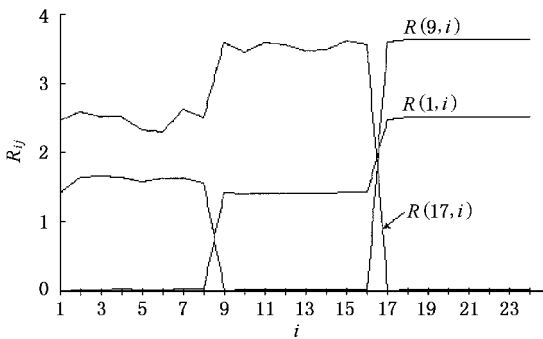


图3 嵌入维数 $m=3$,以 1,9 和 17 为参考窗口,动力学互相关因子指数 $R_{ij}(j=1,9,17; i=1,2,\dots,24)$

种颜色, $C_6(W_i, W_j)$ 大对应的颜色深,横轴和纵轴均代表窗口,在 $(i, j) (i, j = 1, 2, \dots, 140)$ 处标以 $C_6(W_i, W_j)$ 对应的颜色,这样就得到了 $C_6(W_i, W_j)$ 对窗口下标 $(i, j) (i, j = 1, 2, \dots, 140)$ 的函数关系图.图4是静息态(本文中思维状态是指人在正常的思考和计算问题时所处的状态,而静息状态是指一个人尽可能地放松而不对指定任务去思考和计算时所处的状态.很显然,此处所谓的静息状态和思维状态这两个概念并不是十分严密的,只是为了便于对两种不同状态的识别而已)时间序列的 $C_6(W_i, W_j)$ 对窗口下标 $(i, j) (i, j = 1, 2, \dots, 140)$ 的函数关系图.从图中可看出, $C_6(W_i, W_j)$ 对窗口下标 $(i, j) (i, j = 1, 2, \dots, 140)$ 的函数关系图很不均匀,且无明显的规律,这反应了脑电时间序列动力学的复杂性.

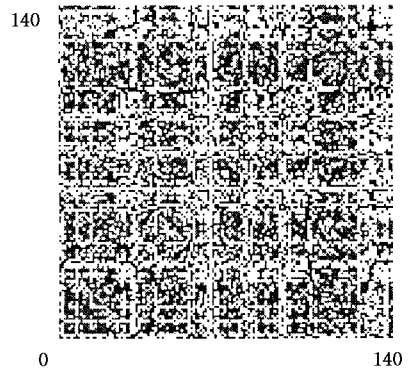


图4 静息态时间序列的 $C_6(W_i, W_j)$ 对窗口下标 (i, j) 的函数关系图(横轴和纵轴均代表窗口,深色处的 $C_6(W_i, W_j)$ 大,浅色处的 $C_6(W_i, W_j)$ 小.重构空间维数为 6, $\epsilon = 0.08$.共 140 个窗口,每个窗口中有 400 个点)

6. 真实数据检验

6.1. 脑电波数据的处理

对脑电时间序列的分析研究一直是很重要的课题^[11].运用本文的方法对脑电时间序列的分析有助于理解思维、睡眠等活动.

这里所用的数据取自北京师范大学电子系孟丽艳 1992 年的硕士论文《脑电波的动力学特征的研究》.受试者:健康女性,31 岁.数据长度 $N:56000$.采样时 $T_s:0.005s$.

经粗略计算发现其关联维数大约在 4—5(约 4.87)之间,故计算时嵌入维数取 $m=6, \epsilon=0.08$.计算中将其分为 140 个窗口,每个窗口包含 400 个数据点.首先计算各个窗口的自关联和 $C_6(W_i, W_j) (i=j)$ 与任意两个窗口之间的交互关联和 $C_6(W_i, W_j) (i \neq j)$ (下标 6 是嵌入维数);然后将 $C_6(W_i, W_j)$ 按大小分成 16 个等级,每一等级对应地赋予一

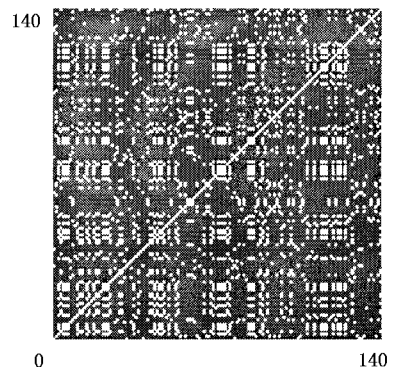


图5 静息态时间序列的符号动力学图(横轴和纵轴均代表窗口,坐标 $(i, j) (i, j = 1, 2, \dots, 140)$ 处为白色,表示该处对应的窗口属于同一动力学.重构空间维数是 6, $\epsilon = 0.08$.共 140 个窗口,每个窗口中有 400 个点)

如果按照本文的方法将窗口分类,就可得到不同动力学对应的窗口族.经计算,静息态时间序列可分成 13 个窗口族.以横轴和纵轴表示窗口,如果窗口 W_i 和 W_j 属于同一个族,那么,点 (i, j) ($i, j = 1, 2, \dots, 140$) 标以白色,否则,标以彩色.得到的图即为符号动力学图.图 5 为静息态时间序列的符号动力学图.图 6 和图 7 分别为思维态的 $C_6(W_i, W_j)$ 对 (i, j) ($i, j = 1, 2, \dots, 140$) 的函数图和符号动力学图.经计算,思维态时间序列可分成 8 个窗口族.

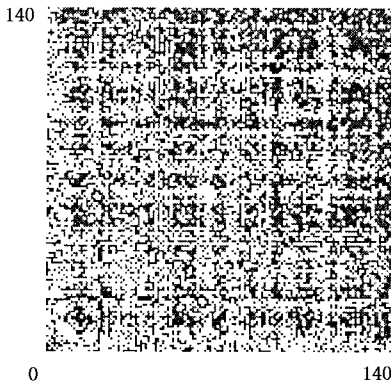


图 6 思维态时间序列的 $C_6(W_i, W_j)$ 对窗口下标 (i, j) 的函数关系图(横轴和纵轴均代表窗口,深色处的 $C_6(W_i, W_j)$ 大,浅色处的 $C_6(W_i, W_j)$ 小.重构空间维数是 6, $\epsilon = 0.08$.共 140 个窗口,每个窗口中有 400 个点)

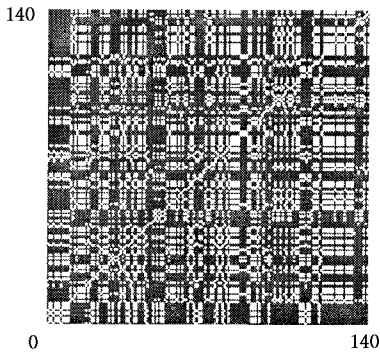


图 7 思维态时间序列的符号动力学图(横轴和纵轴均代表窗口,坐标 (i, j) ($i, j = 1, 2, \dots, 140$) 处为白色,表示该处对应的窗口属于同一动力学.重构空间维数是 6, $\epsilon = 0.08$.共 140 个窗口,每个窗口中有 400 个点)

如果改变窗口的划分情况结果又如何呢?下面我们上述静息态和思维态的 56000 个点都分别分成 90 个窗口,每个窗口含 600 个数据,其他条件都不变.图 8 和图 9 分别为静息态和思维态的 $C_6(W_i, W_j)$ 对 (i, j) ($i, j = 1, 2, \dots, 90$) 的函数图,图 10 和图 11 分别为静息态和思维态的符号动力学图.

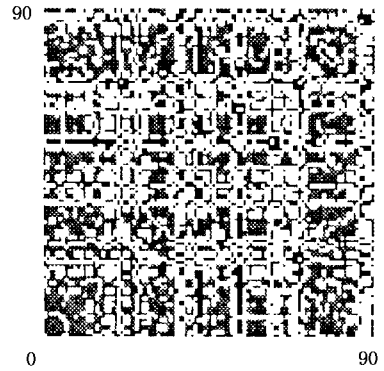


图 8 静息态

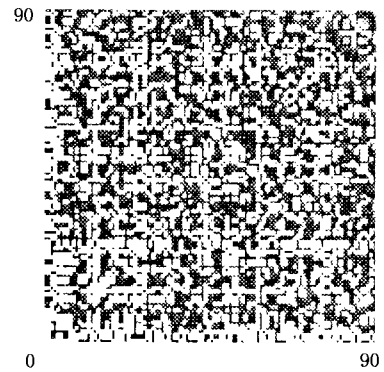


图 9 思维态

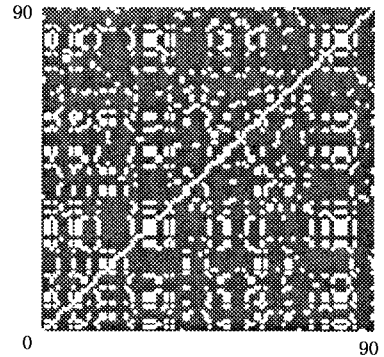


图 10 静息态

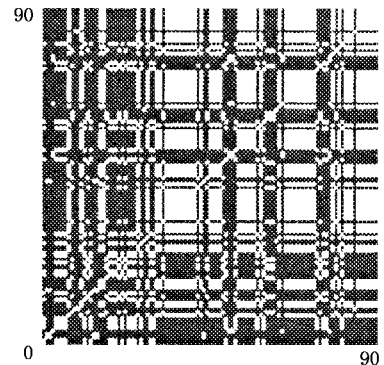


图 11 思维态

经计算,此时,静息态时间序列可分成 12 个窗口族.思维态时间序列可分成 8 个窗口族.

6.2. 对脑电波数据处理结果的讨论

通过上述分析可得出以下几点:

1. 这是最基本的一点.本文的方法不仅能将同一个人的静息态和思维态完全区分开,而且,分别找出了静息态和思维态脑电时间序列中具有相近动力学的区段(窗口)并将具有相近动力学的窗口组成了族.

2. 改变窗口的划分情况后,依然能通过 $C_6(W_i, W_j)$ 函数图反应出脑电时间序列是很复杂的.将上述静息态的 56000 个点分别分成 140 个窗口(每个窗口含 400 个数据)和 90 个窗口(每个窗口含 600 个数据)两种情况下,计算出来的静息态的族数很接近(分别为 13 个和 12 个),而将上述思维态的 56000 个点分别分成 140 个窗口(每个窗口含 400 个数据)和 90 个窗口(每个窗口含 600 个数据)两种情况下,计算出来的思维态的族数相等(都是 8 个).也就是说,窗口划分情况的改变基本不影响通过 $C_6(W_i, W_j)$ 函数图对脑电时间序列复杂性的揭示,族的个数也基本不受窗口划分的影响.这也从一个侧面说明应用本文的方法所得的结果是稳定的,本方法是有效的.

3. 计算所得的静息态和思维态中具有相近动力学属性的族数之多及 $C_6(W_i, W_j)$ 对 (i, j) 函数图的复杂性足以说明:脑电这个复杂的时间序列是由一些有相互耦合的动力学共同作用产生的.

另外,运用本文的方法对脑电时间序列的分析过程中,还发现一个很有意义、很有趣的现象,即同一个人在静息态下具有相同动力学属性的窗口族数比思维态下的多.这说明所谓的“静息态”并不意味着大脑真正地处于一种完全静息的状态之中而不工作,可能只是意味着大脑处于一种无集中的意识指向的状态之中,不具有通常意义上的思维活动,处于一种更加无序的自由状态之中.关于这一点,我们对照人类对睡眠过程的认识便不难理解^[12]:“多年来习惯认为,睡眠时大脑处于安静状态,睡眠是一个被动的过程.然而,最近三十年来的研究表明:睡眠时人的大脑仍在活动,一系列生理和生化的变化仍在进行,心理活动也未停止,梦即是睡眠时心理活动的表现形式.因此,现在认为睡眠同觉醒一样,也是一种主动过程.”尽管我们还不能得出静息态也是一种主动过程的结论,更不能指出静息态的深刻含义.但似乎可肯定的是:“静息态不静息”与“真空并不空”具有同样重要的意义.

[1] Sauer T 1994 *Phys. Rev. Lett.* **72** 3811

[2] Grassberger P and Procaccia I 1983 *Phys. D* **9** 189

[3] Kanta H 1994 *Phys. Rev. E* **49** 5901

[4] Liu H F *et al* 2002 *Acta Phys. Sin.* **51** 1186 [in Chinese] 刘海峰等 2002 物理学报 **51** 1186]

[5] Provenzale A, Smith L A, Vio R and Murante G 1992 *Phys. D* **58** 31

[6] Grassberger P 1990 *Phys. Lett. A* **148** 63

[7] Bingham S and Kot M 1989 *Phys. Lett. A* **140** 327

[8] Theiler J 1987 *Phys. Rev. A* **36** 445

[9] Savit R and Green M 1991 *Phys. D* **50** 95

[10] Manuca R and Savit R 1996 *Phys. D* **99** 134

[11] Xie Y *et al* 2002 *Acta Phys. Sin.* **51** 205 [in Chinese] 谢勇等 2002 物理学报 **51** 205]

[12] Xu S F 1990 *Neurobiology* (Medical University Press of Shanghai) (in Chinese] 许绍芬 1990 神经生物学(上海医科大学出版社)]

A method for distinguishing dynamical species in chaotic time series

Li Chun-Gui¹⁾ Pei Liu-Qing²⁾

¹⁾*(Basic Experiment , Wuhan Polytechnic University ,Wuhan 430023 ,China)*

²⁾*(Department of Electronics ,Beijing Normal University ,Beijing 100875 ,China)*

(Received 19 November 2002 ; revised manuscript received 30 December 2002)

Abstract

Two statistical quantities are proposed in this paper that is Q and R , Q is called the exponent of dynamical autocorrelation factor, and R is called the exponent of dynamical cross-correlation factor. Using Q and R , we can identify the dynamical property of time series or their implicative dynamical structural levels and complexity by estimating the "distance" between different time series or different segments of same time series. This method has been verified to be valid by some typical instances.

Keywords : chaotic time series, the exponent of dynamical autocorrelation factor, the exponent of dynamical cross-correlation factor

PACC : 0545