第 2 章 关联维

2.1 引言

时间序列经过相空间重构后，就可以进行混沌不变量的计算来判断是否具有混沌特性。常用的混沌不变量有关联维[1] 、Kolmogorov熵[2] 和Lyapunov指数[3] 等，本章重点介绍关联维和Kolmogorov熵的计算和应用。本章中各节主要内容如下：2.2 节介绍经典的G-P关联维算法的实现和缺陷，2.3 节介绍高斯核关联维算法的计算及其应用，2.4 节介绍非主观参数选择的G-P关联维算法，2.5 节介绍海杂波关联维的计算及其在目标检测中的应用，2.6 节为本章小结。

2.2 G-P 关联维算法的计算和缺陷

混沌是非周期与非随机的动力学过程，表面上看和研究不平滑、不可微分几何结构的分形学没有联系，但大量研究表明混沌时间序列构造的吸引子就是分形集，分形维数是刻划动力系统是否具有混沌特征的定量指标之一。 对于分形维数，比较严格的数学定义是豪斯道夫维数，但是由于数据量的限制难以在实际中应用。最早用于计算分形维数的简单方法是计盒法，但是计盒法针对高维系统计算速度太慢，并易受噪声的影响[4] 。关联维数 D_2 是比较有效的分形维计算方法，自从 1983 年Grassberger 和Procaccia 提出从时间序列计算关联维[1] 和Kolmogorov熵[2] 的方法后，就被大量研究人员广泛地使用。它的具体定义如下：

设点 $X_1, X_2, \cdots, X_N$ 是相空间内吸引子上的点，用 $B_r\left(X_i\right)$ 表示以参考点 X_i 为中心、半径是 r 的球形盒子，盒子的形状不会影响维数的计算，盒子 $B_r\left(X_i\right)$ 的概率测度为

$$P\left[B_r\left(X_i\right)\right]=\frac{1}{N-1}\sum_{j=1,i\neq j}^{N} H\left(r-\left\|X_i-X_j\right\|\right) \tag{2.1}$$

其中 $\|\bullet\|$ 是 Euclidean 范数，而 H 为 Heaviside 阶跃函数

$$H\left(x\right)=\begin{cases}1 & x\geq 0\\ 0 & x<0\end{cases} \tag{2.2}$$

则关联维数为

$$D_2=\lim_{r\to 0}\frac{\ln C\left(r\right)}{\ln r} \tag{2.3}$$

其中 $C\left(r\right)$ 为关联积分如下

$$C(r)=\frac{1}{N(N-1)}\sum_{i=1,i\neq j}^{N}\sum_{j=1}^{N}H\left(r-\left\|X_i-X_j\right\|\right)=\frac{\text{相点间距离小于r的相点对数目}}{\text{所有相点对数目}} \quad (2.4)$$

当 $r\to 0$ 时，关联积分 $C(r)$ 与 r 之间存在标度关系 $C(r)\propto r^{D_2}$，即有

$$\lim_{r\to 0}C(r)=\lim_{N\to\infty}\frac{1}{N(N-1)}\sum_{i=1,i\neq j}^{N}\sum_{j=1}^{N}H\left(r-\left\|X_i-X_j\right\|\right) \quad (2.5)$$

因此，从理论上说作出 $\ln C(r)$ 对 $\ln r$ 的变化图，则图中曲线的斜率 k 就等于关联维数 D_2。考虑到度量空间距离的对称性，(2.4)式可进一步写成

$$C(N,r)=\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}H\left(r-\left\|X_i-X_j\right\|\right) \quad (2.6)$$

由于动力系统实测数据间不可避免地存在着数据之间的自相关性，为了减少或基本根除由于数据间的自相关性所造成的影响，(2.6)式还可写成

$$C(r,N,w)=\sum_{n=w}^{N}\sum_{i=1}^{N-m}H\left(r-\left\|X_{i+n}-X_i\right\|\right) \quad (2.7)$$

这里 w 应满足：$w>\tau\left(\frac{2}{N}\right)^{\frac{2}{m}}$，$\tau$ 为时间序列的时间间隔，m 为嵌入维数。

二阶Kolmogorov 熵(orde-2 Kolmogorov entropy，简称K2 熵) 定义为[2]

$$K_2=\lim_{r\to 0}\lim_{N\to\infty}\frac{\log C_m(r)}{\log C_{m+1}(r)} \quad (2.8)$$

从上面分析可以看出，采用G-P算法计算动力系统实测数据吸引子的关联维数时，诸多因素可能影响估计精度。误差的来源主要有：实测数据序列的长度 N 有限；采样序列的自相关性；相空间重构参数的选择和实测数据中附加噪声的影响[5, 6] 。

采样序列长度对吸引子关联维数的估计量影响最大，当 $N\to\infty$ 时，关联维数估计的各种偏差都会有所改善，为了获取一个较为可靠的关联维估计值，采样序列的长度必须大于某一最小值 $N_{\min}$ [4, 7] 。但由于关联维算法的复杂度为 $O(N^2)$，因此数据量的大大增加了计算负担，有关学者也提出了改进速度的相应算法[8, 9] 。

若序列的相关时间相对于采样时间较长，采样序列的自相关性会使关联积分产生异常肩峰，导致关联维数估计质量下降，甚至得到虚假的估计值。这个问题的解决方法是增加采样时间[5] ，适当增大采样间隔[7] ，同时采用引入限制短暂分离参数[10] ，使该参数大于序列平均周期时间，去除同一轨道前后点之间的关联。此外无论何时，选择合适的嵌入维和延迟时间进行相空间重构都是必要的[10] 。对任意的动力学系统，简单地选

取延迟时间和重构维数不能准确评价系统的分形特征，并且延迟时间与重构维数对于不同类型的动力系统的作用效果是不同的[4] 。

尽管噪声对某一初值出发的特定轨线是敏感的，但混沌吸引子的整体结构是稳定的，因此动力系统实测数据中的较小噪声对关联维数计算的影响不大。但度量分形特征的尺度 r 应大于噪声幅度，当尺度 r 接近或小于噪声幅度时，关联积分的计算会受到强烈影响。此时需要考虑使用降噪方法，但不宜直接对信号低通滤波来除去高频噪声，因为这有可能人为地提高吸引子的关联维[5] 。

除了上述外在因素会影响关联维数的估计结果以外，G-P算法本身也存在缺陷。在根据关联维数的定义推导出算法的过程中，是用自然测度轨道平均代替相空间平均的假设下得到的，如果自然概率测度在相空间的分布均匀，则假设是成立的。然而当奇怪吸引子在相空间内分布不均匀时，吸引子不同区域在同一尺度下的标度行为将存在显著差异，即不同参考点 X_i 在不同的尺度 r 下出现标度区域，因此轨道平均即在同一 r 下对所有参考点平均将引起关联维数产生误差。另一方面，关联维数对时间序列平稳性的要求也会由于增加数据量而不能得以满足[5] 。而且G-P关联维算法在无标度区的确定和线性回归计算时，受主观因素的影响，不同的人会有不同的分析结果。为了改进G-P算法的诸多缺点，我们在下面章节中介绍高斯核关联维和非主观关联维的计算和应用。

2.3 高斯核关联维的计算和应用

为了消除噪声对关联维计算的影响，同时计算噪声水平，Diks和Yu等使用了高斯核函数来计算关联维数[11-13] ，具体过程如下：

(2.4)式的G-P关联积分可表示成如下的连续形式

$$C_m(r)=\int d\bar{x}\rho_m(\bar{x})\int d\bar{y}\rho_m(\bar{y})\theta(r-|\bar{x}-\bar{y}|) \tag{2.9}$$

因此关联积分的普遍形式为

$$T_m(h)=\int d\bar{x}\rho_m(\bar{x})\int d\bar{y}\rho_m(\bar{y})w(|\bar{x}-\bar{y}|/h) \tag{2.10}$$

其中 $w(\bullet)$ 为核函数，参数 h 为基函数的宽度，在(2.4)式的核函数为 $\theta(1-x)$， $h=1$。

若使用高斯核函数

$$w(x)=e^{-x^2/4} \tag{2.11}$$

可得到高斯核的关联积分函数为[11] ：

$$T_m(h)=\int d\bar{x}\rho_m(\bar{x})\int d\bar{y}\rho_m(\bar{y})e^{-|\bar{x}-\bar{y}|^2/(4h^2)} \tag{2.12}$$

对于无噪声的时间序列，高斯核关联维数可简化为：

$$T_m(h)=\int d\mathbf{x}\rho_m(\mathbf{x})\int d\mathbf{y}\rho_m(\mathbf{y})e^{-\|x-y\|^2/4h^2}\sim e^{-mK\delta t}\left(\frac{h}{\sqrt{m}}\right)^D \quad for \quad h\to 0, m\to\infty \tag{2.13}$$

其中 D 和 K 分别是关联维和 Kolmogorov 熵， h 是基函数带宽， $\rho_m(\mathbf{x})$ 为分布函数。

当有方差为 σ^2 的正态高斯噪声，分布函数 $\tilde{\rho}_m(\mathbf{y})$ 可表示为

$$\tilde{\rho}_m(\mathbf{y})=\int d\mathbf{x}\rho_m(\mathbf{x})\rho_m^g\|x-y\|^2=\frac{1}{\left(\sigma\sqrt{2\pi}\right)^m}\int d\mathbf{x}\rho_m(\mathbf{x})e^{-\|x-y\|^2/2\sigma^2} \tag{2.14}$$

其中$\|\bullet\|$代表欧几里得(L2)范数，将上式化简可得[12]

$$\begin{aligned}T_m(h)&=\int d\mathbf{x}\tilde{\rho}_m(\mathbf{x})\int d\mathbf{y}\tilde{\rho}_m(\mathbf{y})e^{-\|x-y\|^2/4h^2}=\left(\frac{h^2}{h^2+\sigma^2}\right)^{m/2}\int d\mathbf{x}\rho_m(\mathbf{x})\int d\mathbf{y}\rho_m(\mathbf{y})e^{-\|x-y\|^2/4\left(h^2+\sigma^2\right)}\\&=\phi\left(\frac{h^2}{h^2+\sigma^2}\right)^{m/2}e^{-mK\delta t}\left(\frac{h^2}{h^2+\sigma^2}\right)^{d/2}\quad for\sqrt{h^2+\sigma^2}\to 0, m\to\infty\end{aligned} \tag{2.15}$$

式中 D 和 K 分别为关联维和 Kolmogorov 熵两个混沌不变量， σ 是噪声水平，定义为：

$$\sigma=\frac{\sigma_n}{\sigma_s}=\frac{\sigma_n}{\sqrt{\sigma_c+\sigma_n}} \tag{2.16}$$

其中 σ_s 、 σ_c 和 σ_n 分别是混合信号、纯净信号和高斯噪声的标准差[12] 。

通过(2.15)式计算后作数据拟合，可以得到 D 和 σ 高精度的结果，比较不确定的是 ϕ 和 K 。因为对于确定的 m ， ϕ 和 $e^{-mk\delta t}$ 不是独立的，他们之间的关系为 $\beta=\phi e^{-mk\delta t}$ 。Yu 和Small通过对(2.15)式作进一步处理，不但降低了高斯核算法的计算量，而且能够在非线性拟合过程中，只要得到一个稳定的 β 值，就会给出满足条件的任意 ϕ 和 K 值[12] 。

在计算高斯核关联维数时，一个非常重要的参数为基函数带宽 h 的选择。在含有噪声的情况下，若 h 取得过小，受噪声的影响，将无法提取出正确的关联维数；若 h 取得过大，则有可能无法提取出合适的无标度区间。在实际计算中，一般取 $h\geq 3\sigma$ [12] 。

2. 4 非主观关联维的计算

在关联维的计算中，选取不同的参数会得到不同的计算结果，尤其是线性尺度区间的选择具有很强的主观性。若尺度 r 选择过小，实际时间序列中的噪声会影响关联积分的计算，造成结果不准确；而如果尺度 r 过大，关联积分的计算又会受到边界效应的影响。为了客观准确的计算关联维，Harikrishnan等提出一种非主观的关联维计算方法[14] 。该方法通过使用方体来代替球体(即使用最大范数来代替欧几里得范数，这儿我没弄明

白为什么可以这么说，知道的给我讲一下)，并通过识别出 $R_{\min}$ 和 $R_{\max}$ 区间来保证关联维的计算不受噪声和边界效应的影响，具体流程如下[14] ：

1) 将原始时间序列变为均值为 0，方差为 1 归一化序列。
2) 使用自相关函数法计算时间序列的延迟时间 τ 。
3) 嵌入维 M 从 1 开始逐渐增大，在每个嵌入维中随机选择 $N_c = 0.1N_v$ 个点为中心计算关联积分 $C_M(R)$，其中 N_v 为序列长度。
4) 识别出 $R_{\min}$ 和 $R_{\max}$ 尺度区间，并在该区间上计算关联维数 $D_2(M)$
5) 随着嵌入维 M 的增大，$R_{\min}$ 和 $R_{\max}$ 将逐渐靠近，检测 $D_2(M)$ 是否达到饱和。
6) 若饱和，计算出饱和关联维 $D_2^{sat}(M)$；若始终没有饱和，则说明待求数据为噪声。

非主观关联维的计算方法所有的计算过程均通过算法完成，不需要任何主观人为的因素，它适用于数据性质的判断、大数据量含有噪声数据关联维的计算以及原始数据和替代数据之间的性质比较。

2.5 海杂波的关联维及其应用

内容 (略)

2.6 本章小结

关联维和 Kolmogorov 熵是重要的混沌不变量，他们可用来直接判断时间序列的是否具有混沌特性，也是使用替代数据法检测数据是否具有非线性的重要统计量。噪声和主观选择参数是限制关联维应用的两个重要因素，高斯核关联维和非主观关联维的计算部分解决了这一问题，但关联维的计算和混沌吸引子的度量都有待于进一步发展和完善。

2.7 后记

本版本为关联维的最终版本，修订了临时版本的参考文献和语言错误，增加了非主观关联维的计算和小结。

你的关注是我前进的最大动力☺。

任何意见、建议、批评和讨论我都热烈欢迎。

我的信箱：xuxkboy@newmail.dlmu.edu.cn

欢迎大家到研学论坛(http://bbs.matwav.com/)混沌分形版进行讨论

许小可.海杂波的非线性分析与建模:(博士学位论文).大连:大连海事大学,2007

参考文献：

[1] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. Physica D: Nonlinear Phenomena. 1983, 9(1-2): 189-208.

[2] Grassberger P, Procaccia I. Estimation of the Kolmogorov entropy from a chaotic signal. Physical Review A. 1983, 28(4): 2591-2593.

[3] Wolf A, Swift J B, Swinney H L, et al. Determining Lyapunov exponents from a time series. Physica D: Nonlinear Phenomena. 1985, 16(3): 285-317.

[4] 王安良，杨春信. 评价奇怪吸引子分形特征的Grassberger-Procaccia算法. 物理学报. 2002, (12): 2719-2728.

[5] 于青. 关联维数计算的分析研究. 天津理工学院学报. 2004, (04): 60-62.

[6] 党建武，黄建国. 基于G.P算法的关联维计算中参数取值的研究. 计算机应用研究. 2004, (01): 48-51.

[7] 李夕海，刘代志，张斌，等. 基于重采样的混沌时间序列相空间重构研究. 信号处理. 2006, (02): 248-251.

[8] 周越，杨杰. 求解关联维数的快速算法研究. 电子学报. 2002, (10): 1526-1529.

[9] 温晓通，孟丽艳，朱劲松，等. 一种非线性时间序列的关联维快速算法. 北京师范大学学报(自然科学版). 2005, (04): 358-361.

[10] Rosenstein M T, Collins J J, De L C, et al. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D: Nonlinear Phenomena. 1993, 65(1-2): 117-134.

[11] Diks C. Estimating invariants of noisy attractors. Physical Review E. 1996, 53(5): 4263-4266.

[12] Yu D, Small M, Harrison R G, et al. Efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy time series data. Physical Review E. 2000, 61(4): 3750-3756.

[13] Nolte G, Ziehe A, M K R, et al. Noise robust estimates of correlation dimension and K2 entropy. Physical Review E. 2001, 64(1): 016112-1-10.

[14] Harikrishnan K P, Misra R, Ambika G, et al. A non-subjective approach to the GP algorithm for analysing noisy time series. Physica D: Nonlinear Phenomena. 2006, 215(2): 137-145.