

一种基于模糊C均值聚类小数据量计算最大Lyapunov指数的新方法

周双 冯勇 吴文渊 汪维华

A novel method based on the fuzzy C-means clustering to calculate the maximal Lyapunov exponent from small data

Zhou Shuang Feng Yong Wu Wen-Yuan Wang Wei-Hua

引用信息 Citation: *Acta Physica Sinica*, 65, 020502 (2016) DOI: 10.7498/aps.65.020502

在线阅读 View online: <http://dx.doi.org/10.7498/aps.65.020502>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2016/V65/I2>

您可能感兴趣的其他文章

Articles you may be interested in

基于忆阻器的数模混合随机数发生器

[A digital-analog hybrid random number generator based on memristor](#)

物理学报.2015, 64(24): 240503 <http://dx.doi.org/10.7498/aps.64.240503>

空间关联白噪声影响下小世界神经网络系统的同步动力学

[Synchronous dynamics of small-world neuronal network system with spatially correlated white noise](#)

物理学报.2015, 64(22): 220503 <http://dx.doi.org/10.7498/aps.64.220503>

六边形格子态斑图的数值模拟

[Numerical simulations of hexagonal grid state patterns](#)

物理学报.2015, 64(21): 210505 <http://dx.doi.org/10.7498/aps.64.210505>

电压控制正极性输出罗变换器的改进平均模型建模及稳定性分析

[Improved averaged model and stability analysis of voltage-mode controlled positive output super-lift Luo converter](#)

物理学报.2015, 64(21): 210506 <http://dx.doi.org/10.7498/aps.64.210506>

单稳系统的脉冲响应研究

[Pulse response of a monostable system](#)

物理学报.2015, 64(21): 210503 <http://dx.doi.org/10.7498/aps.64.210503>

一种基于模糊C均值聚类小数据量计算最大Lyapunov指数的新方法*

周双^{1)2)†} 冯勇¹⁾ 吴文渊¹⁾ 汪维华¹⁾²⁾

1)(中国科学院重庆绿色智能技术研究院, 自动推理与认知重庆市重点实验室, 重庆 400714)

2)(中国科学院大学, 北京 100049)

(2015年6月3日收到; 2015年10月16日收到修改稿)

在小数据量计算最大Lyapunov指数的过程中, 为了减少人为因素识别线性区域带来的误差, 提出一种基于模糊C均值聚类的新方法. 该方法根据平均发散程度指数曲线的变化特征, 利用分类算法进行识别. 首先, 利用小数据量算法对混沌时间序列进行计算得到平均发散程度指数集合; 其次, 利用模糊C均值聚类算法对平均发散程度指数集合进行分类, 得到饱和数据; 然后, 对饱和的二阶差分数据进行分类, 得到零附近波动数据并剔除粗大误差, 再对保留的有效数据利用统计方法识别出线性区域; 最后, 对线性区域进行最小二乘法拟合得到最大Lyapunov指数. 为了验证该算法的有效性, 对著名Logistic和Hénon混沌系统进行了仿真, 所得结果接近理论值. 实验表明, 所提出的新方法的主观识别方法比较, 计算结果更加准确.

关键词: 最大Lyapunov指数, 线性区域, 模糊C均值聚类

PACS: 05.45.-a, 05.45.Pq, 05.45.Tp

DOI: 10.7498/aps.65.020502

1 引言

混沌是确定系统的内在随机性^[1], 而混沌的离散情况常常表现为混沌时间序列. 这种时间序列是由混沌系统产生的, 其中蕴含着丰富的动力学信息, 在自然界大量存在, 如长江径流混沌时间序列^[2]、太阳黑子混沌时间序列^[3]、电力负荷混沌时间序列^[4]等. 由于它的普遍存在性决定了对其研究的必要性和重要意义. 而用传统的方法去研究混沌时间序列往往导致不能获得重要的信息, 自从1980年Packard等提出重构相空间理论以来, 大大加强了研究混沌时间序列的基础^[5]. 混沌时间序列的一个重要的特征就是初值敏感性, 常用Lyapunov指数进行定量刻画. 由于混沌时间序列的预测以及特征识别分类等问题都是基于Lyapunov指数进行的, 因此, 研究Lyapunov指数的计算问题是有意义的. 目前, 这个重要的指数已经广泛应用

到各个领域, 如在混沌保密通信中应用Lyapunov指数分析混沌电路的动力学行为^[6,7]以及设计多个Lyapunov指数的超混沌信号发生器^[8,9], 在天体物理中利用最大Lyapunov指数定量分析长期太阳活动^[3,10], 在信号识别领域根据Lyapunov指数进行雷达目标检测^[11]等.

虽然人们对Lyapunov指数有了许多认识和应用, 但是针对计算最大Lyapunov指数的问题并没有完全解决, 需要进一步的研究. 一些学者在这方面做了很多贡献, 为此, 出现了很多计算最大Lyapunov指数方法. 对于给定动力学方程, 可以方便地计算出最大Lyapunov指数^[12-14]. 由于实际问题中往往不知道动力系统的方程, 因此使用传统方法比较困难. 为此产生了从时间序列计算最大Lyapunov指数的一些方法, 如基于相空间轨道、相平面、相体积等长期演化的Wolf方法来估计最大Lyapunov指数^[15], 但是需要很长的数据, 计算繁

* 国家自然科学基金(批准号: 11301524)和重庆市基础与前沿研究计划院士专项(批准号: cstc2015jcyjys40001)资助的课题.

† 通信作者. E-mail: zhoushuang@cigit.ac.cn

琐, 受噪声影响. 随后 Briggs 在其工作中提出利用 Jacobi 方法计算 Lyapunov 指数^[16], 此方法适用于噪声大的时间序列, 但是计算量大, 过程繁琐. 1993 年 Rosenstein 等^[17]提出基于轨道跟踪的小数据量法. 同年 Barana 和 Tsuda^[18]提出了 p-范数法, 在 wolf 方法和 Jacobian 方法之间架起了桥梁. 但由于 p-范数的选取和计算比较复杂, 不容易操作. 而小数据量法相对于其他方法, 对相空间的嵌入维数、延迟时间、数据长度、噪声等具有更好的鲁棒性. 此外, 所需要的数据量小, 相对操作简单, 已经广泛应用到很多实际问题, 得到一些有意义的结果, 如蒋海峰等^[19]利用小数据量方法快速识别短时交通流混沌特性; 许小可使用小数据量方法研究海杂波中的弱小目标检测问题^[20]. 为了提高小数据量方法的计算准确性, 主要从以下几个方面进行改进: 1) 减少噪声对结果的影响, 如 Kantz 提出的固定邻域法^[21], 梁勇等的可变邻域法^[22], 姚天亮的基于邻域内不同点对的平均法^[23], 杨绍清等的最大方向上平均估计最大 Lyapunov 指数的方法^[24]等; 2) 减少重构相空间对结果的影响, 如吕金虎等^[4]提出用 C-C 方法求解小数据量中的延迟时间和嵌入维数, 有效减少了盲目选取而导致的繁琐计算, 进而提高了小数据量的计算精度; 王海燕等^[25]提出多变量时间序列重构可以获得比单变量时间序列更好的结果等; 3) 平均周期对于结果是否有影响一直存在争议, 有学者对此进行了研究, 验证了时间序列的平均周期无法确定的情况下, 不考虑限制短暂分离, 仍可以得到比较好的估计^[26]; 4) 线性区域的选择对结果的影响, 由于线性区域对应的斜率就是最大 Lyapunov 指数, 如果不能准确识别线性区域, 会导致计算误差增大. 因此, 线性区域的识别对最大 Lyapunov 指数最终计算结果的准确性十分重要. 为此, Rosenstein 等提出了主观识别法^[17], 效率高, 但每个人的视觉效果可能不一样, 还是会出现很大误差; 而杨永峰等^[27]也对此进行了研究, 提出了最大无波动区间客观寻找法, 避免了主观识别带来的误差, 但对于同一曲线中出现多个线性区域识别可能会陷入局部最优解. 此外, 该算法的时间复杂度为 $O(n^2)$, 导致计算效率不高. 因此, 本文针对 4) 中的问题, 提出一种基于模糊 C 均值聚类客观寻找线性区域的方法. 该方法适用于混沌时间序列, 对小数据量方法进行了改进, 寻找线性区域更加客观准确、运行效率更高、更适用于实际应用, 如水文预报、机械故障诊断以及水面目标识别等领域.

本文对小数据量中的线性区域选择进行了研究, 为了减少人为选择线性区域带来的误差, 提出的新方法可以客观确定线性区域, 提高计算最大 Lyapunov 指数的精确度. 首先, 使用小数据量方法对给定的混沌时间序列计算出平均发散程度指数集合; 其次, 利用分类的算法对得出的集合进行分类, 保留不饱和数据; 再次, 使用分类的算法对保留的数据进行分类得到零附近波动数据并剔除粗大误差; 然后, 用统计方法从零附近波动数据识别出线性区域; 最后, 对该区域进行最小二乘法拟合得到斜率, 即为最大 Lyapunov 指数. 为了验证算法的有效性, 对经典 Logistic 和 Hénon 混沌系统进行了仿真.

2 方 法

由于实测数据带有误差的原因, 小数据量算法中的线性区域对应的并不是一条严格的直线, 而是一些离散点构成波动平缓的区域. 定量地说就是可以通过回归直线方程拟合的区域. 而主观识别容易把多余的点添加到线性区域 (或者漏掉线性区域的一些点), 这样使得计算得到的回归直线方程容易出现较大误差, 进而不能准确计算出最大 Lyapunov 指数. 为了准确提取线性区域, 避免人为因素, 本文提出了一种易于计算机实现、操作简单的方法. 一般情况下, 我们知道线性区域所在回归直线的斜率是常数, 那么回归直线的二阶导数是零, 因此曲线中线性区域内点的二阶导数通常都会为零附近波动, 其他的点会波动较大. 为了准确识别零附近波动的点, 我们使用分类算法在曲线二阶导数中提取零附近波动的点, 再结合统计学提取最佳的线性区域. 这样做既不容易陷入局部最优解, 又可以快捷地识别出线性区域. 为了引进分类思想, 我们将选择模糊 C 均值聚类算法. 又因为线性区域识别是在小数据量算法中体现的, 因此下面先分别介绍小数据量法和模糊 C 均值聚类算法, 最后介绍模糊 C 均值聚类算法在小数据量算法中识别线性区域的应用.

2.1 小数据量法

小数据量法是 1993 年 Rosenstein 等^[17]提出从时间序列计算最大 Lyapunov 指数的简单方法, 其流程图如图 1 所示, 具体步骤如下:

1) 给定时间序列 $\{x(i), i = 1, 2, \dots, N\}$ 重构相空间 $\{X(t), t = 1, 2, \dots, M\}$ 并进行傅里叶变换求出平均周期 p ;

2) 计算相空间中每个点 $X(t)$ 的最近邻点 $X(t')$, 即 $d_t(0) = \min_{t'} \|X(t) - X(t')\|$ 其中 $|t - t'| > p$, $t' = 1, 2, 3, \dots, M$ 且 $t \neq t'$;

3) 对相空间中每个点 $X(t)$, 计算出该邻域点对的 i 个离散时间步后的距离 $d_t(i) = \|X(t+i) - X(t'+i)\|$, $i = 1, 2, 3, \dots, \min(M-t, M-t')$;

4) 对每个 i 计算出所有 t 的 $\ln d_t(i)$ 平均发散程度指数 $y(i)$, 达到饱和前, 选择 $y(i)-i$ 曲线中的线性区域 (如图 2 所示), 并用最小二乘法做出回归直线, 该直线的斜率为最大 Lyapunov 指数.

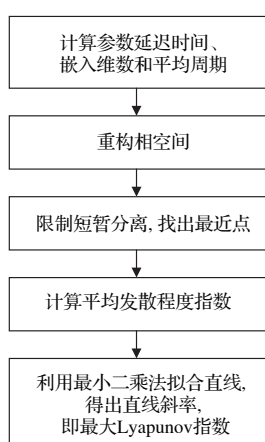


图 1 小数据量算法流程图

Fig. 1. Schematic diagram of the small data algorithm.

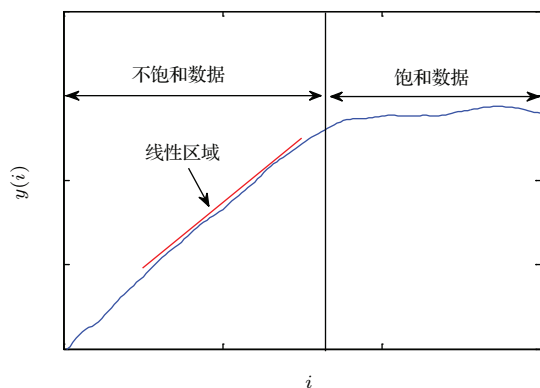


图 2 线性区域示意图

Fig. 2. Schematic diagram of linear zone.

2.2 模糊 C 均值聚类算法

模糊 C 均值聚类 (FCM) 是基于距离的聚类算法, 通过模糊理论对数据进行分析和建模, 建立了数据类属的不确定性描述, 能比较客观地反映现实世界 [28], 具有很快的搜索速度. 该算法的流程图如图 3 所示, 具体步骤如下.

1) 初始化隶属矩阵 $U = (u_j(x_i))_{c \times n}$, 满足约束条件 $\sum_{j=1}^c u_j(x_i) = 1$, 其中 $u_j(x_i) \in [0, 1]$ 为随机数, n 为数据个数, c 为预定的类别数目, $i = 1, 2, \dots, n$.

2) 计算类别中心 c_j ,

$$c_j = \frac{\sum_{i=1}^n u_j(x_i)^b x(i)}{\sum_{i=1}^n u_j(x_i)^b}, \quad (1)$$

其中, b 是一个加权数, 一般取 2 左右, $j = 1, 2, \dots, c$.

3) 计算目标函数

$$J = \sum_{j=1}^c \sum_{i=1}^n [u_j(x_i)]^b \|x_i - c_j\|^2, \quad (2)$$

其中 $u_j(x_i)$ 是第 i 个数据点 x_i 对于 j 类的隶属度函数.

4) 如果目标函数小于设定的阈值, 或者它相对上次目标函数值 J 的改变量 ΔJ 小于某个人为设定的阈值, 算法停止; 否则, 用下面公式重新计算新的 U 矩阵, 返回步骤 2), 直到算法停止.

$$u_j(x_i) = \frac{\left(\frac{1}{\|x_i - c_j\|^2} \right)^{\frac{1}{m-1}}}{\sum_{k=1}^c \left(\frac{1}{\|x_i - c_k\|^2} \right)^{\frac{1}{m-1}}}, \quad (3)$$

其中, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, c$.

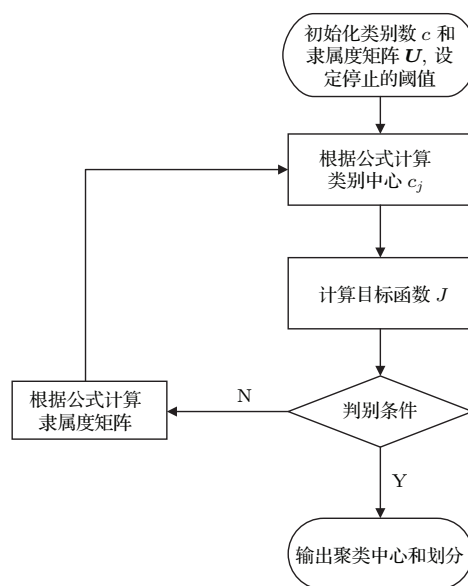


图 3 模糊 C 均值聚类算法流程图

Fig. 3. Flow chart of fuzzy C-means clustering algorithm.

2.3 一种模糊C均值聚类小数据量计算最大Lyapunov指数的新方法

一种模糊C均值聚类小数据量计算Lyapunov指数的新方法见流程图4所示, 包括以下步骤.

1) 利用快速傅里叶变换、互信息以及Cao方法, 分别对去噪后的混沌时间序列 $\{x(i), i = 1, 2, \dots, N\}$ 计算出平均周期 p , 延迟时间 τ 和嵌入维数 m .

2) 采用时间差法重构相空间, 按间隔 τ 从时间序列中取数作为矢量的分量, 因而构造出一批矢量, 即

$$X(t) = [x(t), x(t + \tau), x(t + 2\tau), \dots, x(t + (m - 1)\tau)], \quad (4)$$

其中 m 为嵌入维数, $M = N - (m - 1)\tau$, $t = 1, 2, \dots, M$.

3) 找相空间中每个点 $X(t)$ 的最近邻点 $X(t')$, 即满足最小距离

$$d_t(0) = \min_{t'} \|X(t) - X(t')\|, \quad (5)$$

其中 $|t - t'| > p$, $t' = 1, 2, \dots, M$ 且 $t \neq t'$.

4) 对相空间中每个点 $X(t)$, 计算出该邻点对的 j 个离散时间步后的距离

$$d_t(j) = \|X(t + j) - X(t' + j)\|, j = 1, 2, \dots, \min(M - t, M - t'). \quad (6)$$

5) 对每个 j 求出所有 t 的 $\ln d_t(j)$ 平均 $y(j)$, 得到平均发散程度指数集合 $\{y(j)\}_{j \in J}$.

下面重点讲解应用聚类算法识别线性区域的步骤. 由于混沌时间序列存在最大Lyapunov指数, 因此它的发散程度指数曲线中存在线性区域. 此外, 通过下面步骤可知识别线性区域算法的时间复杂度为 $O(nc^2dlk)$, 其中, n 为数据长度, c 为分类数目 (该算法中 c 取 2 或 3), d 为数据维度 (该算法中 $d = 1$), l 为聚类算法中迭代次数, k 为循环次数. 因此该方法的思想和效率的提高是本文的创新之处.

6) 因为在算法中线性区域是在曲线 $y(j)$ 饱和和之前的不饱和数据中选取, 其中饱和数据中的点会在一条水平直线附近做微小波动 (如图5所示), 其他点会做较大波动. 为了剔除这部分数据, 保留含有线性区域的不饱和数据, 使用模糊C均值聚类算法将平均发散程度指数集合分为饱和与不饱和数据, 保留不饱和数据.

7) 由于不饱和数据中的线性区域所在回归直线的斜率是常数, 因此它的二阶导数是零, 进而线

性区域内点的二阶差分值会在零附近波动 (注: 由于离散数据不能求导数, 因此使用二阶差分近似二阶导数), 其他点会波动较大, 因此利用模糊C均值聚类将不饱和类数据进行分类, 得到正波动、负波动以及零波动三类数据 (注: 二阶差分曲线会出现多种波动形式, 这里假设其中一种波动形式), 如图6所示.

8) 选取零波动数据, 并根据实际工程需要, 应用统计学中的剔除粗大误差的原则, 设定误差范围, 剔除数据中超过误差范围数据, 再从有效数据中选取连续区域数据 $\{y''(j)\}_{j \in F}$, $F \subseteq J$ 对应的点集 $\{(j, y(j))\}_{j \in F}$, $F \subseteq J$. 由于连续区域可能是多个, 因此这样区域对应的点集也是多个, 所以要对其分别进行线性回归分析.

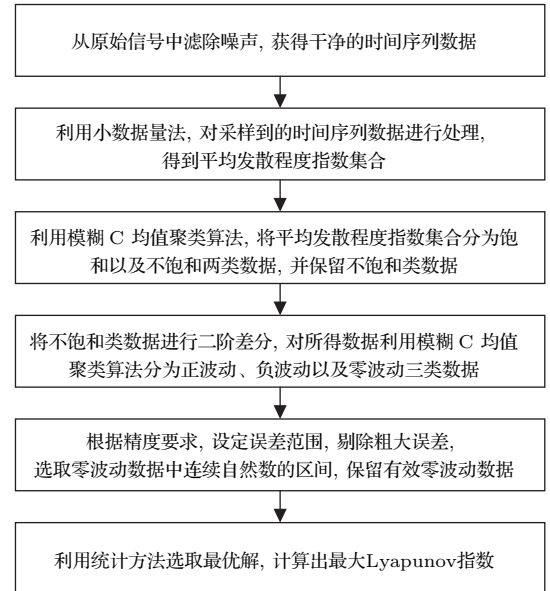


图4 模糊C均值聚类小数据量计算最大Lyapunov指数流程图

Fig. 4. Flow chart of fuzzy C-means small data for calculating the maximal Lyapunov exponent.

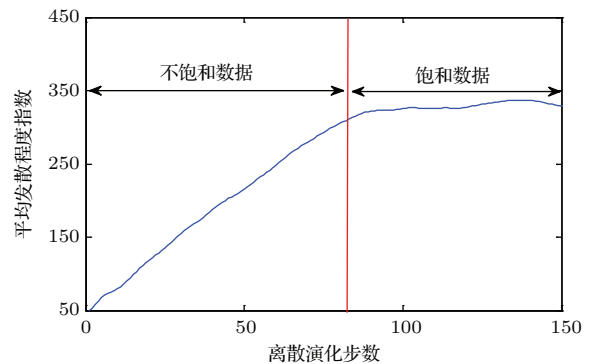


图5 平均发散程度指数示意图

Fig. 5. Schematic diagram of average divergence index.

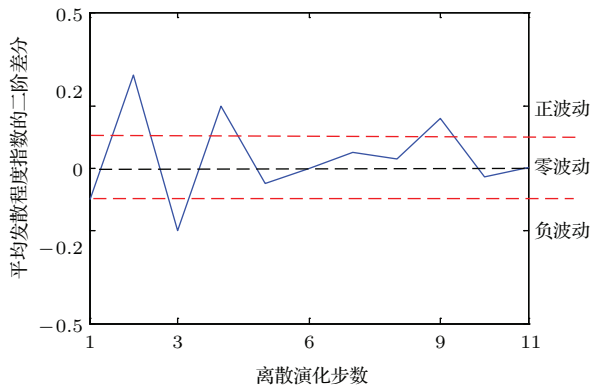


图6 平均发散程度指数二阶差分示意图

Fig. 6. Schematic diagram of the second order of average divergence index.

9) 如果回归分析中通过统计检验的只有一个区域, 那么把该区域作为线性区域, 直接进行下一步; 如果通过统计检验的区域是多个, 那么选取拟

合效果最好区域的作为线性区域(注: 相关系数接近1, 并且相关系数通过统计检验, 这样的回归直线方程拟合效果越好); 否则, 转到步骤7) 对零波动数据再分成三类, 提取更加精确的零波动数据.

10) 最后, 利用最小二乘法对线性区域进行拟合, 求得的斜率即为最大 Lyapunov 指数.

2.4 实验仿真

1976年 May 提出了一个生态学中非常简单的虫口(或人口)模型 Logistic 混沌系统^[29]. 同年 Hénon 考虑标准二次映射迭代系统时获得 Hénon 混沌吸引子^[30,31]. 为了验证算法的有效性, 对这两个经典混沌系统进行仿真, 并与文献^[17]中的结果进行比较(如表1所示). 从表1中可以看出新方法是有效可行的, 比主观方法更加准确.

表1 不同算法比较

Table 1. Comparison of different algorithms.

模型	参数	主观识别法 ^[17]	新算法	理论值 ^[17]
Logistic $x_{n+1} = 4x_n(1 - x_n)$	$x_0 = 0.1$, 500 个数据点, 嵌入维数 $m = 2$, 延迟时间 $\tau = 1$,	0.681	0.690 ± 0.003	0.693
Hénon				
$\begin{cases} x_{n+1} = y_n + 1 - 1.4x_n^2 \\ y_{n+1} = 0.3x_n \end{cases}$	$[x_0, y_0] = [0, 0]$, 500 个数据点, 嵌入维数 $m = 2$, 延迟时间 $\tau = 1$	0.409	0.418 ± 0.010	0.418

3 结 论

本文利用小数据量中发散程度指数曲线变化的几何特征, 提出基于模糊 C 均值聚类识别线性区域的方法. 由于该方法的时间复杂度为 $O(nc^2dlk)$, 因此可以快速客观地识别线性区域. 相对于主观识别法, 该方法提高了小数据量计算最大 Lyapunov 指数的精确度, 而且该方法具有过程简单, 易于计算机实现等优点. 但它也有自身的缺点: 1) 从实验仿真中验证新方法可行性, 并没有严格的数学证明; 2) 新方法中使用了差分运算, 因此会带来一些信息的缺失; 3) 该方法计算的精度有待提高. 因此该方法是一个探索, 对线性区域客观识别起到一定借鉴作用, 只能应用到对于精度要求不高的工程领域. 基于新方法不足的地方, 我们将做进一步的研究探讨, 完善计算最大 Lyapunov 指数的方法, 使其在信号处理领域可以解决检测信号的实时性问题, 在气象学领域可以准确找到气候突变的位置,

在空间天气领域可以提供准确的卫星发射安全期等方面得到应用. 总之, 研究混沌时间序列的最大 Lyapunov 指数计算问题具有广泛的应用前景和实际意义.

参考文献

- [1] Yu S M 2011 *Chaotic Systems and Chaotic Circuits: Principle, Design and Its Application in Communications* (Xian: Xidian University Press) p4 (in Chinese) [禹思敏 2011 混沌系统与混沌电路: 原理、设计及其在通信中的应用 (西安: 西安电子科技大学出版社) 第4页]
- [2] Wang W J, Ye M, Chen X W 1994 *Advances in Water Science* **2** 87 (in Chinese) [王文均, 叶敏, 陈显维 1994 水科学进展 **2** 87]
- [3] Zhou S, Feng Y, Wu W Y, Li Y, Liu J 2014 *Res. Astron. Astrophys.* **14** 104
- [4] Lü J H, Zhan Y, Lu J A 2000 *Proceedings of the CSEE* **20** 80 (in Chinese) [吕金虎, 占勇, 陆君安 2000 中国电机工程学报 **20** 80]
- [5] Packard N H, Crutchfield J, Fammrer J 1980 *Phys. Rev. Lett.* **45** 712

- [6] Yu S M 2008 *Acta Phys. Sin.* **57** 3374 (in Chinese) [禹思敏 2008 物理学报 **57** 3374]
- [7] Yu S M, Tang W K S, Lü J H, Chen G R 2008 *IEEE T. Circuits - II* **55** 1168
- [8] Shen C W, Yu S M, Lü J H, Chen G R 2014 *IEEE T. Circuits - I* **61** 854
- [9] Shen C W, Yu S M, Lü J H, Chen G R 2014 *IEEE T. Circuits - I* **61** 2380
- [10] Li Q X, Li K J 2007 *Publ. Astron. Soc. Jpn.* **59** 983
- [11] Xian M, Zhuang Z W, Xiao S P, Guo G R 1998 *Journal of National University of Defense Technology* **20** 60 (in Chinese) [鲜明, 庄钊文, 肖顺平, 郭桂蓉 1998 国防科技大学学报 **20** 60]
- [12] Lü J H, Lu J A, Chen S H 2002 *Chaotic Time Series Analysis and Its Application* (Wuhan: Wuhan University Press) p73 (in Chinese) [吕金虎, 陆君安, 陈士华 2002 混沌时间序列分析及其应用 (武汉: 武汉大学出版社) 第 73 页]
- [13] von Bremen F H, Udwadia F E, Proskurowski W 1997 *Physica D* **101** 1
- [14] Shimada I, Nagashima T 1979 *Prog. Theor. Phys.* **61** 1605
- [15] Wolf A, Swift J B, Swinney H L, Vastano J A 1985 *Physica D* **16** 285
- [16] Briggs K 1990 *Phys. Lett. A* **151** 27
- [17] Rosenstein M T, Collins J J, de Luca C J 1993 *Physica D* **65** 117
- [18] Barna G, Tsuda I 1993 *Phys. Lett. A* **175** 421
- [19] Jiang H F, Ma R J, Wei X Y, Wen W G 2006 *Journal of the China Railway Society* **28** 63 (in Chinese) [蒋海峰, 马瑞军, 魏学业, 温伟刚 2006 铁道学报 **28** 63]
- [20] Xu X K 2008 *Ph. D. Dissertation* (Dalian: Dalian Maritime University) (in Chinese) [许小可 2008 博士论文 (大连: 大连海事大学)]
- [21] Kantz H 1994 *Phys. Lett. A* **185** 77
- [22] Liang Y, Meng Q, Lu J R 2006 *Technical Acoustics* **25** 463 (in Chinese) [梁勇, 孟桥, 陆信人 2006 声学技术 **25** 463]
- [23] Yao T L 2013 *Ph. D. Dissertation* (Shanghai: East China University of Science and Technology) (in Chinese) [姚天亮 2013 博士论文 (上海: 华东理工大学)]
- [24] Yang S Q, Zhang X H, Zhao C A 2000 *Acta Phys. Sin.* **49** 636 (in Chinese) [杨绍清, 章新华, 赵长安 2000 物理学报 **49** 636]
- [25] Lu S, Wang H Y 2006 *Acta Phys. Sin.* **55** 572 (in Chinese) [卢山, 王海燕 2006 物理学报 **55** 572]
- [26] Lu Z B 2008 *Ph. D. Dissertation* (Wuhan: Chinese People's Liberation Army Navy Engineering University) (in Chinese) [陆振波 2008 博士论文 (武汉: 中国人民解放军海军工程大学)]
- [27] Yang Y F, Wu M J, Gao Z, Wu Y F, Ren X M 2012 *Journal of Vibration Measurement & Diagnosis* **32** 371 (in Chinese) [杨永锋, 仵敏娟, 高喆, 吴亚锋, 任兴民 2012 振动、测试与诊断 **32** 371]
- [28] Jain A K 2010 *Pattern Recogn. Lett.* **31** 651
- [29] May R M 1976 *Nature* **261** 459
- [30] Hénon M 1976 *Commun. Math. Phys.* **50** 69
- [31] Zhang Y P, Sun W H, Liu C A 2010 *Chin. Phys. B* **19** 050512

A novel method based on the fuzzy C-means clustering to calculate the maximal Lyapunov exponent from small data^{*}

Zhou Shuang^{1)2)†} Feng Yong¹⁾ Wu Wen-Yuan¹⁾ Wang Wei-Hua¹⁾²⁾

1) (Chongqing Key Laboratory of Automated Reasoning and Cognition, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China)

2) (University of Chinese Academy of Sciences, Beijing 100049, China)

(Received 3 June 2015; revised manuscript received 16 October 2015)

Abstract

In order to reduce errors caused by human factors to identify the linear region, we propose a new method based on the fuzzy C-means clustering for calculating the maximum Lyapunov exponent from small data. The method based on the changing characteristic of average divergence index curve is used to identify the linear region. Firstly, the average divergence index data are calculated from the small data algorithm for the given chaotic time series. Secondly, the fuzzy C-means clustering method is used for dividing the data into two classes (unsaturated and saturated data), and the unsaturated data are retained. Thirdly, the retained data are divided by the same clustering method into three classes (positive fluctuation data, zero fluctuation data and negative fluctuation data), and the zero fluctuation data are retained. Fourthly, the 3σ criterion is used for excluding gross errors to retain the valid from the selected data. Finally, the regression analysis and statistical test are used to identify the linear region from the valid data. The effectiveness of the proposed method can be demonstrated by the famous chaotic systems of Logistic and Henon. The calculated results are close to the theoretical values than the subjective method. Experimental results show that the proposed new approach is easier to operate, more efficient and more accurate as compared with the subjective recognition. But this method has its own shortcomings. 1) As the new method is verified by the simulation experiment, there exists no strict mathematical proof. 2) Since the difference algorithm is used in this new method, it will miss some detailed information in some cases. 3) The calculation accuracy still needs to be improved, so this method only serves as a reference to detect the linear region, it can not be applied to high precision engineering field. Considering the deficiencies of the new method, we will make further research to improve the calculation method for maximum Lyapunov exponent, so as to make it solve the real-time problem of the signal detection, and find the accurate location of abrupt climate change in the field of meteorology, to provide accurate satellite launch safety period in the field of space weather and other aspects. In short, studying the largest Lyapunov exponent from chaotic time series has a wide application prospect and practical significance.

Keywords: maximal Lyapunov exponent, linear region, fuzzy C-means clustering

PACS: 05.45.-a, 05.45.Pq, 05.45.Tp

DOI: 10.7498/aps.65.020502

^{*} Project supported by the National Natural Science Foundation of China (Grant No. 11301524) and the Chongqing Academicians Special Project Based on the Basic and Frontier Researches, China (Grant No. cstc2015jcyjys40001).

[†] Corresponding author. E-mail: zhoushuang@cigit.ac.cn