

◎ 工程与应用 ◎

非线性动力系统 R/S 法的改进及小波预测

毛军军^{1,2}, 程蓉², 吴涛^{1,2,3}MAO Jun-jun^{1,2}, CHENG Rong², WU Tao^{1,2,3}

1. 安徽大学 数学科学学院, 合肥 230039

2. 安徽大学 智能计算与信号处理教育部重点实验室, 合肥 230039

3. 南京大学 计算机软件新技术国家重点实验室, 南京 210093

1. Department of Mathematical Sciences, Anhui University, Hefei 230039, China

2. Key Laboratory of Intelligent Computing & Signal Processing of Ministry of Education, Anhui University, Hefei 230039, China

3. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China

E-mail: maojunjun@ahu.edu.cn

MAO Jun-jun, CHENG Rong, WU Tao. Improvement of R/S analysis in nonlinear system and wavelet forecast. *Computer Engineering and Applications*, 2009, 45(29): 198-201.

Abstract: By reconstructing the original data by wavelet transform based on R/S analysis and multifractal analysis, it analyses the nonlinear characteristics of the system of fixed-line phone users and mobile phone users and its evolutionary principle. According to fractal dimension, non-periodic cycle, holder index, it points out that the growth of phone users has strong stability, long-term memory and multifractal feature. With the order of partition function increasing, the multifractal strength increases. Moreover, the multifractal strength of mobile phone users is greater. Compared to the original data, the reconstructed data fit better. The number of phone users of the next three months is forecasted by the reconstructed data.

Key words: the number of phone users; Rescaled Range Analysis (R/S); non-periodic cycle; multifractal analysis; wavelet transform; forecast

摘要: 针对固定电话和移动电话用户人数的时间序列, 将小波变换与 R/S 分析法, 多重分形相结合, 探讨了系统的动力学特性和演化规律, 并从分形维数、非周期循环长度、奇异性指数等方面指出电话用户人数增长趋势存在强持续性和长期记忆性以及具有多重分形特征, 随着配分阶数的增大, 多重分形随之增强, 并得出移动电话人数的多重分形特性更强。接着对原始数据和小波重构后的数据进行了回归拟合比较, 结果证明小波重构后的数据拟合程度更高, 并利用重构的数据预测了未来三个月的电话用户人数。

关键词: 电话用户人数; R/S 分析; 非周期循环长度; 多重分形分析; 小波变换; 预测

DOI: 10.3778/j.issn.1002-8331.2009.29.058 文章编号: 1002-8331(2009)29-0198-04 文献标识码: A 中图分类号: TP18

1 序言

R/S 分析方法是一种普适性分析法, 近年来被广泛应用于研究金融市场、图像处理、气候趋势、地震预测等非线性动力系统的分形结构。尤其针对系统的长程相关等一系列中低频时序特征分析具有稳定性^[1]。但是对于短期高频时序数据如何在越来越小的尺度上观察越来越丰富的细节信息? 由于 R/S 分析方法技术内在的统计属性, 使其只能提供信号奇异性的全局描述; 因此在单一分形的基础上引入多重分形分析^[2], 用其来刻画信号的局部分形特征, 并从局部出发来研究其最终的整体特征。

小波分析^[3]是当前应用数学中一个迅速发展的新领域, 与 Fourier 变换、窗口 Fourier 变换 (Gabor 变换) 相比, 小波变换是空间 (时间) 和频率的局部变换, 因而能有效地从信号中提取信息, 通过伸缩和平移等运算功能对函数或信号进行多尺度细化分析, 解决了 Fourier 变换不能解决的许多困难问题, 被誉为“数学显微镜”。

将小波变换与 R/S 分析法, 多重分形分析相结合, 应用于电信行业, 对固定电话用户人数和移动电话用户人数进行了实证分析, 探讨了系统的动力学特性和演化规律, 并从分形维数、

基金项目: 国家自然科学基金 (the National Natural Science Foundation of China under Grant No.60675031); 安徽省高校自然科学研究计划项目 (the Natural Science Research Project of Higher Education of Anhui Province, China under Grant No.KJ2008B093); 中国博士后科学基金 (No.20070411028); 安徽大学人才队伍建设经费资助。

作者简介: 毛军军 (1973-), 女, 副教授, 博士, 主要研究方向是人工智能、信息优化; 程蓉 (1984-), 硕士研究生; 吴涛 (1970-), 教授, 博士后, 主要研究方向是粒度计算、智能优化。

收稿日期: 2008-11-14 修回日期: 2009-02-03

非周期循环长度、奇异性指数等方面指出电话用户人数增长趋势存在强持续性和长期记忆性以及多重分形特征,随着配分阶数的增大,多重分形随之增强,并得出移动电话人数的多重分形特性更强。接着对原始数据与小波重构后的数据进行了回归拟合比较,并利用重构的数据预测了未来三个月的电话用户人数。

2 理论与方法

2.1 R/S 分析法

R/S(Rescaled Range Analysis)分析法,又被称为重标极差法,是英国水文学家 H·E·Hurst 于 1951 年发现的无参数统计方法。R/S 分析法的计算结果如下:

假设 P 为长度为 N 的时间序列。将这 N 个数据划分为 m 个子区间,每个子区间的长度为 n 。每个子区间记为 I_a ,其中 $a=1,2,\dots,m$ 。 I_a 中的元素记为 N_k ,其中 $k=1,2,\dots,n$ 。子区间 I_a 的平均值 M_a 计算如下:

$$M_a = \frac{1}{n} \sum_{k=1}^n N_{k,a} \quad (1)$$

每个子区间 I_a 相对平均值的累计离差 $X_{k,a}$ 为:

$$X_{k,a} = \sum_{i=1}^k (N_{i,a} - M_a), k=1, 2, \dots, n \quad (2)$$

下面定义极差 RI_a 为子区间中最大值和最小值间的差值:

$$RI_a = \max(X_{k,a}) - \min(X_{k,a}), 1 \leq k \leq n \quad (3)$$

子区间 I_a 的标准差记为 SI_a ,计算方法如下:

$$SI_a = \sqrt{\frac{1}{n} \sum_{k=1}^n (N_{k,a} - M_a)^2} \quad (4)$$

由每个子区间的 RI_a 和 SI_a ,计算重标极差 $FI_a = RI_a / SI_a$,由于划分了 m 个连续的子区间,因此必须计算一个平均的重标极差 F_n ,计算方法如下:

$$F_n = \frac{1}{m} \sum_{a=1}^m \left(\frac{RI_a}{SI_a} \right) \quad (5)$$

通过增加 n 的数值, m 的值减少,直到 $n=(m-1)/2$ 。当然, n 要能整除 $(m-1)$ 。

Hurst 指数 H 有关系式:

$$F_n = c \times n^H \quad (6)$$

对上式两端取对数可得:

$$\log(F_n) = \log(c) + H \times \log(n) \quad (7)$$

通过对式(7)用最小二乘法拟合求得 H 估算值。

Mandelbort 指出,若 $0.1 < H < 1.0$,则它与分形维 D 的关系为: $D=2-H$ 。据此,可以求得时间序列的维数。

2.1.1 Hurst 指数

H 指数的取值一般在 0 到 1 之间。 H 的大小可以揭示许多重要的非线性动力学特性和演化规律。

(1) $H \in (0, 0.5)$, 反映的是一个“均值回复”的反持续性时间序列,具有频繁逆向转换和不连续倾向。 H 值越接近 0,反向变化的趋势越明显。

(2) $H=0.5$, 反映的是一个随机的与时间无关联的序列,是标准的布朗运动,事件的过去不会影响未来。

(3) $H \in (0.5, 1)$, 反映的是一个持续性的时序变化过程,具有持续性趋势和循环倾向,系统存在长期和无周期的循环。 H 值越接近 1,趋势性或持续性越强。过去的事件对未来具有影响,存在长程相关特征,会形成一个个大的循环,这些循环没有

固定的周期。

2.1.2 改进

以往都是利用统计量 V_n 来估计非周期循环的长度。计算公式为:

$$V_n = \frac{(R/S)_n}{\sqrt{n}} \quad (8)$$

式(8)中的 n 为观测的天数。用这种方法是先在 V 统计图中找到拐点,由拐点再求出非周期循环的长度,比较费时。

而该文在上述 H 指数算法的基础上,加入了最大非周期循环长度的寻找,在算出 H 指数的同时也算出了非循环周期的长度。具体算法如下:

设算出的平均重标极差 F_n 有 q 个,记为: $F_n(1), F_n(2), \dots, F_n(q)$, 令

$$X(j) = |F_n(j+1) - F_n(j)|, j=1, 2, \dots, q-1 \quad (9)$$

因此可以得到 $X(j)$ 的最大值 $\max_{1 \leq j \leq q-1} X(j)$, 记为 $X(r)$, 其中

的 r 即为非周期循环的长度,也就是说系统平均经过 r 天后失去对初始条件的记忆。

2.2 多重分形

设时间序列为 $\{x_i\}, i=1, 2, \dots, T$, 按大小 $\varepsilon (\varepsilon < 1)$ 均匀等分序列,份数为 N , 令

$$P_j(\varepsilon) = \frac{I_j(\varepsilon)}{\sum I_j(\varepsilon)}, j=1, 2, \dots, N \quad (10)$$

其中 $I_j(\varepsilon)$ 为大小为 ε 时,第 j 份内各数之和, \sum 是对所有各份求和。

定义配分函数 $\chi_q(\varepsilon)$

$$\chi_q(\varepsilon) = \sum P_j^q(\varepsilon) = \varepsilon^{\tau(q)} \quad (11)$$

从 $\tau(q)$ 中可以计算出多重分形谱 $f(\alpha)$, 其计算公式如下:

$$\frac{d}{dq} [\tau(q)] = \alpha(q) \quad (12)$$

$$\tau(q) = q\alpha(q) - f(\alpha) \quad (13)$$

多重分形的奇异性指数 α 描述了每一次分割形成的质量分布概率子集的大小。分形谱的宽度 $\Delta\alpha = \alpha_{\max} - \alpha_{\min}$ 表征了最大、最小概率间的差别。也就是概率变化的不均匀性,表明多重分形的强弱变化程度。相应的最大、最小概率子集分形维数的差别 $\Delta f = f(\alpha_{\min}) - f(\alpha_{\max})$ 反映了高低值出现频率的变化。

2.3 小波分析

小波分析(Wavelets Analysis, WA)是 20 世纪 80 年代中期发展起来的一门新兴的数学理论和方法,通过小波变换可以将研究对象分解到不同尺度的空间进行分析和处理,然后再根据需要进行相应的重构。

小波函数的定义:如果函数 $\Psi(x)$ 是一个平方可积函数,且

Fourier 变换满足, $C_\Psi = \int_R \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$ 称 $\Psi(x)$ 是一个基本小波或母小波,将母小波 $\Psi(x)$ 经伸缩和平移后,就可以得到一个

小波序列。

对于连续的情况,小波序列为:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right) \quad a, b \in R, a \neq 0 \quad (14)$$

其中 a 为伸缩因子, b 为平移因子。

对于离散的情况,小波序列为:

$$\Psi_{j,k}(t) = 2^{-\frac{j}{2}} \Psi(2^{-\frac{j}{2}} t - k) \quad j, k \in \mathbb{Z} \quad (15)$$

分辩分析理论,被处理的信号 A_{j-1} 可分解为低一级逼近(低频)信号 A_j 和细节(高频)信号 D_j 之和。运用 Mallat 算法可以将信号一层层进行分解,每一层分解的结果是将上次分解得到的逼近(低频)信号再分解成逼近(低频)和细节(高频)两部分。Mallat 将多分辩分析和小波变换结合在一起,提出了离散信号系数分解与重建的 Mallat 算法:

$$cA_{j+1}^k = \sum_n \bar{h}_k - 2n \quad cA_j^k \quad (16)$$

$$cD_{j+1}^k = \sum_n \bar{g}_k - 2n \quad cA_j^k$$

式(16)称为信号的分解算法。

$$cA_j^k = \sum_n cA_{j+1}^k h_{k-2n} + \sum_n cD_{j+1}^k g_{k-2n} \quad (17)$$

式(17)称为信号的重建算法。

小波的选择不是唯一的,但它的选择应满足定义域紧支撑条件和小波容许条件,小波函数的消失阶也很重要。必须尽可能选择一个最小的消失阶次数而又能保证其检测出信号的最大奇异性指数。经过计算比较,db4 小波函数不仅可以满足算法的条件,还可以节省内存和计算时间。

3 实例分析

3.1 数据来源及预处理

原始数据(固定,移动电话用户人数(万户))取自信产部官方网站。采样跨度为按每月统计,从 2002 年 1 月到 2008 年 8 月,总计 80 个数据。这段期间,固话和移话用户人数的波动趋势如下图 1,图 2。

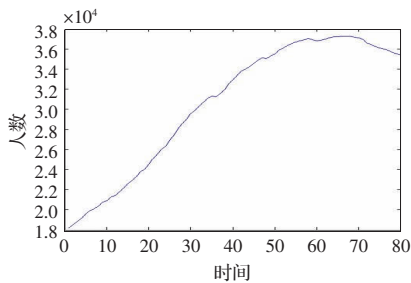


图 1 固定电话用户人数(万户)趋势图

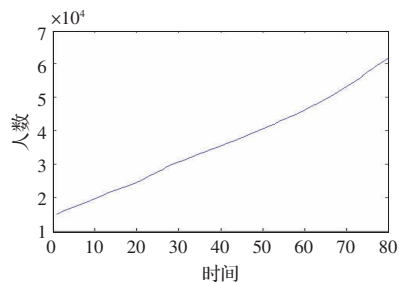


图 2 移动电话用户人数(万户)趋势图

利用 Matlab 软件中的小波工具箱对固定电话用户人数和移动电话用户人数的序列数据分别进行 3 层 Daubechies (db4) 小波分解,然后保留低频部分,将高频部分全置为零进行重构。以下就采用重构后的数据进行 R/S 分析和预测。

3.2 R/S 分形结果及分析

表 1 R/S 分析结果参数表

指标名称	H 值	分形维 D	非周期循环长度/月
固话用户人数	0.722 2	1.277 8	25
移话用户人数	0.723 2	1.276 8	4

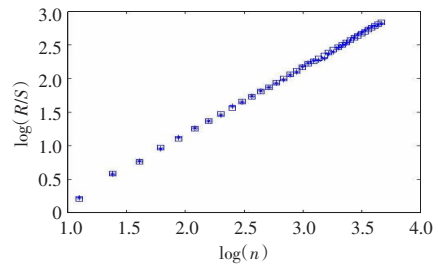


图 3 固话、移话的 R/S 分析

其中的“*”表示固定电话用户人数,“正方形”表示移动电话用户人数。

(1)无论是固定电话用户人数还是移动电话用户人数,其 H 值都是大于 0.5 的,说明从 2002 年 1 月到 2008 年 8 月期间,两者具有强持久性,均为状态持续性的或趋势增强的序列。移动电话用户人数的值 H 大于固定电话用户人数的 H 值,说明移动电话用户人数的趋势增强的强度更大,持续性更强。这是符合实际情况的。因为移动电话可以随身携带,且品种多样,由此不断地吸引了越来越多的用户。

(2)从分形维数来看,两个指标构成的时间序列的 D 值分别为 1.277 8 和 1.276 8,均为分维时间序列,都存在着分维现象,具有非线性特征。分形维数代表非规则性,无序、弯曲或填充空间的能力。一般地,维数越大,弯曲越大,不规则程度越大,复杂性越高。固定电话用户人数的维数要大于移动电话用户人数的维数,说明固定电话用户人数较移动电话人数而言影响因素更为复杂,其不规则性更强。这与近年来固定电话不断推出优惠政策,并且开拓了农村市场,使得固定电话用户人数的构成复杂程度不断加深是十分吻合的。

(3)从表 1 可以知道固定电话用户人数的非周期循环长度为 25 个月,而移动电话用户人数的非周期循环长度为 4 个月。可见,固定电话用户人数的记忆长度大于移动电话用户人数的记忆长度。

3.3 多重分形结果及分析

图 4 和图 5 分别是固定电话人数和移动电话人数的多重分形谱曲线,即 $f(\alpha) - \alpha$ 曲线(横坐标是 α)。

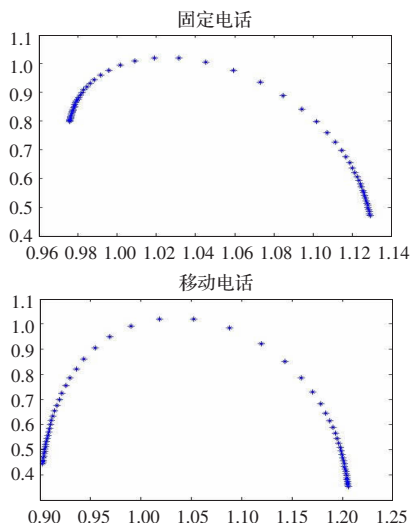
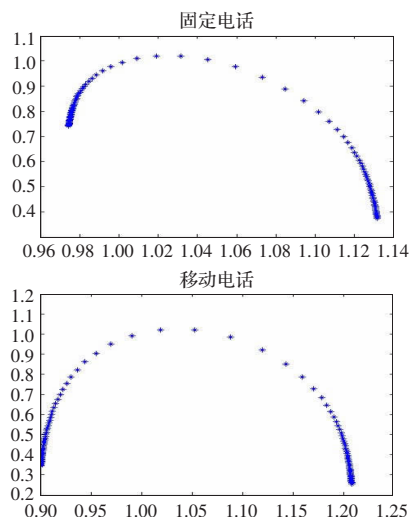
(1)由图 4 和图 5 可以看出奇异谱曲线都呈现左钩状,表明分布是不均匀的,即这两个指标都存在多重分形特征。

(2)根据表 2,不同的 q 值会带来不同的统计结果。随着 q 的增大,一是这两个指标的奇异谱的宽度都逐渐增大,多重分形随之增强;二是相应的分形维的离差也逐渐增大,即左钩状越趋明显。

表 2 分形谱维数($\Delta\alpha$)和相应分形谱离差(Δf)随 q 值的变化

q 的取值范围	-30:30		-50:50	
	固定电话	移动电话	固定电话	移动电话
$\Delta\alpha$	0.153 1	0.302 6	0.157 2	0.307 8
Δf	0.548 0	0.667 0	0.645 7	0.764 0

(3)由表 2 可以看出:无论 q 的取值如何,移动电话用户人

图4 $f(\alpha)$ - α 曲线($q=-30:30$)图5 $f(\alpha)$ - α 曲线($q=-50:50$)

数这个指标的奇异谱宽度都大于固定电话用户人数的奇异谱宽度,这表明前者的多重分形特征更强。

3.4 预测

将时间 2002 年 1 月至 2008 年 8 月按月依次标记为序列 $T=\{1, 2, \dots, 80\}$, 并将其作为自变量, 固定电话用户人数和移动电话用户人数分别作为因变量, 分别对原始数据和小波重构的数据, 利用 SPSS 统计软件进行回归拟合, 比较结果如表 3、表 4:

表3 固定电话用户人数回归拟合结果

	可决系数	标准误差
原始数据	0.865	2 407.643 1
小波重构的数据	0.879	2 253.409

表4 移动电话用户人数回归拟合结果

	可决系数	标准误差
原始数据	0.994	1 014.190 1
小波重构的数据	0.994	1 000.790

由上面两个表可以看出, 无论是固定电话用户人数, 还是移动电话用户人数, 使用小波重构的数据进行回归拟合的拟合程度更高, 标准误差更小。因此, 使用小波变换对原始数据进行处理, 的确是消除了原始数据中的噪声干扰, 有利于提高预测精度。

下面使用小波重构的数据来预测 2008 年 9 月, 10 月, 11 月的固定电话用户人数和移动电话用户人数, 结果如表 5:

表5 预测结果

	9 月	10 月	11 月
固定电话用户人数/万户	41 367.54	41 629.01	41 890.49
移动电话用户人数/万户	58 983.69	59 544.81	60 105.92

4 结束语

对原始数据进行小波变换处理, 对小波重构后的数据应用 R/S 分析法及多重分形分析进行研究。结果表明, 我国固定电话用户人数, 移动电话用户人数增长趋势具有强持续性且存在分维现象以及具有多重分形特征, 随着配分阶数的增大, 多重分形随之增强, 并且移动电话用户人数的多重分形特性更强。选用的小波基函数是 Daubechies(db4), 且进行了三层分解, 然后将高频部分全置为 0, 再进行小波重构, 利用重构后的数据进行预测。不同的小波基函数及分解层数会得出不同的重构数据, 最终会影响预测精度。所以如何选取小波基函数、分解层数以及拟合曲线函数来获得最优预测结果还有待进一步研究。

参考文献:

- [1] Peters E E. Fractal market analysis: Applying Chaos theory to investment and economics[M]. New York: John Wiley & Sons, Inc, 1994.
- [2] 张济忠. 分形[M]. 北京: 清华大学出版社, 1995.
- [3] 赵松年, 熊小芸. 子波变换与子波分析[M]. 北京: 电子工业出版社, 1997.
- [4] Peters E E. Chaos and order in capital markets: A new view of cycles, prices and market volatility[M]. 2nd ed. New York: John Wiley & Sons, Inc, 1996.
- [5] 刘涛, 曾祥利, 曾军. 实用小波分析入门[M]. 北京: 国防工业出版社, 2006.
- [6] 刘秀新, 舒华英. 基于 R/S 分析的通信业务收入波动性研究[J]. 中国管理科学, 2005, 13(Z1): 236-239.
- [7] 周景宏, 郝伟东, 禹海兰. 小波方法在股市分析中的应用[J]. 华北电力大学学报, 2006, 26(2): 47-51.
- [8] 刘晔. 透视投影变换总矩阵的研究[J]. 黑龙江商学院学报: 自然科学版, 1997, 13(2): 48-52.
- [9] 傅雅宁. 计算机图形学教程[M]. 北京: 国防工业出版社, 2005: 57-117.
- [10] 扬饮. 计算机图形学[M]. 北京: 清华大学出版社, 2005: 76-111.
- [11] Hearn D, Baker M P. 计算机图形学[M]. 蔡士杰, 译. 北京: 电子工业出版社, 2005: 188-329.

(上接 186 页)

参考文献:

- [1] 杨燕燕. 基于透视反求的基本体素三维重构关键技术研究[J]. 计算机应用研究, 2006, 23(7): 55-56.
- [2] 杨玉芳. 透视图在反求设计中的应用[J]. 工程图学学报, 2006, 27(5): 121-125.