

利用傅里叶谐波分析法的时序数据周期迭代辨识算法^{*}

黄雄波¹, 胡永健²

(1. 佛山职业技术学院 电子信息系, 广东 佛山 528000; 2. 华南理工大学 电子与信息学院, 广州 510641)

摘要: 针对现有的时序数据其周期辨识算法存在着辨识精度低及计算成本高的问题,在傅里叶谐波分析法的基础上,提出了一种具有基频迭代机制的周期辨识算法。首先,利用方差分析法从原始序列中析出其周期长度的整型估算值;然后,以任意小的频率间隔在估算值区间内进行傅里叶谐波的迭代拟合;最后,基于最小拟合残差和的准则来确定最优的周期成分。实验表明,该算法不仅具有良好的计算效能,而且还能精确地辨识出与序列样本长度无关的周期成分。

关键词: 时序数据; 周期成分; 迭代辨识; 傅里叶谐波分析法; 方差分析法

中图分类号: TP301 **文献标志码:** A **文章编号:** 1001-3695(2018)06-1716-04

doi:10.3969/j.issn.1001-3695.2018.06.025

Iterative identification algorithm for periodic component of time series data using Fourier analysis method

Huang Xiongbo¹, Hu Yongjian²

(1. Dept. of Electronic & Information Engineering, Foshan Professional Technical College, Foshan Guangdong 528000, China; 2. School of Electronic & Information Engineering, South China University of Technology, Guangzhou 510641, China)

Abstract: In view of the existing time series data, the algorithm of periodic identification has the problems of low identification accuracy and high computational cost, on the basis of Fourier harmonic analysis, this paper proposed a periodic identification algorithm with fundamental frequency iteration mechanism. First of all, it used the method of analysis of variance from the original sequence in the precipitation period length integer estimation. Then, to the frequency interval arbitrarily small in the estimated value of iterative fitting, interval Fourier harmonic. Finally, to determine the optimal cycle components based on the criterion of minimum fitting error. The experimental results show that the algorithm not only has good computational efficiency, but also can identify the periodic components independent of the sample length accurately.

Key words: time series data; periodic component; identification of the iteration; Fourier analysis method; variance analysis

0 引言

时序数据就是按照时间的先后顺序对同一物理现象进行观测而得到的一系列数值。由于这些观测数据往往具有一定的前后相依关系,据此,对它们进行有关的统计建模分析后,便可达到发现事物的内在发展规律及揭示其未来发展趋势的目的^[1]。在实际生活中,受自然界固有特性的影响,事物在其发展历程中通常表现出一定的周期行为,例如各种天文和水文历史资料、经济运行数据、设备监控记录等,它们或多或少具有周期性波动的特征^[2,3]。据此,对时序数据进行周期辨识能有利于人们认识事物的本质。

时序数据的周期辨识算法有很多种,如自相关分析法、傅里叶谐波分析法、最大熵谱分析法、逐步回归分析法及小波分析法等^[4-6],其中,傅里叶谐波分析法因具有清晰的物理意义和计算简便等优点,故在实际中得到了广泛的应用^[7-9]。例如,文献[10]根据傅氏级数的定义及有关定理,将时序数据拟合成傅氏级数,并利用有限项三角函数之和来估算各项傅氏级

数的系数,从而设计实现了一种有效的周期成分的提取算法;进一步,文献[11]对多周期时序数据(各个基本周期两两互质)的辨识问题进行了研究,并基于最小二乘法和矩阵计算理论得到了一种高精度的多周期时序数据的傅氏级数拟合法;随后,文献[12]通过对傅里叶谐波系数进行 Taylor 级数的展开,得到了一种具有纳新机制的时序数据周期模式的递推发现算法。

目前,基于傅里叶谐波的周期辨识算法在实际应用中还存在一些着一些问题,较为突出的有,辨识的周期成分受样本长度的影响所导致的拟合误差,以及计算傅里叶谐波系数时所引起的计算成本等问题。为有效地解决上述问题,本文拟设计一种具有基频迭代机制的周期辨识算法,并以太阳黑子数的周期辨识实验来证明算法的有效性及其先进性。

1 问题描述

一般地,时序数据可用式(1)的非平稳数据模型来描述^[13]。

$$Y_t = H_t + P_t + X_t \quad (1)$$

收稿日期: 2017-04-14; **修回日期:** 2017-05-22 **基金项目:** 广东省自然科学基金团队项目(9351064101000003);广东省应用型科技研发专项基金资助项目(2015B010130003);广东省科技计划工业攻关项目(2011B010200031);佛山职业技术学院横向重点资助项目(H201815)

作者简介: 黄雄波(1975-),男,广东南海人,副教授,博士研究生,主要研究方向为时间序列分析、数字图像处理(xb-Huang@hotmail.com);胡永健(1962-),男,湖北武汉人,教授,博导,主要研究方向为多媒体信息安全、信息隐藏、数字图像处理。

其中: $H_t, P_t, X_t (t=1, 2, \dots, n)$ 分别为时序数据的趋势项、周期项和随机项。若周期序列 P_t 满足傅里叶级数的展开条件, 则可将 P_t 以样本长度 n 为周期延拓至 $(-\infty, +\infty)$ 上, 这样便可以在区间 $\left[-\frac{n}{2}, \frac{n}{2}\right]$ 上把 P_t 表示为一系列频率成倍递增的谐波之和, 且谐波数量不超过 $\frac{n}{2}$ 项, 即

$$\hat{P}_t = a_0 + \sum_{i=1}^m \lambda_i \sin(i\omega_0 t + \theta_i) \quad (2)$$

其中: \hat{P}_t 为 P_t 的拟合值, 由直流成分 a_0 和 $m (m = \text{int}(\frac{n}{2}))$, int 为取整运算) 项频率成倍递增的各次谐波线性叠加而成; λ_i 为各次谐波的振幅; $\omega_0 = \frac{2\pi}{n}$ 为各次谐波的基频; θ_i 为各次谐波的初始相角。

根据三角函数公式, 式(2)可转换为

$$\hat{P}_t = \sum_{i=0}^m (a_i \cos i \frac{2\pi t}{n} + b_i \sin i \frac{2\pi t}{n}) \quad (3)$$

其中: $\hat{a}_i = \lambda_i \cos \theta_i, \hat{b}_i = \lambda_i \sin \theta_i$, 它们分别为各次谐波的余弦分量和正弦分量的幅值。为了确定式(3)中的系数, 根据最小二乘法, 应使式(4)中的拟合误差序列 e_t 的平方和最小化。

$$\sum_{t=1}^n |e_t|^2 = \sum_{t=1}^n |P_t - \hat{P}_t|^2 \quad (4)$$

于是, 式(4)右边分别对 \hat{a}_i, \hat{b}_i 求偏导数, 并令它们等于 0, 则有

$$\begin{cases} \sum_{t=1}^n \left[P_t - \sum_{i=0}^m \left(\hat{a}_i \cos i \frac{2\pi t}{n} + \hat{b}_i \sin i \frac{2\pi t}{n} \right) \right] \cos i \frac{2\pi t}{n} = 0 \\ \sum_{t=1}^n \left[P_t - \sum_{i=0}^m \left(\hat{a}_i \cos i \frac{2\pi t}{n} + \hat{b}_i \sin i \frac{2\pi t}{n} \right) \right] \sin i \frac{2\pi t}{n} = 0 \end{cases} \quad (5)$$

利用三角函数的正交特性求解方程组式(5), 便可以得到如式(6)所示的各傅里叶谐波系数的估算表达式。

$$\begin{cases} \hat{a}_0 = \frac{1}{n} \sum_{t=1}^n P_t \\ \hat{a}_i = \frac{2}{n} \sum_{t=1}^n P_t \cos \frac{2\pi i t}{n} \quad i=1, 2, \dots, m \\ \hat{b}_i = \frac{2}{n} \sum_{t=1}^n P_t \sin \frac{2\pi i t}{n} \end{cases} \quad (6)$$

令 $C_i^2 = (\hat{a}_i^2 + \hat{b}_i^2)$, 类似于回归分析, 周期序列 P_t 的整体方差数值 S_p^2 可视为各次谐波的方差和残余序列的方差叠加而成, 即有 $S_p^2 = \frac{1}{n} \sum_{t=1}^n (P_t - \hat{P}_t)^2$, 据此, 可以使用式(7)所示的统计量 t 对各次谐波的方差显著性进行评价^[14]。

$$t = \frac{\frac{1}{2} \frac{C_i^2}{2}}{\left(S_p^2 - \frac{1}{2} C_i^2 \right) / (n-2-1)} \sim F(2, n-3) \quad (7)$$

由于式(7)中的统计量 t 服从分子、分母的自由度分别为 2、3 的 F 分布, 故在某一置信度的条件下, 利用 F 检验, 就可逐一检验各个谐波的显著性。在实际应用中, 可挑选那些振幅幅值 C_i^2 比较大的谐波成分进行周期显著性判别, 若某一对傅里叶谐波系数 \hat{a}_i, \hat{b}_i 通过了周期显著性的判别, 则说明 P_t 具有长度为 $L = \frac{n}{i}$ 的周期成分。

从上述的分析可知, 傅里叶谐波分析法中的拟合序列 \hat{P}_t 其周期长度 L 等于样本长度 n 与某一正整数的商, 而当实际的周期长度 L 不能被样本长度 n 整除时, 易知, 此时 \hat{P}_t 的周期长度 L 将出现偏差, 进而引发周期辨识的精度问题。

2 提出的时序数据周期迭代辨识算法

针对上述问题, 本文拟设计一种具有基频迭代机制的时序数据周期辨识算法。算法的主要思想是, 首先利用方差分析法从周期序列 P_t 中析出其周期长度的整型估算值 $\ell (\ell \in N)$, 然后基于三角函数变换公式推导出具有基频迭代机制的傅里叶谐波系数表达式。在此基础上, 以任意小的频率迭代搜索步长 Δf 在 $[\ell-1, \ell+1]$ 区间内进行相应的傅里叶迭代拟合, 最后根据最小拟合残差和的准则来确定最优的周期成分。该算法以频率迭代的方式来确定最优的周期成分, 这不仅解决了拟合谐波的周期长度受制于样本长度 n 的影响, 而且还具有高效的计算效能。

2.1 利用方差分析法估算周期序列的周期长度

用方差分析法对周期序列 P_t 进行周期分析, 其基本原理就是把序列样本分别分成 ℓ 组, 然后检验各组数据的均值是否有显著性的差异, 并根据检验结果来判定 ℓ 是否为序列的周期长度。具体的处理过程如下:

- a) 按照式(8)所示的划分方式, 把周期序列 P_t 分成 ℓ 组。
- $$\begin{cases} L_1: (P_1, P_{\ell+1}, \dots, P_{\text{int}(n/\ell)+1}) \\ L_2: (P_2, P_{\ell+2}, \dots, P_{\text{int}(n/\ell)+2}) \\ \vdots \\ L_\ell: (P_\ell, P_{\ell+\ell}, \dots, P_{\text{int}(n/\ell)+\ell}) \end{cases} \quad \ell=2, 3, \dots, \text{int}\left(\frac{n}{2}\right) \quad (8)$$

b) 按照式(9)~(11)分别计算各分组子样本的总偏差平方和 SST、组间偏差平方和 SSA 及组内偏差平方和 SSE。其中: n_1, n_2, \dots, n_ℓ 分别表示各组内的样本个数; Σ 表示对所有具有其后面所示形式的数据求和; ΣL_ℓ 表示只对第 ℓ 组内具有一定形式的数据求和。

$$\text{SST} = \sum_{t=1}^n \left[P_t^2 - \frac{(\sum P_t)^2}{n} \right] \quad (9)$$

$$\text{SSA} = \left[\frac{(\sum L_1: P_t)^2}{n_1} + \frac{(\sum L_2: P_t)^2}{n_2} + \dots + \frac{(\sum L_\ell: P_t)^2}{n_\ell} \right] - \frac{(\sum P_t)^2}{n} \quad (10)$$

$$\text{SSE} = \text{SST} - \text{SSA} \quad (11)$$

c) 计算式(12)所示的校验统计量, 对于给定的显著性水平 α , 查 F 分布临界值表得到临界值 $f_{1-\alpha}(\ell-1, n-\ell)$ 。若 $F > f_{1-\alpha}(\ell-1, n-\ell)$, 则表示在显著性水平 α 下, 各组数据的均值有显著性差异, 此时可认为 ℓ 是周期序列 P_t 的周期长度; 否则, ℓ 不是周期序列 P_t 的周期长度。

$$F = \frac{\text{SSA}/(\ell-1)}{\text{SSE}/(n-\ell)} \quad (12)$$

2.2 傅里叶谐波系数的迭代计算表达式

析出周期长度的整型估算值后, 可以在 $[\ell-1, \ell+1]$ 区间内以不同的基频角频率代入式(6)计算出其对应的傅里叶谐波系数, 并可根据最小拟合残差和的准则来确定最优的基频角频率。设基频角频率的迭代搜索步长为 Δf , 则第 p 次傅里叶谐波系数 $\hat{a}_{\ell(p)}, \hat{b}_{\ell(p)}$ 分别为

$$\begin{cases} \hat{a}_{\ell(p)} = \frac{2}{n} \sum_{t=1}^n P_t \cos \left(\frac{2\pi t}{\ell} + p\Delta f \right) \\ \hat{b}_{\ell(p)} = \frac{2}{n} \sum_{t=1}^n P_t \sin \left(\frac{2\pi t}{\ell} + p\Delta f \right) \end{cases} \quad p=1, 2, 3, \dots \quad (13)$$

为了减少重复计算, 有必要对式(13)进行迭代改进, 即推导出 $\hat{a}_{\ell(p+1)}, \hat{b}_{\ell(p+1)}$ 与 $\hat{a}_{\ell(p)}, \hat{b}_{\ell(p)}$ 之间的迭代表达式。

由于

$$\begin{cases} \hat{a}_{\ell(p+1)} = \frac{2}{n} \sum_{i=1}^n P_i \cos\left(\frac{2\pi}{\ell} t + (p+1)\Delta f\right) \\ \hat{b}_{\ell(p+1)} = \frac{2}{n} \sum_{i=1}^n P_i \sin\left(\frac{2\pi}{\ell} t + (p+1)\Delta f\right) \end{cases} \quad p=1,2,3,\dots \quad (14)$$

则利用三角函数的两角和公式对式(14)进行分解后,有

$$\begin{cases} \hat{a}_{\ell(p+1)} = \frac{2}{n} \sum_{i=1}^n P_i \left(\cos\left(\frac{2\pi}{\ell} t + p\Delta f\right) \cos\Delta f - \sin\left(\frac{2\pi}{\ell} t + p\Delta f\right) \sin\Delta f \right) \\ \hat{b}_{\ell(p+1)} = \frac{2}{n} \sum_{i=1}^n P_i \left(\sin\left(\frac{2\pi}{\ell} t + p\Delta f\right) \cos\Delta f + \cos\left(\frac{2\pi}{\ell} t + p\Delta f\right) \sin\Delta f \right) \end{cases} \quad p=1,2,3,\dots \quad (15)$$

联合式(13)和(15),便可得到如下的矩阵递推迭代表达式。

$$\begin{bmatrix} \hat{a}_{\ell(p+1)} \\ \hat{b}_{\ell(p+1)} \end{bmatrix} = \begin{bmatrix} \hat{a}_{\ell(p)} & -\hat{b}_{\ell(p)} \\ \hat{b}_{\ell(p)} & \hat{a}_{\ell(p)} \end{bmatrix} \begin{bmatrix} \cos\Delta f \\ \sin\Delta f \end{bmatrix} \quad p=1,2,3,\dots \quad (16)$$

2.3 最优周期长度的确定

设周期序列 P_t 的拟合残差序列为 e_t , 则有

$$e_t = P_t - \left[a_0 + \sum_{i=1}^m \left(\hat{a}_i \cos i \frac{2\pi}{n} t + \hat{b}_i \sin i \frac{2\pi}{n} t \right) \right] \quad (17)$$

而对应的拟合残差平方和为

$$Q = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left\{ P_i - \left[a_0 + \sum_{i=1}^m \left(\hat{a}_i \cos i \frac{2\pi}{n} t + \hat{b}_i \sin i \frac{2\pi}{n} t \right) \right] \right\}^2 \quad (18)$$

利用式(16),可以在区间 $[\ell-1, \ell+1]$ 内迭代计算出不同基频条件下的傅里叶谐波系数,把这些谐波系数代入式(18)便可得到一系列的拟合残差平方和 Q_1, Q_2, \dots, Q_q , 而周期序列 P_t 的最优周期长度应为 $\min(Q_1, Q_2, \dots, Q_q)$ 所对应的基频频率。

一般地,任意周期序列表示为傅里叶级数时需要无穷项才能完全拟合原序列,但由于傅里叶级数的主要能量集中在基频成分上,故求解 $\min(Q_1, Q_2, \dots, Q_q)$ 的过程可简化为式(19)所示的处理形式。

$$\max \left(\sqrt{(\hat{a}_{\ell(1)})^2 + (\hat{b}_{\ell(1)})^2}, \sqrt{(\hat{a}_{\ell(2)})^2 + (\hat{b}_{\ell(2)})^2}, \dots, \sqrt{(\hat{a}_{\ell(q)})^2 + (\hat{b}_{\ell(q)})^2} \right) \quad (19)$$

其中: $\{(\hat{a}_{\ell(1)}, \hat{b}_{\ell(1)}), (\hat{a}_{\ell(2)}, \hat{b}_{\ell(2)}), \dots, (\hat{a}_{\ell(q)}, \hat{b}_{\ell(q)})\}$ 为各次迭代所得的基频傅里叶谐波系数。

获取最优周期长度 $\hat{\ell}$ 及其基频 $\omega_o = \frac{2\pi}{\hat{\ell}}$ 所对应的傅里叶谐

波系数 $\hat{a}_{\hat{\ell}(1)}, \hat{b}_{\hat{\ell}(1)}$ 之后,可以继续利用式(16)迭代计算出最优周期的其他谐波成分 $(2\omega_o, 3\omega_o, \dots)$ 的傅里叶系数 $(\hat{a}_{\hat{\ell}(2)}, \hat{b}_{\hat{\ell}(2)}, \hat{a}_{\hat{\ell}(3)}, \hat{b}_{\hat{\ell}(3)}, \dots)$, 谐波数量的选取由辨识精度决定。

2.4 时序数据周期迭代辨识算法的实现

对上述的分析进行归纳和总结,便可以设计实现对应的时序数据周期迭代辨识算法。

输入: 样本长度为 n 的周期序列 P_t , 频率迭代搜索步长 Δf , 辨识精度 ε 。

输出: 周期序列 P_t 的最优周期长度 $\hat{\ell}$ 及其对应的傅里叶谐波系数。

a) 按照式(8)所示的分组方法, 分别把 P_t 分为 $2, 3, \dots, \text{int}(\frac{n}{2})$ 组子样本序列, 然后根据式(9)~(12)对各分组子样本的均值进行显著性检验, 若有显著性的差异, 则析出对应的分组长度 ℓ ; 否则, 时序数据 P_t 不含周期成分, 直接转步骤 e)。

b) 如式(16)所示, 在 $[\ell-1, \ell+1]$ 内以 Δf 的步长计算不同基频条件下的首个傅里叶谐波系数 $\{(\hat{a}_{\ell(1)}, \hat{b}_{\ell(1)}), (\hat{a}_{\ell(2)}, \hat{b}_{\ell(2)}), \dots, (\hat{a}_{\ell(q)}, \hat{b}_{\ell(q)})\}$ 。

c) 分别计算步骤 b) 中的傅里叶谐波系数的幅值, 并按照式(19)所示方法确定周期序列 P_t 的最优周期长度 $\hat{\ell}$ 。

d) 令 $\omega_o = \frac{2\pi}{\hat{\ell}}$, 在满足辨识精度 ε 的条件下, 利用式(16)

迭代计算出最优周期的其他谐波成分 $(2\omega_o, 3\omega_o, \dots)$ 所对应的傅里叶谐波系数 $\hat{a}_{\hat{\ell}(2)}, \hat{b}_{\hat{\ell}(2)}, \hat{a}_{\hat{\ell}(3)}, \hat{b}_{\hat{\ell}(3)}, \dots$ 。

e) 算法结束, 输出相应的计算结果。

3 实验及结果分析

为了验证本文算法的有效性及其先进性, 这里选取比利时皇家天文台太阳影响数据分析中心提供的 1701—2000 年的平均太阳黑子数来进行周期迭代辨识的相关实验, 样本长度为 300。实验在 PC 机上进行, 其硬件配置为, Intel 酷睿 i5 4570 四核 CPU、Kingmax DDR3 1600 MHz RAM、Western Digital 500 GB Hard Disk; 操作系统与开发环境为, Microsoft Windows 10、Microsoft Visual Studio 2010 集成开发环境中的 C++。在实验过程中, 着重对比主流算法与本文算法的辨识误差和计算开销等技术指标, 并加以详细的分析和讨论。

3.1 实验过程与方法

这里选取有代表性的传统傅里叶谐波法和小波分析法来与本文算法进行对比实验。其中, 传统傅里叶谐波分析法的实验过程为:

- 用式(6)反复地计算出不同基频的傅里叶系数;
- 按照式(19)的方法把幅值最大的基频确定为最优周期;
- 以 b) 中确定的最优周期为基频, 计算其他谐波成分的傅里叶系数。

小波分析法的实验过程则为^[15,16]:

a) 将周期项 P_t 和 $\bar{\psi}(v) = e^{-2\pi^2(v-v_0)^2}$ (Morlet 小波) 代入式(20), 取不同的 a 和 b 数值, 用 Mallat 快速算法计算相应的小波变换系数 $W_p(a, b)$ 。

$$W_p(a, b) = |a|^{-\frac{1}{2}} \Delta t \sum_{k=1}^n P(k\Delta t) \bar{\psi}\left(\frac{k\Delta t - b}{a}\right) \quad (20)$$

b) 根据式(21)计算出各小波变换系数的方差 $\text{var}(a)$, 并找出极大值点对应的的时间尺度 a_o 。

$$\text{var}(a) = \int_{-\infty}^{\infty} |W_p(a, b)|^2 db \quad (21)$$

c) 把 a 置为周期项 P_t 的周期长度。

注意到基频角频率的迭代搜索步长 Δf 对本文算法的辨识精度和计算成本有着重要的影响, 据此, 分别以 $\Delta f = \{0.01, 0.001, 0.0001\}$ 共三种步长对本文算法进行测试。

实验数据的预处理方面, 则先从平均太阳黑子数序列中剔除趋势项 H_t 和随机项 X_t , 并最后得到周期项 P_t ; 同时, 为了观察样本长度 n 对各种算法的性能影响, 在周期项 P_t 中分别取其前 50、100、150、200、250 和 300 项的样本数量对上述算法进行周期辨识实验。

3.2 实验的结果与分析

为了客观正确地评价各种算法的运行性能, 本文以式(22)(23)所示的样本均方误差 MSE 和平均绝对百分误差 MAPE 作为评价指标, 对传统傅里叶谐波法、小波分析法及本文算法的辨识误差进行比较。

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (P_i - \hat{P}_i)^2 \quad (22)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|P_i - \hat{P}_i|}{P_i} \quad (23)$$

各种算法对平均太阳黑子数的周期辨识结果如表 1 和 2

所示。其中,表1为周期长度的辨识数值,而表2则为样本均方误差 MSE 和平均绝对百分误差 MAPE 的具体数值。

表1 各种算法的周期长度的辨识数值 /年

n	传统傅里叶 谐波法	小波分析法	本文算法		
			$\Delta f=0.01$	$\Delta f=0.001$	$\Delta f=0.0001$
50	10	11.410 2	11.32	11.321	11.321 5
100	11.111 1	11.218 0	11.19	11.199	11.199 6
150	10.714 3	11.195 7	11.20	11.202	11.202 4
200	10.526 3	11.201 3	11.21	11.211	11.211 8
250	11.363 6	11.200 6	11.20	11.202	11.202 2
300	10.344 8	11.209 7	11.20	11.201	11.201 6

表2 各种算法的 MSE 和 MAPE 数值

n	误差	传统傅里叶 谐波法	小波 分析法	本文算法		
				$\Delta f=0.01$	$\Delta f=0.001$	$\Delta f=0.0001$
50	MSE	247.85	46.29	26.11	24.24	22.82
	MAPE	30.22%	13.06%	9.81%	9.45%	9.17%
100	MSE	18.13	9.68	7.29	6.08	4.99
	MAPE	7.92%	5.97%	5.18%	4.74%	4.39%
150	MSE	75.06	7.86	6.53	6.06	4.97
	MAPE	16.63%	5.38%	4.92%	4.73%	4.28%
200	MSE	119.91	3.55	6.98	5.99	4.81
	MAPE	21.02%	3.62%	5.07%	4.70%	4.21%
250	MSE	30.11	3.74	4.05	2.65	2.39
	MAPE	10.53%	3.69%	3.86%	3.12%	2.96%
300	MSE	184.60	10.13	4.76	3.83	3.41
	MAPE	26.08%	6.11%	4.19%	3.76%	3.55%

从表1可知,传统傅里叶谐波法虽然能辨识出显著性的周期成分,但其辨识的周期长度受样本长度 n 的变化而在[10, 12]内波动。相对地,小波分析法和本文算法所辨识的周期长度均能稳定在11.2附近。事实上,本文算法在基于方差分析法的分组处理过程中,当把平均太阳黑子数序列分成11组时,各小组数据的均值出现了显著性的差异,于是,在[10, 12]内基于周期迭代辨识算法并以不同的迭代搜索步长 Δf 进行周期成分的辨识,便可得到一系列不同辨识精度的周期长度。

从表2可知,传统傅里叶谐波法的样本均方误差 MSE 和平均绝对百分误差 MAPE 的数值在不同样本长度情况下出现了显著的波动,而小波分析法和本文算法在样本长度 n 大于50以后,其平均绝对百分误差 MAPE 数值均小于8%,是属于高精度的拟合范畴。在样本长度相同的实验过程中,对于本文算法而言,迭代搜索步长 Δf 越少,其对应的样本均方误差 MSE 和平均绝对百分误差 MAPE 的数值也就越少,从而证明了本文算法是收敛的。此外,由于算法自身的原因,小波分析法和本文算法分别在样本长度为200和250时获得了各自的最佳拟合精度,但两者的 MSE 和 MAPE 的数值仍属同一级别的误差水平。各种算法的计算耗时对比如图1所示。

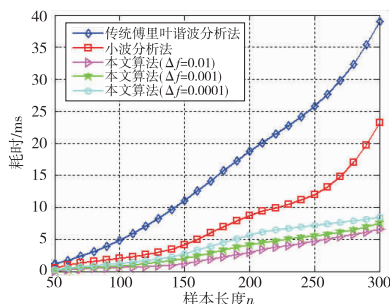


图1 各种算法的计算耗时对比

从图1可知,传统傅里叶谐波分析法和小波分析法均与样

本长度 n 有着明显的线性增长关系,且前者的计算成本较后者高,而本文算法则表现出良好的负载能力。究其原因是,新设计的迭代计算机制能有效地克服样本长度 n 和迭代搜索步长 Δf 所带来的计算开销。

综上所述,传统傅里叶谐波分析法的辨识精度最低,且其计算成本也是最高的;小波分析法具有良好的辨识精度,计算成本虽比传统傅里叶谐波分析法有了一定改善,但仍与样本长度有着线性依赖的关系;而本文算法在辨识精度方面与小波分析法相当,但计算成本则优于上述算法。据此,本文算法是有效和可行的。

4 结束语

时序数据由于混杂了趋势、周期和随机等数据成分,故对它进行精确的周期辨识是一个较为困难的工作。在傅里叶谐波分析法的基础上,提出了一种具有任意搜索步长的时序数据周期迭代辨识算法,由于该算法能以非常低的计算成本进行最优周期成分的精确搜索,故在辨识误差和计算开销方面较现有算法更具有优势。下一步的主要工作是研究融合迭代和递推机制的周期辨识算法,以便进一步提升算法的辨识精度和计算效能。

参考文献:

- [1] 盖哈德·克西盖思纳, 约根·沃特斯, 乌沃·哈斯勒. 现代时间序列分析导论 [M]. 张群群, 刘晓飞, 译. 北京: 中国人民大学出版社, 2015.
- [2] 唐常杰, 于中华, 游志胜, 等. 基于时态数据库的 Web 数据周期规律的采掘 [J]. 计算机学报, 2000, 23(1): 52-59.
- [3] 田端丽, 赵霞, 于重重, 等. 基于凝聚层次聚类的域内交通流周期模式发现 [J]. 计算机应用研究, 2015, 32(8): 2261-2264, 2271.
- [4] 班雷雨, 霍欢, 徐彪. 基于移动数据的异常区域时序分析 [J]. 计算机应用研究, 2017, 34(2): 431-435.
- [5] 程文聪, 邹鹏, 贾焰. 基于 DTW 距离的伪周期数据流异常检测 [J]. 计算机研究与发展, 2010, 47(5): 893-902.
- [6] 王鑫, 吴际, 刘超, 等. 奇异谱分析在故障时间序列分析中的应用 [J]. 北京航空航天大学学报: 自然科学版, 2016, 42(11): 2321-2331.
- [7] 冀振元. 时间序列分析与现代谱估计 [M]. 哈尔滨: 哈尔滨工业大学出版社, 2016.
- [8] Lessard J P, James J D M, Ransford J. Automatic differentiation for Fourier series and the radii polynomial approach [J]. Physica D Nonlinear Phenomena, 2016, 37(3): 174-186.
- [9] Bernas M, Placzek B. Period-aware local modelling and data selection for time series prediction [J]. Expert Systems with Applications, 2016, 27(5): 60-77.
- [10] 蔡智, 岳丽华, 蔡庆生. 时态数据的一种周期函数模式发现研究 [J]. 兰州大学学报: 自然科学版, 1999, 35(8): 117-121.
- [11] 黄雄波. 多周期时序数据的傅氏级数拟合算法 [J]. 计算机系统应用, 2015, 24(7): 142-148.
- [12] 黄雄波. 时序数据的周期模式发现算法的递推改进 [J]. 计算机技术与发展, 2016, 26(2): 47-51.
- [13] 格雷特. 时间序列预测实践教程 [M]. 李洪成, 译. 北京: 清华大学出版社, 2012.
- [14] 李树有, 徐美进, 刘秀娟. 应用数理统计 [M]. 沈阳: 东北大学出版社, 2015.
- [15] 李晓光, 宋宝燕, 于戈, 等. 基于小波的时间序列流伪周期检测方法 [J]. 软件学报, 2010, 21(9): 2161-2172.
- [16] 李晖, 郭晨, 金鸿章. 基于小波变换和均生函数周期外推组合模式的非平稳时间序列分析与长期预测 [J]. 控制理论与应用, 2008, 25(2): 283-288.