

# 多周期时序数据的傅氏级数拟合算法<sup>①</sup>

黄雄波

(佛山职业技术学院 电子信息系, 佛山 528137)

**摘要:** 一般地说, 时序数据通常是由趋势项、随机项及季节周期项三种成分组成的. 通过对已有的时序数据进行分析与建模, 便可以找出事物所蕴含的变化规律. 针对多周期时序数据, 设计实现了一种高精度的数据拟合算法. 该算法首先对被拟合时序数据的趋势成分进行消除, 然后应用自相关函数理论从剩余成分中析出多个两两互质的基本周期, 最后基于最小二乘原理, 用多组傅氏级数对多周期时序数据进行了拟合. 实际应用验证了算法的有效性 & 先进性.

**关键词:** 多周期时序数据; 傅氏级数; 数据拟合; 最小二乘法

## Fourier Series Fitting Method of Multiple Periodic Time-Series Data

HUANG Xiong-Bo

(Department of Electronic and Information Engineering, Foshan Professional Technical College, Foshan 528137, China)

**Abstract:** Generally speaking, time-series data can be divided into three ingredients such as trend, random and season. Analyzing and modeling through the available time-series data, change law can find things contained. A kind of high precision data fitting method was designed for multiple periodic time-series data. Firstly, the algorithm eliminated random component of fitted time-series data. Secondly, several co-primed basic cycles were analyzed by the application of the theory of autocorrelation function. Finally, multiple periodic time-series data was fitting by using multiple sets of Fourier series based on the least squares principle. Practical application proved the effectiveness and progressiveness of the algorithm.

**Key words:** multiple periodic time-series data; fourier series; data-fitting; least square method

所谓时序数据就是指在等时间间隔内所测取到的不间断的一系列数据, 通过对已有的时序数据进行分析与建模, 可以找出事物所蕴含的变化规律, 进而可以对事物的发展进行预测和控制<sup>[1,2]</sup>. 一般地, 时序数据在经过合理的函数变换后总能分解出趋势项、随机项和季节周期项三种数据成分<sup>[3]</sup>, 故其数据建模的过程可以先从原始序列中分离出上述的三种数据成分, 然后运用相应的数学方法分别对这些数据成分进行拟合, 最后对所得的拟合函数进行加性或乘性组合.

在大自然的演变进程中, 事物的发展往往呈现出一定的周期规律特性, 因而研究时序数据的周期特性有着重要的理论和现实意义. 唐常杰等以 Web 数据挖

掘任务为背景, 针对事物在发展过程中的重复趋势和走向现象定义了拟周期概念, 进一步, 设计实现了一种能忽略时间轴上不规则伸缩和幅度上有干扰的拟周期挖掘算法<sup>[4]</sup>. 蔡智等在假设待拟合时序数据周期已知的条件下, 根据傅氏级数的定义及有关定理, 把计算傅氏系数时的积分运算简化为求和运算, 进而得到了一种快捷可行的周期函数模式发现算法<sup>[5]</sup>. 由于文献[4-5]的算法只适用于单周期时序数据, 故本文将运用自相关函数理论从时序数据的季节周期项中析出多个两两互质的基本周期, 在此基础上, 基于最小二乘原理用多组傅氏级数对多周期时序数据进行拟合, 实际应用表明该算法具有良好的拟合精度和收敛速度.

<sup>①</sup> 基金项目: 广东省科技计划工业攻关项目(2011B010200031); 佛山职业技术学院校级重点科研项目(2011KY006)

收稿时间: 2014-11-20; 收到修改稿时间: 2014-12-30

## 1 相关基础知识

### 1.1 时序数据的各种数据成分的拟合处理

对于一般的时序数据  $Y_t$  而言, 由于其趋势项、随机项和季节周期项相互独立, 因而可用式(1)所示的加性模型对其进行描述<sup>[6]</sup>:

$$Y_t = x_t + A_t + B_t \quad (1)$$

式(1)中,  $x_t$  为随机项,  $A_t$  为趋势项, 而  $B_t$  则为季节周期项.

#### 1) 随机项

这里, 我们主要对季节周期项占优的时序数据进行研究, 故其随机项可简要地用式(2)所示的方法进行数据平滑处理<sup>[7]</sup>:

$$\bar{Y}_t = \begin{cases} \frac{y_{t-1} + 2y_t + y_{t+1}}{4}, & 1 < t < n; \\ \frac{3y_1 + y_2}{4}, & t = 1 \\ \frac{y_{n-1} + 3y_n}{4}, & t = n. \end{cases} \quad (2)$$

式(2)中,  $\bar{Y}_t$  是时序数据  $Y_t$  在去除随机项后所剩余的趋势项和季节周期项之和, 而  $y_1, y_2, \dots, y_n$  则为时序数据  $Y_t$  的样本观测值.

#### 2) 趋势项

定义 1. 对于时序数据  $Y_t$  而言, 若

$$A'_t = \frac{\sum_{i=-p}^p i Y_{t+i}}{\sum_{i=-p}^p i^2}, (p \geq 1, p < t < n - p), \text{ 则称 } A'_t \text{ 为 } Y_t$$

的趋势导数.

定义 1 中, 趋势导数  $A'_t$  是以  $t$  为中心, 用前后  $p$  期(共  $2p + 1$  期)  $Y_t$  数据的加权平均值来进行估算的, 在实际拟合中,  $p$  一般取 1 或 2. 进一步, 还可以定义如式(3)所示的  $k$  阶趋势导数:

$$A_t^{(k)} = \frac{\sum_{i=-p}^p i A_{t+i}^{k-1}}{\sum_{i=-p}^p i^2}. \quad (3)$$

由于多项式曲线  $a_0 + a_1 t + \dots + a_m t^m$  的  $m$  阶趋势导数为  $m! a_m$ , 而指数曲线  $ab^t$  的趋势导数与自身

的比等于  $\ln b$ , 故可通过计算并分析时序数据  $Y_t$  的各阶趋势导数, 便可确定趋势拟合曲线的类型.

确定了趋势拟合曲线的类型后, 可基于最小二乘法对相应的拟合参数进行求解, 从而完成趋势项  $A_t$  的数据拟合.

#### 3) 季节周期项

设季节周期项  $B_t$  的周期长度为  $L$ , 且在

$[-\frac{L}{2}, \frac{L}{2}]$  上满足狄氏条件, 则  $B_t$  可由一个三角级数

无限地逼近, 即:

$$B_t = \alpha_0 + \sum_{i=1}^{\infty} \lambda_i \sin(i\omega_0 t + \theta_i) \quad (4)$$

式(4)中,  $\alpha_0$  为常数项,  $\lambda_i$  为各谐波的幅值,  $\omega_0 = 2\pi/L$  为基波角频率,  $\theta_i$  为各谐波的初相位. 根据三角函数的恒等变换公式可得:

$$B_t = \alpha_0 + \sum_{i=1}^{\infty} \{ \alpha_i \cos(i\omega_0 t) + \beta_i \sin(i\omega_0 t) \} \quad (5)$$

式(5)中,  $\alpha_i = \lambda_i \cos(\theta_i)$ ,  $\beta_i = \lambda_i \sin(\theta_i)$ ,

$$\theta_i = \arctan(-\frac{\alpha_i}{\beta_i}).$$

类似地, 可基于最小二乘法对式(5)中的  $\alpha_i$  和  $\beta_i$  等参数进行求解, 从而完成季节周期项  $B_t$  的数据拟合.

### 1.2 自相关函数

季节周期项  $B_t$  的  $k$  阶自相关系数  $\rho_k$  定义为:

$$\rho_k = \frac{Cov(B_t, B_{t+k})}{\sqrt{D(B_t)}\sqrt{D(B_{t+k})}} \quad (6)$$

$\rho_k$  的值表征了季节周期项  $B_t$  的当前项与其后第  $k$  项之间的线性关系的性质及强弱<sup>[8]</sup>. 对于已取得  $n$  个样本观测值的季节周期项  $B_t$  而言,  $\rho_k$  的值可用式(7)中的样本自相关系数  $r_k$  来估计.

$$r_k = \frac{\sum_{j=1}^{n-k} (B_j - \bar{B}') (B_{j+k} - \bar{B}'')}{\sqrt{\sum_{j=1}^{n-k} (B_j - \bar{B}')^2 \sum_{j=1}^{n-k} (B_{j+k} - \bar{B}'')^2}} \quad (7)$$

$$\text{式(7)中, } \bar{B}' = \frac{1}{n-k} \sum_{j=1}^{n-k} B_j, \bar{B}'' = \frac{1}{n-k} \sum_{j=1}^{n-k} B_{j+k}.$$

若一时序数据  $Y_t$  存在周期长度为  $L$  的季节周期项, 则其各个周期内相同位置的数据必然同时大或同

时小, 因而该时序数据的  $L, 2L, \dots, nL$  等各阶的自相关系数将出现正值, 否则, 上述的各阶自相关系数将出现负值, 显然, 利用这一原理便可析出季节周期项的周期长度  $L$ .

定义 2. 若一时序数据  $Y_t$  的季节周期项  $B_t$  是由周期长度为  $L_1, L_2, \dots, L_m (m \geq 2)$  的多个周期数据线性叠加而成, 并且  $L_1, L_2, \dots, L_m$  两两互质, 则称该时序数据  $Y_t$  为多周期时序数据.

多周期时序数据在现实生活中广泛地存在, 以综合性医院的电子病历为例, 由于各种流行病均有不同的发病季节, 且持续的时间长度也不相同, 因而其日(月)门诊量序列便形成了多周期时序数据.

## 2 算法设计

设季节周期项  $B_t$  由  $m$  个两两互质的周期数据线性叠加而成, 它们的周期长度分别为  $L_1, L_2, \dots, L_m$ , 则式(5)可改写为:

$$B_t = \alpha_0 + \sum_{i=1}^n [\alpha_{1i} \cos(i\omega_1 t) + \beta_{1i} \sin(i\omega_1 t)] + \sum_{j=1}^n [\alpha_{2j} \cos(j\omega_2 t) + \beta_{2j} \sin(j\omega_2 t)] + \dots + \sum_{k=1}^n [\alpha_{mk} \cos(k\omega_m t) + \beta_{mk} \sin(k\omega_m t)] \quad (8)$$

式(8)中,  $\omega_1 = 2\pi/L_1, \omega_2 = 2\pi/L_2, \dots, \omega_m = 2\pi/L_m$ . 下面将运用最小二乘法对式(8)中的  $\alpha_0, \alpha_{xy}, x=1, 2, \dots, m; y=1, 2, \dots, \infty$  及  $\beta_{uv}, u=1, 2, \dots, m; v=1, 2, \dots, \infty$  等参数的求解表达式进行推导.

### 2.1 算法的主要推导

设季节周期项  $B_t$  的拟合累计均方误差为  $e$ , 则有:

$$e = \sum_{t=1}^n \left\{ B_t - \left[ \alpha_0 + \sum_{i=1}^n [\alpha_{1i} \cos(i\omega_1 t) + \beta_{1i} \sin(i\omega_1 t)] + \sum_{j=1}^n [\alpha_{2j} \cos(j\omega_2 t) + \beta_{2j} \sin(j\omega_2 t)] + \dots + \sum_{k=1}^n [\alpha_{mk} \cos(k\omega_m t) + \beta_{mk} \sin(k\omega_m t)] \right] \right\}^2 \quad (9)$$

基于最小二乘法, 构造目标函数为:

$$J = \min(e) \quad (10)$$

要使目标函数达到最小, 则有:

$$\begin{cases} \frac{\partial J}{\partial \alpha_0} = 0; \\ \frac{\partial J}{\partial \alpha_{xy}} = 0, x=1, 2, \dots, m; y=1, 2, \dots, \infty; \\ \frac{\partial J}{\partial \beta_{uv}} = 0, u=1, 2, \dots, m; v=1, 2, \dots, \infty. \end{cases} \quad (11)$$

即:

$$\begin{cases} \frac{\partial J}{\partial \alpha_0} = \sum_{t=1}^n \left\{ 2 \left( B_t - \left[ \alpha_0 + \sum_{i=1}^n [\alpha_{1i} \cos(i\omega_1 t) + \beta_{1i} \sin(i\omega_1 t)] + \sum_{j=1}^n [\alpha_{2j} \cos(j\omega_2 t) + \beta_{2j} \sin(j\omega_2 t)] + \dots + \sum_{k=1}^n [\alpha_{mk} \cos(k\omega_m t) + \beta_{mk} \sin(k\omega_m t)] \right] \right) \right\} = 0; \\ \frac{\partial J}{\partial \alpha_{xy}} = \sum_{t=1}^n \left\{ 2 \left( B_t - \left[ \alpha_0 + \sum_{i=1}^n [\alpha_{1i} \cos(i\omega_1 t) + \beta_{1i} \sin(i\omega_1 t)] + \sum_{j=1}^n [\alpha_{2j} \cos(j\omega_2 t) + \beta_{2j} \sin(j\omega_2 t)] + \dots + \sum_{k=1}^n [\alpha_{mk} \cos(k\omega_m t) + \beta_{mk} \sin(k\omega_m t)] \right] \right) \cos(y\omega_x t) \right\} = 0; \\ \frac{\partial J}{\partial \beta_{uv}} = \sum_{t=1}^n \left\{ 2 \left( B_t - \left[ \alpha_0 + \sum_{i=1}^n [\alpha_{1i} \cos(i\omega_1 t) + \beta_{1i} \sin(i\omega_1 t)] + \sum_{j=1}^n [\alpha_{2j} \cos(j\omega_2 t) + \beta_{2j} \sin(j\omega_2 t)] + \dots + \sum_{k=1}^n [\alpha_{mk} \cos(k\omega_m t) + \beta_{mk} \sin(k\omega_m t)] \right] \right) \sin(v\omega_u t) \right\} = 0. \end{cases} \quad (12)$$

把式(12)的后两个分项写成矩阵形式, 有:

$$\begin{bmatrix} \sum_{t=1}^n \cos^2(\omega_1 t) & \sum_{t=1}^n \cos(\omega_1 t) \sin(\omega_1 t) & \dots & \sum_{t=1}^n \cos(\omega_1 t) \cos(\omega_2 t) & \sum_{t=1}^n \cos(\omega_1 t) \sin(\omega_2 t) & \dots & \sum_{t=1}^n \cos(\omega_1 t) \cos(\omega_m t) \\ \sum_{t=1}^n \sin(\omega_1 t) \cos(\omega_1 t) & \sum_{t=1}^n \sin^2(\omega_1 t) & \dots & \sum_{t=1}^n \sin(\omega_1 t) \cos(\omega_2 t) & \sum_{t=1}^n \sin(\omega_1 t) \sin(\omega_2 t) & \dots & \sum_{t=1}^n \sin(\omega_1 t) \cos(\omega_m t) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^n \cos(\omega_2 t) \cos(\omega_1 t) & \sum_{t=1}^n \cos(\omega_2 t) \sin(\omega_1 t) & \dots & \sum_{t=1}^n \cos^2(\omega_2 t) & \sum_{t=1}^n \cos(\omega_2 t) \sin(\omega_2 t) & \dots & \sum_{t=1}^n \cos(\omega_2 t) \cos(\omega_m t) \\ \sum_{t=1}^n \sin(\omega_2 t) \cos(\omega_1 t) & \sum_{t=1}^n \sin(\omega_2 t) \sin(\omega_1 t) & \dots & \sum_{t=1}^n \sin(\omega_2 t) \cos(\omega_2 t) & \sum_{t=1}^n \sin^2(\omega_2 t) & \dots & \sum_{t=1}^n \sin(\omega_2 t) \cos(\omega_m t) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^n \cos(\omega_m t) \cos(\omega_1 t) & \sum_{t=1}^n \cos(\omega_m t) \sin(\omega_1 t) & \dots & \sum_{t=1}^n \cos(\omega_m t) \cos(\omega_2 t) & \sum_{t=1}^n \cos(\omega_m t) \sin(\omega_2 t) & \dots & \sum_{t=1}^n \cos^2(\omega_m t) \\ \sum_{t=1}^n \sin(\omega_m t) \cos(\omega_1 t) & \sum_{t=1}^n \sin(\omega_m t) \sin(\omega_1 t) & \dots & \sum_{t=1}^n \sin(\omega_m t) \cos(\omega_2 t) & \sum_{t=1}^n \sin(\omega_m t) \sin(\omega_2 t) & \dots & \sum_{t=1}^n \sin(\omega_m t) \cos(\omega_m t) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^n \cos(\omega_m t) \cos(\omega_m t) & \sum_{t=1}^n \cos(\omega_m t) \sin(\omega_m t) & \dots & \sum_{t=1}^n \cos(\omega_m t) \cos(\omega_m t) & \sum_{t=1}^n \cos(\omega_m t) \sin(\omega_m t) & \dots & \sum_{t=1}^n \cos^2(\omega_m t) \\ \sum_{t=1}^n \sin(\omega_m t) \cos(\omega_m t) & \sum_{t=1}^n \sin(\omega_m t) \sin(\omega_m t) & \dots & \sum_{t=1}^n \sin(\omega_m t) \cos(\omega_m t) & \sum_{t=1}^n \sin^2(\omega_m t) & \dots & \sum_{t=1}^n \sin(\omega_m t) \cos(\omega_m t) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \beta_{11} \\ \alpha_{11} \\ \beta_{12} \\ \alpha_{12} \\ \vdots \\ \beta_{m1} \\ \alpha_{m1} \\ \beta_{m2} \\ \alpha_{m2} \\ \vdots \\ \beta_{mm} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^n (B_t - \alpha_0) \cos(\omega_1 t) \\ \sum_{t=1}^n (B_t - \alpha_0) \sin(\omega_1 t) \\ \vdots \\ \sum_{t=1}^n (B_t - \alpha_0) \cos(\omega_2 t) \\ \sum_{t=1}^n (B_t - \alpha_0) \sin(\omega_2 t) \\ \vdots \\ \sum_{t=1}^n (B_t - \alpha_0) \cos(\omega_m t) \\ \sum_{t=1}^n (B_t - \alpha_0) \sin(\omega_m t) \\ \vdots \\ \sum_{t=1}^n (B_t - \alpha_0) \cos(\omega_m t) \\ \sum_{t=1}^n (B_t - \alpha_0) \sin(\omega_m t) \end{bmatrix} \quad (13)$$

从文献[8]可知:

$$\begin{cases} \sum_{t=1}^n \hat{a} \sin \omega t = \frac{\sin \frac{1}{2} m \omega \sin \frac{1}{2} (n+1) \omega}{\sin \frac{\omega}{2}}; \\ \sum_{t=1}^n \hat{a} \cos \omega t = \frac{\sin \frac{1}{2} m \omega \cos \frac{1}{2} (n+1) \omega}{\sin \frac{\omega}{2}}. \end{cases} \quad (14)$$

故可以得到:

$$\sum_{t=1}^n \sin x \omega t \cos y \omega t = 0, x, y = 0, 1, 2, \dots \quad (15)$$

$$\sum_{t=1}^n \sin x \omega t \cos y \omega t = \begin{cases} 0, \forall ((x \neq y) \vee (x = y = 0, \frac{n}{2})); \\ \frac{n}{2}, \forall (x = y \neq 0, \frac{n}{2}). \end{cases} \quad (16)$$

$$\sum_{t=1}^n \cos x \omega t \cos y \omega t = \begin{cases} 0, \forall (x \neq y); \\ n, \forall (x = y = 0, \frac{n}{2}); \\ \frac{n}{2}, \forall (x = y \neq 0, \frac{n}{2}). \end{cases} \quad (17)$$

利用式(15)~(17)对式(12)第 1 分项及式(13)进行化简, 分别有:

$$\alpha_0 = \frac{1}{n} \sum_{t=1}^n B_t. \quad (18)$$

和

$$\begin{aligned} & \sum_{t=1}^n \hat{a} (B_t - a_0) \cos(\omega_1 t) \\ & \sum_{t=1}^n \hat{a} (B_t - a_0) \sin(\omega_1 t) \\ & \quad \quad \quad M \\ & \sum_{t=1}^n \hat{a} (B_t - a_0) \cos(\omega_2 t) \\ & \sum_{t=1}^n \hat{a} (B_t - a_0) \sin(\omega_2 t) \\ & \quad \quad \quad M \\ & \quad \quad \quad M \\ & \quad \quad \quad M \\ & \sum_{t=1}^n \hat{a} (B_t - a_0) \cos(\omega_m t) \\ & \sum_{t=1}^n \hat{a} (B_t - a_0) \sin(\omega_m t) \\ & \quad \quad \quad M \\ & \sum_{t=1}^n \hat{a} (B_t - a_0) \cos(\omega_{m+1} t) \\ & \sum_{t=1}^n \hat{a} (B_t - a_0) \sin(\omega_{m+1} t) \end{aligned} \quad (19)$$

## 2.2 算法设计

综上所述, 可设计如下的多周期时序数据的傅氏级数拟合算法.

算法名称: 多周期时序数据的傅氏级数拟合算法

输入: 时序数据  $Y_t$  及平均绝对百分比误差  $MAPE$ ;

输出: 各周期数据成分的周期长度  $L_1, L_2, \dots, L_m$  及它们的傅氏拟合参数  $\alpha_{xy}, \beta_{uv}$ .

步骤 1: 按照式(2)对时序数据  $Y_t$  进行数据平滑处理, 并得到  $\bar{Y}_t$ .

步骤 2: 应用文献[9-10]的算法, 析出并消除  $\bar{Y}_t$  中的趋势项  $A_t$ , 并得到季节周期项  $B_t$ .

步骤 3: 检测周期数据成分  $L_1, L_2, \dots, L_m$ .

$p = 0$ ;  
for( $L = 2; L \leq \text{int}(2/n); L++$ )

```
{
    for( $q = 1; (k = q * L) \leq n; q++$ )
    {
         $\bar{B} = \frac{1}{n-k} \sum_{j=1}^{n-k} B_j; \bar{B}' = \frac{1}{n-k} \sum_{j=1}^{n-k} B_{j+k};$ 
         $r_k = \frac{\sum_{j=1}^{n-k} (B_j - \bar{B})(B_{j+k} - \bar{B}')}{\sqrt{\sum_{j=1}^{n-k} (B_j - \bar{B})^2 \sum_{j=1}^{n-k} (B_{j+k} - \bar{B}')^2}}$ ;
        if( $r_k < 0$ ) break;
    }
    if( $k \geq n$ )
    {
        period[ $p$ ] =  $L$ ; // 记录有效周期
        ++ $p$ ;
    }
}
```

步骤 4: 计算各个周期数据成分的傅氏拟合参数  $\alpha_{xy}, \beta_{uv}$ .

$\omega_1 = 2\pi / L_1; \omega_2 = 2\pi / L_2; \dots; \omega_m = 2\pi / L_m;$

$i = 1; B_i = 0;$

$\alpha_0 = \frac{1}{n} \sum_{t=1}^n B_t;$

do

{

$\alpha_{1i} = \frac{2}{n} \sum_{t=1}^n B_t \cos(i\omega_1 t); \beta_{1i} = \frac{2}{n} \sum_{t=1}^n B_t \sin(i\omega_1 t);$

$\alpha_{2i} = \frac{2}{n} \sum_{t=1}^n B_t \cos(i\omega_2 t); \beta_{2i} = \frac{2}{n} \sum_{t=1}^n B_t \sin(i\omega_2 t);$

.....;

$$\alpha_{mi} = \frac{2}{n} \sum_{t=1}^n B_t \cos(i\omega_m t); \beta_{mi} = \frac{2}{n} \sum_{t=1}^n B_t \sin(i\omega_m t);$$

$$B'_i = B'_i + \left\{ \sum_{i=1}^{\infty} [\alpha_{1i} \cos(i\omega_1 t) + \beta_{1i} \sin(i\omega_1 t)] + \sum_{j=1}^{\infty} [\alpha_{2j} \cos(i\omega_2 t) + \beta_{2j} \sin(i\omega_2 t) + \dots + \sum_{i=1}^{\infty} [\alpha_{mi} \cos(i\omega_m t) + \beta_{mi} \sin(i\omega_m t)]] \right\}$$

$i++;$   
 $\}while((B_i - B'_i) > e);$

### 3 实验及结果分析

为了验证上述算法的有效性及其先进性, 这里将对某三甲医院近 3 年的月门诊量序列进行多周期傅氏级数拟合的实验, 并对相关实验结果进行分析.

#### 3.1 多周期时序数据的拟合实验

1) 对表 1 的月门诊量数据进行数据平滑处理, 得到如表 2 所示的结果, 数据平滑处理的前后曲线如图 1.

表 1 某三甲医院近 3 年的月门诊量序列  $Y_t$  (单位: 人)

201101	201102	201103	201104	201105	201106	201107	201108	201109	201110	201111	201112
163929	156356	163003	160952	166016	160341	162281	162603	164610	161755	162459	163086
201201	201202	201203	201204	201205	201206	201207	201208	201209	201210	201211	201212
168743	165627	161845	165394	169236	166651	172152	169528	171790	169380	170135	174893
201301	201302	201303	201304	201305	201306	201307	201308	201309	201310	201311	201312
175960	177530	178794	173169	179697	172171	176080	179722	176380	182652	184884	183018

表 2 表 1 进行数据平滑处理后的得到的  $\bar{Y}_t$  (单位: 人)

201101	201102	201103	201104	201105	201106	201107	201108	201109	201110	201111	201112
162035	159911	160828	162730	163331	162244	161876	163024	163394	162644	162439	164343
201201	201202	201203	201204	201205	201206	201207	201208	201209	201210	201211	201212
166549	165460	163677	165467	167629	168672	170120	170749	170622	170171	171135	173970
201301	201302	201303	201304	201305	201306	201307	201308	201309	201310	201311	201312
176085	177453	177071	176207	176183	175029	176013	177976	178783	181642	183859	183484

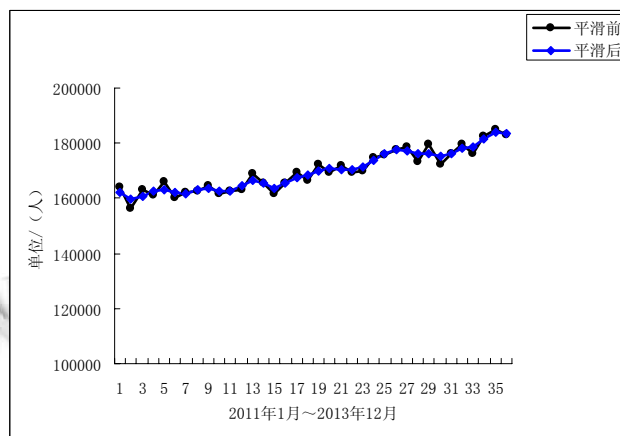


图 1 平滑处理前后的某三甲医院近 3 年的月门诊量序列

2) 对  $\bar{Y}_t$  进行趋势拟合, 得到式(20)所示的拟合函数  $A_t$ , 在此基础上, 用  $\bar{Y}_t - A_t$  析出如表 3 所示的季节

周期项  $B_t$ .

$$A_t = 0.014t^5 - 1.272t^4 + 41t^3 - 547.25t^2 + 3194t + 153802 \quad (20)$$

表 3 季节周期项  $B_t$  (单位: 人)

201101	201102	201103	201104	201105	201106	201107	201108	201109	201110	201111	201112
5138	2594	1194	1024	3902	978	2610	2570	4164	850	1054	1144

201201	201202	201203	201204	201205	201206	201207	201208	201209	201210	201211	201212
5730	2514	1108	1012	4194	936	2758	2450	4030	944	1032	1136
201301	201302	201303	201304	201305	201306	201307	201308	201309	201310	201311	201312
5568	2526	1204	1024	4032	1026	2498	2752	4074	1066	1080	1060

3) 计算季节周期项  $B_t$  的自相关系数  $r_k$ , 相关计算结果如表 4 所示, 从表 4 易知, 该季节周期项  $B_t$  含有  $L = 4$  和  $L = 6$  两种周期成分的数据.

表 4 季节周期项  $B_t$  的各阶自相关系数  $r_k$  的值

$r_2$	$r_3$	$r_4$	$r_5$	$r_6$	$r_7$	$r_8$	$r_9$	$r_{10}$
-0.447	-0.409	0.627	-0.279	0.191	-0.139	0.669	-0.404	-0.359
$r_{11}$	$r_{12}$	$r_{13}$	$r_{14}$	$r_{15}$	$r_{16}$	$r_{17}$	$r_{18}$	—
0.254	0.981	-0.155	-0.276	-0.405	0.614	-0.317	0.148	—

4) 在平均绝对百分比误差  $MAPE \leq 5\%$  的计算精度下求解  $L = 4$  和  $L = 6$  两种周期成分的傅氏拟合参数, 结果分别如式 (21) 和式 (22) 所示, 而  $B_t, B_{(L=4)t}, B_{(L=6)t}$  所对应的曲线见图 2 所示.

$$B_{(L=4)t} = 1416.556 + 58.806 \cos\left(\frac{2\pi}{4}t\right) + 1453.368 \sin\left(\frac{2\pi}{4}t\right) - 741.622 \cos\left(2 \cdot \frac{2\pi}{4}t\right) + 43.645 \sin\left(2 \cdot \frac{2\pi}{4}t\right) + 0.379 \cos\left(3 \cdot \frac{2\pi}{4}t\right) + 0.163 \sin\left(3 \cdot \frac{2\pi}{4}t\right). \quad (21)$$

$$B_{(L=6)t} = 888.333 + 5.84 \cos\left(\frac{2\pi}{6}t\right) + 895.777 \sin\left(\frac{2\pi}{6}t\right) - 479.356 \cos\left(2 \cdot \frac{2\pi}{6}t\right) + 6.743 \sin\left(2 \cdot \frac{2\pi}{6}t\right) - 0.019 \cos\left(3 \cdot \frac{2\pi}{6}t\right) - 0.501 \sin\left(3 \cdot \frac{2\pi}{6}t\right). \quad (22)$$

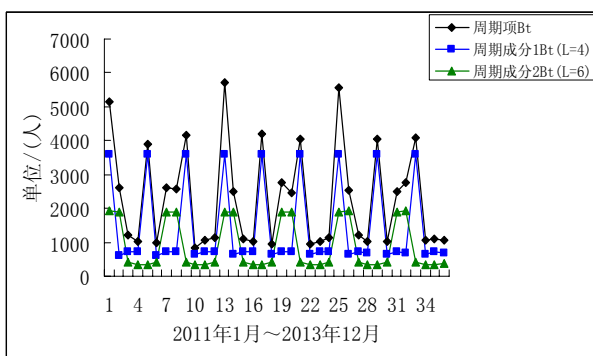


图 2 周期项及两种周期成分的傅氏拟合效果图

### 3.2 实验结果分析

从图 2 可知, 该三甲医院近 3 年的月门诊量序列的周期项  $B_t$  可分解出长度  $L = 4$  和  $L = 6$  两种周期成分数据. 其中, 长度  $L = 4$  的周期成分数据在 1 月、5

月及 9 月均产生了门诊峰值, 经调查后得知, 这是由于季节转换引发了流感、腹泻、呼吸道感染等患者比平时剧增所导致的; 而  $L = 6$  的周期成分数据在 1 月和 2 月、7 月和 8 月也产生了门诊峰值, 这是由于该医院是骨科专科医院, 上述时间段适逢寒暑假, 学龄儿童因意外受伤就医的情况也比平时倍增.

用式(21)和式(22)之和对表 3 的周期项  $B_t$  进行拟合, 其绝对误差函数  $e_t = |B_t - B_{(L=4)t} - B_{(L=6)t}|$ , 所对应的曲线如图 3 所示.

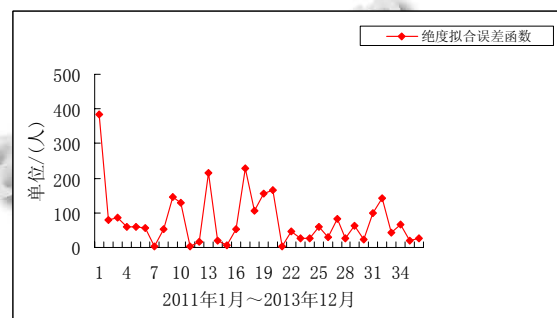


图 3 季节周期项  $B_t$  的绝对拟合误差函数  $e_t$  的示意图

而平均绝对百分比误差 MAPE 为:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{B_t} = \frac{1}{36} \sum_{t=1}^{36} \frac{|B_t - B_{(L=4)t} - B_{(L=6)t}|}{B_t} \approx 3.854\% \quad (23)$$

从上述实验结果易知, 本文所设计的算法具有较高的拟合精度, 且有较快的拟合收敛速度.

### 4 结语

设计实现了一种多周期时序数据的傅氏级数拟合

算法, 实验结果验证了新算法的有效性和先进性. 下一步的主要工作有, 引入模糊聚类分析方法进行更为合理的周期发现、研究更为有效的傅氏级数的系数求解方法, 以便进一步提升拟合算法的拟合精度及计算效能.

#### 参考文献

- 1 Shmueli G. 时间序列预测实践教程. 北京: 清华大学出版社, 2012.
- 2 Box GEP, Jenkins GM. 时间序列分析: 预测与控制. 北京: 机械工业出版社, 2011.
- 3 王燕. 应用时间序列分析(第 3 版). 北京: 中国人民大学出版社, 2012.
- 4 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. 计算机学报, 2000, 23(1): 52-59.
- 5 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. 兰州大学学报(自然科学版), 1999, 35(8): 117-121.
- 6 Min GN, Yu CC, Kai L, et al. Seasonal and trend time series forecasting based on a quasi-linear autoregressive model. Applied Soft Computing, 2014, 24(10): 13-18.
- 7 宋少鹏, 杨英杰, 汪永伟. 基于周期规律的加权时态关联规则挖掘算法. 计算机工程, 2013, 39(3): 41-45.
- 8 侯文超. 经济预测——理论、方法及应用. 北京: 商务印书馆, 1993.
- 9 李光辉, 叶绪国. 基于 Legendre 多项式的分段拟合. 统计与决策, 2013, 30(18): 28-31.
- 10 黄雄波. 时序数据趋势项的分段拟合. 计算机系统应用, 2015, 24(2): 174-179.