第 27 卷　第 3 期　华侨大学学报(自然科学版)　Vol. 27　No. 3
2006 年 7 月　Journal of Huaqiao University (Natural Science)　Jul. 2006

文章编号　1000-5013(2006)03-0313-04

Fabonacci 法应用于曲线拟合优化新算法

汪　岚　　金福江

(华侨大学信息科学与工程学院, 福建 泉州 362021)

摘要　由最小二乘法得到的曲线拟合的效果是较好的,但却不一定是最优的. 文中提出将最小二乘法与 Fabonacci 法结合,得到一种曲线拟合的新优化算法. 该算法的思路是在最小二乘法的基础之上,结合 Fabonacci 法优化已得到的拟合曲线的方程系数,使其更加合理化,从而获得最佳拟合曲线. 该方法适用优化区间为单峰函数的任何数据,实例表明方法是切实可行的.

关键词　曲线拟合, 最小二乘法, Fabonacci 优化法, Matlab

中图分类号　TB 114.1　　**文献标识码**　A

通常情况下,曲线拟合的性能通过计算近似函数值和给定数据集中数值的差的平方和来度量. 根据最小二乘法,当平方和最小时,曲线拟合的效果最好,但其拟合出的曲线方程式的系数虽然接近最优,却不一定是最优值. 文[1,2]提出采用 0.618 法,对拟合曲线的方程式系数进行优化,所得到的最终区间较大,若是在精度高的情况下,难以较准确地得到最佳值. 因此,本文提出一种新的优化拟合方法,即在最小二乘法曲线拟合的基础之上,采用 Fabonacci 优化法对已得到的拟合曲线方程式的系数进行优化. 最终得到的区间较 0.618 法缩短了 15 %,能够较准确地获得最佳值.

1　模型的建立

1.1　最小二乘法分段拟合

例如,一组数据(x_i, y_i), $(i = 1, 2, \cdots, n)$而各 x_i 各不相同的,且设 $y_i = f(x)$. 根据其散点图发现前 m 个点较接近 p 阶曲线,后 $n - m$ 个点呈现出 q 阶的非线性关系,则可将其分两段拟合. 分别用 p 次多项式 y_1 和 q 次多项式 y_2 进行拟合,即

$$\left.\begin{aligned} y_1 &= a_0 + a_1 x + a_2 x^2 + \cdots + a_p x^p, \\ y_2 &= b_0 + b_1 x + b_2 x^2 + \cdots + b_p x^p. \end{aligned}\right\} \tag{1}$$

在式(1)中, $p > q$. 同时,要保证在边界点(x_m, y_m)连续光滑,则必须满足边界约束条件

$$y_{1m} = y_{2m}, \qquad y'_{1m} = y'_{2m}, \qquad y''_{1m} = y''_{2m}, \qquad \cdots . \tag{2}$$

由此可知,式(1)中两个方程的系数是相关的,系数 a_i $(i = 1, 2, \cdots, p)$可由 b_i $(i = 1, 2, \cdots, q)$表示出来. 假设一曲线 $S_n(x)$逼近 $y_i = f(x)$,由于观测必有误差,因此 $\delta_i = S_n(x_i) - y_i$ 并不为零,则最小平方误差 S 为

$$\sum_{i=1}^{n} [S_n(x) - y_i]^2. \tag{3}$$

根据曲线拟合的最小二乘法要求,有

$$S = \sum_{i=1}^{m} [\delta(x)^2] = \min \sum_{i=1}^{m} [S_n(x) - y_i]^2,$$

收稿日期　2005-12-03

作者简介　汪　岚(1978-),女,硕士研究生,主要从事智能建模的研究; 通信作者:金福江(1965-),男,副教授,E-mail:fjjin@yahoo.com.cn

基金项目　福建省自然科学基金资助项目(A0504002)

则可以通过列出$\frac{\partial S}{\partial b_i}=0$的方程,求出在最小方差 S 的条件下 b_i 的值,进而求出 a_i. 最终,可以确定 y_1 和 y_2 的系数,并在图中作出拟合曲线. 通过一个决定系数

$$R^2=\frac{\sum(y_t-\bar{y}_i)^2}{\sum(y_i-\bar{y}_i)^2},\tag{4}$$

判断出给定数据的拟合曲线的拟合质量. 式中,y_t 为已知数据点处拟合方程 y 的值,即

$$y_t=c_nx^n+\cdots+c_2x^2+c_1x+c_0,$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i.$$

这个 R^2 的值可以清楚的表达拟合质量. 当 R^2 值越接近 1 时,表示拟合质量越高.

1.2　Fabonacci 法优化系数

用最小二乘法拟合出来的效果是较好的,但却不一定是最佳的. 因此,可在最小二乘法拟合的基础之上,采用 Fabonacci 法[3]进一步的优化已求出的各项系数,使其更加接近最优值. 以 $y=c_nx^n+\cdots+c_1x+c_0$ 为例,首先要求出 c_n 的取值区间(初始区间),然后采用 Fabonacci 法对其取值区间进行优化,确定出最佳值. 在此最佳值条件下,求出 c_{n-1} 的最佳值. 按此方法可逐一求出其他系数的最佳值.

1.3　c_m 的初始区间$[a_1,b_1]$的算法

设第 j 个离散点与拟合曲线上对应值的偏差为

$$\Delta y_i=y_i-y_T=y_j-\sum_{i=1}^{n}c_ix_i^i-c_0,\qquad j=1,2,\cdots,n.\tag{5}$$

根据式(5)可以求出,n 个离散点中的最大正偏差点$(x_{+\Delta,\max},y_{+\Delta,\max})$和最大负偏差点$(x_{-\Delta,\max},y_{-\Delta,\max})$,则第 m 次方系数 c_m 的初始区间$[a_1,b_1]$为

$$\left.\begin{aligned}a_1(c_m)&=\frac{y_{-\Delta,\max}-\sum\limits_{i=1,i\neq m}^{n}c_ix_i}{(x_{-\Delta,\max})^m},\\ b_1(c_m)&=\frac{y_{+\Delta,\max}-\sum\limits_{i=1,i\neq m}^{n}c_ix_i}{(x_{+\Delta,\max})^m}.\end{aligned}\right\}\tag{6}$$

1.4　最佳系数

Fabonacci 法适应于单峰函数,因此必须先判断 c_i 在初始区间$[a_1,b_1]$是单峰函数. 根据单峰函数的定义[3]知道,求出的最佳 c_T 就是 c_i 在区间$[a_1,b_1]$中的近似极小(大)值,即 c_i 在$[a_1,c_T]$区间应严格递减(增),在$[c_T,b_1]$上应严格递增(减). 判断 c_i 为单峰函数后,采用 Fabonacci 法对其进行优化. Fabonacci 法计算步骤参见文[4].

1.5　常系数 c_0 的算法

根据最佳拟合曲线的最大正、负偏差的绝对值近似相等,则偏差取两者之和的一半,即

$$\Delta_{\max}=y_j-y_T=y_i-\sum_{i=1}^{n}c_ix_i-c_0,\qquad j=1,2,\cdots,n.$$

由此可得最佳常系数为

$$c_{0,T}=y_j-\sum_{i=1}^{n}c_ix_i-\Delta_{\max}=y_j-\sum_{i=1}^{n}c_ix_i-\frac{1}{2}[|\Delta_{+\max}|+|\Delta_{-\max}|],\qquad j=1,2,\cdots,n.\tag{7}$$

2　建立最佳拟合曲线方程的实例

实例为福建晋江某包装厂五层瓦楞纸机,宽幅为 1.8 m,最高车速为 200 m·s⁻¹的 B 机的进纸速度(v)进行曲线拟合. 拟合的离散点如图 1 所示.

2.1　最小二乘法曲线拟合

由图 1 可以看出,前 7 个点较接近直线,而后 7 个点则呈现出非线性关系这个特点,故采用将这个区间分为两段进行拟合. 分别以一阶多项式 y_1 和二阶多项式 y_2 进行拟合,根据前面所介绍的方法,按照式(1)设 $y_1=kx+b$, $y_2=c_2x^2+c_1x+c_0$, $x_m=37$. 同时,还要保证在边界点(x_m,y_m)连续光滑,则必须

满足边界约束条件 $y_{1m} = y_{2m}$, $y'_{1m} = y'_{2m}$. 即 $kx_m + b = c_2 x_m^2 + c_1 x_m + c_0$, $k = 2c_2 x_m + c_1$. 则假设条件转化为 $y_1 = (2c_2 x_m + c_1) x + (c_0 - c_2 x_m^2)$, $y_2 = c_2 x^2 + c_1 x + c_0$. 最小平方误差 S 为

$$S = \sum_{i=1}^{7} [2c_2 x_m + c_1) x + c_0 - c_2 x_m^2 - y_i]^2 + \sum_{i=8}^{14} [c_2 x^2 + c_1 x + c_0 - y_i]^2.$$

通过求 $\frac{\partial S}{\partial C_i} = 0$, 可以求出 $c_2 = -1.292\,4$, $c_1 = 131.018\,9$, $c_0 = -3\,224.8$, 进而可以求出 $k = -3.390\,4$ 和 $b = 269.856\,0$, 最后求出两段的拟合曲线分别为 $y_1 = -3.390\,4x_1 + 269.856\,0$ 和 $y_2 = -1.292\,4x_2^2 + 131.018\,9x_2 - 3\,224.8$. 拟合效果如图 1 所示. 可以看出拟合效果还是较好的. 最后, 根据已求出的两段拟合曲线可求出 $R^2 = 0.990\,2$.

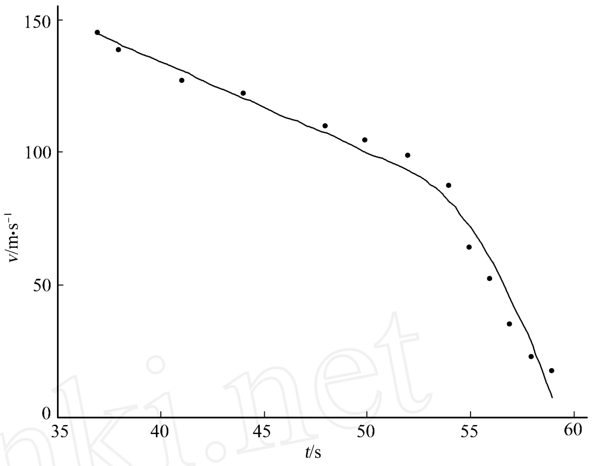


图 1　B 机[37,59]区进纸线速度离散点和拟合曲线图

2.2　采用 Fabonacci 法优化拟合曲线方程系数

在采用分段最小二乘法拟合出曲线方程后,用优化法优化已求出的拟合曲线方程式的系数. 前面已求出拟合直线为 $y_1 = -3.390\,4x + 269.856\,0$, 则拟合直线数据如表 2 所示. 表 2 中的 k_i 和 y_i 的离散点,如图 2 所示. 由表 2 可看出, [52, 98.43] 和 [41, 127.38] 两点数据分别为最大正误差和最大负误差.

表 2　拟合直线数据表

x_i	y_i	y_t	$y_i - y_t$	$k_i = (y_i - b)/x_i$
37	144.75	144.411 2	0.388 8	-3.381 2
38	138.96	141.020 8	-2.060 8	-3.444 6
41	127.38	130.849 6	-3.469 6	-3.475 0 (a_1)
44	121.59	120.678 4	0.911 6	-3.369 7
48	110.01	107.116 8	2.893 2	-3.330 1
50	104.22	100.336 0	3.884 0	-3.312 7
52	98.43	93.555 2	4.874 8	-3.296 7 (b_1)

它们分别对应的斜率为 -3.296 7 和 -3.475 0, 则斜率 k 的初始范围为 [-3.475 5, -3.296 7]. 令 $a_1 = -3.475\,0$, $b_1 = -3.296\,7$, 由图 2 可看出, 斜率 k 在其优化区间为单峰函数.

2.3　计算函数值的迭代次数 n

令最终区间长度 $L = 0.001$ 时, 以及已确定的斜率 k 的初始范围为 [-3.475 0, -3.296 7], 可得 $F_n \geq \frac{b_1 - a_1}{L} = \frac{-3.692\,7 - (-3.475\,0)}{0.001} = 178.3$. 查 Fabonacci 数列表, 得 $n = 12$, 即迭代次数为 12 次.

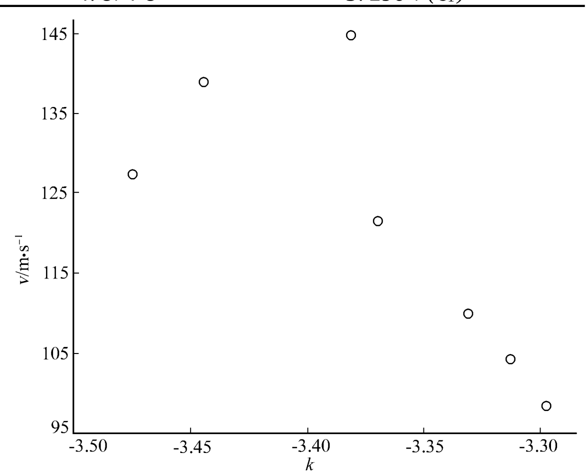


图 2　表 2 中 k_i 和 y_i 的离散点

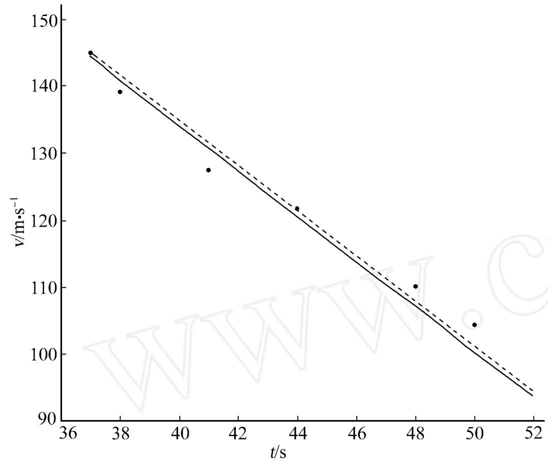
2.4　最佳斜率 k_T

前面已计算出斜率优化范围为 [-3.475 0, -3.269 7], 根据 Fabonacci 优化算法步骤, 最终求出 k 的优化区间为 [-3.375 9, -3.375 5], 取 k_T 的最佳值为 -3.375 7.

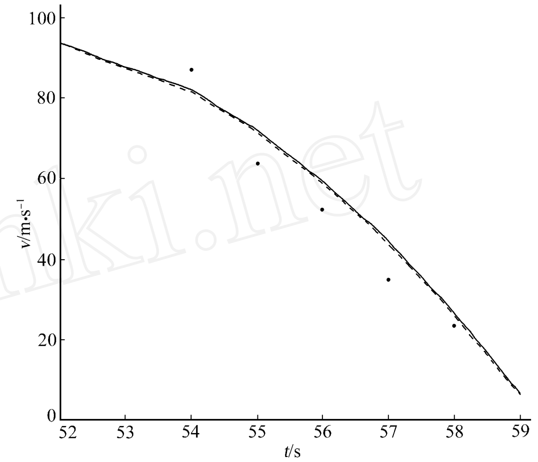
2.5　最佳截距 b_T

根据式(7)可得, 最佳曲线拟合方程为 $y_1 = -3.375\,7x + 269.875$. 同理, 对于 y_2 曲线方程只对的二次系数 c_2 进行优化. 在等精度 $L = 0.001$ 条件下, 迭代次数仅需 6 次, 得到最佳值 $c_{2T} = -1.292\,6$, 则 $y_2 = -1.292\,6x_2^2 + 131.018\,9x_2 - 3\,224.8$.

由于分段曲线方程的系数是相关的,因此在进行分段优化时,势必会导致边界点的不连续.但是,若保证分段与整体的拟合质量都提高的情况下,边界点造成的曲线不光滑影响可不考虑[5].可计算出当只优化 y_1 决定系数 R^2 的值由原来的0.9902增加到了0.9964,再优化 y_2 中的 c_2, R^2 增加到了0.9971,整体拟合质量提高.同时,可发现在精度不变的情况下,迭代次数较少表明拟合的曲线方程的系数已经很逼近最佳值了.因此,可通过计算迭代次数有选择的对次数较多的系数进行优化,效果会更明显.图3



(a) B机[37,52]区间



(b) B机[52,59]区间

图3 B机进纸速度的两种优化方法曲线比较图

中实线为最小二乘拟合曲线,虚线为优化后的拟合曲线,由图可看出优化后拟合质量提高.

3 结束语

本文在原有的最小二乘法的基础之上,提出对数据进行拟合后再采用Fabonacci优化法对已有的拟合曲线方程式的系数进行优化,从而提高曲线的拟合质量的曲线拟合新方法.对于拟合曲线方程式系数,可根据实际情况所要求的精度计算其迭代次数,有选择优化其中的几个或是全部系数,从而提高整体的拟合质量.通过应用实例,证明该方法是有效可行的.

参 考 文 献

1 薛亚琴.电涡流传感器特性曲线拟合的新方法[J].传感器技术,2003,22(7):42～44
2 陈宝林.最优化理论与算法[M].北京:清华大学出版社,2003.307～311
3 Mathews J H, Kurtis D F.数值分析[M].周 璐,等译.北京:电子工业出版社,2004.254～259
4 袁亚湘.最优化理论与方法[M].北京:科学出版社,1999.73～75
5 金福江.制浆生产蒸煮过程多目标优化控制[J].华侨大学学报(自然科学版),2004,25(4):422～427

Research and Application of Optimum Curve-Fitting Method Based on Fabonacci Algorithm

Wang Lan Jin Fujiang
(College of Information Science and Engineering, Huaqiao University, 362021, Quanzhou, China)

Abstract According to classic experiment, using the least-square, law, better fitting-curve can be obtained, but it isn't optimum. A new algorithm based on least-square law combined with Fabonacci optimum was proposed. To optimize the constant of imitative equation, it shows that this algorithm becomes more reasonable. Using this new algorithm, optimum fitting-curve can be obtained. This algorithm can be applied to all data of single-peak function and its applications are introduced. The feasibility and validity of this new algorithm have been verified by real examples.

Keywords curve-fitting, least-square law, Fabonacci algorithm, matlab