基于改进进化策略的线性参数估计方法*

Linear Parameter Estimation Method Based on Improved Evolutionary Strategy

郭德龙[1],夏慧明[2],周永权[3]
GUO De-long[1],XIA Hui-ming[2],ZHOU Yong-quan[3]

(1.黔南民族师范学院数学系,贵州都匀 558000;2.南京师范大学泰州学院数学系,江苏泰州 225300;3.广西民族大学数学与计算机科学学院,广西南宁 530006)

(1. Department of Maths Qiannan Normal University for Nationalities, Duyun, Guizhou, 558000,China;2. Department Math Taizhou College of Nanjing Normal University, Taizhou, Jiangsu,225300, China;3. College of Math and Computer Science,Guangxi University for Nationalities,Nanning,Guangxi,530006,China)

摘要:针对线性回归模型,在最小二乘意义下,提出一种基于改进进化策略的线性参数估计方法并用仿真实例检验其有效性.该方法是将进化策略与最小二乘法相结合并且对算法中的变异算子作了相应改进.算法参数估计精度较高,收敛速度快,自适应性强,能有效处理线性回归参数估计问题.

关键词:参数估计　进化策略　回归分析　最小二乘法

中图法分类号:O212,TP181　　**文献标识码**:A　　**文章编号**:1005-9164(2010)04-0318-03

Abstract:In terms of the least squares in linear regression modeling,a linear parameter estimation method is proposed basing on improved evolutionary strategy and its effectiveness is tested with simulation examples. The method combines evolutionary stratagy with the least square method,and the mutation operators of the algorithm are improved corespondingly. The parameter estimation algorithm has high precision,fast convergence speed,adaptabilily and can deal effectively with the problem of linear regression parameter estimation.

Key words:parameter estimation, evolution strategy,regression analysis,least squares method

回归分析是实验数据处理中使用最为普遍的数学方法之一.物理学、化学、地质学和地理学等学科大量实验中的数据分析与计算都可能应用到线性回归分析方法.我们根据机理分析各种模型的结构时,有时需要对模型中的参数进行估计.线性参数估计是参数估计中的重要内容之一,传统的参数估计方法有直接搜索法[1]、梯度法[1]、牛顿法[2]等.线性回归分析中参数估计使用方法最多的是最小二乘法[3].由于数学模型的建立与求解都是在已有历史数据的前提下进行的,预测精度不仅取决于统计数据的准确性,还取决于模型随新增信息的动态性.也就是说作为预测的数学模型,必须随着新信息的不断涌现,及时修正模型中的参数,否则将难以保证预测的精度.本文提出一种新的改进进化策略[4]的线性参数估计方法,并用仿真实例来检验其有效性.

1　线性回归模型

针对线性回归模型,在最小二乘意义下应用改进的进化策略来进行参数估计.线性回规模型一般可以表示为

$$Y=f(X,\theta)+e\ , \tag{1}$$

其中 e 服从正态分布 $N(0,\sigma)$, $X\in R^m$, $Y\in R$, $\theta\in R^p$, P 为参数个数.

设已知观测值为 $\{(x_i,y_i),i=1,2,\cdots,n\}$,线性回归模型参数估计问题就是如何求出 θ 的最小二乘估计值,即求 $\hat{\theta}\in R^p$ 使得对任何 $\theta\in R^p$ 都有 $s(\hat{\theta})\leqslant s(\theta)$,

收稿日期:2009-12-24
修回日期:2010-03-17
作者简介:郭德龙(1976-),男,硕士,讲师,主要从事于进化计算及应用研究。
*国家民委科研基金项目(08GX01),广西自然科学基金项目(0832082),黔南民族师范学院院级科研项目(QNSY0905)资助。

$$s(\theta)=\sum_{i=1}^{n}(y_i-f(x_i,\theta))^2. \quad (2)$$

有关最小二乘估计量的存在性，Jennrich[5]给出了如下定理.

定理1 假设Θ为R^p上的紧子集，$f(X,\theta)$关于θ在Θ上连续，则必存在R^p上的可测函数$\theta=\hat{\theta}(Y)$使得

$$\| Y-f(X,\hat{\theta}(Y)) \|^2=\inf_{\theta\in\Theta}\| Y-f(X,\theta)\|^2. \quad (3)$$

2 改进进化策略的线性参数估计方法

(1)确定问题的表达方式. 这种表达式中个体由目标变量X和标准差两部分组成，每部分又可以有n个分量，即

$$(\theta,\sigma)=((\theta_1,\theta_2,\cdots,\theta_i,\cdots,\theta_n),(\sigma_1,\sigma_2,\cdots,\sigma_i,\cdots,\sigma_n)), \quad (4)$$

θ和σ之间的关系是

$$\begin{cases}\sigma_i'=\sigma_i\exp(r'\cdot N(0,1)+r\cdot N_i(0,1)),\\ \theta_i'=\theta_i+\sigma_i\cdot N_i(0,1),\end{cases} \quad (5)$$

式中，(θ_i,σ_i)是父代个体的第i个分量；(θ_i',σ_i')是子代新个体的第i个分量；$N(0,1)$是服从标准正态分布的随机数；$N_i(0,1)$是针对第i个分量重新产生一次符合标准正态分布的随机数。r'是全局系数，等于$(\sqrt{2\sqrt{n}})^{-1}$，常取1；r是局部系数，等于$(\sqrt{2\sqrt{n}})^{-1}$，常取1，$\sigma(0)=3.0$.

(2)随机生成初始群体. 进化策略中初始群体由μ个个体组成，每个体(θ,σ)内又可以包含n个θ_i,σ_i分量. 产生初始个体的方法是随机生成.

(3)计算初始个体的适应度. 这里适应度函数取

$$F(\theta)=S(\theta), \quad (6)$$

其中，$s(\theta)=\sum_{i=1}^{n}(y_i-f(x_i,\theta))^2$. 如若满足条件，终止；否则，往下进行. 终止条件一般选择最大迭代次数T_0，若运行次数超过T_0，程序终止.

(4)根据进化策略，用下述操作产生新群体：

1)重组：由两个父代个体交换目标变量和标准差产生新个体. 一般目标变量采用离散重组，标准差采用中值重组.

①离散重组：

从μ个父代个体中用随机的方法任选两个个体：

$$(\theta^1,\sigma^1)=((\theta_1^1,\theta_2^1,\cdots,\theta_n^1),(\sigma_1^1,\sigma_2^1,\cdots,\sigma_n^1)),$$

$$(\theta^2,\sigma^2)=((\theta_1^2,\theta_2^2,\cdots,\theta_n^2),(\sigma_1^2,\sigma_2^2,\cdots,\sigma_n^2)). \quad (7)$$

然后从这两个个体中组合出新个体：

$$(\theta,\sigma)=((\theta_1^{q_1},\theta_2^{q_2},\cdots,\theta_n^{q_n}),(\sigma_1^{q_1},\sigma_2^{q_2},\cdots,\sigma_n^{q_n})), \quad (8)$$

式中$q_i=1$或2. 它以相同的概率针对$i=1,2,\cdots,n$随机选取.

②中值重组：

从μ个父代个体中用随机的方法任选两个个体，如(5)式，然后将父代个体各分量的平均值作为子代新个体的分量. 构成的新个体为

$$(\theta,\sigma)=((\theta_1^1,\theta_1^2)/2,(\theta_2^1+\theta_2^2)/2,\cdots,(\theta_n^1+\theta_n^2)/2),$$

$$(\sigma_1^1+\sigma_1^2)/2,(\sigma_2^1,\sigma_2^2)/2,\cdots,(\sigma_n^1+\sigma_n^2)/2)). \quad (9)$$

2)突变：采用目标变量柯西突变[6]，将原来进化策略中的变异算子中目标变量的修正，由原来的高斯分布产生的随机数改为由柯西分布所产生的随机数来修正. 即可以用下式表示

$$\sigma'_i(j)=\sigma'_i(j)\cdot\exp(\tau_2\cdot\eta+\tau_1\cdot\eta_j),$$

$$\theta'_i(j)=\theta'_i(j)+\sigma'_i(j)\cdot N_j(0,1),i=1,\cdots,n. \quad (10)$$

这里η,η_j均为$t=1$的柯西随机变量比例参数，其中η_j用于更新每一个分量，(10)式中标准差保持原来的高斯变异. 对重组后的个体添加随机量，按照(2)式产生新个体.

3)计算新个体适应度.

4)选择：采用精英策略[7]，将这一代最优个体与上一代最优个体做比较，如果它比上一代优就保留这一代最优个体；否则保留上一代最优个体组成下一代群体，按(μ,λ)选择策略，挑选出优良个体组成下一代群体.

(5)反复执行(4). 当运行次数超过T_0时程序终止，选择最佳个体作为进化策略的结果.

3 仿真实例

例1[8,9] 杨氏弹性模量的估计问题.

采用光杠杆法测量7次，所加力F(N)和对应标尺的变化量Δn (cm)分别为9.8012,0.77;19.6024,1.55;29.4036,2.30;39.2048,3.10;49.2048,3.87;58.8072,4.66;68.6084,5.42. 适应度函数取为

$$f=\sum_{i=1}^{7}(\hat{F}-F)^2,$$

式中的d,b,L,R应用测量工具测得分别为$d=0.0005$(m),$b=0.067$(m),$R=1.16$(m),$L=0.96$(m). 由实验知待估计参数杨氏弹性模量E的取值范围为$1.85\times10^{11}<E<2.0\times10^{11}$.

参数设置分别为，随机生成初始群体规模$\mu=30$，突变产生的新个体$\lambda=210$，初始时标准差取为

$\sigma(0)=3.0$，r' 全局系数取 1；r 局部系数取 1. 目标变量采用柯西突变的 η,η_j 均为 $t=1$ 柯西随机变量；选择：采用 (30,210) 选择策略，挑选优良的个体组成下一代群体，终止条件设为最大迭代次数 $T_0=200$ 和解的精度 $\varepsilon=0.1$.

由表 1 可知，新算法收敛速度快，效果也比逐差法好，可以将进化策略应用在测量数据的处理中去.

表 1　两种不同算法所得结果

Table 1　Results of two different algorithms

方法 Methods	E (N/m^2)	$\sum_{i=1}^{7}(\hat{F}-F)^2$
逐差法 By difference method	1.93×10^{11}	0.122717
改进进化策略算法 Improved evolution ary strategy algorithm	1.92876×10^{11}	0.119764

例 2[3]　在不同温度下，测定铜棒的长度，试估计 0℃时的铜棒长度 y_0 和铜的线性膨胀系数 a. 测量方程为 $y=y_0(1+at)$，最小二乘估计为

$$\sum v^2=\sum_{i=1}^{6}(l_i-y_0(1+at))^2,$$

适应度函数取

$$f=\sum_{i=1}^{6}(l_i-y_0(1+at_i))^2.$$

参数设置与例 1 相同。确定个体的取值范围：在 T(℃)分别为 10,20,25,30,40,45 时，测得 L (cm)对应为 2000.36,2000.72,2000.80,2001.07,2001.48,2001.60，长度分量 y_i (mm)取值范围为 $1990<y_i<2010$，线性膨胀系数分量为 [0,1] 的值. 应用进化策略估计的参数值与传统的方法比较(结果见表 2)，改进进化策略算法求解精度较高.

表 2　3 种不同算法所得结果

Table 2　Results of three different algorithms

方法 Methods	y_0 (mm)	a (℃)	$\sum v^2$
矩阵计算 Matrix Calculation	1999.97	0.0000183	0.010533
文献[10]方法 Reference [10]	1999.98	0.0000182	0.010507
改进进化策略算法 Improved evolution ary strategy algorithm	1999.96559	0.000018253	0.0105064

例 3[9]　对两个电容 C_1 和 C_2 进行等精度测量，第一次测量 C_1，第二次测量 C_2，第三次将之串联，得到的测量结果分别为 $l_1=0.2071\mu F$，$l_2=0.2056\mu F$，$l_3=0.4111\mu F$. 求最佳估计电容值.

待估计电容的取值范围为 [0,1]，测量方程为 $f_1=C_1$，$f_2=C_2$，$f_3=C_1+C_2$，$v_i=l_i-f_i$. 适应度函数取为 $F=\dfrac{1}{(1+\sum_{i=1}^{3}v_i^2)}$.

参数设置的终止条件设为最大迭代次数 $T_0=100$ 和解的精度 $\varepsilon=0.999$，其余参数与例 1 相同.

从表 3 结果可以看出，改进进化策略，进行测量数据中的参数估计比其余两种方法好.

表 3　3 种不同算法所得结果

Table 3　Results of three different algorithms

方法 Methods	C_1 (μF)	C_2 (μF)	$\sum_{i=1}^{3}v_i^2$
矩阵计算法 Matrix Calculation	0.206613	0.205115	0.0000011905
文献[10]方法 Reference [10]	0.206613	0.205115	0.0000011906
改进进化策略算法 Improved evolution ary strategy algorithm	0.206611	0.205116	0.0000011901

4　结束语

本文针对线性回归模型，应用改进进化策略进行线性参数估计，充分利用进化策略这种启发式的随机搜索方法，仿效生物的进化与遗传，根据生存竞争和优胜劣汰的原则，使所要解决的问题从初始解逐渐逼近最优解. 仿真实例结果表明，该方法简单通用，鲁棒性强，高效实用，能处理线性回归参数估计问题.

参考文献：

[1] Nash J C, Walker-Smith M. Nonlinear regression modeling-a unified practical approach[M]. New York: Macel Dekker Inc, 1983.

[2] 陈金山，韦岗. 一种新的非线性回归模型参数估计算法[J]. 控制理论与应用，2001,18(5):808-810.

[3] 何长英. 演化计算在参数估计中的应用[D]. 武汉大学硕士学位论文，2004.

[4] 云庆夏. 进化算法[M]. 北京：冶金工业出版社，2000.

[5] Jennrich RI. Asymptotic properties of nolinear least squares estimation [J]. Ann math statist, 1969, 40: 633-643.

[6] 王战权，唐春安，云庆夏，等. 进化规划和进化策略中变异算子的改进研究[J]. 华东冶金学院学报，1999,16(4):295-300.

[7] 张德勇，黄莎白. 一种改进的基于精英策略的 EDPGA 算法[J]. 控制与决策，2004,19(4):465-467.

[8] 桂传友. 杨氏弹性模量测量实验的优化设计[J]. 巢湖学院学报，2005,7(3):63-66.

[9] 龙作友，王丰. 大学物理实验[M]. 武汉：湖北科学技术出版社，2003.

[10] 石玉. 提高实数遗传算法数值优化效率的研究[D]. 南京航空航天大学博士学位论文，2002.

（责任编辑：尹　闯）