

DOI:10.16356/j.1005-2615.2019.05.018

基于遗传算法的混合威布尔分布参数最小二乘估计

董力¹ 陆中¹ 周伽²

(1. 南京航空航天大学民航学院, 南京, 211106; 2. 东方航空江苏有限公司飞机维修部, 南京, 211113)

摘要: 混合威布尔分布模型常用来分析具有多种失效模式的复杂系统的可靠性数据, 由于模型中包含较多参数, 与单一威布尔分布相比, 混合威布尔分布的参数估计更为复杂。利用遗传算法为优化方法, 提出了一种混合威布尔分布参数估计的最小二乘方法。以残差平方和最小为优化目标, 以各参数取值范围为约束条件, 构建了混合威布尔分布的非线性最小二乘优化模型; 通过变换决策变量上下限、引入惩罚因子和保存最优个体等策略改进传统遗传算法以提高算法的性能, 进而利用改进后的遗传优化算法对混合威布尔分布的非线性最小二乘优化模型进行求解。实例分析表明本文方法有效, 利用本文方法计算得到的可靠度估计值与真实值之间的最大偏差和标准均方根误差, 相对于图估计法分别减少了 0.028 4 与 0.032 8, 相对于极大似然估计法分别减少了 0.000 8 与 0.003 6。

关键词: 混合威布尔分布; 参数估计; 非线性最小二乘法; 遗传算法; 参数优化

中图分类号: V355 **文献标志码:** A **文章编号:** 1005-2615(2019)05-0711-08

Least Square Estimation for Mixed Weibull Distribution Based on Genetic Algorithm

DONG Li¹, LU Zhong¹, ZHOU Jia²

(1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China;

2. Department of Aircraft Maintenance, China Eastern Airlines Jiangsu Limited, Nanjing, 211113, China)

Abstract: Mixed Weibull distributions are widely used in the reliability data analysis of complex systems with multiple failure modes. Since the distribution functions contain much more parameters, the parameter estimation of the mixed Weibull distribution is more complicated than the single Weibull distribution. A least square method for parameter estimations of mixed Weibull distributions is proposed by using a genetic algorithm as the optimization method. The minimum residual sum of square is taken as the optimization objective. The ranges of each parameter are taken as the constraints. And a nonlinear least square model of the mixed Weibull distribution is developed. The traditional genetic algorithm has been improved by altering the upper and lower bounds, introducing the penalty factor and preserving the optimal individual. And the nonlinear least squares model is solved based on the improved genetic algorithm. A case study is presented to illustrate the effectiveness of our proposed method. The maximum deviation and the normalized root mean square error obtained by our method are reduced by 0.028 4 and 0.032 8 compared with those of the graph estimation method, and reduced by 0.000 8 and 0.003 61 compared with those of the maximum likelihood estimation method.

Key words: mixed Weibull distribution; parameter estimation; nonlinear least square method; genetic algorithm (GA); parameter optimization

基金项目: 国家自然科学基金(U1733124, 61403192)资助项目; 江苏省自然科学基金(BQ20130811)资助项目; 南京航空航天大学研究生创新基地(实验室)开放基金(kfj20180709)资助项目。

收稿日期: 2018-08-31; **修订日期:** 2019-01-30

通信作者: 陆中, 男, 副教授, E-mail: luzhong@nuaa.edu.cn。

引用格式: 董力, 陆中, 周伽. 基于遗传算法的混合威布尔分布参数最小二乘估计[J]. 南京航空航天大学学报, 2019, 51(5): 711-718. DONG Li, LU Zhong, ZHOU Jia. Least Square Estimation for Mixed Weibull Distribution Based on Genetic Algorithm[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5): 711-718.

复杂机械产品通常包含多个零部件,这些零部件会在多种故障机理的相互作用下发生失效,从而形成多种失效模式,并且每种失效模式都对应于不同的失效分布。威布尔分布是描述机械产品寿命常用的分布模型,但由于复杂机械产品包含多种失效模式,用单一威布尔分布对其寿命分布进行描述会出现很大的误差。混合威布尔分布能够准确描述产品在多种失效原因或模式下的寿命变化规律,在复杂机械可靠性领域有着广泛的应用。

Attardi等^[1]利用混合威布尔模型分析了汽车零部件的寿命变化规律,验证了故障数据中弱子样的存在,并在模型中引入协变量,提出一种混合威布尔分布的回归模型;Kumaravel等^[2]运用混合威布尔分布模型预计了某地区风电场的风速概率分布和年发电量,较单一威布尔分布模型其结果更接近实际情况;Zhou等^[3]利用混合威布尔分布分析高速铁路设备的故障数据,并对比几种传统随机失效分布函数模型,通过K-S检验得出混合威布尔分布模型具有更精确的结果。

相比于单一威布尔分布的参数估计,混合威布尔分布的参数估计较为复杂。传统参数估计方法中常用到图估计法和极大似然法。图估计法是将失效数据绘制在概率纸上,具有计算简单、直观的优点,但是估计的结果受主观因素影响比较大,估计的精度较差,因此图估计法常常被用来处理现场数据以及确定优化算法的初始值。极大似然估计是一种具有渐近性的优化方法,对于多参数分布模型的参数估计求解过程比较复杂,特别是求解非中心混合分布参数时需要联立一组超越方程。因此,混合威布尔分布模型中未知参数的估计一直以来都是研究的热点。凌丹等^[4]借助非线性最小二乘法法和L-M(Levenberg-Marquardt)法估计5参数混合威布尔分布的未知参数,相比于图解法其具有更高的精度;王松岩等^[5]利用了数据布点法对双重威布尔分布模型中的未知参数进行了估计;Gong^[6]利用一种复杂进化算法对混合威布尔分布参数进行估计,并与极大似然方法估计的结果进行对比,验证了方法具有更高的精度;Nagode和Fajdiga^[7-8]针对多参数形式的混合威布尔分布的参数估计,给出了一种未知参数估计的替代过程,又通过对参数粗略估计和误差处理的优化修改,对其替代过程进行了进一步改进;蒋卉等^[9]利用ECM算法研究了混合威布尔分布在正常工作条件下,完全数据场合、I-型截尾和II-截尾场合的参数估计问题;袁忠大等^[10]将威布尔混合分布模型应用于航空发动机失效数据的处理分析中,并提出了动态权重系数的概念,提高了分布模型的精度;Ling等^[11]利用非线性

最小二乘法理论建立一个双重威布尔分布的参数估计模型,并借助拟牛顿法解决模型的优化问题,通过数值算例,说明该方法相比传统的图形方法具有更高的精度。同时,人们也将不断发展的智能优化算法应用于混合威布尔分布模型未知参数的求解中,Krohling等^[12]利用似然函数和骨干粒子群算法对混合威布尔分布的未知参数进行优化求解;Zhai等^[13]借助遗传优化算法,通过最大化对数似然函数的方法对混合威布尔分布模型中的未知参数进行了估计求解;Chi等^[14]提出了一种基于贝叶斯定理的混合威布尔优化模型,并借助布谷鸟搜索算法(Cuckoo search algorithm, CSA)求解模型中未知参数;付涛等^[15]运用极大似然估计求解混合威布尔分布中的未知参数,并借助混合粒子群优化算法求解联立的超越方程,计算结果相比图估计法具有更高的精度。

非线性最小二乘法是解决非线性最优化问题的一类重要方法,目前广泛应用在试验数据的拟合分析上,对于数据点较少的小样本数据仍可以获得较好的拟合结果^[4]。本文综合利用遗传优化算法与非线性最小二乘法,提出了一种混合威布尔分布的非线性最小二乘参数估计方法,并通过变换决策变量上下限、引入惩罚因子和保存最优个体等策略对传统遗传算法进行改进以提高算法性能,相比于传统参数估计方法中的图估计法、极大似然估计方法,本文方法计算得到的可靠度估计值与真实值之间的最大偏差、标准均方根误差都较小。

1 混合威布尔分布的非线性最小二乘估计模型

假设随机变量 T 服从混合威布尔分布,定义混合威布尔分布的概率密度函数为

$$f(t) = \sum_{i=1}^n p_i \frac{\beta_i}{\eta_i^{\beta_i}} t^{\beta_i-1} e^{-\left(\frac{t}{\eta_i}\right)^{\beta_i}} \quad (1)$$

式中: n 表示混合威布尔分布模型的重数; p_i 表示第 i 个子分布权重; η_i 表示第 i 个子分布的尺度参数; β_i 表示第 i 个子分布的形状参数。

混合威布尔分布的可靠度函数 $R(t)$ 可表示为

$$R(t) = \sum_{i=1}^n p_i e^{-\left(\frac{t}{\eta_i}\right)^{\beta_i}} \quad (2)$$

式(2)中,未知参数的数量为

$$n_p = 3n - 1 \quad (3)$$

一般的威布尔分布可通过简单的数学变换,将样本数据 t 和对应的函数值 $F(t)$ 之间的非线性关系转化成线性关系,因而能够借助线性回归法进行参数估计。但是对于混合威布尔分布而言,无法将 t

和 $F(t)$ 变换成线性关系,故不能用线性回归的方法估计未知参数。非线性最小二乘法是最小二乘法的一种,基本思想是“误差平方总和最小”,常用于解决一些非线性优化问题,在参数估计、拟合分析方面有着广泛应用。针对混合威布尔分布模型,可借助非线性最小二乘法,构建出多参数 Weibull 分布的非线性最小二乘法模型。

以 $t_1 \leq t_2 \leq \dots \leq t_j \leq \dots \leq t_m; j=1, 2, \dots, m$ 表示寿命样本,可通过式(4)求出对应可靠度的估计值

$$\tilde{R}(t_j) = 1 - \frac{j}{m} \quad (4)$$

当样本数量小于 20 时,常用式(5)计算可靠度的近似中位秩估计值

$$\tilde{R}(t_j) = 1 - \frac{j - 0.3}{m + 0.4} \quad (5)$$

式中: j 为数据样本序号; m 为样本容量。

用 $\theta = (p_1, \eta_1, \beta_1, p_2, \eta_2, \beta_2, \dots, p_n, \eta_n, \beta_n)$ 表示混合威布尔分布中未知参数组成的向量,建立混合威布尔分布非线性最小二乘参数估计方法的数学模型

$$\begin{aligned} \text{ming}(\theta) &= \sum_{j=1}^m \left\{ \tilde{R}(t_j) - \sum_{i=1}^n p_i e^{-\left(\frac{t_j}{\eta_i}\right)^{\beta_i}} \right\}^2 \\ \text{s.t.} \quad &\begin{cases} 0 \leq p_i \leq 1 \\ \eta_{\text{down}i} \leq \eta_i \leq \eta_{\text{up}i} \\ \beta_{\text{down}i} \leq \beta_i \leq \beta_{\text{up}i} \end{cases} \end{aligned} \quad (6)$$

式中: $g(\theta)$ 表示混合威布尔分布的非线性最小二乘法参数估计模型的目标函数; $\eta_{\text{up}i}, \eta_{\text{down}i}$ 分别表示第 i 个子分布的尺度参数的约束上限和下限; $\beta_{\text{up}i}, \beta_{\text{down}i}$ 分别表示第 i 个子分布的形状参数的约束上限和下限。

通过借助优化算法可找到一组最优的 θ ,使对应的目标函数 $g(\theta)$ 的值最小,并将此时的 θ 作为模型中未知参数的估计值。

2 基于改进遗传算法的最小二乘法

遗传算法(Genetic algorithm, GA)是借鉴生物界的进化规律演化而来的自适应全局优化概率搜索算法,具有很强的适应性,在组合优化等方面得到了广泛的应用^[16-17]。本文以混合威布尔分布非线性最小二乘估计模型的残差平方和最小作为算法的优化目标,借助遗传算法对未知参数进行优化求解,通过改变上下限,引入惩罚因子和保存最优个体策略来对传统遗传算法进行改进以提高算法性能,改进后遗传算法的构造过程如图 1 所示。

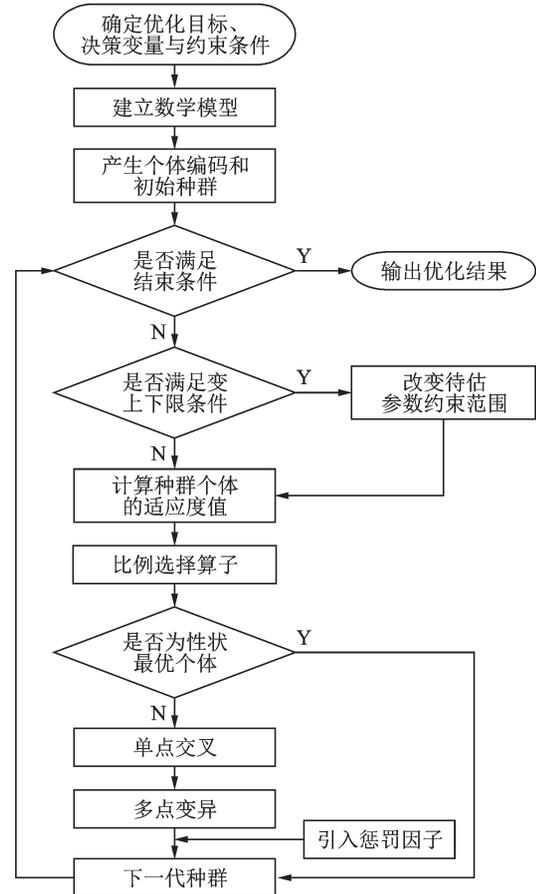


图 1 改进 GA 流程图

Fig. 1 Flowchart of the improved GA

2.1 初始种群的产生

本文采用多参数级联的方法对待估参数进行二进制编码,以 $l_i (i=1, 2, \dots, n_p)$ 表示第 i 个待估参数的编码长度,可通过式(7)计算 l_i 的值

$$l_i = \left\lceil \log_2 \frac{s_{\text{up}} - s_{\text{down}}}{p_{\text{rec}}} \right\rceil + 1 \quad (7)$$

式中: $s_{\text{up}}, s_{\text{down}}$ 表示待估参数约束的上限和下限; p_{rec} 表示定义的待估参数的求解精度;符号 $\lceil \cdot \rceil$ 表示取符号中的整数部分。

例如,当待估参数的上限 s_{up} 取 15, 下限 s_{down} 取 5, 精度 p_{rec} 取 0.01, 通过式(7)可计算该待估参数的二进制编码长度为 10。

以 L 表示参数级联后的二进制编码串的总长度, L 可通过下式求得

$$L = \sum_{i=1}^{n_p} l_i \quad (8)$$

利用图估计法可以求出待估参数的粗略估计值,也可借助于常用的非参数估计方法,通过观察样本数据点的分布范围、变化趋势,粗略估计出待估参数的大致范围,然后进行编码生成初始种群。

2.2 遗传算子的设计

2.2.1 选择算子

比例选择算子是传统遗传算法中常用的选择

算子,但在处理不同的算法时并不是最合适的,有时候会出现性状较好的个体不能被选择的问题^[18]。为了避免这种情况的出现,将比例选择算子与最优个体保存策略相结合,以 $g(\theta_k)(k=1,2,\dots,N)$ 表示第 k 个种群个体的适应度值, N 表示种群规模,过程如下:

步骤1 计算种群中每一个体的相对适应度值

$$f_i(k) = \frac{g(\theta_k)}{\sum_{i=1}^N g(\theta_i)} \quad (9)$$

步骤2 对个体的相对适应度值进行累计求和

$$F_s(k) = \frac{\sum_{i=1}^k g(\theta_i)}{\sum_{i=1}^N g(\theta_i)} \quad (10)$$

步骤3 生成0到1之间的随机数 r_{rd} ,然后依次对 N 个个体进行判断。若 $F_s(k-1) < r_{rd} \leq F_s(k)$,则个体 k 被选中,其中定义 $F_s(0)=0$,选中个体将被遗传到下一代,多次执行步骤3,直至下一代新的种群产生。

步骤4 保存最优个体,找出种群中性状最优的个体,并将其直接作为下一代种群中的个体。

2.2.2 交叉算子

设计交叉算子时,多点交叉^[19]可能破坏一些好的个体编码串,不易找到问题的最优解,因此本文采用单点交叉,过程如下:

步骤1 对种群中的个体进行两两随机配对。

步骤2 对每一对配对个体进行操作,产生0到1之间的随机数 r_{rd} ,然后对其进行判断。若 r_{rd} 小于预先设定的交叉概率 P_c ,则配对个体进行第3步交叉过程,否则不进行交叉过程。

步骤3 交叉过程,随机产生一个1到 L 之间的整数,将其对应的编码点作为交叉的起始点,在配对个体的二进制编码中,从交叉点到最后的编码点之间的片段作为交叉片段进行互换。

2.2.3 变异算子

考虑到目标函数中待估参数较多,个体的二进制编码串较长,本文采用多点变异,将待估参数分开考虑,对每一待估参数所对应的编码串进行并行式的变异操作,过程如下:

步骤1 变异对象的选择,对每一个体进行变异操作,随机产生 n_p 个0到1之间的随机数 $[\delta_1, \delta_2, \dots, \delta_{n_p}]$,依次比较预先设定的变异概率 P_m 与 $[\delta_1, \delta_2, \dots, \delta_{n_p}]$ 的大小。若小于 P_m ,则对其待估参数对应的编码串进行变异操作;若大于 P_m ,则该

待估参数不发生变异。

步骤2 变异操作, l_i 表示待估参数对应的二进制编码长度,产生一个1到 l_i 的随机整数 Z 。若对应编码串中第 Z 个二进制位为1则变为0,若为0则变为1。

2.3 算法的改进

传统的遗传算法在实际的运算过程中存在一些不足,比如会出现过早收敛,最优个体不被选中的现象,为改善传统遗传算法的性能,对其进行以下改进。

(1)变上下限

由于遗传算法在进化过程中能够自动获取和引导优化的搜索空间,当算法运算到若干代后,种群个体的性状趋于相近,相比于整个算法搜索区域,大量种群个体会相对集中在一定的区域范围内,以上下限为边界的约束范围内的某些空间被遗弃,一定程度上造成了空间的浪费。对于一些需要进行很多次(大于500次)进化代数的遗传优化算法,为了提高搜索空间的利用率,并考虑局部最优与全局最优的因素,在算法运算到一定程度后对算法的搜索区域进行调整,改变当前的约束范围,因而式(6)给出的待估参数的上下限在运算过程中将会发生变化。变上下限的具体操作:在算法运算达到给定代数时,以当前种群中的最优个体为基准,采用不同的调整值对每个待估参数约束的上下边界进行调整,改变和缩小参数的约束范围,使种群密度不断增大,一定程度上也有利于提高优化算法的精度。

(2)最优保存策略

在算法搜索过程中,设计的选择算子操作存在一定的随机性,有时候会出现性状最差的个体有可能被选中或者性状最优的个体不能被选中的现象,为避免最优个体丢失,采用最优保存策略,将最优个体直接遗传到下一代种群中,不必再参与交叉和变异过程。

(3)惩罚因子的引入

遗传算法搜索过程中,有时候会出现过早收敛的现象,造成种群基因在演变过程早期变得单一化,种群个体之间的相互竞争减弱,从而导致算法陷入局部最优,不利于找到最优值。为了避免此类现象的出现,在种群中引入惩罚因子,增加种群的多样性,避免过早的进入收敛。引入惩罚因子的具体操作:在算法演变过程的每一代种群中随机加入一定数量的新个体,加入的新个体尽可能均匀地分布在种群中,并删除种群中相同数量的个体,保持种群规模大小不变,加入的新个体可以随机生成也可以人为选择一些个体,删除的原有个体不能是种

群的最优个体。随机加入的新个体以及在算法演变过程中的多点变异操作可能使一些种群个体处在规定约束范围边界附近或处在规定约束范围的边界上,为保证种群个体在演变过程中始终处于规定的约束范围内,可利用内部惩罚函数法^[20],引入惩罚因子 r ,并定义障碍函数

$$G(\theta, r) = g(\theta) + rB(\theta) \quad (11)$$

式中: $B(\theta)$ 表示连续函数, r 是极小的正数,当 θ 趋向约束边界时, $B(\theta)$ 将趋向于无穷大, $B(\theta)$ 可用以下形式表示

$$B(\theta) = \sum_{i=1}^{n_p} \frac{1}{b_i} \quad (12)$$

式中, b_1, b_2, \dots, b_{n_p} 表示 θ 所对应的每一待估参数取值与待估参数约束边界之间的距离,当 θ 趋向于边界时, $G(\theta, r)$ 将趋向于无穷大,否则 $G(\theta, r) \approx g(\theta)$ 。

3 算 例

为验证提出的方法是有效的,以一个包含 6 参数的双重威布尔分布为例,假定模型中的参数 $\theta = (0.15, 20, 0.6, 0.85, 500, 5)$,利用文献[21]给出的逆变换技术产生一组以 θ 为参数,容量为 200 的 6 参数混合威布尔分布的数据样本,样本中的数据如表 1 所示。

3.1 待估参数的初始值范围

将 t 与 $F(t)$ 之间的关系转化为 x 与 y 之间的关系,对应关系如下式所示

$$\begin{cases} x = \ln t \\ y = \ln[-\ln(1 - F(t))] \end{cases} \quad (13)$$

绘制失效概率分布变换后的曲线,如图 2 所示。图中,数据样本所对应失效概率的变化趋势存在明显转折,转折前后对应的失效概率变化趋势有不同的斜率,因此可假定数据样本分布服从双参数威布尔分布,计算折点附近数据点对应的失效概率并结合式(4),可粗略估计特征寿命较小的子分布权重 p_1 在 0.1 到 0.2 附近。

假设 $\eta_1 < \eta_2$, 特征寿命 η_1 所对应的失效概率为 $p_1(1 - 1/e)$, 特征寿命 η_2 所对应的失效概率为 $p_2(1 - 1/e) + p_1$, 根据两个子分布的特征寿命对应的失效概率和式(4)给出的公式,粗略估计 η_1 的范围在 5 到 30 之间, η_2 值的范围在 460 到 550 之间;子分布形状参数的初值范围并不影响计算过程^[22],可假定形状参数的范围在 0 到 10 之间,若最后的计算结果在预先设定的边界附近,则应当加宽此约束边界,重新进行优化求解。

表 1 寿命样本

Tab. 1 Lifetime samples

0.2	0.3	0.5	1	1.4	1.7	1.9	1.9
1.9	2.2	4.8	5.6	6.4	9.9	12.7	14.9
16.4	16.9	17.0	22.5	22.6	24.6	26.4	29.4
35.4	40.1	41.2	59.7	127.3	167.1	169.6	194.5
196.6	221.4	223.2	257.9	263.3	263.4	271.0	271.3
273.8	293.2	296.5	296.7	300.2	302.1	305.4	307.2
314.9	319.0	322.2	324.9	327.6	331.0	335.9	337.0
337.2	345.7	348.8	355.3	367.8	369.5	369.9	373.4
374.0	377.2	382.5	382.6	387.7	388.0	391.5	395.4
395.9	397.0	398.5	399.2	400.4	401.3	403.2	404.0
408.6	409.3	410.3	410.8	411.0	411.3	414.5	417.0
417.3	421.2	425.2	425.8	425.8	426.0	426.6	427.0
427.0	429.4	430.6	431.0	436.9	440.5	441.0	448.0
449.7	450.1	451.3	459.1	459.5	459.5	460.2	460.6
464.7	466.7	469.6	469.7	471.5	472.2	475.4	477.1
477.8	478.0	481.1	483.4	484.3	485.9	486.9	488.3
490.4	491.9	492.0	494.1	494.5	494.9	497.7	502.4
505.2	506.3	507.0	508.2	511.5	512.3	512.9	521.6
521.8	522.1	522.6	522.7	525.3	525.4	526.0	527.7
528.7	530.7	533.3	533.6	535.2	536.3	536.3	537.0
538.9	542.8	543.4	547.3	548.3	548.9	550.4	554.2
554.3	555.0	559.5	559.7	560.6	563.7	564.6	567.5
568.3	572.3	575.5	580.8	580.9	584.6	586.2	587.2
588.5	590.1	597.9	598.4	613.6	617.3	621.1	628.6
632.4	635.0	647.9	648.3	652.0	668.2	673.6	675.6

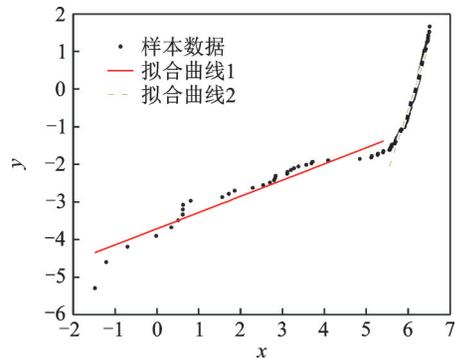


图 2 失效概率分布变换后的曲线

Fig. 2 Curves transferred from the failure probability function

3.2 优化结果

由式(4)和(6)可确定优化的目标函数表达式为

$$\begin{aligned} \text{ming}(\hat{\theta}) &= \sum_{j=1}^m \left(\tilde{R}(t_j) - \sum_{i=1}^n p_i e^{-\left(\frac{t_j}{\eta_i}\right)^{\beta_i}} \right)^2 = \\ &= \sum_{j=1}^{200} \left(1 - \frac{j}{200} - \left[p_1 e^{-\left(\frac{t_j}{\eta_1}\right)^{\beta_1}} + p_2 e^{-\left(\frac{t_j}{\eta_2}\right)^{\beta_2}} \right] \right)^2 \end{aligned} \quad (14)$$

在表1中给出的数据样本的基础上,利用文中提出的基于遗传算法的混合威布尔分布参数最小二乘估计方法对模型中的未知参数进行优化求解,分别以传统的遗传算法和改进后的遗传算法进行

4次优化求解,计算结果如表2和表3所示。

图3给出了传统遗传算法与改进遗传算法的优化过程对比,图3(a—d)分别给出了与表2、表3相对应的4次参数估计的优化过程。

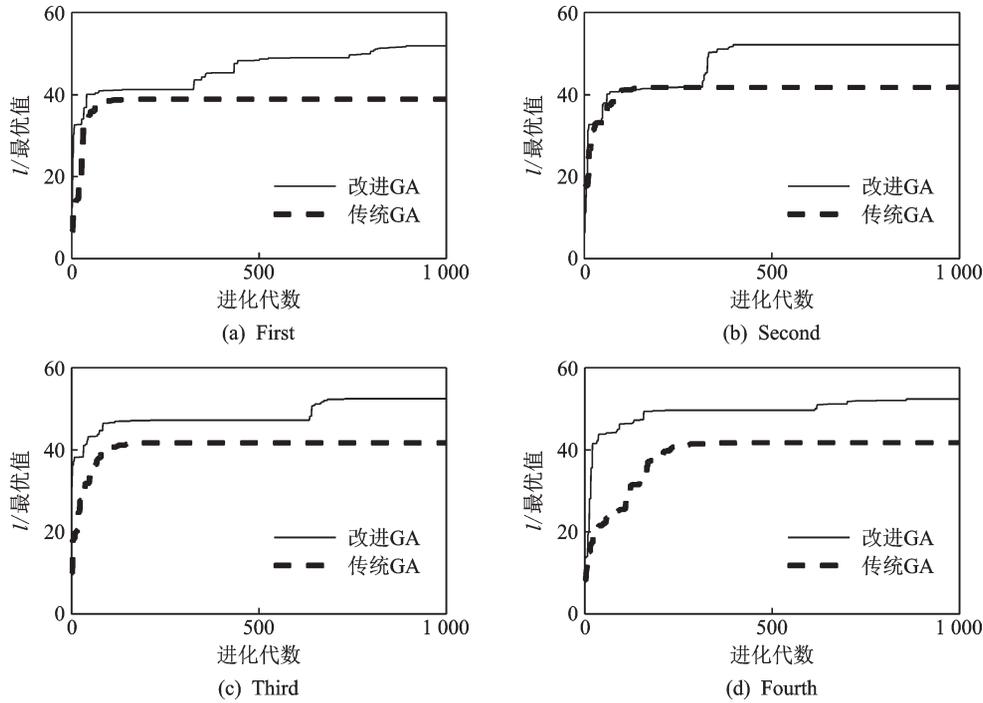


图3 传统GA与改进GA的优化过程

Fig. 3 Optimization comparison between the traditional GA and the improved GA

表2 传统遗传算法的优化结果

Tab. 2 Results of the traditional GA

组序	参数估计值 $\hat{\theta}$						$g(\hat{\theta})$
	p	η_1	β_1	p_2	η_2	β_2	
第1次	0.150	16.399	0.657	0.850	501.320	4.875	0.0224
第2次	0.148	15.795	0.656	0.852	499.951	4.875	0.0239
第3次	0.148	15.938	0.657	0.852	499.951	4.874	0.0239
第4次	0.149	15.959	0.657	0.851	499.951	4.874	0.0239
平均值	0.149	16.023	0.657	0.851	500.293	4.875	0.0233

表3 改进后遗传算法的优化结果

Tab. 3 Results of the improved GA

组序	参数估计值 $\hat{\theta}$						$g(\hat{\theta})$
	p_1	η_1	β_1	p_2	η_2	β_2	
第1次	0.164	21.832	0.603	0.836	503.079	5.072	0.0192
第2次	0.169	23.645	0.590	0.831	504.057	5.116	0.0191
第3次	0.166	22.622	0.619	0.834	503.226	5.101	0.0192
第4次	0.164	21.499	0.605	0.836	503.320	5.055	0.0193
平均值	0.166	22.400	0.604	0.834	503.421	5.086	0.0191

3.3 拟合优度检验

以改进后遗传算法的五组优化结果的平均值 $\hat{\theta}$ 作为模型中未知参数估计值,在显著性水平为 $\alpha = 0.05$ 下进行威尔逊 χ^2 检验,以数据样本分布服从参数 $\hat{\theta}$ 的双重威布尔分布为原假设 H_0

$$\chi_q^2 = \sum_{i=1}^m \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} \quad (15)$$

式中: m 表示样本组数, f_i 表示每组的频数, \hat{f}_i 表示在原假设 H_0 条件下第 i 组的预期样本数。计算有

$$\chi_q^2 = 2.447 < \chi_{0.05}^2 = 7.815$$

表明原假设 H_0 是被接受的。

3.4 结果分析

在本文提出的方法中,对传统遗传算法进行了

改进,在优化过程中引入了惩罚因子,从而增强了种群的多样性,避免算法陷入局部最优,易于找到全局最优解。比较传统的遗传算法和改进的遗传算法,不难发现:在优化过程中,改进的遗传算法相比传统的遗传算法出现过早收敛(“早熟”现象)的几率明显有所改善,如图3所示,在算法迭代到100~200代时,传统的遗传算法对应的曲线开始保持水平,直到算法结束,而改进的遗传算法的这一过程明显靠后;在优化结果上,如表2和表3所示,改进的遗传算法相比传统的遗传算法有更高的精度,表明了对传统遗传算法的改进是有效的,改善了算法的性能。

针对表1所给数据,利用本文方法、图估计法、

极大似然估计方法求得的未知参数估计值见表 4。

参数估计的优劣可由可靠度估计值与真实值的最大偏差以及标准均方根误差(Normalized root mean squared error, NRMSE)^[23]来衡量。可靠度估计值与真实值的最大偏差可由式(13)求得

$$D = \max |R(t_i) - \hat{R}(t_i)| \quad (16)$$

式中: $R(t_i)$ 表示可靠度的真实值; $\hat{R}(t_i)$ 表示根据参数估计结果求出的可靠度值。

可靠度估计值与真实值的标准均方根误差可由式(14)求得

$$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^m [R(t_i) - \hat{R}(t_i)]^2}{\sum_{i=1}^m R(t_i)^2}} \quad (17)$$

表 4 同样给出了由上述 3 种方法进行参数估计后求得的可靠度估计值与真实值的最大偏差以及标准均方根误差。

由表 4 可以看出,本文方法计算得到的可靠度估计值与真实值之间的最大偏差、标准均方根误差,相对于图估计法减少了 0.028 4, 0.032 8;相对于极大似然估计方法减少了 0.000 8, 0.003 6。

表 4 3 种参数估计方法的估计结果

Tab. 4 Estimation of three kinds of parameter estimation methods

参数	P_1	η_1	β_1	P_2	η_2	β_2	D	NRMSE
图估计法	0.175	27.865	0.736	0.825	507.029	6.133	0.0424	4.21e-2
MLE	0.150	14.809	0.752	0.850	497.898	4.874	0.0148	1.29e-2
本文	0.166	22.400	0.604	0.834	503.421	5.086	0.0140	9.33e-3

4 结 论

本文提出了基于 GA 的混合威布尔分布最小二乘估计方法,利用该方法求解混合威布尔分布模型中的未知参数所得到的计算结果与未知参数的真实值很接近,表明了本文中提出的方法是有效的,能够实施的。

相比于传统的遗传算法,改进的遗传算法在精度、过早收敛方面进行了改善,提高了算法的性能。

相比于传统参数估计方法中的图估计法、极大似然估计法,本文提出的基于 GA 的混合威布尔最小二乘估计方法减少了主观因素对计算结果的影响,同时又避免了极大似然估计求解多参数混合威布尔分布时复杂的求解过程,易于开展实施,并且具有更高的求解精度。

多组算例表明,本文提出的方法在小样本、大样本情况下均具有很好的适用性,能够得出较高精度的计算结果,并且通过适当变换,本文提出的方法同样可以应用到其他几种改进威布尔分布模型的参数估计中。

参考文献:

[1] ATTARDI L, GUIDA M, PULCINI G. A mixed-Weibull regression model for the analysis of automotive warranty data[J]. Reliability Engineering and System Safety, 2005, 87: 265-273.
 [2] KUMARAVEL R, VARUN C, SARFUDEEN M. Mixed Weibull distribution: A case study on Ichanda, India[J]. Wind Engineering, 2014, 38(6): 613-620.

[3] ZHOU F F, XIA L, ZHAO Q C, et al. Statistical analysis of the random failure model of high-speed railway equipments[C]//2014 IEEE International Conference on Networking, Sensing and Control. [S. l.]: IEEE, 2014: 215-220.
 [4] 凌丹,黄洪钟. 混合威布尔分布参数估计的 L-M 算法[J]. 电子科技大学学报, 2008, 37(4): 634-636, 640. LING Dan, HUANG Hongzhong. Parameters estimation for mixed weibull distribution using L-M algorithm [J]. Journal of University of Electronic Science and Technology of China, 2008, 37(4): 634-636, 640.
 [5] 王松岩,于继来. 短时风速概率分布的混合威布尔逼近方法[J]. 电力系统自动化, 2010, 34(6): 89-93. WANG Songyan, YU Jilai. Approximation of two-peak wind speed probability density function with mixed Weibull distribution[J]. Automation of Electric Power Systems, 2010, 34(6): 89-93.
 [6] GONG Z J. Estimation of mixed Weibull distribution parameters using the SCEM-UA algorithm: Application and comparison with MLE in automotive reliability analysis[J]. Reliability Engineering & System Safety, 2006, 91(8): 915-922.
 [7] NAGODE M, FAJDIGA M. A general multi-modal probability density function suitable for the rainflow ranges of stationary random processes[J]. International Journal of Fatigue, 1998, 20(3): 211-223.
 [8] NAGODE M, FAJDIGA M. An improved algorithm for parameter estimation suitable for mixed Weibull distributions[J]. International Journal of Fatigue, 2000, 22(1): 75-80.

- [9] 蒋卉, 汤银才. 混合 Weibull 分布参数估计的 ECM 算法[J]. 系统科学与数学, 2010, 30(1): 79-88.
JIANG Hui, TANG Yincai. ECM algorithm for parameter estimation of mixed Weibull distribution [J]. Journal of System Science and Mathematical, 2010, 30(1): 79-88.
- [10] 袁忠大, 符双学, 王大伟. 基于威布尔混合分布的航空发动机可靠性评估[J]. 航空计算技术, 2015, 45(5): 65-67.
YUAN Zhongda, FU Shuangxue, WANG Dawei. Evaluation of aero-engine reliability based on Weibull mixed distribution [J]. Aeronautical Computing Technique, 2015, 45(5): 65-67.
- [11] LING D, HUANG H Z, LIU Y. A method for parameter estimation of mixed Weibull distribution [C]//Annual Reliability & Maintainability Symposium. [S.l.]: IEEE, 2009: 129-133.
- [12] KROHLING R, CAMPOS M, BORDES P. Bare bones particle swarm applied to parameter estimation of mixed weibull distribution [M]. Germany: Springer Berlin Heidelberg, 2010: 53-60.
- [13] ZHAI X T, WANG J X, CHEN J S. Parameter estimation method of mixture distribution for construction machinery [J]. Mathematical Problems in Engineering, 2018(1/2): 3124048.
- [14] CHI K, KANG J S, WU K, et al. Bayesian parameter estimation of weibull mixtures using cuckoo search [C]//2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS). Ostrava, Czech Republic: IEEE, 2016: 411-414.
- [15] 付涛, 王大镇, 弓清忠, 等. 基于混合粒子群算法的柴油机喷油器可靠性分布的参数估计[J]. 机械强度, 2014, 36(4): 539-542.
FU Tao, WANG Dazhen, GONG Qingzhong, et al. Diesel injector reliability distribution of parameter estimation based on hybrid particle swarm algorithm [J]. Journal of Mechanical Strength, 2014, 36(4): 539-542.
- [16] 雷秀娟. 群智能优化算法及其应用 [M]. 北京: 科学出版社, 2012: 46-52.
LEI Xiujuan. Swarm intelligent optimization algorithms and their applications [M]. Beijing: Science Press, 2012: 46-52.
- [17] 马永杰, 云文霞. 遗传算法研究进展 [J]. 计算机应用研究, 2012, 29(4): 1201-1206.
MA Yongjie, YUN Wenxia. Research progress of genetic algorithm [J]. Application Research of Computers, 2012, 29(4): 1201-1206.
- [18] 毛昭勇, 宋保维, 李正, 等. 基于遗传算法的最大似然参数优化估计 [J]. 机械强度, 2006, 28(1): 79-82.
MAO Zhaoyong, SONG Baowei, LI Zheng, et al. Optimization method of maximum likelihood estimation parameter estimation based on genetic algorithms [J]. Journal of Mechanical Strength, 2006, 28(1): 79-82.
- [19] 周明, 孙树栋. 遗传算法原理及应用 [M]. 北京: 国防工业出版社, 1999: 11-15.
ZHOU Ming, SUN Shudong. Principle and application of genetic algorithm [M]. Beijing: National Defense Industry Press, 1999: 11-15.
- [20] 曹卫华, 郭正. 最优化技术方法及 MATLAB 的实现 [M]. 北京: 化学工业出版社, 2005.
CAO Weihua, GUO Zheng. Optimization of technical methods and implementation of MATLAB [M]. Beijing: Chemical Industry Press, 2005.
- [21] ELSAYED E. Reliability engineering [M]. 2nd ed. New Jersey: Wiley, 2012.
- [22] 赵继俊, 邹经湘, 张锡清. 可靠性实验数据混合威布尔分布的参数优化估计 [J]. 农业机械学报, 2000, 31(5): 106-109.
ZHANG Jijun, ZOU Jingxiang, ZHANG Xiqing. Parameter estimation of parameters for hybrid Weibull distribution of reliability experiment data [J]. Mechanical Science and Technology, 2000, 31(5): 106-109.
- [23] NOCEDAL J, WRIGHT S J. Numerical optimization [M]. Berlin: Springer, 1999.

(编辑: 张蓓)