

文章编号: 2096-1618(2016)02-0223-05

免疫进化算法在复合极值分布参数估计中的应用

任照环¹, 倪长健²

(1. 南川区气象局, 重庆 南川 408400; 2. 成都信息工程大学大气科学学院, 四川 成都 610225)

摘要:参数估计是复合极值模型应用的难点之一, 针对这一问题, 利用免疫进化算法来估计复合极值模型的参数, 给出在气候统计方面的两个应用实例。结果表明: 免疫进化算法应用于复合极值模型参数求解是合理和可行的; 相比于L-矩估计法及复合矩估计法, 免疫进化算法所求参数拟合效果更好; 免疫进化算法具有高效、普适的优点, 在其他参数优化问题上也有广泛的应用前景。

关键词:气象学; 气候统计; 免疫进化算法; 复合极值分布; 参数估计

中图分类号: P456.8

文献标志码: A

0 引言

在全球变暖的气候背景下, 各种极端气候事件发生日益频繁, 给各国人民造成极大的经济损失和人员伤亡^[1-3]。关于极端气候事件和气候极值问题的研究一直是科学家们关注的热点之一, 在目前对极端气候事件预测能力有限的情况下, 利用频率分析推断一定重现期的气候极值就成为极端气候研究的重要方向^[4-6]。1979年马逢时等^[7-8]提出复合极值模型, 发现复合极值模型克服了传统计算方法在资料年限较短的情况下, 计算结果不稳定, 计算误差大的问题^[9-10], 引起国内外学者的重视^[11-13]。目前复合极值模型的参数估计方法主要有以下几种: 马逢时^[8]假设服从复合极值分布的两个随机变量相互独立, 利用耿贝尔估计法分别对两种随机分布的参数进行估计; 赵维谦^[14]认为, 二者构成的复合极值分布是某一个随机变量的分布, 而非二者的联合分布, 并在泊松分布参数已知时, 利用复合矩估计法计算泊松-耿贝尔复合极值分布的参数; 刘晶^[15]则对比分析了概率权重矩法、极大似然法、复合矩法的计算结果, 认为3种方法中, 极大似然法对泊松-耿贝尔复合极值分布拟合效果最好且表现最稳定; 张香云等^[16]用概率加权矩法给出二项-广义帕累托复合极值分布的参数估计公式, 并在二项分布参数 $m=2$ 时, 给出了形状参数 δ 和尺度参数 σ 的解析表达式。由于复合极值分布模型较复杂, 上述研究给出的参数估计公式也较为繁琐, 计算不便。因此, 寻找一种简单快捷的复合极值模型参数估计方法就显得十分必要。免疫进化算法是在现有进化算法基础上发

展的一种新算法, 灵感来自生物免疫机理, 其诸多优秀的性能已分别从理论上和在应用中得到了证明和验证^[17-19], 文中将其应用于复合极值模型参数的求解, 力图寻找一种普适、高效的参数优化的方法。

1 方法介绍

1.1 免疫进化算法

1.1.1 免疫进化算法的思想及实现技术

借鉴生物免疫机制, 倪长健^[13]提出免疫进化算法, 它具有原理简单易懂、搜索效率高、不易陷入局部最优解的优点。免疫进化算法中的最优个体是每次进化中适合度最好的可能解。从统计上来看, 一方面, 每次进化的最优个体和整体最优解之间的空间距离可能是最小的; 另一方面, 和最优个体之间距离较近的个体也可能有较高的适应度。免疫进化算法中子代个体的生成方法为

$$\begin{cases} x^{t+1} = x_{best}^t + \sigma^t \times N(0, 1) \\ \sigma^{t+1} = \sigma^t \times e^{-\frac{Ax^t}{T}} \end{cases} \quad (1)$$

式中 x^{t+1} 为子代个体的可能解; x_{best}^t 为父代最优个体; σ^{t+1} 为子代个体的方差; σ^t 为父代个体的方差; A 为方差动态调整系数; T 为总的进化代数; $N(0, 1)$ 为产生的服从标准正态分布的随机数。 t 是进化的代数, σ^0 对应于初始群体的方差, A 和 σ^0 具体取值根据被研究的问题确定, 通常 $A \in [1, 10]$, $\sigma^0 \in [1, 3]$ 。免疫进化算法的本质在于充分利用最优个体的信息, 以最优个体的进化代替群体的进化。

1.1.2 免疫进化算法的表述

如果考虑如下优化问题:

$$\max\{f(x) \mid x \in X\} \quad (2)$$

收稿日期: 2015-05-10

基金项目: 公益性行业(气象)资助项目(GYHY201006023)

式(2)中 $f(x)$ 为适应度函数。群体的初始规模设为 S ,且群体规模固定不变。免疫进化算法工作步骤如下:

- (1)确定适应度函数的具体表达式。
- (2)在取值空间内按照正态分布随机生成初始群体,并计算每个取值的适应度,确定最优个体 x_{best}^0 ,给出 σ^0 的取值。
- (3)根据公式式(1)进行进化,在取值空间内生成规模为 S 的子代群体。
- (4)分别计算每个子代个体的适应度,选取适应度最高的个体为最优个体 x_{best}^{t+1} 。
- (5)当达到预设的条件时,进化终止,选择最后一次进化的最优个体作为最优解。否则继续循环。

1.2 复合极值模型

定义有一种离散分布 P_k

$$\begin{bmatrix} 0, & 1, & 2, & \cdots, & k, & \cdots \\ p_0, & p_1, & p_2, & \cdots, & p_k, & \cdots \end{bmatrix}$$

有一种连续分布 $G(x)$, 记

$$F_0(x) = \sum_{k=0}^{\infty} p_k [G(x)]^k \tag{3}$$

称 $F_0(x)$ 为由这两种分布构成的复合极值分布。其中, $n(n = 0,1,2,\cdots,k,\cdots)$ 为极端事件出现的频数, $p(n = k) = p_k$ 为该极端事件出现的概率。在实际问题中,常常是给定一个 $R(0 < R < 1)$, 求解方程 $F_0(x) = R$ 。通常记设计频率 $P = 1 - R$, 故重现期为

$$T = \frac{1}{P} = \frac{1}{1 - R} \tag{4}$$

若离散分布取泊松分布时,则复合极值分布的分布函数为

$$F_0(x) = \sum_{k=0}^{\infty} \frac{\lambda}{k!} e^{-\lambda} [G(x)]^k = e^{-\lambda[1-G(x)]} \tag{5}$$

其中 $k = 0,1,2,\cdots$, $G(x)$ 可为任意连续分布的分布函数,文中分别利用广义帕累托分布和韦伯分布与泊松分布复合,建立泊松-广义帕累托复合极值分布和泊松-韦伯复合极值分布,公式如下,泊松-广义帕累托复合极值分布:

$$F_0(x) = e^{-\lambda(1-\frac{x-\beta}{\alpha})^{\frac{1}{k}}} \quad k \neq 0, \beta \leq x \leq \frac{\alpha}{k} \tag{6}$$

泊松-韦伯复合极值分布:

$$F_0(x) = e^{-\lambda e^{-(\frac{x-\beta}{\beta})^\alpha}} \quad \alpha > 0, \beta > 0, \delta < x_{\min} \tag{7}$$

2 实例应用

2.1 目标函数的建立

目前的分布拟合检验,其基本思想都是用样本数据计算的频率与用特定理论分布计算的频率相比较,如相差较小,则认为样本数据取自某一特定的分布族;且相差越小,则认为可信度越高。因此构建离差绝对值之和为目标函数:

$$\min f(u) = \sum_{i=1}^n |F_i - P_i| \tag{8}$$

式中: F_i 为第 i 个样本值对应的理论频率,由式(6)和式(7)求的; P_i 为第 i 个样本值对应的经验频率。根据目标函数(8),利用 IEA 对复合极值分布参数寻优,搜索较优解。

2.2 算例

在实例应用时,除利用免疫进化算法求解参数外,还利用 L-矩估计法求解泊松-广义帕累托复合极值分布参数,用复合矩估计法求解泊松-韦伯复合极值分布参数,然后进行对比分析。

2.2.1 泊松-广义帕累托复合极值分布参数估计

以岳池站(站号:57414)为例,利用 1960 ~ 2010 年逐日降水观测资料,提取出极端降水序列(阈值 $\beta = 50 \text{ mm}$),利用泊松-广义帕累托复合极值分布进行模拟,分别用免疫进化算法和 L-矩估计法估算参数,并利用 Kolmogorov-Smirnov (K-S) 检验统计量(越小越优)、相关系数 R (越大越优)、最小离差平方和 OLS(越小越优)3 个指标来评价两种参数估计方法的优劣。免疫进化算法有关参数设置: $\sigma_1^0 = 3, \sigma_2^0 = 0.1, \sigma_3^0 = 1.5, A = 3, T = 300$, 初始群体规模为 200, 其计算结果见表 1。

表 1 泊松-广义帕累托复合极值分布的参数估计结果

参数	方法	
	L-矩估计法	免疫进化算法
λ	2.785	7.187
α	21.182	19.407
k	0.020	0.051
K-S 检验统计量	0.135	0.078
R	0.995	0.996
OLS	0.086	0.025

由表 1 可见,L-矩估计法估计参数的 K-S 检验统计量为 0.135,只通过了显著水平为 0.01 的检验(临界值为 0.137),免疫进化算法的 K-S 检验统计量为 0.078,通过了显著水平为 0.05 的检验(临界值为

0.113);并且,从 K-s 检验统计量、相关系数、最小平方离差和等指标来看,免疫进化算法都好于传统的 L-矩估计法。

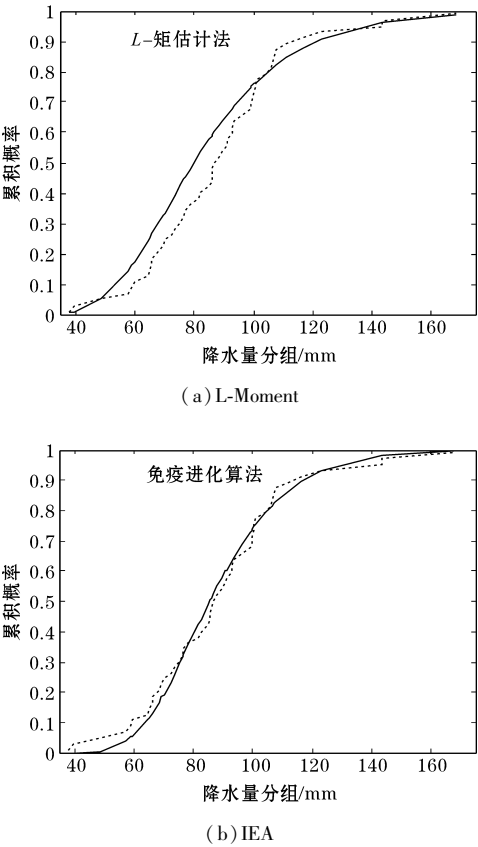


图1 累积频率分布曲线

图1是L-矩估计法和免疫进化算法所求参数的拟合效果图,从图中可以看出L-矩估计法所求参数在降水量分组数两端拟合较好,中间较差;免疫进化算法所求参数拟合效果优于L-矩估计法。

由上面分析可知,岳池站极端降水序列服从泊松-广义帕累托复合极值分布,则由式(6)可得岳池站设计暴雨量计算公式:

$$x_p = \beta + \frac{\alpha(1 - (-\frac{\ln(F)}{\lambda})^k)}{k}$$

(9)

式中, $F = 1 - p$, p 为设计频率,是重现期 T 的倒数($p = \frac{1}{T}$),将两组参数值带入式(9),即可求的重现期极值 x_p ,具体计算结果见表2。

表2 泊松-广义帕累托复合极值分布的设计频率分位值

参数估计方法 降水量/mm 重现期	2	5	10	20	50	100
L-矩估计法	79.05	102.13	117.13	131.31	149.36	162.66
免疫进化算法	85.42	104.79	117.02	128.32	142.33	152.41

表2为L-矩法与免疫进化算法估计的参数所求的不同重现期降水量极值,由表可见在较低重现期时(10年,5年),两者计算的结果基本一致,在重现期较大时,L-矩估计法计算结果比免疫进化算法结果更大,100年一遇大10.25mm。

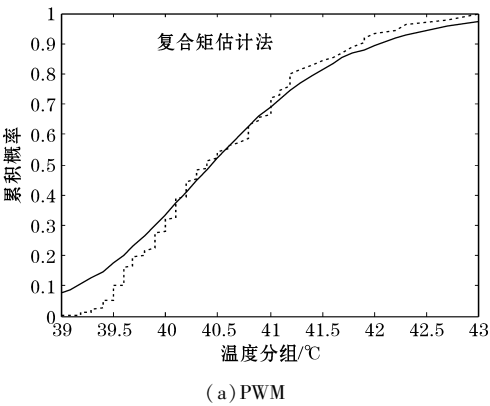
2.2.2 泊松-韦伯复合极值分布的参数估计

利用泊松-韦伯复合极值分布模拟重庆站(站号:57516)1951~2012年极端高温序列,分别采用复合矩估计法和免疫进化算法求解模型参数。极端高温序列采用超阈值取样,取极端高温的阈值为39℃,同样利用K-S检验统计量、相关系数 R 、最小离差平方和OLS 23个指标评价两种参数估计方法的优劣。免疫进化算法有关参数设置: $\sigma_1^0 = 3$, $\sigma_2^0 = 1.5$, $\sigma_3^0 = 1.5$, $\sigma_4^0 = 2$, $A = 3$, $T = 300$,初始群体规模为500,其计算结果见表3。

表3 泊松-韦伯复合极值分布的参数估计结果

参数	方法	
	复合矩估计法	免疫进化算法
λ	3.048	6.740
δ	38.699	38.464
β	1.277	1.164
α	1.263	0.964
K-S 检验统计量	0.100	0.047
R	0.996	0.997
OLS	0.073	0.028

由表3可见,复合矩估计法估计的参数对应的K-S检验统计量为0.100,只通过显著水平为0.01的检验(临界值为0.118),免疫进化算法的K-S检验统计量为0.047,通过了显著水平为0.05的检验(临界值为0.094);并且,免疫进化算法所求参数K-S检验统计量更小,相关系数更大,最小平方离差和更小,可见免疫进化算法所求参数更优。



(a) PWM

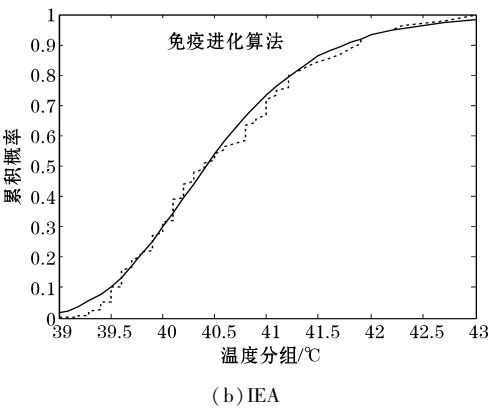


图 2 累积频率分布曲线

图 2 是复合矩估计法和免疫进化算法所求参数的拟合效果图,从图中可以看出复合矩估计法所求参数在温度分组数两端拟合效果较差,中间拟合效果较好;而免疫进化算法所求参数拟合效果明显要好于复合矩估计法。

由上面的分析可知,重庆站极端高温序列服从泊松-韦伯复合极值分布,则由式(7)可得重庆站温度设计极值计算公式:

$$x_p = \delta + \beta \left[-\ln\left(-\frac{\ln(F)}{\lambda}\right) \right]^{\frac{1}{\alpha}} \quad (10)$$

将两组参数值带入式(10),即可求的重现期极值 x_p ,具体计算结果见下表。

表 4 泊松-韦伯复合极值分布的设计频率分位值

参数估计方法 温度/°C 重现期	2	5	10	20	50	100
复合矩估计法	40.44	41.43	42.04	42.59	43.28	43.77
免疫进化算法	40.45	41.26	41.77	42.25	42.86	43.31

由表 4 可知,在重现期较小时,两种方法计算结果,相差较小;随着重现期的增大,复合矩估计法计算结果增大更快,50 年一遇值比免疫进化算法计算结果要大 0.42℃,100 年一遇值大 0.46℃。

3 结束语

(1)文中两个算例中,免疫进化算法所求参数拟合效果都较好,样本都通过显著水平为0.05的 K-S 检验,表明它应用于复合极值分布参数求解是合理和可行的;

(2)与 L-矩估计法及复合矩估计法相比,免疫进化算法所求参数 K-S 检验统计量更小,相关系数更大,最小平方离差和更小;从拟合效果图可知,免疫进化算法在整个样本区域拟合效果都较好,拟合效果优于其他两种方法。

(3)应用表明,免疫进化算法具有高效、普适的优点,在其他参数优化问题上具有广泛的应用前景。

致谢:感谢重庆市气象局青年基金项目(QNJJ-201507)对本文的资助

参考文献:

[1] IPCC. Climate Change The Science of Climate Change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change[M]. Cambridge, UK; New York, USA: Cambridge University Press, 1995: 1-572.

[2] IPCC. Climate Change The Science of Climate Change. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change[M]. Cambridge UK: Cambridge University Press, 2001:1-785.

[3] 国家防汛抗旱总指挥部办公室,水利部南京水文水资源研究所. 中国水旱灾害[M]. 北京:中国水利水电出版社, 1997.

[4] Pickands J. Statistical inference using extreme order statistics[J]. Ann Stat,1975,3:119-131.

[5] Hosking J R M, Wallis J R. Parameter and quantile estimation for the generalized pareto distribution[J]. Technometrics, 1987, 29: 339-349.

[6] 毛慧琴,杜尧东,宋丽莉. 广州短历时降水极值概率分布模型研究[J]. 气象, 2000, 30(10): 3-6.

[7] Liu T F, Ma F S, Prediction of Extreme Wave Heights and Wind Velocities[J]. Journal of the Waterway Port Coastal and Ocean Division. ASCE, 1980, 106.

[8] 马逢时,刘德辅. 复合极值分布理论及其应用[J]. 应用数学学报,1979,2(4):366-375.

[9] Langley R M , El-Shaarawi A H. On the calculation of extreme wave height: A review[J]. Ocean Engineering, 1986, 13(1) : 93 - 118.

[10] 刘德辅,王莉萍. 复合极值分布理论及其工程应用[J]. 中国海洋大学学报, 2004, 34(5): 892-902.

[11] W H Kirby, M E Moss. “Summary” of Flood - Frequency Analysis in The United States”, Journal of Hydrology,1987,96:5-14.

[12] 任照环,倪长健. 二维复合极值模型在暴雨多

- 变量联合分布中的应用[J]. 成都信息工程学院学报, 2014, 20(3): 59-65.
- [13] Liu T F. Long term distributions of Hurricane characteristics Pro[J]. Offshore Technology Conference, 1982, 4325:305-313.
- [14] 赵维谦, 张大错, 李辉权, 等. 复合极值分布的参数估计及其在工程设计中的应用[J]. 山东海洋学院学报, 1986, 16(3): 98-106.
- [15] 刘晶, 吴新荣, 李素红. Poisson-Gumbel 复合极值分布的参数估计[J]. 统计与决策, 2007, 18: 17-19.
- [16] 张香云, 程维虎. 二项-广义 Pareto 复合极值分布模型的统计推断[J]. 应用数学学报, 2012, 35(3): 560-572.
- [17] 倪长健, 丁晶, 李祚泳. 免疫进化算法[J]. 西南交通大学学报, 2003, 38(1): 87-91.
- [18] 倪长健, 王顺久, 丁晶. 免疫进化算法及其在水面曲线计算中的应用[J]. 水利水电技术, 2012, 12(33): 8-10.
- [19] 李祚泳, 汪嘉杨, 程会珍. 基于免疫进化算法优化的地下水水质评价普适公式[J]. 水科学进展, 2008, 19(5): 707-713.

Application of Immune Evolutionary Algorithm to Parameter Estimate of Compound Extreme Model

REN Zhao-huan¹, Ni Chang-jian²

(1. Nanchuan Meteorological Observatory, Nanchuan 408400, China; 2. College of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Parameter optimization is becoming one of the difficulties for the application of compound extreme model. To solve this problem, immune evolutionary algorithm (IEA) is proposed to optimize the parameter of compound extreme model, which was demonstrated to be effective using two concrete examples. The results show that: It is reasonable and feasible to optimize the parameter of compound extreme model with the proposed IEA; The fit model parameters calculated from IEA is better than that calculated from L-Moment and PWM; With the advantages of highly efficient and general, IEA is potential to be used in solving other parameters optimization problems.

Key words: meteorology; climatological statistics; immune evolutionary algorithm; compound extreme model; parameter estimation