

遗传算法在马斯京根模型参数估计中的应用

陆桂华¹ 酆建强¹ 杨晓华²

(1. 河海大学水文水资源及环境学院, 江苏 南京 210098; 2. 河海大学数学物理系, 江苏 南京 210098)

摘要:在马斯京根模型参数估计问题中, 引进了绝对残差绝对值和最小准则及相对残差绝对值和最小准则, 直接优选参数. 探讨了遗传算法的原理及本身参数优化特性, 总结了参数的简便设置技术, 并针对该法的寻优效果明显依赖于模型参数的初始变化区间的大小, 及可能会出现过早收敛的问题, 提出了自适应加速遗传算法. 通过实例应用, 进行了本文方法与传统优化方法的比较. 结果表明, 本文方法具有直观、简便、快速、适用性强等特点, 可广泛应用于各种优化问题中.

关键词:自适应遗传算法; 优秀个体; 马斯京根模型; 参数估计

中图分类号: P333 **文献标识码:** A **文章编号:** 1000-1980(2001)04-0009-04

河道洪水演算是根据河道上游水情推求下游水情的一种洪水预报方法. 至今为止, 马斯京根模型仍是一种常用的洪水演算方法, 它的基本方程是:

$$\left. \begin{aligned} \frac{dW}{dt} &= I - Q \\ W &= kQ' = k[xI + (1-x)Q] \end{aligned} \right\} \quad (1)$$

式中: W ——河段的槽蓄量, $\text{h} \cdot \text{m}^3/\text{s}$; t ——时间, h ; I, Q ——河段的入流量及出流量, m^3/s ; Q' ——示储流量, m^3/s ; x ——流量比重因子; k ——槽蓄系数, h . 式(1)的差分分解为

$$\left. \begin{aligned} \tilde{Q}(1) &= Q(1) \\ \tilde{Q}(i) &= aQ(i) + bK(i-1) + c\tilde{Q}(i-1) \quad i = 2, \dots, n \end{aligned} \right\} \quad (2)$$

式中: $\tilde{Q}(i), Q(i)$ ——第 i 个演算时段的出流量与实测出流量; $K(i)$ ——第 i 个演算时段的入流量; n ——演算时段个数; a, b, c ——演算时段系数, 满足:

$$a + b + c = 1.0 \quad (3)$$

式(2)和(3)组成了马斯京根模型流量演算公式. 马斯京根模型在实际应用中的一个重要问题是模型参数 x, k 或 a, b, c 的估计, 目前常用的方法有试错法、最小二乘法、改进图解法和矩法等. 其中最小二乘法因 x, k 交互作用, 使正规方程变为非线性方程组, 给参数的求解带来一定的困难. Jys. Wu 等 1985 年提出的改进图解法^[1], 虽能克服试错法中最优参数选择时的主观性, 但仍需作较多的运算. 文献[1]直接以演算出流量与实测出流量的离差平方和最小为优化准则函数来估计 a, b, c , 但它采用的是传统的非线性规化方法, 求得的是否是全局最优解没有理论根据.

马斯京根模型中的参数估计问题实际上是一个非线性优化问题. 基于生物界自然选择思想和自然遗传机制的遗传算法 (genetic algorithm) 简称 (GA), 是一种全局随机搜索优化算法, 它模拟自然界中生物从低级向高级的进化过程, 其主要优点是优化求解过程与梯度信息无关, 对于复杂的优化问题只需选择、杂交、变异 3 种遗传算子就能得到优化解, 它对问题是否线性、连续、可微等不作限制, 也不受优化变量数目、约束条件的束缚, 直接在优化准则函数引导下进行全局寻优. 基于这些显著优点, 遗传算法已引起人们的广泛应用和研究.

针对 GA 的计算量大、对变量取值范围大小变化适应性较差、在进化后期搜索效率较低及易发生过早收敛等问题, 本文提出了自适应加速遗传算法 (AAGA).

1 AAGA 的计算技术

不失一般性,设模型的参数优化问题为

$$\begin{aligned} \min f &= \sum_{i=1}^m \| F(c, x_i) - y_i \|^q \\ \text{s.t. } &a_j \leq c_j \leq b_j \quad (j = 1, 2, \dots, P) \end{aligned} \tag{4}$$

式中: c_j —— p 个待优化变量; x_i, y_i —— m 对模型输入、输出的观测数据; q ——任意常数.

自适应加速遗传算法包括 8 步. 步 1: 变量取值范围的二进制数编码. 步 2: 初始父代的随机生成. 步 3: 父代个体的适应能力评价. 步 4: 父代个体的概率选择. 步 5: 父代个体依概率 P_c 进行杂交, 得到两组子代个体. 步 6: 任取上步中的一组子代个体, 依概率 P_m (即变异率) 进行变异. 步 7: 进化迭代. 由上步得到的子代个体作为新的父代个体. 算法转入步 3, 进入下一次进化过程. 重新评价、选择、杂交、变异, 如此循环往复, 优秀个体将逼近最优点. 以上 7 步构成经典的遗传算法 (GA). 步 8: 加速循环. 笔者提出了把第一、第二次进化迭代所产生的优秀个体的变化范围作为变量的新取值范围, 算法转入步 1, 如此循环往复, 优秀个体的变化范围将逐步收缩, 与最优点的距离越来越近, 直至达到预定加速 (循环) 次数或最优个体的目标函数值小于某一设定值, 结束整个算法的运行.

以上 8 步构成加速遗传算法 (AGA). 它自身的参数有长度 e , 父代个体数目 n , 优秀个体数目 n_s 和变异率 P_m [详见文献 6]. 一般地, AGA 算法的杂交概率 P_c 、变异率 P_m 通常取常数. 这种算法收敛速度可能很快, 但可能会出现局部收敛 (尽管这种概率较小). 在文献 [4] 中给出了种群过早收敛程度的定量评价指标. 种群在进化过程中发生过早收敛的主要表现是: 种群内适应度暂时最大的一些个体相互重复或趋同, 且它们之间杂交后的子代也不会与父代有太大的变化, 这导致一般遗传算法搜索过早收敛. 因此, 要正确判断一个种群是否会发生过早收敛, 主要应看这个种群中当前适应度最大的那些个体是否重复或相互趋同. 本文引进 Δ 作为表征种群过早收敛程度的指标. 设第 t 代种群由个体 $x_t^1, x_t^2, \dots, x_t^N$ 构成, 适应度分别为 $f_t^1, f_t^2, \dots, f_t^N$, 个体总平均适应度

$$\bar{f}_t = \frac{1}{N} \sum_{i=1}^N f_t^i$$

最优个体的适应度为 $f_{\max}$, 将适应度大于 $\bar{f}_t$ 的个体的适应度再做平均得 $\bar{f}$, 定义 $\Delta = f_{\max} - \bar{f}$, 则指标 Δ 可用来表征种群的过早收敛程度. 在传统的遗传算法中, 杂交率 P_c 、变异率 P_m 等控制参数与种群进化过程无关, 从始至终都保持定值. 近年来的研究表明^[5], 控制参数对系统性能有重要影响, P_c 及 P_m 在进化过程中, 根据种群的实际情况随时调整大小. 本文把这一思想应用于加速遗传算法中, 使 P_c 和 P_m 随 Δ 的增大而减小, 提出了基于动态选择、变异操作的自适应加速遗传算法, 取杂交率和变异率分别为

$$P_c = 1.5 - \frac{1}{1 + \exp(-\Delta)} \quad P_m = 1.0 - \frac{1}{1 + \exp(-\Delta)}$$

2 应用实例

例 1 根据南运河称钩湾至临清河段某年的一次洪水过程资料^[1], 演算时段 Δt 取 12 h, 马斯京根模型参数估计问题可改写为 (取相对残差绝对值和准则):

$$\min F = \sum_{i=2}^n | [aK(i) + bK(i-1) + (1-a-b)\tilde{Q}(i-1) - Q(i)] / Q(i) | \tag{5}$$

或 (取绝对残差绝对值和准则):

$$\min F1 = \sum_{i=2}^n | aK(i) + bK(i-1) + (1-a-b)\tilde{Q}(i-1) - Q(i) | \tag{6}$$

$$\text{s.t. } g_1 : a \in [0, 1] \quad g_2 : b \in [0, 1] \quad g_3 : 1-a-b \in [0, 1] \tag{7}$$

现取优化准则函数为

$$\min f = F + h(g_3(a, b)) \tag{8}$$

或

$$\min f1 = F1 + h(g_3(a, b)) \tag{9}$$

式中 $h(g_3(a, b))$ 为罚项,当约束 g_3 满足时取值为 0,否则取值为 10^5 . 由约束 g_1, g_2 构成 a, b 的初始取值范围,取优化准则函数为(8),用 AAGA 求解,加速寻优 4 次,得近似解 $a = 0.473\ 116, b = 0.031\ 262$,结果见表 1、表 2. 有关文献^[1]的结果是根据文献[1]所给参数由笔者计算的. 可见,就演算出流量与实测出流量的绝对误差而言, AAGA 的结果优于文献[1]的非线性规化方法以及试错法、最小二乘法的结果,且 AAGA 历时仅 2 s. AAGA 法, $a = 0.473\ 1, b = 0.031\ 3$;非线性规化方法, $a = 0.426\ 5, b = 0.126\ 4$;最小二乘法, $a = 0.462\ 6, b = 0.084\ 3$;试错法^[1], $a = 0.285\ 7, b = 0.428\ 6$.

例 2 取文献[3]的一个例题,用 AAGA 求解,加速寻优 6 次,得近似解 $a = 0.129\ 579, b = 0.369\ 400$,其结果见表 3. 由表 3 可见就计算流量与实测出流量的绝对误差而言, AAGA 的结果优于文献[2]的最小面积法以及试错法. 文献[2]最小面积法求得公式(1)中的参数 $x = 0.123, k = 20.92$;试错法求得参数 $x = 0.10, k = 18$,且 AAGA 历时仅 2 s,优化速度大大提高.

另外,就例 1、例 2 的优化参数值是大于 0、小于 1 的正数来看,这样的优化参数也是合理的.

表 1 用 AAGA 估计马斯京根模型参数
Table 1 Parameters estimated by use of AAGA in the Muskingum Routing Model

加速次数	优秀个体各变量的变化范围		绝对残差绝对值之和 f_1
	a	b	
0	[0.000 000 1.000 000]	[0.000 000 1.000 000]	143.7
2	[0.408 629 0.479 438]	[0.004 858 0.163 275]	141.5
4	[0.464 385 0.477 164]	[0.019 874 0.047 031]	141.2

表 2 例 1 的各准则计算比较
Table 2 Comparison of criteria for example 1

误 差	AAGA	非线性规化方法	最小二乘法	试错法 ^[1]
平均绝对误差/($\text{m}^3 \cdot \text{s}^{-1}$)	4.87	5.18	5.63	7.40
平均相对误差/%	1.10	1.16	1.29	1.79
绝对残差绝对值之和/($\text{m}^3 \cdot \text{s}^{-1}$)	141.2	145.0	157.5	207.1

表 3 例 2 的各准则计算比较
Table 3 Comparison of criteria for example 2

误差	AAGA	文献[3]法	试错法 ^[3]
$\sum d^2$	29 877 780	31 225 100	221 564 000
$\sum d $	14 272	14 590	50 660

注 $d = \tilde{Q}(i) - Q(i)$.

3 结 语

本文提出马斯京根模型中参数估计的一种新方法——自适应加速遗传算法,分析了该算法控制参数的优化特性,总结出这些参数的简便设置技术.在马斯京根模型参数估计问题中,本文引进绝对残差绝对值和最小准则及相对残差绝对值和最小准则,直接优选流量演算系数,且结合遗传算法,探讨了遗传算法的原理步骤及算法本身的参数优化特性,针对该法的寻优效果明显依赖于模型参数的初始变化区间的大小,及可能会出现过早收敛的现象,提出了自适应加速遗传算法.自适应杂交算子及变异算子的引进在理论上是收敛的^[4],但自适应遗传算法可能运行时间很长,本文将自适应遗传算法第一、第二次进化迭代所产生的优秀个体的参数变化区间作为模型参数的新的初始变化区间,再重新运行,由此形成的自适应加速遗传算法大大缩短了寻优时间.文中给出了这一改进的遗传算法的详细步骤和简便的算法参数的设置技术,并在马斯京根模型参数优化问题中得到满意的全局最优解.

参考文献：

- [1] 瞿国静. 马斯京根模型参数估计方法探讨[J]. 水文 ,1977(3) :40 ~ 43 ,36.
- [2] 杨荣富. 马斯京根模型最优参数估计探讨[J]. 水文 ,1988(4) :18 ~ 20.
- [3] 吴浩扬 ,朱长纯 ,常炳国. 基于种群过早收敛程度定量分析的改进自适应遗传算法[J]. 西安交通大学学报 ,1999 ,33(11) :27 ~ 30.
- [4] 孙建永 ,申建中 ,徐宗本. 一类自适应遗传算法的理论分析与数值模拟[J]. 西安交通大学学报 ,2000 ,34(12) :84 ~ 87.
- [5] QI Xiao-feng. Theory analysis of evolution algorithms with an infinite population size in continuous space[J]. IEEE Trans Neural Network ,1994 ,5(1) :120 ~ 129.
- [6] 金菊良 ,杨晓华 ,储开凤. 加速基因算法在海洋环境预报中的应用[J]. 海洋环境科学 ,1997(4) :7 ~ 11.

Applications of Genetic Algorithm to Parameter Estimation in Muskingum Routing Model

LU Gui-hua¹ , LI Jian-qiang¹ , YANG Xiao-hua²

(1. College of Water Resources and Environment , Hohai Univ. , Nanjing 210098 , China ;

2. Department of Mathematics and Physics , Hohai Univ. , Nanjing 210098 , China)

Abstract The minimum criterion and absolute value of absolute errors and the minimum criterion and absolute value of relative errors are introduced into the direct optimum selection of parameters in the Muskingum Routing Model. The theory of genetic algorithm and the optimal characteristics of parameters are discussed ; the convenient configuration technique for the selection of parameters is summarized. In consideration of the facts that the characteristics of the optimal result apparently depends on the size of the initial variation region of parameters and too early convergence may appear , an accelerated adaptive genetic algorithm is proposed. Through application in practice , the present method is compared with other conventional methods , and the result shows that the method is objective , convenient , fast and highly-adaptive , and can be widely applied to various kinds of optimization.

Key words adaptive genetic algorithm ; superior individuals ; Muskingum routing model ; parameter estimation