

## 遗传算法的基因定位算子

熊伟清<sup>1</sup>, 刘明达<sup>2</sup>, 魏平<sup>1</sup>

(1. 宁波大学 计算机科学与技术研究所, 浙江 宁波 315211; 2. 北京大学 软件学院, 北京 102600)

**摘要:** 针对遗传算法局部搜索能力弱, 求解精度不高的缺陷提出了一个基因定位算子. 该算子的思路是进化一定代数(L)后通过对最优的若干个(N)染色体 基因位从高位到低位逐次进行比较, 如果当前的基因位都相同时便把该基因位确定下来, 以后的交叉、变异操作都不让该基因位参考, 随着算法的进行, 染色体基因便从高位到底位逐渐地确定下来. 其次, 通过在基因定位过程中引入模拟退火思想和小生境技术等局部搜索能力的算法, 提高该算子的全局优化能力. 最后, 通过几个非常容易陷入局部最优的测试函数测试表明几乎所有的峰值都得到了理论值.

**关键词:** 遗传算法; 小生境技术; 函数优化; 基因定位

**中图分类号:** TP18      **文献标识码:** A

## Gene-orientation operator for genetic algorithm

XIONG Wei-qing<sup>1</sup>, LIU Ming-da<sup>2</sup>, WEI Ping<sup>1</sup>

(1. Institute of Computer Science and Technology, Ningbo University, Ningbo Zhejiang 315211, China

2. Software School, Beijing University, Beijing 102600, China)

**Abstract:** A gene-orientation operator is proposed to solve the problem of weak ability and low precision in local searching in genetic algorithm. The operator makes a comparison among the several best (N) chromosomes from the high position to the low position after several generations (L). If all the genes on the current position happen to be the same, this gene is locked and not allowed to take part in the genetic operations such as crossover and mutation. In the process of algorithm, every gene of the chromosome will be determined gradually from the high position to the low position. Some algorithms with strong ability in local searching such as the Simulating Anneal and the Niche technology in the process of gene orientation are also used to improve the global optimizing ability of the operator. Finally, several test functions which are easy to fall into the local optimization are implemented to show that almost all the extremes reach their theoretical values.

**Key words:** genetic algorithm; niche technology; function optimization; gene orientation

### 1 引言(Introduction)

遗传算法(Genetic Algorithm, 简称 GA)由美国 Michigan 大学的 Holland 提出, 是模拟达尔文的遗传选择和自然淘汰的生物进化论的计算模型. 它作为一种有效的全局并行优化搜索工具, 具有简单、通用、鲁棒性强和适于并行分布处理的特点. 但遗传算法也会由于各种原因过早向目标函数的局部最优解收敛, 从而很难找到全局最优解. 其中有些是由于目标函数的特性造成的, 例如函数具有欺骗性, 不满足构造模块假说等等; 另外一些则是由于算法设计不当. 自从 Holland 系统的提出遗传算法的完整结构和理论以来, 众多学者一直致力于推动遗传算法的发展, 对编码方式、控制参数的确定、选择方式和交叉

机理进行了深入的探究提出了各种变形的遗传算法<sup>[1,2]</sup>.

针对遗传算法局部搜索能力不强, 大多数改进是利用函数的特征比如梯度信息、群体的情况等. 实际使用情况表明这样的改进针对不同的情况有效. 作者认为既然是采用染色体表示应该在染色体内部来考虑可能更合理. 对此, 本文为遗传算法提出一个基因定位算子, 通过函数优化证明适用的范围更广受到具体函数特征的影响小求解精度大大提高, 通过几个非常容易陷入局部最优的测试函数测试证明几乎所有的函数得到了理论值.

在大量的实际优化问题的求解计算中不仅要求在可行域内寻找全局最优解, 从而为决策者提供多

种选择或者多方面的信息,这类问题一般称为多峰搜索问题或者多模态函数优化问题.如何构造一种优化算法使之能够搜索到尽量多的或者全局最优解和有意义的局部最优解问题,已成为一个重要的研究方向<sup>[3]</sup>.基因定位算子与小生境计算的结合效果非常好,在下面的多模态峰函数求解测试中所有峰值均得到理论值.

## 2 基因定位算子的设计 (Design about the gene-orientation operator)

遗传算法有一个特点就是全局搜索能力很强而局部搜索能力不足,对于遗传算法解决函数优化问题来说它大体上都能找到最优点的近似值,但是想进一步得到精确值就非常的困难.

定义有向图  $G = (C, L)$ , 其中顶点集  $C$  为  $\{c_0(v_s), c_1(v_N^0), c_2(v_N^1), c_3(v_{N-1}^0), c_4(v_{N-1}^1), \dots, c_{2N-3}(v_2^0), c_{2N-2}(v_2^1), c_{2N-1}(v_1^0), c_{2N}(v_1^1)\}$ .  $v_s$  为起始顶点, 顶点  $v_j^0$  和  $v_j^1$  分别用于表示二进制码串中位  $b_j$  取值为 0 和 1 的状态. 有向弧集合  $L$  为

$$\{(v_s, v_N^0), (v_s, v_N^1), (v_N^0, v_{N-1}^0), (v_N^0, v_{N-1}^1), \dots, (v_2^0, v_1^0), (v_2^0, v_1^1), (v_2^1, v_1^0), (v_2^1, v_1^1)\}.$$

即对于  $j = 2, 3, \dots, N$ , 在所有顶点  $v_j^0$  和  $v_j^1$  处, 分别指向  $v_{j-1}^0$  和  $v_{j-1}^1$  的两条有向弧. 将二进制编码的问题求解转变为有向图的遍历. 为此本文在遗传算法中提出一个基因定位算子, 该算子的思路是进化一定代数 ( $L$ ) 后通过对最优的若干个 ( $N$ ) 染色体的基因位从高位到底位逐次进行比较, 如果当前的基因位都相同时便把该基因位确定下来, 以后的交叉、变异操作都不让该基因位参与, 随着算法的进行, 染色体基因便从高位到底位逐渐的确定下来. 另外, 当最低位也确定下来后算法便可结束了, 不必等到要满足其它的终止条件.

具体步骤为:

- 1) 统计相应的数值;
- 2) 个体按适应度排序;

3) if ( $gen > L$ ) 若按序选出进行比较的前  $N$  个体当前基因位都相同, 锁定该基因位, 要比较的基因位后移一位如果所有基因位都锁定 算法结束, 否则群体继续执行进化操作;

但纯粹运用基因定位算子, 虽然精度很高但通过实验发现有些函数容易陷入局部最优解, 因而提出了以下一些改进策略.

**改进 1** 利用算法计算速度快的优点让其一开始便进行基因定位运算, 使其迅速收敛于某一点, 记

录下来后重新开始让其收敛到另一点, 经过数次运算后比较各点大小得到一个总体最优点. 该算法与多种群进化有些类似, 但它能迅速收敛而结束当群体, 很快进行另一群体运算. 算法速度比多种群进化快得多, 精度也会有所提高.

多次试探染色体逐渐定位 (试探次数为  $\beta$ ). 其主要思想是产生初始种群马上进行染色体逐渐定位操作, 完成后再重复以上操作直到设定的试探次数, 最终记录最优个体.

具体步骤为

```
for( $i = 0; i < \beta; i ++$ )
{ 执行染色体逐渐定位算法;
  记录迄今为止产生的最好个体;
  报告当前状态及最优个体;
}
```

**改进 2** 与其他算法如 Tabu 搜索、模拟退火等局部搜索能力强的算法结合. 本文以与模拟退火思想结合为例说明. 采用当个体适应度平均值波动较大且退火温度较高时以很小的概率进行定位操作, 反之便以较大概率进行定位操作.

具体步骤为:

- 1) 初始化退火温度 ( $T$ );
- 2) 执行遗传算法的进化操作;
- 3) 用  $T$  控制的概率  $P$  执行纯染色体逐渐定位算法;
- 4) 记录迄今为止产生的最好个体;
- 5) 报告当前状态及最优个体;
- 6) 更新退火温度 ( $T$ );
- 7) 满足结束条件程序结束, 否则转到 2);

**改进 3** 与小生境技术的结合.

生物学上, 小生境是指环境中的一字组织功能. 在自然界中, 往往特征、性状相似的物种相聚在一起, 并在同类中交配繁衍后代. 小生境技术就是将每一代个体划分为若干类, 每个类中选出若干适用度较大的个体作为一个类的优秀代表组成一个种群, 再在种群以及不同种群之间通过杂交、变异产生新一代个体群, 同时采用预选择机制或排挤机制或分享机制完成选择操作.

基于小生境技术的遗传算法, 可以更好地保持种群的多样性, 同时具有很高的全局寻优能力和收敛速度, 特别适合于求解多模态函数优化问题. 小生境虽然有能力强求解多模态函数优化问题, 但精度不高, 而基因定位算子精度很高, 因而可以把两种思想结合, 称之为高精度多模态函数求解遗传算法.

具体步骤为:

1) 设置进化计数器  $t \leftarrow 0$ , 随机生成  $M$  个初始个体组成初始群体  $P(t)$ , 并且求出各个个体的适应度  $F_i (i = 1, 2, \dots, M)$ ;

2) 依据各个个体的适应度对其进行排序, 记忆前  $N$  个个体 ( $N < M$ );

3) 选择、交叉、变异运算;

4) 小生境淘汰运算: 将 3) 得到的  $M$  个个体和 2) 记忆的  $N$  个个体合并在一起求出两个个体之间的海明距离. 如果比较的两个个体的海明距离较小的便对其中适应度较低的个体处以罚函数  $F_{\min}(x_i, x_j) = \text{Penalty}$ ;

5) 依据  $M + N$  个个体的适应度进行排序, 记忆前  $N$  个个体;

6) 终止条件判断: 若不满足终止条件则  $t \leftarrow t + 1$ , 并将 5) 排序中的前  $M$  个个体作为新的下一代群体  $P(t)$ , 然后转到 3), 若满足终止条件则输出计算结果, 算法结束.

另外采用保留精英 GA 是依概率 1 收敛<sup>[4]</sup>, 在算法中还是需要的.

### 3 测试实验 (Test example)

为证明该算法是否有效, 选用了一下几个函数进行测试.

测试函数 1<sup>[1]</sup>

$$f_1 = 100(x_1^2 - x_2)^2 + (1 - x_1)^2, \\ -2.048 \leq x_1, x_2 \leq 2.048.$$

它具有一个全局也是唯一的极小点  $f(1, 1) = 0$ . 虽然在求极小值时它是单峰值的函数, 但它却是病态的, 难以进行全局极小化.

本算法得到的解为

$$f(1.000000, 1.000001) = 0.000000.$$

其中  $P_c = 0.9, P_m = 0.05$ , 编码长度 60, 种群规模 100, 进化代数 400.

测试函数 2<sup>[1]</sup>

$$f_2 = (4 - 2.1x_1^2 + \frac{1}{3}x_1^4)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2, \\ -3 \leq x_1 \leq x, -2 \leq x_2 \leq 2.$$

该函数共有 2 个全局最大值和 2 个全局最小值:  $f(3, 2) = f(-3, -2) = 162.9; f(-0.089842, 0.712656) = f(0.089842, -0.712656) = -1.031628$ ; 另外该函数还有较多的局部极值点, 如果算法的全局搜索能力不强会很容易的陷入局部极值点.

实验结果 求最大值时本算法得到的解为

$$f(-3.000000, -2.000000) = f(3.000000, 2.000000) =$$

162.900000; 求最小值时本算法得到的解为  $f(-0.0898, 0.7126) = f(0.0898, -0.7126) = -1.031628$ .

测试函数 3<sup>[1]</sup>

$$f_3 = \sum_{i=1}^5 i \cos[(i+1)x_1 + i] \cdot \sum_{i=1}^5 i \cos[(i+1)x_2 + i], \\ -10 \leq x_1, x_2 \leq 10.$$

该函数是一个多模态函数, 在其定义域内共有 760 个局部最小点, 其中的 18 个点是全球最小点  $f = -168.731$ , 因而该函数对于测试小生境遗传算法性能是非常适合的.

求最大值时, 本算法与小生境技术结合得到的解为

$$f(-0.800321, -0.800321) = \\ f(-7.083506, 5.482864) = \\ f(-0.800321, 5.482864) = \\ f(-0.800321, -7.083506) = \\ f(-7.083506, -7.083506) = \\ f(-7.083506, -0.800321) = \\ f(5.482864, 5.482864) = \\ f(5.482864, -0.800321) = \\ f(5.482864, -7.083506) = 210.482294.$$

求最小值时, 本算法和小生境技术结合得到的解为

$$f(-7.083506, -1.425128) = \\ f(-0.800321, -7.708314) = \\ f(-7.708314, -7.083506) = \\ f(4.858057, -0.800321) = \\ f(-7.083506, -7.708314) = \\ f(-7.708314, -0.800321) = \\ f(-7.708314, 5.482864) = \\ f(-0.800321, 4.858057) = \\ f(5.482864, -1.425128) = \\ f(4.858057, -7.083506) = \\ f(-7.083506, 4.858057) = \\ f(5.482864, 4.858057) = \\ f(5.482864, -7.708314) = \\ f(-1.425128, -7.083506) = \\ f(-1.425128, 5.482864) = \\ f(-1.425128, -0.800321) = \\ f(4.858057, 5.482864) = \\ f(-0.800321, -1.425128) = -186.730909.$$

其中  $P_c = 0.9, P_m = 0.05$ , 编码长度 100, 种群规模

100, 进化代数 1000.

**测试函数 4** 文献[5]构造的一类大海捞针问题

$$\max f(x, y) = \left( \frac{a}{b + (x^2 + y^2)} \right)^2 + (x^2 + y^2),$$

$$x, y \in [-5.12, 5.12].$$

其中:  $a = 3.0; b = 0.05; \max f(0, 0) = 3600$ ; 4个局部极值点为  $(-5.12, 5.12), (-5.12, -5.12), (5.12, -5.12)$  和  $(5.12, 5.12)$ ; 函数值为 2748.78. 这是一类全局最优解被次优解所包围, 隔断了模式的重组过程, 使得 GA 搜索长期陷入局部极值点, 随着参数  $a, b$  的变化, 该函数将形成不同严重程度的 GA 欺骗问题. 本算法得到  $f(0.000000000000, 0.000000000000) = 3600.000000000$ . 其中  $P_c = 0.9, P_m = 0.05$ , 编码长度 100, 种群规模 100, 进化代数 400.

**测试函数 5** Bobachevsky 函数<sup>[1]</sup>

$$x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7,$$

$$x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1)\cos(4\pi x_2) + 0.3,$$

$$x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1 + 4\pi x_2) + 0.3,$$

$$-50 \leq x_1, x_2 \leq 50.$$

3个函数在  $(x_1, x_2) = (0, 0)$  处有一个全局最小值 0.

通过实验均得到理论值. 其中  $P_c = 0.9, P_m = 0.05$ , 编码长度 100, 种群规模 100, 进化代数 1000, 就可得到本算法下面的值(理论值):

$$f(0.0000000045, -0.0000000045) = 0.000000000000,$$

$$f(0.000000848, 0.0000000274) = 0.000000000000,$$

$$f(0.0000000045, -0.0000000045) = 0.000000000000.$$

#### 4 结论(Conclusions)

从实验结果看基因定位算子无论是在求解精度上还是在算法的运行时间上, 比起基本算法改进都是非常明显的, 特别是与小生境技术结合构成的高精度多模态函数优化的遗传算法在所有的测试函数都得找到了所有的全局极值点, 且精度都达到相应

的理论值如对于函数  $f_3$  其全部 18 个全局最小点和 9 个全局最大点都找到, 而且都精确到了理论值, 这对于遗传算法局部搜索能力不强的弱点是一个非常大的改进.

基因定位算子提高求解精度是通过基因定位来实现的, 对问题的特征依赖程度要小, 具有好的通用性, 我们所用的测试函数均得到理论值也证明了这一点. 基因定位算子的提出, 使得各种局部搜索能力强的搜索算法如模拟退火、Tabu 搜索等算法, 融合在基因定位上而不是在种群选择上, 为各类领域搜索算法相互融合提供了一种接口方式.

#### 参考文献(References):

- [1] MICHALEWICZ Z. *Genetic Algorithms + Data Structures = Evolution Programs* [M]. Beijing: Science Press, 2000: 1 - 128.
- [2] MICHALEWICZ Z, FOGEL D B. *How to Solve It: Modern Heuristics* [M]. Beijing: China Waterpower Press, 2003: 126 - 139.
- [3] 李敏强, 寇纪松. 多模态函数优化的协同多群体遗传算法[J]. 自动化学报, 2002, 28(4): 497 - 504.  
(LI Minqiang, KOU Jisong. Coordinate multi-population genetic algorithms for multi-modal function optimization [J]. *Acta Automatica Sinica*, 2002, 28(4): 497 - 504.)
- [4] 张文修, 梁怡. 遗传算法的数学基础[M]. 西安: 西安交通大学出版社, 2000: 146 - 200.  
(ZHANG Wenxiu, LEUNG Yee. *Mathematical Foundation of Genetic Algorithms* [M]. Xi'an: Xi'an Jiaotong University Press, 2000: 146 - 200.)
- [5] 李敏强, 寇纪松. 遗传算法的模式欺骗性[J]. 中国科学(E辑), 2002, 32(1): 95 - 102.  
(LI Minqing, KOU Jisong. Genetic algorithms deceptive problem analysis [J]. *Science in China (Series E)*, 2002, 32(1): 95 - 102.)

#### 作者简介:

熊伟清 (1966—), 男, 副教授, 研究领域为人工智能、进化计算和软件工程, E-mail: Xiongweiqing@nbu.edu.cn;

刘明达 (1981—), 男, 研究生, 研究领域为进化计算、软件工程;

魏平 (1965—), 女, 副教授, 研究领域为人工智能、进化计算, E-mail: weiping@nbu.edu.cn.