

文章编号: 1002-0411(2001)05-470-04

一种多模态单亲遗传算法

余文 李人厚

(西安交通大学系统工程研究所 710049 西安)

摘要: 遗传算法在处理复杂的、多模态优化问题时常不十分有效, 很难同时搜索多个峰点. 这主要是由全局选择机制和交叉算子引起的. 针对上述不足, 本文提出了一种多模态单亲遗传算法, 目标不是发现一个最优解而是多个最优或次优解的集合. 主要是对交叉算子和选择机制作了改进, 群体中个体能较好地保留自己的遗传特性, 大大增强了种群个体的分散性. 该方法不仅易实现并行或分布计算, 且群体规模可以任意选取. 仿真结果验证了算法的有效性.

关键词: 单亲遗传; 多模态; 遗传算法

中图分类号: TP13

文献标识码: B

A MULTIMODAL PARTHENO-GENETIC ALGORITHM

YU Wen LI Ren-hou

(Xi'an Jiaotong University, Xi'an 710049 China)

Abstract: Often a classical GA is not flexible or effective for a optimizing application of complicated multimodal problems. It is difficult to reach more than one of the global maximum, simultaneously. Considering the above deficiency of GA, the paper presents a multimodal partheno-genetic algorithm (MPGA). The aim of the proposed algorithm is to find not only a single solution but also a set of the optimal. Both new a crossover operator and selection schemes are introduced so that every population individual inherits more genetic material of its own forefather. The diversity of population individual can be improved greatly in this approach. The size of population is very flexible and the architecture of computation is well suited for parallel implementation. The simulating cases show the effectiveness of this approach.

Keywords: partheno-genetic, multimodal problem, GA

1 引言(Introduction)

自 Holland 提出遗传算法(GA)以来, GA 以它善于处理传统搜索算法难于解决的优化问题和固有的并行特征等优点, 在众多工程应用领域, 得到了广泛的研究和应用^[1]. 但是, 对于复杂的、多模态优化搜索问题, GA 存在着明显的不足: 往往难以发现多个最优或次优的峰点, 甚至会陷入局部最优或成熟前收敛. 因此, 解决这类问题, 通常需要对 GA 进行改进. 常见的方法有: 由 Goldberg^[3]等首先提出的基于“共享函数”的途径; 另一个是基于并行处理的方法^[1]. 一般来说, 前种方法需较多的计算代价, 而后种方法在实现中会遇到大量的通讯开销^[2]. 在某种程度上, 两种方法都需要先验知识. 一般而言, 这种计算花费或通讯开销每代都有, 当群体规模或子群划分较大时, 会导致系统效率十分低下.

作者受生物 DNA 分子结构模型和繁殖机理的启示, 提出了一种多模态单亲遗传算法(MPGA). 在 MPGA 中, 每个个体与它自身的一条互补单链配对成一个 DNA 双链分子, DNA 分子的双链在繁殖过程中可以进行遗传信息的局部交换, 并将交换后的结果遗传给他的后代. 文[6]论证了这种特定的(互补)交换方式能够使个体具有全局搜索性能. 通常, 每个单亲个体允许通过多次交叉操作, 繁殖产生多个不同的后代, 单亲及它的后代属于同一个家庭. 通过选择, 每个家庭的一个成员能够成为下代的种群个体. 这种单系繁殖和遗传方式一个优点就是它几乎不受其它种群个体的影响, 能够使每个个体保持自己的遗传特性, 以致使种群个体具有多模式或多样性, 不仅容易搜索到全局最优点, 而且还可同时发现多个有效的模态峰点.

2 基本概念和定义(Fundamental concepts and definitions)

定义 1 一条 DNA 单链(单亲), 与它的另一条互补链发生遗传信息的局部交换并生成新的后代的过程, 称为繁殖. 单亲及其所繁殖的后代共同组成一个家庭.

定义 2 个体 x 称做是优良的, 如果满足 $f(x) \geq \bar{f}$, 其中 $f(x)$ 是 x 的适应度, \bar{f} 是其单亲的适应度.

定义 3 个体 x 称做最佳个体, 如果 x 是家庭中最优的个体.

定义 4 MPGA 中, 对于多模态的优化问题, 称 M 为最优集, 为 B 满意集, 如果 M 是全局最优点的集合. B 是局部极值点的集合.

定义 5 种群个体根据其适应度的值, 分配一个允许其繁殖后代的个数, 称为繁殖指标.

定义 6 MPGA 中, 记 $\{X(t)\} \equiv \{x_1(t), x_2(t), \dots, x_n(t)\}$ (n 为群体规模) 为一群体转移序列, 称 $\{X(t)\}$ 分别为概率(弱)收敛、处处概率(弱)收敛和几乎完全(弱)收敛, 如果分别满足条件 1)、2)、3):

- 1) $\exists x_i(t) \lim_{t \rightarrow \infty} P\{x_i(t) \in M(B)\} = 1$ (1)
- 2) $\forall x_i(t) \lim_{t \rightarrow \infty} P(x_i(t) \in M(B)) = 1$ (2)
- 3) $P\{\lim_{t \rightarrow \infty} \{X(t)\} \subset M(B)\} = 1$ (3)

3 多模态单亲遗传算法 (Multimodal partheno-genetic algorithm)

3.1 算法描述

算法步骤如下:

Step 1 随机产生初始遗传种群, 置 $t=0$.

Step 2 确定各初始遗传个体的适应度.

Step 3 重复下列操作直到满足停机准则.

- 1) 确定每个种群个体的繁殖指标.
- 2) 通过繁殖产生各种群个体的家庭(成员).
- 3) 计算适应度并选出各家庭中的遗传个体.
- 4) 置 $t=t+1$.

Step 4 结束

群体中个体依据其繁殖指标繁殖产生各自的家庭(见图 1), 再经各家庭内部选择, 构成新的下代种群. 图 2 是一单亲个体通过多次交叉操作所繁殖的后代(子链).

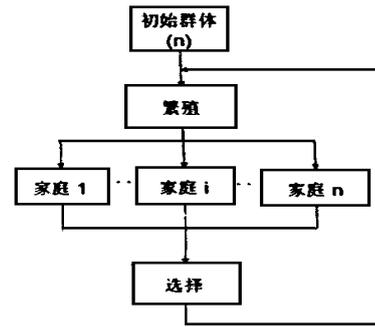


图 1 算法模块图

Fig. 1 Block diagram of the algorithm

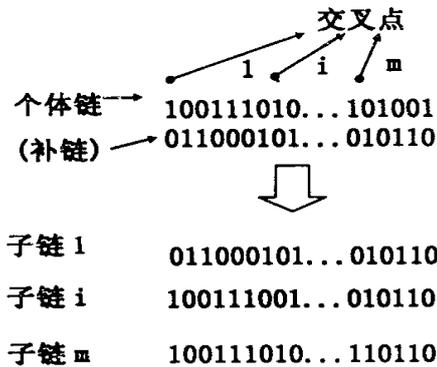


图 2 交叉操作实例

Fig. 2 Examples of crossover operation

3.2 繁殖和选择策略

3.2.1 繁殖指标

在种群个体的单亲进化中, 作者试图通过引入“繁殖指标”, 使得群体个体的演化相互关联. 繁殖指标是与个体适应度相关的动态指标, 其核心是鼓励适应环境的个体繁殖较多的后代. 从另一方面, 由于优良的个体更难繁殖出更优的个体, 因而需要较多的后代以供选择. 繁殖指标的一般表示形式为

$$m = \text{int}[g(s)] \tag{4}$$

其中 $g(s)$ 称为繁殖函数, $s=f(x)$ 代表个体 x 的适应度. $\text{int}[\]$ 代表取整函数.

3.2.2 选择机制

从每个家庭中选出一个代表作为下代遗传个体, 通常可采用三种不同的选择策略.

(1) 最佳个体选择法

由当前各家庭中的最佳个体, 组成下代种群.

(2) 优良选择法

从各家庭的优良个体中非确定性地选出一个代表, 组成下代群体.

(3) 最优值保留的比例选择法

按比例法从各家庭成员中选出各自的遗传个体, 但保留各家庭中曾出现过的最优个体.

4 收敛性分析(Convergence analysis)

4.1 预备知识

设 A 是一个集合, $R \subseteq A \times A$ 是某种关系, 下面是几种特殊二元关系的定义

- (1) 自反关系: 若 $\forall a \in A \Rightarrow aRa$
- (2) 传递关系: 若 $aRb, bRc \Rightarrow aRc$
- (3) 反对称关系: 若 $aRb, bRa \Rightarrow a = b$
- (4) 偏序关系: 若关系是自反、传递和反对称关系; 此时 A 称为由 R 规定的偏序集合.

(5) 全序关系: 若 R 是偏序关系, 且 $\forall a, b \in A$, 要么 aRb , 要么 bRa .

(6) 最大元: 若 $b \in A$, 对任意 $x \in A$, 且 $bRx \Rightarrow x = b$, 则 b 称为 A 的一个最大元.

Zorn 引理[7] 设 A 是一个非空偏序集合, 如果 A 的每个全序子集都有上界, 则 A 至少含有一个最大元.

4.2 收敛性定理

传统 GA 常会遇到成熟前收敛, 甚至收敛点并不是局部极值点, MPGA 不会出现这种情况. MPGA 至少会收敛到局部极值点, 即本文定义的“满意集”. 针对多模态的优化问题, 我们把这种收敛称为“弱”收敛, 前面几个收敛性定义之间的关系由下面结论给出.

定理 1 概率(弱)收敛、处处概率(弱)收敛和几乎完全(弱)收敛的一个关系是几乎完全(弱)收敛 \Rightarrow 处处概率(弱)收敛 \Rightarrow 概率(弱)收敛

文献[5]中对 GA 各种收敛性定义有专门讨论. 也可以依定义直观地证明上面结论.

定理 2 对多模态的优化问题

(1) 按优良或最佳个体选择. MPGA 几乎完全弱收敛.

(2) 按比例优选择法且保留最佳个体. MPGA 几乎完全收敛.

证明 按优良或最佳个体选择, 对任意时刻的单系遗传个体 i (记作 $x_i(t)$), 其适应度始终随时间(代)增加, 即有界序列 $\{f(x_i(t))\}$ 是一个 t 的单调序列. 因此, 极限值(记 z)存在.

记 $A = \{y: y = f(x_i(t)), (i = 0, 1, 2, \dots)\}$, A 是由偏序关系“ \leq ”上的一个偏序集合, 且它的每个全序子集均有上限, 由引理: A 有一个最大元素, 即有: $z = \lim_{t \rightarrow \infty} f(x_i(t))$.

又因 $z \in A \Rightarrow \exists N > 0$, 使 $z = f(x_i(N))$; 考虑到序列 $\{f(x_i(t))\}$ 的单调性 $\Rightarrow f(x_i(j)) = z$, 对 $\forall j \geq$

N

故 $X = \{x; x = x_i(j), j \geq N\}$, 以下证明: 对任意 $x \in X$, 有 $x \in B$

(以一维情形编码为例) 令 x 的编码为 $* \dots * ab \dots bb$, 其中, $a, b \in \{0, 1\}$, $k \geq 0$, $*$ 为通配符. 显然, 在最末 $k+1$ 位互补交叉操作能产生一个后代 $* \dots * ba \dots aa$, 而对最后一位互补交叉能得到另一个后代 $* \dots * ab \dots b$, 而这两点都是 x 的最小邻域中点的编码, 即 x 的相邻点(记作 x_+ 和 x_-), 而 $f(x) = z$, 也就是说

$$f(x) \geq f(x_+) \text{ 和 } f(x) \geq f(x_-)$$

(否则 z 不是 A 的最大元)

即 x 是一个局部极值点. 由 x (及 i) 的任意性, $\Rightarrow X \subset B$, MPGA 几乎完全弱收敛.

对于多维的数串, 只要对交叉操作略做扩展, 结论依然成立(限于篇幅, 从略).

当按比例优选择法且保留最佳个体时当按精英保留选择法, 演化序列是 $\{x_i(t)\}$ 都具有遍历性, 因而极限状态所保留的个体, 必属于最优解集, 故结论成立.

5 实例计算(Simulating cases)

通过对下面两个二维多模态函数的实例计算, 以说明算法的特点和计算能力

$$(1) f_1 = \cos(2\pi \cdot x_1) \cos(2\pi \cdot x_2) e^{-(x_1^2 + x_2^2)/10} - 1 \leq x_1, x_2 \leq 1 \quad (5)$$

$$(2) f_2 = 0.002 + \sum_{j=1}^{25} \frac{1}{j + \sum_i (x_i - a_{ij})^6} - 65.5 \leq x_1, x_2 \leq 65 \quad (6)$$

函数 f_1 定义域中心有 13 个起伏变化、适应度相近的模式峰点. 最优点在 $(0, 0)$, 适应度为 1. 函数 f_2 是著名的多模态测试函数, 有 25 个局部峰点. 简单遗传算法很难同时搜索到这两个函数的多个峰点, 并容易陷入局部模态. 两个函数的图形分别见图 3 及图 4. 本算法对它的测试中, 每个变量的编码长度为 16. 采用最佳个体选择, 随机产生初始群体, 算法收敛后, 群体中包含多个相异的局部模态. 当群体规模从 30 逐渐增大到 100 时, 函数 f_1 的 13 个模态点几乎已能够同时找出; 函数 f_2 的模态点也随群体规模增加而增加. 10 次测试的均值结果见图 5. 最大进化代数为 100. 繁殖函数选取如下

$$m(x) = \text{int}[4.0e^{f(x)}] \quad (7)$$

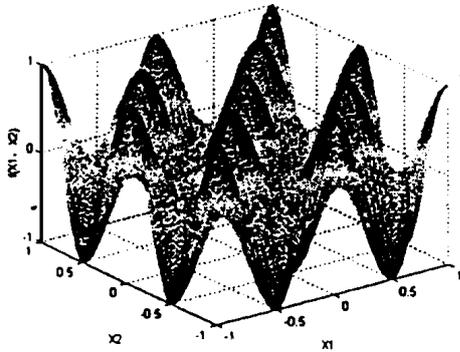


图 3 函数 $f_1(x_1, x_2)$ 的图形

Fig. 3 Graphics of function $f_1(x_1, x_2)$

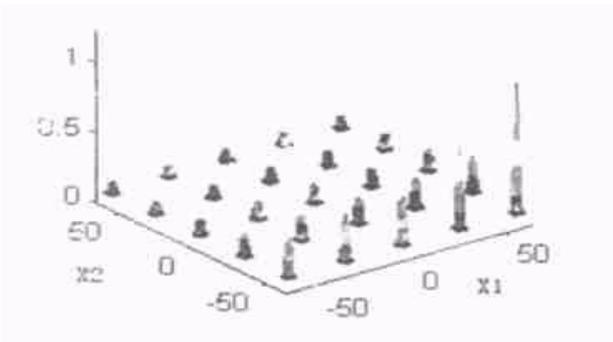


图 4 函数 $f_2(x_1, x_2)$ 的图形

Fig. 4 Graphics of function $f_2(x_1, x_2)$

6 结论(Conclusion)

本文通过对遗传算法操作算子和选择策略的改进, 提出一种多模态单亲遗传算法(MPGA). 目标不是为发现一个最优模态点而是发现多个最优或次优模态点. 该算法方法简便、易于实现. 适合并行或分布处理, 群体规模能灵活选择. 不担心陷入“成熟前”收敛问题. 它为处理复杂的多模态优化问题提供了一个选择的途径. 仿真计算验证了算法的有效性.

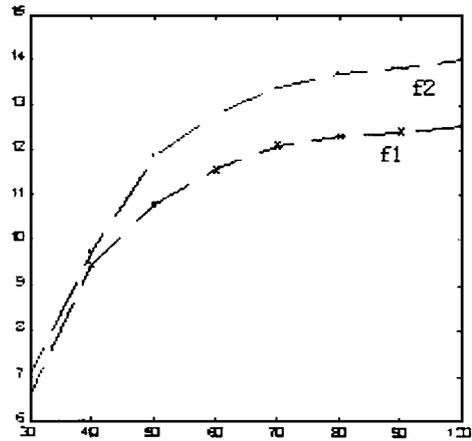


图 5 模态点数量与群体规模的关系

Fig. 5 Relations between modal number achieved and population size

参 考 文 献 (References)

- 1 Parallel Genetic Algorithms theory and Applications, ed. by J. Stender, Amsterdam IOS Press, 1993
- 2 Abramson D, Mills G, Perkins S. Parallelisation of a Genetic Algorithm for the Computation of Efficient Train Schedules. In: Proceeding of the 1993 Parallel Computing and Transputers Conference, 1993. 139~ 149
- 3 Goldberg D E, Richardson J. Genetic Algorithms with Sharing for Multimodal Function Optimization, Proceedings of the Second International Conference on Genetic Algorithms 1987, 41~ 49
- 4 刘国良等编著. 遗传算法及其应用, 人民邮电出版社, 1996
- 5 张文修, 梁怡编著. 遗传算法的数学基础. 西安交通大学出版社, 2000
- 6 孟祥武, 张玉洁. 遗传算法交换操作的遍历性. 北京邮电大学学报, 1999, 22(1): 34~ 37
- 7 韩崇昭, 胡保生编著. 泛函分析极其在自动控制中的应用. 西安交通大学出版社, 1990

作者简介

余 文(1964-), 男, 博士生, 毕业于兰州大学数学力学系, 获硕士学位. 研究领域为复杂系统智能控制理论和方法, 遗传算法等.

李人厚(1935-), 男, 博士生导师. 研究领域为系统智能控制理论和方法, 优化算法等.