

改进的模糊交叉算子及其在 CGA 中的应用

颜 颖, 缙 锦

(华侨大学信息科学与工程学院, 泉州 362021)

摘 要: 基于标准化适应值信息, 提出改进的模糊交叉算子, 并应用到细胞状遗传算法(CGA)中。在具有局部搜索倾向的交叉操作中, 该算子能使后代更偏向于适应值高的父体。在具有全局搜索倾向的交叉操作中, 能使较差个体在更大范围内进行搜索, 有效地引导 CGA 算法向全局最优解的方向收敛。仿真实验结果表明, 基于改进模糊交叉算子的 CGA 算法性能更好。

关键词: 模糊交叉算子; 多峰分布; 三角概率分布; 细胞状遗传算法

Advanced Fuzzy Recombination Operator and Its Application in CGA

YAN Ying, GOU Jin

(College of Information Science and Engineering, Huaqiao University, Quanzhou 362021)

【Abstract】 An advanced fuzzy recombination operator named SFFRO is proposed based on standardized fitness and applied to Cellular Genetic Algorithm(CGA). The exploitative SFFRO has much more probability to generate offspring closer to the parent with higher fitness, and in the other hand, the explorative SFFRO tends to search in a larger scale for the parent with lower fitness. Therefore, SFFRO indicates the potential search direction and accelerates the convergence to global optimum. In the simulation research, experimental results show that CGA based on SFFRO obviously outperforms others in terms of efficiency and reliability.

【Key words】 fuzzy recombination operator; multimodal distribution; triangular probability distribution; cellular genetic algorithm

1 概述

遗传算法作为一种随机化搜索算法, 以获取最优解或准最优解为目的, 其行为很大程度上取决于全局搜索与局部搜索间的平衡。交叉算子是最主要的遗传算子之一, 其根本作用是使子代继承父代的优秀基因, 提高算法的搜索能力, 对种群的搜索性能起着重要作用。

不良的交叉算子和参数设置会影响选择压力、选择噪音和模式干扰, 而造成关键等位基因的缺失, 使“全局搜索/局部搜索”不成比例, 导致种群多样性不足, 使搜索易于陷入不包含全局最优解的区域, 即“过早收敛”。

为了解决上述问题, 使算法收敛到全局最优解, 文献[1]结合模糊集理论提出了模糊交叉算子(Fuzzy Recombination Operator, FRO), 由一个双峰分布决定特定后代基因出现的概率。文献[2]在其基础上提出了扩展的模糊交叉算子(Extended Fuzzy Recombination Operator, EFRO), 将基因的活动区间划分为全局搜索区和局部搜索区, 并将后代基因出现的概率改由多峰(3 峰)分布来决定, 通过参数调节, 使交叉算子产生不同的全局搜索度或局部搜索度, 从而提高算法的搜索能力。

但是, 上述两种算子都不包含任何父体适应值信息, 无法根据父体间的适应值差异, 为生成更优质子代提示潜在的搜索方向。为了更好地利用父代的适应值信息, 使适应值更高的父体特征更多地保留在子代中, 本文提出了基于标准化适应值的模糊交叉算子。

2 改进的模糊交叉算子

假设 $X=(x_1, x_2, \dots, x_n)$ 和 $Y=(y_1, y_2, \dots, y_n)$ 是两个被选择用于交叉操作的父体, 其中, $x_i, y_i \in [a_i, b_i] \subset R, i=1, 2, \dots, n, n$ 是个

体所含的基因个数。首先, 对父体 X 和 Y 的适应值 F_x 和 F_y 进行极差正规化处理, 使处理后得到的 $F'_x, F'_y \in [0, 1]$, 且与量纲无关。这是为了避免种群演化过程中的适应值标度变换、两个父体间适应值差异过大以及负适应值带来的算子失效。不妨设 $F'_y > F'_x$, 则交叉后代中第 i 个基因具有值 z_i 的概率由多峰分布 $p(z_i) = \{\varphi(x_i), \varphi(av), \varphi(y_i)\}$ 给出, 其中, $\varphi(x_i), \varphi(av), \varphi(y_i)$ 为三角概率分布, 特性如表 1 所示。

表 1 三角概率分布 $\varphi(x_i), \varphi(av), \varphi(y_i)$ 的特性

分布	最小值	典型值	最大值
$\varphi(x_i)$	$\max\{x_i - e \cdot d_x, a_i\}$	x_i	$\min\{x_i + e \cdot d_x, b_i\}$
$\varphi(av)$	$\min\{x_i + e \cdot d_x, av\}$	av	$\max\{y_i - e \cdot d_y, av\}$
$\varphi(y_i)$	$\max\{y_i - e \cdot d_y, av\}$	y_i	$\min\{y_i + e \cdot d_y, b_i\}$

其中,

$$e \in [0, 1] \quad (1)$$

$$I = |y_i - x_i| \quad (2)$$

$$d_x = \frac{2F'_x}{F'_x + F'_y} \quad (3)$$

$$d_y = 2 - d_x \quad (4)$$

$$av = x_i + \frac{F'_y}{F'_x + F'_y} \cdot I \quad (5)$$

基金项目: 国家自然科学基金资助项目(60573056); 福建省自然科学基金资助项目(A0710013); 福建省青年人才基金资助项目(2006F3085)

作者简介: 颜 颖(1980 -), 女, 助教、硕士, 主研方向: 并行演化计算, 神经网络; 缙 锦, 讲师、博士

收稿日期: 2007-08-15 **E-mail:** yanying@hqu.edu.cn

结合表 1 及式(1)~式(5)观察上述定义,可以发现 SFFRO 算子具有以下几个重要特性:

(1)可以通过参数 e 调节算法进行全局搜索或局部搜索,这与 EFRO 算子类似,而 FRO 算子则无此特性。当 $0 < e < 0.5$ 时,后代基因 z_i 属于局部搜索区域 $[x_i, y_i]$ 的概率高于落入全局搜索区域 $[a_i, x_i]$ 和 $[y_i, b_i]$ 的概率,此时算法具有局部搜索倾向,如图 1(a)所示;当 $0.5 < e < 1$ 时,则产生相反的效果,算法趋向于全局搜索,如图 1(b)所示。

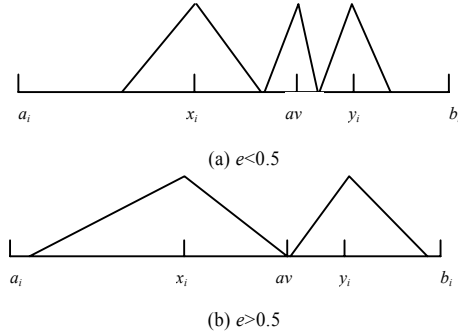


图 1 SFFRO 算子的 2 个实例

(2)根据式(3)~式(5),SFFRO算子在 d_x , d_y 和 av 中引入了标准化适应值信息。由于 $F'_y = F'_x$ ($F'_x, F'_y \in [0, 1]$), 因此可得 $d_x = d_y$, 且 $d_x \in [0, 1]$, $d_y \in [1, 2]$ 。更进一步分析可以发现,在具有局部搜索倾向的交叉操作中,如图 1(a)所示,当 $e < 0.25$ 时,有 $x_i + e \cdot d_y \cdot I < av$, $y_i - e \cdot d_x \cdot I > av$ 且 $av > x_i + I/2$, 因此,后代基因 z_i 落在 y_i 附近的概率高于落在 x_i 附近的概率,即 z_i 在局部搜索区间 $[x_i, y_i]$ 中会更偏向于适应值高的父体,从而保留了较优父体的更多特征。相比之下,由于 EFRO 算子中并不包含父体适应值信息,且 $av = (x_i + y_i)/2$, 因此 z_i 在局部搜索区间 $[x_i, y_i]$ 中落在 y_i 附近的概率与落在 x_i 附近的概率相等,无法根据父体的适应值信息,更深入地引导潜在的搜索方向。

(3)与特性(2)相对应,在具有全局搜索倾向的交叉操作中,如图 1(b)所示,SFFRO算子产生的后代基因 z_i 在 $[a_i, x_i]$ 区间中远离 x_i 的概率高于其在 $[y_i, b_i]$ 区间中远离 y_i 的概率,使得适应值较差的个体比适应值较高的个体能够在更大范围内进行全局搜索,从而在更大程度上得到改善。而分析 EFRO 算子,则不具备这样的特性。

综上所述,SFFRO 算子引入了父体适应值信息,在引导潜在搜索方向上更具优势。不仅能够通过参数 e 的调节很好地控制种群全局搜索和局部搜索的平衡,而且能够根据交叉父体的适应值信息对搜索的方向和强度进行自适应调节,从而更有效地引导算法向全局最优解收敛。

3 在 CGA 算法中的应用

CGA 算法是一种结构化的遗传算法,其种群具有特定的空间结构,能够通过更多的迭代来获取适应性和基因型的多样性。研究证实,CGA 算法在处理复杂优化任务时优于其它的遗传算法模型。此外,该算法与细胞自动机的相似性以及局部搜索嵌入算法的适用性使它极具研究价值。因此,本文将 SFFRO 算子应用到 CGA 算法中,并在仿真实验中与 FRO, EFRO 进行性能比较。

3.1 SFFRO 算子参数的自适应调节

SFFRO 算子不仅能够很好地平衡全局搜索和局部搜索,而且能够根据交叉父体的适应值信息更深入有效地引导搜索方向和搜索强度。这种能力是通过参数 e 的自适应调节来

实现的。

在演化过程中,如果算法的收敛速度衰退到一个给定的阈值以下,则说明该种群很可能已经陷入了某种局部最优无法跳出,此时就需要在 $(0.5, 1]$ 的范围内增大参数 e 的值,以增强全局搜索的范围和强度。由于 SFFRO 算子中引入了父体适应值信息,适应值较差的个体将比适应值较高的个体在更大范围内进行全局搜索,从而使算法能够有效地避开局部最优解,继续向全局最优解收敛。

相反,如果算法的收敛速度提高到某个阈值以上,则说明该种群目前的演化方向有值得开发的潜力,因此,就可以在 $[0, 0.25)$ 的范围内减小参数 e 的值,以增强算法局部搜索的能力。在 $[0, 0.25)$ 的范围内,由 SFFRO 算子生成的后代更偏向于适应值高的父体,因此,能够更有效地引导算法朝潜在的搜索方向收敛。

在 SFFRO 算子自适应调节的基本原理中,需要判定的一个前提是算法的收敛速度。本文采用的是基于平均适应值(Average Fitness, AF)的判定准则,它对收敛速度的度量建立在表现型多样性的基础上,并且能够以较低的量度代价指示当前的收敛速度,计算复杂度为 $O(u)$, 其中 u 是种群的大小。假设种群在第 t 代的平均适应值为 \bar{f}_t , 则第 t 代平均适应值与第 $t-1$ 代平均适应值的差异为 $\Delta \bar{f}_t = \bar{f}_t - \bar{f}_{t-1}$ 。如果 2 个连续演化代的差异 $\Delta \bar{f}_t$ 下降了至少 ε 倍,即 $\Delta \bar{f}_t - \Delta \bar{f}_{t-1} < \varepsilon \cdot \Delta \bar{f}_{t-1}$, 则判定种群收敛速度过慢。如果差异增大了至少 $(1-\varepsilon)$ 倍,即 $\Delta \bar{f}_t - \Delta \bar{f}_{t-1} > (1-\varepsilon) \cdot \Delta \bar{f}_{t-1}$, 则判定种群的收敛速度达到阈值。

根据 AF 准则,SFFRO 算子中参数 e 自适应调节的算法用伪代码描述如下:

```
(1) if (  $\bar{f}_t < (1 + \varepsilon) \bar{f}_{t-1}$  )
(2)  $e \leftarrow e\_global$ ;
(3) else if (  $\bar{f}_t > (2 - \varepsilon) \bar{f}_{t-1}$  )
(4)  $e \leftarrow e\_local$ ;
(5) else
(6) 保持参数  $e$  的值不变
(7) end if
```

3.2 基于 SFFRO 算子的 CGA 算法

描述了 SFFRO 算子参数 e 进行自适应调节的原理和算法后,本文将 SFFRO 算子应用到 CGA 算法中。假设 x, y 为种群区域的横坐标和纵坐标, (x, y) 表示种群拓扑中横坐标为 x 、纵坐标为 y 的一个位置, Width, Height 分别表示该种群区域的宽度和高度(以个体数为单位),则基于 SFFRO 算子的 CGA 算法步骤用伪代码描述如下:

```
(1) 种群、参数初始化;
(2) 评价初始种群;
(3) 对初始种群个体适应值进行极差正规化处理;
(4) while(不满足停机条件)
(5) SFFRO 算子参数  $e$  自适应调节;
(6) for  $x \leftarrow 0$  to Width-1 do
(7) for  $y \leftarrow 0$  to Height-1 do
(8) 以位于  $(x, y)$  的个体作为一个交叉父体;
(9) 计算该个体的邻域;
(10) 从邻域中选取一个个体作为另一交叉父体;
(11) 使用 SFFRO 算子进行交叉操作;
(12) 交叉后代变异操作,并放入临时种群;
(13) end for
(14) end for
```

- (15)评价临时子种群，更新原种群；
- (16)评价更新后的种群；
- (17)定位当前种群的最优个体；
- (18)end while。

在算法第(1)行种群初始化中采用的种群拓扑结构是二维超环面栅格。第(2)行和第(16)行中评价种群的主要的操作，是计算当前演化代种群的平均适应值以及所有个体的适应值。第(9)行计算个体邻域时采用的邻居关系模型为 NEWS 模型。第(10)行将线性排序选择法与轮盘赌选择法相结合，从计算好的邻域中选择交叉操作的另一个父体。第(12)行的变异操作采用的是自适应变异。在第(15)行的更新操作中，先对临时子种群进行评估，计算其适应值，如果优于原子种群的相应个体，则新个体取代原个体，否则保留原个体。

4 仿真研究

为了说明SFFRO算子的性能，本文在仿真研究中采用了一组被广泛使用的测试问题集^[3]，对基于SFFRO算子的CGA算法进行测试，并将SFFRO算子分别与FRO算子和EFRO算子做了详细的比较。

4.1 测试问题集与参数设置

(1)Sphere 模型是一个连续、严格凸面的单峰函数，比较简单，主要用于测试算法的寻优精度，其定义如下：

$$f_{Sph}(\mathbf{x}) = \sum_{i=1}^n x_i^2 \quad -5.12 \leq x_i \leq 5.12$$

(2)Griewangk 函数是一个连续多峰函数，由于它不可分离，且搜索算法必须爬山到达下一个峡谷才能得到全局最优解，使得该函数难以优化。但在这组测试问题中相对比较简单，定义如下：

$$f_{Gri}(\mathbf{x}) = 1 + \frac{1}{4000} \cdot \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) \quad -600.0 \leq x_i \leq 600.0$$

(3)Ackley 函数从总体上看，似乎是单峰函数，但由于指数第二项生成了许多小峰和小谷，使得函数具有大量的局部最小值。因此，该函数本质上也是一个连续多峰函数，其定义如下：

$$f_{Ack}(\mathbf{x}) = 20 + e - 20 \cdot e^{-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}} - e^{\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)} \quad -32.768 \leq x_i \leq 32.768$$

(4)一般Rastrigin函数是一个可扩展的连续多峰函数，是通过使用 $a \cdot \cos(\omega \cdot x_i)$ 对Sphere函数进行调整演化而来的，是GA-难度非常大的问题^[4]，其定义如下：

$$f_{Ras}(\mathbf{x}) = a \cdot n + \sum_{i=1}^n (x_i^2 - a \cdot \cos(\omega \cdot x_i)) \quad a=10; \omega=2\pi; -5.12 \leq x_i \leq 5.12$$

当且仅当 $\mathbf{x}^* = (0, 0, \dots, 0)$ 时，这 4 个测试函数的全局最小值均为 0。实验参数如表 2 所示。

表 2 算法参数

参数名称	参数值	参数名称	参数值
子种群规模	180	最大误差	0.000 01
种群拓扑宽度	15	交叉概率	0.700 00
种群拓扑高度	12	变异概率	0.001 00
基因个数	30	最差个体期望拷贝数	0.200 00
邻域大小	5	全局/局部搜索度 e	0.600 00
自适应程度参数	2	e_{local}	0.200 00
最大演化代数	10 000	e_{global}	0.900 00

4.2 仿真结果及其分析

在 CGA 算法中，对上述每个测试函数分别采用 FRO 算子、EFRO 算子和 SFFRO 算子各自独立运行 50 次。若一次

运行中，算法能够在最大演化代数内收敛到指定误差范围中的全局最优解，则认为算法的本次运行是“成功的”，否则认为是“失败的”。由于算法迭代的终止条件为“达到最大演化代数”或“达到指定误差”，只要算法的运行是成功的，其平均最优解就没有显著差异，因此在报告实验结果时仅选取平均演化代数和成功次数来衡量算法的性能。仿真结果如表 3、表 4 所示。

表 3 实验结果 1(平均演化代数)

测试问题	FRO	EFRO	SFFRO
f_{Sph}	602	356	346
f_{Gri}	2 577	2 131	822
f_{Ack}	5 286	4 757	2 140
f_{Ras}	9 166	8 100	4 285

表 4 实验结果 2(成功次数)

测试问题	FRO	EFRO	SFFRO
f_{Sph}	50	50	50
f_{Gri}	50	50	50
f_{Ack}	47	50	50
f_{Ras}	38	42	50

从指标“平均演化代数”(表 3)上看，对于测试问题 f_{Sph} ，SFFRO算子与EFRO算子性能接近；对于其余 3 个测试问题 f_{Gri} 、 f_{Ack} 和 f_{Ras} ，SFFRO算子的收敛效率均明显优于EFRO算子和FRO算子。

从指标“成功次数”(表 4)来看，对于测试问题 f_{Sph} 和 f_{Gri} ，采用FRO算子、EFRO算子和SFFRO算子均能够让CGA算法在各自的 50 次测试运行中求解成功；对于测试问题 f_{Ack} ，只有EFRO算子和SFFRO算子能够在测试中全部运行成功，FRO算子曾经出现 3 次失败；而对于最后一个测试问题 f_{Ras} ，只有SFFRO算子的测试全部获得了成功，FRO算子和EFRO算子分别出现了 12 次和 8 次的失败。

综上所述，SFFRO 算子不仅能平衡算法的全局搜索和局部搜索，而且它引入的适应值信息能为算法指明潜在的搜索方向及搜索范围，从而有效地提高算法的收敛速度。

5 结束语

本文基于标准化适应值信息，提出了改进的模糊交叉算子 SFFRO，并从理论上分析了该算子的 3 个重要特性，并将其应用到 CGA 算法中。仿真实验显示，对于所用的测试问题，SFFRO 算子在平均演化代数和算法成功次数方面均表现出明显的优越性。

参考文献

- [1] Voigt H M, Mühlenbein H, Cvetković D. Fuzzy Recombination for the Breeder Genetic Algorithm[C]//Proc. of the 6th Int'l. Conf. on Genetic Algorithms. San Mateo, CA: Morgan Kaufmann, 1995: 104-111.
- [2] Herrera F, Lozano M. Gradual Distributed Real-coded Genetic Algorithms[J]. IEEE Transactions on Evolutionary Computation, 2005, 9(2): 126-142.
- [3] Clerc M, Kennedy J. The Particle Swarm-explosion, Stability, and Convergence in a Multidimensional Complex Space[J]. IEEE Trans. on Evolutionary Computation, 2002, 6(1): 58-73.
- [4] 李敏强, 寇纪淞, 林 丹, 等. 遗传算法的基本理论与应用[M]. 北京: 科学出版社, 2002.