

一种改进的遗传算法

梁影 金铭 乔晓林

(哈尔滨工业大学 哈尔滨 150001)

摘要 针对遗传算法(Genetic Algorithm, GA)存在的未成熟收敛现象,提出一种改进的遗传算法(IGA)。该算法采用双变异算子,即正交变异和多位点变异两种变异算子联合作用,提高了算法的全局寻优能力。仿真实验表明,对遗传算法的改进是有效的。改进后的算法与标准遗传算法相比具有更高的全局收敛性,并在一定程度上克服了未成熟收敛。

关键词 遗传算法 双变异算子 正交变异 全局收敛性

中图分类号 TP301.6;

文献标志码 A

遗传算法^[1]自20世纪70年代提出以来,因其在求解各类复杂问题过程中表现出的简单易行、全局寻优、鲁棒性强等特点日益受到人们的关注,现已成为研究的热点,目前在很多领域都有成功的应用。但在解决大规模、复杂和精度要求较高的优化问题时,遗传算法会出现未成熟收敛现象^[2-3]。对此已有学者进行了相关研究,但大部分的研究都集中在对交叉概率及变异概率的改进上^[4-7],也有少数的研究考虑了对遗传算子的改进^[8-9]。相关的研究成果都在一定程度上改善了遗传算法的未成熟收敛问题,但也使算法的计算复杂度有所增加,牺牲了算法原有的简洁。

针对标准遗传算法的未成熟收敛问题,文献[10]中指出造成这一问题的主要原因是染色体基因位上基因值的缺失,对此本文提出一种应用于二进制编码的正交变异算子。该算子均衡了各基因位上不同基因值的比例,防止了基因值的过早丢失,提高了算法跳出局部最优解的能力。该正交变异算子虽在一定程度上拓展了搜索空间,但对染色体的破坏力较大,局部搜索能力很差。因此本文将该算子与多位点变异算子相互结合使用,先采用正交变异算子进行全局变异,再用多位点变异算子进行局部变异,形成一种改进的遗传算法(Improved

Genetic Algorithm,以下简称IGA)。与标准遗传算法相比,该算法能够在一定程度上克服未成熟收敛,具有更好的全局寻优能力,相对于其他改进措施而言,该改进方法更加简单,计算复杂度更低。计算机仿真结果验证了本文算法改进的有效性。

1 遗传算法概述

遗传算法是模拟自然界生物进化过程的一种仿生优化算法。根据优胜劣汰的生存法则,种群中适应度值较低的个体逐渐被淘汰,适应度值较高的个体被保留,经过遗传操作产生子代群体,经多代繁殖,使种群不断进化,逐渐趋于最优。

在实际应用中,遗传算法根据问题的目标函数构造适应度函数,并产生一个由多个解(每个解对应一个染色体,即一个编码组合)构成的初始种群,根据适应度值大小对该种群进行选择、遗传运算,繁殖出下一代群体,多代繁殖后将获得的适应度值最好的个体作为问题的最优解。

下面简要介绍标准遗传算法(Standard Genetic Algorithm, SGA)的基本流程:

- (1) 编码方法的选择;
- (2) 产生初始种群;
- (3) 计算适应度值;
- (4) 选择个体形成交配池;
- (5) 对交配池中的个体进行遗传运算产生子代群体,更新种群;
- (6) 判断是否满足停止准则,若满足则输出最

2012年3月12日收到

2009 航天支撑技术基金课题资助

第一作者简介:梁影(1987—),女,硕士。研究方向:阵列信号处理、智能优化算法等。

优解,否则返回步骤(3)。

编码方法也称为基因表达方法,在遗传算法中需对染色体进行编码,使一个染色体对应优化问题的一个解,主要的编码方法有二进制编码、实数编码、序列编码等。编码对遗传算法的搜索效果和效率有重要影响,本文中采用二进制编码方法,二进制编码具有便于位值计算,包括的实数范围较广,硬件实现简单等优点。

初始种群的产生一般为随机产生,种群的大小依赖于计算机的计算能力及所要求解问题的复杂度,一般建议取 20~200。

选择操作的种类主要有轮盘赌选择、联赛选择、比例选择、精英选择等,不同的选择操作选择压的大小不同,可根据需要选择合适的方法。

遗传运算包括交叉和变异,常用的交叉算子主要有单切点交叉、双切点交叉、均匀交叉。传统的变异算子为多位点变异,多位点变异的变异概率 p_m 一般取值范围为 0.001~0.1,变异概率太小,可能使某些基因位过早丢失的信息无法恢复;过高,则遗传搜索将变成随机搜索。

虽然传统的变异算子可以通过变异概率的自适应调整来改善算法的未成熟收敛情况,但每代都要进行变异概率的计算,增加了很多额外的计算量。本文提出的正交变异算子具有操作简单,全局搜索能力强的特点,但其局部搜索能力较差,因此将其与传统变异算子联合使用,采用双变异算子进行变异操作。

2 算法改进

2.1 正交变异算子

根据模式理论^[10],遗传算法中造成未成熟收敛的主要原因是模式的丢失,也即遗传编码中某些基因值的过早丢失,那么均衡各基因值在种群中所占的比例,降低基因值的丢失机率,就能够在一定程度上克服未成熟收敛,正交变异算子就是在此基础上提出来的。

正交变异算子主要由两部分操作组成:异或操作和取反操作。对染色体编码长度为 L 的二进制编码串 $S_1 = a_1 a_2 \cdots a_i \cdots a_L$, $S_2 = b_1 b_2 \cdots b_i \cdots b_L$, 其中 $a_i, b_i \in \{0, 1\}$ 进行异或操作得到的新编码串为 $S'_1 = c_1 c_2 \cdots c_i \cdots c_L$, 其中 $c_i = a_i \oplus b_i$ 。对染色体编码串 S'_1 进行取反操作,得到的新染色体编码记为 $S'_2 = d_1 d_2 \cdots d_i \cdots d_L$, 其中 d_i 为对 c_i 取反后的值,即“0”→“1”,“1”→“0”。

经正交变异算子作用后的种群是由染色体编码相互正交的个体对组成。正交变异算子的作用过程为:①选取种群中两个个体,对这两个个体的二进制编码相同基因位上的基因值进行异或,得到一个新的染色体编码;②对异或操作产生的新染色体编码进行取反操作得到另一个新的染色体。由以上操作就得到一个具有正交关系的染色体对,用该染色体对作为子代个体替换原来的两个父代个体。这样经过正交变异算子作用后的种群就是由具有正交关系的染色体对组成,各基因位上基因值“0”和“1”存在的比例都达到最大值 50%,均衡了各基因位上基因值“0”,“1”的比例,防止了基因值的缺失,一定程度上增加了种群的多样性。

2.2 双变异

仿真发现采用正交变异算子单独进行变异作用的遗传算法,其收敛速度较慢,类似于随机搜索。正交变异算子虽然加强了遗传算法变异操作的全局性,拓展了搜索范围。但由于该变异算子局部搜索能力差的缺点,降低了算法的收敛速度,甚至较难收敛。为了加快算法的收敛速度,使变异操作同样具有局部搜索作用,在算法后期的寻优中采用多位点变异算子进行变异操作。

2.2 双变异

IGA 算法采用双变异算子进行变异操作,正交变异承担全局搜索的任务,多位点变异算子承担局部搜索的任务。在早期的繁殖中采用正交变异算子,而在后期的繁殖中采用传统的变异算子。进化早期的正交变异能够提高基因值的比例,在大范围进行寻优,从而使在进化后期多位点变异算子的作用中,由于前期基因值比例的提高,使该操作中基因值的丢失机率也降低,多位点变异具有较强的局部搜索能力,能够加快算法的收敛速度。

2.3 改进遗传算法步骤

步骤 1: 根据目标函数构造适应度函数,并随机产生满足二进制编码规则的初始种群;

步骤 2: 对染色体进行解码(将二进制编码转换为十进制数)并计算种群中各个体的适应度值;

步骤 3: 对种群进行选择操作;

步骤 4: 对步骤 3 中选择出来的个体进行交叉操作。本文采用双切点交叉,交叉概率 $p_c \in (0, 1)$,交叉概率越高,新结构引入越快,已获得的优良基因结构的丢失速度越高,太低可能导致搜索阻滞;

步骤 5: 若进化代数小于 N ,对交叉后的个体进行正交变异,否则进行多位点变异,从而产生出新的子代群体。

步骤 6: 判断是否达到最大进化代数,若达到最大进化代数则停止,并输出问题的解,否则返回步骤 2。

步骤 5 中 N 的选取较关键,取得太小改进的效果不明显,太大算法收敛速度慢。 N 的取值与目标函数复杂度,以及最大进化代数的大小有关,针对不同的问题可通过仿真来选取合适的 N 值。

3 仿真及分析

为了验证改进算法的有效性,采用如下算例进行仿真分析,仿真中交叉概率 $p_c = 0.7$,变异概率 $p_m = 0.03$,染色体编码长度 $L = 12$ 。

算例 1

采用一个简单的一元多峰函数寻优的例子来验证改进算法的有效性,分别用标准遗传算法和改进的遗传算法求解下面函数的最大值:

$$f_1(x) = x \sin(10\pi x) + 2.0; x \in [-1, 2]。$$

该函数共有 15 个峰值点,其中最大峰值约为 3.8503,其他次优峰值分布在最优解的一侧,峰值依次递减,受其他峰值影响,对该函数的寻优很容易收敛到局部最优点,可用于测试算法全局寻优的能力。仿真中种群规模 $PopSize = 30$,参数 $N = 10$,采用 2-联赛选择机制。最大进化代数 $MaxGen$ 为 40 代,70 代,100 代时的仿真结果如图 1、图 2、图 3 所示。

图 1、图 2、图 3 分别对随机抽取的连续 5 次 IGA 与 SGA 的寻优结果进行了显示,从图中可以很直观的看出 SGA 在 5 次寻优中出现了明显的未成熟收敛现象,很容易收敛到两个较大的次优点,相

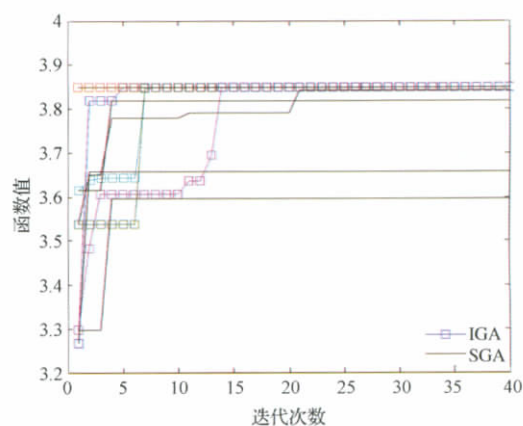


图 1 进化 40 代时 IGA 与 SGA 寻优结果

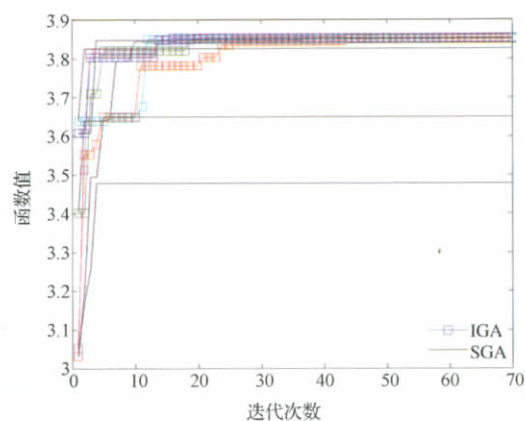


图 2 进化 70 代时 IGA 与 SGA 寻优结果

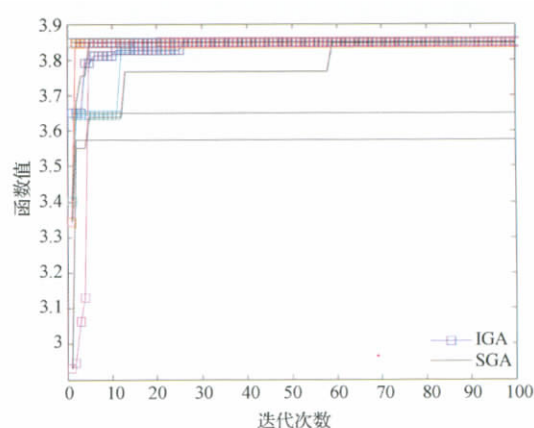


图 3 进化 100 代时 IGA 与 SGA 寻优结果

比之下 IGA 的改进效果较为明显,其在很大程度上改善了遗传算法的未成熟收敛情况,具有更强的全局寻优能力。

算例 2

对二元单峰函数 $f_2(x, y)$ 进行寻优,搜索该函数的最小值:

$$f_2(x, y) = 100(x^2 - y)^2 + (1 - x)^2, -2.048 \leq x, y \leq 2.048.$$

函数 f_2 在 (1, 1) 点取得最小值 0, 其最小值点位于一个狭长的 U 型面平坦底部的凹陷处, 一般很难找到, 该函数一般用于测试算法的未成熟收敛情况。仿真中参数设置为种群规模 $PopSize = 50$, 最大进化代数 $MaxGen = 50$, 参数 $N = 10$, 采用 2-联赛选择机制与精英保留策略。若在最大进化代数之内得到的最优值与实际的最优值相差小于 $1e^{-2}$, 即误差精度为 $1e^{-2}$, 则认为算法收敛到全局最优解, 即寻优成功, 否则认为是未成熟收敛。用 SGA 和 IGA 算法分别对该函数连续进行 100 次蒙特卡洛仿真, 仿真结果如表 1 所示。

表 1 函数 f_2 仿真结果

算法	最优值/ $\times 10^{-6}$	最差值	平均值	成功率/%
SGA	6.611 7	1.003 0	0.048 2	61%
IGA	7.075 3	0.130 5	0.010 7	73%

表 1 给出了 IGA 与 SGA 对函数 f_2 寻优的对比结果, 由于是求 f_2 的最小值, 表 1 中的最优值、最差值分别对应 100 次寻优结果中的最小值、最大值。IGA 算法所得到的最优值、最差值、平均值以及寻优的成功率上都要明显优于 SGA 算法, 我们很清楚的看到 IGA 算法的优越性。IGA 算法在计算量增加不大的情况下使寻优性能有了很大的提高, 具有克服未成熟收敛的能力。

算例 3

对二元多峰函数 $f_3(x, y)$ 进行寻优, 搜索该函数的最大值

$$f_3(x, y) = 0.5 - \frac{\sin^2 \sqrt{x^2 + y^2} - 0.5}{[1.0 + 0.001(x^2 + y^2)]^2}.$$

其中 $-10 \leq x, y \leq 10$ 。

函数 f_3 在 (0, 0) 点取得最大值 1, 该函数的最大值点被很多的次优点包围, 并且由于该函数值具有很强的震荡性, 对该函数寻优很容易收敛到距离最大值点约 3.138 5 处的次优点, 因此可用该函数检测寻优算法跳出局部最优解的能力。仿真中种群规模 $PopSize = 70$, 最大进化代数 $MaxGen = 200$,

参数 $N = \frac{MaxGen}{3} \approx 66$, 采用 2-联赛选择及精英

保留策略。在最大进化代数之内得到的最大值点的坐标位置与实际的最优位置 (0, 0) 距离小于 1×10^{-4} , 则认为算法成功收敛到全局最优解。用 SGA 和 IGA 算法分别对该函数连续进行 100 次蒙特卡洛仿真, 仿真结果如表 2 所示。

表 2 函数 f_3 仿真结果

算法	最优值	最差值	平均值	成功率/%
SGA	1.000 0	0.990 3	0.993 0	28%
IGA	1.000 0	0.990 3	0.996 3	62%

SGA、IGA 对函数 f_3 进行寻优时, 最终的收敛结果只有两种, 一个是全局最优值 1.000 0, 另一个是包围全局最优值的次优点 0.990 3。从寻优结果的寻优成功率上来看, IGA 算法相比 SGA 算法搜索的成功率提高了一倍, IGA 算法在对函数 f_3 寻优的过程中展现了其很好的跳出局部最优解的能力。

4 结论

遗传算法是一种应用范围最广的最优化方法, 传统的遗传算法容易陷入局部最优, 出现未成熟收敛现象。针对该问题本文提出一种具有很强的全局搜索能力的正交变异算子, 并将该算子与传统变异算子相结合提出了一种新的改进遗传算法。仿真结果表明改进算法不仅具有克服未成熟收敛的能力, 还能够很好地摆脱局部最优解, 有效地提高了全局收敛性, 证明了本文方法的优越性。

参 考 文 献

- Holland J H. Adaptation in natural and artificial system: an introduction analysis with applications to biology, control and artificial intelligence. Michigan: The University of Michigan press, 1975
- 周克民, 胡云昌. 遗传算法计算效率的改进. 控制理论与应用, 2002; 19(5): 812—814
- Lin F, Yang Q W. Improved genetic operator for genetic algorithm. Journal of Zhejiang University Science, 2002; 3(4): 431—434
- Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in genetic algorithm. IEEE Trans on Systems, Man and Cybernetics, 1994; 24(4): 656—667
- 吴志远, 邵惠鹤, 吴新余. 一种新的自适应遗传算法及其在多峰值函数优化中的应用. 控制理论与应用, 1999; 16(1): 127—129

(下转第 3644 页)

- 法. 科学技术与工程 2012; 12(4): 809—812
- 6 张学工. 模式识别(第三版). 北京: 清华大学出版社, 2010: 153—157
- 7 张祥德, 张大为, 唐青松, 等. 仿生算法与主成分分析相融合的人脸识别方法. 东北大学学报(自然科学版), 2009; 30(7): 972—975
- 8 Kennedy J, Eberhart R C. Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks, Perth Australia, 1995: 1942—1948
- 9 Kennedy J, Eberhart R. C. A discrete binary version of the particle swarm algorithm. Proc of the Conf on Systems, Man, and Cybernetics. Piscataway: IEEE Press, 1997: 4104—4108
- 10 张焕萍, 宋晓峰, 王惠南. 基于离散粒子群和支持向量机的特征基因选择算法. 计算机与应用化学, 2007; 24(9): 1159—1161
- 11 程国, 丁正生, 刘亚亚. 基于混合投影峰分析的人眼定位方法. 五邑大学学报(自然科学版) 2010; 23(3): 50—55

Face Feature Extraction Combined MMC with BPSO

CHENG Guo

(Department of Mathematics and Computer Science, Shangluo University, Shangluo 726000, P. R. China)

[Abstract] In order to improve the ability of Maximum Margin Criterion (MMC), a face recognition method which fusion the MMC and binary particle swarm optimization algorithm (BPSO) is proposed. Then BPSO is used in feature selection after the transformation of MMC, which can find out feature subspace that is beneficial to classification. Experimental results on ORL face database and Yale face database show that the proposed method not only reduces the dimensions of face feature space, but also expresses the advantages of MMC, improves the speed and accuracy of face recognition.

[Key words] face recognition maximum margin criterion binary particle swarm optimization feature extraction

(上接第 3639 页)

- 6 陶林波, 沈建京, 韩强. 一种解决早熟收敛的自适应遗传算法设计. 微计算机信息, 2006; 22(34): 268—270
- 7 刘学增, 周敏. 改进的自适应遗传算法及其工程应用. 同济大学学报(自然科学版), 2009; 37(3): 303—305
- 8 帅训波, 马书南, 周相广, 等. 一种基于矩阵遗传算子的优化组合遗传算法. 小型微型计算机系统, 2009; 30(5): 951—954
- 9 帅训波, 马书南, 邵艳伟, 等. 基于两种新型遗传算子的优化组合遗传算法. 计算机系统应用, 2010; 19(7): 98—102
- 10 Holland J. Adaptation in nature and artificial systems. 2nd edition, Cambridge, MA: MIT Press, 1992

An Improved Genetic Algorithm

LIANG Ying, JIN Ming, QIAO Xiao-lin

(Harbin Institute of Technology, Harbin 150001, P. R. China)

[Abstract] In order to avoid premature convergence, an improved genetic algorithm (IGA) is presented. The algorithm adopts orthogonal mutation operator and Multi-locus mutation operator that improve the ability of global optimization. The results of simulation experiments show the efficiency of the Improved Algorithm. Compared with standard genetic algorithms, the improved algorithm has a higher degree of convergence, and to a certain extent, overcome the premature convergence.

[Key words] genetic algorithm dual mutation orthogonal mutation global convergence