

# 多点正交交叉的遗传算法研究

刘 清

(南京师范大学 数学与计算机科学学院, 江苏 南京 210042)

[摘要] 利用正交实验法的全局均衡思想, 提出一种采用多点正交交换的遗传算法. 算法通过正交表安排遗传算法的交换运算, 并在所产生的多个子代中选择适应度大的进入下一次进化, 这样既加快了算法的收敛速度又保证了种群的多样性. 实验证明, 该算法不但可以有效地克服标准遗传算法的缺陷, 而且计算速度、精度和算法稳定性也得到了显著提高.

[关键词] 遗传算法, 正交试验, 多点交叉, 仿真

[中图分类号] TP301.6 [文献标识码] A [文章编号] 1672-1292(2005)02-0042-04

## Research on Genetic Algorithm with Multi-Point Orthogonal Crossover Operation

LU Qing

(School of Mathematics and Computer Science, Nanjing Normal University, Jiangsu Nanjing 210042, China)

**Abstract** By using the globalequilibrium design ideobgy of orthogonal experimentmethod, the paper proposes a genetic algorithm with multi-point orthogonal crossover operation. Crossover operation of the algorithm is based on the orthogonal array, and the two of many offspring that have bigger fitness are chosen to put in next evolution. The algorithm can ensure population multiformity and convergence speed rapidly. The research results show that the algorithm can not only overcome the shortcomings of SGA effectively, but also evidently improve the computing speed, computing precision and computing stability.

**Key words** genetic algorithm, orthogonal experiment, multi-point crossover, simulation

## 0 引言

遗传算法(GA)是一种借鉴自然界中“适者生存, 优胜劣汰”思想的全局优化算法. 它把求解问题的可能解看做是用染色体串编码的个体, 由多个个体(或称为染色体)构成种群, 用个体的适应度函数值作为个体优劣的评价指标, 在种群的进化过程中, 不断使用选择、交换、变异这 3 种遗传算子, 使问题的解不断进化, 直至产生全局最优解. 遗传算法特别适合处理传统搜索方法所不能解决的复杂问题和非线性问题. 遗传算法收敛由 Holland<sup>[1]</sup> 用模式定理证明了, 但算法收敛的条件严格, 文献[2]研究了遗传算法的收敛并给出收敛的充分条件. 遗传算法虽然可以稳定收敛, 但是, 它局部搜索能力差, 收敛速度很慢. 为此, 改善遗传算法全局搜

索性, 已成为是遗传算法领域的研究热点.

目前, 一种把遗传算法与其它优化算法相结合, 以利用启发式信息以及与领域有关知识的混合遗传算法已成为改善遗传算法性能的主要策略之一. 其中一种常用的方法就是把局域搜索技术嵌入到遗传算法的基本环节中, 同变异和交换算子一起作用, 在评估后, 对后代施行快捷的局部优化, 这样所得到的算法既具有遗传算法的整体优化特性, 又能加快算法的收敛速度. 如在文献[3]中, 把遗传算法与准牛顿法、爬山法和 Nelder-Mead's Simplex Method 结合起来, 发现遗传算法与准牛顿法的结合使全局优化问题既有遗传算法全局收敛可靠性又有准牛顿法的高精度.

正交试验设计法是一种试验的优化设计技术, 将该技术引入到遗传算法中, 可以改善遗传算法的

收稿日期: 2004-08-25

基金项目: 国家“十五”“211”一期学科建设基金资助项目.

作者简介: 刘清(1962-), 博士, 副教授, 主要从事人工智能理论及测量技术的教学与研究. E-mail: njulq@163.com

局部搜索能力和收敛速度. 目前, 所取得的成果主要有: 把正交设计引入 GA, 提出一种基于家族优生学的进化算法<sup>[4]</sup>; 用实数交换操作的混合遗传算法<sup>[5]</sup>; 利用正交试验设计法求解遗传算法的欺骗问题<sup>[6]</sup>; 用正交设计确定区间水平, 在每个区间水平按遗传算法进行搜索<sup>[7]</sup>. 根据遗传算法和正交试验法的特点, 本文研究了一种采用多点正交交换的遗传算法. 算法通过正交表安排遗传算法的交换运算, 并在所产生的多个子代中选择适应度大的进入下一次进化, 这样既加快了算法的收敛速度又保证了种群的多样性. 实验证明, 该算法在参数寻优中, 收敛速度比简单遗传算法 (SGA) 快, 且不陷入局部最优点.

1 多点正交交换的遗传算法

遗传算法是一种全局并行搜索的寻优方法, 它通过选择复制、交叉、变异等 3 种遗传算子的作用, 使优化群体不断进化, 最终收敛于最优状态. 遗传算法虽然可以稳定收敛, 但它局部搜索能力差, 收敛速度很慢.

1.1 正交试验的原理

正交试验设计法是解决多因素、多水平试验问题的一种有效方法. 在实际问题中, 试验往往是多因素多水平的, 如果对每个因素的每个水平都相互搭配进行全组合试验, 然后从所有的搭配方案中找出最优方案, 则总的试验次数将会爆炸式地增长. 利用正交表的均匀分散性和整齐可比性<sup>[8]</sup> 来安排多因素、多水平试验, 可以实现对试验方案的最优设计. 正交表如表 1 所示.

表 1 正交表  $L_4(2^3)$

实验次数 / 实验因素数	X	Y	Z
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

$A$  是一个  $n \times m$  矩阵, 它的第  $j$  列的元素为由数码  $1, 2, \dots, t_j (j = 1, 2, \dots, m)$  所构成. 它的任两列中, 同行元素所构成的元素对是一个完全对, 而且, 每对出现的次数相同, 则称  $A$  是正交表. 简记为  $L_n(t_1 \times t_2 \times \dots \times t_m) = L_n(t_m)$ ,  $n$  为试验数,  $m$  为因素数,  $t$  为水平数. 这里所谓的完全对是指: 有两组元素  $A_1, A_2, \dots, A_r$  与  $B_1, B_2, \dots, B_s$ , 把  $rs$  个元素对  $(A_1, B_1), (A_1, B_2), \dots, (A_1, B_s), (A_2, B_1), (A_2, B_2), \dots, (A_2, B_s), \dots, (A_r, B_1), (A_r, B_2), \dots, (A_r, B_s)$ , 叫做由元素  $A_1, A_2, \dots, A_r$  与  $B_1, B_2, \dots, B_s$  所构

成的完全对. 正交表有下列性质: (1) 它的任两个元素都构成有相等重复的完全试验, 但是, 一般来说, 却不能构成有  $m$  个因素的完全试验. (2) 在第  $i, j$  两列所构成的水平对中, 每个水平对都重复出现  $n/t_i t_j$  次, 其中  $n$  为试验数. (3) 每一列中各水平出现的次数相同, 如第  $j$  列中, 每个水平都出现  $n/t_j$  次 ( $j = 1, \dots, m$ ). 表 1 是一个简单的正交表  $L_4(2^3)$ , 它用来安排 3 因素 2 水平试验. 表中  $X, Y$  和  $Z$  代表 3 因素; 数字“1”和“2”代表因素的 2 个不同的水平, 从表中可知因素的  $2^3$  种组合就只要 4 重组合进行试验.

1.2 多点正交交叉算子

遗传算法的局部搜索能力与交叉和变异算子有关, 其中, 交叉运算的目的是把染色体中性能优良的组块, 遗传到下一代某个染色体中, 使之具有父辈染色体的优良性能. 但是, 在简单遗传算法 (SGA) 中采用一点交叉运算, 使得许多情况下无法达到这个目的, 而采用多点交叉运算可以避免这个问题<sup>[9]</sup>. 在多点交叉运算中, 其交叉组合有多种方式存在, 且随着交叉点的增多, 组合方式的数量将会爆炸式地增长. 为了解决这个问题, 本文研究了一种正交多点交叉算子, 通过正交表安排多点交叉运算, 以保证交叉运算后能够得到优良染色体, 提高遗传算法的局部搜索能力和收敛速度.

定义 1 设染色体进行  $N - 1$  点交叉运算, 则交叉的染色体子串的个数为  $N$ , 将这  $N$  个子串看成  $N$  个因素.

定义 2 将染色体的每一个交叉子串, 设置为同一水平, 这样, 两个的染色体的交叉运算, 就是一个  $N$  因素、2 水平试验.

定义 3 用  $N$  因素、2 水平正交试验表安排染色体的子串交叉, 得到  $2^N$  个新染色体, 通过适应度函数选取两个适应度最大的作为子代染色体, 进入下一次进化算法. 该生成子代染色体的过程是多点正交交叉算法.

例如, 设一个长度为 24 的染色体, 如进行 2 点交叉运算, 则要交叉的染色体子串为 3 将染色体 1 的每一个子串设置为水平 1 染色体 2 的每一个子串设置为水平 2 如表 2 所示. 这样, 2 点交叉运算就成为一个二水平三因素的试验.

表 2 染色体参数设置

	水平 / 因素	X	Y	Z
染色体 1	1	01010101	10101110	01001011
染色体 2	2	11001100	00101011	11010101

但因素间可能会有交互作用, 一个二水平三因素的试验, 可用正交表  $L_8(2^7)$  来安排试验. 将  $L_8(2^7)$  表中的水平用 “- 1” 和 “1” 来表示, 如表 3 所示.

表 3 正交表  $L_8(2^7)$

实验次数 / 列号	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	- 1	- 1	- 1	- 1
3	1	- 1	- 1	1	1	- 1	- 1
4	1	- 1	- 1	- 1	- 1	1	1
5	- 1	1	- 1	1	- 1	1	- 1
6	- 1	1	- 1	- 1	1	- 1	1
7	- 1	- 1	1	1	- 1	- 1	1
8	- 1	- 1	1	- 1	1	1	- 1

在表 3 中的 7 个 8 维向量相互正交, 且第 1 和第 2 列的元素相乘, 是第 3 列; 第 1 和第 4 列的元素相乘, 是第 5 列; 第 2 和第 4 列的元素相乘, 是第 6 列. 从试验设计的角度, 若将因素  $X$  和  $Y$  分别放在第 1、2 列, 则它们的交互作用  $X \times Y$  反映在第 3 列. 所以, 该列不能再排其他因素, 否则主效应与交互效应将混在一起, 两者都无法估计. 考虑因素间可能会有交互作用的一个二水平三因素试验的设计方案如下:

列号	1	2	3	4	5	6	7
因素	$X$	$Y$	$X \times Y$	$Z$	$X \times Z$	$Y \times Z$	

该方案中, 每个主效应和交互效应各占 1 列. 但在试验安排时, 只需要  $X$ 、 $Y$ 、 $Z$  所在的第 1、2、3、4 列, 将该 3 列的 2 个水平换算成实际的水平, 得到遗传算法的两点交叉算子的试验方案.

新定义的 2 点交叉算法的遗传基本结构如下:

```
(1) 初试始化种群随机产生  $N$  个个体.
(2) do{
    for ( $i = 0$   $i < N$ ;  $i++$ )
        {计算第  $i$  个个体的适应度值; }
    根据适应值通过轮盘赌选择算法选出  $N$  个父体;
    按照交叉概率  $P_c$  取出 2 个父体, 根据新定义的交叉操作产生 8 个子代, 并计算它们的适应度值, 保留适应度值大的 2 个个体;
    按照变异概率对新种群的个体进行变异操作;
    while( (不满足性能指标要求) && (没有完成规定代数的寻优) ); }
(3) 从群体中选择最优个体, 输出寻优结果.
在交叉运算得到 8 个新染色体中, 含了原两个
```

父代染色体 (第 1 次和第 8 次试验), 且选取两个适应度最大的作为子代. 所以, 该算法既保证了种群的多样性, 又是一种精英策略的局部寻优算法.

2 试验与仿真

由于遗传算法中大量的随机操作, 理论上进行严格分析是非常困难的. 通常人们采用一些典型的测试函数来检测算法的实际效率<sup>[10]</sup>. 为验证多点正交交叉的遗传算法的有效性, 本文用函数优化问题进行仿真实验, 以多峰 Schaffer 和 Rosenbrock 函数作为测试函数, 分别用简单遗传算法 (SGA: 采用赌轮法选择, 单点交叉和单点变异, 固定交叉和变异概率)<sup>[9]</sup>、自适应遗传算法 (AGA: 采用赌轮法加精英最优保留法选择, 单点交叉和单点变异, 自适应改变交叉和变异概率)<sup>[9]</sup> 以及本文提出的多点正交交叉的遗传算法 (这里采用 2 点交叉) 对上述函数进行优化对比. 在性能测试中, SGA、AGA 和改进算法参数取为: 群体规模 popsize = 100, Schaffer 函数编码长度  $L = 48$  位 (两变量的编码长度分别为  $l = 22$  位), 最大迭代代数 MaxGenerations = 200 代. Rosenbrock 函数编码长度  $L = 32$  位 ( $l = 16$  位);  $P_m = 0.04$ ,  $P_c = 1$ ; 最大迭代代数 MaxGenerations = 1 000 代.

Rosenbrock 函数定义为:

$$\min f_1(x) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2 - 2.048 \leq x_i \leq 2.048$$

函数  $f_1$  是一个二维单峰值函数, 具有一个全局极小点  $f_1(1.0, 1.0) = 0.0$  但它却是病态的, 在函数曲面上沿着曲线  $x_2 = x_1^2$  有一条较为狭窄的山谷, 传统的梯度优化方法搜索到山谷边缘时, 往往会发生振荡, 难以进行全局优化.

Schaffer 函数定义为:

$$\max f_2(x) = 0.5 - \frac{\sin^2 \sqrt{x_1^2 + x_2^2} - 0.5}{[1 + 0.001(x_1^2 + x_2^2)]^2} - 100 \leq x_i \leq 100$$

函数  $f_2$  在其定义域有无数个局部极大点, 只有一个全局极大点  $f_2(0.0, 0.0) = 1.0$  但由于变量的取值范围大, 最大值峰周围有一圈脊, 它们的取值均为 0.990 283, 因此很容易使算法陷入此局部极大点.

本文采用 Matlab 编制了相应各个算法的仿真试验程序, 为了消除随机干扰, 每个试验都重复进行 50 次, 具体试验统计结果如表 4 所示.

表 4 算法性能比较

函数	平均迭代数			平均时间 /s			收敛率 /%			最优解		
	本文	SGA	AGA	本文	SGA	AGA	本文	SGA	AGA	本文	SGA	AGA
$f_1$	21.3	135.3	50.7	5.4	32.5	15.7	100	100	100	0.0	0.0	0.0
$f_2$	421.3	/	898.1	92.2	/	154.2	72	0	45	1	/	1

从实验结果中可以看出, 本论文所研究的多点正交遗传算法, 具有更好的收敛性能.

3 结语

针对传统的遗传算法存在许多局部寻优能力差、收敛速度慢的不足, 本文把正交设计引入 GA 中, 提出了一种多点正交交叉的交叉遗传算法, 算法通过正交表安排遗传算法的交换运算, 并在所产生的多个子代中选择适应度大的进入下一次进化, 这样既加快了算法的收敛速度又保证了种群的多样性. 研究结果表明, 该算法性能得到了很大提高, 因此具有良好的应用前景和推广价值.

[参考文献]

[ 1 ] Holland JH. Adaptation in Natural and Artificial System [M]. Ann Arbor: Michigan Univ Press, 1975  
[ 2 ] 王丽薇, 洪勇, 洪家荣. 遗传算法的收敛性研究 [J].

计算机学报, 1996, 19( 10): 794- 797.

[ 3 ] JeanMichel Renders, Stéphane P Fasse. Hybrid methods using genetic algorithms for global optimization[ J]. IEEE Trans. System, Man and Cybernetics part-B Cybernetics, 1996, 26( 2): 234- 248  
[ 4 ] 吴少岩, 张青富, 陈火旺. 基于家族优生学的进化算法 [J]. 软件学报, 1997, 8( 2): 137- 144.  
[ 5 ] 彭伟, 卢锡城. 一种函数优化问题的混合遗传算法 [J]. 软件学报, 1999, 10( 8): 819- 823  
[ 6 ] 吴浩扬, 常炳国, 朱长纯. 遗传算法的一种特例——正交试验设计法 [J]. 软件学报, 2001, 12( 1): 148- 153  
[ 7 ] 史奎凡, 董吉文, 李金屏, 等. 正交遗传算法 [J]. 电子学报, 2002, 30( 10): 1501- 1504  
[ 8 ] 方开泰, 马长兴. 正交与均匀试验设计 [M]. 北京: 科学出版社, 2001. 40- 63  
[ 9 ] 李人厚. 智能控制理论和方法 [M]. 西安: 西安电子科技大学出版社, 1999. 178- 180  
[ 10 ] 牛志华, 李乃成. 一种新的混合遗传算法及其性能分析 [J]. 工程数学学报, 2003, 20( 4): 54- 58

[ 责任编辑: 严海琳 ]