

多群体阶段性杂交遗传算法

董安邦, 李军军, 王 嵩

(西安交通大学管理学院, 710049, 西安)

摘要: 借鉴生物遗传学提出了一种多群体阶段性杂交遗传算法. 引入相对顺序交叉算子对标准遗传算法进行了改进. 为验证该算法的性能, 对旅行商问题进行了求解, 采用多群体和阶段性杂交的改进策略, 并分别和标准遗传算法进行了比较. 计算结果表明, 该方法能较好地保证个体多样性, 并能促进优秀基因型的杂交和遗传, 在收敛和鲁棒性方面优于一般的单群体、非杂交算法. 另外, 将其应用于水电站优化调度也取得了较好的效果.

关键词: 遗传算法; 杂交遗传算法; 多群体; 阶段性杂交

中图分类号: TP18 **文献标识码:** A **文章编号:** 0253 - 987X(2004)02 - 0128 - 04

Stages Hybrid Genetic Algorithm with Multi-Group

Dong Anbang, Li Junjun, Wang Song

(School of Management, Xi an Jiaotong University, Xi an 710049, China)

Abstract: Drawing on idea of genetics, a staged hybrid genetic algorithm (SHMGA) with multi-group (SHMGA) is proposed. Firstly, the relative sequential crossover operator is introduced into the standard genetic algorithm (GA) to improve its performance, then a SHMGA is designed to solve the traveling salesman problem for testifying its ability with the adoption of the multi-group and staged hybrid policies. Computational results show that this method can ensure the diversity of individuals and promote the crossbreed and heredity of good genotypes so it has better convergence and robustness compared with the standard genetic algorithm, i. e. single-group or non-hybrid GA. Additionally it is applied to the schedule optimization of hydropower stations and shown to be valid.

Key words: genetic algorithm; hybrid genetic algorithm; multi-group; stages hybrid

遗传算法 (GA) 的缺陷^[1] 主要在于不能保证收敛至全局最优解^[2]. 这些未成熟收敛现象^[3,4] 与算法处理中的每个环节都有关系.

本文先对标准 GA 的主要算子进行优化, 设计了求解旅行商问题 (TSP) 的基本算法, 然后借鉴生物遗传学的理论并应用如下策略来抑制未成熟收敛.

(1) 可在 GA 中引入分群的概念使各 (子) 群体独立进行遗传操作, 以保障遗传个体的多样性.

(2) 引入杂交算子, 把所有子群体中的最优个体引入至其他群体, 这将加快成熟收敛^[5~7].

(3) 可以分阶段使用不同的杂交概率进行杂交,

以提高搜索性能.

(4) 生物体提高适应力的另一理论就是显性基因假说^[8]. 由于显、隐两种基因的存在, 使得生物体的后代以 3:1 的比例表现出母体的性状. 我们将这一假说应用于单倍体遗传算法, 优化效果差强人意.

1 求解 TSP 的基本遗传算法^[9]

初始种群 随机产生 3 个个体; 分别计算 3 个个体的适应度函数, 从中选择函数值最大的个体放入初始群体; 若达到群体规模数 M , 则初始种群设定完毕, 否则继续步骤 .

适应度函数 $f(x) = 1/l \sum_{i=1}^{n-1} d(v_i, v_{i+1}) + d(v_1, v_n)$, $d(v_i, v_{i+1})$ 为城市 v_i 到城市 v_{i+1} 的距离。

迭代停止条件 若群体中的最佳个体保持 N 代不变,则遗传停止,其中 $N = n^2/4$, n 为城市个数。

选择 设群体规模为 M , 其中个体 j 的适应度值为 f_j , 则 j 被选中的概率 $P_j = f_j / \sum_{k=1}^M f_k$ 。为了保证收敛,采用最佳个体保留法,即若上一代的最佳个体的适应度大于新一代的,则用上一代的最佳个体替换新一代中的最差个体进入遗传群体。

交叉 设计了相对顺序交叉法,描述如下。

(1) 随机在串 $A = 1\ 2\ | \ 3\ 4\ 5\ 6\ | \ 7\ 8\ 9$ 和串 $B = 5\ 7\ | \ 1\ 6\ 9\ 8\ | \ 3\ 4\ 2$ 中选择一个交叉区域(以 9 个城市遍历为例)。

(2) A 中交叉区域的第一个城市代码为 3, 城市 3 在 B 中的前接城市为 8, 则将 A 的交叉区域插在 B 中城市 8 的后面; B 中交叉区域的第一个城市代码为 1, 城市 1 在 A 中的前接城市为 9, 则将 B 的交叉区域插在 A 中城市 9 的后面。因此,有 $A = 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ | \ 1\ 6\ 9\ 8\ |$; $B = 5\ 7\ 1\ 6\ 9\ 8\ | \ 3\ 4\ 5\ 6\ | \ 3\ 4\ 2$ 。

(3) 删除和串中交叉区域相同的城市,得到两个最终子串,即 $A = 234571698$; $B = 719834562$ 。

变异 采用逆转、对换变异,交叉概率为 95%, 对换变异概率为 0.8%, 逆转变异概率为 100%。

该法较好地体现了子代个体对父代个体特性的继承性,维持了遗传的相对稳定,结果(见表 1)表明其性能优于类 OX 交叉法^[2]。

2 多群体算法和阶段杂交策略

传统的 GA 都是单群体遗传操作,易造成群体内基因类型的同化,而群体间的间隔有利于形成各自不同的基因库,可保障个体多样性促进成熟收敛。将群体中的个体划分成若干个子群体,对各子群体分别进行操作,保持各自的遗传独立性;用 Delphi 编程来实现该多群体遗传算法。将群体划分成 2 个和 3 个子群体,各子群体的规模数 $M = 80$, 进行 n 为 20、30、40 的 TSP 求解,各进行 5 次(相当于样本数为 5 的抽样),并与单群体进行比较,结果见表 2。可以看出,无论是最短路径,还是各最短路径的离散程度均明显优于单群体算法。当群体数 g 由 2 增为 3 时,其收敛效果反而不如群体数 g 从 1 增加到 2 时的状况。这可能与群体间模式信息的交流更有利于群体社会进化有关,因此有必要引入杂交优势算法,使群体间发生一定的联系。

采用最佳个体引入法,即按照一定的杂交概率,用所有子群体中最佳的一个个体替代各子群体中的最差个体。考虑到不同的杂交概率会影响到群体间基因类型差异的大小,以 $n = 30$ 的 TSP 为例,分别进行不同杂交概率的 2 个群体、3 个群体的实验,结果(见表 3)发现,杂交算法基本上优于非杂交算法,且引入杂交算法的较优结果基本上是在小杂交概率上取得的,当杂交概率为 90% 时,杂交效果最差。因为在算法迭代的初期,各子群体还未形成各自的基因库,此时若以大概率杂交,容易破坏各群体的自我发展;当各群体形成自己的基因库时,以较大的概率

表 1 相对顺序交叉法与类 OX 交叉法求解 TSP 最佳路径长度的结果

n	类 OX 交叉法				相对顺序交叉法			
	l_{max}/km	l_{min}/km	l_{mean}/km	方差	l_{max}/km	l_{min}/km	l_{mean}/km	方差
20	444.821	437.658	439.731	8.633	438.115	437.658	437.841	0.063
30	551.172	523.050	529.706	145.206	549.659	517.561	526.731	172.079
40	591.262	555.963	577.135	248.949	575.886	544.603	555.391	227.521

注: $M = 80$; l 表示路径长度。

表 2 群体数为 1、2 和 3 的遗传算法的优化路径长度比较

n	$g = 1$		$g = 2$		$g = 3$		最优结果	
	l_{mean}/km	方差	l_{mean}/km	方差	l_{mean}/km	方差	l_{mean}/km	方差
20	437.840	0.030	437.658	0	437.658	0	437.658	0
20	526.731	172.079	522.454	56.693	522.363	57.067	522.363	56.693
40	555.391	227.521	553.815	83.764	558.006	144.216	553.815	83.764

注: l 表示路径长度。

表 3 $n = 30$ 时不同杂交概率的优化路径长度比较

杂交概率/ %	$g = 2$				$g = 3$			
	l_{max}/km	l_{min}/km	l_{mean}/km	方差	l_{max}/km	l_{min}/km	l_{mean}/km	方差
0	535.611	517.561	522.454	56.693 0	535.611	517.561	522.363	57.066 8
10	520.541	517.561	518.347	1.559 8	520.541	517.561	518.848	2.242 8
20	522.234	517.561	519.779	4.482 6	520.541	517.561	518.344	1.563 4
40	520.541	517.561	518.936	2.182 0	535.611	517.561	522.454	56.011 3
90	520.541	517.561	518.753	2.664 1	535.611	518.017	525.564	85.178 2

注: l 表示路径长度.

杂交有利于产生体现不同群体优势的杂交后代,可使算法向更优的方向收敛,所以采用阶段性杂交是必要的.

阶段性杂交遗传算法是指,在运算中根据遗传群体世代数的不同采用不同的杂交概率,在迭代初期采用小杂交概率,随着世代数的增加逐渐增大杂交概率,从而实现具有阶段性的群体间杂交.我们采用了两种阶段性杂交策略(见表 4),进行了 $n = 30$ 、群体数分别为 2 和 3 的实验.实验证明 2 个群体使用策略 2、3 个群体使用策略 1 可以达到各自最优迭代.

表 4 两种阶段性杂交策略下的杂交概率

	杂交概率/ %				
	G1	G2	G3	G4	G5
策略 1	10	30	50	70	
策略 2		20	40	60	80

注: G1 表示 1~99 代; G2 表示 100~199 代; G3 表示 200~299 代; G4 表示 300~399 代; G5 表示 400 代.

对 2 个群体使用策略 2、3 个群体使用策略 1,分别进行 n 为 20、30、40 的 TSP 求解,然后将结果与非杂交的多群体算法进行对比(见表 5).可见,采用阶段性杂交使杂交优势理论得以体现,这表明对于多群体采用合适的阶段杂交策略有利于群体间遗传信息的共享,改善收敛质量.

表 5 阶段性杂交算法与非杂交多群体算法的优化路径长度的比较

n	非杂交多群体算法				阶段性杂交算法				最优结果	
	$g = 2$		$g = 3$		$g = 2$		$g = 3$			
	l_{max}/km	方差	l_{max}/km	方差	l_{max}/km	方差	l_{max}/km	方差	l_{max}/km	方差
20	437.658	0	437.658	0	437.749	0.042	437.658	0	437.658	0
30	522.454	553.815	522.363	558.001	517.656	558.198	518.157	550.045	517.656	550.045
40	553.815	83.764	558.001	144.216	558.198	154.029	550.045	96.517	550.045	83.764

注: l 表示路径长度.

表 6 群体数为 3 的显性算法和多群体阶段性杂交算法的优化结果比较

n	平均世代数			平均时间/ s		
	非显性算法	显性算法		非显性算法	显性算法	
		变异概率 0.8 %	变异概率 0.30 %		变异概率 0.8 %	变异概率 0.30 %
20	197	189	146	29	28	25
30	429	424	350	84	84	69
40	686	673	538	187	184	112

传信息的共享,改善收敛质量.

3 显性遗传算法的检验

按 3 1 的比例分别随机产生高适应度后代和低适应度后代,采用阶段性杂交策略 1,群体数为 3,各子群体 $M = 80$,进行 n 为 20、30、40 的 TSP 求解,结果(见表 6)并不理想.由于所采用的显性算法仅仅是从现象上对生物体的显性遗传进行模拟,相当于增加了高适应度个体的生存率而使收敛速度有所提高.但是,这会引入群体中个体的同化现象,加大未成熟收敛的可能.为了提高群体的多样性,还进行了变异概率为 3 % 的显性遗传实验,结果表明变异概率的增加在一定程度上缓解了同化现象,但效果仍难以令人满意.以往的研究表明了,显性算法在变动性较大的搜索中效果较好.单倍体显性算法在不稳定系统中的应用则有待于进一步研究,所以暂不将显、隐机制引入该算法.综上,我们得到了一种在收敛速度和搜索能力上均具有较多优势的多群体阶段性杂交遗传算法.算法求解的结果示例见图 1a(城市数为 30)和图 1b(城市数为 40).

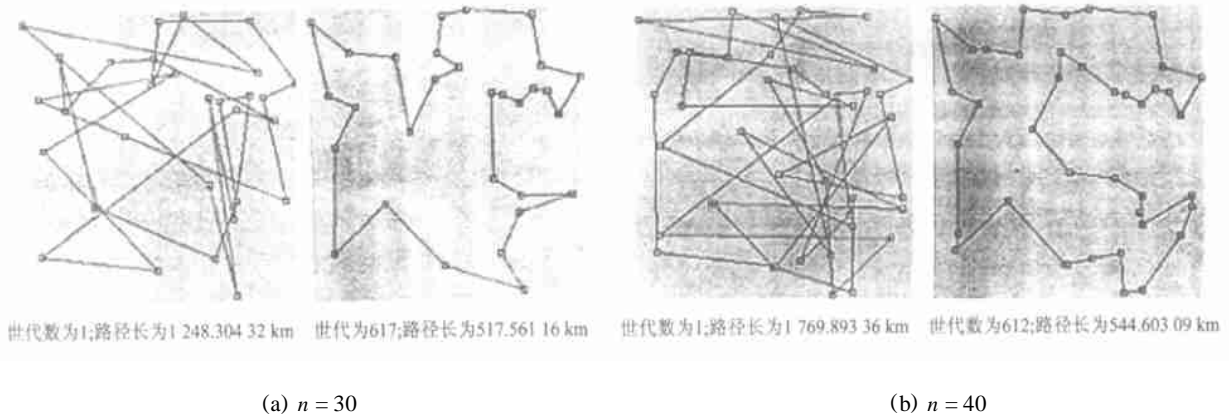


图 1 多群体阶段性杂交遗传算法求解结果

4 结 论

由于目前仍缺乏缜密的数学理论来分析和评价 GA 的各执行环节,本文立足于实践,借鉴生物进化理论分别对多群体、阶段性杂交、显性遗传机制的优化效果进行了验证.实验表明,多群体和杂交策略有助于抑制未成熟收敛,提高收敛效果,而单纯配比的显性遗传算法的效果并不理想.最后,得到了多群体阶段性杂交遗传算法.

针对待求解问题,设定了相应的种群划分方法和选择算子,经验表明这种方法有助于有种群划分倾向的群体能较快地淘汰劣解且又不失个体的多样性,但其能否对所有解具有普遍意义还需进一步研究.此外,2 个群体优化结果优于 3 个群体的现象是在实施单纯多群体算法中出现的(表 2 和表 3),当采用阶段性杂交策略(表 4)时,3 个群体的结果均优于 2 个群体的,其中的部分机制可由文献[7]得到解释,但还需从理论上进行严格的证明.当然,还有其他内容要作进一步的研究,如在阶段性杂交遗传算法中,究竟如何根据世代数的增加和问题复杂度来确定杂交概率,还要进行大量的探索.

本文所提方法已应用于龙羊峡水电站的优化调度^[10],结果证明该方法在处理高复杂性问题时是成功的.限于篇幅,这里仅给出应用 SHMGA 按照当时调度期始末的实际水位对龙羊峡水电站进行模拟调度的结果(见表 7).SHMGA 的比较优势表明了

表 7 龙羊峡水电站模拟调度结果

发电量/ 10^9 xkW·h		提高率/%
实验结果数据	实际运行数据	
51.25	50.2	2.09

该方法在处理复杂多维问题的优化领域具有良好的应用前景.

参考文献:

- [1] 王 健,王建华.标准遗传算法的研究进展 [J]. 华东船舶工业学院学报,2000,14(3):28~29.
- [2] 陈国良.遗传算法及其应用 [M].北京:人民邮电出版社,1996.
- [3] Campbell P.J. Reviews: Genetic algorithms in search, optimization and machine learning by David E. Goldberg [J]. Mathematics Magazine, 1989,62(3):206~207.
- [4] Rudolph G. Convergence analysis of canonical genetic algorithms [J]. IEEE Transactions on Neural Networks, 1994,5(1):96~101.
- [5] 张尤凯,吴昌谋.遗传学 - 群体分化与新物种的形成 [M].广州:暨南大学出版社,1996.
- [6] Chu P.H., Dudley S.A. The effect of population structure on the rate of convergence of genetic algorithms [A]. ACM-SAC, Indianapolis, USA, 1993.
- [7] 曾建潮,徐玉斌.多群体优良模式自学习遗传算法 [J].系统工程与电子技术,2002,22(6):81~83.
- [8] 褚 折.遗传学 [M].上海:上海教育出版社,1980. 220~221.
- [9] 朱剑英.智能系统非经典数学方法 [M].武汉:华中科技大学出版社,2001.
- [10] 李军军.多群体阶段性杂交遗传算法及其在水电站优化调度中的应用 [D].西安:西安交通大学管理学院,1998.

(编辑 苗 凌)