

文章编号:1006-5911(2005)12-1743-04

基于双螺旋染色体和分层结构的遗传算法

吴家宏, 雷毅

(北京航空航天大学 华正软件研究所, 北京 100083)

摘要:为满足制造业相关项目中寻找最优化解的需求,提出了一种基于双螺旋染色体和分层种群结构的遗传算法模型。在此模型中,解染色体数据结构采用的是类似 DNA 的双螺旋对称结构,一个解染色体中包含 2 份完全相同的解信息,但在遗传运算中各自承担不同的角色;种群采用优、中、差 3 类解按一定比例组成的分层结构,使得新生的解种群能够涵盖更广的求解范围;遗传运算参数中的交叉和变异概率随种群结构特征动态变化。基于此模型的遗传算法,能够保护优秀解,避免算法早熟,同时还具有收敛速度快、稳定性强的特点。

关键词:遗传算法;双螺旋对称结构;分层种群

中图分类号:O224

文献标识码:A

Research on genetic algorithm based on double helix chromosome & multi-layers population structure

WU Jia-hong, LEI Yi

(Huazheng Software Research Inst., Beihang Univ., Beijing 100083, China)

Abstract: To deal with the problem of seeking optimal solution for mechanical projects in manufacturing, a new model of genetic algorithms based on double helix chromosome and multi-layer population structure was proposed. The gene data structure of chromosome of the proposed model representing the real solution was based on double symmetrical chains structure, just like the double helix spiral structure of DNA in nature. There were two kinds of completely identical solution information in every solution chromosome but played different roles in genetic algorithm. The population of chromosomes was proportionally made up of excellent, medium and bad layers so that each layer could be prorated by the user. And the probability of crossing and aberrance would be dynamically changed with the structure of population. This model could keep the excellent solution genes and avoid the precocity of computing process with fast astringency and better stability.

Key words: genetic algorithm; double helix spiral data structure; multi-layers population structure

0 引言

遗传算法 (Genetic Algorithms, GA)^[1-2] 是一种崭新的全局优化算法,在制造工程领域有很大的应用前景,如在计算机辅助制造 (Computer Aided Manufacturing, CAM) 领域求解刀具加工的最短路

径问题、基于资源约束的生产计划安排问题等。这些都是典型的 NP 问题,用人力去解决几乎是不可能的;而通过利用计算机的运算能力,再结合遗传算法模型,就可以得到较为理想的答案。

目前,广大工程技术人员在应用遗传算法时所做的改进主要集中在染色体编码的改进、进化参数

收稿日期:2004-10-19;修订日期:2005-01-07。Received 19 Oct. 2004; accepted 07 Jan. 2005.

基金项目:国家 863/CIMS 主题资助项目(2001AA412020,2003AA414041)。Foundation item: Project supported by the National High-Tech. R&D Program for CIMS, China(No. 2001AA412020,2003AA414041).

作者简介:吴家宏(1979-),男,福建福州人,北京航空航天大学华正软件研究所硕士研究生,主要从事中小制造企业信息化、生产作业调度等的研究。E-mail: wuahong_cn@hotmail.com。

模型的改进以及与其他启发式搜索算法的结合等。文献[3]通过改进交叉和变异算子的模型,以及使用单纯形局部搜索操作来增强算法的搜索能力;文献[4]将遗传算法和模拟退火算法相结合,提出遗传算法模拟退火(Genetic Algorithms Simulated Annealing, GASA)模型来改进全局寻优能力和进化速度;文献[5]提出了均匀两点交叉模型。以上文献通过各自的模型来克服标准遗传算法模型的不足,都取得了较好的结果。这些文献的共同点是:在基因结构方面采用的均是“单链结构”;在种群模型方面都采用只保留最优,摒弃最差的方式。自然界中, DNA 存在的物理形式是空间双螺旋结构,种群的繁衍和进化常常依赖分层的结构,从植物群落到动物群体,再到人类社会,均是如此。受此启发,本算法模型中的染色体采用对称的数据结构模拟双螺旋结构;种群模型使用按设定比例的分层结构,而不是采用单纯的择优原则。

1 基于双螺旋染色体结构和分层种群结构的遗传算法模型

1.1 算法流程

算法总体上采用顺序结构的流程,如图 1 所示。

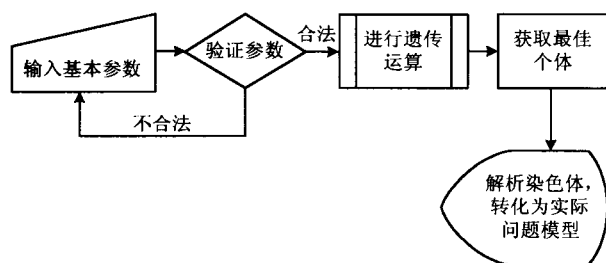


图1 算法的顺序结构

系统从用户操作界面获得计算参数之后,首先进行参数的合法性验证。例如种群的规模和进化代数只能是正整数,种群的进化参数只能介于 0 和 1 之间等。参数验证为合法后,就进入遗传运算,直到达到设定的计算终止条件。遗传运算结束后,挑选出适应值最佳的个体,并把所携带的染色体编码“翻译”为实际问题的解答模型。进行遗传运算是其中最重要的环节,主要步骤如下:

步骤 1 对解决问题进行编码设计。

步骤 2 $t=0$, 随机初始化群体 $X(0)=(x_1, x_2, \dots, x_n)$ 。

步骤 3 对当前群体 $X(t)$ 中每个个体 x_i 计算其适应度 $F(x_i)$, 适应度表示该个体所代表的解的

优劣。

步骤 4 应用选择算子产生中间代 $X'(t)$ 。

步骤 5 应用交叉算子和变异算子处理 $X'(t)$, 这时候的 $X'(t)$ 包含了新产生的个体和父代个体。

步骤 6 计算 $X'(t)$ 每个个体的适应度 $F(x_i)$, 排序, 按照预先设定的比例分为 3 段, 取出每段中按比例计算出数量的个体, 组合成新一代群体 $X(t+1)$ 。

步骤 7 整理个体, 使之变为双螺旋对称结构。

步骤 8 $t:=t+1$; 如果 t 大于预先设定的世代数, 则终止; 否则, 转步骤 3。

1.2 染色体采用双螺旋的双链对称结构

自然界中, 染色体的空间物理结构是双螺旋结构, 2 条螺旋之间通过 A, T, C 和 G 4 种碱基来配对, 由于 4 种碱基的配对关系是固定的 (A-T, C-G), 使得其中任一条螺旋链可以完整地表示另一条的信息。这种结构不仅使染色体在复制过程中可以很好地保留父代信息, 也能为子代的变异提供条件。

基于该思想, 本算法模型中的染色体采用双链对称结构, 每一个染色体均保存 2 份解信息, 在进行进化操作时, 只选择其中的一条进行。这样处理的优势在于可以很好地保留某条染色体对应的父代信息, 为在后续的种群重组中实现跨代择优选择提供条件, 也为在需要对父代信息做额外处理的实际应用场景下打下良好的基础。

基于对称结构的染色体在进行进化操作之前呈双链对称形态。进入进化模块之后, 首先进行解链操作, 一个染色体的双链分离; 其次, 选取其中的一条链与其他的染色体发生交叉、变异操作, 形成新的一条染色体链。在算法模型进入种群重组之前, 每个染色体所带的双链呈不对称结构, 其中一条代表父代信息, 另一条代表子代信息。染色体在“进化”操作前后的变化如图 2 所示。

1.3 分层结构的种群模型

自然界中存在的很多群体都呈分层的结构, 如蜂群社会以蜂后、雄蜂和工蜂的 3 层结构形式存在, 原始森林一般都由高大的乔木、低矮的灌木和贴地的草类植物组成。在这些种群中, 各层次之间的个体互相依存, 互相促进。正是由于这种分层的结构, 种群才得以生存、延续和进化。

一般遗传算法都按照染色体适应值的降序, 选择最优的 N 个的染色体重组为新的种群, 进入下一轮运算。这样处理的优点是可以确保保留最好的染

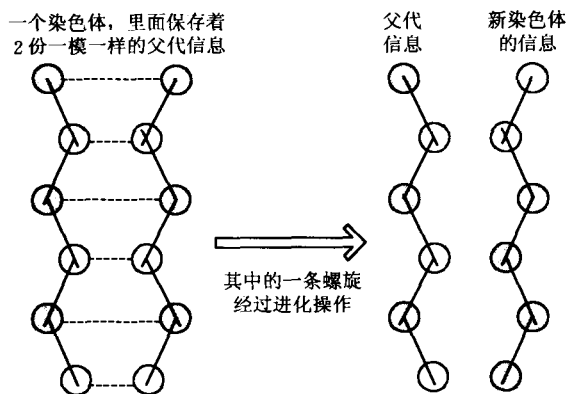


图2 双螺旋对称结构染色体在“进化”操作中的变化

染色体, 算法收敛速度快; 但是这种模式很容易使算法陷入局部欺骗区域, 出现进化进程缓慢, 或者个别个体早熟, 导致运算过早结束, 得到效果不佳的解。使用分层结构的种群模式能够尽量地保留种群中染色体的多样性, 不仅为跳出局部欺骗区域提供条件, 还能有效地保护优秀的染色体, 使之能在种群中保留下来。这样的模式不仅没有牺牲收敛速度, 还大大增加了遗传算法模型的容错性, 使之不容易“早熟”。

本算法模型中的种群模型在初期呈离散状态, 经过进化操作之后, 按照适应值的降序排序。这时的临时种群由于包括了父代信息, 个体数量大于预先设定的种群规模数量。通过采用分段比例选取的方法, 将临时种群压缩至预先设定的规模。图 3 为本模型的种群构建方案。

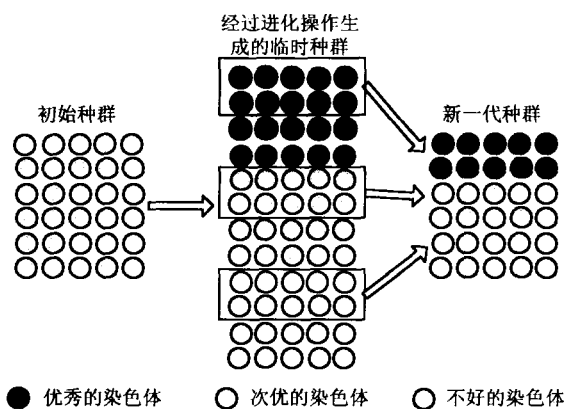


图3 分层结构的种群在进化过程中的构建示意图

1.4 进化参数的动态设定

遗传算法中, 交叉概率 P_c 和变异概率 P_m 是影响算法性能的两个重要因素, P_c 越大, 算法收敛越快; P_m 过大, 会使遗传算法变为随机搜索。所以本算法模型采用动态改变 P_c 和 P_m 的方式, 当种群的最优适应度和平均适应度小于设定值 ξ 时, 种群趋

于稳定, 采用小的交叉概率和大的变异概率有利于扩大搜索范围, 防止早熟; 当种群的最优适应度和平均适应度相差甚远时, 采用大的交叉概率和小的变异概率有利于提高算法收敛速度。在本算法模型中, 初期使用预先设定的 $P_c (P_c > 0.5)$ 和 $P_m (P_m < 0.5)$; 模型稳定时, 交叉概率取 $1 - P_c$, 变异概率取 $1 - P_m$ 。这样处理不但简单, 运算速度快, 而且能够达到动态进化参数的目的。

2 实例验证

通常采用求解函数极值的方法来验证遗传算法模型的性能。这样的验证方法具有普适性, 不仅能够验证算法的收敛速度, 还能验证算法的“容错”能力和稳定性。为了验证本文提出的算法模型, 选取 2 个典型的测试函数, 分为 2 维和 3 维 2 种类型。选取的 2 个函数的共同特征就是存在多个局部极值, 全局只有一个最优值。下面分别给出统计结果。

2.1 2 维振荡函数

$$f(x) = x \sin(10\pi \times x) + 2.0, -1 < x < 2.$$

在这次测试中, 参考对比的对象是文献[6], 采用的参数如表 1 所示。

表 1 本文算法与文献[6]算法中的参数

	遗传代数	种群规模	交叉概率	变异概率	其他
本算法	50	30	0.80	0.05	分层比例为 3:4:3
文献[6]	150	30	0.80	0.05	无

本算法运行 40 次, 结果对比如表 2 所示。

表 2 本文算法与文献[6]算法运行 40 次的结果对比

	最优解	最优解出现代数
本算法	3.850 272 366 355 279 8	平均第 16 代
文献[6]	3.850 274	第 89 代

从实验数据可以看出, 本文算法模型的收敛速度优于文献[6]中的模型, 同时获得了期望的结果。

2.2 Shaffer's F6 函数

$$f(x, y) = 0.5 - \frac{\sin^2 \sqrt{x^2 + y^2} - 0.5}{(1 + 0.001 \times (x^2 + y^2))^2},$$

$$-100 < x, y < 100.$$

这个函数是著名的遗传算法测试函数, 有无数个局部极大点, 其中只有 (0, 0) 为全局最大, 最大值为 1, 函数最大值周围有 1 个圈脊, 很容易停滞在此区域, 其 3 维局部形状如图 4 所示。

在这次测试中,采用的参数如表 3 所示。

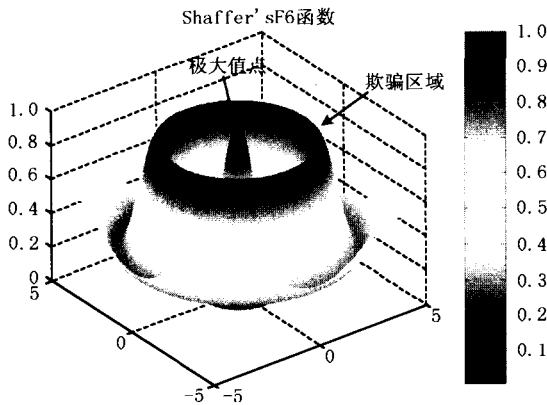


图4 Shaffer's F6函数3维局部形状

表 3 Shaffer's F6 函数的参数

总体参数	运算次数	14
	每次运算遗传代数	100
	种群规模大小	30
进化参数	交叉概率	0.8
	变异概率	0.2
种群设置 比例/%	优秀解比例	30
	次优秀解比例	40
	较差解比例	30

测试标准是如果大于“圈脊”区域的值,就认为此次收敛了。

结果运行 50 次,和文献[7]的比较如表 4 所示。

表 4 简单遗传算法与文献[7]改进算法的结果比较

	平均进化 代数	未收敛 次数	提高率(和简单遗传 算法相比的 收敛速度)/%
简单遗传算法 ^[7]	70.6	4	
文献[7]改进算法	30.8	0	128.5
本文提出算法	11.7	2	503.4

3 结束语

上述试验表明:①基于双螺旋染色体结构和分

层种群结构的遗传算法模型能够很好地保护优秀解,保持种群的多样性,避免种群“早熟”,算法提前结束;②算法的收敛速度很快;③算法的稳定性很好。本算法模型已被应用于“面向中小制造企业信息化的集成系统开发与应用”课题中的车间作业安排模块,取得了良好的效果。实例证明,本文所提出的遗传算法模型在计算效率和计算效果方面都有很大的优势,为制造领域中需要进行大规模工程优化问题的解决提供了一种可行方案。

参考文献:

[1] HOLLAND J H. Adaptation in nature and artificial system [M]. MI, USA: The University of Michigan Press, 1975.

[2] BROOKER L B, GOLDBERG D E, HOLLAND J H. Classifier system and genetic algorithm [J]. Artificial Intelligence, 1989, 40(2): 235—282.

[3] PENG Xinzhu. Improved genetic algorithm and its application [J]. Journal of East China Shipbuilding Institute (Natural Science Edition), 2002, 16(3): 53—58 (in Chinese). [彭新竹. 遗传算法的改进策略及其应用[J]. 华东船舶工业学院学报(自然科学版), 2002, 16(3): 53—58.]

[4] FU Yongfeng. Performance appraisalment of the simulated annealing genetic algorithms [J]. Journal of Changchun Teachers College, 2003, 22(2): 9—12 (in Chinese). [付永锋. 一种改进的遗传算法[J]. 长春师范学院学报, 2003, 22(2): 9—12.]

[5] YANG Dadi, ZHANG Chuntao. Genetic algorithm of uniform two—point crisscross [J]. Journal of Chongqing Normal University (Natural Science Edition), 2004, 21(1): 26—29 (in Chinese). [杨大地, 张春涛. 均匀两点交叉遗传算法[J]. 重庆师范大学学报(自然科学版), 2004, 21(1): 26—29.]

[6] WANG Xiaoping, CAO Liming. Genetic algorithm—theory, application and software development [M]. Xi'an: Xi'an Jiaotong University Press, 2002. 20—49 (in Chinese). [王小平, 曹立明. 遗传算法——理论、应用与软件实现[M]. 西安: 西安交通大学出版社, 2002. 20—49.]

[7] HAN Wanlin, ZHANG Youdi. Improvement of genetic algorithm [J]. Journal of China University of Mining & Technology, 2000, 29(1): 102—105 (in Chinese). [韩万林, 张幼蒂. 遗传算法的改进[J]. 中国矿业大学学报, 2000, 29(1): 102—105.]