

遗传算法综述

刘立平 牛 熠

(东莞理工学院计算机系, 广东东莞 523106)

摘要: 遗传算法是一种自适应全局优化概率搜索算法, 在各种科学工程领域得到了广泛的应用, 文章介绍了遗传算法的发展历程、理论研究和应用研究状况。

关键词: 遗传算法; 进化算法; 概率搜索算法

中图分类号: O211 **文献标识码:** A **文章编号:** 1009-0312 (2005) 01-0048-05

遗传算法 (Genetic Algorithms 简称GA) 是由美国Michigan大学的John Holland教授于20世纪60年代末创建的。它来源于达尔文的进化论和孟德尔、摩根的遗传学理论, 通过模拟生物进化的机制来构造人工系统。从1985年在美国卡耐基·梅隆大学召开的第一届国际遗传算法会议到1997年5月IEEE的Transactions on Evolutionary Computation创刊, 遗传算法作为具有系统优化、适应和学习的高性能计算和建模方法的研究渐趋成熟。

遗传算法是一种自适应全局优化概率搜索算法^[1], 它不同于枚举法、启发式算法、搜索算法等传统的优化方法, 主要有以下特点:

(1) 自组织、自适应和学习性 (智能性)。遗传算法消除了算法设计中的一个最大障碍, 即需要事先描述问题的全部特点, 并要说明针对问题的不同特点算法应采取的措施, 因此, 它可用来解决复杂的非结构化问题, 具有很强的鲁棒性。

(2) 直接处理的对象是参数的编码集而不是问题参数本身。

(3) 搜索过程中使用的是基于目标函数值的评价信息, 搜索过程既不受优化函数连续性的约束, 也没有优化函数必须可导的要求。

(4) 具有显著的隐并行性。遗传算法按并行方式搜索一个种群数目的点, 而不是单点。它的并行性表现在两个方面: 一是遗传算法是内在并行的、二是遗传算法的内含并行性。

(5) 形式上简单明了。

1 标准遗传算法

生物进化过程的发生一般需要四个基本条件: 1) 存在由多个生物个体组成的种群; 2) 生物个体之间存在着差异, 或群体具有多样性; 3) 生物能够自我繁殖; 4) 不同个体具有不同的环境生存能力, 具有优良基因结构的个体繁殖能力强, 反之则弱。生物群体的进化机制包括三种基本形式: 1) 自然选择; 2) 杂交; 3) 突变; 另外, 外界对生物的评价反映了生物的生存价值和机会。John Holland^[2]教授通过模拟生物进化过程设计了最初的遗传算法, 我们称之为标准遗传算法。标准遗传算法给出了遗传算法的基本框架, 以后对于遗传算法的改进, 都是基于此种算法。标准遗传算法的流程如下:

- (1) 使用二进制编码对搜索空间进行编码
- (2) 随机产生包含n个个体的初始群体
- (3) 适应度评估检测个体适应度 (个体适应度反映了个体好坏的情况)
- (4) WHILE <未满足迭代终止条件> DO

收稿日期: 2004-12-23

作者简介: 刘立平 (1975-), 男, 湖南湘乡人, 硕士, 主要从事数据挖掘等的研究。

- (5) 用赌轮选择方法选出若干个体进行繁殖, 个体可以重复
- (6) 随机配对, 按一定概率(交叉概率)进行一点交叉操作, 并生成两个子个体
- (7) 按照一定概率(变异概率)变异二进制个体串中某个(些)位
- (8) 适应度评估检测个体适应度
- (9) END DO

遗传算法是从代表问题可能潜在解集的一个种群开始的, 一个种群由经过基因编码的一定数目的个体组成。初始种群产生之后, 按照适者生存和优胜劣汰的原理, 逐代演化产生出越来越好的近似解。在每一代, 根据问题域中个体的适应度大小挑选个体, 并借助自然遗传学的遗传算子进行交叉和变异, 产生出代表新的解集的种群。这个过程将导致种群象自然进化一样的后代种群比前代更加适应环境, 末代种群中的最优个体经过解码, 可以作为问题近似最优解。

2 遗传算法理论研究

遗传算法所追求的是当前群体产生比现有个体更好个体的能力, 即遗传算法的可进化性或称群体可进化性。因此, 遗传算法的理论和研究方法也就围绕着这一目标展开。

2.1 遗传算法的基础理论研究

遗传算法的基础理论主要以收敛性分析为主, 即群体收敛到优化问题的全局最优解的概率。从整体上讲, 可以分为基于随机过程的收敛性研究和基于模式理论的收敛性分析。

(1) 随机模型理论

对于有限的编码空间和有限的群体, 遗传算法的搜索过程可以表示为离散时间的马尔可夫链模型(Markov chain model), 从而可以采用已有的随机过程理论进行严密分析:

- ① Rudolph^[3]用奇次有限马尔可夫链证明了标准遗传算法收敛不到全局最优解;
- ② Eiben^[4]等用马尔可夫链证明了保留最优个体的GA的概率性全局收敛;
- ③ Qi和Palmieri^[5]用马尔可夫链证明可浮点编码的遗传算法是收敛的, 但此结果是基于群体数为无穷大这一假设。

(2) 进化动力学理论

基于某种具体运算形式的遗传算法的进化行为分析构成了进化动力学理论的基本内容。

- ① 由Holland提出的模式定理^[1]可以称为遗传算法进化动力学的基本定理;
- ② 积木块假设^[1]描述了遗传算法的重组功能;
- ③ 模式定理和积木块假设构成了求解优化问题时遗传算法具备发现全局最优解的充分条件, 也是分析遗传算法的进化行为的基本理论, 统称为模式理论。

模式理论奠定了GA的数学基础, 根据隐形并行性得出了每一代处理有效模式的下限值是 $n^l/(C_l^{1/2})$, 其中 n 是群体大小, l 是二进制个体的串长, C_l 是小整数, 一些研究把结果简化为 n^l , 恽为民、席裕庚^[6]等获得了每代至少产生 $O(2^n)$ 数量级的新模式, 即一次迭代, 相当于对 n 的指数次方大小的子空间进行了搜索, 表明了遗传算法搜索的高效性。

2.2 遗传策略研究与设计

为了维持群体可进化性并最终搜索到问题的全局最优解, 遗传算法必须采用合适的运算形式(编码、计算流程、种群规模、种群初始化策略、GA算子、终止条件等), 这就是遗传策略研究与设计的主要研究内容。遗传策略研究与设计可以分为微观遗传策略研究和宏观遗传策略研究两部分。微观遗传策略: 主要讨论群体规模、遗传算子的形式和参数设计, 及其对GA求解能力的影响。宏观遗传策略: 主要讨论关于通过GA流程的再设计, 改变GA的宏观特征, 或者以GA流程为基础, 引入其他算法构成混合GA, 以期提高GA求解问题全局最优解的能力。下面我们讨论改进遗传算法的一些方法:

2.2.1 改变遗传算法的组成成分或使用技术

(1) 编码

编码是由问题空间向GA空间的映射, De Jong的编码原理包括两条规则:

- ① 有意义积木块编码规则: 所定编码应当易于生成与所求问题相关的短距和低阶的积木块。
- ② 最小字符集编码规则: 所定编码应采用最小字符集以使问题得到自然的表示或描述。

目前主要的编码技术有^[1]: 一维染色体编码(包括二进制编码、实数编码、格雷编码等)、多参数映射编码、可变染色体长度编码、二倍体、多倍体编码和树形编码等等。Holland模式定理建议使用二进制编码, 但二进制编码精度不高。双倍体二进制编码比单倍体的动态跟踪能力更强。浮点数编码精度高, 便于大空间搜索, 是现在研究热点。

(2) 群体设定

群体设定主要指如何维护群体规模, 即群体中包括的个体数目如何设定, 包括两个方面的问题: 1) 初始群体的设定; 2) 进化过程中各代的群体规模如何维持。初始群体设定的一般解决办法是: 1) 根据问题固有知识, 设法把握最优解所占空间在整个问题空间中的分布范围, 然后, 在此分布范围内设定初始群体; 2) 先随机生成一定数目的个体, 然后从中挑出最好的个体加到初始群体中。这种过程不断迭代, 直到初始群体中个体数达到了预先确定的规模。一般情况下, 群体规模为几十至几百, 各代群体规模相同。

(3) 适应度函数

适应度函数是评价个体适应环境的能力, 是选择操作的依据, 它的好坏能直接影响遗传算法的性能。适应度函数是由目标函数变换而成的。对适应度函数唯一的要求是: 结果为非负值。这使得遗传算法能够应用于不连续、不能求导的情况, 大大增加了它的适用性。对目标函数值域的某种映射变换称为适应度的尺度变换。采用尺度变换可克服未成熟收敛和随机漫游现象^[1]。常用的适应度函数尺度变换方法主要有线性变换、幂函数变换和指数变换^[1]。

2.2.2 采用非标准的遗传操作算子

(1) 选择算子(selection)

选择是用来确定交叉个体, 以及被选个体将产生多少个子代个体。常用的方法有^[1]:

① 适应度比例方法(赌轮或蒙特卡罗选择), 各个个体的选择概率和其适应度值成比例; ② 最佳个体保存方法, 把群体中适应度最高的个体不进行配对交叉而直接复制到下一代中。一般和其它方法结合使用; ③ 排序选择方法, 指在计算每个个体的适应度后, 根据适应度大小顺序对群体中个体排序, 然后把事先设计好的概率表按序分配给个体, 作为各自的选择概率。所有个体按适应度大小排序, 而选择概率和适应度无直接关系而仅与序号有关。④ 联赛选择方法, 其操作思想是, 从群体中任意选择一定数目的个体(称为联赛规模), 其中适应度最高的个体保存到下一代。这一过程反复执行, 直到保存到下一代的个体数达到预先设定的数目为止。联赛规模一般取2。⑤ 排挤方法, 采用该方法时, 新生成的子代将替代或排挤相似的旧父代个体。该方法可以提高群体的多样性

(2) 交叉算子(crossover)

交叉指把两个父代个体的部分结构加以替换重组而生成新个体的操作。交叉操作的作用是组合出新的个体, 在串空间进行有效搜索, 同时降低对有效模式的破坏概率。交叉操作是GA区别于其它进化算法的重要特征, 遗传算法中起核心作用的是遗传操作的交叉算子。设计交叉算子需考虑如下几点: 1) 需保证前一代中优秀个体的性状能够在后一代的新个体中尽可能得到遗传和继承; 2) 交叉算子设计和编码设计需协调操作。对于占主流地位的二值编码而言, 各种交叉算子都包括两个基本内容: 1) 从由选择操作形成的群体中, 对个体随机配对并按预先设定的交叉概率来决定每对是否需要交叉操作; 2) 设定配对个体的交叉点, 并对这些点前后的配对个体的部分结构(或基因)进行相互交换。常用的交叉操作方法^[7]有一点交叉(在个体串中随机设定一个交叉点)、二点交叉、一致交叉(通过设定屏蔽字来决定新个体的基因继承两个父个体中哪个个体的对应基因)、二维交叉、树结构交叉、部分匹配交叉、顺序交叉和周期交叉等等。

(3) 变异算子(mutation)

变异即对群体中的个体串的某些基因座上的基因值作变动。变异的目的是有两个: 1) 使遗传算法具有局部的随机搜索能力; 2) 保持群体的多样性。但常规位突变的效果是不明显又很慢的, 突

变在GA中的作用是第二位。变异算子的操作一般分两步：1) 在群体中所有个体的码串范围内随机地确定基因座；2) 以事先设定的变异概率来对这些基因座的基因值进行变异。变异算子有很多方式^[7]，如基本变异算子（对群体中的个体码串随机挑选一个或多个基因座并对这些基因座的基因值做变动）、逆转算子（在个体码串中随机挑选两个逆转点，然后将两个逆转点间的基因值以逆转概率 p 逆向排序）、自适应变异算子（与基本变异算子的操作内容类似，唯一的不同点是交叉概率 p 不是固定不变，而是随群体中个体的多样性程度而自适应调整）等等。

(4) 算法参数的选择

遗传算法中需要确定一些参数的取值，这些参数对GA性能影响很大。选择的参数主要有串长 l ，群体大小 n ，交叉概率 p_c ，以及变异概率 p_m 等。参数的确定分为静态方法和动态方法，静态方法指参数事先就确定好，算法执行过程中不变；动态方法指参数在算法执行过程中根据情况进行调整变化，以适应环境的变化。目前对动态方法的研究较多。对于静态参数：恽为民等建议 n 的选择与所求问题的非线性程度相关，非线性越大， n 越大。目前常用的参数范围是： $n=20-200$ ， $p_c=0.5-1.0$ ， $p_m=0-0.05$ 。

2.2.3 采用混合遗传算法

基本遗传算法的局部搜索能力较差，主要原因是遗传算法的交叉操作容易破坏当前模式，使得小范围的精确搜寻很困难。一般对标准遗传算法进行适当的改进或将遗传算法与传统的、基于问题知识的启发式搜索技术相结合（即混合遗传算法）来解决这个问题。其中将全局搜索能力强的遗传算法和局部搜索能力较强的传统启发式搜索方法相结合的混合遗传算法能产生很好的优化效果。设计混合遗传算法有三条指导性原则：1) 采用原有算法中的编码技术；2) 吸收原有算法的优点（a. 如果原有算法是快速算法，则把它产生的解添加到混合遗传算法的初始群体中；b. 把原有算法中的一系列变换结合到混合遗传算法）；3) 改造遗传算子。混合遗传算法的框架如图1：

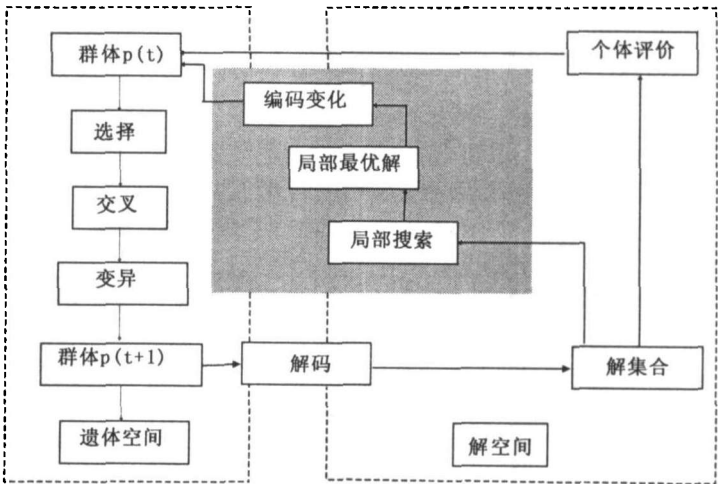


图1 混合遗传算法构成示意图

常见的混合遗传算法有：1) 遗传算法与模拟退火相结合的混合遗传算法；2) 遗传算法与最速下降相结合的混合遗传算法；3) 遗传算法与爬山法相结合的混合遗传算法

2.2.4 调整算法控制参数和编码粒度

Srinivas^[11]等提出的一种自适应遗传算法AGA，交叉概率 P_c 和变异概率 P_m 能够随适应度自动改变。Davis^[12]提出了一种 p_c ， p_m 随基因操作的在线性能自适应变化的有效方法，性能提高，则 p_c 增加，反之则 p_m 增加。Fogarty研究了 p_m 随遗传代数变化的效果，例子显示 p_m 随指数下降有较好性能。

3 遗传算法的应用

遗传算法提供了一种求解复杂系统优化问题的通用框架，它不依赖于问题的具体领域，对问题

的种类有很强的鲁棒性,所以广泛应用于很多学科,主要有以下应用领域:控制方面有瓦斯管道控制,防避导弹控制,机器人控制等;规划方面有生产规划,并行机任务分配等;设计方面有VLSI布局,通信网络设计,喷气发动机设计等;组合优化方面有TSP问题,背包问题,图划分问题等;图像处理方面有模式识别,特征抽取等;信号处理方面有滤波器设计等;机器人方面有路径规划等;人工生命方面有生命的遗传进化等;知识发现方面有规则提取,数据挖掘等。

4 总结

遗传算法经过30多年的发展,在基础理论、算法的设计上都取得了长足的发展,尤其是在越来越多的领域得到了广泛的应用。近年来,许多冠以“遗传算法”的研究与Holland最初提出的算法已少有雷同:不同的遗传算法表达方式,不同的交叉和变异算子,特殊算子的引用,以及不同的再生和选择方法,这些改进方法产生的灵感都来自大自然的生物进化,可归结为进化计算这样的遗传“算法簇”。遗传算法本身的发展也是不断进化的过程,需要引入新的数学工具,吸收生物学的最新成果。

参考文献

- [1] 王小平,曹立明,等.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.
- [2] Holland J H. Concerning Efficient Adaptive Systems[J]. In Yovirs.M.C.Eds. Self-Organizing Systems. 1962: 215-230.
- [3] Rudolph G. Convergence Properties of Caucritical Genetic Algorithms[J]. IEEE Trans on Neorel Networks. 1994.5(1):96-101.
- [4] Eiben A E,Art E h, Van Hee K M. Global Convergence of Genetic Algorithms: An Infinite Markov Chain Analysis[J]. Parallel Problem Solving, 1-12.
- [5] Li X ,Palmieri F. Adaptive Mutation in the Genetic Algorithm. Proc of the Sec Ann.Conf on Evolutionary Programming[M]. Fogel D B. Atmar.w.,Eds. GA: Evolutionary Programming Society, 1993.
- [6] 恽为民,席裕庚.遗传算法的运行机理分析[J].控制理论与应用,1996.13(3): 289-297.
- [7] 李敏强,等.遗传算法的基本理论与应用[M].北京:科学出版社,2002.
- [8] Davis L D. Genetic Algorithms and Simulated Annealing[M]. Morgan Kaufmann. Los Altos.1987.

A Summary of Genetic Algorithm

Liu Liping Niu Yi

(DongGuan University of Technology,Dongguan 523808,China)

Abstract Genetic Algorithm is a self-adapted probability search method used to solve optimisation problems, which has been applied widely in science and engineering. This paper mainly introduces the history, theory and application of Genetic Algorithm.

Keywords genetic algorithm; evolutionary algorithm; probability search method