

遗传算法综述

常洪江*

摘 要：本文主要回顾了遗传算法的发展历程,并对遗传算法的基本原理及特点作了简要阐述。进一步指出了遗传算法存在的问题及相应的改进措施,讨论了遗传算法在实际中的应用。

关键词：遗传算法 选择 交叉 变异 适应度函数

中图分类号：TP273 **文献标识码：**A **文章编号：**1002-2422(2010)03-0115-02

An Overview of Genetic Algorithm

Chang Hongjiang

Abstract: The paper reviews the short history of GA's development and simply expatiates on the basic principle and characteristics on GA. The paper also points out GA's limitation and improving measures, finally gives a simple discussion about GA's application.

Keyword: GA Selection Crossover Mutation Fitness Function

遗传算法广泛应用于自动控制、计算科学、模式识别、工程设计、智能故障诊断管理科学和社会科学领域,适用于解决复杂的非线性和多维空间寻优问题。

1 遗传算法的特点

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。遗传算法具有进化计算的所有特征,同时又具有自身的特点:

(1) 搜索过程既不受优化函数的连续性约束,也没有优化函数导数必须存在的要求。

(2) 遗传算法采用多点搜索或者说是群体搜索,具有很高的隐含并行性,因而可以提高计算速度。

(3) 遗传算法是一种自适应搜索技术,其选择、交叉、变异等运算都是以一种概率方式来进行,从而增加了搜索过程的灵活性,具有较好的全局优化求解能力。

(4) 遗传算法直接以目标函数值为搜索信息,对函数的性态无要求,具有较好的普适性和易扩充性。

(5) 遗传算法更适合大规模复杂问题的优化。

度模型三种主要框架。在主从式模型中,主处理器控制整个 GA 的运行,其他的从处理器则进行个体的初始化、交叉、变异和选择操作,然后由主处理器进行总体调控。在粗细粒度模型中,每个处理器进行整体算法的独立操作,然后各个处理器之间进行个体交流。在细粒度模型中,一个处理器只能与其周围的处理器之间进行个体交换。在并行算法中,处理器之间的个体交换称之为“迁移”,是并行 GA 中独特的操作算子。通过迁移算子,可实现最佳个体在不同子种群之间的流动,从而提高各个子种群之间,加快子种群的收敛。

5 结束语

遗传算法经过几十年的发展,其应用从工程科学到社会科学的诸多领域,今后,拓展更加多样的应用领域将是遗传算法发展的主流。

收稿日期 2010-04-20

* 常洪江 黑龙江亚电鑫宝热电有限公司工程师(黑龙江 哈尔滨 150001)。

2 遗传算法的基本原理

GA 研究的问题是搜索候选假设空间并确定“最佳假设”。在 GA 中,“最佳假设”被定义为是使适应度最优的假设,适应度是为当前问题预先定义的数字度量。

2.1 遗传算法的原型

John Holland 教授通过模拟生物进化过程设计了最初的遗传算法,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。尽管遗传算法的实现在细节上有所不同,但都具有以下的共同结构:算法迭代更新一个假设池,这个假设池称为群体。在每一次的迭代中,根据适应度函数评估群体中的所有成员,然后从当前群体中用概率方法选取适应度最高的个体产生新一代群体。在这些选中的个体中,一部分保持原样地进入下一代群体,其他的被用作产生后代个体的基础,其中应用交叉和变异这样的遗传方法。标准遗传算法的流程如表 1。

2.2 选择算子

参考文献

- [1] Pan Z J, Kang L S. An Adaptive Evolutionary Algorithms for Numerical Optimization. Proc.of SEAL'96, Taejon, Korea, 1996: 53-60.
- [2] L. Davis, Handbook of Genetic Algorithms, van Nostrand Reinhold, New York, 1991.
- [3] 席裕庚,柴天佑,恽为民. 遗传算法综述. 1996, 13(60):69-7-708.
- [4] 刘勇,等. 非数值并行算法(二)—遗传算法. 北京: 科学出版社, 1995.
- [5] 陈国良,等. 遗传算法及应用. haviour of A Class of Genetic Adaptive Systems. PhD thesis, University of Michigan, 1975.
- [6] 孙艳丰,王众托. 自然数编码遗传算法的最优种群规模. 沈阳: 信息与控制, 1996(5).

表 1 遗传算法原型

GA(Fitness, Fitness_threshold, p , r , m)
 Fitness :适应度评分函数,为给定假设赋予一个评估分数。
 Fitness_threshold :指定终止判据的阈值。
 p :群体中包含的假设数量。
 r :每一步中通过交叉取代群体成员的比例。
 m :变异率。
 初始化群体 : $P \leftarrow$ 随机产生的 p 个假设。
 评估 :对于 P 中的每一个 h , 计算 $\text{Fitness}(h)$,
 当 $[\max \text{Fitness}(h)] < \text{Fitness_threshold}$ 时 ,
 产生新一代 P_s :
 (1) 选择 :用概率方法选择 P 的 $(1-r)p$ 个成员加入 P_s 。从 P 中选择假设 h_i 的概率 $\text{Pr}(h_i)$ 用下面公式计算 :

$$\text{Pr}(h_i) = \frac{\text{Fitness}(h_i)}{\sum_{j=1}^p \text{Fitness}(h_j)}$$

 (2) 交叉 :根据上面给出的 $\text{Pr}(h_i)$,从 P 中按概率 $r \cdot p/2$ 选择对假设。对于每对假设 $\langle h_1, h_2 \rangle$,应用交叉算子产生两个后代,把所有的后代加入 P_s 。
 (3) 变异 :使用均匀的概率从 P_s 中选择成员。对于选出的每个成员,在其表示中随机选择一个位取反。
 (4) 更新 : $P \leftarrow P_s$
 (5) 评估 :对于 P 中 h 的每个计算 $\text{Fitness}(h)$
 从 P 中返回适应度最高的假设。

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) 适应度比例方法,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。

(3) 排序选择方法,指在计算每个个体的适应度后,根据适应度大小对群体中个体排序,并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,其中适应度最高的个体保存到下一代。并反复执行,直到保存到的个体数达到预先设定的数目为止。

2.3 交叉算子

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,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。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进行相互交换。常用的交叉操作方法有一点交叉、二点交叉、一致交叉、二维交叉、树结构交叉等等。

2.4 变异算子

变异即对群体中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。

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,如基本变异算子、逆转算子、自适应变异算子等等。

3 遗传算法存在的问题

(1) 编码表示:Holland在运用模式定理分析编码机制时建议使用二进制编码,但二进制编码不能直接反映问题的固有结构、精度不高、个体长度大和占用计算机内存多。

解决这个问题的措施有:

① 动态编码即在保持串长不变的前提下减小搜索区域,当算法收敛到某局部最优时增加搜索的精度,从而使使得在全局最优附近可以进行更精确的搜索。

② 对于问题的变量是实向量的情形,可以直接采用实数进行编码,便于引入与问题领域相关的启发式信息以增加算法的搜索能力。

③ 复数编码本是为了描述和解决二维问题,还可以推广到多维问题的描述中。

(2) 适应度函数:适应度函数是用来区分群体中个体好坏的标准,选择的好坏直接影响算法的优劣,选择的不好容易引起两种不利于优化的现象:

① 异常个体引起早熟收敛,影响求得全局最优解。这种现象常出现在小规模群体中。

② 个体差距不大引起搜索成为随机漫游,特别是平均适应度已接近最佳适应度时,最佳个体与其他许多个体在选择过程中就会有大体相等的选择机会,影响求得最优解。

(3) 选择策略:优胜劣汰的选择机制使得适应值大的个体有较大的存活机会,不同的选择策略对算法性能有较大的影响。轮盘赌法是使用最多的选择策略,但这种策略可能会产生较大的抽样误差,对此提出了很多的改进方法,如繁殖池选择、非线性排名选择、基于局部竞争机制的选择等。

(4) 控制参数:群体大小、交换概率、变异概率等这些参数对遗传算法性能影响较大,会影响算法的全局最优性和收敛性。Davis提出自适应算子概率的方法,用自适应机制把算子概率与算子产生的个体适应性结合。而且高适应性值被分配高算子概率这种方法较好地解决了这一问题。

参考文献

- [1] 陈文清. 遗传算法综述[J]. 洛阳: 洛阳高等专科学校学报, 20-03.
- [2] 徐清振, 肖成林. 遗传算法的研究与应用[J]. 广州: 现代计算机, 2006.
- [3] 章晓军, 戴冠中, 徐乃平. 一种新的优化搜索算法—遗传算法[J]. 广州: 控制理论与应用, 1999, 12(3): 365-273.
- [4] Tom M. Michell. 机器学习[M]. 北京: 机械工业出版社, 2003.
- [5] (美) 卡伦著. 人工智能. 黄厚宽, 等译. 北京: 电子工业出版社, 2004.