

标准遗传算法的原理及算例

唐穗欣

(武汉科技大学中南分校 现代教育技术学院, 湖北 武汉 430223)

摘要:由于遗传算法求解复杂优化问题的巨大潜力及其在工业工程领域的成功应用,这种算法受到广泛关注。叙述了遗传算法的原理,并以实例介绍了遗传算法的具体用法、步骤,对那些数学方法难以奏效的问题提出了一种新的处理方法。

关键词:遗传算法;再生;交叉;变异;算例

中图分类号:TP301.6

文献标识码:A

文章编号:1672-7800(2007)01-0100-02

0 前言

遗传算法(Genetic Algorithms 简称 GA)是由美国 Michigan 大学的 John Holland 教授于 1975 年创建的。它来源于达尔文的进化论、孟德尔的群体遗传学说和魏茨曼的物种选择学说。其基本思想是模拟自然界遗传机制和生物进化论而形成的一种过程搜索最优解的算法。

作为一种新的优化算法,遗传算法的特点是几乎不需要所求问题的任何信息,仅需要目标函数的信息,不受搜索空间是否连续或可微的限制就可找到最优解。依据它的并行性,非常适用于大规模并行计算机。因此,遗传算法广泛地应用于自动控制、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]。

1 遗传算法的基本原理

遗传算法(GA)是建立在自然选择和群体遗传学机理基础上的随机、迭代、进化、具有广泛适应性的搜索方法。GA 搜索结合了达尔文适者生存和随机信息交换的思想,前者消除了解中不适应因素,后者利用了原有解中已知的知识,从而有力地加快了搜索过程。

与传统搜索算法不同,遗传算法从一组随机产生的初始解(称为群体)开始搜索过程,群体中的每个个体是问题的一个解(称为染色体),这些染色体在后续迭代中不断进化(称为遗传)^[4]。

遗传算法主要通过交叉、变异、选择运算实现。交叉或变异运算生成下一代染色体(称为后代),染色体的好坏用适应度来衡量,根据适应度的大小从上一代和后代中选择一定数量的个体,作为下一代群体,再继续进化,这样经过若干代之后,算法收敛于最好的染色体,它很可能就是问题的最优解或次优解。遗传算法中使用适应度这个概念来度量群体中的各个个体。

2 标准算法的基本步骤

GA 的基本思想是把待优化问题的参数编码成二进制位串的形式,然后由若干个位串形成一个初始种群作为待求问题的候选解,使用再生(Reproduction)、交叉(Crossover)和变异(Mutation)3 种算子进行操作,不断迭代优化,直到找到最优解。

经典遗传算法的计算流程如图 1 所示。从图中可以看出,遗传算法是一种种群型操作,该操作以种群中的所有个体为对象。具体求解步骤如下:

(1)编码。确定用何种码制,然后将问题参数编码形成基因码链,每一个码链代

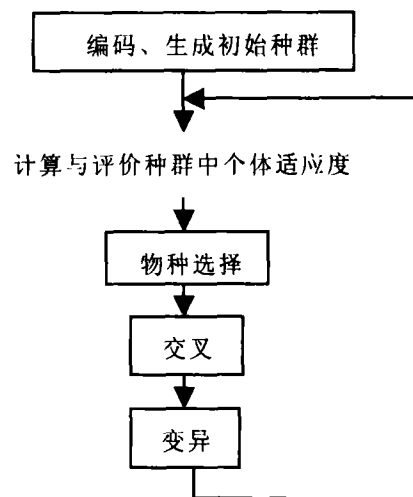


图 1 遗传算法结构图

表一个个体,表示优化问题的一个解。

(2)初始化。随机产生一个规模为 P 的初始种群,其中每个个体为一定长度的码链,该群体代表优化问题的一些可能解的集合。

(3)估计适应度。计算种群中每个个体的适应度,适应度为群体进化时的选择提供了依据。一般来说适应度越高,解的素质越好。适应度函数可以根据目标函数而定。

(4)再生(选择)。根据每个个体的相对适应度,计算每个个体的再生次数,并进行再生操作,产生新的个体加入下一代群

作者简介:唐穗欣(1973-),男,广东汕尾人,武汉理工大学自动化学院硕士研究生,研究方向为智能控制。

体中,一般再生的概率与其适应度成正比。

(5)交叉。从种群中随机选择两个染色体,按一定的概率进行基因交换,交换位置的选取是随机的。

(6)变异。从种群中随机地选择一个染色体,按一定的变异概率 P 进行基因变异;GA 的搜索能力主要是由选择与交叉赋予的,变异算子则保证了算法能搜索到问题空间的每一点,从而使算法具有全局最优性,它进一步增强了 GA 的能力。

(7)若发现最优解,则算法停止,否则转 3,对产生的新一代群体进行重新评价、选择、交叉、变异操作,如此循环往复,使群体中最优个体的适应度和平均适应度不断提高。

3 应用实例

用遗传算法求解 $f(x) = x^2$ ($0 \leq x \leq 31$), x 为整数时 $f(x)$ 的最大值。

(1)编码。在区间 $[0,31]$ 上的整数可以用一个 5 位的二进制位串进行编码, x 的值直接对应二进制位串的数值,如:9 对应 01001,31 对应 11111。

(2)初始化。在 $[0,31]$ 范围内按某种分布,随机产生一个规模为 P 的初始种群,本例子中假定初始种群取为 4 个二进制串, $x_1=01100, x_2=11001, x_3=01000, x_4=10010$ 。以上 5 位字串称为染色体,每个分量称为基因,每个基因有两种状态 0 或者 1。

(3)计算适应度。计算种群中每个个体的适应度,本题中,适应度函数可定为: $\text{fitness}(x) = f(x) = x^2$, 于是, $\text{Fitness}(x_1) = 144$, $\text{Fitness}(x_2) = 625$, $\text{Fitness}(x_3) = 64$, $\text{Fitness}(x_4) = 400$ 。

(4)再生(选择)。初始种群中每个个体入选种群的概率为: $p(x_i) = \text{fitness}(x_i) / \sum \text{fitness}(x_i)$ 。利用上述概率,对这四个整数进行四次有放回的随机抽取,产生四个新的整数,显然概率大的有更多的机会被选中,四个新的整数为 $x_1'=01100, x_2'=11001, x_3'=11001, x_4'=10010$ 。经前四步后,具体情况如表 1 所示:

(5)交叉。将 $x_i'(i=1,2,3,4)$ 随机结合成两组,每组两个数;对每组再进行一次随机抽样,以等概率从 1,2,3,4 中选取一个数 n 。假设在上述例子中恰恰 $x_1'=01100, x_2'=$

表 1

序号	初始种群	X 值	适应度	入选种群概率	期望复制数	实际复制数
1	01100	12	144	12%	0.48	1
2	11001	25	625	51%	2.04	2
3	01000	8	64	5%	0.2	0
4	10010	20	400	32%	1.28	1
合计			1233	100%	4	4
平均			308.3	25%	1	1

11001 分为一组,随机抽取 $n=3$,那么将由二进制表示的 x_1', x_2' 从后三位开始互换,得到两个新的二进制整数,记为 x_1'', x_2'' 。同样对另外一组数作类似处理,假设另一组抽得 $n=2$,这样得到四个新的整数,结果为: $x_1''=01001, x_2''=11100, x_3''=11010, x_4''=10001$ 。具体情况如表所示:

表 2

序 号	复制后 种群	复制 对象	交叉 点 n	交叉后 种群	X 值	适应度
1	01100	2	3	01001	9	81
2	11001	1	3	11100	28	784
3	11001	4	2	11010	22	484
4	10010	3	2	10001	17	289
合计						1638
平均						409.5

(6)变异。将交叉得到的四个二进制数,每个数每个二进制位进行一次随机抽样,以一个小概率 P ,例如 $1/1000$ 将该位取反,即由 0 变为 1,由 1 变为 0,以概率 $1-P$ 保持该位数字不变。这样得到四个新的数记为 $x_i^{*}(i=1,2,3,4)$,计算其函数值 $f(x_i^{*})$,回到步骤(4)。对产生的新一代群体进行重新计算适应度、选择、交叉、变异操作,如此循环往复,使群体中最优个体的适应度和平均适应度不断提高。变异可保持群体的多样性,可使遗传算法跳出局部极值点^[9]。

以上的简单例子,从开始随机产生的种群到经过一次再生、交叉、变异操作后的新种群中,我们不难看出初始种群的平均适应度为 308.3,新种群的平均适应度为 409.5,最大值从初始种群的 625 增加到新种群的 784。可见经过一次遗传算法的操作后,问题的解向最优解的方向前进了一大步。因此经过以上多次类似计算,问题的解最终会接近最优解。

当然在应用遗传算法时,可以事先规定一个最大允许循环次数,以使计算得到

终结,而以计算过程达到的最大函数值的 x^n 作为所要的解,遗传算法实际上是一种搜索方式,对那些标准数学方法难以奏效的问题给出了一种处理办法^[6]。

4 结束语

遗传算法具有计算简单及功能强的特点,它对于搜索空间基本上不需要什么限制性假设,和传统的优化算法相比,有以下优点:

(1)遗传算法的处理对象不是参数本身,而是对参数集进行了编码的个体,这种操作使遗传算法可直接对结构对象进行操作。这样一个特点,使得遗传算法具有广泛的应用领域。

(2)采用同时处理群体中多个个体的方法,这一特点使遗传算法具有较好的全局搜索性能,减少了局部优解的风险,

(3)通过适应度函数这一信息进行搜索,对寻优的函数基本无限制,既不要求函数连续,也不要要求函数可导,而且其定义域可以任意设定。

(4)使用的复制、交叉、变异这三个算子都是随机操作,而不是确定规则,该法虽采用随机操作,但并不是盲目的随机漫游,而是一种启发式搜索,其效率优于其它方法。

(5)遗传算法具有隐含并行性,鲁棒性强,可以通过大规模并行计算提高计算速度,使其更适合大规模复杂问题的优化。

参考文献:

- [1]周明,孙树栋.遗传算法原理及应用[M].北京:国防工业出版社,2001.
- [2]陈国良等.遗传算法及其应用[M].北京:人民邮电出版社,2001.
- [3]施光燕,董加礼.最优化方法[M].北京:高等教育出版社,2003.
- [4]胡玉锁,陈宗海.基于混合遗传算法的聚类分析,模式识别与人工智能[J].2001.
- [5]全惠云,邹秀芬,康立山.计算方法与应用软件[M].武汉大学出版社,2002.
- [6]Liu Li, Chen Xueyun. Reconfiguration of distribution networks based on fuzzy genetic algorithms [J] Zhong-guo Dianji Gongcheng Xuebao, 2003, (2).

(责任编辑:杜能钢)