

文章编号 :1009 - 2269(2006)03 - 0001 - 04

遗传算法中适应度函数的研究^{*}

刘 英

(西北师范大学 数学与信息科学学院,甘肃 兰州 730070)

摘要:通过分析遗传算法中常见的几种适应度函数的不足,论证了适应度函数在遗传算法中的重要性,提出了设计适应度函数应满足的标准,在此基础上给出了一适应度函数公式.实验结果表明:此适应度函数的性能明显优于其它函数,对提高遗传算法的整体性能也有重要意义.

关 键 词: 遗传算法;适应度函数;最优设计;性能分析

中图分类号: TP 301.6

文献标识码: A

0 引 言

遗传算法是模拟生物在自然环境中遗传和进化过程而形成的一种自适应全局化概率搜索算法.它最早是由美国密执安大学的 Hollan 教授提出,起源于 60 年代,对自然和人工自适应系统的研究^[1].遗传算法对包含可能解的群体反复使用遗传学的基本操作,不断生成新的群体,使种群不断进化,同时以全局并行搜索技术来搜索优化群体,以取得满足要求的最优个体,得到满足要求的最优解.因此求解复杂函数的最优化问题是遗传算法(Genetic Algorithms, GAs)的一个重要研究方向.

众所周知,用 GAs 求解函数最优化问题时,是依靠适应度函数值的大小来区分每个个体的优劣的.适应度值大的个体将有更多的机会繁衍下一代,通常取高于群体平均适应度值的个体做交叉,而低于平均适应度值的个体做变异,从而一代一代地提高群体的平均适应度值和最优个体的性能.可见,适应度函数在 GAs 中起着决定性作用.为此,本文以函数最优化问题为背景,在现有结果的基础上对适应度函数做进一步研究.

1 适应度函数分析

适应度函数(Fitness Function)的选取直接影响到遗传算法的收敛速度以及能否找到最优解^[2],因为遗传算法在进化搜索中基本不利用外部信息,仅以适应度函数为依据,利用种群每个个体的适应度来进行搜索.因为适应度函数的复杂度是遗传算法复杂度的主要组成部分,所以适应度函数的设计应尽可能简单,使计算的时间复杂度最小.遗传算法评价一个解的好坏不是取决于它的解的结构,而是取决于该解的适应度值,这正体现了遗传算法“优胜劣汰”的特点.遗传算法不需要适应度函数满足连续可微等条件,唯一要求是针对输入可计算出能加以比较的非负结果.这一特点使得遗传算法具有广泛的适用性.在实际问题中,适应度函数与问题的目标函数是不完全一致的,如有的问题的目标是要求得最小值(费用问题),而有的问题的目标是要求得最大值(利润函数).因此在不少场合,将目标函数映射成求最大值形式而且函数值非负的适应度函数是必要的.

^{*} 收稿日期:2006 - 03 - 22
基金项目:甘肃省自然科学基金(3ZS051 - A25 - 047)
作者简介:刘 英(1979 -),女,甘肃兰州人,硕士生.

1.1 几种常见的适应度函数及其不足

在函数优化中,适应度函数可由目标函数变换得到,一般而言有以下 3 种定义形式:

1) 直接以待求解的目标函数转化为适应度函数,即

若目标函数为最大化问题 $f(x) = g(x)$;

若目标函数为最小化问题 $f(x) = -g(x)$.

这种适应度函数简单直观,但实际应用时,存在以下 2 个问题:第一,不满足常用的赌盘选择非负的要求;第二,某些待求解的函数值可能彼此相差十分悬殊,由此得到的平均适应度值可能不利于体现群体的平均性能,将影响算法的效果.

2) 若目标函数为最小问题,则

$$f(x) = \begin{cases} c_{\max} - g(x), & g(x) < c_{\max} \\ 0, & \text{其他情况} \end{cases}$$

其中系数 c_{\max} 存在多种选择方法,它可以是一个合适的输入值,也可以采用迄今为止进化过程中 $g(x)$ 的最大值,但 c_{\max} 最好与群体无关.由于参数 c_{\max} 需事先预估,不可能精确,其结果常常是适应度函数不灵敏,影响了算法的性能.

若目标函数为最大问题,则

$$f(x) = \begin{cases} g(x) - c_{\min}, & g(x) > c_{\min} \\ 0, & \text{其他情况} \end{cases}$$

其中系数 c_{\min} 存在多种选择方法,它可以是一个合适的输入值,也可以采用迄今为止进化过程中 $g(x)$ 的最小值,但 c_{\min} 最好与群体无关.

3) 若目标函数为最小问题,则

$$f(x) = \frac{1}{g(x) + c + 1}, \quad c = 0, c + g(x) = 0$$

若目标函数为最大问题,则

$$f(x) = \frac{1}{1 + c - g(x)}, \quad c = 0, c - g(x) = 0$$

上两式 c 为目标函数界限的保守估计值.由于事先不知道 $\min g(x)$,故 c 的取值只能采取保守的估计值,存在和第二种适应度函数相似的问题.适应度函数对 GA 的收敛速度和结果影响很大.如果过分强调当前的较优点,就可能很快降低种群的多样性,造成不成熟收敛;如果对当前较优点强调不够,算法就很容易丢失已经找到的较优点信息,从而不能在合理的时间内收敛到较好的点.

1.2 适应度函数的作用

在遗传算法中,适应度是描述个体性能的主要指标^[3].根据适应度的大小,对个体进行优胜劣汰.适应度是驱动遗传算法的动力.从生物学角度讲,适应度相当于“生存竞争、适者生存”的生物生存能力,在遗传过程中具有重要意义.将优化问题的目标函数与个体的适应度建立映射关系,即可在群体进化过程中实现对优化问题目标函数的寻优.

适应度函数也称评价函数,是根据目标函数确定的用于区分群体中个体好坏的标准,总是非负的,任何情况下都希望它的值越大越好.在选择操作中,会出现 2 个成为遗传算法欺骗的问题:

1) 在遗传算法初期,通常会产生一些超常个体,按照比例选择法,这些超常个体会因竞争力突出,而控制选择过程,影响到算法的全局优化性能;

2) 遗传算法后期,当算法趋于收敛时,由于种群中个体适应度差异较小,继续优化的潜能降低,可能获得某个局部最优解.

因此,如果适应度函数选择不当就会产生以上的欺骗问题.可见适应度函数的选择对于遗传算法的意义重大.

2 适应度函数的设计

适应度函数的设计主要满足以下条件:

1) 单值、连续、非负、最大化.

适应度函数 $Fit(f(x))$ 应该是实函数,并且单值、连续,但不要求可导.不过, $Fit(f(x))$ 的曲线在重要部位,特别在最优解附近一般不宜太陡也不宜过于平缓.

2) 合理、一致性.

是指适应度函数曲线上,各点的适应度值应与解的优劣成反比例,即

$x_1, x_2, (x_1, x_2) \in [lmin, lmax] \quad f(x_1) < f(x_2) \quad Fit(f(x_1)) > Fit(f(x_2))$, 其中 $[lmin, lmax]$ 是函数 $f(x)$ 的定义域.

3) 计算量小.

$Fit(f(x))$ 不应设计得过于繁复,应在上述条件下越简单越好.

4) 通用性.

一个适应度函数的好坏,还应满足尽可能广泛的通用性,使用户在求解种种问题时,最好无需改变适应度函数中的参数.通用性要求是对适应度函数设计的更高一层的要求.它能使用户在对所求解函数的全局最优解的性质完全“无知”的情况下,由算法在运行过程中自动修正其中的参数值,从而一步一步接近最优解.从另一种意义上说,这样的适应度函数具有自适应性.本文建议的一类适应度函数定义为

$$Fit(f(x)) = \begin{cases} 1 - 0.5 \times [| \frac{f(x) - b}{a} |], & | f(x) - b | < a \\ \frac{1}{1 + [| \frac{f(x) - b}{a} |]}, & | f(x) - b | \geq a \end{cases}$$

理想情况下: b 的值是 $\min f(x) = y^*$, 当适应度值为 0.5 时, 是 $f(x)$ 到 $\min f(x)$ 的距离. 考虑到适应度函数的不同应用场合, 本文将 a 值取为 2, 将 b 值分别取为 1, 1.5, 0.5. 即可在 b, a 取定的情况下得到 3 种适应度函数.

当取 $a = 1$ 时, 适应度值在 $[0.5 \sim 1]$ 之间是线性的; 而对于在全局最优解 y^* 附近变化比较缓慢的函数, 用 $a = 0.5$ 可以使适应度函数较灵敏地反映出 y 值的变化情况. 在算法的后期, 则可以有效地拉开最优解附近点的适应度值, 便于做出敏感选择, 从而有利于以后的选择;

当 $a = 1.5$ 时的适应度函数则有相反的适应情况.

本文建议公式里的 b 和 a 随遗传算法的下一代进化而不断地修正. b 的值可取当前第 i 代中的最小值, 即 $b = f(x) * i$; 而 a 则建议用公式

$$a = \max \left[0.5, \frac{(\max f_i - f_i^*)}{30} \right]$$

求取, 若要求适应度函数曲线变化得“陡”一些, 只需把公式中的相关系数调节得再小一些即可 (例如将分母变大); 同样, 若要求适应度函数曲线“缓”一点, 可以把系数适当调大一些 (例如将分母变小).

3 实验结果及结论

为考察该适应度函数的性能, 本文选用了一组测试 GAs 性能的经典函数, 包括多峰、非凸、病态等常规方法难于求解的函数对适应度函数进行测试. 算法对每个测试函数均分别独立地运行 60 次, 求出的均为全局最优解. 实验结果表明, 采用文中的适应度函数的算法性能比一般算法性能有很大的提高. 限于篇幅, 这里不再给出具体的实验结果.

4 结论

本文考察了几种常见的适应度函数各自存在的不足以后,提出了设计适应度函数的几条标准,并在此基础上依据智能化和自适应的思想,提出了本文建议的一种动态变化的适应度函数公式.理论分析与各项测试都表明,本文建议的适应度函数的性能比以往的几种适应度函数均有提高,并且对遗传算法的整体性能的提高同样有重要作用.

参考文献:

- [1] 陈国良,王煦法,庄镇泉,等.遗传算法及其应用[M].北京:人民邮电出版社,1999.
- [2] 刘勇,康立山,陈镇屏.非线性并行算法——遗传算法:第2版[M].北京:科学出版社,1995.
- [3] 边肇祺.模式识别[M].北京:清华大学出版社,1988.

Research on Fitness Function in Genetic Algorithm

LIU Ying

(The Mathematics and Information College of Northwest Normal University ,Lanzhou 730070 ,China)

Abstract : After analysing some drawback of familiar fitness functions ,this paper discusses the importance of fitness function in GAs. The most important is that it proposes four standards that fitness function must satisfy. At last ,this article gives its own fitness function ,and the testing results show that the performance is much better than the others and it is important to improve the whole performance of genetic algorithm.

key words : GAs ;fitness function ;optimised design ;performance analysis