

遗传算法理论研究综述^{*}

陈建安

郭大伟

徐乃平

孙云芝

(西安电子科技大学电子工程学院 西安 710071)

(西北工业大学自动控制系统 西安 710072)

摘要 通过遗传算法的描述,介绍了有关遗传算法理论问题的研究现状,如:参数设置的选择、编码方法、遗传算子、模式定理、收敛性分析、欺骗问题、并行计算、局部改进和混合方法等。尽管遗传算法在其应用中还存在某些问题,但通过遗传算法的理论研究将有助于遗传算法理论及其应用的发展。最后,简单地总结了遗传算法的研究趋势。

关键词 遗传算法 理论进展 研究综述

中图分类号 TP18

遗传算法作为一种非确定性的拟自然随机优化算法在过去 20 年中得到了广泛的应用。例如:复杂函数系统优化、机器学习、系统识别、故障诊断、分类系统、控制器设计、神经网络设计、自适应滤波器设计等。Goldberg 的专著和 Davis 的专著都对遗传算法的应用有更详细的论述。遗传算法的原理和数学基础是 Holland 在 1975 年发表的。

尽管遗传算法在很多领域具有广泛的应用价值,但它也存在一些问题。这些问题迫使研究者对遗传算法的理论需要作进一步的研究。目前遗传算法的主要研究方向为应用研究和理论研究。文中首先对遗传算法进行简单的描述,然后对目前遗传算法自身理论、遗传算法与其他技术混合方法及相关功能扩展技术等综述,最后,简单地总结了遗传算法的研究趋势。

1 遗传算法的描述

遗传算法可以认为是一个进化过程(即迭代过程)。首先根据具体问题建立评价群体的评价函数。为了测试个体的适应性,可以用评价函数对染色体进行评价,也可以用目标函数的数学公式对染色体进行评价。遗传算法的描述如下:

(1) 随机选择 N 个初始点组成一个群体,群体内的每个点叫一个个体,或叫染色体。群体内个体的数量 N 就是群体规模。群体内每个染色体必须以某种编码形式表示。编码的内容可以表示染色体的某些特征,随着求解问题的不同,它所表示的内容也是不同。通常染色体表示被优化的参数,每个初始个体就表示着问题的初始解。

(2) 按照一定的选择策略选择合适的个体,选择体现“适者生存”的原理。根据群体中每个个体的适应性值,从中选择具有最好的 M 个个体作为重新繁殖的下一代群体。

(3) 以事先给定的杂交概率 P_c 在选择出的 M 个个体中任意选择两个个体进行杂交运算或重组运算,产生两个新的个体,重复此过程直到所有要求杂交的个体杂交完毕。杂交是两个染色体之间随机交换信息的一种机制。

(4) 根据需要可以以事先给定的变异概率 P_m 在 M 个个体中选择若干个体,并按一定的策略对选中的个体进行变异运算。变异运算增加了遗传算法找到最优解的能力。

^{*} 收稿日期: 1997-05-16

(5) 检验停机条件,若满足收敛条件或固定迭代次数则停机;若不满足条件则转(2)重新进行进化过程.每一次进化过程就产生新一代的群体.群体内个体所表示的解经过进化最终达到最优解.

遗传算法是基于自然的选择和遗传机制的搜索算法,它与传统的搜索算法之间有很多差别,归纳为(1)遗传算法以编码形式工作,可以并行搜索多个峰值而不是一个点.(2)遗传算法把问题的参数集表示成个体,并以编码的形式运行,而不是对参数本身进行求解.(3)遗传算法用概率传递规则代替确定性规则.(4)遗传算法只使用目标函数信息,而不使用推导过程或其他辅助知识.遗传算法具有运算并行性,它可以在复杂的搜索空间内同时评价多个点,这样有利于在多值空间寻找全局最优解.遗传算法关心每次进化群体中个体的质量,即问题解的质量.这不同于许多优化方法需要递推信息或需要知道问题的结构和参数的全部知识.因此,遗传算法尤其适合不确定问题或非线性复杂问题的求解.

2 遗传算法的主要研究内容

遗传算法强有力的搜索和优化技术使得它被广泛地应用到许多领域.但在不同的应用领域中,人们发现使用遗传算法会有很多问题,为此人们提出各种解决方法.下面把遗传算法的理论研究归纳几个部分分别论述.

2.1 参数设置的选择

遗传算法的参数设置包括群体规模、迭代次数、杂交概率和变异概率等.目前对遗传算法的参数设置的合理选择还缺少相应的理论作为指导.由于参数设置关系到遗传算法的精度、可靠性和计算时间等诸多因素,并且影响结果的质量和系统性能,因此,参数设置的研究受到重视.

孙艳丰证明^[1],当采用自然数编码时,从理论上可以证明遗传算法的最优群体规模的存在性,并给出相应的计算方法.但在大多数情况下,群体规模的大小是很难进行计算,理论上讲不同的解题过程应该有不同最佳群体规模.如果群体规模太小,由于群体中的模式缺少使得遗传算法受到搜索空间的限制.如果群体规模太大,则搜索和优化需要花费大量的时间.

迭代次数的设置分为固定和不固定两种设置.固定迭代次数有利于遗传算法的处理,但设置选择困难,并且不利于产生最优解.不固定迭代次数通过对个体解的判断自动进行迭代有利于产生最优解,并且解决了参数选择的困难,但容易增加遗传算法的处理时间,尤其是在算法发散时.最佳群体规模和迭代次数的设置研究目前很少有报道.通常,群体规模和迭代次数的设置是根据经验或实验进行的.

对杂交概率 P_c 和变异概率 P_m 的设置,由于它们关系到遗传算法的收敛性和群体中个体的多样性,即探索和开发问题,因此倍受重视.传统的 P_c 和 P_m 的设置是静态人工设置.现在有人提出动态参数设置的方法,以减少人工选择参数的困难和盲目性. Srinivas et al. 提出用自适应杂交概率和变异概率来实现两个目的^[2]: 维持群体的多样性和保证遗传算法的收敛性.探索问题是指在寻找全局最优解时能探索新的解空间区域;开发问题是指在探索到最优解区域时能收敛到最优解.两者的平衡由 P_c 和 P_m 值决定. P_c 和 P_m 值增加,则以开发为代价提高探索过程.增加探索过程也就增加个体的变化,却降低了个体适应性值.这有利于探索各种解空间,保证解的全局性,但不利于群体收敛到最优解.而 P_c 和 P_m 值降低,则以探索为代价提高开发过程.增加开发过程即增加个体适应性值,减少个体的变化,有助于群体收敛到最优解,但不能保证全局最优解.因此由算法自动设置 P_c 和 P_m 是必要的. Davis 提出自适应算子概率的方法^[3],即用自适应机制把算子概率的选择与算子产生的个体适应性结合,高适应性值被分配高算子概率. Fogarty 研究了变异率对整数编码的影响^[4]. Whiley 研究了由父辈解之间的 Hamming 距离来决定 P_m 的设置^[5]. 这是一个变化的参数设置.

2.2 编码方法

遗传算法对被优化的参数(用个体表示)进行编码,并以编码方式运算.对参数编码的目的是把一个优化问题变成一个组合问题.根据应用的不同,遗传算法的搜索空间可以分为连续空间和离散空间,分别称为连续遗传算法和离散遗传算法(即标准遗传算法).在离散遗传算法中,个体编码常用的方法是有限长二进制编码.其优点是有助于遗传算法建立数学模型,并便于诸如模式定理、收敛性分析、欺骗问题

等理论分析和计算机处理,但不利于人们的习惯表达;在连续遗传算法中,个体的编码有整数编码和实数编码.从提高系统精度考虑使用实数编码更适合连续问题的数据优化、加速搜索、更容易开发与其他技术的混合方法.但使用实数编码缺少理论支持,并且原有遗传算子不再适用,需要开发新的遗传算子. Feng 对连续遗传算法的理论进行了分析^[6]. Janikow et al.给出二进制编码和实数编码表示的实验比较^[7].孟庆寿提出对称编码^[8],并通过优良型保护、希望型成员移民和部分基因保留等措施,以期加快算法的收敛速度.

2.3 遗传算法的 3 个算子

遗传算法的 3 个算子分别是选择、杂交和变异.选择体现“适者生存”的原理,通过适应性值选择优质个体而抛弃劣质个体.杂交能使个体之间的遗传物质进行交换从而产生更好的个体.变异能恢复个体失去的或未开发的遗传物质,以防止个体在形成最优解过程中过早收敛.下面分别介绍选择、杂交和变异 3 个算子的主要研究内容.

2.3.1 选择算子

大量事实证明改进选择过程的方法可以提高遗传算法的性能,而选择过程不是唯一的.常用的选择方法有:精华选择、重组选择、均分选择、适应性调整、线性排序、比例选择、自适应选择等.选择是一个重要过程,因为选择不当,容易失去许多重要的个体,从而影响算法的收敛性. Potts et al.概括了大约 20 多样选择方法^[9].为了防止算法过早收敛的发生, Potts 提出微进化结构和人工选择等新算子来改进遗传算法. Kuo et al.采用分裂选择方法^[10]使新的遗传算法比传统的遗传算法在优化欺骗函数上具有更好的特性. Back 在研究各种不同的选择机制中提出扩展选择和偏置选择^[11]. Back 把选择阶段分为选择算法和采样算法.选择算法给每个个体指定一个被叫做目标采样率的实数,然后,根据目标采样率的大小,采样算法再生复制个体形成一个新的群体.

2.3.2 杂交算子

杂交是两个个体之间随机交换信息的一种操作.当两个个体之间进行杂交操作时,由于杂交通过个体传播可以发生模式的破坏作用,因此研究杂交技术对减少杂交的破坏作用具有重要意义.常用的杂交技术有:一点杂交、二点杂交、均匀杂交、多点杂交、启发式杂交、顺序杂交、混合杂交等. Potts 概括大约 20 种已有的杂交技术^[9]. Feng et al.对多维连续空间的遗传算法的杂交多样性进行了分析^[6].通过建立相应的数学模型, Feng 解释了在多维连续空间和大规模群体中使用均匀杂交算子是如何探索新的解空间区域.

2.3.3 变异算子 (Mutation Operation)

变异是随机改变某个个体遗传信息的一种操作.即随机地对个体某一位或某几位进行变异 $\theta \rightarrow 1$ 或 $\downarrow \rightarrow 0$.变异修改个体的适应性值. Potts 总结了 3 种变异技术:管理变异、变化的变异概率和单值运算^[9].张良杰等人通过引入 i 位改进子空间的概念,采用模糊推理技术来确定选取突变概率的一般性原则^[12]. Feng et al.通过连续遗传算法的理论分析,提出随时间变化的变异技术^[6],即根据群体的平均适应性值的高低确定变异变化率.类似的研究还有 Fogarty 提出的变化的变异概率^[4].有关变异算子的研究大部分是与变异概率的研究一起进行的.

2.4 模式定理

模式定理的一般数学表达式为

$$M[H(t+1)] \geq M[H(t)] \frac{f(H(t))}{f(t)} \left[1 - \frac{P_c d(H)}{L-1} \right] (1 - P_m)^{O(H)}.$$

其中 $M[H(t+1)]$ 是 $t+1$ 次迭代群体中模式 H 的期望的个数; $M[H(t)]$ 是 t 次迭代模式 H 的期望的个数; P_c 是杂交概率; P_m 是变异概率; $d(H)$ 是模式 H 的长度; $O(H)$ 是模式 H 的阶; L 是每个个体编码后的长度; $f(t)$ 是整个群体的平均适应性值; $f(H)$ 是群体中模式 H 的平均适应性值:

$$f(H) = \frac{\sum_{\{i, x_i^t \in H\}} f_i}{M[H(t)]},$$

f_i 是第 i 个个体的适应性值. 当误差项 $\{1 - [P_c d(H) / (L - 1)]\} (1 - P_m)^{O(H)}$ 很小时, 如果 $f(H)$ 大于 $f(t)$, 则 $M[H(t+1)]$ 增加. 误差项表示由杂交和变异引起模式 H 的破坏效应, 是负效应, 而模式定理强调选择的作用. 模式定理只对二进制编码适用, 而对其他编码形式不适合.

模式定理是分析遗传算法的主要方法. 一个模式 H 是指编码空间 (即描述解空间区域) 的编码的子集. 一个模式的变化受到诸多因素的影响, 如适应性函数、编码长度、群体规模、杂交率、变异率等. 随着逐步迭代的模式进化可使遗传算法达到最优解 (即使染色体可以获得较高的适应性值). 尽管遗传算法仅作用于 N 个编码组成的种群, 但这 N 个编码实际上包含 $O(N^3)$ 阶个模式的信息. 这一性质被称为遗传算法的隐含并行性. 孙艳丰^[13] 对模式定理进行了分析: 模式定理揭示经过选择、杂交、变异后, 模式 H 在 $K+1$ 代的数目与第 K 代的数目的关系, 并讨论无杂交时的模式定理. Srinivas 讨论了自适应 P_c 和 P_m 下的模式定理的表示方法^[2].

2.5 收敛性分析

一般收敛性分析是通过计算群体的收敛性进行的. 设适应性函数 $f(x)$, $\forall x \in H, \exists X > 0$, 使 $\|x - x^*\| \leq X$ 有 $f(x) \geq f(x^*)$, $x^* \in H$. 其中 $f(x^*)$ 是局部最小值, x^* 是群体中具有局部最小值的个体. 若有 $f(x) \leq f(x^*)$, $x^* \in H$, 那么 $f(x^*)$ 是局部最大值, x^* 是群体中局部最大值的个体; $\forall x \in H$, 有 $f(x) \geq f(x^*)$, $x^* \in H$, 其中 $f(x^*)$ 是全局最小值, x^* 是群体中具有全局最小值的个体. 若有 $f(x) \leq f(x^*)$, $x^* \in H$, 那么 $f(x^*)$ 是全局最大值, x^* 是群体中全局最大值的个体.

Eiben 等证明了遗传算法在精英选择方式下的收敛性^[14]. Rudolph et al. 给出一般遗传算法的收敛性分析^[15], 并证明如果选择算子采用自然选择策略, 则算法不收敛, 若记录每次迭代中的最佳个体, 则改进的算法收敛; Kuo 针对分裂选择在解决欺骗问题时给出其收敛分析^[10]; Feng 给出连续遗传算法的收敛性分析^[6]. 徐宋本等分析过早收敛的原因是由于杂交算子引起^[16], 并基于这一理论的解释提出了对遗传算法的修正. 遗传算法的收敛性分析有助于遗传算法的性能评价, 如收敛速度、全局收敛等问题. 目前, 较多的收敛性分析和研究都是针对具体问题进行的.

2.6 欺骗问题

遗传算法依据个体的选择、杂交、变异找到最优解. 如果一个问题的个体编码满足建筑块条件, 那么用遗传算法求解效率较高, 否则用遗传算法求解效率较低. 低阶的建筑块错误地引导搜索过程, 使遗传算法不能发现高阶建筑块, 最终导致算法发散, 找不到最优解. 这一现象称为欺骗问题. 欺骗问题实际就是要预测给定问题用遗传算法求解的难易程度. Goldberg 曾设计一个离散遗传算法的最小欺骗问题^[17]; 目前遗传算法的欺骗问题研究主要集中在 3 个方面: 设计欺骗函数; 理解欺骗函数对遗传算法的影响; 修改遗传算法以解决欺骗函数的影响.

2.7 并行计算

遗传算法在求解过程中, 其计算过程可以是并行计算, 而且遗传算法本身就具有并行计算能力^[19], 只是在应用遗传算法中人们较少使用并行计算. Zeigler 把高性能并行遗传算法应用到大型并行计算机^[18]. Bertoni et al. 研究遗传算法中的内在并行性^[19].

2.8 遗传算法的局部改进与混合方法

由于遗传算法涉及精度、可靠性、计算时间、探索与开发等诸多问题, 通过改进遗传算法本身在某种程度上可以提高遗传算法求解问题的性能, 但仅仅这样改进是不够的. Miller 使用混合方法^[20], 即用两种或多种不同算法组成的方法, 来提高全局优化和改进局部优化; Renders 把传统的方法与遗传算法结合用于解决计算精度、可靠性和计算时间之间的矛盾^[21]. Grefenstette 提出了把特殊知识与遗传算法相结合的方法^[25].

遗传算法与进化策略、进化程序设计被认为是进化算法的 3 个主要算法. 由于遗传算法的并行特点, 使得遗传算法与计算机的结合可以产生巨大的优越性. 陈火旺论述了遗传程序设计的方法和理论^[22]. Fogel 在进化算法的综述中认为遗传算法、进化策略和进化程序设计分别强调自然进化的不同方面^[23]. 遗传算法强调个体的运算; 进化策略强调个体水平的行为变化; 进化程序设计强调种类水平的行

为变化, 可以认为进化策略和进化程序设计都是遗传算法功能的扩展和补充. 随着模糊逻辑的应用, Buckley 提出了模糊遗传算法并应用其解决实际问题^[24].

3 结 束 语

遗传算法作为一种非确定性的拟自然算法为复杂系统的优化提供了一种新的方法, 并且经过实践证明效果显著. 尽管遗传算法在很多领域具有广泛的应用价值, 但它仍存在一些问题, 如遗传算法很多机制和现象还缺少理论的指导, 这都需要对遗传算法做进一步的研究. 目前, 遗传算法的研究可以归纳为以下几点:

(1) 如果说目前遗传算法很多机制和现象还缺少理论的指导, 那么发展一些启发式策略或其他方法对搜索过程给予指导, 则是遗传算法的研究方向之一.

(2) 加强遗传算法自身理论的研究, 如参数的自动设置; 编码方式及相应的模式定理; 收敛性分析; 选择策略; 并行计算; 欺骗问题等, 是完善遗传算法理论不可缺少的内容.

(3) 补充和扩展遗传算法与其他算法或其他技术的混合方法, 扩展遗传算法的功能和应用, 则是遗传算法研究的另一方向.

(4) 遗传算法理论研究的目的很大程度上是为了使遗传算法具有更广泛、更容易、更可靠的应用. 因此, 遗传算法作为人工生命中关键技术, 加强其应用研究, 使其发挥更大的作用, 是研究者非常重要的研究内容之一.

参 考 文 献

- 1 孙艳丰, 王众托. 自然编码遗传算法的最优群体规模. 信息与控制, 1996, 25(5): 317~ 320
- 2 Srinivas M, Patnaik L M. Adaptive Probabilities of Crossover and Mutation in Genetic Algorithm. IEEE Trans Syst Man, Cybern, 1994, 24(4): 656~ 667
- 3 Davis L. Adaptive Operator Probabilities in Genetic Algorithms. In Proc 3rd Int Conf Genetic Algorithms, 1989. 61 ~ 69
- 4 Fogarty J C. Varying the Probability of Mutation in Genetic Algorithms. In Proc 3rd Int Conf Genetic Algorithms, 1989. 42~ 50
- 5 Whitley D, Starkweather D. Genetic II: A Distributed Genetic Algorithm. Expt J Theory, Art Intell. 1990, 2: 189~ 214
- 6 Feng Q X, Francesco P. Theoretical Analysis of Evolutionary Algorithms with an Infinite Population Size in Continuous Space Part I and II: Basic Properties of Selection and Mutation. IEEE Trans Neural Networks, 1994, 5(1): 102~ 129
- 7 Janikow C I, Michalewicz J. An Experimental Comparison of Binary and Floating Point Representation in Genetic Algorithms. In Proc 4th Int Conf Genetic Algorithms, 1991. 31~ 36
- 8 孟庆寿. 带有对称编码的基因算法的研究. 电子学报, 1996, 24(10): 27~ 31
- 9 Potts J C, et al. The Development and Evaluation of an Improved Genetic Algorithm Based on Migration and Artificial Selection. IEEE Trans Syst Man, Cybern, 1994, 24(1): 73~ 86
- 10 Kuo T, Hwang S Y. A Genetic Algorithm with Disruptive Selection. IEEE Trans Syst Man, Cybern, 1996, 26(2): 299~ 306
- 11 Back T, Hoffmeister F. Extended Selection Mechanisms in Genetic Algorithms. In Proc 4th Int Conf Genetic Algorithms, 1991. 89~ 99
- 12 张良杰, 毛志宏, 李衍达. 遗传算法中突变算子的数学分析及改进策略. 电子科学学刊, 1996, 18(6): 590~ 595
- 13 孙艳丰, 王众托. 关于遗传算法模式定理的分析研究. 控制与决策, 1996, 11(1): 221~ 224
- 14 Eiben A E, Aarts E H L, Hee K M V. Global Convergence of Genetic Algorithms A Markov Chain Analysis. Schwefel H P, Manner R. Eds, Springer-Verlag, 1990. 4~ 12

- 15 Rudolph G. Convergence Analysis of Canonical Genetic Algorithms. IEEE Trans Neural Networks, 1994, 5(1): 96~101
- 16 徐宋本,高 勇. 遗传算法过早收敛现象的特征分析及其预防. 中国科学(E), 1996, 26(4): 364~375
- 17 Goldberg D E. Genetic Algorithms and Wolsh Function: Part II. Deception and Its Analysis, Complex Syst, 1989, 3: 153~171
- 18 Zeigler B P, Kim J. Asynchronous Genetic Algorithms on Parallel Computers. In Proc 5th Int Conf Genetic Algorithms, 1993
- 19 Bertoni A, Dorigo M. Implicit Parallelism in Genetic Algorithms. In Art Intell, 1993
- 20 Miller J A, et al. An Evaluation of Local Improvement Operators for Genetic Algorithms. IEEE Trans Syst Man, Cybern, 1993, 23(5): 1340~1351
- 21 Renders J M, Flasse S P. Hybrid Methods Using Genetic Algorithms for Global Optimization, IEEE Trans Syst Man, Cybern, 1996, 26(2): 243~258
- 22 陈火旺,吴少岩,罗铁庚. 遗传程序设计(之二). 计算机科学, 1996, 23(1): 12~15
- 23 Srinival M, Patnaik L M. Genetic Algorithms: A Survey. Computer, 1994, 23: 17~26
- 24 Buckley T J, Hayashi Y. Fuzzy Genetic Algorithms and Applications, Fuzzy Sets and Systems, 1994, 61: 129~136
- 25 Grefenstette J J. Incorporating Problem Specific Knowledge into Genetic Algorithms. In Genetic Algorithms and Simulating Annealing Davis L, Ed. San Mateo, CA: Morgan Kaufmann, 1987. 42~60

(编辑: 李维东)

A review on the theory for the genetic algorithm

Chen Jian-an

(School of Electronic Engineering, Xidian Univ., Xi'an, 710071)

Guo Dawei Xu Naiping Sun Yunzhi

(Dept. of Automatic Control, Northwestern Polytechnical Univ., Xi'an, 710072)

Abstract By describing the genetic algorithm, we introduce the development of the following researches in the genetic algorithm: selection for parameter sets, encoding methods, genetic operators, schema theorems, convergence analysis, deception problems, parallel computation, local improvement and hybrid methods. Although there exist some problems in the application of the genetic algorithm, these researches will stimulate the development of theories and applications of the genetic algorithm. Finally, a research trend is summarized briefly.

Key Words genetic algorithm theory development research review