

遗传算法及其研究与应用新进展

Genetic Algorithm and Its New Progress in Research and Application

施光林

(西安交通大学机械工程学院, 博士研究生 西安 710049)

史维祥

(西安交通大学机械工程学院, 教授 西安 710049)

近些年, 一种新的优化算法——遗传算法(Genetic Algorithm, 即 GA)正在迅速发展。这一算法最早是由美国的密西根(Michigan)大学的霍兰(Holland)教授和他的学生们在 70 年代初提出而创立的。该算法植根于自然进化与遗传机理, 原先是用于模拟自然界的自适应(适者生存)现象, 后来被引向于广泛的工程问题, 而快速发展成一种“自适应启发式概率性迭代式全局搜索算法”。由于遗传算法具有不依赖于问题模型的特性、全局最优性、隐含并行性、高效率及解决不同非线性问题的鲁棒性的特点, 正越来越激发起人们研究与应用的兴趣。

自 80 年代以来, 对 GA 的研究与应用进入了一个新高潮, 每年在国际上都有大量关于这方面的论文及研究成果问世。目前, 它已被广泛用于自动控制、机器人学、计算机科学、模式识别、模糊与人工神经网络和工程优化设计等领域。可以说对遗传算法的研究与应用现今已成为国际学术界关注的一大热点之一。

一、遗传算法的机理与实现步骤

1. 机理

遗传算法的基本思想来源于达尔文(Darwin)的进化论和门德尔(Mendel)的遗传学说。达尔文的进化论认为: 每一物种在不断的发展过程中都是越来越适应环境; 物种的每个个体的基本特征被其后代所继承, 但后代又不完全同于自己的父代; 在个体的生存与发展中那些更能适应环境的个体特征则能被保留下来, 体现了适者生存的原理。与此相应, 门德尔的遗传学说则认为: 遗传是作为一种指令遗传码封装在每个细胞中, 并以基因的形式包含在染色体中; 每个基因有特殊的位置并控制某个特殊的性质, 而由每个基因产生的个体对环境有一定的适应性; 通过基因杂交和基因突变可能产生对环境适应性强的后代, 并通过优胜劣汰的自然选择,

适应值高的基因结构则被保存下来。

霍兰等人正是综合了上述两种学说的基本思想创立了遗传算法。该算法借助于计算机编程, 一般是将待求问题表示成串(或称染色体), 即为二进制码或整数码串, 从而构成一群串, 并将它们置于问题的求解环境中, 根据适者生存的原则, 从中选择出适应环境的串进行复制(reproduction), 且通过交换(crossover)、变异(mutation)两种基因操作产生出新的一代更适应环境的串群, 经这样一代代地不断变化, 最后收敛到一个最适应环境的串上, 即求得问题的最优解。

2. 实现步骤

GA 的求解分为下列六步:

(1) 选择编码方式和编码

GA 在执行求解之前, 首先要选择合适的编码方式, 将待求解问题的所有参量编码成一个定长的字符串, 设串长为 l 。目前, 常使用的编码方式主要是二进制码。

(2) 产生初始群体

即由计算机随机产生 N 个初始字符串, 每个字符串称为一个个体, 这 N 个个体就构成了一个群体。GA 正是以这 N 个字符串作为初始点开始迭代计算。

(3) 计算适应度函数值

适应度函数反映了个体对环境适应能力的强弱, 根据对它的计算值, 可以很好地控制个体生存的机会, 以体现适者生存的自然法则。一般来说, 对于不同的问题, 适应度函数的定义方式也不尽相同。

(4) 复制

一个群体由 N 个个体组成, 这其中哪个个体保留用以繁殖后代, 哪个被淘汰, 是根据对它们各自的适应度函数值的计算来决定的。个体的适应度函数值大, 说明其对环境的适应能力强, 它也就有更多的机会被保留下来用以复制后代; 反之, 则其被

淘汰。可见,复制依据的原则是适应度大的个体为下一代贡献一个或多个后代的概率大。

(5) 交换

对于依据复制的原则而被选中、用于繁殖的两个个体,随机选择位置 $i (1 \leq i \leq l)$, 交换两个字符串位置 i 左边(或右边)的部分,产生两个新个体,这两个新个体分别组合了其父辈个体的特性。显然,交换的目的在于产生新的基因组合,形成新的个体,而不是代复一代地重复同一个群体。图 1 是按交换位置 i 右边部分的交换操作示意图。

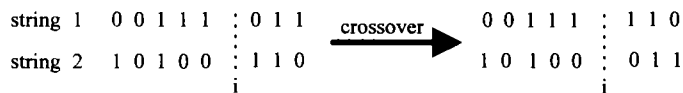


图 1 交换操作示意图

(6) 变异

即首先在群体中随机地选择一个个体,对于选中的个体以一定的概率随机地改变字符串中某位上的字符的值。当采用二进制码串时就是将所选中位上的 1 变为 0 或 0 变为 1。变异操作如图 2 所示。

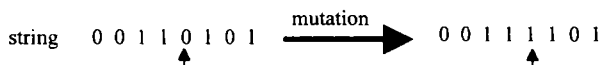


图 2 变异操作示意图

值得一提的是,单靠变异一般不能使问题求解取得进展,其发生的概率很低,但它能确保不产生不能继续进化的单一群体,且为新的个体的产生提供了机会。

遗传算法正是经历了上述 6 个步骤并执行 6 至 3 步的循环操作,才求出问题的最优解。

二、遗传算法的特点和存在的问题

1. 遗传算法的特点

通过对 GA 的机理、实现过程的分析与介绍,人们不难看出由霍兰等人提出的这一基于生物进化和遗传的优化算法与传统的优化算法相比,有着许多鲜明的特点,这些特点是:

(1) 优化过程针对由参数编码所形成的串,而不是参数本身

即 GA 要将所求问题的参数依据它们各自的约束条件,按一定的编码方式将它们形成一个串,并对串群进行操作,从中找出高适应度值的串,而不是对问题或它们的参数直接操作。因此,用传统方法很难解决的问题,一般用 GA 都能解决。这是 GA 不受问题性质(如连续、可导、凸性等)限制的结

果。

(2) 优化搜索始于问题的所有可能解,而不是从单个解开始

也就是说 GA 同时从所有可能解组成的串集中的每个串开始搜索,犹如像一张网罩在所有可能的搜索方向上,使数量极大的串同时在很多区域中被进行搜索,大大减少了陷入局部解的可能性。因而,可以避免传统优化方法是从一个点开始搜索,当遇到复杂问题所形成的“多峰”搜索路径时往往会陷入某个“谷底”,而不能到达全局最优点的弱点。

(3) 求解所需问题的信息量少

这是因为 GA 是利用所求问题的适应度函数值这唯一信息来进行搜索求解的,而不像传统优化方法需要一些辅助信息(如导数等),一旦所需信息不存在,这些方法就会失败而无法被执行。从这个意义上说,遗传算法几乎可以处理任何问题。

(4) 具有隐含的并行性

所谓遗传算法的隐含并行性即是指在搜索空间里使用相对少的串,就可以检验表示数量极大的区域的特性。这可以解释为,当计算机只执行了 N 个串的操作,但却好像是在没有占用多于 N 个串的内存的情况下,并行地实现了对 $O(N^3)$ 个串的处理。这种隐含的并行性是遗传算法优于其它算法的关键所在。

(5) 所进行的操作简单和可靠

因为 GA 仅使用复制、交换和变异这三种手段进行随机操作,而非确定性规则,所以实际操作简单、方便。遗传算法的这一随机操作并不意味着它是一种简单的随机搜索,但它又正是使用这些简单的随机操作来指导搜索向着问题的一个最优解前进。

(6) 适于求解复杂问题而且效率高

正是由于 GA 具有上述的特点,所以它最善于搜索求解复杂问题,并可以很快地从这一问题所形成的复杂地域中找出期望值高的区域,最终快速找出复杂问题的最优解。值得注意的倒是遗传算法在解决简单问题时显现的效率并不高。

2. 存在的问题

一般把霍兰等人提出的遗传算法称之为简单的遗传算法。尽管这一算法已具有较多的优点,业已在实际中得到了大量应用,但它也存在着许多急待解决的问题,归纳起来主要有:

- 如何进行算法本身的参数优化选择;
- 如何避免算法过早收敛的产生;
- 如何改进操作手段或引入新操作手段来提高算法的执行效果;
- 如何将该算法与其它传统优化方法有机结合起来;

——如何从理论上更加完善地对该算法进行定量分析;

——如何更广泛地开辟该算法及其改进算法的实际应用新领域。

上述问题中,参数优化选择是指对群体的规模 N 、交换概率 P_c 和变异概率 P_m 进行优化选择。因为实践发现这些参数的选取直接关系着GA求解问题的成败。过早收敛是指GA在执行过程中会出现群体中的个体过早地在一个非最优点上达到完全相同或接近完全相同的现象。一旦出现该现象,利用GA就不能求得问题的全域最优解。以上存在的问题有的已基本获得解决,而有的则正在解决当中。

三、遗传算法的研究与应用新进展

自从70年代初遗传算法创立以来,先是经过了一段时间的沉寂,接着在80年代初,随着柯克帕特里克(Kirkpatrick)等人提出模拟退火算法(Simulated Annealing Algorithm)后激起了人们研究与应用遗传算法的热潮。这一时期,遗传算法本身所存在的许多理论问题和实际应用问题开始得到了人们的关注。应该说在其研究与应用上都取得了丰硕的成果。然而,更为可喜的是,进入90年代以后,学术界对遗传算法的研究与应用的兴趣有增无减,显现了良好的势头。所取得的新进展主要包括对算法本身的理论及改进研究、实际应用两大方面。

1. 理论及改进研究方面

(1)理论上,成功地应用齐次有限马尔科夫链理论分析了简单遗传算法、最优保存简单遗传算法(OMSGA)和自适应遗传算法(AGA)的收敛性。从而得到简单遗传算法不是全域收敛,而OMSGA和AGA是全域收敛的重要结论。此外,有人利用马尔科夫链理论对浮点数编码的遗传算法进行了严密的全域收敛性分析;另有学者也研究了达到全局最优解的遗传算法的时间复杂性问题。这些理论问题的相继解决,为遗传算法获得更好的实际应用奠定了基础。

(2)针对简单遗传算法本身所存在的缺陷,有关该算法的改进研究工作也有许多。如有人就提出了多点交换操作、倒位操作来改进已有的操作模式;也有的学者指出应该引入某些特定信息来指导交换和变异操作,以使遗传算法更易于某些特定问题的求解。另外,编码方式更趋多样化,现已有了整数编码方式和浮点数编码方式。目前,对简单遗传算法改进的新算法主要有启发式遗传算法、递级分布式遗传算法、连续繁殖式遗传算法、基于分裂选

择的遗传算法、自适应遗传算法、并行遗传算法、基于迁移和人工选择的遗传算法。

(3)这一时期,尤为成功的是在遗传算法与其它方法的混合研究上。这一研究的目的在于既发挥遗传算法的全局性特点又发挥某类特定方法对于特定问题的有效收敛性之特长,且能快速、稳定地搜索到问题的全局最优解。现有的混合方法主要有并行遗传/神经网络混合学习方法、遗传/进化编程混合方法、遗传算法/遗传编程混合方法、“爬山”法/遗传算法混合方法、模拟退火/遗传算法混合方法、模糊/遗传算法混合方法。上述这些混合方法在实际中都已得到了较好的应用。

2. 应用方面

遗传算法的应用领域从函数优化、机器学习、模式识别、图像处理、模糊与神经网络一直到包括了工业控制和工程优化设计等在内的诸多方面。近些年,该算法在下列方面的应用有了新的进展。

(1)在函数优化方面,已从离散区域的优化迈向连续区域优化,并采用了动态参数编码技术,初步解决了连续区域优化在采用定长的二进制编码时所出现的精度不高、搜索空间有限和计算效率低的弊端。另外,多目标函数优化也是遗传算法应用的一大热点。

(2)在工业控制方面,遗传算法除继续应用工业过程、工业机器人等的自适应控制、最优控制外,更多的是用于对模糊控制、神经网络控制和模糊——神经网络控制的优化设计上。这里主要是利用遗传算法来优化模糊规则数、隶属函数形状以及神经网络的各连接权值、学习参数和网络的拓扑结构。现在,国际上每年都有大量关于这方面应用的成果面世。

(3)在工程优化设计方面,已开始在大工程结构件的拓扑结构优化设计和传统机械装置如减速箱的优化设计中应用了遗传算法。实践证明,采用遗传算法的优化结果要普遍优于用传统优化方法所得到的结果,并且求解简便,适应求解问题的范围宽。

此外,在金属材料 and 食品成分配方优化、图像处理、故障诊断、社会科学和人工生命等方面,遗传算法也已经或正在开始发挥其巨大的潜力。

可以相信,随着对遗传算法所存在问题的逐步解决以及计算机在硬件技术上的发展和突破,只要人们勇于实践,遗传算法及其改进算法会在更多的领域得到进一步的应用。

(本文引用了不少有关内容和成果,在此不一一列出,谨向这些中外作者致谢——笔者)

(责任编辑 王宏章)