

# 遗传算法原理与发展方向综述

赵宜鹏 孟磊 彭承靖

(云南民族大学数计学院,云南 昆明 650031)

**摘要:**遗传算法是一种基于生物自然选择与遗传机理的随机搜索与优化方法。近年来,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,这种算法受到了国内外学者的广泛关注。介绍了遗传算法的研究现状和基本原理,概述了它的理论和技术,并对遗传算法的性能作了分析。

**关键词:**遗传算法;遗传算子;进化计算;编码

## 1 概述

目前,遗传算法正在迅速发展。遗传算法以其很强的解决问题能力和广泛的适应性渗透到研究与工程的各个领域,取得了良好的效果。近年来,在多种问题的解决方法上,遗传算法都得到广泛应用。

## 2 遗传算法的基本原理

遗传算法是建立在自然选择和群众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。他最先由 John Holland 于 1975 年提出。从此以后,它逐渐发展成为一种通过模拟自然进化过程解决最优化问题的计算模型。

遗传算法的研究主要包括三个领域:遗传算法的理论与技术;用遗传算法进行优化;用遗传算法进行分类系统的机器学习。其中,遗传算法的理论和技术研究主要包括编码、交叉运算、变异运算、选择运算以及适应度评价等问题。

### 2.1 基本原理

在自然界,由于组成生物群体中各个体之间的差异,对所处环境有不同的适应和生存能力,遵照自然界生物进化的基本原则,适者生存,优胜劣汰,将要淘汰哪些最差个体,通过交配将父本优秀的染色体和基因遗传给子代,通过染色体核基因的重新组合产生生命力更强的新的个体与他们组成的新群体。在特定的条件下,基因会发生突变,产生新基因和生命力更强的新个体,但突变是非遗传的。随着个体不断更新,群体不断朝着最优方向进化,遗传算法是真实模拟自然界生物进化机制进行寻优的。

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。遗传算法主要通过交叉、变异、选择运算实现。交叉或便宜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体的优劣计算中有可能达到最优解的优良程度。度量个体适应度的函数称为适应度函数。适应度函数的定义一般与具体问题有关。

### 2.2 典型的算法步骤

#### 2.2.1 初始化种群

#### 2.2.2 计算种群上每个个体的适应度;

2.2.3 按由个体适应度值所决定的某个规则选择将进入下一代的个体;

#### 2.2.4 按概率 $P_c$ 进行交叉操作;

#### 2.2.5 按概率 $P_m$ 进行变异操作;

2.2.6 若没有满足某种停止条件,则转入(2.2.2),否则进入下一步;

2.2.7 输出种群中适应度最优的染色体作为问题的满意解或最优解。

### 2.3 编码问题

编码是遗传算法要解决的首要问题。常用的编码方法有二进制编码、格雷码编码、实数编码、符号编码等。针对不同的问题要采用不同的编码方法。

### 2.4 群体设定

遗传操作是对众多个体同时进行的,这众多个体组成了群体。在遗传算法处理流程中,继续编码设计后的任务是初始群体的设定,并以次为起点一代代进化知道按某种进化停止准则终止进化过程,由此得到最后一代。关键问题是,群体规模,即群体中包括的个体数目如何确定。其中有两个需要考虑的因素:a. 初始群体的设定,首先根据问题固有知识,设法把握最优解所占空间在整个问题空间的分布范围,然后在此分布范围内设定初始群体。然后先随机生成一定数目的个体,然后从中挑出最好的个体加到初始群体中。这种过程不断迭代,直到初始群体中个体数达到了预先确定的规模。b. 进化过程周各代的规模如何维持。群体规模的确定受遗传操作中选择操作的影响很大。一般而言,群体规模越大,群体中个体的多样性越高,算法陷入局部解的危险就越小。所以,从考虑群体多样性出发,群体规模应较大。但是,群体规模太大会带来若干弊病。一是计算效率,群体越大,其适应度评估次数增加,所以计算量也增加,从而影响算法效能;二是群体中个体生存下来概率大多采用和适应值成比例的方法,当群体中个体非常多时,少量适应值很高的个体会被选择而生存下来,但大多数个体却被淘汰,这会影响到库的形成,从而影响到交叉操作。另一方面,群体规模太小,会使遗传算法的搜索空间中分布范围有限,因而搜索有可能停止在未成熟阶段,引起未成熟收敛现象。

适应度函数:遗传算法在进化搜索中基本上不用外部信息,仅用适应度函数作为依据。遗传算法的目标函数不受连续可微的约束且定义域可以为任意集合。对目标函数的唯一要求是针对输入值,可计算出能加以比较的非负结果。这一特点使得遗传算法应用范围很广。

### 2.5 遗传操作

遗传操作包括 3 个基本遗传算子,交叉、变异、选择。这 3 个算子有以下特点:a. 这 3 个遗传算子的操作都是随机化操作,因此,个体向最优解的迁移也是随机的。b. 遗传操作的效果和上述 3 个遗传算子所取的操作概率、编码方法、群

体大小,初始群体以及适应度函数的设定密切相关。c. 3 个基本遗传算子的操作方法或操作策略和个体的编码方式直接有关。

### 2.5.1 交叉运算

所谓交叉运算,是指两个相互配对的染色体按某种方式相互交换其部分基因,从而形成两个新的个体。交叉运算是遗传算法区别于其他进化算法的重要特征,它决定了遗传算法的全局搜索能力,是产生新个体的主要方法。常用的交叉算子有 a. 单点交叉。又称简单交叉,它是指在个体编码串中随机设置一个交叉点,然后在该点相互交换两个配对个体的部分基因。b. 双点交叉。它的具体操作过程是 1 在相互配对的两个个体中编码串中设置两个交叉点,2 交换两个交叉点之间的部分基因。c. 均匀交叉。它是指两个配对个体的每一位基因都以相同的概率进行交换,从而形成两个新个体。具体操作过程是 1 随机产生一个与个体编码长度相同的二进制屏蔽字  $w=w_1w_2\cdots w_x$ 。2 按下列规则从 A、B 两个父代个体中产生两个新个体 X、Y;若  $w_i=0$ ,则 X 的第 i 个基因继承 A 的对应基因, Y 的第 i 个基因继承 B 的对应基因;若  $w_i=1$ ,则 A、B 的第 i 个基因相互交换,从而生成 X、Y 的第 i 个基因。

### 2.5.2 变异运算

所谓变异运算,是指将个体编码串中的某些基因值用其他基因值来替换,从而形成一个新的个体,遗传算法中的变异运算是产生新个体的辅助方法,但必不可少,因为它决定了遗传算法的局部搜索能力。常用的方法有 a. 基本位变异。它是指对个体编码串以变异概率 p 随机制定某一位或某几位基因作变异运算。b. 均匀变异。它是指分别用符合某一范围内均匀分布的随机数,以某一较小的概率来替换个体中的每个基因。c. 二元变异。它的操作需要两条染色体参与,两条染色体通过二元变异操作后生成两条新个体,新个体中的各个基因分别取原染色体对应基因值的同或/异或。

### 2.5.3 选择运算

选择运算就是对群体中的个体进行优胜劣汰操作:适应度高的个体被遗传到下一代群体中的概率大,适应度低的个体,被遗传到下一代群体中的概率小。它的任务就是按某种方法从父代群体中选取一些个体,遗传到下一代群体。

## 3 遗传算法的发展过程与现状

遗传算法研究的兴起是在 20 世纪 80 年代末和 90 年代初,但它的历史起源可追溯到 20 世纪 60 年代初期。早期的研究大多以对自然系统的计算机模拟为主。如 Fraser 的模拟研究,他提出了和现在的遗传算法十分相似的概念和思想。Holland 和 DeJong 的创造性研究成

果改变了早期遗传算法研究的无目标性和理论指导的缺乏。其中, Holland 于 1975 年出版的著名著作《自然系统和人工系统的适配》系统地阐述了遗传算法的基本理论和方法, 并提出了对遗传算法的理论研究和发极为重要的模式理论。这一理论首次确认了结构重组遗传操作对于获得隐并行性的重要性。同年, DeJong 的重要论文《遗传自适应系统到的行为分析》将 Holland 的模式理论与他的计算实验结合起来, 并提出了诸如代沟等新的遗传操作技术。可以认为, DeJong 所作的研究工作是遗传算法发展过程中的一个里程碑。进入 20 世纪 80 年代, 遗传算法迎来了兴盛发展时期, 无论是理论研究还是应用研究都成了十分热门的课题。尤其是遗传算法的应用领域也不断扩大。目前遗传算法所涉及的主要领域有自动控制、规划设计、组合优化、图象处理、信号处理、人工生命等。可见, 遗传算法的应用研究已从初期的组合优化求解拓展到了许多更新、更工程化的应用方面。

#### 4 遗传算法的应用

遗传算法提供了一种求解复杂系统优化问题的通用框架, 它不依赖于问题的具体领域, 对问题的种类有很强的鲁棒性, 所以广泛应用于很多学科。目前主要应用的领域有函数优化、生产调度问题、自动控制、机器学习、图像处理、人工生命、遗传编程、机器学习、数据挖掘等等。

#### 5 遗传算法的发展方向

##### 5.1 遗传算法自身的优化

自从遗传算法在各种问题中得到广泛的应用以来, 遗传算法的优化问题就成了人们研究的焦点。专家学者们从各个方面, 在各种细节上采用各种方法试图来改进遗传算法。从编码方法、控制种群、控制交叉、控制变异等来改进遗传算法。根据不同问题的需要来选择编码方案, 是改进遗传算法最初的手段。随着编码方案的不断完善, 现在从编码方案上来改进遗传算法的意义已经很小。目前对遗传算法的优化主要有两大手段。一是利用对种群的控制, 在选取种群的时候, 在种群的规模、种群的多样性上下功夫。有的是利用加入种间竞争的手段。另一种是通过控制交叉方法和变异的概率, 根据问题的实际情况来设定一个线性或非线性的函数来控制变异的概率, 使得遗传算法在执行的时候能够在加快收敛效率的时候同时保持个体的多样性, 有利于找出全局最优解。这两种优化算法在不同问题的应用中都成功的优化了遗传算法的效率。

##### 5.2 遗传算法与其他计算智能方法的相互渗透和结合

遗传算法是一种通用而有效的求解最优化问题的方法, 然而, 单用简单的遗传算法在许多情况下不是十分有效, 容易产生早熟现象以及局部寻优能力较差等问题, 于是提出了多种混合算法。遗传算法在日益和神经网络、模糊推理以及混沌理论等其他智能计算方法相互渗透和结合, 必能达到取长补短的作用。近年来在这方面已取得了不少研究成果, 并形成了“计算智能”的研究领域, 这对开拓 21 世纪新的智能计算技术具有重要的意义。例如 Ackley 推荐的遗传爬山法, Mathefoud 提出的遗传模拟退火算法, Miller 等提出的对于 NP 难问题的优化问

题, 采用遗传算法中增加局部改善运算等等。混合遗传算法的基本思想是: 对于每个新产生的后代在其进入下一代群体之前应用局部优化技术, 使之移动到最近的局部最优值。在混合遗传操作中, 运用启发式方法作局部优化, 采用遗传算法作全局最优值的探索。由于遗传算法与传统优化方法的互补性, 混合遗传算法通常比单一算法优越。

#### 5.3 并行处理遗传算法

并行处理的遗传算法的研究不仅是遗传算法本身的发展, 而且对于新一代智能计算机体系结构的研究都是十分重要的, 遗传算法在操作上具有高度的并行性, 许多研究人员都在搜索在并行机上搞紧执行遗传算法的策略。并行遗传算法分为两点, 一粗粒度并行遗传算法, 它主要开发群体间的并行性; 另一类是细粒度并行遗传算法, 它主要开发一个群体中的并行性。

遗传算法在解决一些实际问题时, 由于它一般具有较大的群体规模, 需要对较多的个体进行大量的遗传和进化操作, 特别是要对大量的个体进行适应度计算或评价, 从而使得算法的进化运算过程进展缓慢, 难以达到计算速度的要求, 因而遗传算法的并行计算问题受到重视。人们认识到对遗传算法进行并行处理的可能性, 于是提出了多种基于各种并行计算机或局域网的并行遗传算法。这些并行遗传算法主要从下列四个方面对其进行改进和发展。

##### a. 个体适应度评价的并行性

个体适应度的评价或计算在遗传算法的运行过程中所占的运行时间比较长。通过对个体适应度并行计算方法的研究可找到并行评价个体适应度的算法。

##### b. 整个群体中各个个体的适应度评价的并行性

群体中各个个体适应度之间无相互依赖关系, 这样各个个体的适应度计算过程就可以相互独立、并行地进行。即不同个体的适应度计算可以再以不同的处理机上同时进行。

##### c. 群体产生过程的并行性

在父代群体产生下一代群体过程中, 选择只与个体的适应度有关, 而交叉和变异操作只与参加运算的个体编码有关。这样, 产生群体过程中的选择、交叉、变异操作就可以相互独立的并行进行。

##### d. 基于群体分组的并行性

可以对群体按一定得方式进行分组, 分组后各组的个体遗传进化过程可以再以不同的处理机上相互独立的进行, 在适当的时候, 各处理机之间相互交换信息。

遗传算法用于优化的研究成果目前已经发表了较多的论文。这些成果有的是对此算法进行改进, 如李通化等给出了一种动态的遗传算法, 详细讨论了变异函数对优化进程的影响, 并用红外光谱数据和二极管阵列检测液相色谱数据进行了验证。杨立民等运用双极性压缩适应度定标和基于排挤方法的选择算子改进标准遗传算法, 使其成为简单通用、快速收敛的并行全局搜索算法; 利用该算法优化误差反馈神经网络, 克服了 BPN 收敛慢和不具有全局收敛性的缺陷, 并在此基础上, 建立大气环境质量评价

模型。张昆等为了使金属切削加工中切削参数能实现实时优化以保证产品质量和设备效率, 提出采用遗传算法, 结合现场实际工况的反馈信息实现了实时优化, 在任一不同的生产条件下均能达到最优值。也有应用遗传算法去解决某一实际问题, 如宋仁国等将人工神经网络用于建立 7175 铝合金的性能预测模型, 在此基础上, 采用遗传算法对其工艺进行优化, 王举等将间歇化工过程的最优设计问题, 分解为只包含离散变量的主导问题和只含连续变量的子问题, 把遗传算法和现行规划法结合起来对其进行求解, 并在算法和线性规划法结合起来进行求解, 并在算法中引入了一些新的算子, 显著地提高了收敛效率。蔡煜东等将遗传算法用于化学校正, 运用改进的遗传算法拟合离子选择电极工作曲线, 并以半对数线性函数模型和二阶幂函数模型尝试了该算法的效果, 结果表明, 改进的遗传算法性能较好, 优于一般遗传算法和直线回归法。他还提出非线性多元函数拟合的遗传算法, 为分析化学中非线性函数拟合、曲线校正提供了一种性能较好的方法。

#### 6 结论

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解决方法, 并且经过实践证明效果显著。遗传算法是一种全局最优化方法, 在优化过程中, 它无需体系的先验只是, 能在许多局部较优中找到全局最优值, 能有效的处理复杂的非线性问题。遗传算法的理论研究需要进一步深入, 应用领域有待进一步开拓。但可以相信, 遗传算法必将在以后得到更为广泛的应用。

#### 参考文献

- [1] 吴家英, 郑金华. 遗传算法的研究与发展动向[J]. 濮阳师范学院学报, 2003.
- [2] 李华昌, 谢淑兰, 易忠胜. 遗传算法的原理与应用[J]. 矿业, 2005, 3.
- [3] 张玲, 张钊. 遗传算法机理的研究[J]. 软件学报, 2000, 11.
- [4] 吉根林. 遗传算法研究综述[J]. 计算机应用与软件, 2004, 2.
- [5] 王春水, 肖学柱, 陈汉明. 遗传算法的应用举例[J]. 计算机仿真, 2005, 6.
- [6] 陈根社, 陈新海. 遗传算法的研究与发展[J]. 信息与控制, 1994, 23.
- [7] 张丽萍, 柴跃进. 遗传算法的现状与发展动向[J]. 信息与控制, 2001, 30.
- [8] 丁承民, 张传生, 刘辉. 遗传算法纵横谈[J]. 信息与控制, 1997, 26.

责任编辑: 鲁艳