

生物智能与算法课程报告

遗传算法

陈琛 3140104329

一、绪论

遗传算法（Genetic Algorithms，简称GA）是由美国密歇根大学的John Holland教授于1975年在他的专著《自然界和人工系统的适应性》中首先提出的。它模拟达尔文生物进化论的自然选择和孟德尔、摩根遗传学机理的生物进化过程来建立计算模型，通过对自然进化过程的模拟来搜索最优解。

遗传算法主要应用于各种优化问题，通过有限的代价来解决搜索和优化问题。区别于传统优化算法，遗传算法具有随机性和非线性。通过仿效自然选择下的生物进化，遗传算法形成了随机化的搜索方法。而相比于传统的启发式优化搜索算法，遗传算法的本质特征在于群体搜索策略和简单的遗传算子，这使得遗传算法可以实现整个解空间上的信息采集和搜索，而且因为作为度量标准的适应度函数的存在，遗传算法可以很大程度上的降低人机交互的依赖。

遗传算法从随机形成的问题的解集的一个种群开始，每个个体都通过固定格式的一组编码表示。从初代种群开始，按照适者生存和优胜劣汰的原理，

逐代演化出越来越好的近似解。在每一代中，对每个个体的适应度进行评估并进行淘汰和复制，之后利用遗传算子对个体之间进行交叉组合和变异，产生下一代种群。经历预先确定代数的演化后，得到的种群被视为最优解。

二、基本原理及生物学背景

1、生物学背景

达尔文进化论认为，在一个种群中，个体的特征由父代继承而来同时产生一定几率的变异。大多数变异是有害的（即不适应环境的），少部分变异是有利于个体生存的。自然选择会淘汰不适应环境的个体从而达到基因筛选的目的。

孟德尔遗传学认为每个个体的基因型在一定程度上决定外在的表现型。种群之内存在以繁殖、交配、基因突变及变异形式的基因流动。

2、GA术语解释及生物学对应

- 染色体：种群中一个个体的基因编码组合，通常是一个二进制串或浮点数串
- 基因：染色体中的最基本元素，用于表示个体的特征。每个元素的值被称为等位基因。
- 基因位点：某个基因在串中的位置。
- 适应度：个体对环境的适应程度。GA中引入适应度函数来计算某个解的优化程度。

3、GA的一般算法流程

- 确定编码方式——使用二进制编码（或浮点数串编码）对搜索空间（解空间）进行编码
- 确定停止条件——迭代一定代数后停止或者达到一定的效果（如种群的稳定）后停止
- 产生初代种群——随机产生包含n个个体的初始种群

- 判断是否达到停止条件。若达到停止条件则停止迭代，否则向下进行
- 利用适应度函数评估每个个体
- 选择要进行繁殖的若干个个体，复制这些个体使得种群数量不变
- 对要进行繁殖的种群进行随机配对，按一定概率进行交叉操作形成下一代个体
- 按照一定概率对某些个体进行变异
- 回到第四步

4、运算过程具体介绍

(1) 编码方式

编码是将解空间投射到GA空间的过程，De Jong提出过两条关于GA的编码原理的规则：

- 有意义积木块编码规则:所定编码应当易于生成与所求问题相关的短距和低阶的积木块。
- 最小字符集编码规则:所定编码应采用最小字符集以使问题得到自然的表示或描述。

目前主要采用的编码方式一般为二进制编码以及浮点数编码。前者简单易操作，后者则更加精确，更利于大空间搜索。

(2) 适应度函数

适应度函数是用于评价个体适应环境的能力，是选择操作的依据，直接影响遗传算法的性能。一般来说，适应度函数是目标函数或者由目标函数变换而成。对适应度函数唯一的要求是：结果为非负值。这使得GA可以应用于不连续、不能求导的情况，提高了GA的适用性。

(3) 选择算子

选择算子作用于适应度函数的基础上，用于确定交叉个体以及被选个体将产生多少个复制个体。常见的方法包括：轮盘赌选择、最佳个体保存、排序选择、联赛选择方法。

最常用的方法是轮盘赌选择方式。在该方法中，各个个体的选择概率和其适应度值成比例。显然，概率反映了个体 i 的适应度在整个群体的个体适应度总中所占的比例。个体适应度越大，其被选择的概率就越高、反之亦然。计算出群体中各个个体的选择概率后，为了选择交配个体，需要进行多轮选择。每一轮产生一个 $[0, 1]$ 之间均匀随机数，将该随机数作为选择指针来确定被选个体。个体被选后，可随机地组成交配对，以供后面的交叉操作。

(4) 交叉算子

交叉是指将两个父代个体的部分结构加以替换重组而生成新的个体。交叉操作的作用是组合出新的个体，在解空间进行有效搜索，同时降低对已有有效模式的破坏。因此对交叉算子的要求是：保证前一代中优秀个体的性状能够得到充分的遗传和继承。常用的交叉算子包括：一点交叉、两点交叉、一致交叉等。

(5) 变异算子

变异是指对已形成的种群的部分个体的某些基因位上的基因值进行变动。变异的作用是增加种群的多样性，在一定几率下可以使得当前解跳出局部最优解。

三、遗传算法的优缺点

1、优点

- 直接对结构对象进行操作，不存在求导和函数连续性的限定；
- 具有内在的隐并行性和更好的全局寻优能；

- 采用概率化的寻优法，能动获取和指导优化的搜索空间，自适应地调整搜索方向，不需要确定的规则。

2、缺点

- 收敛慢
- 有可能陷入局部最优

四、相关理论基础

有关遗传算法的分析，主要以收敛性分析为主。从整体上讲，可以分为基于随机过程和模式理论两种收敛性分析

1、基于随机过程的收敛性分析

遗传算法的搜索过程可以表示为马尔可夫链模型，从而利用已有的随机过程理论进行分析：

- 1Rudolph“用奇次有限马尔科夫链证明了标准遗传算法收敛不到全局最优解；
- 2Eiben…等用马尔科夫链证明了保留最优个体的GA的概率性全局收敛；
- 3Qi和Palmieril用马尔科夫链证明可浮点编码的遗传算法是收敛的，但此结果是基于群体数 为无穷大这一假设。

2、基于模式理论的收敛性分析

五、改进方法

1、自适应遗传算法

鉴于遗传算法易出现“早熟”或陷入局部最优的现象。在搜索过程中适时调整其使用的各个参数是非常有意义的。

Srinivas.M and Patnaik.L.M (1994)提出来自适应遗传算法的。在论文中， p_c 和 p_m 的计算公式如下：

六、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，主要有以下应用领域：控制方面有瓦斯管道控制，防避导弹控制，机器人控制等；规划方面有生产规划，并行机任务分配等；设计方面有VLSI布局，通信网络设计，喷气发动机设计等；组合优化方面有TSP问题，背包问题，图划分问题等；图像处理方面有模式识别，特征抽取等；信号处理方面有滤波器设计等；机器人方面有路径规划等；人工生命方面有生命的遗传进化等；知识发现方面有规则提取，数据挖掘等。

参考文献：

- [1]马永杰,云文霞. 遗传算法研究进展[J]. 计算机应用研究,2012,(04):1201-1206+1210.
- [2]李欣. 自适应遗传算法的改进与研究[D].南京信息工程大学,2008.
- [3]刘立平,牛熠. 遗传算法综述[J]. 东莞理工学院学报,2005,(03):48-52.
- [4]马玉明,贺爱玲,李爱民. 遗传算法的理论研究综述[J]. 山东轻工业学院学报(自然科学版),2004,(03):77-80.
- [5]周芳,邓乐斌. 免疫算法与遗传算法的比较[J]. 鄱阳师范高等专科学校学报,2004,(03):8-10.
- [6]A genetic algorithm tutorial.