

遗传算法

姓名：马春雨 学号：21721195

May 9, 2018

目录

1	遗传算法简介	2
1.1	遗传算法的产生与发展	2
1.2	生物进化理论和遗传学的基本知识	2
1.3	遗传算法的思路与特点	2
1.4	遗传算法的基本操作	3
1.5	遗传算法的应用	3
2	基本遗传算法	3
2.1	模式定理	3
2.2	基本遗传算法	4
2.3	简单函数优化的实例	4
3	遗传算法的具体应用	6
3.1	解决带约束的函数优化问题	6
3.2	解决多目标优化问题	6
3.3	解决组合优化问题	6
3.4	遗传算法在过程建模中的应用	6
3.5	遗传算法在模式识别中的应用	7
4	遗传算法的改进	7

1 遗传算法简介

1.1 遗传算法的产生与发展

早在50年代，一些生物学家开始研究运用数字计算机模拟生物的自然遗传与自然进化过程；1963年，德国柏林技术大学的I. Rechenberg和H. P. Schwefel，做风洞实验时，产生了进化策略的初步思想[1]；60年代，L. J. Fogel在设计有限态自动机时提出进化规划的思想[2]。1966年Fogel等出版了《基于模拟进化的人工智能》，系统阐述了进化规划的思想；60年代中期，美国Michigan大学的J. H. Holland教授提出借鉴生物自然遗传的基本原理用于自然和人工系统的自适应行为研究和串编码技术；1967年，他的学生J. D. Bagley在博士论文中首次提出“遗传算法（Genetic Algorithms）”一词；1975年，Holland出版了著名的“Adaptation in Natural and Artificial Systems”，标志遗传算法的诞生[3]。70年代初，Holland提出了“模式定理”（Schema Theorem），一般认为是“遗传算法的基本定理”，从而奠定了遗传算法研究的理论基础；1985年，在美国召开了第一届遗传算法国际会议，并且成立了国际遗传算法学会（ISGA, International Society of Genetic Algorithms）；发展1989年，Holland的学生D. J. Goldberg出版了“Genetic Algorithms in Search, Optimization, and Machine Learning”，对遗传算法及其应用作了全面而系统的论述；1991年，L. Davis编辑出版了《遗传算法手册》，其中包括了遗传算法在工程技术和社会生活中大量的应用实例。

1.2 生物进化理论和遗传学的基本知识

达尔文的自然选择说中遗传（heredity）是指子代和父代具有相同或相似的性状，保证物种的稳定性；变异（variation）指子代与父代，子代不同个体之间总有差异，是生命多样性的根源；生存斗争和适者生存是指具有适应性变异的个体被保留，不具适应性变异的个体被淘汰。自然选择过程是长期的、缓慢的、连续的过程。在遗传学中染色体是遗传物质的载体，其决定形状的外部表现即表现型，脱氧核糖核酸是大分子有机聚合物，遗传因子是DNA或RNA 长链结构中占有一定位置的基本遗传单位，遗传因子组合成基因型。染色体带有特征的实体构成个体，个体的集合构成种群，个体的数目称为种群的大小。

进化（evolution）指生物在其延续生存的过程中，逐渐适应其生存环境，使得其品质不断得到改良，这种生命现象称为进化；适应度（fitness）指度量某个物种对于生存环境的适应程度，对生存环境适应程度较高的物种将获得更多的繁殖机会，而对生存环境适应程度较低的物种，其繁殖机会就会相对较少，甚至逐渐灭绝；选择（selection）指决定以一定的概率从种群中选择若干个体的操作；复制（reproduction）指细胞在分裂时，遗传物质DNA通过复制而转移到新产生的细胞中，新的细胞就继承了旧细胞的基因；交叉（crossover）指在两个染色体的某一相同位置处DNA被切断，其前后两串分别交叉组合形成两个新的染色体，又称基因重组，俗称“杂交”；变异（mutation）指在细胞进行复制时可能以很小的概率产生某些复制差错，从而使DNA发生某种变异，产生出新的染色体，这些新的染色体表现出新的性状；编码（coding）是表现型到基因型的映射，而解码（decoding）是从基因型到表现型的映射。

1930至1947年，达尔文进化论与遗传学走向融合，Th. Dobzhansky 1937年发表的《遗传学与物种起源》是融合进化论与遗传学的代表作。生物物种作为复杂系统，具有奇妙的自适应、自组织和自优化能力，这是一种生物在进化过程中体现的智能，也是人工系统梦寐以求的功能。

1.3 遗传算法的思路与特点

遗传算法特点包含自组织、自适应和自学习性，在编码方案、适应度函数及遗传算子确定后，算法将利用进化过程中获得的信息自行组织搜索；内在并行性与内含并行性决定了本质并行性；不需求导，只需目标函数和适应度函数；强调概率转换规则，而不是确定的转换规则。

1.4 遗传算法的基本操作

遗传算法基本操作包含适应度计算、根据适应度进行选择、交叉或基因重组、变异，重复执行基本操作直至满足终止条件。遗传算法流程如图1所示

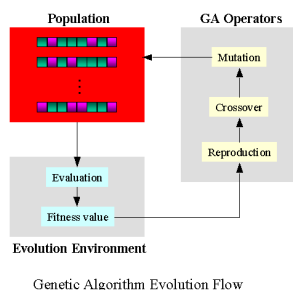


图 1: 遗传算法流程

1.5 遗传算法的应用

遗传算法可应用于以下领域: 函数优化, 是遗传算法的经典应用领域; 组合优化, 实践证明, 遗传算法对于组合优化中的NP完全问题非常有效; 自动控制, 如基于遗传算法的模糊控制器优化设计、基于遗传算法的参数辨识、利用遗传算法进行人工神经网络的结构优化设计和权值学习等; 机器人智能控制, 遗传算法已经在移动机器人路径规划、关节机器人运动轨迹规划、机器人逆运动学求解、细胞机器人的结构优化和行动协调等; 组合图像处理和模式识别, 目前已在图像恢复、图像边缘特征提取、几何形状识别等方面得到了应用; 人工生命, 基于遗传算法的进化模型是研究人工生命现象的重要理论基础, 遗传算法已在其进化模型、学习模型、行为模型等方面显示了初步的应用能力; 遗传程序设计, Koza发展了遗传程序设计的概念, 他使用了以LISP语言所表示的编码方法, 基于对一种树型结构所进行的遗传操作自动生成计算机程序。

2 基本遗传算法

2.1 模式定理

将种群中的个体即基因串中的相似样板称为模式。在二进制编码的串中, 模式是基于三个字符集(0, 1, *)的字符串, 符号*代表任意字符, 即0或1。如模式*1* 描述了一个四个元的子集010, 011, 110, 111。模式H 中确定位置的个数称为模式H 的模式阶, 记作 $O(H)$, 如 $O(011 * 1 *) = 4$ 。模式阶用来反映不同模式间确定性的差异, 模式阶越高, 模式的确定性就越高, 所匹配的样本个数就越少。

模式H 中第一个确定位置和最后一个确定位置之间的距离称为模式的定义距, 记作 $\delta(H)$, 如 $\delta(011 * 1 *) = 4$ 。阶数相同的模式会有不同的性质, 定义距就反映了这种性质的差异。

模式定理为在遗传算子选择、交叉、变异的作用下, 具有低阶、短定义距以及平均适应度高于种群平均适应度的模式在子代中呈指数增长。具有低阶、短定义距以及平均适应度高于种群平均适应度的模式被定义为积木块。遗传算法通过短定义距、低阶以及高平均适应度的模式(积木块), 产生适应度更高的个体, 从而找到更优的可行解。

2.2 基本遗传算法

遗传基因型：编码原则包含完备性，问题空间的所有解都能表示为所设计的基因型；健全性，任何一个基因型都对应于一个可能解；非冗余性，问题空间和表达空间一一对应。

编码方式包含二进制编码、浮点数编码、格雷码编码、符号编码、复数编码、DNA编码等。在交叉操作时，二进制编码比浮点数编码产生新个体的可能性多，而且产生的新个体不受父个体所构成的超体的限制；在变异操作时，二进制编码的种群稳定性比浮点数编码差。

适应度函数及其尺度变换：适应度函数的选取直接影响遗传算法的收敛速度以及能否找到最优解，设计不当有可能出现欺骗问题，进化初期，个别超常个体控制选择过程，进化末期，个体差别太小导致陷入局部极值。一般而言，适应度函数是由目标函数变换而成的，对目标函数值域的某种映射变换称为适应度的尺度变换（fitness scaling）。

常见的适应度函数包含直接转换、界限构造法、线性变换法、幂函数变换法、指数变换法。适应度函数设计需要满足单值、连续、非负、最大化；合理、一致性；计算量小；通用性强。

遗传操作：个体选择概率包含按比例的适应度分配，适应度高则被选择的概率高；基于线性排序的适应度分配与基于非线性排序的适应度分配。常用选择方法包含轮盘赌选择法、随机遍历抽样法、局部选择法、截断选择法、锦标赛选择法。重组方法包含离散重组、中间重组、线性重组。交叉包括单点交叉、多点交叉、均匀交叉。变异包含实值变异与二进制变异。

2.3 简单函数优化的实例

遗传算法能够解决函数优化，以下通过一元函数求最大值的实例介绍遗传算法。一元函数形式为 $f(x) = x \sin(10\pi \times x) + 2.0$ ，定义域为 $[-1, 2]$ 。函数形状如图2所示。

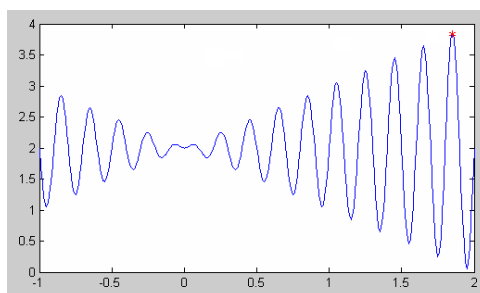


图 2: 一元函数

我们若采用微分法求 $f(x)$ 的最大值，即

$$f'(x) = \sin(10\pi \times x) + 10\pi \times x \times \cos(10\pi \times x) = 0$$

$$\tan(10\pi \times x) = -10\pi \times x$$

得到的解有无穷多个

$$\begin{cases} x_i = \frac{2i-1}{20} + \epsilon_i, i = 1, 2, \dots \\ x_0 = 0 \\ x_i = \frac{2i+1}{20} + \epsilon_i, i = -1, -2, \dots \end{cases}$$

当 i 为奇数时 x_i 对应局部极大值点， i 为偶数时 x_i 对应局部极小值。 x_{19} 即为区间 $[-1, 2]$ 内的最大值点： $x_{19} = \frac{37}{20} + \epsilon_{19} = 1.85 + \epsilon_{19}$

此时，函数最大值 $f(x_{19})$ 比 $f(1.85) = 3.85$ 稍大。

编码：对表现型 x 进行二进制编码，二进制串长取决于求解精度。若要求求解精度到6位小数，区间长度为 $2 - (-1) = 3$ ，需要将区间分为 $3/0.000001 = 3 \times 10^6$ 等份，编码的二进制串长为22位。

产生初始种群：随机产生长度为22的二进制串，产生数目为种群的大小。

如1111010011100001011000
1100110011101010101110
1010100011110010000100
1011110010011100111001
0001100101001100000011
0000011010010000000000
...

计算适应度：不同的问题有不同的适应度计算方法，本实例直接用目标函数作为适应度函数。

将一个二进制串 $(b_{21}b_{20} \dots b_0)$ 转化为10进制数 $(b_{21}b_{20} \dots b_0)_2 = (\sum_{i=0}^{21} b_i \cdot 2^i)_{10} = x'$ 例如某个个体转化为 $[-1,2]$ 区间的实数：

$$s = < 1000101110110101000111 > \rightarrow x = 0.637197$$

计算 x 的函数值（适应度）： $f(x) = x \sin(10\pi x) + 2.0 = 2.586345$

遗传操作：使用轮盘赌选择法进行选择，单点交叉进行交叉，小概率变异进行变异。

实验结果：种群大小设置为50，交叉概率为0.75，变异概率为0.05，最大代数200。
实验得到的最佳个体为

$$s_{max} = < 1111001100111011111100 >$$

$$x_{max} = 1.8506$$

$$f(x_{max}) = 3.8503$$

进化过程如图3所示

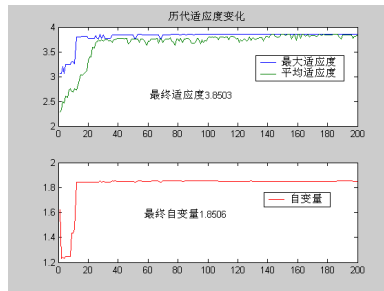


图 3: 进化过程

种群进化过程中选择、交叉、变异操作对种群的影响如图4、5、6所示

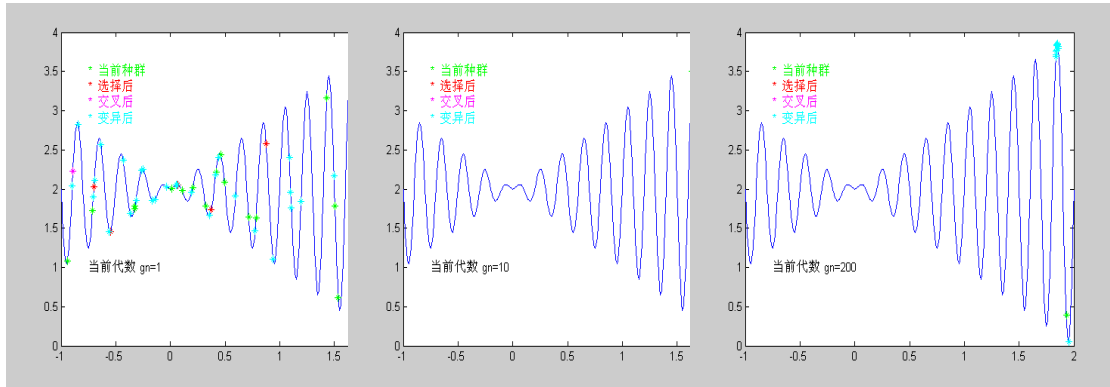


图 4: 第1代

图 5: 第10代

图 6: 第200代

3 遗传算法的具体应用

3.1 解决带约束的函数优化问题

遗传算法解决有约束问题的关键是对约束条件的处理（直接按无约束问题处理是行不通的：随机生成的初始点中可能有大量不可行解；遗传算子作用于可行解后可能产生不可行解）。解决方法是罚函数法，将罚函数包含到适应度评价中。对于线性约束优化问题，目标函数可以是线性函数或非线性函数，思路为消除可能的变量，消除等式约束；设计罚函数；设计特别的遗传操作。

3.2 解决多目标优化问题

使用将多个目标函数表示为一个目标函数作为遗传算法的适应函数的聚合函数法、基于子种群的产生根据每一个目标函数分别进行选择的向量评价遗传算法；基于排序的多目标遗传算法；基于小生境Pareto遗传算法。

3.3 解决组合优化问题

旅行商问题(TSP Benchmark)可使用遗传算法解决。

编码：直接采用解的表示形式，30位（30个城市）长，每位代表所经过的城市序号（无重复）

适应度函数：个体所代表的路径距离的倒数

选择：轮盘赌方法

交叉：有序交叉法,随机选取两个交叉点,两个父个体交换中间部分,替换交换后重复的城市序号

变异：随机选择同一个个体的两个点进行交换

初始参数：种群规模为100，交叉概率为0.8，变异概率为0.8，终止代数为2000。

实验结果如图7、8、9、10所示：

3.4 遗传算法在过程建模中的应用

利用遗传算法可以优化神经网络的权值，将权值表示为个体，适应度设计为样本数据与神经网络预测数据的误差和的倒数，可得到优化的参数。

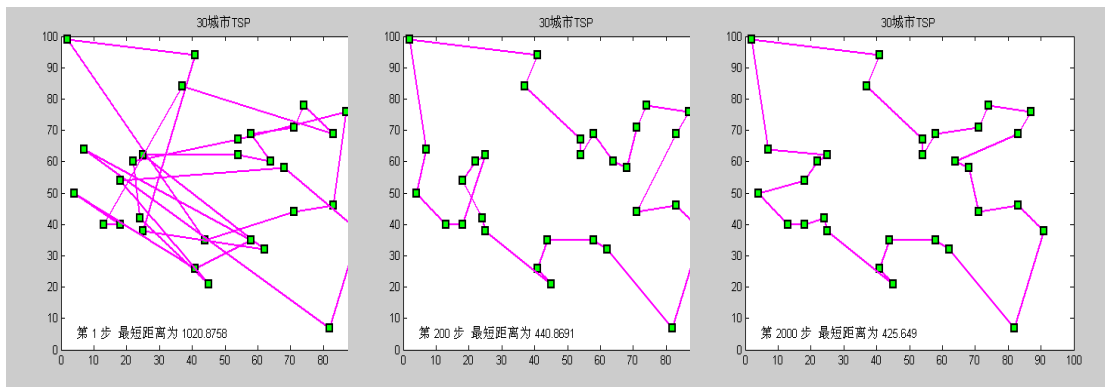


图 7: 第1步

图 8: 第200步

图 9: 第2000步

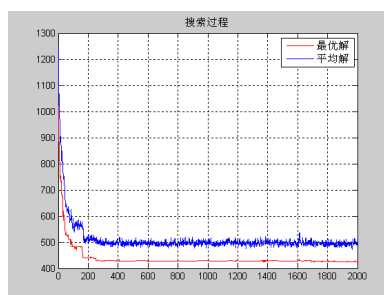


图 10: 搜索过程

3.5 遗传算法在模式识别中的应用

从可能的 m 个特征中依据某个评价标准选出 d 个特征($m > d$)，个体的表示方法为一个长度为 L 的个体对应于一个 L 维的二进制特征矢量，它的每一位就表示包括或排除一个相应的特征。一个个体即是一个可能的最优特征子集；适应度函数的设计：个体所代表的特征子集进行分类时的识别率；遗传算子：可采用各种方法。

4 遗传算法的改进

遗传算法在优化识别方面的效率非常高，而且受到越来越广泛的研究，然而遗传算法自身也有一些缺点，在寻优过程中易出现早熟，设计变量增多时效率较低以及结构分析时间长，可采用CHC算法、自适应遗传算法、基于小生境技术的遗传算法进行改进。

参考文献

- [1] J. Klockgether and H.-P. Schwefel, “Two-phase nozzle and hollow core jet experiments,” in *Proc. 11th Symp. Engineering Aspects of Magnetohydrodynamics*, pp. 141–148, Pasadena, CA: California Institute of Technology, 1970.
- [2] L. J. Fogel, A. J. Owens, and M. J. Walsh, “Artificial intelligence through simulated evolution,” 1966.
- [3] J. H. Holland, “Genetic algorithms and classifier systems: foundations and future directions,” tech. rep., Michigan Univ., Ann Arbor (USA), 1987.