



智能优化方法主要特点

- 1) 基于**群体**的搜索
- 2) 个体之间相互**共享信息**，同时相互**竞争资源**
- 3) **优胜劣汰**，**适者生存**
- 4) **隐并行性**



智能优化技术

遗传算法 (Genetic Algorithms)

龚文引 (教授、博士生导师)

中国地质大学 (武汉) 计算机学院

April 27, 2020

1. 大纲

从生物进化到进化计算

遗传算法

小结

2. 从生物进化到进化计算

自然进化（演化）

生物界是一种自然进化（趋优）的系统。

- 生物为什么会进化？（生物是如何进化的？）
- 只有生物界是自然进化的吗？
- 人工系统能否自然进化？如何实现？

2. 从生物进化到进化计算

进化的基本条件

生物进化过程的发生需要的基本条件：

- 存在由多个生物个体组成的**种群**；
- 生物个体之间存在着差异，或群体具有**多样性**；
- 生物能够**繁殖**；
- 不同个体具有不同的环境生存能力，**具有优良基因结构的个体繁殖能力强**，反之则弱（即：**存在竞争，优胜劣汰**）。

2. 从生物进化到进化计算

进化机制的基本形式

生物群体的进化机制包括三种基本形式：

- ① 自然选择（优胜劣汰，适者生存）
- ② 杂交
- ③ 突变

另外，**外界**对生物的评价反映了生物的生存**价值和机会**。

遗传算法（Genetic Algorithm）

- 遗传算法是一种随机的启发式搜索策略（randomized heuristic search strategy）；
- 基本思想：模拟自然进化，由多个候选解组成群体；
- 通过杂交和变异繁衍后代。

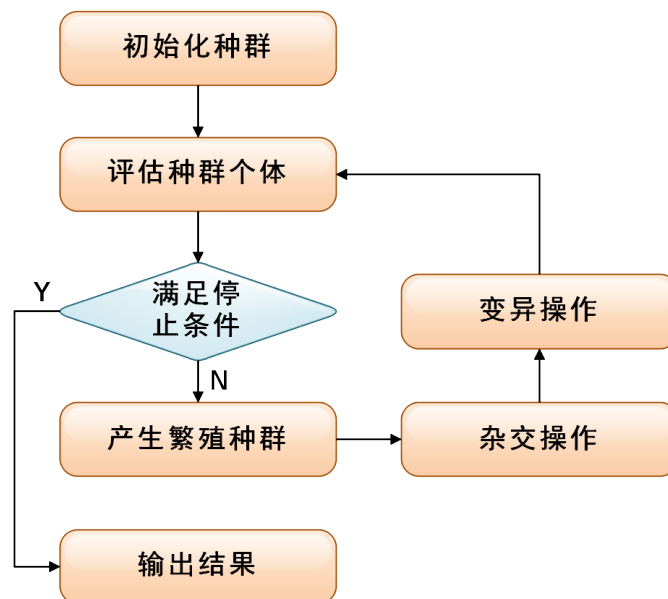
遗传算法（Genetic Algorithm）

- 提出：20 世纪 70 年代
- 早期主要研究者：J. Holland, K. DeJong, D. Goldberg
- 典型应用：离散优化（discrete optimization）
- 主要特点：1）不是很快；2）比较适合求解组合优化问题（combinatorial problems）

基本要素

遗传算法的基本要素：

- ① **参数编码**：群体中的个体怎么表示成问题的解；
- ② **初始群体**：如何产生初始种群；
- ③ **适应度函数的设计**：如何评价个体的好坏；
- ④ **父个体选择**：如何选择父个体用于产生子个体；
- ⑤ **遗传算子设计**：如何产生新个体；
- ⑥ **控制参数设定**：杂交概率 p_c ，变异概率 p_m 的设定；
- ⑦ **迭代终止条件**。



简单 GA 伪码流程

- ① **[Start]**: 产生初始种群;
- ② **[Fitness]**: 计算每个个体的适应值;
- ③ **[New population]**: 产生新种群, 直到达到停止条件
 - **[Selection]**: 从上一代种群中选择合适的个体形成繁殖种群;
 - **[Crossover]**: 按照杂交概率 p_c 对选择的父个体进行杂交;
 - **[Mutation]**: 按照变异概率 p_m 对选择的父个体进行变异;
 - **[Accepting]**: 利用新个体形成子种群;
- ④ **[Replace]**: 通过生存选择在子种群和原始父种群形成新的父种群;
- ⑤ **[Test]**: 如果满足终止条件则输出最优结果, 并退出程序; 否则执行后续步骤;
- ⑥ **[Loop]**: 返回步骤 2.

遗传算法的特点

- 基于**群体**的搜索。
- 直接处理的对象是参数的**编码集**而不一定是问题参数本身。
- 搜索过程中使用的是基于**目标函数值的评价信息**，搜索过程既不受优化函数连续性的约束，也没有优化函数必须可导的要求。
- 具有显著的**隐并行性**。
- 形式上简单明了。
- 具有很强的**鲁棒性**。

问题

求函数 $f(x) = x^2$ 的最大值， $x = \{0, 1, \dots, 30, 31\}$ 。

编码 (Encoding)

编码：定义问题参数空间到编码空间的映射

13 → 01101

24 → 11000

8 → 01000

19 → 10011

初始种群 (population) 的产生

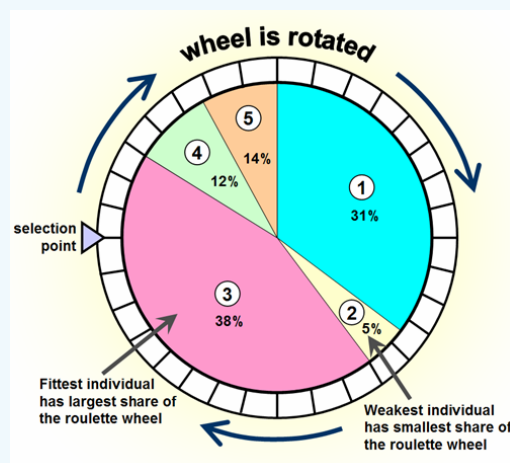
$$\left\{ \begin{array}{l} 01101 \\ 11000 \\ 01000 \\ 10011 \end{array} \right\}$$

适应值评价 (Fitness evaluation), 解码 (Decoding)

01101 $\rightarrow x = 13 \rightarrow f(x) = 169$
11000 $\rightarrow x = 24 \rightarrow f(x) = 576$
01000 $\rightarrow x = 8 \rightarrow f(x) = 64$
10011 $\rightarrow x = 19 \rightarrow f(x) = 361$

遗传算子：选择 (Selection)

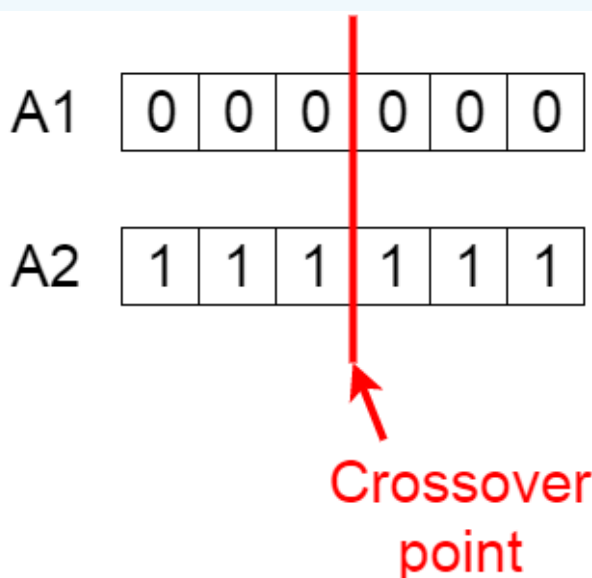
简单遗传算法一般采用轮盘赌选择 (Roulette wheel selection)



遗传算子：选择（Selection）

String no.	Initial population	x Value	Fitness $f(x) = x^2$	$Prob_i$	Expected count	Actual count
1	0 1 1 0 1	13	169	0.14	0.58	1
2	1 1 0 0 0	24	576	0.49	1.97	2
3	0 1 0 0 0	8	64	0.06	0.22	0
4	1 0 0 1 1	19	361	0.31	1.23	1
Sum			1170	1.00	4.00	4
Average			293	0.25	1.00	1
Max			576	0.49	1.97	2

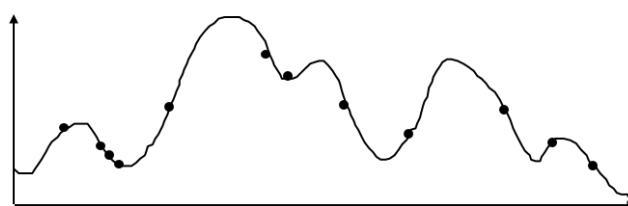
遗传算子：杂交（Crossover）



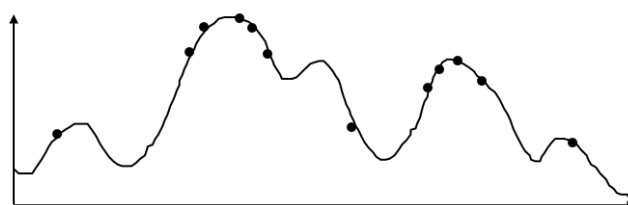
遗传算子：变异（Mutation）

String no.	Offspring after xover	Offspring after mutation	x Value	Fitness $f(x) = x^2$
1	0 1 1 0 0	1 1 1 0 0	26	676
2	1 1 0 0 1	1 1 0 0 1	25	625
2	1 1 0 1 1	1 1 0 1 1	27	729
4	1 0 0 0 0	1 0 1 0 0	18	324
Sum				2354
Average				588.5
Max				729

一个抽象示例



第0代群体的分布



第n代群体的分布

个体编码

- 二进制编码: 0101...1100
- 实数编码: 43.2, -33.1, ..., 0.0, 89.2
- 互斥 (permutation) 元素编码: [1, 2, 3, 4]; [4, 3, 2, 1]
- 规则列表编码: $R_1, R_2, R_3, \dots, R_{22}, R_{23}$
- 树编码: 主要用于遗传规划算法 (Genetic Programming)
- ...

父个体选择

遗传算法的父个体选择主要是为了产生繁殖种群, 用于后续杂交算子和变异算子产生新种群。

- 适应值比例选择 (Fitness proportionate selection), 也叫轮盘赌选择 (roulette-wheel selection)
- 基于奖励的选择 (Reward-based selection)
- 随机广义抽样 (Stochastic universal sampling)
- 锦标赛选择 (Tournament selection)
- 截断选择 (Truncation selection)
- ...

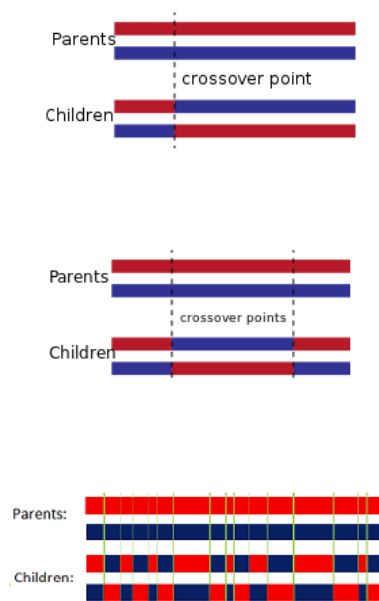
[https://en.wikipedia.org/wiki/Selection_\(genetic_algorithm\)](https://en.wikipedia.org/wiki/Selection_(genetic_algorithm))

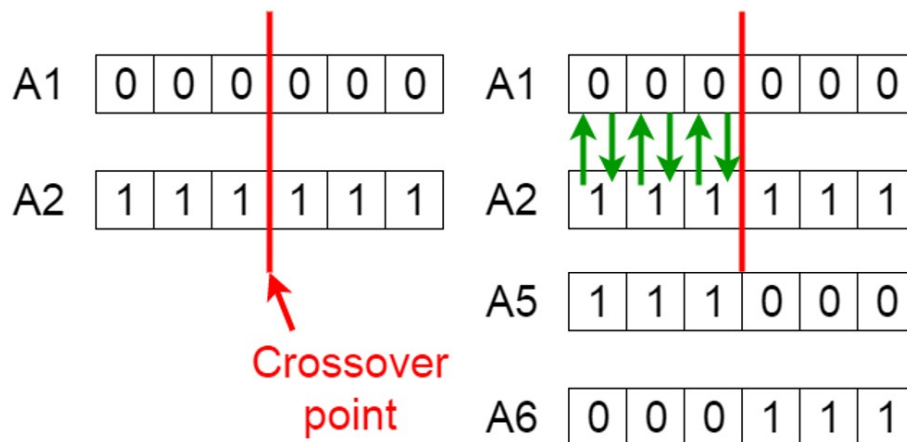
杂交算子

遗传算法的杂交算子利用多个父个体产生一个或多个子个体。一般地，针对不同的编码技术有不同的杂交算子。

- 单点杂交 (Single-point crossover)
- 两点杂交 (Two-point crossover)
- 均匀杂交 (Uniform crossover)
- 多父体杂交 (Multi-parent crossover)
- 针对边个体的杂交 (For edge chromosomes)
- ...

- G. Pavai & T. V. Geetha, "A survey on crossover operators," *ACM Comput. Surv.* vol. 49, no. 4, Article 72, 2016.
- [https://en.wikipedia.org/wiki/Crossover_\(genetic_algorithm\)](https://en.wikipedia.org/wiki/Crossover_(genetic_algorithm))
- <http://www.obitko.com/tutorials/genetic-algorithms/crossover-mutation.php>





变异算子

遗传算法的变异算子利用一个父个体产生一个或多个子个体。一般地，针对不同的编码技术有不同的变异交算子。

- Bit 串变异 (Bit string mutation)
- 边界变异 (Boundary)
- 均匀变异 (Uniform mutation)
- 高斯变异 (Gaussian mutation)
- 收缩变异 (Shrink mutation)
- ...

- [https://en.wikipedia.org/wiki/Mutation_\(genetic_algorithm\)](https://en.wikipedia.org/wiki/Mutation_(genetic_algorithm))
- <http://www.obitko.com/tutorials/genetic-algorithms/crossover-mutation.php>

Before Mutation

A5

1	1	1	0	0	0
---	---	---	---	---	---

After Mutation

A5

1	1	0	1	1	0
---	---	---	---	---	---

作业

在网上搜索遗传算法相关代码，要求提交以下子程序，**每行需要中文注释：**

- 群体初始化
- 轮盘赌选择
- 单点杂交



把函数曲线理解成一个一个山峰和山谷组成的山脉，那么我们可以设想所得到的每一个解就是一只袋鼠，我们希望它们不断的向着更高处跳去，直到跳到最高的山峰。所以求最大值的过程就转化成一个“袋鼠跳”的过程。

爬山算法

一只袋鼠朝着比**现在**高的地方跳去。它找到了不远处的最高的山峰。但是这座山不一定是最高峰。这就是爬山算法，它不能保证局部最优值就是全局最优值。

模拟退火算法

袋鼠喝醉了。它随机地跳了很长时间。这期间，它可能走向高处，也可能踏入平地。但是，它渐渐清醒了并朝最高峰跳去。这就是模拟退火算法。

遗传算法

有很多袋鼠，它们降落到喜马拉雅山脉的任意地方。这些袋鼠并不知道它们的任务是寻找珠穆朗玛峰。但每过几年，就在一些海拔高度较低的地方射杀一些袋鼠。于是，不断有袋鼠死于海拔较低的地方，而越是在海拔高的袋鼠越是能活得更久，也越有机会生儿育女。就这样经过许多年，这些袋鼠们竟然都不自觉地聚拢到了一个个的山峰上，可是在所有的袋鼠中，只有聚拢到珠穆朗玛峰的袋鼠被带回了美丽的澳洲。

4. 小结

本章小结

- ① 遗传算法相关概念
- ② 简单遗传算法示例
- ③ 遗传算法扩展介绍

4. 小结

思考

从网上收集有关遗传算法的相关代码，看懂代码，并运行。思考以下几个问题：

- 和课程中所介绍的简单遗传算法的区别与联系如何？
- 运行结果如何？能否达到预期要求吗？
- 有哪些可能存在的问题？
- 个体编码和问题参数之间如何转化？

4. 小结

进一步阅读

- S. Forrest, "Genetic algorithms: principles of natural selection applied to computation," *Science*, vol. 261, no. 5123, pp. 872-878, 1993.

5. 致谢

Thank you!

AUTHOR: GONG, Wenyin
ADDRESS: School of Computer Science,
China University of Geosciences,
Wuhan, 430074, China
E-MAIL: wygong@cug.edu.cn
HOMEPAGE: <http://www.escience.cn/people/wygong>