
遗传算法

王晓丽

计算机科学与技术学院 副教授

遗传算法

■ 遗传算法基本思想：

- ✓ 首先对问题进行**编码**，产生初始种群。
- ✓ 然后对个体进行**交叉**、**变异**等遗传操作，产生出新的个体。
- ✓ 再按照优胜劣汰的原则对个体进行**选择**
- ✓ 如此往复，逐代演化产生出越来越好的个体。

1. 编码

- 将问题的解变换为位串形式表示的过程叫编码；

问题的解 X $\xrightarrow{\text{编码}}$ 个体 / 染色体

- 将位串形式表示变换为原问题的解的过程叫解码。

- 把位串形式编码叫染色体(chromosome)或个体(individual)，如：0100111010100011

- 染色体的每位称为基因(gene)。
-

常用的编码方案

- ✓ 二进制编码
- ✓ 实数编码
- ✓ 整数编码
- ✓ 排列编码
- ✓ 有限状态机编码
- ✓ 树编码
- ✓

二进制编码

□ 函数 $f(x)=x^2$, 其自变量 x 在区间 $[0, 31]$ 取整数值, 求该函数的最大值。

✓ **确定适当的编码方案:**

二进制编码

- 函数 $f(x)=x^2$, 其自变量 x 在区间 $[0, 31]$ 取整数值, 求该函数的最大值。
- ✓ **确定适当的编码方案:**
 - 采用二进制数来对其编码, 由 $2^5 = 32$, 故使用5位无符号二进制数构成染色体。

二进制编码

□ 函数 $f(x)=x^2$, 其自变量 x 在区间 $[0, 31]$ 取整数值, 求该函数的最大值。

✓ 种群初始化:

- 通过随机的方法来产生 N 个个体组成初始种群。
- 使用计算机在 $0\sim 1$ 之间产生随机数 K , 并按照数 K 的值初始化基因位:
 - $0\leq K<0.5$, 基因为置为1
 - $0.5\leq K\leq 1$, 基因为置为0

二进制编码

□ 函数 $f(x)=x^2$, 其自变量 x 在区间 $[0, 31]$ 取整数值, 求该函数的最大值。

✓ **种群初始化:**

■ 随机生成 $N = 4$ 个个体:

➤ 01101 11000 01000 10011

■ 构造了初始种群, 完成了遗传算法的准备工作。

Question: 哪个个体最好?

适应度

- 函数 $f(x)=x^2$, 其自变量 x 在区间 $[0, 31]$ 取整数值, 求该函数的最大值。

标号	个体	解码	适应值 $f(x) = x^2$
1	01101	13	169
2	11000	24	576
3	01000	8	64
4	10011	19	361

适应度

- 函数 $f(x)=x^2$, 其自变量 x 在区间 $[0, 31]$ 取整数值, 求该函数的最大值。

标号	个体	解码	适应值 $f(x) = x^2$
1	01101	13	169
2	11000	24	576
3	01000	8	64
4	10011	19	361

整数编码

货郎担问题(Travelling Salesman Problem, 简记为TSP): 设有n个城市, 城市i和城市j之间的距离为 $d(i,j)$ 。TSP问题是寻找最短的一条回路, 要求该回路能够遍访每个城市且每个城市仅访问一次。

TSP问题可以表述为:

$$\min \left\{ \sum_{i=1}^{n-1} d(C_i, C_{i+1}) + d(C_n, C_1) \right\}$$

其中 $C = (C_1, C_2, \dots, C_n)$ 是 $1, 2, 3, \dots, n$ 的一个全排列。

2. 交叉

- ✓ 单点交叉
- ✓ 两点交叉
- ✓ 多点交叉
- ✓ 部分匹配交叉
- ✓ 顺序交叉

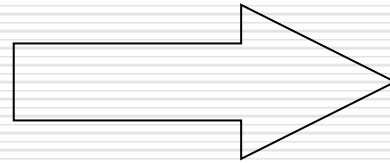
单点交叉

- 具体操作是：在个体基因串中随机设定一个交叉点，该点前或后的两个个体的基因互换，并生成两个新个体。

P_1 A_1 B_1
011 | 0011

P_2 A_2 B_2
101 | 0110

单点交叉

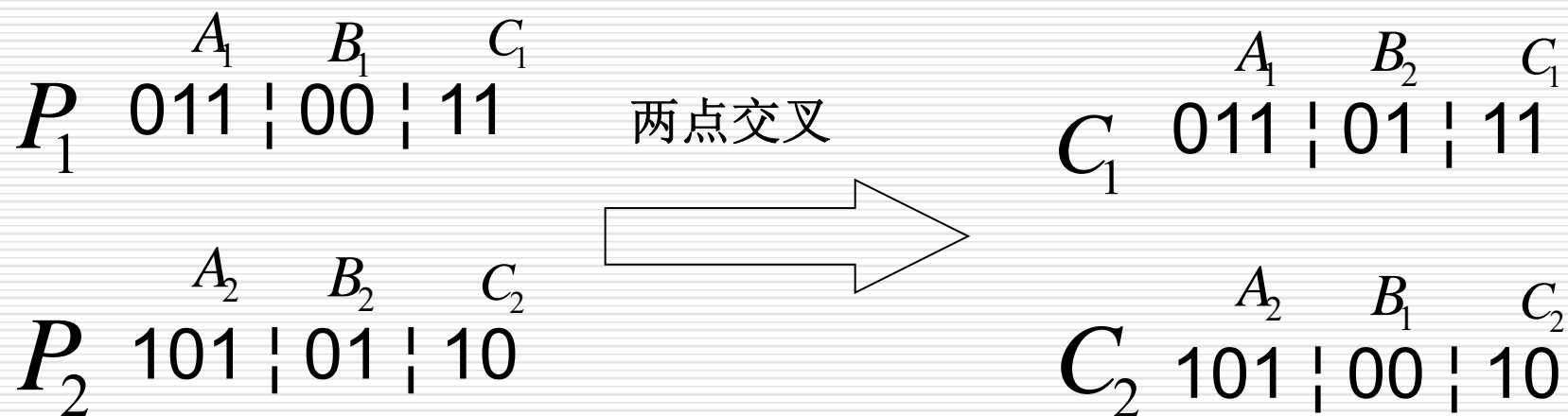


C_1 A_1 B_2
011 | 0110

C_2 A_2 B_1
101 | 0011

两点交叉

□ 随机产生两个交叉点



多点交叉

$$P_1 \quad \begin{array}{ccccc} A_1 & B_1 & C_1 & D_1 & E_1 \\ 0110 & | & 001 & | & 11011 & | & 01 & | & 010 \end{array}$$

$$P_2 \quad \begin{array}{ccccc} A_2 & B_2 & C_2 & D_2 & E_2 \\ 101 & 1 & | & 011 & | & 11110 & | & 01 & | & 001 \end{array}$$

- 多点交叉不常被采用，因为当基因链码的长度 n 较小，或交叉点数 c 较大时，**具有优良特性的模式很容易被破坏**。另外出于计算速度的考虑，基因链码的长度 n 通常不会太长。

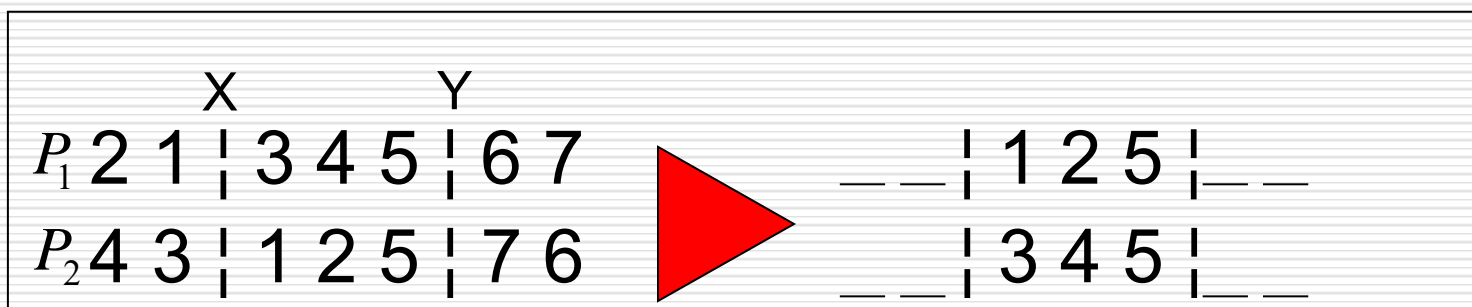
2. 交叉

- ✓ 单点交叉
- ✓ 两点交叉
- ✓ 多点交叉
- ✓ 部分匹配交叉
- ✓ 顺序交叉

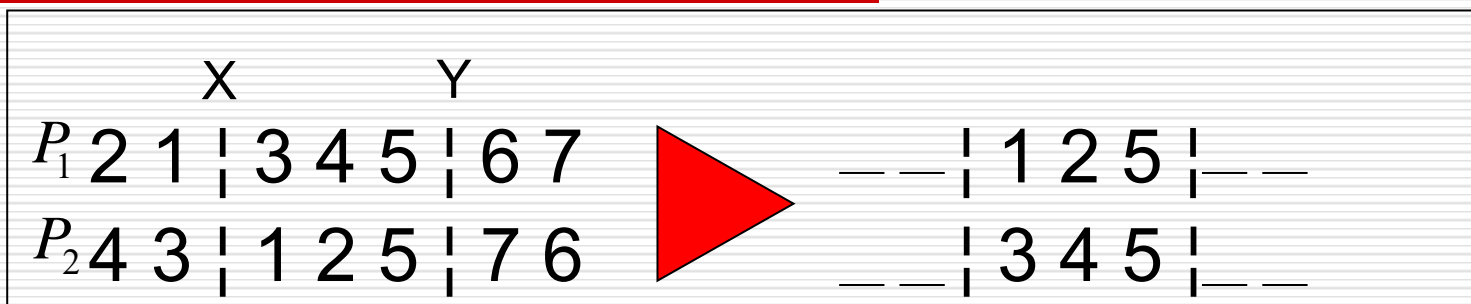
部分匹配交叉

Partial-Mapped Crossover

- 选切点X,Y;
- 交换中间部分;
- 确定映射关系;
- 将未换部分按映射关系恢复合法性。



部分匹配交叉 Partial-Mapped Crossover



映射关系：3-1, 4-2, 5-5

则： c_1 4 3 | 1 2 5 | 6 7
 c_2 2 1 | 3 4 5 | 7 6

部分匹配交叉

Partial-Mapped Crossover

- 选切点X,Y;
- 交换中间部分;
- 确定映射关系;
- 将未换部分按映射关系恢复合法性。

$$I_1 = \left(12, 10, 6, 9, 4, \overset{k_1}{:} 1, 13, 0, 11, 5, 3, 7, \overset{k_2}{:} 14, 2, 8 \right)$$

$$I_2 = (14, 12, 10, 8, 4, : 11, 6, 13, 5, 2, 1, 0, : 9, 3, 7)$$

部分匹配交叉

Partial-Mapped Crossover

$$I_1 = \left(12, 10, 6, 9, 4, \overset{k_1}{:} 1, 13, 0, 11, 5, 3, 7, \overset{k_2}{:} 14, 2, 8 \right)$$
$$I_2 = (14, 12, 10, 8, 4, : 11, 6, 13, 5, 2, 1, 0, : 9, 3, 7)$$

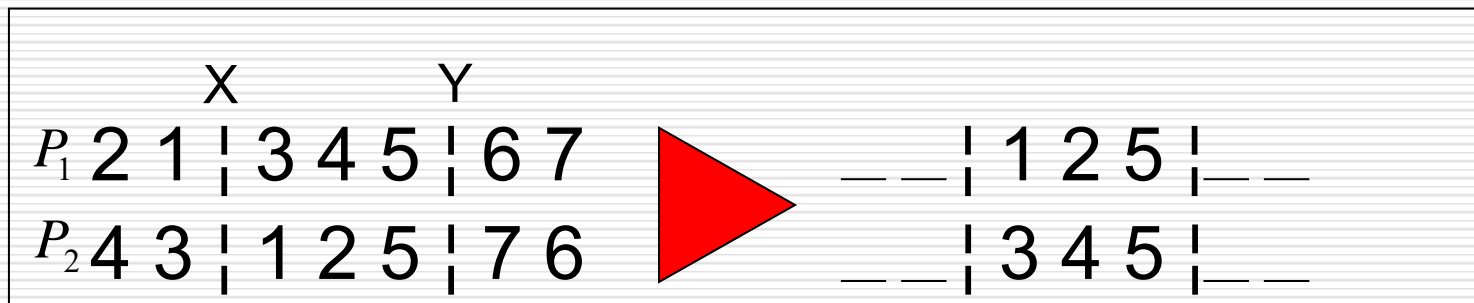
$$I_1'' = (12, 10, 7, 9, 4, 11, 6, 13, 5, 2, 1, 0, 14, 3, 8)$$

$$I_2'' = (14, 12, 10, 8, 4, 1, 13, 0, 11, 5, 3, 7, 9, 2, 6)$$

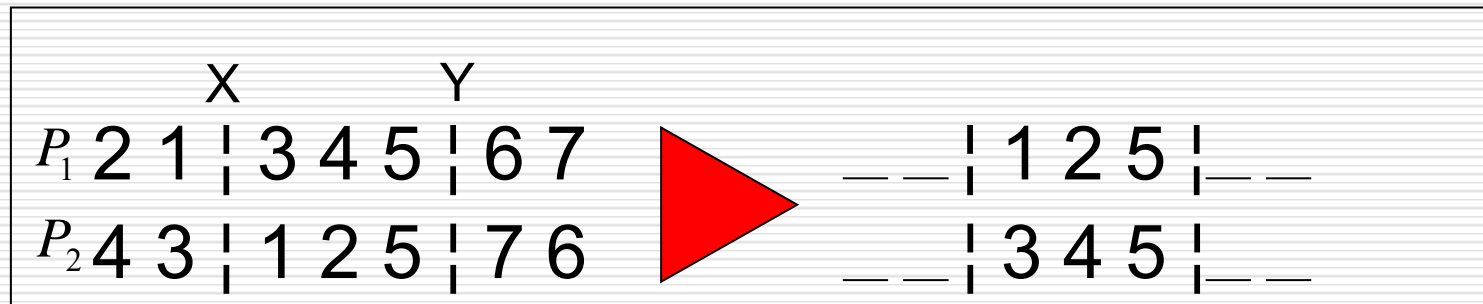
顺序交叉

Order Crossover

- 选切点X,Y;
- 交换中间部分;
- 从切点Y后第一个基因起列出原顺序, 去掉已有基因;
- 从切点Y后第一个位置起, 按顺序填入。



顺序交叉 Order Crossover



列出基因: 6 7 ~~2~~ ~~1~~ 3 4 ~~5~~ 7 6 ~~4~~ ~~3~~ 1 2 ~~5~~



则: C_1 3 4 | 1 2 5 | 6 7
 C_2 1 2 | 3 4 5 | 7 6

3. 变异

- 变异：先取定变异概率 p_m （一般较小， $p_m \leq 0.05$ ），对交叉后代集中的每个后代

$$O = \overbrace{** \dots *}^m$$

的每位基因生成一个随机数 $r \in [0,1]$ ，若 $r \leq p_m$ 则将此基因的*改为 $1 - *$ ，否则*保持不变，*代表0或1。

- 即：若 $r \leq p_m$ 则将数字0变为1，1变为0；否则保持不变。
- 例如： O_1 中的第5位和第12位需要改变：

$$O_1 = 0000100000101111000101$$

$$\overline{O_1} = 000000000011111000101$$

3. 变异

- 对换变异：随机地在染色体上选取两个位置，交换两位置的基因。

✓ 4 3 1 2 5 6 7 --》 4 5 1 2 3 6 7

- 移位变异：任选一位移到最前

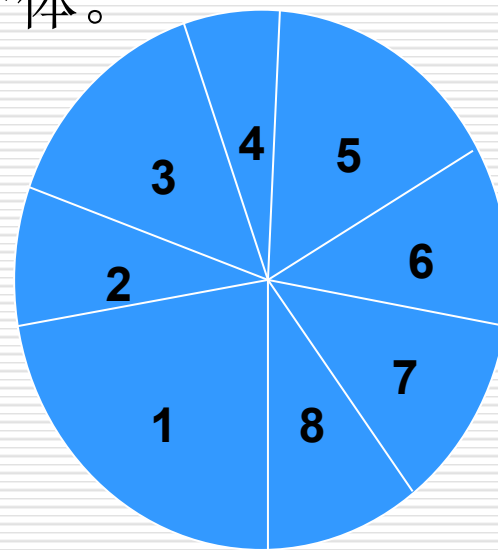
✓ 4 3 1 2 5 6 7 --》 5 4 3 1 2 6 7

4. 选择

- ✓ 轮盘赌选择
- ✓ 两两竞争法选择
- ✓ 锦标赛选择
- ✓ 精英保留

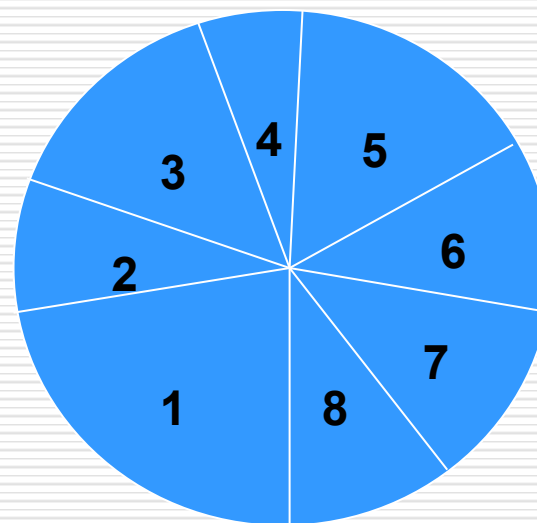
轮盘赌选择

- 为群体中每个个体指定饼图中一个小块。块的大小与个体的适应度成比例，适应度愈高，它在饼图中对应的小块所占面积也愈大。为了选取一个个体，要做的就是旋转这个轮子，直到轮盘停止时，看指针停止在哪一块上，就选中与它对应的那个个体。



轮盘赌选择

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} = \frac{f_i}{f_{sum}}$$



- ✓ f_i 为个体的适应度； f_{sum} 为种群的总适应度； p_i 为个体*i*的选择概率。
- ✓ 举例： $N=4$, $f_1=1$, $f_2=4$, $f_3=2$, $f_4=3$

轮盘赌选择

1) 计算各个体适应度值

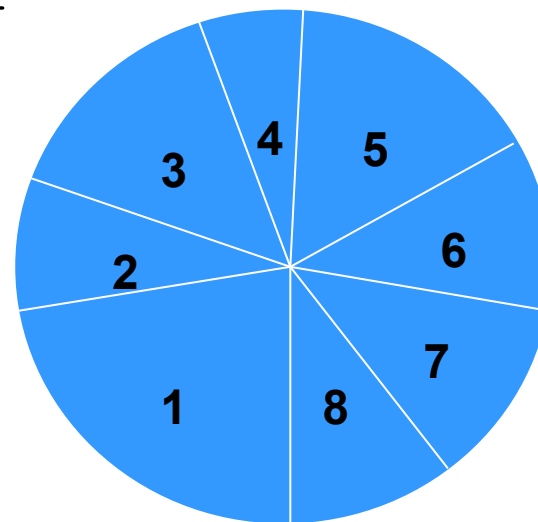
2) 计算各个体的选择概率,
$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} = \frac{f_i}{f_{sum}}$$

3) 计算每个个体的概率累加值 $q_k = \sum_{i=1}^k p_i$

4) 产生一个随机数 r , $0 \leq r \leq 1$

5) 若 $q_{k-1} < r \leq q_k$, 则选择第 k 个个体。

6) 重复 (4) 和 (5), 直到获得足够的个体。



4. 选择

- ✓ 轮盘赌选择
- ✓ 两两竞争法选择
- ✓ 锦标赛选择
- ✓ 精英保留

两两竞争法选择

- 随机地在种群中取两个个体；
- 选择两个个体中适应度大的个体；若两者相同，只取其中之一；
- 重复上述步骤直到选出的个体数满足需求。

4. 选择

- ✓ 轮盘赌选择
- ✓ 两两竞争法选择
- ✓ 锦标赛选择
- ✓ 精英保留

锦标赛选择

- 随机地在当前种群中取 S 个个体；
- 从 S 个个体中选择适应度最大的一个个体到下一代群体中。
- 重复上述步骤，直到选择的个体数达到预先设定的数目为止。 S 规模一般不要太大，取 2~4 即可。
- **优点：**对个体适应度取正负值无要求，但此方法的随机性更强，但是有较高概率保证最优个体被选择，最差个体被淘汰。

4. 选择

- ✓ 轮盘赌选择
- ✓ 两两竞争法选择
- ✓ 锦标赛选择
- ✓ 精英保留

精英保留选择

- 将适应度值最高的前 E 个个体直接保留到下一代种群中。
- **优点：**进化过程中到目前为止的最好解不会被破坏掉。