

遗传算法

陈志华



主要内容

- 遗传算法概述
- 遗传算法基本概念
- 简单遗传算法
- 遗传算法应用举例

1. 概述

□ 遗传算法 (Genetic Algorithm, GA)

- 是基于达尔文生物进化论的**自然选择和遗传学机理**的**计算模型**
- 是一种通过模拟自然进化过程**搜索最优解**的方法

遗传算法之父

- **约翰·霍兰德** (John Holland, 1929.2.2—2015.8.9),
美国密歇根大学心理学教授和电子工程及计算机科学教授
- 1975年出版了专著《自然与人工系统中的适配》
(Adaptation in Natural and Artificial Systems), **正式创立遗传算法的理论**

遗传算法之父



□ GEC-summit-2009 (2009遗传与进化计算国际峰会)

- 由ACM/SIGEVO主办的一个高水平国际学术会议
- 于2009年6月12-14日首次在中国上海召开
- <http://www.sigevo.org/gec-summit-2009/>

算法思路

- 遗传算法（Genetic Algorithm, GA）是一种基于**模拟进化**的学习方法
 - 假设通常被描述为**二进制位串**
 - 从若干**初始假设的群体或集合**，开始搜索合适的假设
 - 当前群体的成员通过**模拟生物进化的方式**来产生**下一代群体**，比如随机变异和交叉

算法思路

- 在每一次的迭代中，根据给定的适应度评估当前群体中的假设，而后使用概率方法**选出适应度最高的假设**作为产生下一代的种子

算法应用领域

- 遗传算法在人工智能的众多领域得到了广泛应用
 - 机器学习、聚类、控制（如煤气管道控制）、规划（如生产任务规划）、设计（如通信网络设计、布局设计）、调度（如作业车间调度、机器调度、运输问题）、配置（机器配置、分配问题）、组合优化（如TSP、背包问题）、函数的最大值

算法应用领域

- 另一方面，人们又将遗传算法与其他智能算法和技术相结合，使其问题求解能力得到进一步扩展和提高。例如，将遗传算法与模糊技术、神经网络相结合，已取得了不少成果
- 此外，GA也在自动控制、机器人学、图象处理、人工生命等方面获得了广泛的运用

算法说明

- 遗传算法是一种受生物进化启发的学习方法，它不再是从一般到特殊或从简单到复杂地搜索假设，而是通过变异和重组当前已知的最好假设来生成后续的假设
- 每一步更新被称为当前群体的一组假设，方法是使用当前适应度最高的假设的后代替代群体的某个部分

算法说明

□ 适应度实例：

- 如果学习任务是在给定一个未知函数的输入输出训练样例后逼近这个函数，适应度可被定义为**假设在训练数据上的精度**
- 如果是学习下国际象棋的策略，适应度可被**对弈的胜率**

算法说明

□ 遗传算法具有以下共同结构：

- 算法迭代更新一个**假设池**(称为群体)
- 在每一次迭代中，根据适应度评估群体中的所有成员，然后用概率方法选取适应度最高的个体**产生新一代群体**
- 在被选中的个体中，**一部分保持原样**，进入下一代群体；**其他被用作产生后代个体的基础**，采用交叉和变异方法

2.基本概念

1. 问题的解空间

遗传算法主要用来针对问题搜索**它的最优解或次优解**。问题的最优解和次优解都包含在一个**庞大的解集合**中，即问题的解空间。

2.基本概念

2. 个体与种群

- 个体就是模拟生物个体，对问题中的对象（一般就是问题的解）的一种称呼，一个个体也就是解空间中的一个点。
- 种群是模拟生物种群，由若干个体组成的群体，它一般是整个解空间的一个很小的子集。

2.基本概念

3. 适应度(fitness)与适应度函数

- 适应度就是借鉴生物个体对环境的适应程度，而对问题中的个体对象所设计的**表征其优劣的一种测度。**

- 适应度函数就是问题中的全体个体与其适应度之间的一个对应关系。它一般是一个实值函数。该函数就是**遗传算法中指导搜索的评价函数**，需要有效反映任一个染色体和最优解染色体的差距。

2.基本概念

4. 染色体与基因

染色体 (chromosome) 就是问题中个体的某种字符串形式的编码表示。字符串中的字符也就称为基因 (gene)，即染色体上的任一位。

例如：

| 个体 | | 染色体 |
|-----------|------|-------------|
| 9 | ---- | 1001 |
| (2, 5, 6) | ---- | 010 101 110 |

2.基本概念

5. 遗传操作

亦称遗传算子(genetic operator), **就是关于染色体的运算**。遗传算法中有三种遗传操作:

- 选择-复制(selection-reproduction)
- 交叉(crossover, 亦称交换、交配或杂交)
- 变异(mutation, 亦称突变)

2.基本概念

选择-复制 对于一个规模为 N 的种群 S , 按每个染色体 $x_i \in S$ 的选择概率 $P(x_i)$ 所决定的选中机会, 分 N 次从 S 中随机选定 N 个染色体, 并进行复制。

这里的选择概率 $P(x_i)$ 的计算公式为

$$P(x_i) = \frac{f(x_i)}{\sum_{j=1}^N f(x_j)}$$

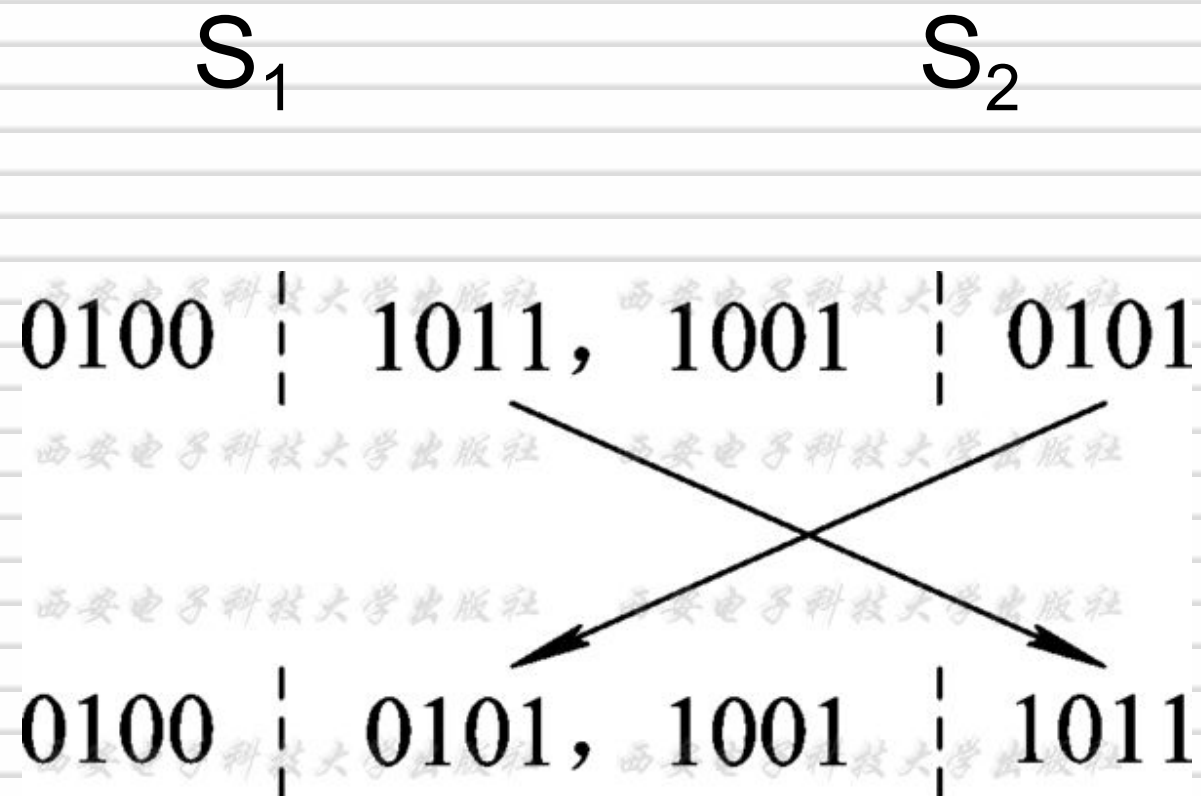
2.基本概念

交叉 就是互换两个染色体某些位上的基因

例如, 设染色体 $s_1=01001011$, $s_2=10010101$,

交换其后4位基因

2.基本概念



$$s_1' = 01000101, \quad s_2' = 10011011$$

可以看做是原染色体 s_1 和 s_2 的**子代染色体**。

2.基本概念

变异：就是改变染色体某个(些)位上的基因。

例如，设染色体 $s=11001101$

将其第三位上的0变为1, 即

$$s=11\underline{0}01101 \rightarrow 11\underline{1}01101 = s'$$

s' 也可以看做是原染色体 s 的子代染色体。

3. 简单遗传算法

□ 简单遗传算法又称为 SGA

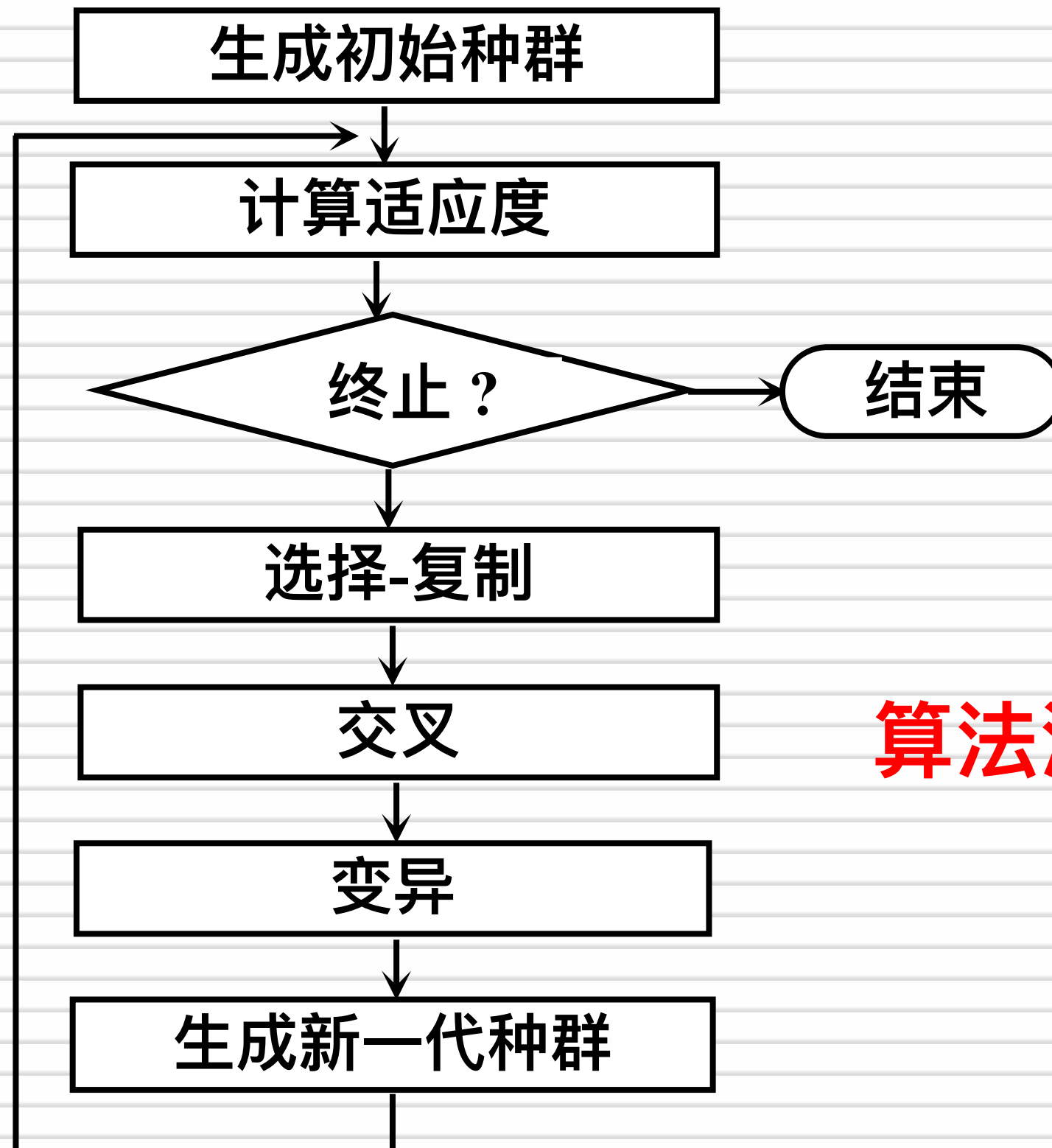
■ Simple Genetic Algorithm

■ Holland Genetic Algorithm

□ 思想

■ 从初始种群出发，采用基本的遗传算子进行运算，
产生下一代种群；反复进行，直到满足终止条件

3. 简单遗传算法



算法流程图

3. 简单遗传算法

算法中的一些控制参数：

- **种群规模**：种群包含的染色体数目
- **最大换代数**：染色体更新换代的最大数目
- **交叉率(crossover rate)**就是参加交叉运算的染色体个数占全体染色体总数的比例，记为 P_c ，取值范围一般为0.4~0.99。□
- **变异率(mutation rate)**是指发生变异的基因位数所占全体染色体的基因总位数的比例，记为 P_m ，取值范围一般为0.0001~0.1。

3. 简单遗传算法（基本遗传算法）

- 步1 在搜索空间 U 上定义一个**适应度函数** $f(x)$,
给定种群规模 N , 交叉率 P_c 和变异率 P_m , 代数 T ;
- 步2 随机产生 U 中的 N 个个体 s_1, s_2, \dots, s_N , 组成
初始种群 $S=\{s_1, s_2, \dots, s_N\}$, 置代数计数器 $t=1$;
- 步3 计算 S 中每个个体的**适应度**;
- 步4 若终止条件满足, 则取 S 中适应度最大的
个体作为所求结果, 算法结束。

3. 简单遗传算法（基本遗传算法）

步5 按选择概率 $P(x_i)$ 所决定的选中机会，每次从 S 中随机选定1个个体并将其染色体复制，共做 N 次，然后将复制所得的 N 个染色体组成群体 S_1 ；

步6 按交叉率 P_c 所决定的参加交叉的染色体数 c ，从 S_1 中随机确定 c 个染色体，配对进行交叉操作，并用产生的新染色体代替原染色体，得群体 S_2 ；

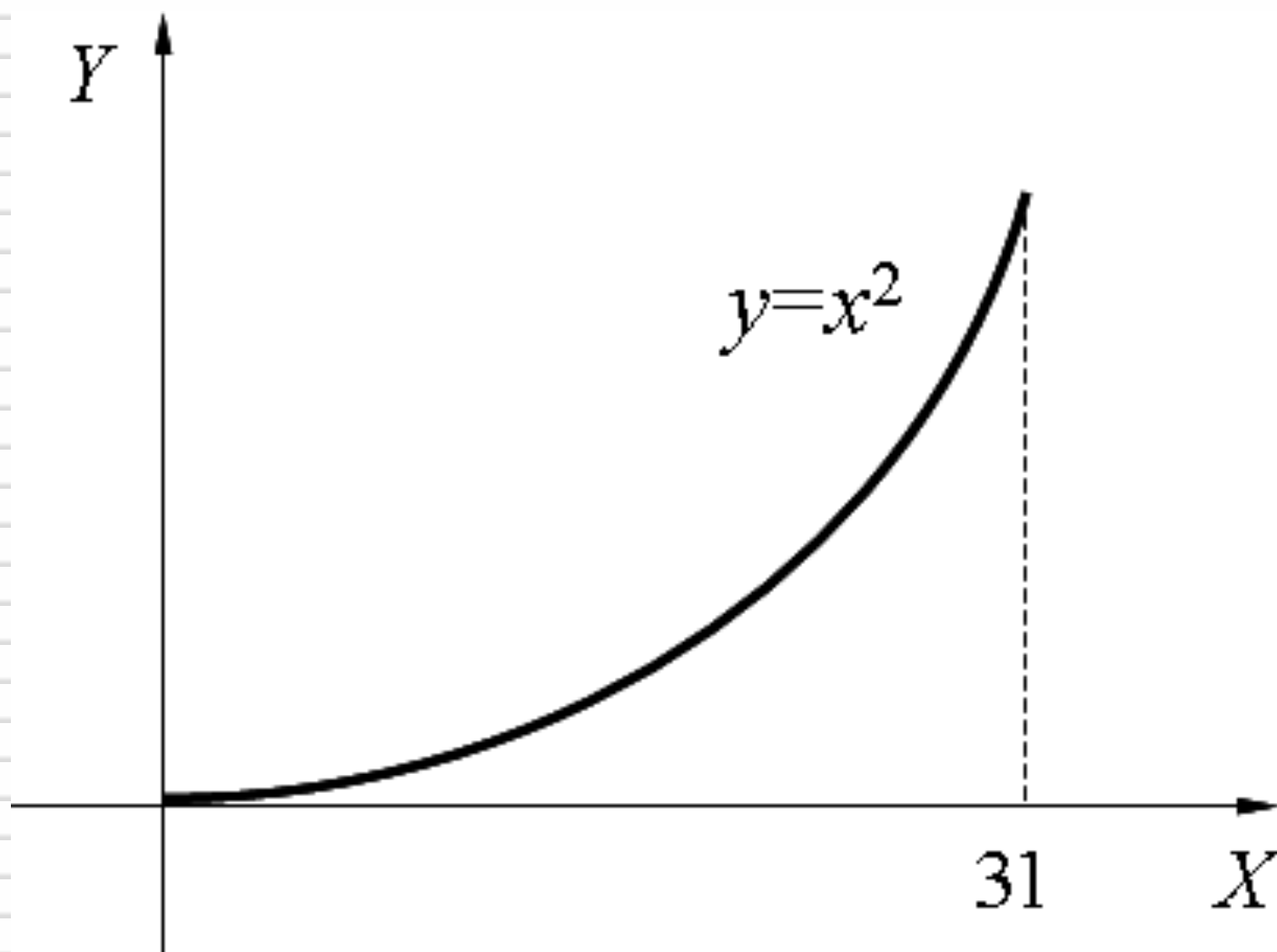
3. 简单遗传算法（基本遗传算法）

步7 按变异率 P_m 所决定的变异次数 m ，从 S_2 中随机确定 m 个染色体，分别进行**变异操作**，并用产生的新染色体代替原染色体，**得群体 S_3** ；

步8 将群体 S_3 作为新一代种群，即用 S_3 代替 S ， $t = t+1$ ，转步3；

4. 遗传算法应用举例

例1 利用遗传算法求解区间 $[0,31]$ 上的二次函数 $y=x^2$ 的最大值。



4. 遗传算法应用举例

分 析

- 1 原问题可转化为在区间 $[0, 31]$ 中搜索能使 y 取最大值的点 a 的问题
- 2 $[0, 31]$ 中的点 x 就是个体, 函数值 $f(x)$ 恰好就可以作为 x 的适应度, 区间 $[0, 31]$ 就是一个(解)空间
- 3 这样, 只要能给出个体 x 的适当染色体编码, 该问题就可以用遗传算法来解决。

4. 遗传算法应用举例

解 (1) 设定种群规模,编码染色体, 产生初始种群

将种群规模设定为4；用5位二进制数编码染色体；

取下列个体组成初始种群 S_1 ：

$$s_1 = 13 \text{ (01101)}, \quad s_2 = 24 \text{ (11000)}$$

$$s_3 = 8 \text{ (01000)}, \quad s_4 = 19 \text{ (10011)}$$

4. 遗传算法应用举例

(2) 定义适应度函数

取适应度函数： $f(x) = x^2$

(3) 计算各代种群中的各个体的适应度, 并对其染色体进行遗传操作, 直到适应度最高的个体 (即31 (1111))出现为止。

4. 遗传算法应用举例

首先计算种群 S_1 中各个体

$$s_1 = 13(01101), \quad s_2 = 24(11000)$$

$$s_3 = 8(01000), \quad s_4 = 19(10011)$$

的适应度 $f(s_i)$ 。

容易求得

$$f(s_1) = f(13) = 13^2 = 169$$

$$f(s_2) = f(24) = 24^2 = 576$$

$$f(s_3) = f(8) = 8^2 = 64$$

$$f(s_4) = f(19) = 19^2 = 361$$

4. 遗传算法应用举例

再计算种群 S_1 中各个体的选择概率

选择概率的计算公式为

$$P(x_i) = \frac{f(x_i)}{\sum_{j=1}^N f(x_j)}$$

由此可求得

$$P(s_1) = P(13) = 0.14$$

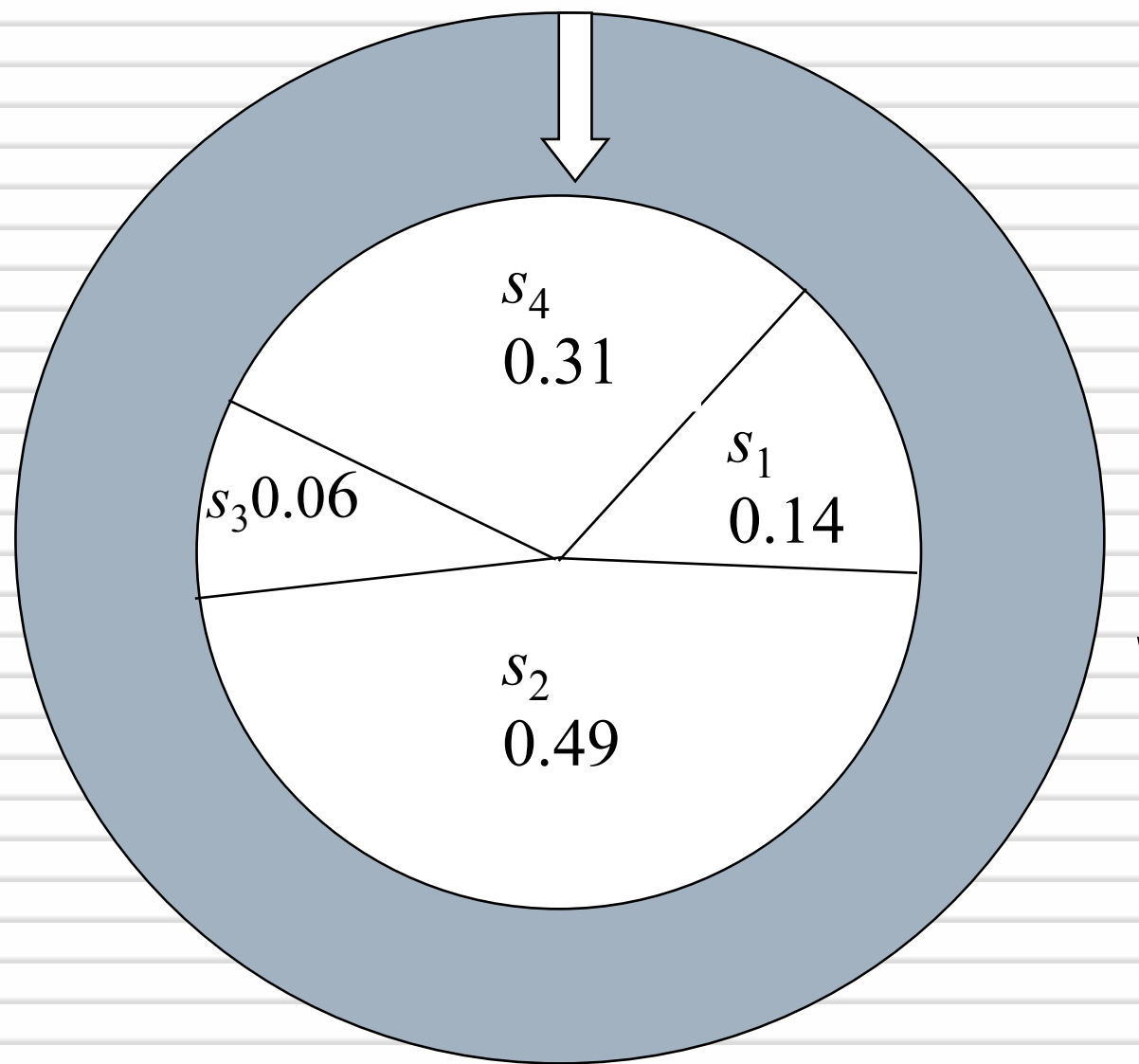
$$P(s_2) = P(24) = 0.49$$

$$P(s_3) = P(8) = 0.06$$

$$P(s_4) = P(19) = 0.31$$

4. 遗传算法应用举例

❑ 赌轮选择法



赌轮选择示意

4. 遗传算法应用举例

在算法中，赌轮选择法可用下面的子过程模拟：

□ ① 在 $[0, 1]$ 区间内产生一个均匀分布的随机数 r 。

② 若 $r \leq q_1$, 则染色体 x_1 被选中。

③ 若 $q_{k-1} < r \leq q_k (2 \leq k \leq N)$, 则染色体 x_k 被选中。其中的 q_i 称为染色体 $x_i (i=1, 2, \dots, n)$ 的**积累概率**, 其计算公式为

$$q_i = \sum_{j=1}^i P(x_j)$$

4. 遗传算法应用举例

选择-复制

设从区间 $[0, 1]$ 中产生4个随机数如下:

$$r_1 = 0.450126, \quad r_2 = 0.110347$$

$$r_3 = 0.572496, \quad r_4 = 0.98503$$

| 染色体 | 适应度 | 选择概率 | 积累概率 | 选中次数 |
|-------------|-----|------|------|------|
| $s_1=01101$ | 169 | 0.14 | 0.14 | 1 |
| $s_2=11000$ | 576 | 0.49 | 0.63 | 2 |
| $s_3=01000$ | 64 | 0.06 | 0.69 | 0 |
| $s_4=10011$ | 361 | 0.31 | 1.00 | 1 |

4. 遗传算法应用举例

于是，经复制得群体：

$$s_1' = 11000 \quad (24) , \quad s_2' = 01101 \quad (13)$$

$$s_3' = 11000 \quad (24) , \quad s_4' = 10011 \quad (19)$$

4. 遗传算法应用举例

交叉

设交叉率 $p_c=100\%$ ，即 S_1 中的全体染色体都参加交叉运算。

设 s_1' 与 s_2' 配对， s_3' 与 s_4' 配对。分别交换后两位基因，得新染色体：

$$s_1''=11001 \quad (25), \quad s_2''=01100 \quad (12)$$

$$s_3''=11011 \quad (27), \quad s_4''=10000 \quad (16)$$

4. 遗传算法应用举例

变异

设变异率 $p_m=0.001$ 。

这样，群体 S_1 中共有

$$5 \times 4 \times 0.001 = 0.02$$

位基因可以变异。

0.02位显然不足1位，所以本轮遗传操作不做变异。

4. 遗传算法应用举例

于是，得到第二代种群 S_2 ：

$$s_1=11001 \quad (25) , s_2=01100 \quad (12)$$

$$s_3=11011 \quad (27) , s_4=10000 \quad (16)$$

4. 遗传算法应用举例

第二代种群 S_2 中各染色体的情况

| 染色体 | 适应度 | 选择概率 | 积累概率 | 估计的选中次数 |
|-------------|-----|------|------|---------|
| $s_1=11001$ | 625 | 0.36 | 0.36 | 1 |
| $s_2=01100$ | 144 | 0.08 | 0.44 | 1 |
| $s_3=11011$ | 729 | 0.41 | 0.85 | 1 |
| $s_4=10000$ | 256 | 0.15 | 1.00 | 1 |

4. 遗传算法应用举例

假设这一轮选择-复制操作中，种群 S_2 中的
4个染色体都被选中，则得到群体：

$$s_1' = 11001 \quad (25), \quad s_2' = 01100 \quad (12)$$

$$s_3' = 11011 \quad (27), \quad s_4' = 10000 \quad (16)$$

做交叉运算，让 s_1' 与 s_2' ， s_3' 与 s_4' 分别交换
后三位基因，得

$$s_1'' = 11100 \quad (28), \quad s_2'' = 01001 \quad (9)$$

$$s_3'' = 11000 \quad (24), \quad s_4'' = 10011 \quad (19)$$

这一轮仍然不会发生变异。

4. 遗传算法应用举例

于是，得第三代种群 S_3 ：

$$s_1=11100 \quad (28) , s_2=01001 \quad (9)$$

$$s_3=11000 \quad (24) , s_4=10011 \quad (19)$$

4. 遗传算法应用举例

第三代种群 S_3 中各染色体的情况

| 染色体 | 适应度 | 选择概率 | 积累概率 | 估计的选中次数 |
|-------------|-----|------|------|---------|
| $s_1=11100$ | 784 | 0.44 | 0.44 | 2 |
| $s_2=01001$ | 81 | 0.04 | 0.48 | 0 |
| $s_3=11000$ | 576 | 0.32 | 0.80 | 1 |
| $s_4=10011$ | 361 | 0.20 | 1.00 | 1 |

4. 遗传算法应用举例

设这一轮的选择-复制结果为：

$$s_1' = 11100 \quad (28), \quad s_2' = 11100 \quad (28)$$

$$s_3' = 11000 \quad (24), \quad s_4' = 10011 \quad (19)$$

做交叉运算，让 s_1' 与 s_4' ， s_2' 与 s_3' 分别交换后两位基因，得

$$s_1'' = 11111 \quad (31), \quad s_2'' = 11100 \quad (28)$$

$$s_3'' = 11000 \quad (24), \quad s_4'' = 10000 \quad (16)$$

这一轮仍然不会发生变异。

4. 遗传算法应用举例

于是，得第四代种群 S_4 ：

$$s_1=11111 \quad (31), \quad s_2=11100 \quad (28)$$

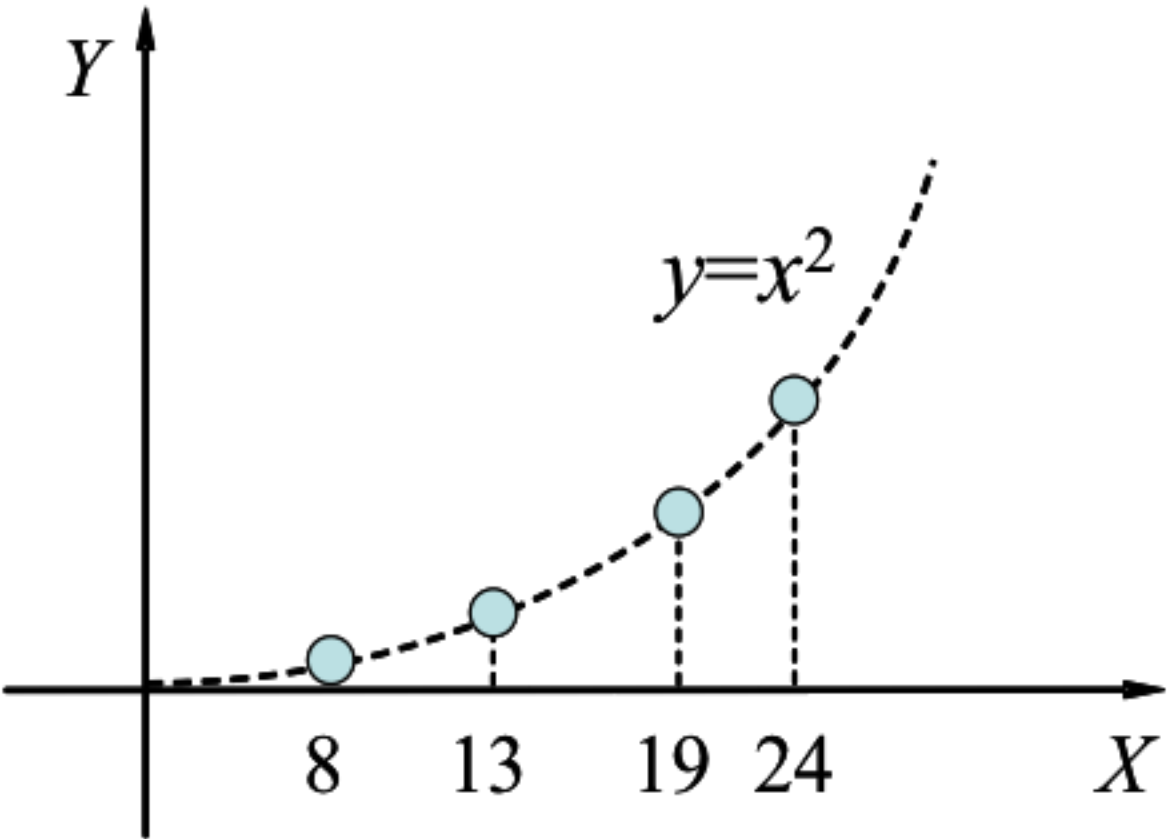
$$s_3=11000 \quad (24), \quad s_4=10000 \quad (16)$$

4. 遗传算法应用举例

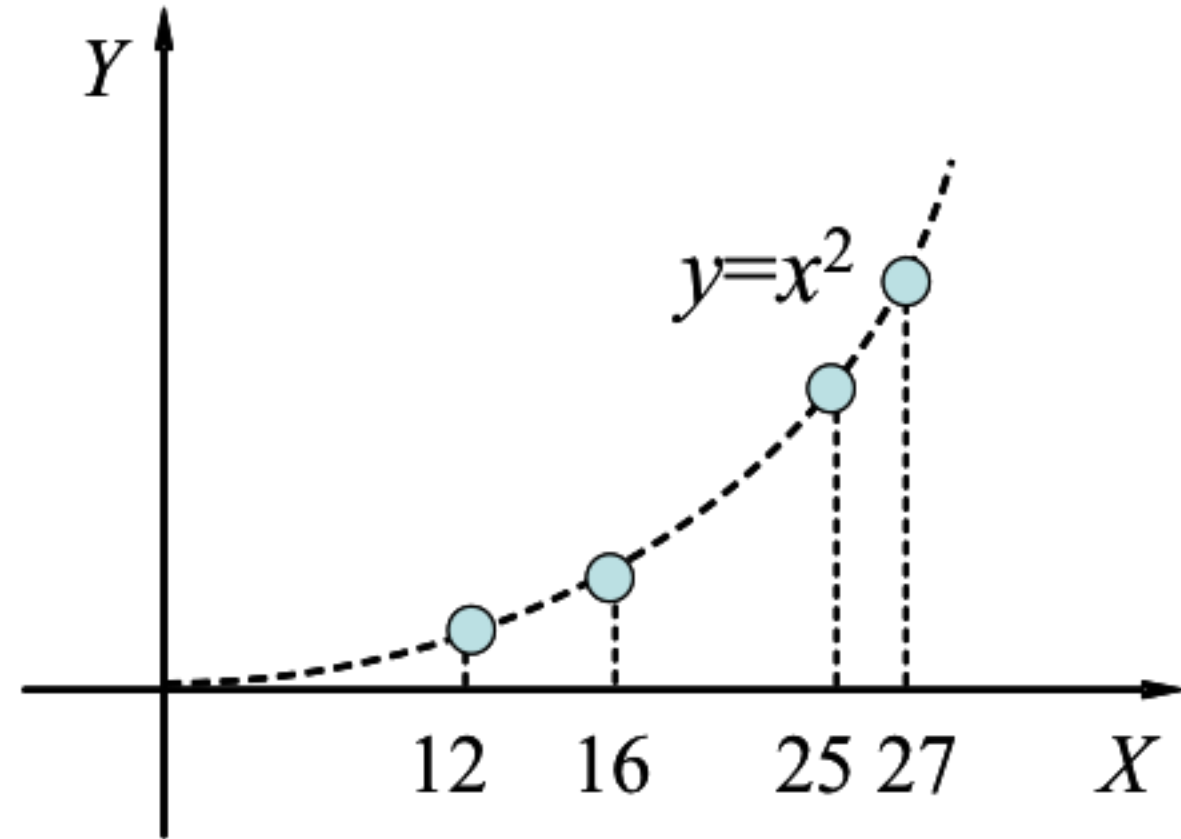
显然，在这一代种群中已经出现了适应度最高的染色体 $s_1=11111$ 。于是，遗传操作终止，将染色体“11111”作为最终结果输出。

然后，将染色体“11111”解码为表现型，即得所求的最优解：31。

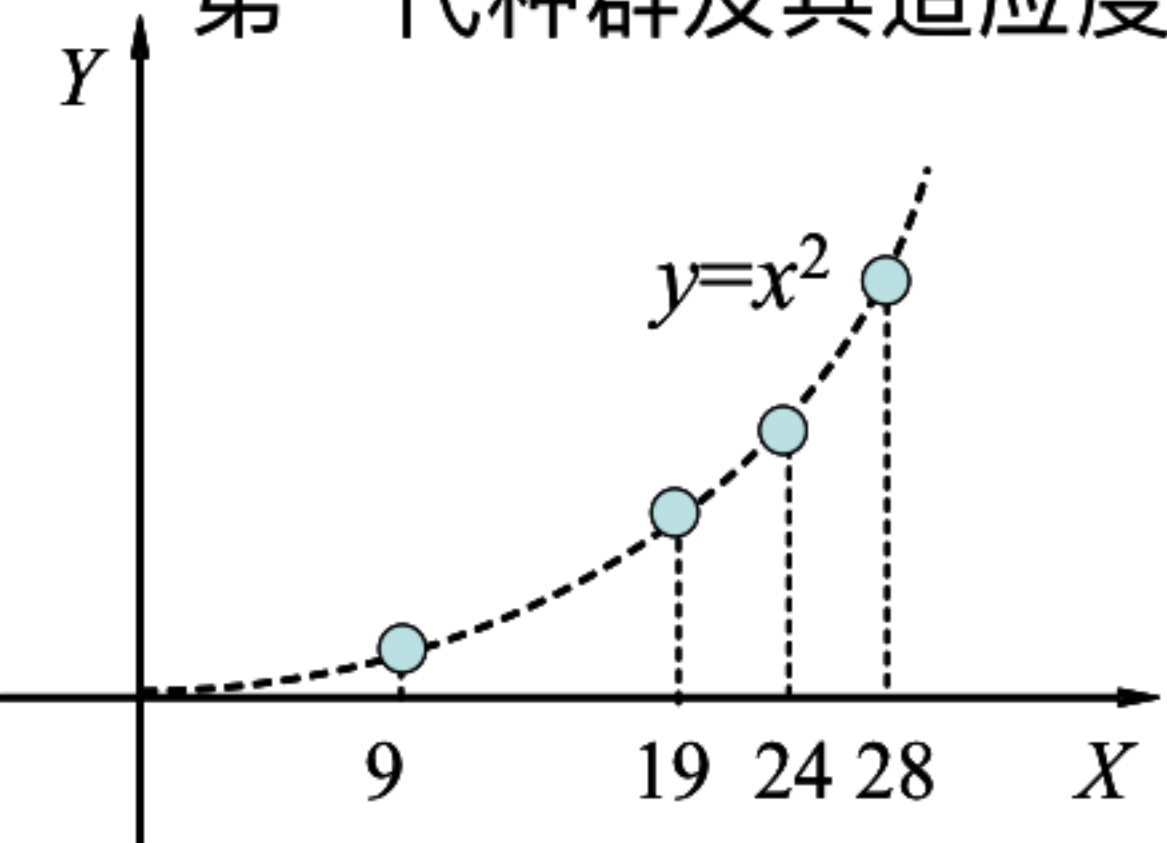
将31代入函数 $y=x^2$ 中，即得原问题的解，即函数 $y=x^2$ 的最大值为961。



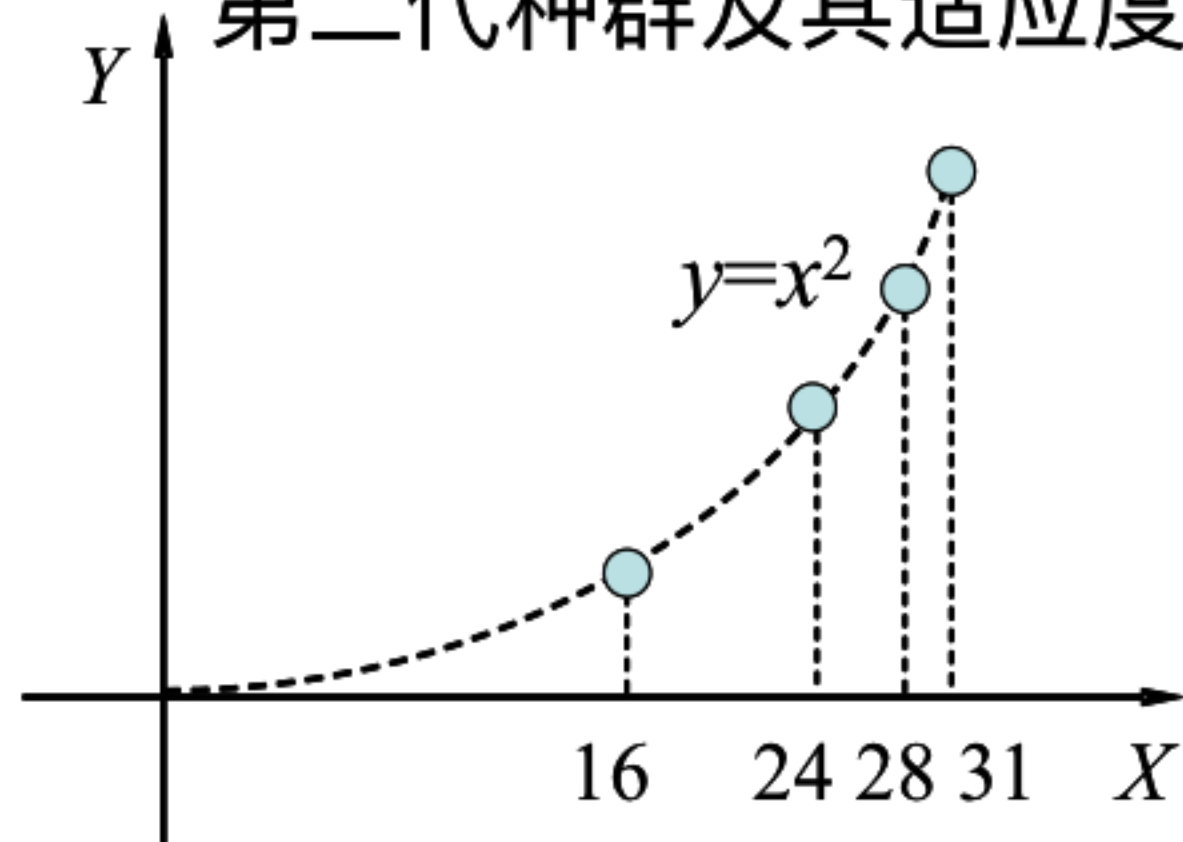
第一代种群及其适应度



第二代种群及其适应度



第三代种群及其适应度



第四代种群及其适应度

小 结

- 遗传算法概述
- 遗传算法的基本操作
- 简单遗传算法

练习 1

- 利用遗传算法，求解区间 $[1,16]$ 上的函数 $y = 2x$ 的最大值。假设初始种群为 $(2, 5, 8, 15)$ ，随机数假设为 $(0.1, 0.3, 0.5, 0.8)$ ，要求写出具体步骤。

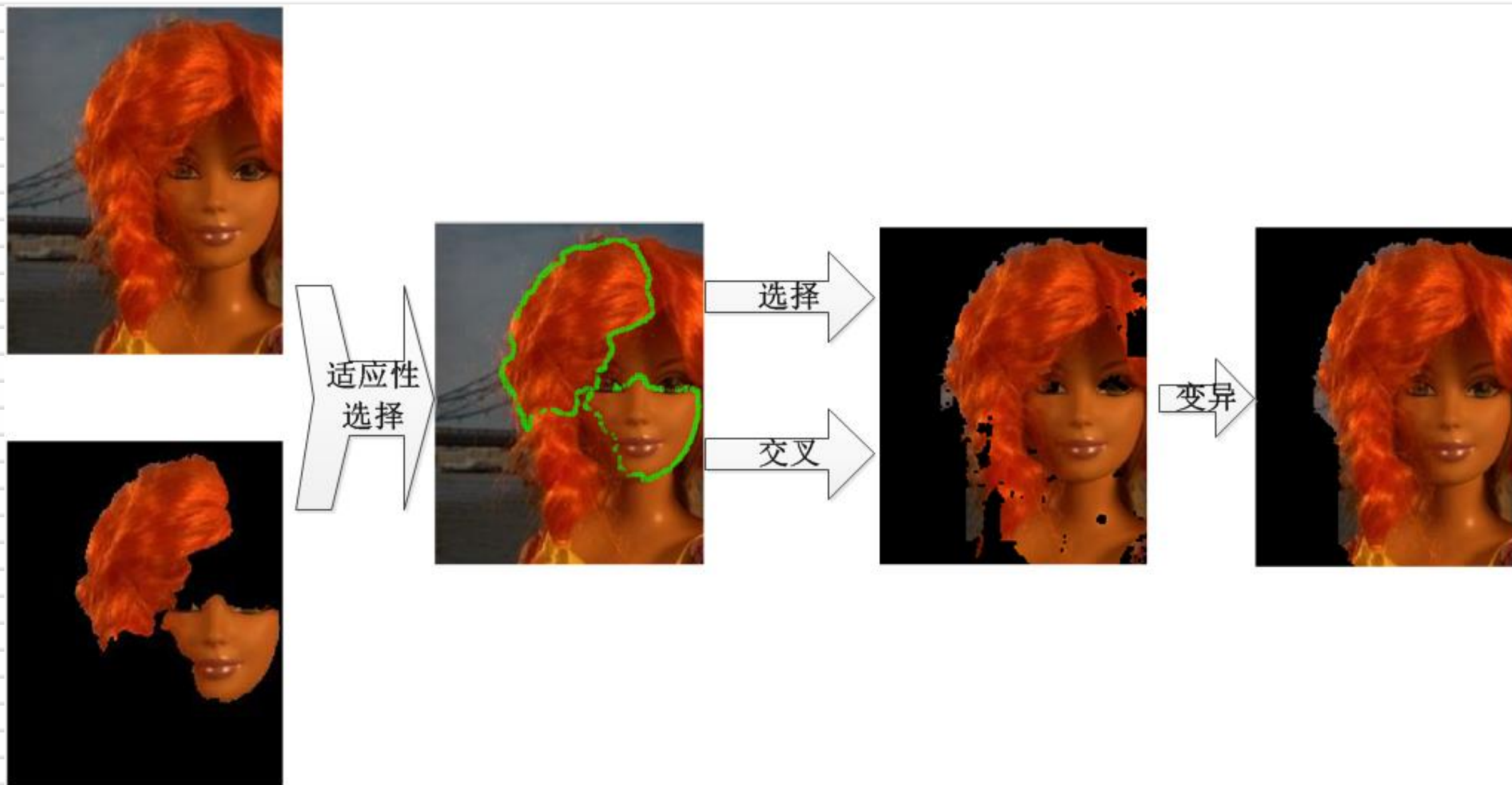
实例简介：基于遗传算法的图像区域生成技术

- 利用遗传算法处理图像区域增长的问题
 - 通过选择种子点，交叉和变异等遗传过程来
 - 再生出图像中的前景区域

研究问题



技术路线



算法伪代码

| 步骤 | 操作 |
|----|---|
| 1. | 输入：原始图像 I ，初始图像前景区域 $I_foreground$ ； |
| 2. | 在前景区域内选择父节点 Fa 和母节点 Ma ； |
| 3. | 根据遗传机制获得子节点 $Child$ ； |
| 4. | 通过比较 $Child$ 节点和 Fa 八邻域内的非种群内的像素点的相似性改变其 $GA_{(x,y)}$ 的值； |
| 5. | 完成像素点的变异； |
| 6. | 判断遗传终止条件是否成立，如果成立则执行第 7 步，如果不成立则执行第 2 步； |
| 7. | 输出：区域生成后的图像的前景区域 $I_foreground$ 。 |

部分实验结果展示

