

第 5 章 模拟生物进化的遗传算法

教材：

王万良 《人工智能应用教程》（2版）

清华大学出版社，2025



第5章 模拟生物进化的遗传算法

- 受自然界和生物界规律的启迪，人们根据其原理模仿设计了许多求解问题的算法，包括人工神经网络、模糊逻辑、遗传算法、粒子群优化算法、蚁群算法等，这些算法称为智能计算也称为计算智能(computational intelligence, CI)。
- 智能优化方法通常包括进化计算和群智能等两大类方法，已经广泛应用于组合优化、机器学习、智能控制、模式识别、规划设计、网络安全等领域，是21世纪有关智能计算中的重要技术之一。
- 本章首先简要介绍进化算法的概念，详细介绍基本遗传算法，这是进化算法的基本框架。然后介绍双倍体、双种群、自适应等比较典型的改进遗传算法及其应用。

第 5 章 模拟生物进化的遗传算法

- 5.1 进化算法的生物学背景
- 5.2 遗传算法
- 5.3 遗传算法的主要改进算法
- 5.4 基于遗传算法的生产调度方法

第 5 章 模拟生物进化的遗传算法

✓ 5.1 进化算法的生物学背景

□ 5.2 遗传算法

□ 5.3 遗传算法的主要改进算法

□ 5.4 基于遗传算法的生产调度方法

5.1 进化算法的生物学背景

- 进化算法(evolutionary algorithms, EA)是基于自然选择和自然遗传等生物进化机制的一种搜索算法。
- 生物进化是通过繁殖、变异、竞争和选择实现的；而进化算法则主要通过选择、重组和变异这三种操作实现优化问题的求解。
- 进化算法是一个“算法簇”，包括遗传算法(GA)、遗传规划、进化策略和进化规划等。
- 进化算法的基本框架是遗传算法所描述的框架。

5.1 进化算法的生物学背景

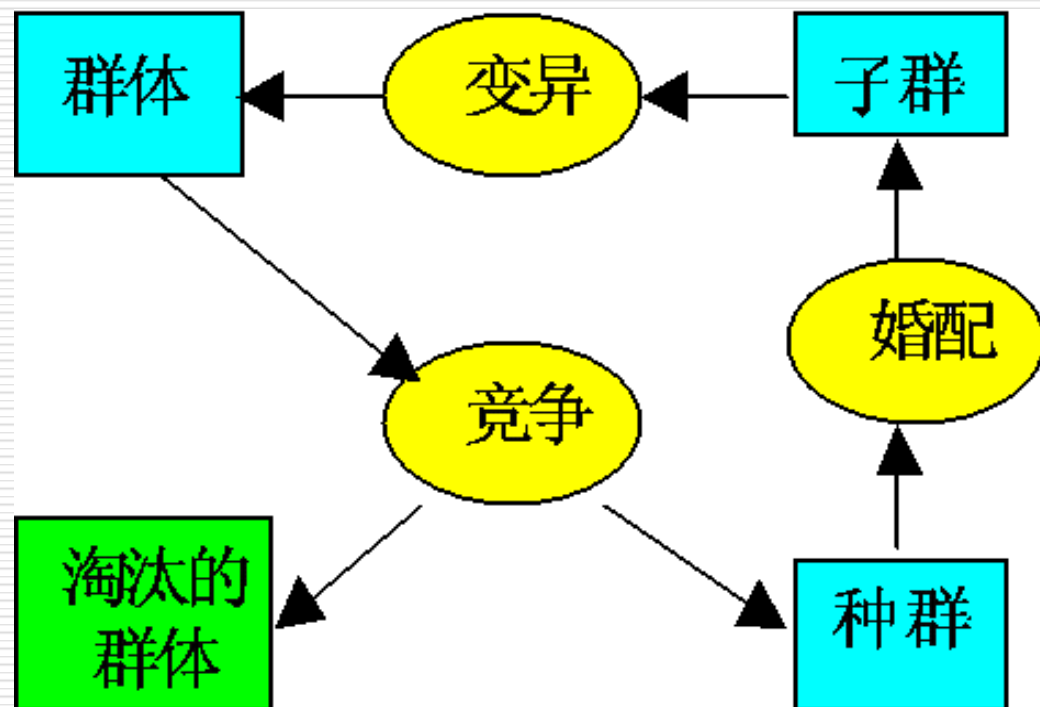
- **适者生存**：最适合自然环境的群体往往产生了更大的后代群体。
- **生物进化的基本过程**：

染色体(chromosome)：生物的遗传物质的主要载体。

基因(gene)：扩展生物性状的遗传物质的功能单元和结构单位。

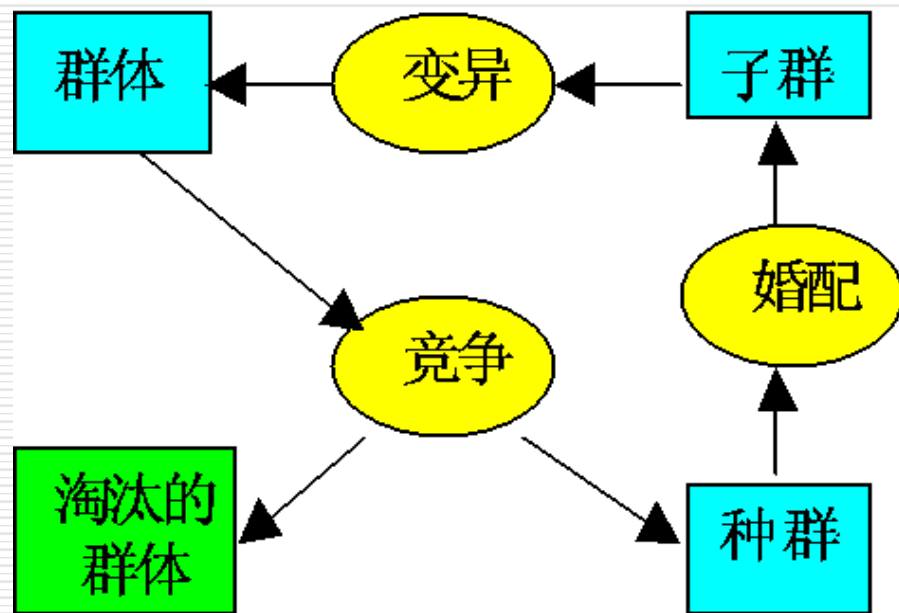
基因座 (locus)：染色体中基因的位置。

等位基因 (alleles)：基因所取的值。



5.1 进化算法的生物学背景

□ 生物进化的基本过程:



第 5 章 模拟生物进化的遗传算法

- 5.1 进化算法的生物学背景
- 5.2 遗传算法
- 5.3 遗传算法的主要改进算法
- 5.4 基于遗传算法的生产调度方法

5.2 遗传算法

- 遗传算法（genetic algorithms, GA）：一类借鉴生物界自然选择和自然遗传机制的随机搜索算法。
- 遗传算法非常适用于处理传统搜索方法难以解决的复杂和非线性优化问题。
- 遗传算法可广泛应用于组合优化、机器学习、自适应控制、规划设计和人工生命等领域。

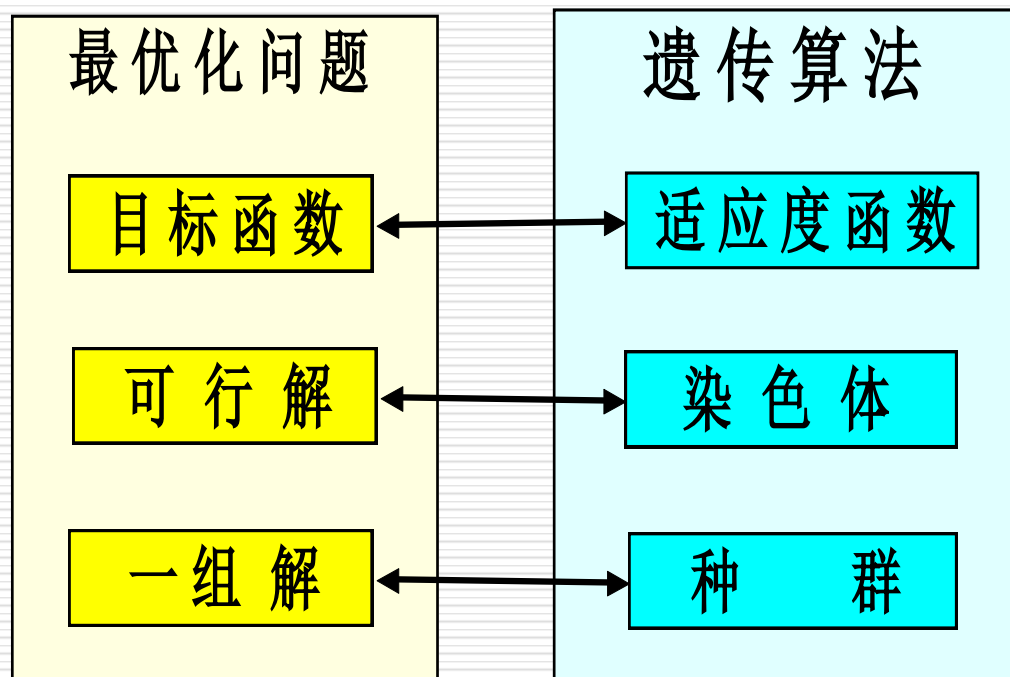
5.2.1 遗传算法的发展历史

- ❑ 1962年，Fraser提出了自然遗传算法。
- ❑ 1965年，Holland首次提出了人工遗传操作的重要性。
- ❑ 1967年，Bagley首次提出了遗传算法这一术语。
- ❑ 1970年，Cavicchio把遗传算法应用于模式识别中。
- ❑ 1971年，Hollstien在论文《计算机控制系统中人工遗传自适应方法》中阐述了遗传算法用于数字反馈控制的方法。
- ❑ 1975年，美国J. Holland出版了《自然系统和人工系统的适配》；DeJong完成了重要论文《遗传自适应系统的行为分析》。
- ❑ 20世纪80年代以后，遗传算法进入兴盛发展时期。

5.2.2 遗传算法的基本思想

生物遗传概念	遗传算法中的概念
适者生存	目标值比较大的解被选择的可能性大
个体 (Individual)	问题的解
染色体 (Chromosome)	解的编码 (字符串、向量等)
基因 (Gene)	解的编码中的每一分量
适应性 (Fitness)	适应度函数值
群体 (Population)	一组解 (解的个数为群体的规模)
婚配 (Marry)	交叉 (Crossover) 选择两个染色体进行交叉产生一组新的染色体的过程
变异 (Mutation)	染色体的某一分量发生变化的过程

5.2.2 遗传算法的基本思想



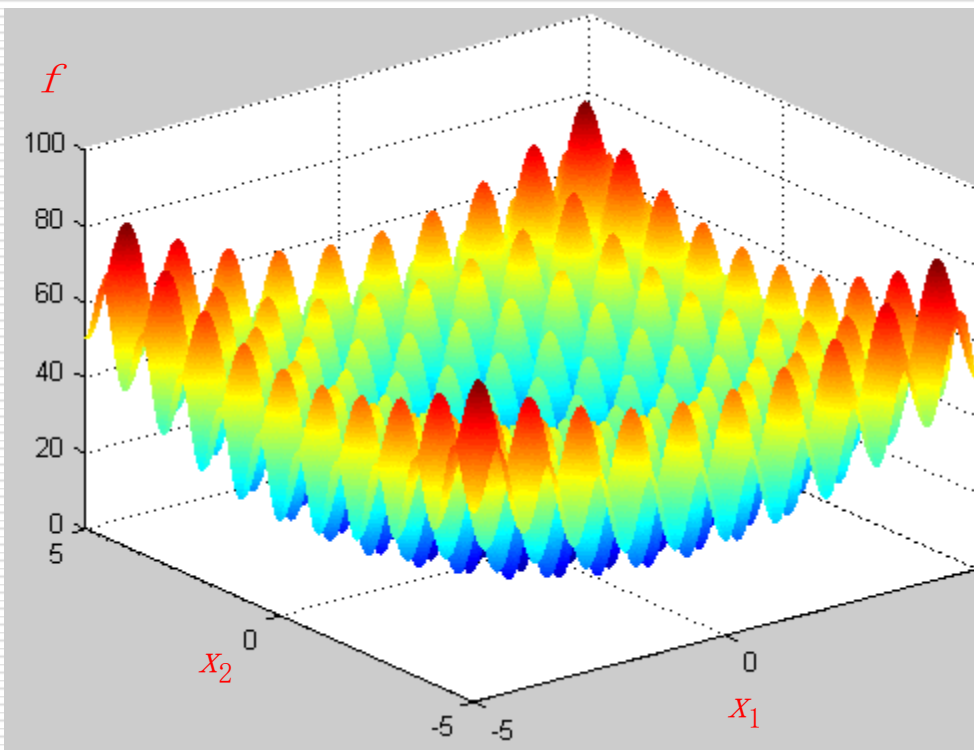
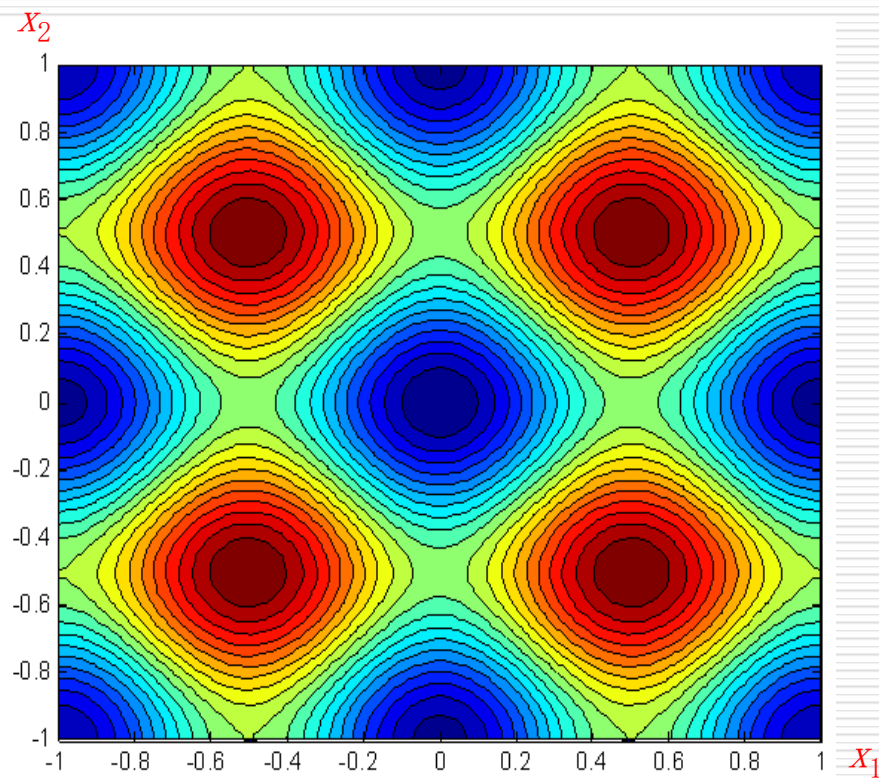
■ 遗传算法的基本思想:

在求解问题时从多个解开始，然后通过一定的法则进行逐步迭代以产生新的解。

遗传算法搜索过程演示

□ 例： 用遗传算法求解下面Rastrigin函数的最小值。

$$f(x_1, x_2) = 20 + x_1^2 + x_2^2 - 10(\cos 2\pi x_1 + \cos 2\pi x_2)$$
$$-5 \leq x_i \leq 5 \quad i = 1, 2$$

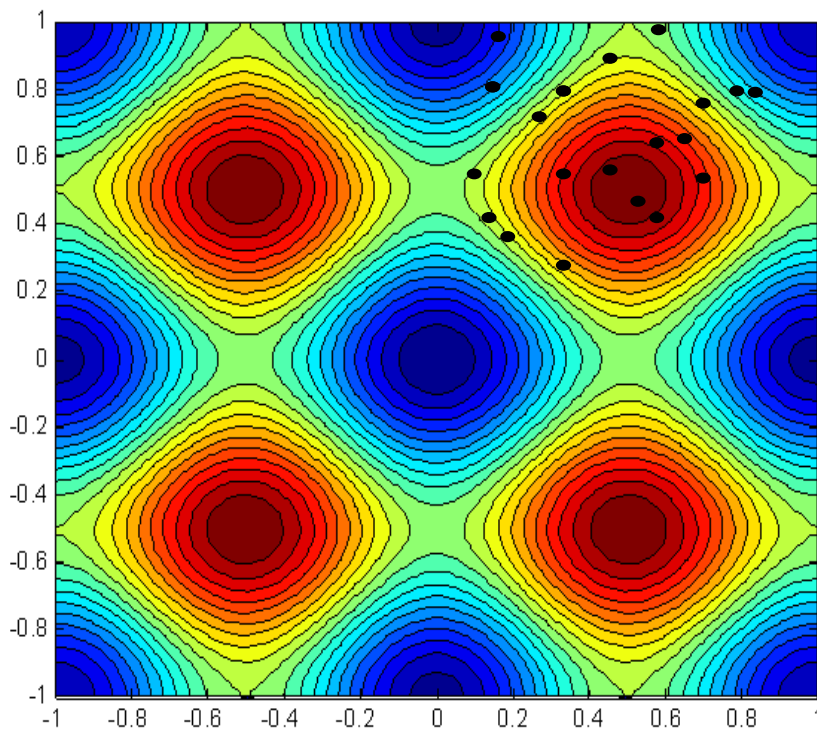


遗传算法搜索过程演示

□ 例： 用遗传算法求解下面Rastrigin函数的最小值。

$$f(x_1, x_2) = 20 + x_1^2 + x_2^2 - 10(\cos 2\pi x_1 + \cos 2\pi x_2)$$
$$-5 \leq x_i \leq 5 \quad i = 1, 2$$

□ 初始种群:



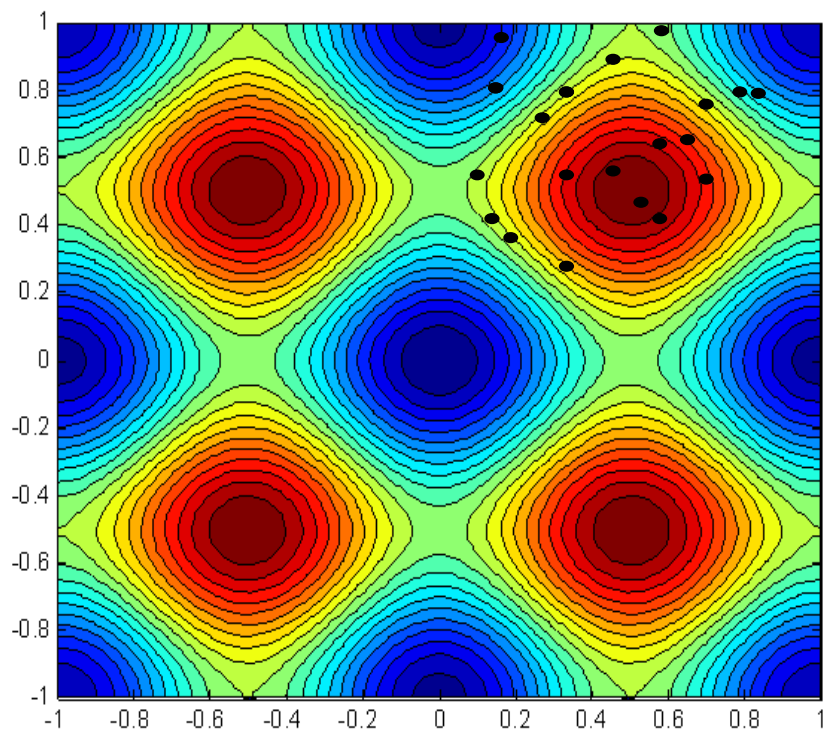
遗传算法搜索过程演示

□ 例： 用遗传算法求解下面Rastrigin函数的最小值。

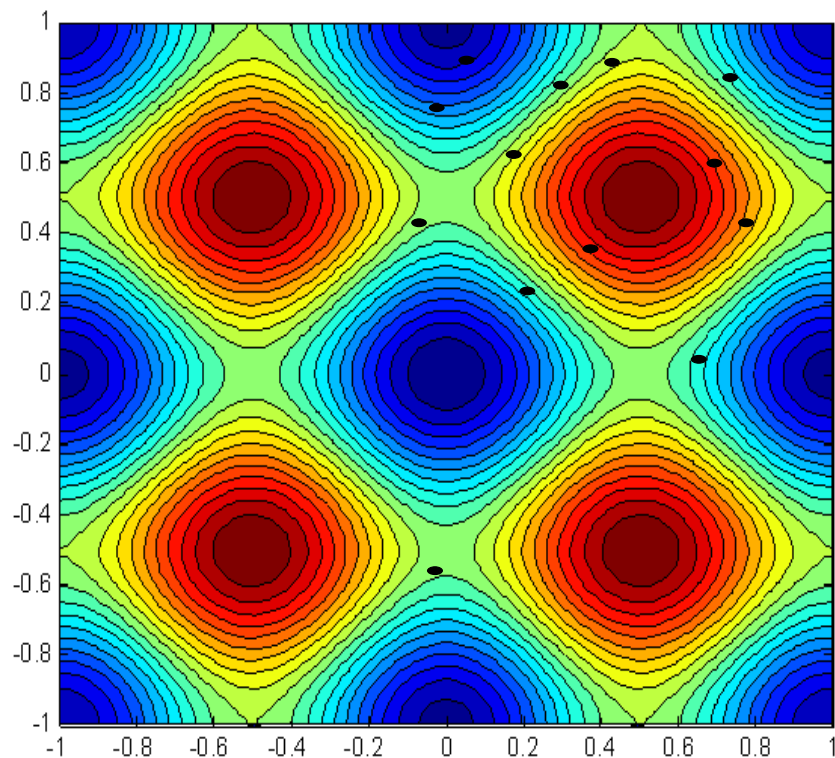
$$f(x_1, x_2) = 20 + x_1^2 + x_2^2 - 10(\cos 2\pi x_1 + \cos 2\pi x_2)$$

$$-5 \leq x_i \leq 5 \quad i = 1, 2$$

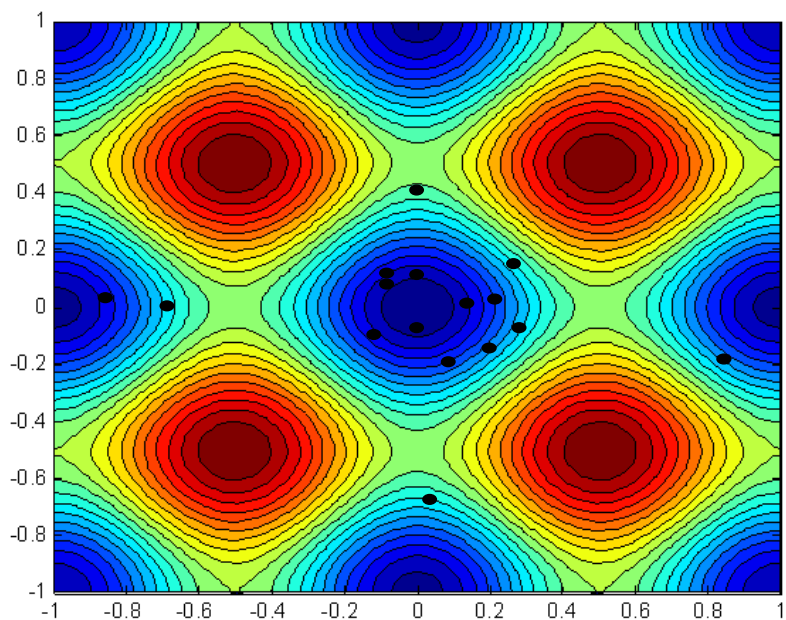
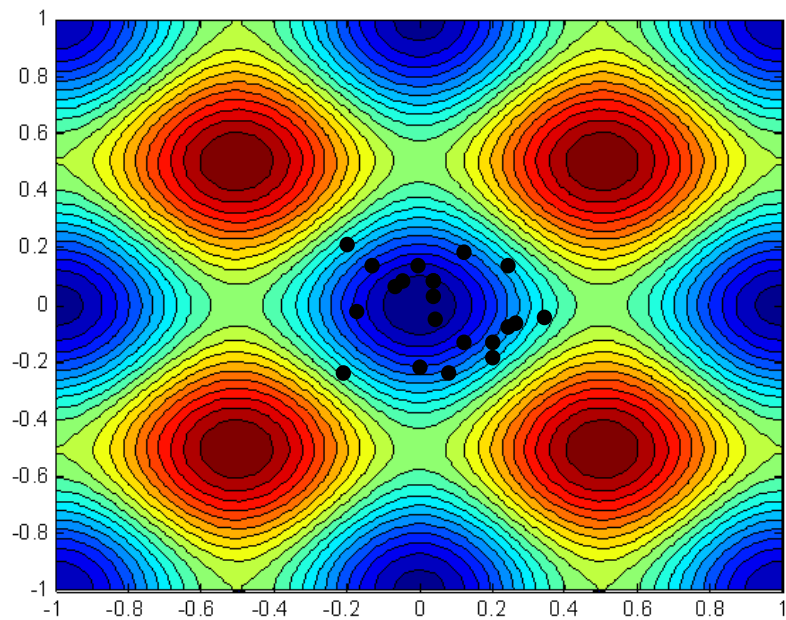
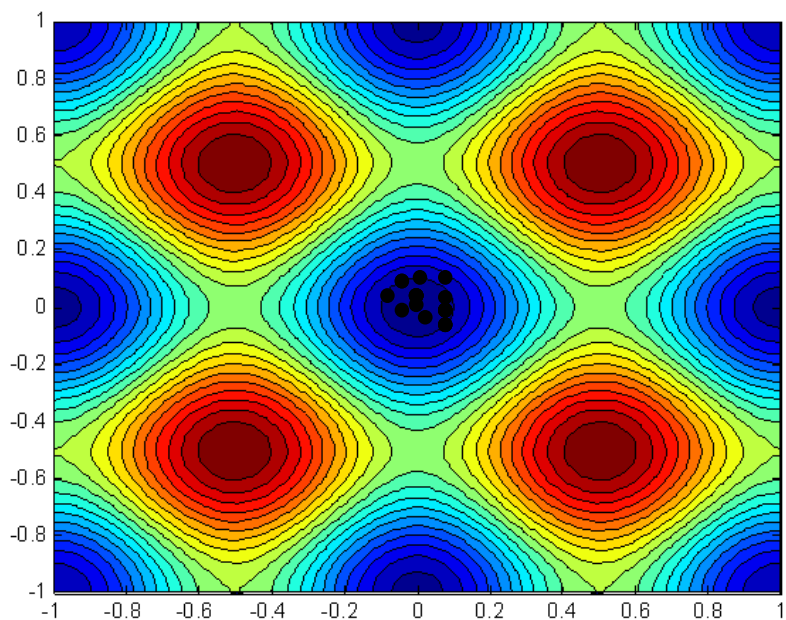
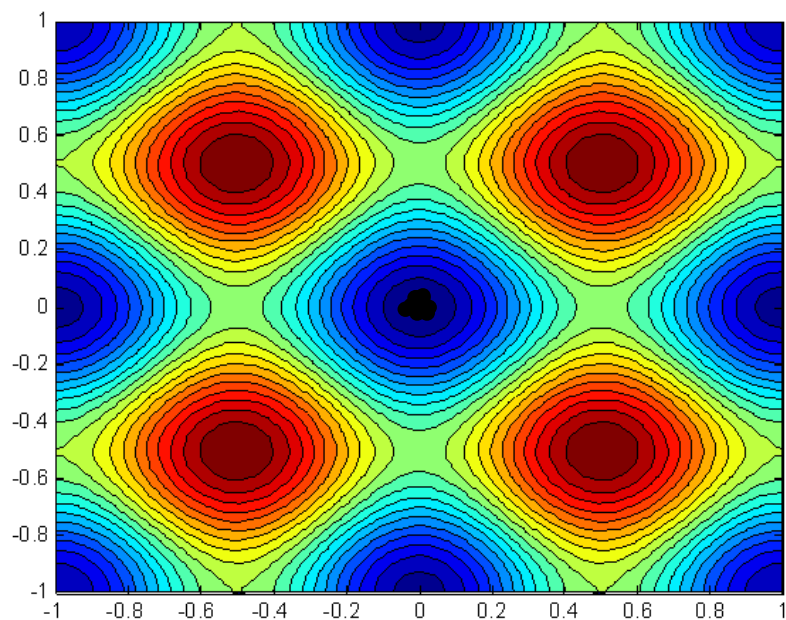
□ 初始种群



第二代种群



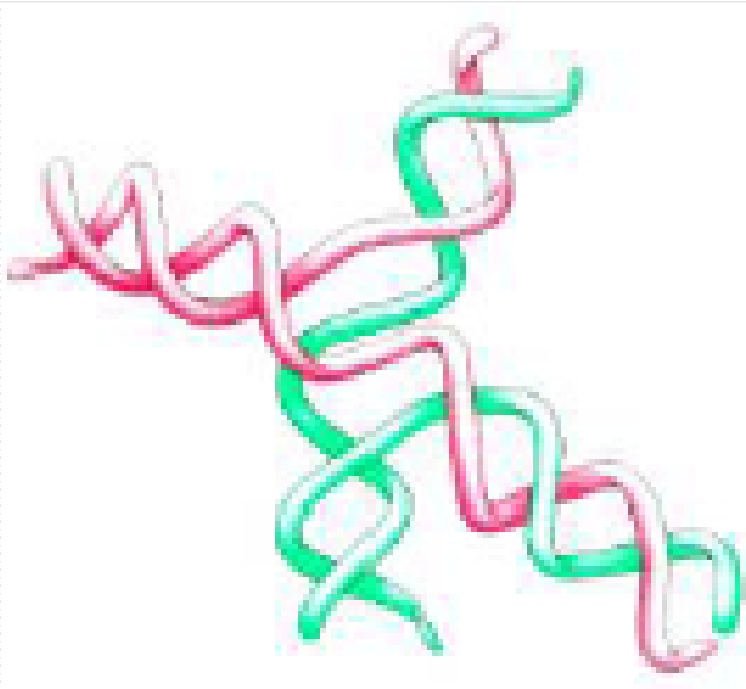
在迭代60、80、95、100次时的种群



5.2.3 编码

位串编码

一维染色体编码方法：将问题空间的参数编码为一维排列的染色体的方法。



5.2.3 编码

1. 二进制编码



二进制编码：用若干二进制数表示一个个体，将原问题的解空间映射到位串空间 $B=\{0, 1\}$ 上，然后在位串空间上进行遗传操作。

Channel ID: c1 c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12 c13 c14 c15 c16

Chromosome:

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

5.2.3 编码

1. 二进制编码（续）

● 优点：

类似于生物染色体的组成，算法易于用生物遗传理论解释，遗传操作如交叉、变异等易实现；算法处理的模式数最多。

● 缺点：

① 相邻整数的二进制编码可能具有较大的Hamming距离，降低了遗传算子的搜索效率。

Hamming距离是指两个数码串中对应的位的值不相同的数量。

15: 01111

16: 10000

② 要先给出求解的精度。

③ 求解高维优化问题的二进制编码串长，算法的搜索效率低。

5.2.3 编码

2. 实数编码

- 采用实数表达法不必进行数制转换，可直接在解的表现型上进行遗传操作。

3. 多参数级联编码

- **多参数映射编码的基本思想：**把每个参数先进行二进制编码得到子串，再把这些子串连成一个完整的染色体。
- 多参数映射编码中的每个子串对应各自的编码参数，所以，可以有不同的串长度和参数的取值范围。

5.2.4 种群设定

1. 初始种群的产生

- **基本方法：** 随机产生一定数目的个体。
- **改进方法：** 随机产生一定数目的个体，从中挑选最好的个体加到初始群体中。这种过程不断迭代，直到初始群体中个体数目达到了预先确定的规模。

5.2.4 群体设定

2. 种群规模的确定

- 群体规模太小，遗传算法的优化性能不太好，易陷入局部最优解。
- 群体规模太大，计算复杂。

思考题：

为什么动物群体规模越大越好？

而遗传算法中群体规模不能太大？

5.2.5 适应度函数

1. 将目标函数映射成适应度函数的方法

基本原则：

目标函数值好的个体的适应度函数值高。

适应度函数值只能是正值，越大越好。

● 若目标函数为最大化问题，则 $Fit(f(x)) = f(x)$

● 若目标函数为最小化问题，则 $Fit(f(x)) = \frac{1}{f(x)}$

5.2.5 适应度函数

学生考试成绩没有拉开，会有什么影响？如何处理？

在遗传算法中有什么表现？如何处理？

2. 适应度函数的尺度变换

- 在遗传算法中，将所有妨碍适应度值高的个体产生，从而影响遗传算法正常工作的问题统称为**欺骗问题**。
- **尺度变换（fitness scaling）**或**定标**：对适应度函数值域的某种映射变换。
- **变换的原则**：原来适应度值高的个体，变换后应该仍然高。

5.2.5 适应度函数

适应度函数的尺度变换常用方法

$$f' = af + b$$

满足 $f'_{avg} = f_{avg}$, $f'_{max} = C_{mult} \cdot f_{avg}$

$$a = \frac{f_{avg}}{f_{avg} - f_{min}}$$

$$b = \frac{-f_{min}f_{avg}}{f_{avg} - f_{min}}$$

5.2.6 选择

1. 个体选择概率分配方法

- 选择操作也称为复制（**reproduction**）操作：从当前群体中按照一定概率选出优良的个体，使它们有机会作为父代繁殖下一代子孙。
- 判断个体优良与否的准则是各个个体的适应度值：个体适应度越高，其被选择的机会就越多。

5.2.6 选择

1. 个体选择概率分配方法

(1) 适应度比例方法 (fitness proportional model) 或蒙特卡罗法 (Monte Carlo)

- 各个个体被选择的概率和其适应度值成比例。
- 个体 i 被选择的概率为：

$$p_{si} = \frac{f_i}{\sum_{i=1}^M f_i}$$

5.2.6 选择

1. 个体选择概率分配方法

(2) 排序方法 (rank-based model)

- 群体成员按适应值大小从好到坏依次排列: $x_1, x_2, \dots, x_N$
- 个体 x_i 分配选择概率 p_i

$$p_i = \frac{a - bi}{M(M + 1)}$$

- 按转盘式选择的方式选择父体

5.2.6 选择

2. 选择个体方法

基本原则：

适应度高的个体被选择的可能性大，但不一定被选择。

为什么？

(1) 轮盘赌选择

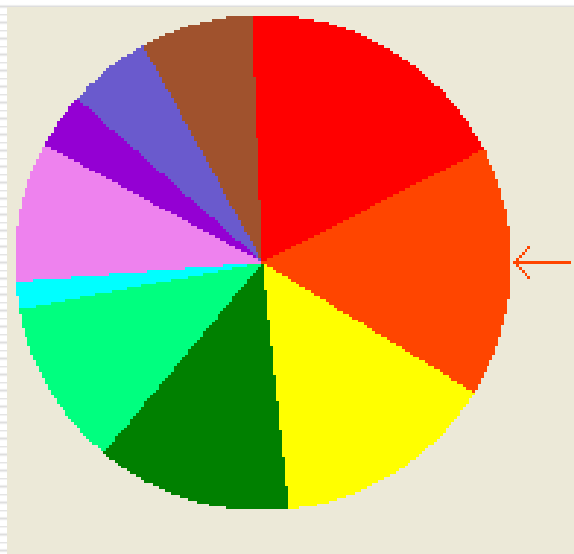
- 按个体的选择概率产生一个轮盘，轮盘每个区的角度与个体的选择概率成比例。
- 产生一个随机数，它落入转盘的哪个区域就选择相应的个体交叉。



5.2.6 选择

2. 选择个体方法

(1) 轮盘赌选择



个体	1	2	3	4	5	6	7	8	9	10	11
适应度	2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2	0.1
选择概率	0.18	0.16	0.15	0.13	0.11	0.09	0.07	0.06	0.03	0.02	0.0
累积概率	0.18	0.34	0.49	0.62	0.73	0.82	0.89	0.95	0.98	1.00	1.00

第1轮产生一个随机数: **0.81**

第2轮产生一个随机数: **0.32**

5.2.6 选择

2. 选择个体方法

(2) 锦标赛选择方法 (tournament selection model)

● 锦标赛选择方法：从群体中随机选择个个体，将其中适应度最高的个体保存到下一代。这一过程反复执行，直到保存到下一代的个体数达到预先设定的数量为止。

(3) 随机竞争方法 (stochastic tournament)

每次按赌轮选择方法选取一对个体，然后让这两个个体进行竞争，适应度高者获胜。如此反复，直到选满为止。

5.2.6 选择

2. 选择个体方法

体育比赛规则：上届冠军球队
直接进入决赛阶段比赛。



(4) 最佳个体保存方法

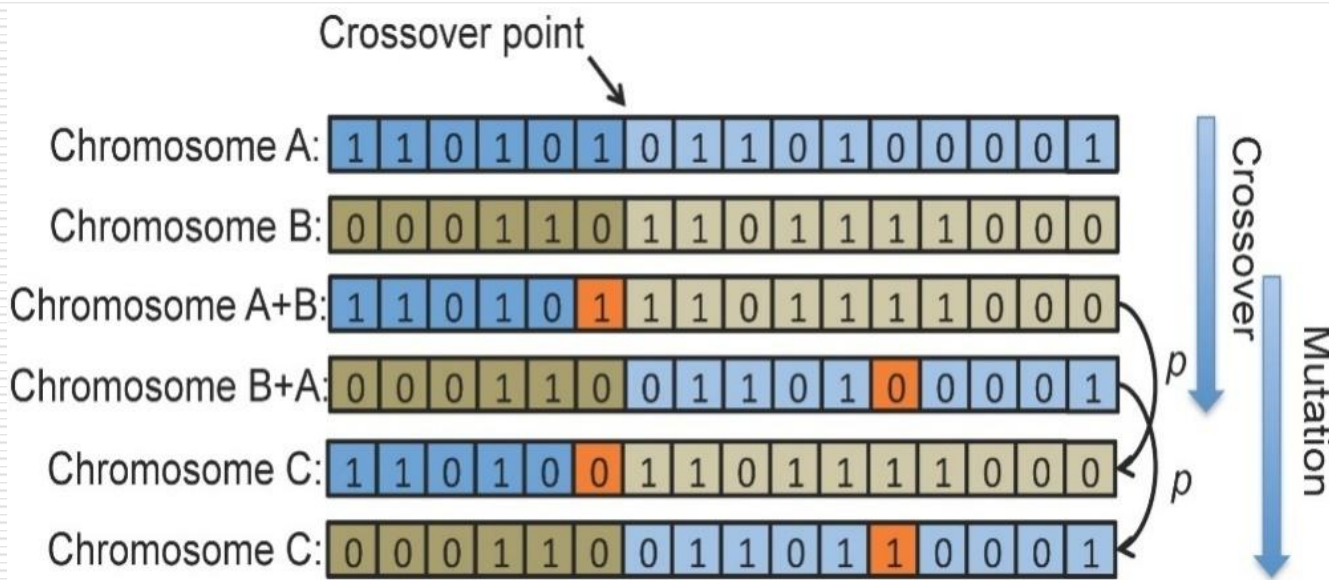
● 最佳个体（elitist model）保存方法：把群体中适应度最高的个体不进行交叉而直接复制到下一代中，保证遗传算法终止时得到的最后结果一定是历代出现过的最高适应度的个体。

5.2.7 交叉

1. 基本的交叉算子

(1) 一点交叉 (single-point crossover)

● 一点交叉：在个体串中随机设定一个交叉点，实行交叉时，该点前或后的两个个体的部分结构进行互换，并生成两个新的个体。

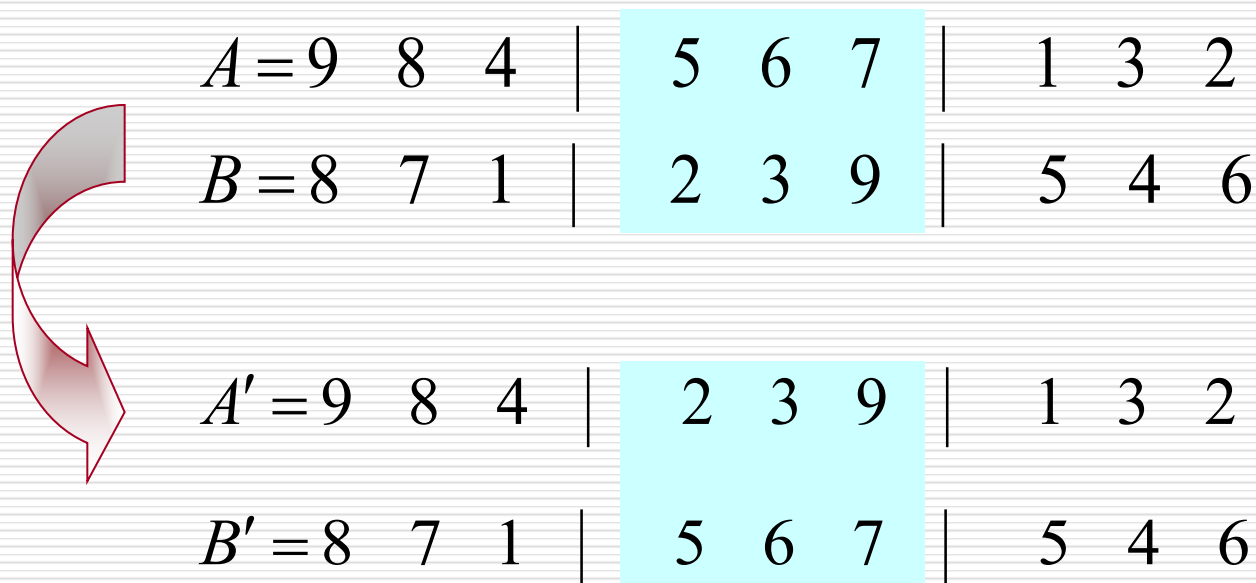


5.2.7 交叉

1. 基本的交叉算子

(2) 二点交叉 (two-point crossover)

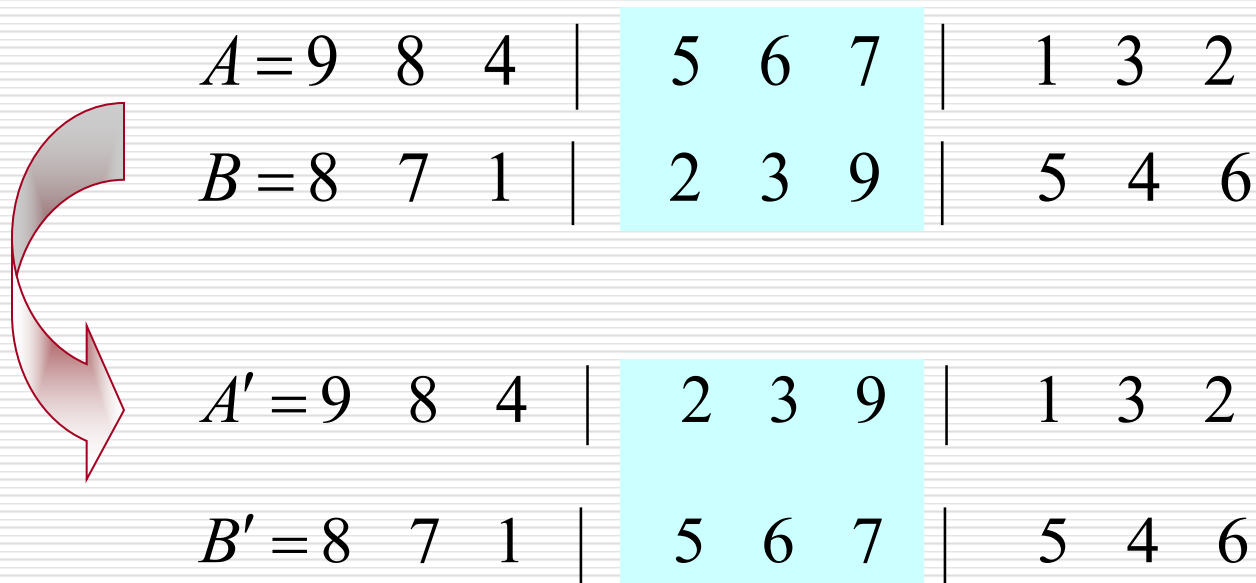
● 二点交叉：随机设置两个交叉点，将两个交叉点之间的码串相互交换。



5.2.7 交叉

2. 修正的交叉方法

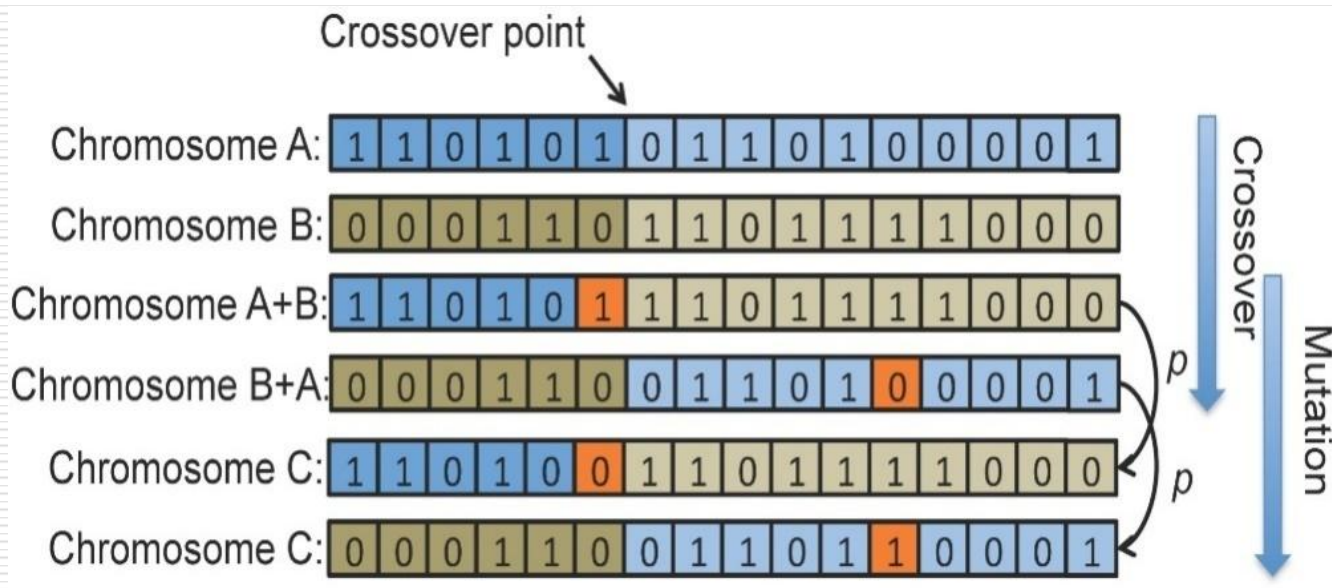
部分匹配交叉PMX: Goldberg D. E.和R. Lingle(1985)



5.2.8 变异

动物眼睛怎么形成的？

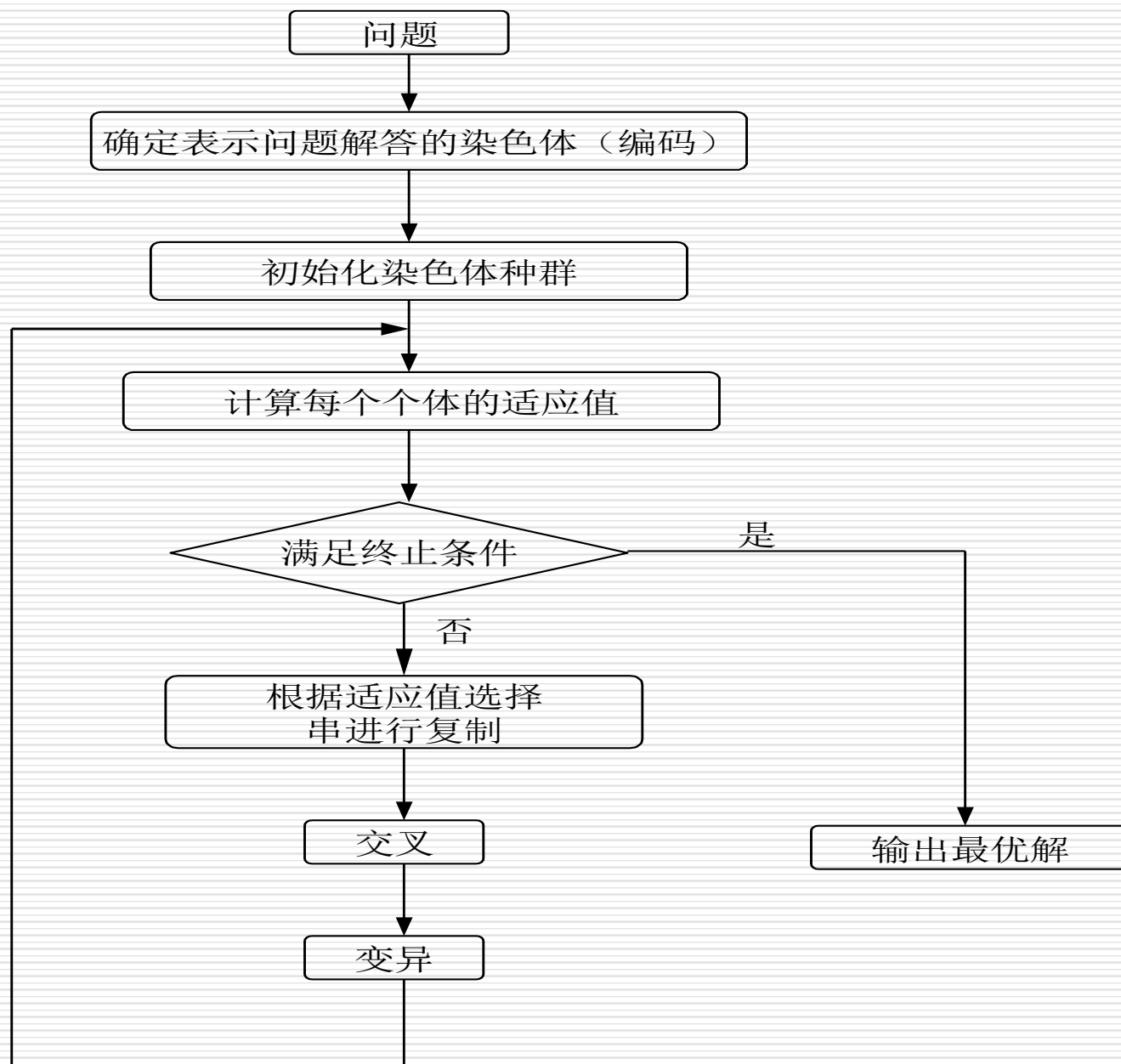
- **位点变异**：群体中的个体码串，随机挑选一个或多个基因座，并对这些基因座的基因值以变异概率作变动。



5.2.8 变异

- **逆转变异**：在个体码串中随机选择两点（逆转点），然后将两点之间的基因值以逆向排序插入到原位置中。
- **插入变异**：在个体码串中随机选择一个码，然后将此码插入随机选择的插入点中间。
- **互换变异**：随机选取染色体的两个基因进行简单互换。
- **移动变异**：随机选取一个基因，向左或者向右移动一个随机位数。

遗传算法的一般步骤



第 5 章 模拟生物进化的遗传算法

- 5.1 进化算法的生物学背景
- 5.2 遗传算法
- 5.3 遗传算法的主要改进算法
- 5.4 基于遗传算法的生产调度方法

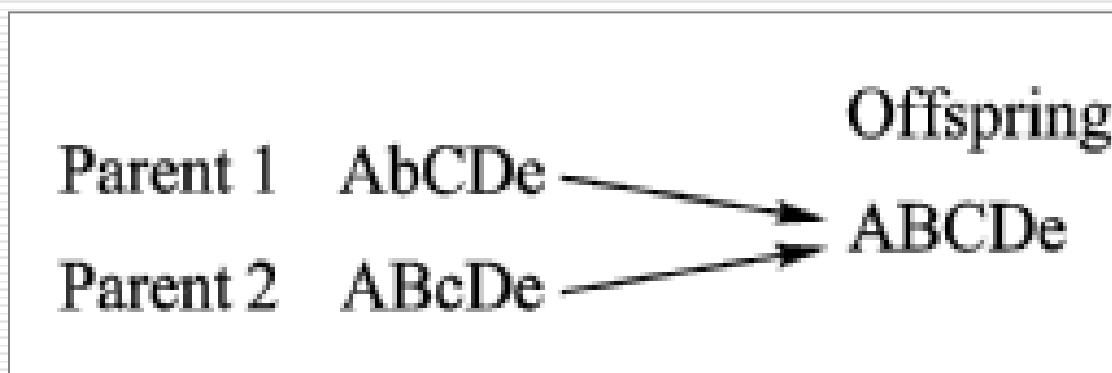
5.3 遗传算法的主要改进算法

- 5.3.1 双倍体遗传算法
- 5.3.2 双种群遗传算法
- 5.3.3 自适应遗传算法

5.3.1 双倍体遗传算法

1. 基本思想

- 双倍体遗传算法采用**显性**和**隐性**两个染色体同时进行进化，提供了一种记忆以前有用的基因块的功能。
- 双倍体遗传算法采用**显性遗传**。

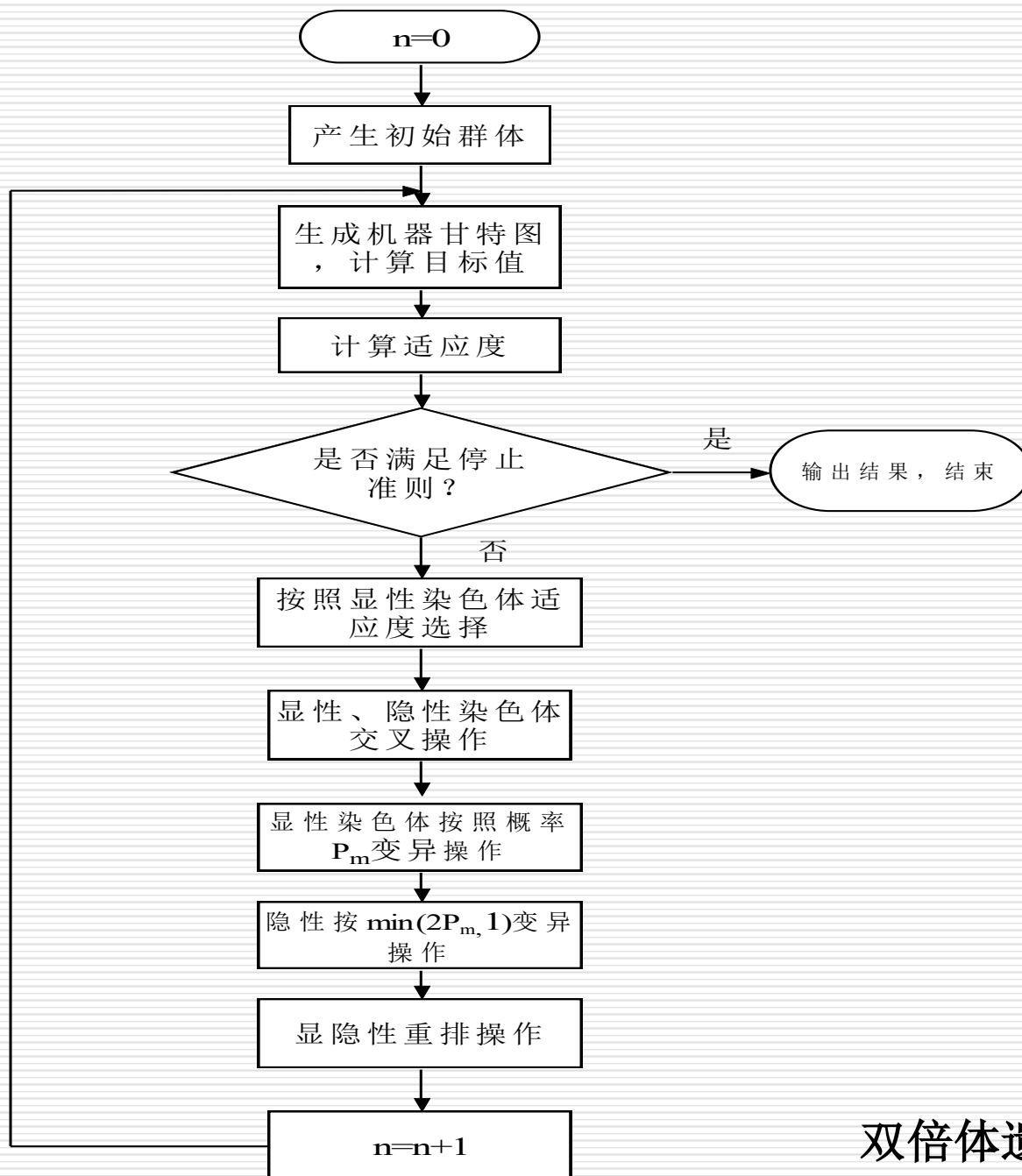


- 双倍体遗传延长了有用基因块的寿命，提高了算法的收敛能力，在变异概率低的情况下能保持一定水平的多样性。

5.3.1 双倍体遗传算法

2. 双倍体遗传算法的设计

- (1) **编码/解码**: 两个染色体（显性、隐性）
- (2) **复制算子**: 计算显性染色体的适应度，按照显性染色体的复制概率将个体复制到下一代群体中。
- (3) **交叉算子**: 两个个体的显性染色体交叉、隐性染色体也同时交叉。
- (4) **变异算子**: 个体的显性染色体按正常的变异概率变异；隐性染色体按较大的变异概率变异。
- (5) **双倍体遗传算法显隐性重排算子**: 个体中适应值较大的染色体设为显性染色体，适应值较小的染色体设为隐性染色体。



双倍体遗传算法程序流程图

5.3.2 双种群遗传算法

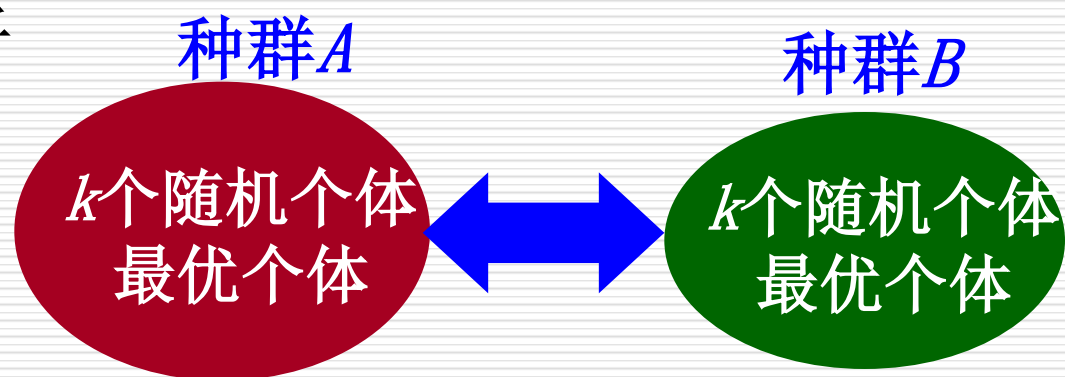
1. 基本思想

- 在遗传算法中使用多种群同时进化，并交换种群之间优秀个体所携带的遗传信息，以打破种群内的平衡态达到更高的平衡态，有利于算法跳出局部最优。
- **多种群遗传算法**：建立两个遗传算法群体，分别独立地运行复制、交叉、变异操作，同时当每一代运行结束以后，选择两个种群中的随机个体及最优个体分别交换。

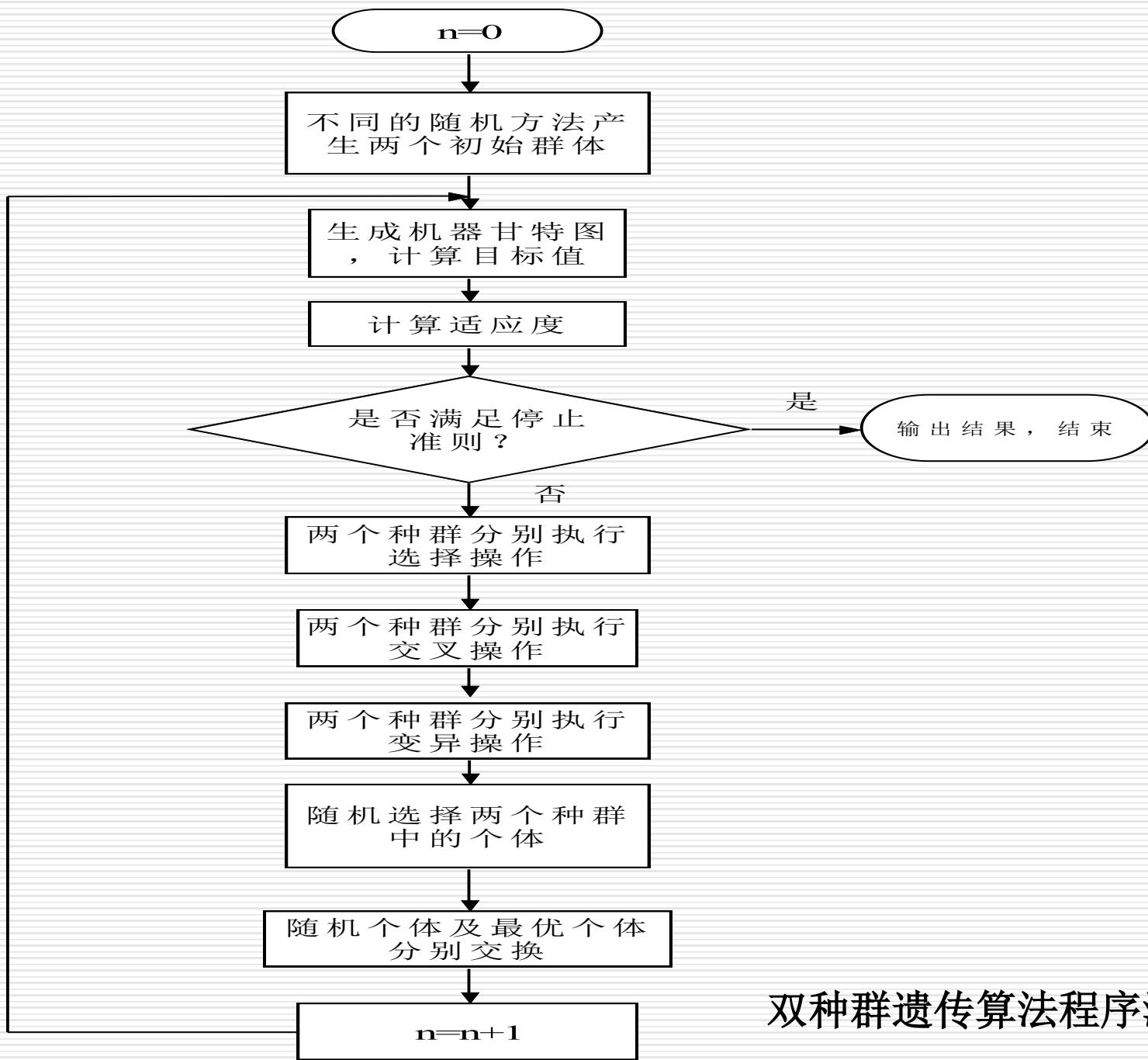
5.3.2 双种群遗传算法

2. 双种群遗传算法的设计

- 编码/解码设计
- 交叉算子、变异算子
- 杂交算子



设种群 A 与种群 B ，当 A 与 B 种群都完成了选择、交叉、变异算子后，产生一个随机数 num ，随机选择 A 中 num 个个体与 A 中最优个体，随机选择 B 中 num 个个体与 B 中最优个体，交换两者，以打破平衡态。



双种群遗传算法程序流程图

5.3.3 自适应遗传算法

● Srinivas M., Patnaik L. M.等在1994年提出一种自适应遗传算法(adaptive genetic algorithms, AGA): P_c 和 P_m 能随适应度自动改变。

● AGA: 当种群各个体适应度趋于一致或者趋于局部最优时, 使 P_c 和 P_m 增加, 以跳出局部最优; 而当群体适应度比较分散时, 使 P_c 和 P_m 减少, 以利于优良个体的生存。

● 同时, 对于适应度高于群体平均适应值的个体, 选择较低的 P_c 和 P_m , 使得该解得以保护进入下一代; 对低于平均适应值的个体, 选择较高的 P_c 和 P_m 值, 使该解被淘汰。

第 5 章 模拟生物进化的遗传算法

- 5.1 进化算法的生物学背景
- 5.2 遗传算法
- 5.3 遗传算法的主要改进算法
- 5.4 基于遗传算法的生产调度方法

5.4.1 基于遗传算法的流水车间调度方法

1. 流水车间调度问题

■ **问题描述：** n 个工件要在 m 台机器上加工，每个工件需要经过 m 道工序，每道工序要求不同的机器， n 个工件在 m 台机器上的加工顺序相同。工件在机器上的加工时间是给定的，设为

$$t_{ij} (i = 1, \dots, n; j = 1, \dots, m)$$

■ **问题的目标：** 确定 n 个工件在每台机器上的最优加工顺序，使最大流程时间达到最小。



5.4.1 基于遗传算法的流水车间调度方法

1. 流水车间调度问题

■ 假设:

- (1) 每个工件在机器上的加工顺序是给定的。
- (2) 每台机器同时只能加工一个工件。
- (3) 一个工件不能同时在不同的机器上加工。
- (4) 工序不能预定。
- (5) 工序的准备时间与顺序无关，且包含在加工时间中。
- (6) 工件在每台机器上的加工顺序相同，且是确定的。

5.4.1 基于遗传算法的流水车间调度方法

2. 求解流水车间调度问题的遗传算法设计

(1) FSP的编码方法

对于FSP，最自然的编码方式是用染色体表示工件的顺序。

对于有四个工件的FSP，第 k 个染色体 $v_k = [1, 2, 3, 4]$ ，表示工件的加工顺序为： j_1, j_2, j_3, j_4 。

(2) FSP的适应度函数

$c_{\max}^k$: k 个染色体 v_k 的最大流程时间，

FSP的适应度函数：

$$eval(v_k) = \frac{1}{c_{\max}^k}$$

5.4.1 基于遗传算法的流水车间调度方法

3. 求解FSP的遗传算法实例

例5.1 Ho 和 Chang(1991) 给出的5个工件、4台机器问题。

加工时间表

工件 j	t_{j1}	t_{j2}	t_{j3}	t_{j4}
1	31	41	25	30
2	19	55	3	34
3	23	42	27	6
4	13	22	14	13
5	33	5	57	19

5.4.1 基于遗传算法的流水车间调度方法

用穷举法求得最优解：4-2-5-1-3，加工时间：213；

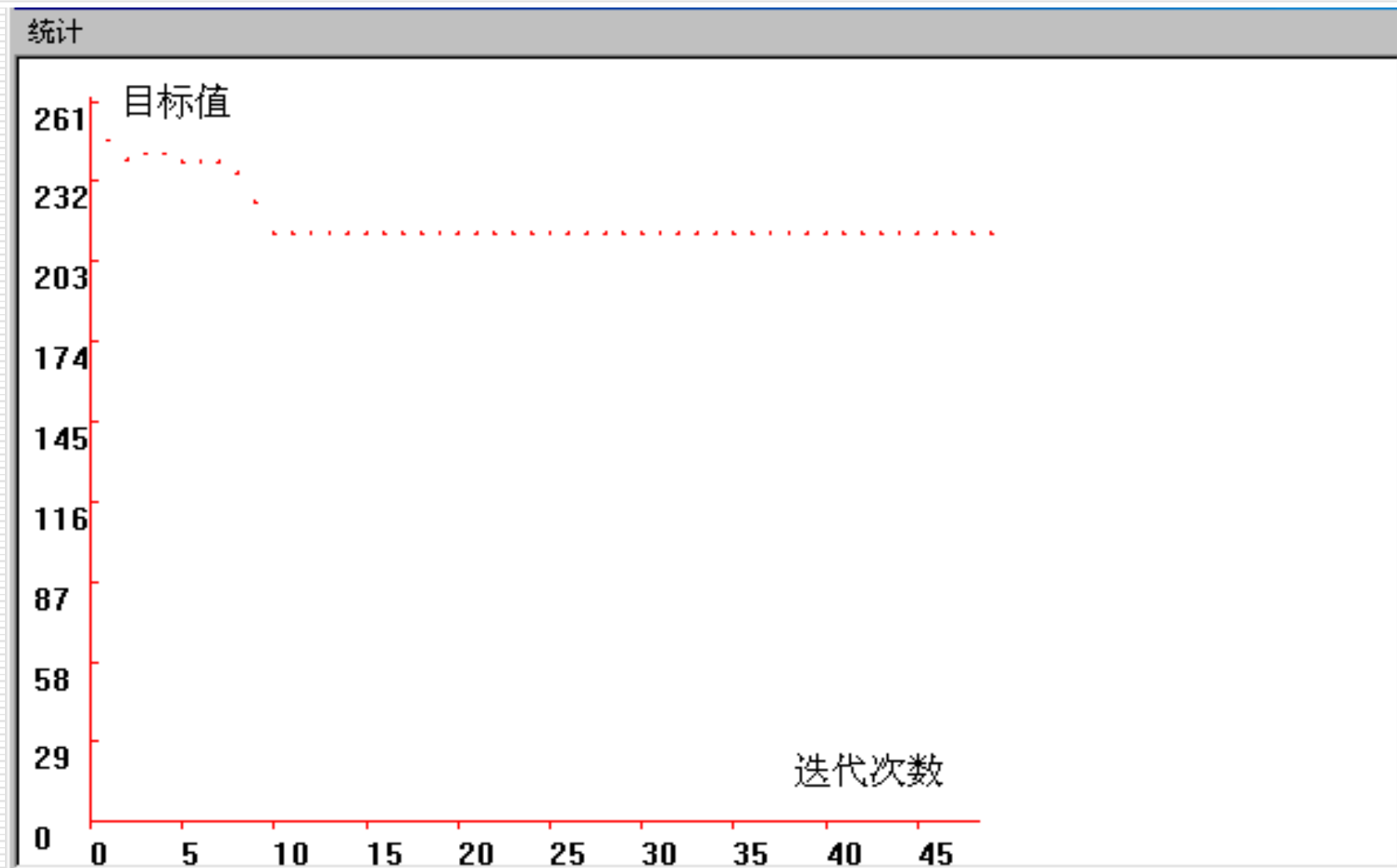
最劣解：1-4-2-3-5，加工时间：294；平均解的加工时间：265。

用遗传算法求解。选择交叉概率 $p_c = 0.6$ ，变异概率 $p_m = 0.1$ ，种群规模为20，迭代次数 $N = 50$ 。

遗传算法运行的结果

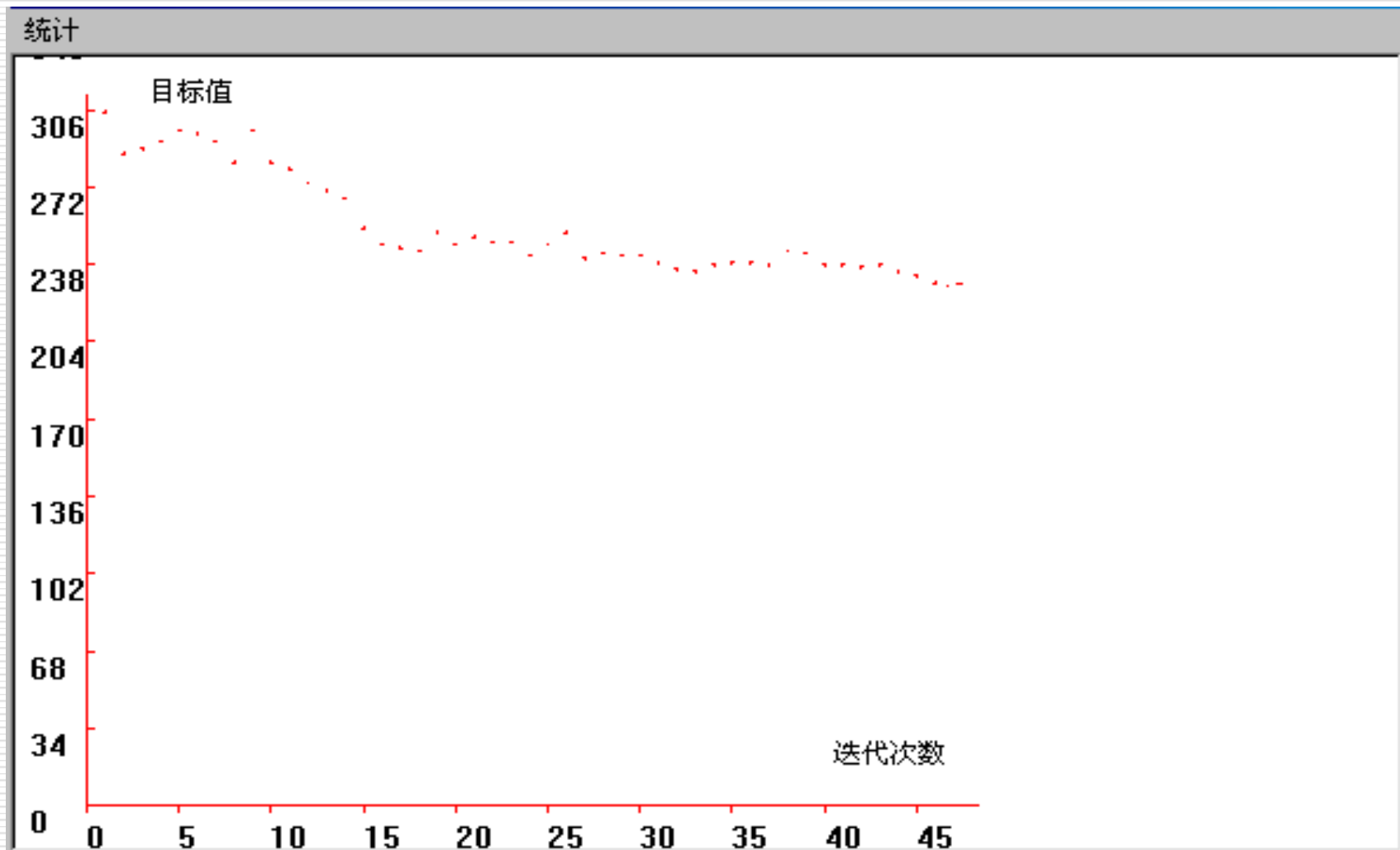
总运行次数	最好解	最坏解	平均	最好解的频率	最好解的平均代数
20	213	221	213.95	0.85	12

5.4.1 基于遗传算法的流水车间调度方法



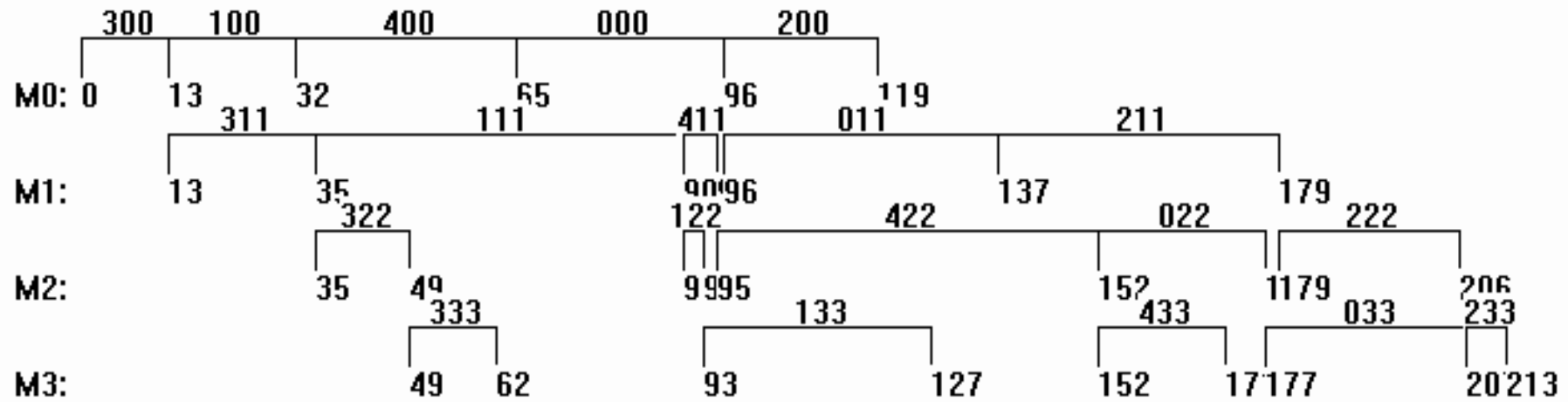
最优解收敛图

5.4.1 基于遗传算法的流水车间调度方法



平均值收敛图

5.4.1 基于遗传算法的流水车间调度方法



机器甘特图

5.4.2 基于遗传算法的混合流水车间调度方法

1. 混合流水车间调度问题

- 问题的特征：在某些工序上存在并行机器。
- **问题的描述**：需要加工多个工件，所有工件的加工路线都相同，都需要依次通过几道工序，在所有工序中至少有一个工序存在着多台并行机器。
- 需要解决的问题：确定并行机器的分配情况以及同一台机器上工件的加工排序。
- 目标：最小化最大流程时间。

5.4.2 基于遗传算法的混合流水车间调度方法

2. 混合流水车间调度问题的编码

■ 假设加工 N 个工件，每个工件都要依次经过 S 个加工工序，每个工序的并行机器数为 M_i ($i=1, \dots, S$)。所有工序中至少有一个工序存在并行机，即至少有一个 M_i 大于1。

■ HFSP的编码矩阵

$$A_{S \times N} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{S1} & a_{S2} & \dots & a_{SN} \end{bmatrix}$$

a_{ij} : $(1, M_i + 1)$ 上的一个实数，表示 j 工件的第 i 个工序在第 $Int(a_{ij})$ 台并行机上加工。

$Int(a_{ij}) = Int(a_{ik})$, $j \neq k$: 多个工件在同一台机器上加工同一个工序

5.4.2 基于遗传算法的混合流水车间调度方法

- $i=1$, 按 a_{1j} 的升序来加工工件。
- $i>1$, 根据每个工件的前一个工序的完成时间来确定其加工顺序, 前一个工序先完成的先加工。
- 假如完成时间相同, 也按 a_{ij} 的升序来加工。

■ 染色体:

$$Ind_k = [a_{11}, a_{12}, \dots, a_{1N}, 0, a_{21}, \dots, a_{2N}, 0, \dots, 0, a_{S1}, a_{S2}, \dots, a_{SN}]$$

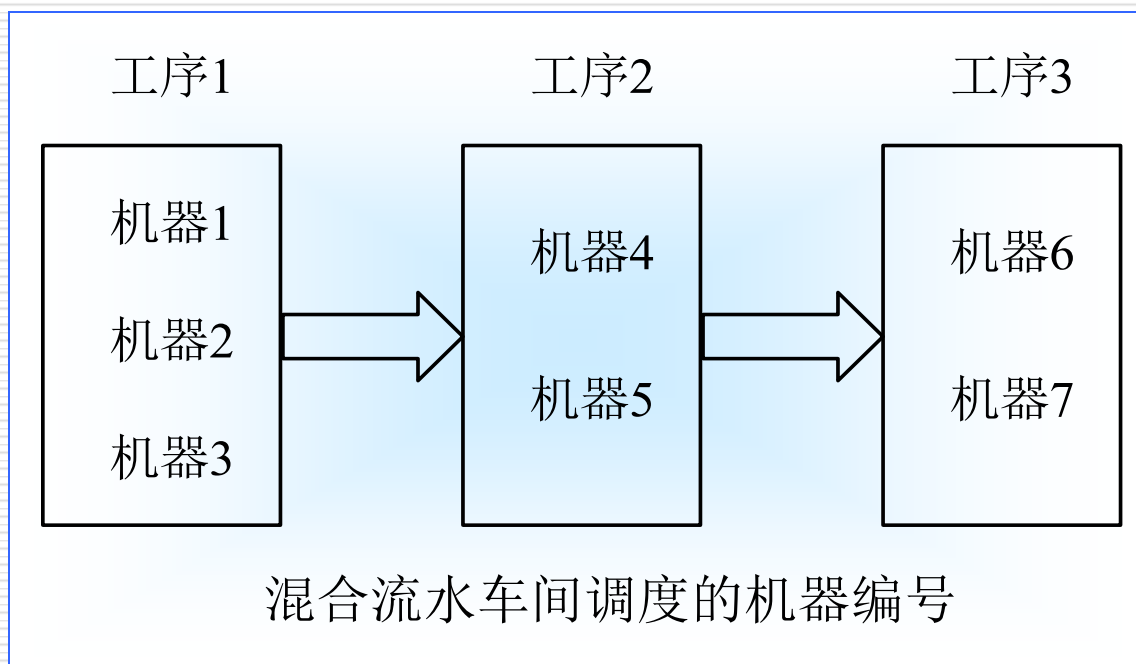
染色体的长度: $S \times N + N - 1$ 。

5.4.2 基于遗传算法的混合流水车间调度方法

- 例如，对于有3个工件、3道工序、各工序的并行机器数分别为3，2，2的混合流水车间调度问题。

■ 编码矩阵:

$$A = \begin{bmatrix} 2.1 & 2.4 & 1.9 \\ 1.6 & 2.1 & 2.3 \\ 1.1 & 2.4 & 1.2 \end{bmatrix}$$



■ 染色体:

[2.1, 2.4, 1.9, 0, 1.6, 2.1, 2.3, 0, 1.1, 2.4, 1.2]

5.4.2 基于遗传算法的混合流水车间调度方法

3. 调度实例

■ 某汽车发动机厂金加工车间要加工12个工件，每个工件都有车、刨、磨3个工序，现有3台车床，2台刨床，4台磨床，每台机床的加工能力不同。



5.4.2 基于遗传算法的混合流水车间调度方法

工件在每个机器上的加工时间

工件	工序1			工序2		工序3			
	机器1	机器2	机器3	机器4	机器5	机器6	机器7	机器8	机器9
1	2	2	3	4	5	2	3	2	3
2	4	5	4	3	4	3	4	5	4
3	6	5	4	4	2	3	4	2	5
4	4	3	4	6	5	3	6	5	8
5	4	5	3	3	1	3	4	6	5
6	6	5	4	2	3	4	3	9	5
7	5	2	4	4	6	3	4	3	5
8	3	5	4	7	5	3	3	6	4
9	2	5	4	1	2	7	8	6	5
10	3	6	4	3	4	4	8	6	7
11	5	2	4	3	5	6	7	6	5
12	6	5	4	5	4	3	4	7	5

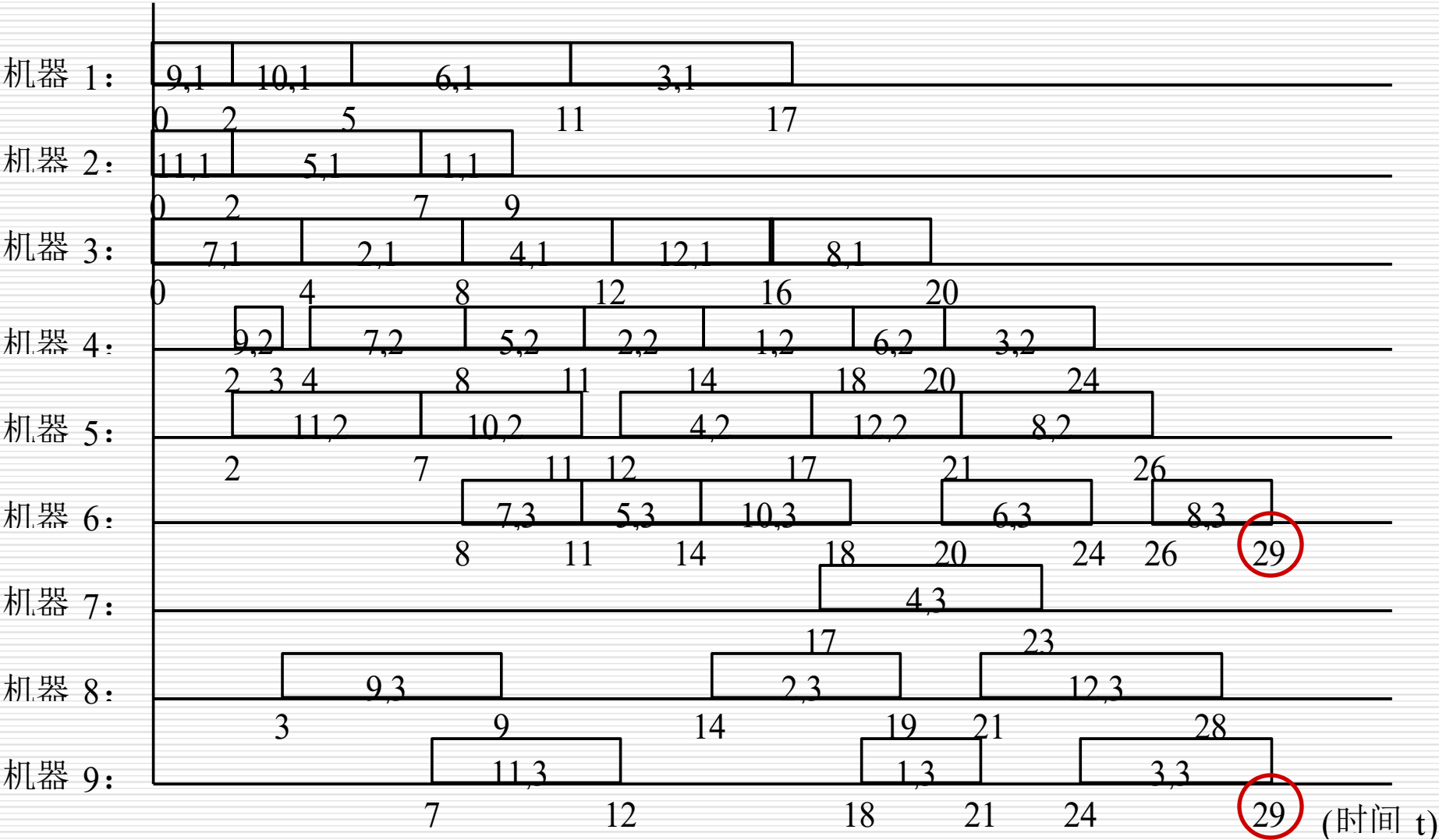
5.4.2 基于遗传算法的混合流水车间调度方法

算法中使用的参数为 $\alpha = 0.07$, $P_c = 0.80$, $P_m = 0.01$, 种群规模为30, 种群经过100代的进化, 目标函数最小值随着种群的进化逐渐地减小, 最后收敛于极值, 目标函数平均值也随着群体的进化逐渐减少, 最后趋近于最优值。

最好的染色体:

[2.77, 3.51, 1.74, 3.52, 2.42, 1.36, 3.28, 3.94, 1.09,
1.22, 2.24, 3.64, 0, 1.60, 1.13, 1.24, 2.97, 1.73, 1.88,
1.08, 2.68, 1.16, 2.69, 2.51, 2.96, 0, 4.99, 3.29, 4.95,
2.35, 1.10, 1.01, 1.73, 1.35, 3.06, 1.20, 4.13, 3.67]

5.4.2 基于遗传算法的混合流水车间调度方法





THE END

