可求解与难求解问题

战德臣

哈尔滨工业大学 教授.博士生导师 教育部大学计算机课程教学指导委员会委员



Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

可求解与难求解问题 (1)什么是可求解与难求解问题?



现实世界中的问题分类

- ●计算机在有限时间内能够求解的(可求解问题)
- ●计算机在有限时间内不能求解的(难求解问题)
- ●计算机完全不能求解的(不可计算问题)

问题的计算复杂性

计算复杂性是指问题的一种特性,即利用计算机求解问题的难易性或难易程度,其衡量标准:

- ◆计算所需的步数或指令条数(即时间复杂度)
- ◆计算所需的存储空间大小(即空间复杂度)
- ----通常表达为关于问题规模n的一个函数 O(f(n))

可求解与难求解问题 (2)什么是有限时间内不能求解?



O(n³)与O(3ⁿ)的差别,O(n!)与O(n³)的差别

问题规模n	计算量
10	10!
20	20!
100	100!
1000	1000!
10000	10000!

$$20! = 1.216 \times 10^{17}$$

$$20^3 = 8000$$

$O(n^3)$	O (3 ⁿ)
0.2秒	4×10 ²⁸ 秒
	=1015年

注: 每秒百万次, n=60, 1015年相当于10 亿台计算机计算一百万年

O(1), O(log n), O(n), O(n log n), O(n^b) O(bⁿ), O(n!)

可求解与难求解问题 (3)怎样以复杂性来划分问题?



P类问题,NP类问题,NPC类问题

- ■P类问题: 多项式问题(Polynomial Problem),指计算机可以在有限时间内求解的问题,即: P类问题是可以找出一个呈现O(n²)复杂性算法的问题,其中a为常数。
- ■NP类问题:非确定性多项式问题(Non-deterministic Polynomial)。有些问题,其答案是无法直接计算得到的,只能通过间接的猜算或试算来得到结果,这就是非确定性问题(Non-deterministic)。虽然在多项式时间内难于求解但不难判断给定一个解的正确性的问题,即:在多项式时间内可以由一个算法验证一个解是否正确的非确定性问题,就是NP类问题。
- ■NPC问题:完全非确定性多项式问题(NP-Complete)。如果NP问题的所有可能答案都可以在多项式时间内进行正确与否的验算的话就叫做完全非确定性多项式问题,即NP-Complete问题。

问:加密算法应该设计成一个什么问题呢?

可求解与难求解问题 (4)NPC类问题如何求解?



NPC类问题求解

TSP问题的 遍历算法 **穷举法**或称**遍历法**:对解空间中的每一个可能解进行验证,直到所有的解都被验证是否正确,便能得到精确的结果---**精确解**



可能是O(n!) 或O(aⁿ)

TSP问题的 贪心算法

仿生学算法

近似解求解算法---近似解

 $\Delta = |$ 近似解 - 精确解 | 满意解: Δ 充分小时的近似解 应该是O(na)

进化算法,遗传算法,蚁群算法,蜂群算法,***

可求解与难求解问题 (5)小结?





算法复 计算复杂性

P类 问题 NDC-

NPC类问题

NPC类 问题求解 -计算

精确解 近似解

遗传算法的缘起--生物学中的遗传与进化

战德臣

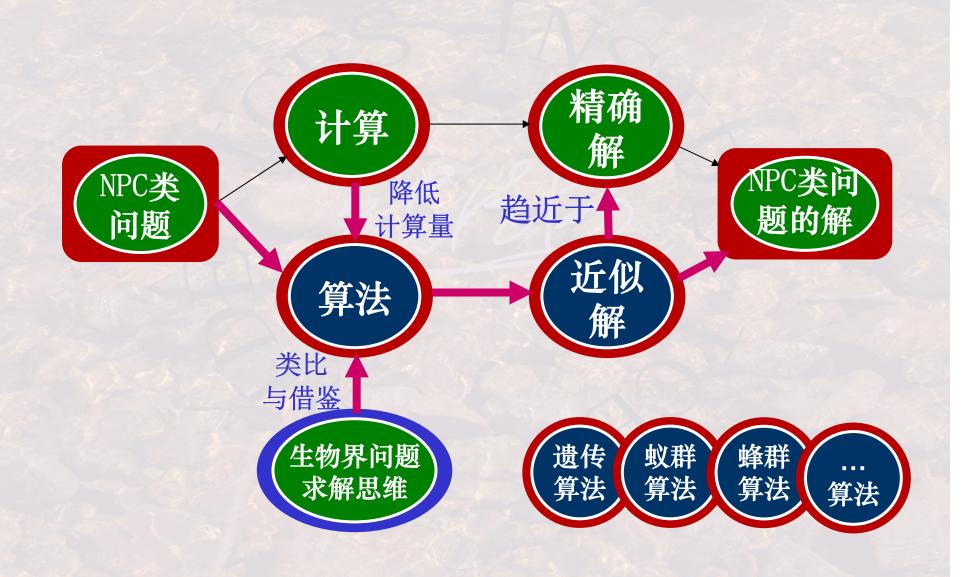
哈尔滨工业大学 教授.博士生导师 教育部大学计算机课程教学指导委员会委员



Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

遗传算法的缘起--生物学中的遗传与进化 (1)为什么引入生物学的内容?





遗传算法的缘起--生物学中的遗传与进化 (2)生物体中是怎样遗传的?



生物学中的遗传和进化

关于生物遗传进化的基本观点

- (i) 生物的所有遗传信息都包含在其染色体中,染色体决定了生物的性状;
- (ii) 染色体是由基因及其有规律的排列所构成的,遗传和进化过程发生在染色体上;
- (iii) 生物的繁殖过程是由其基因的复制过程来完成的;
- (iv) 通过同源染色体之间的交叉或染色体的变异会产生新的物种,使生物呈现新的性状。
- (v) 对环境适应性好的基因或染色体经常比适应性差的基因或染色体有更多的机会遗传到下一代。

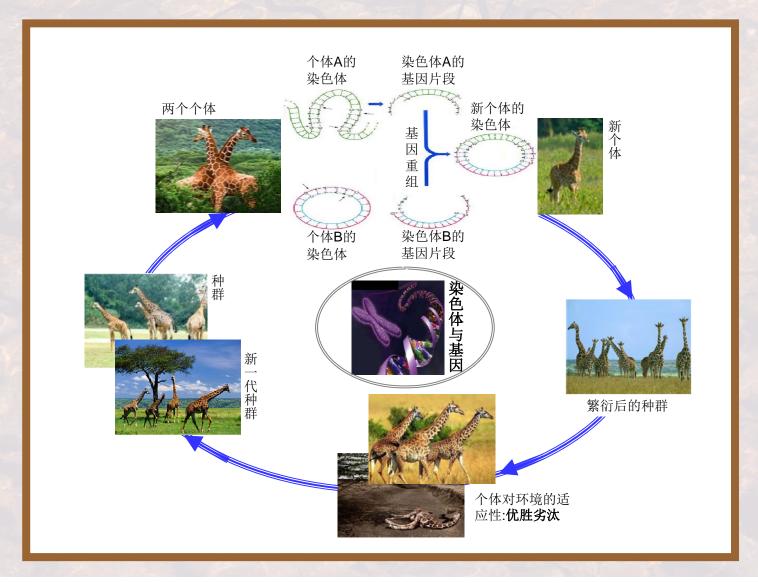
遗传与进化

优胜劣汰

遗传算法的缘起--生物学中的遗传与进化 (3)生物体遗传与进化的过程是怎样的?



生物学中的遗传和进化



遗传算法的缘起--生物学中的遗传与进化 (4)什么是染色体和基因,它们与遗传有什么关系?



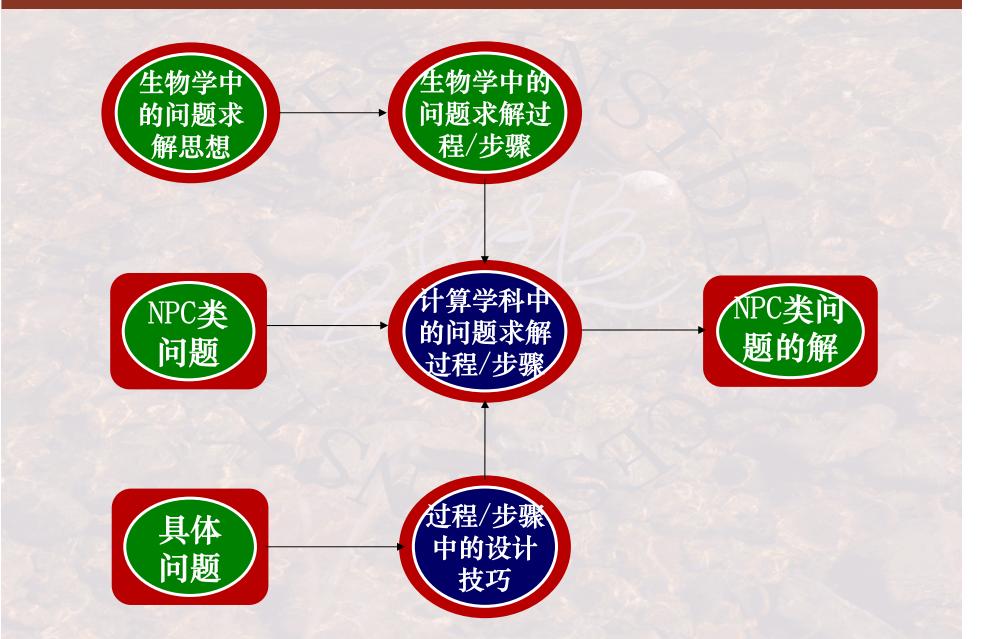
生物学中的遗传和进化

基本概念

- ■种群(Population) vs. 个体(Individual) vs. 染色体(chromosome)
- ■染色体(chromosome) vs.基因(gene)
- ■基因型(Genotype) vs.表现型(Phenotype)
- ■个体的适应度(Fitness)
- ■选择(Selection)
- ■复制((Reproduction)
- ■交配/杂交(Crossover)
- ■突变(Mutation)

遗传算法的缘起--生物学中的遗传与进化 (5)小结?





计算学科的遗传算法

战德臣

哈尔滨工业大学 教授.博士生导师 教育部大学计算机课程教学指导委员会委员



Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

计算学科的遗传算法 (1)怎样用遗传算法求解问题?



一个简单的遗传算法应用示例

求多项式函数的最小值:

Min $F(X) = X^2 - 19X + 20$, 其中X=1,...,64之间的整数。

注:此题不难求出精确解,其精确解为X=9或者X=10。



如何用遗传算法进行求解?

计算学科的遗传算法

(2)生物学中的概念如何与计算学科中的概念映射?



生物学中的概念与计算学科中的概念映射

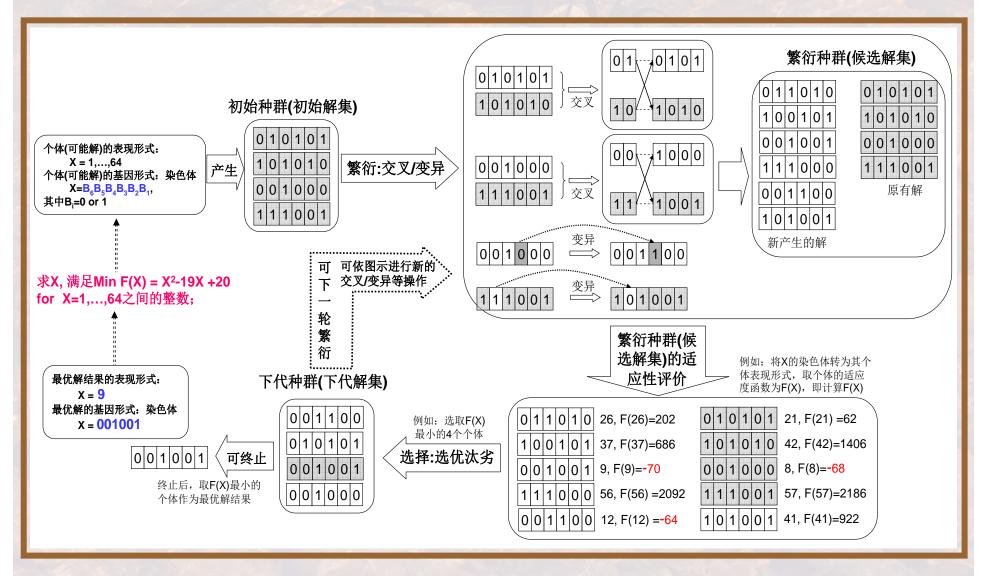
生物学中 的概念	计算学科 中的概念	解释
种群	解集/种群	若干可能解的集合
个体	解/个体	一个可能解的表现型,本例中即是十进制的X
染色体	编码解 /染 色体	一个可能解的基因型,本例即是 X 的二进制编码。 $X=b_6b_5b_4b_3b_2b_1$,其中 b_i =0或1。
基因	基因	解编码中的若干位的b _i
适应度	适应度	一个可能解接近最优解的一个度量,本例直接用F(X)作为其适应度的 度量,其值越小越接近最优解
选择	选择	指从种群(解集)中依据适应度按某种条件选择某些个个体(可能解)
复制	复制	将一个解从一个解集复制到另一个解集
交配/ 杂交	交叉	即交配/杂交,是新可能解的一种形成方法,即是对两个可能解的编码通过交换某些编码位而形成两个新的可能解的遗传操作
突变	变异	也是新可能解的一种形成方法,它是通过随机地改变一个可能解的编码的某些片段(或基因)而使一个可能解变为一新的可能解的遗传操作

计算学科的遗传算法

(3)怎样模拟遗传算法进行求解?



遗传算法求解过程的模拟



计算学科的遗传算法 (4)如何设计遗传算法?



遗传算法基本框架及其设计要点

begin /* 遗传算法 */

 $t \leftarrow 0$; /* 进化的种群代数 */

生成初始种群P(t);

计算初始种群P(t)中每个个体的适应值;

while(不满足终止条件)do

/* 利用下述操作生成新个体,并选择更优个体组成新种群 */

(1)通过复制、交叉或变异操作重组种群P(t)中的个体,产生新个体,形成候选种群C(t);

/*注意此处C(t)并未包含P(t)中的个体 */

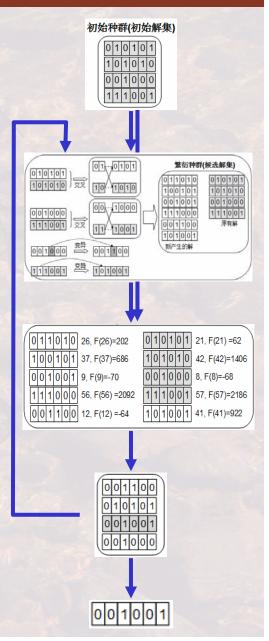
- (2)计算C(t)中每个个体的适应值;
- (3)根据适应值从C(t)和P(t)中选择更优的个体组成新种群P(t+1);

 $(4) t \leftarrow t+1;$

end while

选择P(t)中最优个体为所求的解;

end begin

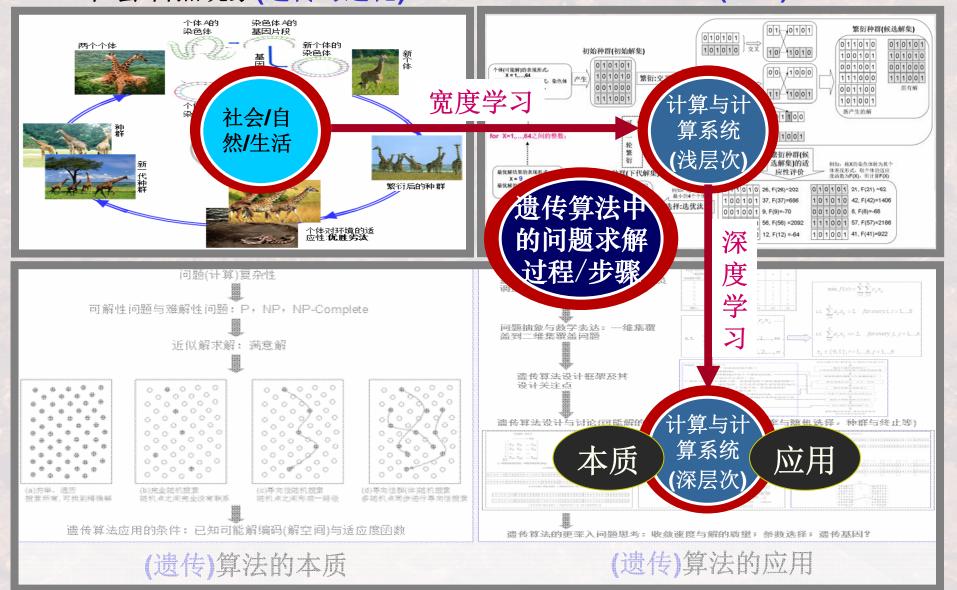


计算学科的遗传算法 (5)小结?



社会/自然现象(遗传与进化)

计算学科的(遗传)算法



遗传算法 为什么可以求解NPC问题

战德臣

哈尔滨工业大学 教授.博士生导师 教育部大学计算机课程教学指导委员会委员



Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

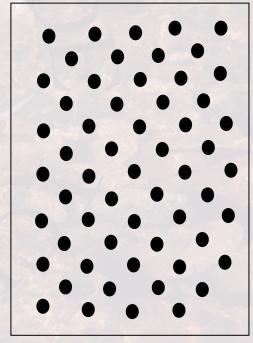
遗传算法为什么可以求解NPC问题 (1)NPC问题的求解



■NPC问题理论上可通过枚举-验证的遍历算法来进行

穷举法或称**遍历法**:对解空间中的每一个可能解进行验证,直到所有的解都被验证是否正确,便能得到精确的结果---精确解

可能是O(n!) 或O(aⁿ)



(a)穷举,遍历 搜索所有,可找到精确解

遗传算法为什么可以求解NPC问题 (2)怎样用遗传算法求解问题?

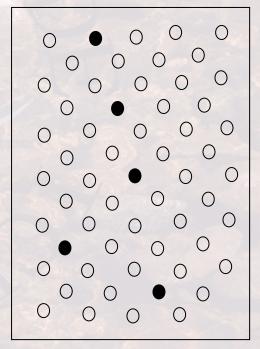


遗传算法的基本思想

■不求精确解,而求近似解-满意解,可采取随机搜索方法

随机搜索法: 在解空间中随机选择一些可能解进行验证,求出所选择可能解(子解空间)中的最优解---**近似解**

- •基于概率论: 随机选择
- •子解空间越大,求得满意解的可能 性越大,但耗时也会加长
- •求出的近似解并不能保证是满意解



(b)完全随机搜索 随机点之间完全没有联系

遗传算法为什么可以求解NPC问题 (2)怎样用遗传算法求解问题?

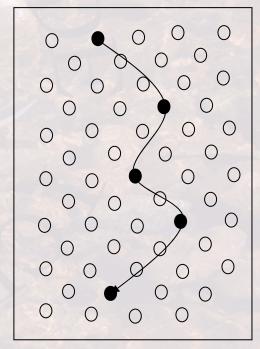


遗传算法的基本思想

■为提高近似解的质量,可采取导向性随机搜索方法

导向性随机搜索法:对随机点的选取进行导向(导向到接近最优解的方向或路径)

- •基于概率论: 随机选择
- •随机选择的可能解与前一可能解相比,更偏向于满意解
- •万一初始解就很差怎么办?



(c)导向性随机搜索 随机点之间形成一路径

遗传算法为什么可以求解NPC问题 (2)怎样用遗传算法求解问题?

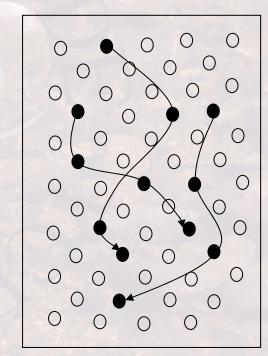


遗传算法的基本思想

■为进一步提高近似解的质量,可采取导向性群(体)随机搜索方法

导向性群(体)随机搜索法:同时对多个随机点的选取进行导向(导向到接近最优解的方向或路径),多条搜索路径

- •基于概率论: 随机选择
- •多条路径下的最优,总比一条路径的最优要更优一些。
- •遗传算法就是这样一种导向性群随机搜索算法。
- •同一时刻多条路径上的解集合即为一个种群。多次选择,即多代进



(d)导向性群(体)随机搜索 多随机点同步进行导向性搜索

遗传算法为什么可以求解NPC问题 (3)什么情况下可用遗传算法求解问题?



遗传算法的适用条件

- ■对于NP问题,当没有其他更好的算法可以使用时,可以 考虑选择遗传算法。
- ■遗传算法的适用条件:
- (1)已知"解空间",即可能解的表现型和基因型
- (2)关于可能解的"适应度"函数的计算方法(适应度用于判断一个可能解接近精确解的程度或方向)。

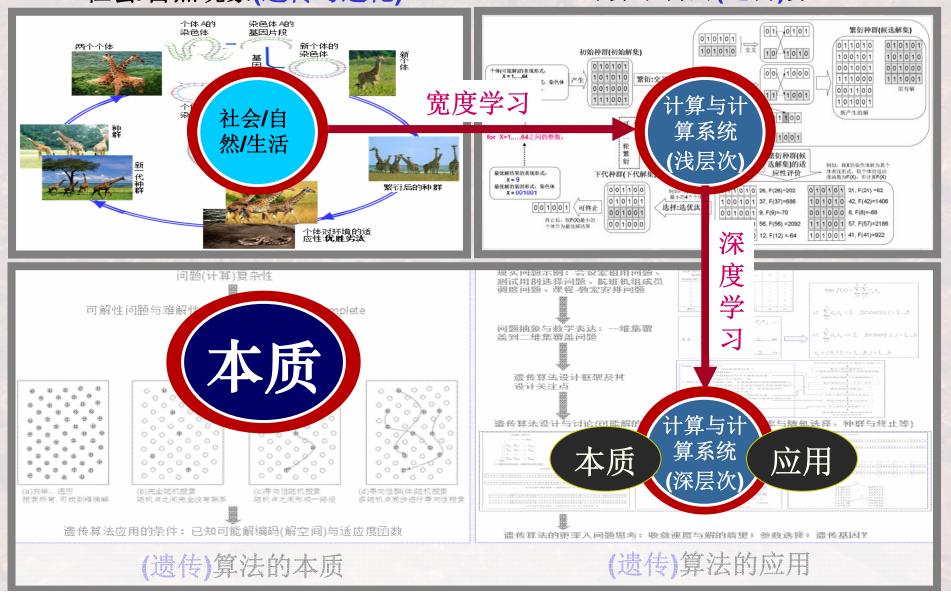
遗传算法提供了一种求解复杂系统问题的通用框架。

遗传算法为什么可以求解NPC问题 (4)小结?



社会/自然现象(遗传与进化)

计算学科的(遗传)算法



怎样用遗传算法求解具体的应用问题(I)—问题及其建模

战德臣

哈尔滨工业大学 教授.博士生导师 教育部大学计算机课程教学指导委员会委员



Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

怎样用遗传算法求解具体的应用问题(I)—问题及其建模(1)你能举出一些需要求解的现实问题吗?



现实世界的几个具体问题分析

例1: "会议室"租用问题

例2: "航班机组成员"选择问题

例3: "软件测试用例"选择问题

NAME OF THE PARTY		约束矩阵		车		
测试用例 软件功能	X 1	X ₂	X 3	X ₄	X 5	X ₆
T ₁	0	0	0	1	1	0
T ₂	_ 0	1	1	0	0	1
T_3	1	0	a	0	1	0
T ₄	0	1	0	1 1	0	1
T ₅	0	1	1	0	0	0
T ₆	1	0	0	0	1	0
费用 C _j	500	250	250	400	600	400

 $< x_1, x_2, \dots, x_n >$

其实它们是同一个问题: 一维的集覆盖问题

$$\min \ z(x) = \sum_{j=1}^{n} c_j x_j$$

s.t.
$$\sum_{j=1}^{n} a_{ij} x_j \ge 1$$
, $i = 1, 2, ..., m$

$$x_{j} \in \{0,1\}, j = 1,2,...,n$$

怎样用遗传算法求解具体的应用问题(I)—问题及其建模(2)例4和例1,2,3有何不同?



例4"课程表"优化问题

- •有8门课程需要安排教室,记为L;, i=1,...,8;
- •有6个教室可被使用,记为R_i, j=1,...,6;
- •要求每门课只能安排在一个教室,而每个教室最多只能安排两门课。
- •教室与课程班之间的约束矩阵A如下表给出,其中a_{ij}=1表示该课程班可以安排在该教室,a_{ii}=0则不可安排。
- •费用为教室容纳人数/课程班人数,即C_{ij}=K_j/L_i。

教室(最大人数)	R1	R2	R3(多媒体)	R4	R5(多媒体)	R6(多媒体)
课程(选课人数)	(100 人)	(50 人)	(50 人)	(80 人)	(100 人)	(80 人)
L ₁ (85 人)	1	0	0	0	1	0
L ₂ (40 人)	1	1	1	1	1	1
L ₃ (95 人)	1	0	0	0	1	0
L ₄ (60 人)	1	0	0	1	1	1
L ₅ (45 人)	1	1	1	1	1	1
L ₆ (90 人)	1	0	0	0	1	0
L ₇ (76 人)	1	0	0	1	1	1
L ₈ (56 人)	1	0	0	1	1	1
K	100	50	70	80	120	100

 $C_{ii} = K_i / L_i$

怎样用遗传算法求解具体的应用问题(I)—问题及其建模 (3)例4和例1,2,3有何不同?



例1,例2,例3

例4

一维集覆盖问题

可能解是 一维向量

$$< x_1, x_2, \dots, x_n >$$

使每一行都被选出的列覆盖, 被哪一列或几列覆盖不重要, 要满足约束矩阵

$$\min \ z(x) = \sum_{j=1}^{n} c_j x_j$$

s.t.
$$\sum_{j=1}^{n} a_{ij} x_j \ge 1$$
, $i = 1, 2, ..., m$

$$x_j \in \{0,1\}, j = 1,2,...,n$$

二维集覆盖问题

二维矩阵

使每一行都确定安排在某一列 上,使被选中的行-列覆盖所有 的行,要满足约束矩阵

$$\min f(x) = \sum_{i=1}^{8} \sum_{j=1}^{6} c_{ij} x_{ij}$$

s.t.
$$\sum_{j=1}^{6} a_{ij} x_{ij} = 1$$
, for every $i, i = 1,...,8$

s.t.
$$\sum_{i=1}^{8} a_{ij} x_{ij} \le 2$$
, for every $j, j = 1,...,6$

$$x_{ij} \in \{0,1\}; i = 1,...,8; j = 1,...,6$$

怎样用遗传算法求解具体的应用问题(I)—问题及其建模 (4)小结?



例1,例2,例3 一维集覆盖问题 二维集覆盖问题 可能解是 <X4,X2,***,Xn> 一维细醛 问题理解很重要,从问题出发寻找算法 一列 要满足约 数学建模对于正确理解问题很关键 可从问题角度寻找针对问题的特定的算法 s.t. $\sum a_{ij}x_{ij} \ll 2$, for every j, j = 1,...,6 $x_{j} \in \{0,1\}, j = 1,2,...,n$ $x_{ij} \in \{0,1\}; i = 1,...,8; j = 1,...,6$

怎样用遗传算法求解具体的应用问题(II)—算法设计要点及解的编码

战德臣

哈尔滨工业大学 教授.博士生导师 教育部大学计算机课程教学指导委员会委员



Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

怎样用遗传算法求解具体的应用问题(II)—算法设计要点及解的编码(1)遗传算法的设计要点是什么?



遗传算法设计要点

解的表现型和基因型

NPC求解:

- ●产生一个或一批可能解
- ●判断可能解是否是问题的解

可能解的形式是怎样的?

怎样产生待判定的可能解?

产生多少个待判定可能解?

交叉、变异随机(概率)

怎样判定一个解是否是所求的解?

适应度 选择(标准) 满意解

解集的规模进化的代数随机(概率)

怎样用遗传算法求解具体的应用问题(II)—算法设计要点及解的编码(1)遗传算法的设计要点是什么?



一组关于解的概念需要区分

<x₁,x₂,···,x_n>

$$\min \ z(x) = \sum_{j=1}^{n} c_j x_j$$
s.t.
$$\sum_{j=1}^{n} a_{ij} x_j \ge 1, \ i = 1, 2, ..., m$$

$$x_j \in \{0, 1\}, \ j = 1, 2, ..., n$$

假设一个问题的解的形式为x,由x的取值空间或定义域给定的任何一个x值被称为可能解

而一个问题通常有很多个关于可能解的约束,即不是任何一个**x**值都满足约束,我们可将满足问题约束的那些**x**值称为**可行解** 而由一个算法在任何一组可行解中求出的最优解被称为是**近似解** 而符合用户期望的近似解被称为是**满意解**

所有可行解中的最优解是问题的最优解。

"可能解集合"⊃"可行解集合"⊃"近似解集合"⊃"满意解集合"⊃"最优解集合"

怎样用遗传算法求解具体的应用问题(Ⅱ)—算法设计要点及解的编码 (1)遗传算法的设计要点是什么?



遗传算法设计要点

begin /* 遗传算法 */

t ← 0; /* 进化的种群代数 */

生成初始种群P(t);

计算初始种群P(t)中每个个体的适应值;

while (不满足终止条件) do

/* 利用下述操作生成新个体,并选择更优个体组成新种群 */ 通过复制、交叉或变异操作重组种群P(t)中 的个体,产生新个体,形成候选种群C(t);

/*注意此处C(t)并未包含P(t)中的个体 */

计算*C(t)*中每个个体的适应值; 根据适应值从*C(t)*和*P(t)*中选择更优的个体 组成新种群*P(t*+1);

 $t \leftarrow t+1$;

end while

选择P(t)中最优个体为所求的解;

end begin

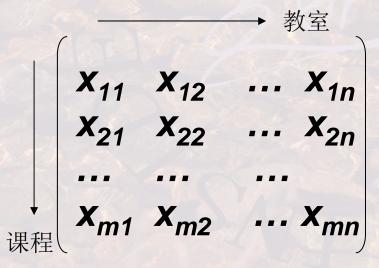
实际问题分析:解的形式,解的约束,解空间 问题解的编码与解码规则设计: 个体(解)的表现型与基因型的变换函数 初始种群的规模与生成规则设计 遗传规则的设计:交叉、变异 繁衍种群的策略设计 种群个体的适应度函数设计 优质个体的选择(汰选)方法设计 终止条件及最终解产生 解的满意度评估,算法效率的评 估以及算法的改进

怎样用遗传算法求解具体的应用问题(II)—算法设计要点及解的编码(2)为什么要考虑解的形式与解的编码?



例4问题的解的形式(表现型)与编码(基因型)

可能解的一般形式



 X_{ij} =1课程i被安排在教室j; =0课程未被安排在教室j

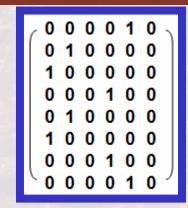
可行解1

可行解2

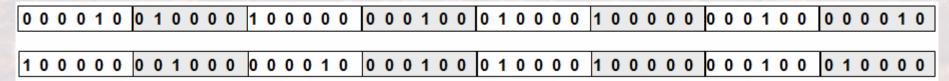
怎样用遗传算法求解具体的应用问题(II)—算法设计要点及解的编码(2)为什么要考虑解的形式与解的编码?



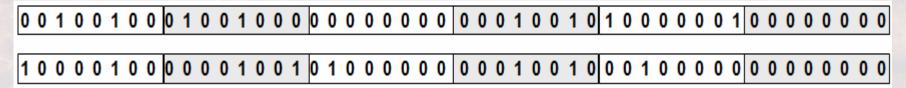
例4问题的解的形式(表现型)与编码(基因型)针对具体问题,可以有不同的编码方案



行优先编码:课程分段编码,一门课程相关的划为一段,有多少课程就有多少段



列优先编码: 教室分段编码, 一个教室相关的划为一段, 有多少教室就有多少段

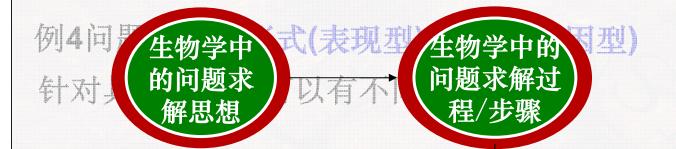


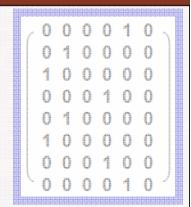
(解的)编码方案不同,交叉、变异(产生新可能解)的规则也可能不同

怎样用遗传算法求解具体的应用问题(II)—算法设计要点及解的编码(3)小结?



战德臣 教授

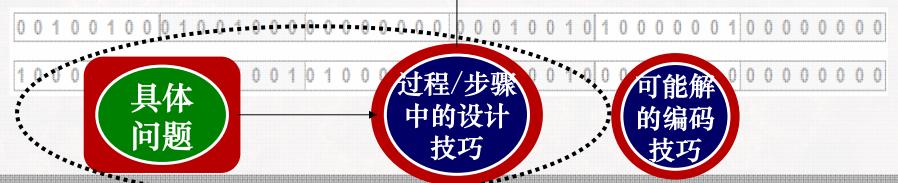




行优先编码: 课程分段编码,一门课程担当的为一段,有多少课程就有多少段



列优先编码: 教室分段编码, 一个教室相关的划为一段, 有多少教室就有多少段



(解的)编码方案不同,交叉、变异(产生新可能解)的规则也可能不同