

遗传算法理论与实现技术

(一)

引言

- 遗传算法(genetic algorithms, GA)是J Holland于1975年受生物进化论的启发而提出的。
- GA的提出一定程度上解决了传统的基于符号处理机制的人工智能方法在知识表示、信息处理和解决组合爆炸等方面所遇到的困难，其自组织、自适应、自学习和群体进化能力使其适合于大规模复杂优化问题。
- GA是基于“适者生存”的一种高度并行、随机和自适应优化算法，它将问题的求解表示成“染色体”的适者生存过程，通过“染色体群(population)的一代代不断进化，包括复制(reproduction)、交叉(crossover)和变异(mutation)等操作，最终收敛到“最适应环境”的个体，从而求得问题的最优解或满意解。
- GA是一种通用的优化算法，其编码技术和遗传操作比较简单，优化不受限制性条件的约束，而其两个最大的显著特点则是隐含并行性和全局解空间搜索。
- 目前，随着计算机技术的发展，GA愈来愈得到重视，并在机器学习、模式识别、图象处理、神经网络、优化控制、组合优化、VLSI设计、遗传学、日程安排(如1992年奥运会)等领域得到了成功应用，尤其在生产调度领域。

学习内容

- 遗传算法的基本流程
- 模式定理和隐含并行性
- 遗传算法的马氏链描述及其收敛性、收敛速度估计
- 遗传算法参数与操作的设计
 - 编码
 - 适配值函数
 - 算法参数
 - 遗传操作
 - 算法终止条件
- 遗传算法的改进
- 免疫遗传算法
- 并行遗传算法

基本概念

- 染色体(Chromosome), 个体(Individual), 串(String)
- 基因(Gene)
- 种群(Population)
- 适配值(Fitness value)
- 选择(Selection), 复制(Reproduction)
- 交叉(Crossover)
- 变异(Mutation)

遗传算法的机制和基本操作

- 机制：生物进化思想的启迪，即 适者生存, 优生劣汰, 群体进化
- 选择: 适配值高或目标值好的个体在下一代中被复制的概率大, 提高种群的平均适配值.
- 交叉: 交换两父代个体的部分信息来构成后代, 使后代继承父代的有效模式, 有助于产生优良个体.
- 变异: 随机改变个体的某些基因来产生新个体, 有助于增加种群的多样性.

EC(Evolutionary Computation)

- 遗传算法(Genetic Algorithm, GA), 强调基因层次[J.H Holland, D.E Goldberg, Z Michalewicz]
- 进化规划(Evolutionary Programming, EP), 强调群体遗传结构[L.J Fogel, D.B Fogel]
- 进化策略(Evolutionary Strategy, ES), 强调个体遗传结构[H. P Schwefel, I Rechenberg]
- 遗传编程(Genetic Programming, GP), 机器学习和建模 [J.R Koza]

遗传算法的基本流程

步骤 1: 令 $k=0$ ，随机产生 N 个初始个体构成初始种群 $P(0)$ 。

步骤 2: 评价 $P(k)$ 中各个体的适配值(fitness value)。

步骤 3: 判断算法收敛准则是否满足。若满足则输出搜索结果；否则执行以下步骤。

步骤 4: 令 $m = 0$ 。

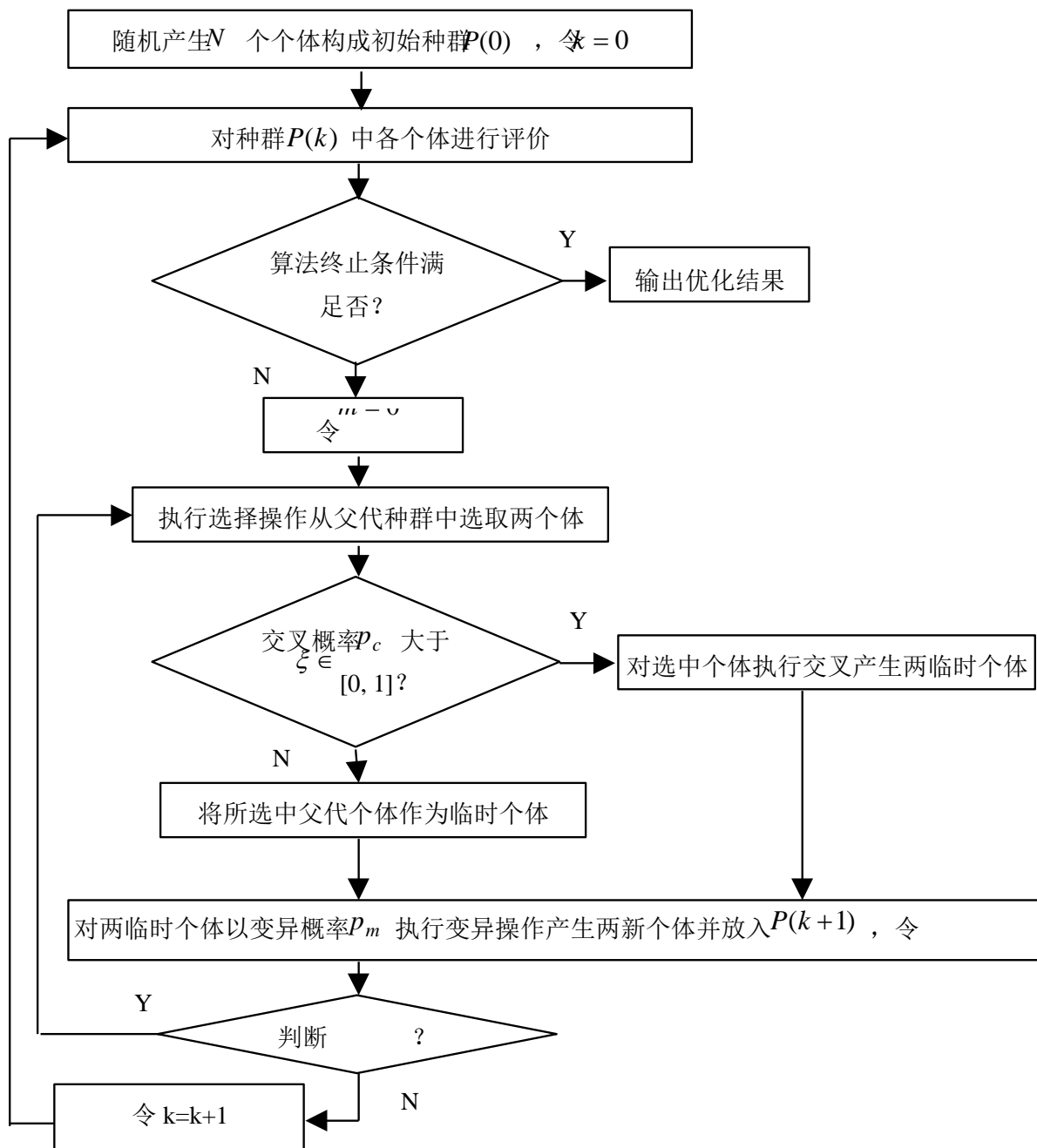
步骤 5: 根据适配值大小以一定方式执行复制操作来从 $P(k)$ 中选取两个个体。

步骤 6: 若交叉概率 $p_c > \xi \in [0, 1]$ ，则对选中个体执行交叉操作来产生两个临时个体；否则将所选中父代个体作为临时个体。

步骤 7: 按变异概率 p_m 对临时个体执行变异操作产生两个新个体并放入 $P(k+1)$ ，并令 $m = m + 2$ 。

步骤 8: 若 $m < N$ ，则返回步骤 5；否则令 $k = k + 1$ 并返回步骤 2。

- 适配值是对个体进行评价的一种指标，是GA进行优化所用的主要信息，它与个体的目标值存在一种对应关系；
- 复制操作(也称选择操作)通常采用比例复制，即复制概率正比于个体的适配值，如此意味着适配值高的个体在下一代中复制自身的概率大，从而提高了种群的平均适配值；
- 交叉操作通过交换两父代个体的部分信息构成后代个体，使得后代继承父代的有效模式，从而有助于产生优良个体；
- 变异操作通过随机改变个体中某些基因而产生新个体，有助于增加种群的多样性，避免早熟收敛。



遗传算法的特点

- GA对问题参数编码成“染色体”后进行进化操作，而不是针对参数本身，这使得GA不受函数约束条件的限制，如连续性、可导性等。
- GA的搜索过程是从问题解的一个集合开始的，而不是从单个个体开始的，具有隐含并行搜索特性，从而大大减小了陷入局部极小的可能。
- GA使用的遗传操作均是随机操作，同时GA根据个体的适配值信息进行搜索，而无需其它信息，如导数信息等。
- GA具有全局搜索能力，最善于搜索复杂问题和非线性问题。

遗传算法的优越性

- 算法进行全空间并行搜索，并将搜索重点集中于性能高的部分，从而能够提高效率且不易陷入局部极小。
- 算法具有固有的并行性，通过对种群的遗传处理可处理大量的模式，并且容易并行实现。

SGA的模式Schema定理

- SGA(二进制编码), 基因(0 or 1), 通配符*
- X_l 为长度为 l 的二进制串的全体
- 模式空间 $V^l = \{0, 1, *\}^l$
- 例模式 $H = 01**1$ 表示集合 $\{01001, 01011, 01101, 01111\}$
- 阶 $o(H)$: 模式中取确定值的位置的数目, 上例中 $o(H) = 3$.
- 模式的对应长度 $\delta(H)$: 模式中第一个取确定值的位置与最后一个取确定值的位置之间的距离, 上例中 $\delta(H) = 4$.

模式定理(续)

- 种群数目 N
- 种群空间 $X=(B^l)^N$
- H 为 V^l 中的任一模式
- 第 t 代种群 $P(t) = \{x_1(t), x_2(t), \dots, x_N(t)\} \in X$
- 适配值函数 $f : (B^l)^N \rightarrow R^1$
- H 的平均适配值
$$f(H, t) = \frac{1}{|H \cap P(t)|} \sum_{x \in H \cap P(t)} f(x)$$

$|H \cap P(t)|$ 表示 $P(t)$ 中含有 H 中元素的个数

- 第 t 代种群的平均适配值
$$\bar{f}(t) = \sum_{x \in P(t)} f(x) / N$$

模式定理(续)

模式定理 设 SGA 采用比例选择策略，交叉概率和变异概率分别为 p_c 和 p_m ，且 p_m 取值较小，模式 H 的定义长度为 $\delta(H)$ ，阶为 $o(H)$ ，第 $t+1$ 代种群 $P(t+1)$ 含有 H 中元素个数的期望为 $E[|H \cap P(t+1)|]$ ，则以下不等式成立：

$$E[|H \cap P(t+1)|] \geq |H \cap P(t)| \cdot \frac{f(H,t)}{\bar{f}(t)} \cdot [1 - p_c \frac{\delta(H)}{l-1} - o(H)p_m]$$

- **推论** 在SGA中，阶次低、定义长度短且适配值超过平均适配值的模式在种群中的数目的期望值以指数级递增。
- 模式定理一定意义上解释了SGA的有效性，但还存在以下缺点：
 - 模式定理仅适用于基于二进制编码的SGA，对其它编码方式此定理未必成立；
 - 模式定理仅提供一个期望值的下界，但仍不能说明算法的收敛性；
 - 模式定理对算法参数的选取不能够提供实用的指导，对算法操作也有依赖性。

隐含并行性

隐含并行性定理 设 ε 为一小正数, $l_s < \varepsilon(l-1)+1$, $N = 2^{l_s/2}$, 则 SGA 一次处理的存活概率不小于 $1-\varepsilon$ 且定义长度不大于 l_s 的模式数为 $O(N^3)$ 。

- 此定理说明SGA表面上每代仅对 N 个个体作处理, 但实际上并行处理了大约 $O(N^3)$ 个模式, 并且无须额外的存储, 这正是遗传算法具有高效搜索能力的所在, 即隐含并行性。

遗传算法的收敛性

- 基本数学定义
- SGA的马氏链模型
- SGA的收敛定理
- SGA的收敛速度估计
- 一般遗传算法的收敛性讨论

基本数学定义

定义 2.3.1 称 $n \times n$ 方阵 $A = (a_{ij})$ 为

(1) 非负的(non-negative), 记 $A \geq 0$, 若 $a_{ij} \geq 0$, $i, j = 1, 2, \dots, n$ 。

(2) 严格正的(positive), 记 $A > 0$, 若 $a_{ij} > 0$, $i, j = 1, 2, \dots, n$ 。

(3) 正则的(primitive), 若 $A \geq 0$, 且存在自然数 k , 使 $A^k > 0$ 。

(4) 随机的(stochastic), 若 $A \geq 0$, $\sum_{j=1}^n a_{ij} = 1$, $i = 1, 2, \dots, n$ 。

(5) 归约的(reducible), 若 $A \geq 0$ 且经过相同的行和列初等变换后可转化为 $\begin{bmatrix} C & 0 \\ R & T \end{bmatrix}$ 的形式,

其中 C 和 T 均为方阵。

(6) 不可约的(non-reducible), 若 $A \geq 0$ 且不归约。

(7) 稳定的(stable), 若 A 是随机的且所有行相同。

(8) 列容的(column-allowable), 若 A 是随机的且每一列中至少有一个正数。

马氏链模型

在 SGA 中，令所有个体形成的有限空间为 Ω ，所有种群对应的整个状态空间为 G ，其中每一种群为一个状态，一旦染色体长度 l 和种群数目 N 给定且有限时，则 Ω 的维数 $|\Omega|$ 有限， G 的维数 $|G| = |\Omega|^N$ 也有限。由于算法中每一代状态的转移依赖于选择(复制)、交叉和变异操作，且与进化代数无关，因此 SGA 可视为一个有限状态的齐次马氏链。

鉴于算法中三个遗传操作是循环往返执行的，我们按“交叉→变异→选择”顺序来考虑算法中状态的转移。从而，表征马氏链的状态转移矩阵 P 为矩阵 C 、 M 和 S 的乘积，其中 C 、 M 和 S 分别为交叉、变异和选择操作所决定的状态转移矩阵。

马氏链模型(续)

- 交叉操作

交叉操作可视为状态空间上的随机函数，记交叉操作决定的状态转移矩阵为 $C = (c_{ij})_{|G| \times |G|}$ ，其中 c_{ij} 为从状态 i 经交叉操作转移到状态 j 的概率。由于一个状态通过交叉操作总要转移到状态空间中的另一个状态，因此 $\sum_{j=1}^{|G|} c_{ij} = 1$ 显然成立，即 C 为随机矩阵。

马氏链模型(续)

- 变异操作

在 SGA 中, 变异操作是独立作用于种群内各个体的每一位基因上, 记变异操作决定的状态转移矩阵为 $M = (m_{ij})_{|G| \times |G|}$, 其中 m_{ij} 为从状态 i 经变异操作转移到状态 j 的概率。由于染色体的每个基因具有相同的变异概率 $p_m > 0$, 则 $m_{ij} = p_m^{H_{ij}} (1 - p_m)^{N \cdot l - H_{ij}} > 0$, 其中 H_{ij} 为状态 i 与状态 j 之间具有不同基因的位置的总数。例如, 状态 $\{(1000), (0111), (1010)\}$ 与状态 $\{(1100), (1110), (0101)\}$ 之间的 H_{ij} 为 7。因此, M 为严格正的随机矩阵。

马氏链模型(续)

• 选择操作

在 SGA 中，交叉和变异操作后得到的种群经选择操作后又将转移到另一个种群，记选择操作决定的状态转移矩阵为 $S = (s_{ij})_{|G| \times |G|}$ 。考虑比例选择 (轮盘赌) 策略，种群

中染色体 X_i 被选中的概率为 $\frac{f(X_i)}{\sum_{k=1}^N f(X_k)} > 0$ ，则种群 i 经选择操作后保持不变

的概率 $s_{ii} \geq \prod_{j=1}^N \left(\frac{f(X_j)}{\sum_{k=1}^N f(X_k)} \right) > 0$ 。因此，转移矩阵 S 是随机且列容的。

SGA的收敛定理

定理 2.3.1 若 SGA 中交叉概率 $p_c \in [0,1]$ ，变异概率 $p_m \in (0,1)$ ，并且算法采用比例选择策略，则 SGA 的状态转移矩阵 $P = CMS$ 是严格正的。

定理 2.3.2 SGA 不能够以概率 1 收敛到全局最优解。

注：上述定理从概率意义上说明了 SGA 不能够收敛到全局最优，其原因在于算法中最优解的概率遗失。因此，只要在算法中每代保留当前最优解，无论是在选择之前还是在选择之后，算法将最终收敛到全局最优，从而有以下定理。

定理 2.3.3 每代在选择操作后保留最优解的 SGA 以概率 1 收敛到全局最优解。

定理 2.3.4 每代在选择操作前保留最优解的 SGA 以概率 1 收敛到全局最优解。

保优SGA的收敛速度估计

设带有保优操作的遗传算法对应于有限状态维马氏链 $\{Z_k\}$ ，行向量 π 为给定每一状态的概率， P 为一步转移矩阵，则经 n 次转移后，各状态的概率为 πP^n 。若种群中全部个体相同且均为最优，则称此状态为吸收态。显然，马氏链 $\{Z_k\}$ 中有如下三种性质的状态：

- (1) 吸收态；
- (2) 可一步转移到吸收态的非吸收态；
- (3) 经 P 可一步转移到另一状态，而此状态转移到吸收态的概率为零。从而，一步

转移矩阵 P 可分解为 $P = \begin{bmatrix} I_k & 0 \\ R & Q \end{bmatrix}$ 。其中， I_k 为描述吸收态的 k 阶单位矩阵， R 为描述

能转移到吸收态的状态的 $(|G|-k) \times k$ 阶矩阵， Q 为描述其它状态的 $(|G|-k) \times (|G|-k)$ 阶矩阵。

定理 2.3.5 设 $\|Q\| = \lambda < 1$ ，则概率分布 $z(i)$ 收敛到 z^* 的收敛速度估计为 $\|z(i) - z^*\| \leq O(\lambda^i)$ 。

一般GA收敛性的条件

- 优化空间可测(measurable, 针对连续优化)或有限(finite, 针对离散优化)
- 各遗传操作具有保优性(elitist strategy)
- 对任意初始种群, 最优集在有限步数内可达(accessibility)