

能够寻求到最优样本。而积木块假设却说明了遗传算法的这种能力。

【积木块假设】个体的基因块通过选择、交叉、变异等遗传算子的作用，能够相互拼接在一起，形成适应度更高的个体编码串。

积木块假设说明了用遗传算法求解各类问题的基本思想，即通过基因块之间的相互拼接能够产生出问题更好的解。基于模式定理和积木块假设，就使得我们能够在很多应用问题中广泛地使用遗传算法的思想。

需要说明的是，虽然积木块假设并未得到完整而严密的数学证明，但大量的应用实践说明了其有效性。

6.2.2 遗传算法欺骗问题

提到积木块假设，这里就必须说明一下遗传算法欺骗问题^[75,76] (GA Deceptive Problem)。应用实践表明，存在着一类用遗传算法难以求解的问题，这类称为“GA-难”的问题往往不满足积木块假设，即由基因块之间的拼接，往往会欺骗遗传算法，使其进化过程偏离最优解。

各种研究结果表明，属于“GA-难”的问题一般包含有孤立的最优点，即在这个最优点周围是一些较差的点，从而使得遗传算法较难通过基因之间的相互拼接而达到这个最优点的模式。实际上，目前也尚无解决这类问题的较好的方法或策略。所幸的是，现实所遇到的各种应用问题中，很少有这种奇怪的性质。

6.3 隐含并行性

在遗传算法的运行过程中，每代都处理了 M 个个体，但由于一个个体编码串中隐含有多种不同的模式，所以算法实质上却是处理了更多的模式。

以二进制编码符号串为例，长度为 l 的编码串中隐含有 2^l 种模式，这样，规模为 M 的群体中就可能隐含有 $2^l \sim M \cdot 2^l$ 种

由此我们可以得出如下结论：遗传算法所处理的有效模式总数约与群体规模 M 的立方成正比^[3,5,20]。

也就是说，虽然在进化过程的每一代中只处理了 M 个个体，但实际上我们并行处理了与 M 的立方成正比例的模式数。这种并行处理过程有别于一般意义下的并行算法的运行过程，是包含在处理过程内部的一种隐含并行性（Implicit Parallelism）。通过这种隐含并行性，使得我们可以快速地搜索出一些比较好的模式。

6.4 遗传算法的收敛性分析

模式定理虽然定量地估算出了具有较优结构特点的模式在进化过程中的增长规律，但并未导出遗传算法能够收敛于问题最优解的概率。在遗传算法的进化过程中，个体集合一代一代地变化着，若把每一代群体看作为一种状态的话，则可以把整个进化过程作为一个随机过程来加以考察，并可利用 Markov 链来对进化过程进行理论分析，从而得到遗传算法收敛性方面的重要结论。

6.4.1 Markov 链

这里先定义几个随机过程中的术语^[77]。

【定义 6.4】 设随机过程 $\{X(n), n \geq 0\}$ 只能取可列个值 $I = \{i_0, i_1, \dots\}$ ，并且满足条件：对任意 n 及 i_0, i_1, \dots, i_n ，如果

$$P\{X(0) = i_0, X(1) = i_1, \dots, X(n) = i_n\} > 0 \quad (6-10)$$

必有

$$\begin{aligned} P\{X(n+1) = i_{n+1} | X(0) \\ = i_0, X(1) = i_1, \dots, X(n) = i_n\} \\ = P\{X(n+1) = i_{n+1} | X(n) = i_n\} \end{aligned} \quad (6-11)$$

则称 $\{X(n), n \geq 0\}$ 为时间离散状态离散的 Markov 链，简称 Markov 链。

【定义 6.5】 称 $P\{X(n)=j|X(m)=i, n>m\}$ 为 Markov 链的转移概率, 记为 $P_{ij}(m, n)$ 。

$P_{ij}(m, n)$ 具有下面几条性质:

$$(1) P_{ij}(m, n) \geq 0 \quad (6-12)$$

$$(2) \sum_{j \in I} P_{ij}(m, n) = 1 \quad (6-13)$$

【定义 6.6】 对于 Markov 链, 如果

$$\begin{aligned} P_{ij}(m, m+1) &= P\{X(m+1)=j|X(m)=i\} \\ &= P_{ij} \quad (i, j \in I) \end{aligned} \quad (6-14)$$

即从状态 i 出发转移到状态 j 的转移概率与时间起点 m 无关, 则称这类 Markov 链为齐次 Markov 链。

【定义 6.7】 对于齐次 Markov 链, 称 P_{ij} 为一步转移概率, 全部 $P_{ij} (j, i \in I)$ 所组成的一个矩阵 $P = (P_{ij})$ 称为一步转移概率矩阵, 或称为随机矩阵。

6.4.2 遗传算法的收敛性分析

为简单起见, 我们只对基本遗传算法的收敛性进行分析。

基本遗传算法可描述为一个齐次 Markov 链 $P_t = \{P(t), t \geq 0\}$, 因为基本遗传算法的选择、交叉和变异操作都是独立随机进行的, 新群体仅与其父代群体及遗传操作算子有关, 而与其父代群体之前的各代群体无关, 即群体无后效性, 并且各代群体之间的转换概率与时间的起点无关。

【定理 6.1】 基本遗传算法收敛于最优解的概率小于 1。

证明 将群体的各种可能状态 I 分为包括最优个体的状态 I_o 和不包括最优个体的状态 I_n :

$$I = I_o \cup I_n \quad (I_o \cap I_n = \phi)$$

本定理是要证明 P_t 进入 I_o 状态的稳定概率小于 1。

用反证法。假设基本遗传算法能收敛于最优解的概率等于 1, 则进入 I_n 状态的稳定概率应等于 0, 即:

$$\lim_{t \rightarrow \infty} P\{P_t \in I_n\} = 0 \quad (6-15)$$

在基本遗传算法的进化过程中, 群体从某一状态 $i \in I$, 经

过选择、交叉和变异算子的连续作用而转变为状态下 $j \in I$ 。这三种遗传算子的转移概率分别为 s_{ij} 、 c_{ij} 、 m_{ij} ，它们可分别构成相应的随机矩阵 $S = \{s_{ij}\}$ 、 $C = \{c_{ij}\}$ 、 $M = \{m_{ij}\}$ ，则遗传算法的群体状态变换矩阵为： $R = SCM = \{r_{ij}\}$ 。

由于 S 、 C 、 M 都是随机矩阵，并且 $m_{ij} = P_m^{H(i,j)} (1 - p_m)^{1-H(i,j)} > 0$ ($H(i, j)$ 为状态 i 和状态 j 之间的海明距离)，容易证得 $r_{ij} > 0$ ，即 R 是正定的。

在第 t 时刻，群体是状态 j 的概率 $P_j(t)$ 为：

$$P_j(t) = \sum_{i \in I} P_i(0) r_{ij}^t \quad (t = 0, 1, 2, \dots) \quad (6-16)$$

由齐次 Markov 链的性质可知， $P_j(t)$ 的稳定概率分布与其初始概率分布无关^[77]，即有：

$$P_j(\infty) = \sum_{i \in I} P_i(\infty) r_{ij} > 0 \quad (6-17)$$

注意到上式中的状态 $j \in I$ ，即 j 也可能是 I_q 中的一个状态，从而可知：

$$\lim_{t \rightarrow \infty} P\{P_t \in I_n\} > 0 \quad (6-18)$$

上式与式 (6-15) 的假设相矛盾，从而定理得证。

显然，对于这种收敛于最优解的概率小于 1 的基本遗传算法，其应用可靠性就值得怀疑。从理论上来说，仍希望遗传算法能够保证收敛于最优解，这就需要对基本遗传算法进行改进，如使用保留最佳个体的策略就可达到这个要求。

【定理 6.2】 使用保留最佳个体策略的遗传算法能收敛于最优解的概率为 1。

证明 考察这样所组成的个体集合 $P^+(t) = (A(t), P(t))$ ，其中 $A(t)$ 是当前群体中适应度最高的个体。这一个体集合的状态转移规则是^[78]：

(1) 依据定理 6.1 中的状态转移矩阵 R 由 $P(t)$ 产生 $P(t+1)$ 。

(2) $A(t+1)$ 是从上一代群体中和本代群体中挑出的一个

具有最大适应度的个体, 即:

$$A(t+1) = \max \{A(t), A_o\} \quad (A_o \text{ 是群体 } P(t+1) \text{ 中适应度最高的个体})$$

这样所构造出的随机过程 $\{P^+(t), t \geq 0\}$ 仍然是一个齐次 Markov 链, 即有:

$$P^+(t) = P^+(0) (R^+)^t \quad (6-19)$$

假设个体集合状态中包括有最优解的状态为 I_o , 则该随机过程的状态转移概率为:

$$r_{ij}^+ > 0 \quad (\forall i \in I, \forall j \in I_o) \quad (6-20)$$

$$r_{ij}^+ = 0 \quad (\forall i \in I_o, \forall j \notin I_o) \quad (6-21)$$

即从任意状态向含有最优解的状态转移的概率大于 0, 而从含有最优解的状态向不含有最优解的状态转移的概率等于 0。

此时, 对于 $\forall i \in I, \forall j \notin I_o$, 下式都成立:

$$(r^+)^t_{ij} \rightarrow 0 \quad (t \rightarrow \infty) \quad (6-22)$$

$$P_j^+(\infty) = 0 \quad (j \notin I_o) \quad (6-23)$$

亦即个体集合收敛于不含有最优解的状态的概率为 0。换句话说, 算法总能够以概率 1 找到最优解。

定理 6.2 说明了这种使用保留最佳个体策略的遗传算法总能够以概率 1 搜索到最优解。这个结论除了理论上具有重要意义之外, 在实际应用中也为最优解的搜索过程提供了一种保证。

6.5 适应度函数的自相关分析

经过改进的遗传算法虽然能够以概率 1 收敛于问题的最优解, 但若这个收敛过程进行得比较缓慢的话, 也会使其毫无应用价值。所以从应用的角度来说, 需要定量分析求解效率和解的质量, 往往还需要在求解效率和解的质量之间达到一种平衡。

6.5.1 适应度函数的景象

前面对遗传算法的理论分析是基于模式的概念来进行的, 未

涉及到适应度函数。但是，遗传算法的运行过程中毕竟主要是依据个体的适应度来进行优胜劣汰操作的，所以有必要对适应度函数的特性进行研究。

若某一个体 A_i 在一系列遗传算子的作用下被转化为另一个不同的个体 A_j ，那么在解空间中这两个个体之间就有一种很自然的邻接关系。这样，在解空间中就可以定义一种邻接结构。例如，变异算子使某一个体的某一基因座上的基因值反转，由此而产生出一个新的个体，由新旧两个个体之间的海明距离就可在解空间中定义出一种距离空间结构。同样，其他种类的遗传算子也可在解空间中定义出其他类型的邻接空间结构。由此看来，解空间就不仅仅是表示可行解的一系列点的集合，与解空间中的邻接结构相对应，各点都有不同的适应度，这样，在解空间中也可定义一种适应度分布函数，从而构成了一种适应度函数的景象 (Fitness Landscape)^[79]。

之所以研究适应度函数的景象，是因为适应度函数的景象与最优化的难易程度密切相关。有些最优化问题，在所考虑的邻接结构下其适应度函数的景象凸凹不平，有很多局部最优解，如图 6-4 所示，这类函数优化起来就比较困难。而另一些最优化问题，其适应度函数的景象起伏一致，如图 6-5 所示，这类函数优化起来就比较方便。

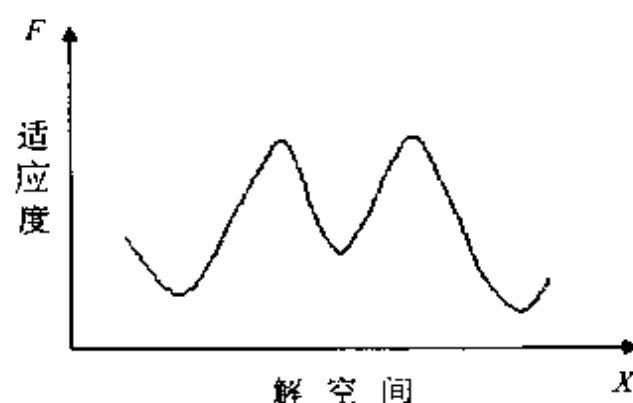


图 6-4 凸凹型适应度函数景象

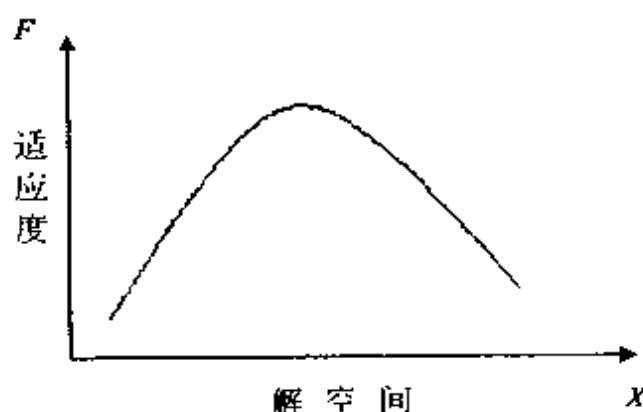


图 6-5 山峰型适应度函数景象

6.5.2 适应度函数的自相关函数

从解空间中的某一点 x 出发, 作向其邻接点的随机漫游 $\{x_i\}$, 并确定各漫游点的适应度 $\{F(x_i)\}$ 。若以适应度 F 为随机变量, 则该随机漫游过程中所得到的适应度函数的自相关函数可定义为^[80]:

$$\rho(s) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{[F(x_{i+1}) - \bar{F}][F(x_i) - \bar{F}]}{\sigma(F)^2} \quad (6-24)$$

式中, \bar{F} 表示随机变量 F 的平均值; $\sigma(F)$ 表示随机变量 F 的标准偏差。

随机漫游的观测结果表明, 对于很多组合优化的问题, 其适应度函数的自相关函数是以某一相关长度 l_c 为参数而指数衰减的, 即:

$$\rho(s) \approx \exp(-s/l_c) \quad (6-25)$$

其中的相关长度 l_c 是该适应度函数的一个重要特征参数。

对于图 6-4 所示的具有凸凹型景象的适应度函数, 相关长度 l_c 就意味着当前搜索点的适应度函数值与相关长度以前的点的适应度函数值无关, 此时它相对于相关长度以前的点来说, 是呈现出一种随机摆动的趋势。对这类函数进行优化计算时, 最少应把可行解区域按相关长度 l_c 分割为一些小的区域, 各个区域分别

进行求解，并且相关长度越短，需分割的份数越多。而对于图 6-5 所示的具有山峰型景象的适应度函数，可认为其相关长度很大，这样，邻近点之间的适应度有一定的关系，若当前搜索点的适应度较大的话，则其邻近点的适应度也会较好。

6.5.3 遗传算子的相关系数

遗传算法中，一个或多个父代个体通过遗传算子的作用可生成一个或多个子代个体，这个遗传算子的相关系数可定义为：

$$\rho(F_p, F_c) = \frac{Cov(F_p, F_c)}{\sigma(F_p) \sigma(F_c)} \quad (6-26)$$

式中， F_p 和 F_c 分别为父代个体和子代个体的适应度或其平均值； $Cov(F_p, F_c)$ 是父代个体和子代个体适应度之间的协方差。

上式定义的相关系数的含义是：相关系数的绝对值越小，该遗传算子产生新个体的能力越接近于随机搜索算法产生新个体的能力。由此可见，相关系数是评价遗传算子搜索能力的一种有效标准。