

## 4 选择

在第一节“概述”中提到在进化算法中存在两个阶段的选择。第一次是参与进化操作的个体的选择。这个阶段的选择可以是基于个体适应度的、也可以是完全随机地选择交配个体。一旦个体被选中，那么它们就会参与交叉、变异等进化操作。未被选中的个体不会参与到进化操作中。

第二次是常被称为“重插入”或“环境选择”的选择，它是指在个体经过交叉、变异等进化操作所形成的子代（或称“育种个体”）后用某种方法来保留到下一代从而形成新一代种群的过程。这个选择过程对应的是生物学中的”自然选择”。它可以是显性地根据适应度（**再次注意**：适应度并不等价于目标函数值）来进行选择的，也可以是隐性地根据适应度（即不刻意去计算个体适应度）来选择。例如在多目标优化的 NSGA-II 算法 [1] 中，父代与子代合并后，处于帕累托分层中第一层级的个体以及处于临界层中的且拥挤距离最大的若干个个体被保留到下一代。这个过程就没有显性地去计算每个个体的适应度。

重插入方案有以下两种：

### 1) 全局重插入 (Global reinsertion):

这种插入算法适用于在第一次选择阶段使用除了本地选择外的其他选择算法的情况。假如在第一次选择阶段使用了本地选择算法，那么重插入中就不能用全局重插入算法了。

- 育种个体与父代个体一样多，直接用所有的育种个体生成新一代种群 (完全重插入)。
- 育种个体比父代个体少，把育种个体随机均匀地替换父代的个体 (均匀重插入)。
- 育种个体比父代个体少，取代适应度较低的父代个体 (精英重插入)。
- 育种个体比父代个体多，重新插入适应度较高的育种个体 (精英保留重插入)
- 父子两代合并的竞争选择，即在父代个体和育种个体同时并存的情况下从两代种群的所有个体中选择合适数量的个体保留到下一代。
- 一对一生存者竞争选择。

完全重插入是最简单的重插入算法。最经典最简单的遗传算法 (Simple GA) 用的就是这种重插入方法。然而最坏的情况是：父代中最优秀的个体并没有繁殖产生优秀的个体，此时在重插入时又被最差的个体替代，因此种群丢失了精英，使得算法的收敛能力

强。

因此在多数情况下，推荐使用精英重新插入、精英保留重新插入或是父子两代合并竞争选择等带精英策略的方法。

### 2) 本地重插入 (Local reinsertion):

在前面章节中介绍过本地选择算法，它在有界领域中选择个体。那么育种后代的重插入也应该发生在相同的邻域中。这是一种基于位置信息的重插入算法，重插入时使用的邻域要跟本地选择中的一样。

以下是可行的：

- 把育种个体随机均匀地替换其所在邻域内的父代个体。
- 把育种个体替换其所在邻域内的最差的父代个体。
- 把更适合插入到相邻邻域的育种个体替换相邻邻域中的最差的父代个体。
- 把更适合插入到相邻邻域的育种个体随机均匀地替换相邻邻域中的父代个体。

下面是一些选择操作有关的术语：

**选择压力 (selective pressure):** 与所有个体的平均选择概率相比，最优个体被选中的概率。

**偏差 (bias):** 个体标准化的适应度预期期望再生概率之差的绝对值。

**个体扩展 (spread):** 在进行选择时，比较优秀个体可能会被多次选择。该参数就限制了最多可以被重复选择的个数。

**多样性损失 (loss of diversity):** 进行选择操作后，未被选择的个体数目占种群个体总数的比例。

**选择强度 (selection intensity):** 把标准高斯分布用在选择操作后的种群个体平均适应度的期望值。选择强度越大，进化算法的最优化搜索的收敛速度就越快，但也往往意味着多样性的损失。

**选择方差 (selection variance):** 把标准高斯分布用在选择操作后的种群个体适应度的方差。

选择操作一般是基于个体的适应度来计算的 (详见上一节)，下面介绍一些经典的选择算子：

### 1) 轮盘赌选择 (Roulette wheel selection):

轮盘赌选择是一种有回放的随机抽样选择法。种群的个体被映射到区间的连续片段，每个个体所在片段的长度与其适应度成比例。生成随机数，根据其所落在的片段选择对应的个体，并重复该过程直到获得所需数量的个体。

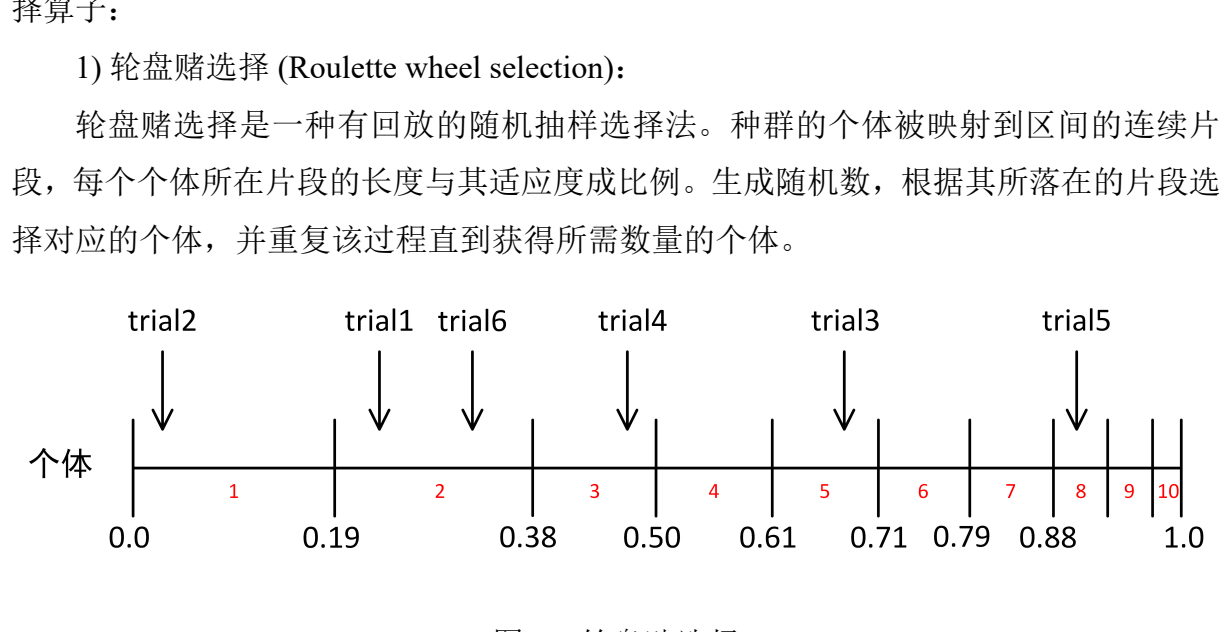


图 1 轮盘赌选择

所选择的个体为：2, 1, 5, 3, 8, 2。

轮盘赌选择算法实现了零偏差 (bias) 但不保证最小个体扩展 (spread)，因此，上面的例子中，2 号个体被选择了两次 (有可能所选的个体全部都是同一个)。

### 2) 随机抽样选择 (Stochastic universal sampling):

随机抽样选择实现了零偏差同时保证最小个体扩展。种群的个体被映射到区间的连续片段，每个个体所在片段的长度与其适应度成比例 (这里跟轮盘赌是相同的)。用一组等距离的指针随机地指向上述区间，每个指针所指的个体即为要选择的个体。指针间的距离是  $1/N_{sel}$  ( $N_{sel}$  为要选择的个体数)，第一个指针的位置由  $[0, 1/N_{sel}]$  范围内的随机数给出。

假设要选择 6 个个体，那么指针间的距离则为  $1/6 \approx 0.167$ ，选择情况如下图：

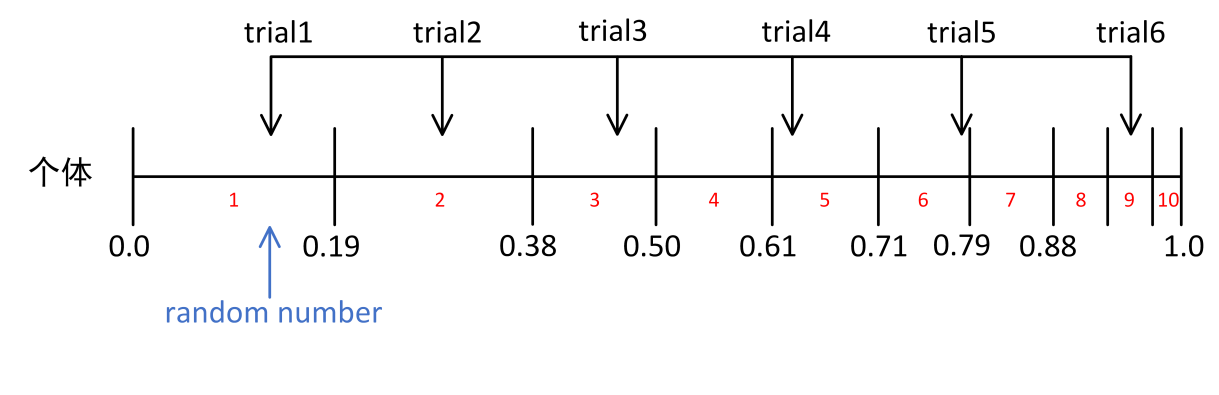


图 2 随机抽样选择

所选择的个体为：1, 2, 3, 5, 6, 9。

轮盘赌选择和随机抽样选择都属于轮盘赌模型 (或称基于适应度比例的选择模型)，但后者通常能够得到更想要的结果，种群多样性也得以保证。

某些人认为随机抽样选择是不考虑个体的适应度、纯粹依靠随机数的结果来进行个体选择，这种想法是极其错误的。这种选择方法实际上会让进化算法变为盲目随机搜索，而无法发挥适应度应该起到的指导搜索方向的作用。

### 3) 锦标赛选择 (Tournament selection):

该选择算法模仿淘汰赛制，每次从种群中挑选  $Tour$  个个体进行比较，从中选出一个最好的个体加入被选集合。重复该操作，直到被选集合的大小达到需要选择的个体数。

$Tour$  即为竞赛规模。在文献 [2] 中，可以看到锦标赛选择的相关理论分析。

选择强度：

$$SelInt(Tour) \approx \sqrt{2 \left( \ln(Tour) - \ln \left( \sqrt{4.14 \cdot \ln(Tour)} \right) \right)}$$

多样性损失：

$$LossDiv(Tour) = Tour^{-\frac{1}{Tour-1}} - Tour^{-\frac{Tour}{Tour-1}}$$

选择方差：

$$SelVar(Tour) \approx \frac{0.918}{\ln(1.186 + 1.328 \cdot Tour)}$$

### 4) 截断选择 (Truncation selection):

截断选择是与前面的几种选择算法比较不一样的算法。它更接近于生物学里的”人工育种”，适用于大规模种群的个体选择。

在截断选择中，根据适应度对种群个体进行分类。设定截断阈值为  $Trunc$  (表示选择作为父母的个体的比例)，低于阈值的个体不会产生后代。”选择强度”这个术语就经常用在截断选择中，”选择强度”和截断阈值之间的关系非常密切。

在文献 [1] 中可以看到截断选择的相关理论分析。

### 5) 本地选择 (Local selection):

在本地选择算法中，每个个体处在一个约束环境中 (称为”本地邻域”)。个体仅与同一邻域内的个体相互作用。邻域是根据种群的结构来定义的，可以是线性的、二维的或者是三位及更加复杂的。邻域之间的距离决定了邻域的大小，而邻域的大小决定了种群个体之间遗传信息的传播速度，即决定了种群是快速繁殖还是维持高度的种群多样性。[VBS91] 的模拟中得出了类似的结果，在较小的邻域中进行的本地选择比在较大的邻域中要好。

选择算法依赖于适应度的计算，不同的适应度计算方法会使得种群个体的适应度呈现不一样的分布特征，从而影响到选择操作的效果。

### 6) 基向量选择:

在差分进化算法中，差分进化操作是由差分变异基向量  $X_{r0}$  和差分向量  $X_{r1}$  以及  $X_{r2}$  共同作用得到。其中差分变异基向量的索引  $r_0$  是从个体索引序列  $[0, 1, 2, \dots, N-1]$  中抽选出来的。在差分变异前一共要抽选出与种群个体数等量的基向量。因此在这种情况下下的选择会跟前面的选择方法有一个不同之处：它不能控制选择的数目。除了可以直接使用前面提到的轮盘赌选择、随机抽样选择以及锦标赛选择等选择算法来得到基向量的索引序列之外，在差分进化中还常使用精英个体选择以及随机补偿选择两种选择方式：

#### A. 精英个体索引复制:

它将根据种群个体的适应度得到精英个体的索引，然后复制该索引若干份直至达到所需选择的个体数，最后返回一个全是精英个体索引的序列。

#### B. 随机补偿选择:

在差分进化算法中，非常鼓励索引值互异，即各个索引在相同位置上的序号与原种群个体的索引序列是不一致的。例如基向量索引的第一位不能为 0，第二位不能为 1，第三位不能为 2……以此类推 (索引序列从 0 数起)。这样能够避免差分进化发生退化现象。因此随机补偿法根据以下公式产生基向量索引  $r_0$ ：

$$r_0 = (i + r_g) \% N$$

其中  $N$  为种群的个体数， $i$  为种群个体索引， $i = 0, 1, 2, \dots, N$ ， $r_g$  为  $[1, N-1]$  之间的整数。它在每一代的进化中会被重新生成一次。根据该式得到的  $r_0$  可以保证与目标索引  $i$  互异。

### 7) 一对一生存者竞争选择:

这种选择算法需要两个种群。两个种群的每个个体只与在另一个种群中相同位置的个体进行适应度比较，适应度高者被选中。当然这个算法可以扩展到多个种群。或者是可以把一个大种群等分为若干个小个子种群，让子种群的个体参与一对一生存者竞争选择。如图所示：

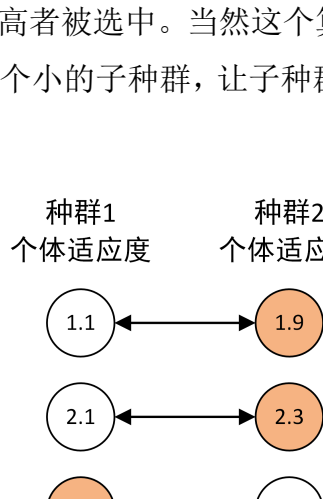


图 3 一对一生存者竞争选择

图中有两个种群，圆圈内标注的是个体的适应度，橙色的便是根据一对一生存者竞争选择的被选中的个体。

## 参考文献

[1] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2):0-197.

[2] Blikle T, Thiele L. A Comparison of Selection Schemes Used in Evolutionary Algorithms[J]. Evolutionary Computation, 1996, 4(4):361-394.