文章编号:1001-5132（2009）03-0374-04

选择算子与遗传算法的计算效率分析

张松艳

（宁波大学 理学院，浙江 宁波 315211）

摘要：指出了遗传算法的选择方式与其全局收敛性和收敛速度的关系. 常用的选择算子不能保证算法的全局收敛性，在引入改进选择策略后轮盘赌选择方式能保证算法的全局收敛性，但收敛速度较慢. 同时给出了遗传算法选择操作的若干策略.

关键词：选择算子；遗传算法；计算效率；收敛性定理

中图分类号：TP301.6 **文献标识码：**A

遗传算法(Genetic Algorithm)是一种模拟自然界生物进化过程中“优胜劣汰”选择机制的随机搜索算法，它将问题的求解表示成“染色体”的适者生存过程，通过种群的一代代不断进化，包括选择、交叉和变异等操作，最终收敛到“最适应环境”的个体，从而求得问题的最优解或者满意解. 其自组织、自适应、自学习和种群进化能力使其适合于大规模复杂优化问题，应用于全局搜索等参数优化计算领域. 随着计算机技术的发展，遗传算法越来越受到人们的重视，并在机器学习、模式识别、神经网络、优化控制、组合优化等领域得到了成功的应用.

选择操作是遗传算法中环境对个体适应性的评价方式，也是实现群体优良基因传播的基本方式. 遗传算法使用选择、交叉、变异等遗传算子繁殖后代，交叉和变异算子用于产生新的个体，而选择算子保证了 GA 迭代中“适者生存”的群体进化现象，体现了群体中个体求同的意向，在很大程度上决定了 GA 收敛的效果和速度. 个体适应值是选择算子运行的基础，而选择算子直接影响着 GA 的性能，所以合理确定目标函数值到适应值的映射关系具有重要意义.

1 遗传算法的选择算子

选择即从当前群体中选择适应值高的个体以生成交配池的过程. 遗传算法中最常用的选择方式是轮盘赌(Roulette Wheel)选择方式，也称比例选择或复制. 在该方法中，各个个体被选择的概率和其适应度值成比例. 设群体规模大小为 N，个体 i 的适应度值为 f_i，则这个个体被选择的概率为:

$$p_{si} = f_i / \sum_{i=1}^{N} f_i .$$

显然，个体适应度越大，其被选择的概率越高，反之亦然.

遗传算法另一种常用的选择方式是联赛选择方式，其基本思想是将上一代群体中的个体和本次遗传操作产生的所有新个体放到一起按适值从

收稿日期：2008-11-06. 宁波大学学报（理工版）网址: http://3xb.nbu.edu.cn
基金项目：国家自然科学基金（10771110, 10471069）.
作者简介：张松艳（1966－），女，浙江舟山人，副教授，主要研究方向：遗传算法. E-mail: zhangsongyan@nbu.edu.cn

大到小的顺序排队, 然后取排在前面的 N 个(N 为群体规模)个体组成新一代群体.

遗传算法的交叉算子作用于某 2 个父代个体时, 会产生 2 个子代个体, 父子 2 代共 4 个个体平等竞争, 淘汰 2 个低适值个体, 保留 2 个高适值个体. 遗传算法的变异算子作用于某一父代个体时, 会产生一个子代个体, 如果子代个体的适值比父代个体的高, 则用子代个体取代父代个体; 否则保留父代个体淘汰子代个体, 这就是父子竞争选择.

遗传算法初始群体中的个体一般是随机产生的, 初始群体中的个体均匀地分布于整个串空间. 在遗传迭代的早期, 群体中个体适值差别很大, 按上述 3 种选择方式容易出现的问题是: 在选择下一代群体时, 适值低的个体被选中的机会很小, 最佳个体在下一代的生存机会将显著增加, 而最差个体的生存机会将被剥夺, 低适值个体淘汰太快容易使算法收敛于局部最优解. 群体中的最佳个体快速充满整个群体, 导致群体多样性降低, GA 也过早地丧失了进化能力. 而到了遗传迭代的晚期, 群体中个体适值差别不大, 算法收敛速度慢. 此外, 遗传算法只有在引入了最优保持操作后才是全局收敛的. 因此, 我们提出改进的选择策略, 先对群体中个体的适值进行变换, 再按个体适值大小的比例进行选择. 具体方法是: 先将参与选择的 X 个个体按适值从小到大顺序编号(相同适值的个体可随意排列), 然后以个体的序号作为其变换后的适值, 即 X 个个体的适值分别变换为 $1, 2, 3,\cdots, X$. 编号为 m 的个体被选中的概率为 $p=m/X$, $1\leqslant m\leqslant X$. 显然, 这种改进的选择与个体的适应值无直接关系, 仅仅与个体之间的适应值相对大小有关. 这种策略一方面通过对群体中个体适值的变换, 使群体中的个体在遗传迭代的整个过程中都能保持良好的多样性, 既保证了算法具有较快的收敛速度, 又能防止算法收敛于局部最优解; 另一方面能使上一代的最优个体一定会被选择到下一代, 即这种选择策略隐含了最优保持操作, 保证了算法的全局收敛性. 由于选择概率比较容易控制, 所以适用于动态调整选择概率, 根据进化效果适时改变群体选择压力.

2　选择算子对遗传算法收敛性的影响

由文献[1]可知, 不论群体初始分布如何, 马尔可夫链中的任何状态均有一大于零的唯一的极限分布, 从任意状态 i 出发, 可以在有限时间内到达任意状态 j, 自然 $k\to\infty$, 简单遗传算法(SGA)能遍历状态空间, 但这不意味着 SGA 能收敛于全局最优解.

下面给出全局收敛概念.

定义 1　令 $F_k=\max\{f(s_{ik_j}\,|\,j\in[1,n])\}$ 是时刻 k, 状态为(群体) λ_i 时, 群体中的最大适应度. 令 $F^*=\max\{f(s^j\,|\,j\in[1,2^{nl}])\}$ 是 GA 所求问题的全局最优适应度, 当且仅当条件 $\lim\limits_{k\to\infty}P(F_k=F^*)=1$ 成立时, GA 是全局收敛的.

定理 1　简单遗传算法(SGA)不是全局收敛的.

证明　设 λ_i 是满足 $F_k<F^*$ 的一个任意状态, 并设 P_{ik} 是 SGA 在时刻 k 处于状态 λ_i 时的概率, 易知: $P_{ik}=1/2^{nl-1}$, 若在状态空间中满足 $F_k<F^*$ 的状态数为 m, 则 $m\geqslant 1$, 所以

$$P(F_k<F^*)=m/2^{nl-1}\geqslant 1/2^{nl-1}=P_{ik}.$$

显然满足 $F_k=F^*$ 状态的概率为:

$$P(F_k=F^*)\leqslant 1-P_{ik}.$$

由马尔科夫链的基本极限定理及文献[1]可知, $\lim\limits_{k\to\infty}P_{ik}>0$. 因此, $\lim\limits_{k\to\infty}P(F_k=F^*)\leqslant 1-\lim\limits_{k\to\infty}P_{ik}<1$.

由全局收敛概念可知, SGA 不是全局收敛的.

定义 2　设 λ_k^* 是 SGA 马尔科夫链在 k 时刻具有最优适应度 F_k^* 的状态, 若 $F_{k+1}>F_k^*$, 则令 $\lambda_{k+1}^*=\lambda_{k+1}$, $F_{k+1}^*=F_{k+1}$, 以上操作称为最优保存操作, 加入最优保存操作的 SGA 称为最优保存 SGA, 简称为 OMSGA(Optimum Maintaining Simple Genetic Algorithm).

定理 **2** 最优保存简单遗传算法(OMSGA)是全局收敛的[2].

证明 设 Λ_0 是满足 $F_0 = F^*$ 状态的集合, 转移矩阵 P' 扩展为:

$$P' = \begin{bmatrix} Q & 0 \\ T & P \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ T & P \end{bmatrix},$$

P' 中 Q 是一个封闭类, 是 Λ_0 的转移矩阵, Q 只有一个元素 1, 即它是自吸引的, 这与最优保存操作的概念是一致的. P 是过渡类, 设 λ_i 是属于 Λ 的任意状态, 由于 SGA 构成的马尔科夫链是遍历的, 不论初始分布如何, λ_i 经有限步后总能到达 Λ_0, 也即能到达封闭类, 因此, $\lim\limits_{k\to\infty} P(F_k = F_u = F^*) = 1$. 所以, OMSGA 是全局收敛的.

在假定上一代群体中的所有个体被选择到下一代的概率都大于 0 时, 遗传算法在引入了最优保持操作后是全局收敛的. 按个体适值的比例进行选择, 在任何个体的适值都大于 0 的情况下, 任何个体被选中的概率也大于 0. 如果在按个体适值的比例进行选择的操作过程中引入最优保持操作, 遗传算法能收敛于全局最优解. 联赛选择方式和父子竞争选择方式虽然隐含了最优保持操作, 但不能保证上一代群体中的个体被选择到下一代的概率大于 0, 因此遗传算法的全局收敛性问题有待进一步研究. 改进选择策略既能保证上一代群体中的任何个体被选择到下一代的概率大于 0, 又隐含了最优保持操作, 因此保证了算法的全局收敛性, 并使算法具有较快的收敛速度.

3 选择算子对算法计算效率的影响

初始群体随机产生时, 初始群体中存在最优个体的概率为:

$$p_0 = 1 - (1 - r/s)^N,$$

这里 s 为个体空间的大小, r 为个体空间中最优个体的数目, N 为群体规模. 一般情况下 $r << s$, N 为有限数, 因此 $p_0 \approx 0$, 即在初始群体中几乎不存在最优个体, 因此必须借助于遗传算子的作用, 经过多次遗传迭代, 才可能产生最优个体. 遗传算法主要通过交叉算子繁殖后代. 对第 t 代群体中的 N 个(假定 N 为偶数)个体进行交叉操作能生成个体的平均数目为 $N_t' = Np_e$, 其中 p_e 为交叉算子的作用概率.

考虑到生成的个体可能重复, 对第 t 代群体中的 2 个体进行交叉操作后能产生新个体的数目为:

$$N_t = p_e g_1 g_2 N, \tag{1}$$

这里 g_1 为第 t 次遗传迭代生成的个体中彼此不重复的概率, g_2 为第 t 次遗传迭代生成的个体与历代群体中的个体不重复的概率. 第 t 次遗传迭代过程中能生成最优个体的概率为:

$$p_t = (1 - (1 - r/s)^{N_t})c_t, \tag{2}$$

其中: c_t 为考虑选择算子的作用后一次遗传迭代比前一次遗传迭代更容易生成最优个体的比率.

由(1)式和(2)式可得:

$$P_t = (1 - (1 - r/s)^{p_e g_1 g_2 N})c_t. \tag{3}$$

为了使遗传迭代能尽快生成最优个体, 必须尽量提高 p_t. 从(3)式可以看出, 增大 p_e、g_1、g_2、N 和 c_t 都能提高 p_t, 其中 p_e 一般接近于 1. g_1、g_2 会随遗传迭代的进行而下降, 所以要提高 p_t 主要靠增大群体规模 N 和 c_t, 而增大 N 会增大每一代的计算时间, 过大的 N 对提高算法的计算效率并不利, 因此增大 c_t 对提高算法的计算效率非常重要, 而 c_t 的大小主要取决于选择方式. 联赛选择方式和父子竞争选择方式都有较大的 c_t, 但这 2 种选择方式不能保证上一代群体中的个体都有机会被选择到下一代, 算法的全局收敛性有待进一步研究, 即这 2 种选择方式尽管有较快的收敛速度, 但很可能是收敛于局部最优解; 轮盘赌方式在引入了最优保持操作后能保证算法的全局收敛性, 但只有较小的 c_t, 即这种选择方式的收敛速度较慢.

作者给出了遗传算法的 3 种常用的选择方式, 即轮盘赌选择方式、联赛选择方式和父子竞争选择方式, 前一种选择方式在引入了最优保持操作后

能保证算法的全局收敛性，但收敛速度较慢；后 2 种选择方式不能保证算法的全局收敛性，很可能收敛于局部最优解，但有较快的收敛速度．因此，适当选择遗传算法的选择方式对提高算法的计算效率很有意义．

以上讨论的 GA 选择操作是定向选择．定向选择方式保证了群体的进化收敛性，但也导致了 GA 早熟或者欺骗问题[3]，为了克服 GA 进化过程中的这种现象，需要改变传统的选择策略，这里提出以下策略：

第一，在进化初始阶段，可以保持较大的群体规模和较小的选择压力，充分发挥交叉算子的模式重组能力，避免产生模式欺骗问题，或者过早地陷于局部最优解，最大程度上扩大 GA 可以搜索到的问题的可行解子空间．

第二，在进化中期阶段，可以采用较小的群体规模和较大的选择压力，以及较高的变异概率；或者采用较大的群体规模和较小的选择压力；或者采用新的适应值标度变换(非线性变换)，通过交叉算子克服可能存在的模式欺骗性，或者突破局部最优解的限制．

第三，在进化后期阶段，采用较小的群体规模和较大的选择压力，促进 GA 局部搜索，尽快完成当前最优解的求解．

参考文献：

[1] 恽为民，席裕庚．遗传算法的全局收敛性和计算效率分析[J]．控制理论与应用, 1996, 13(4):456-460.

[2] 李茂军，童调生．单亲遗传算法及其全局收敛性分析[J]．自动化学报, 1999, 25(1):68-72.

[3] 李敏强，寇纪淞，林丹，等．遗传算法的基本理论与应用[M]．北京：科学出版社, 2002.

Operator Selection and Computational Efficiency Analysis on Genetic Algorithm

ZHANG Song-yan

(Faculty of Science, Ningbo University, Ningbo 315211, China)

Abstract: The relationship between the selection of genetic algorithm and its global convergence and convergence rate is identified. It is found that the commonly used selection operator can not guarantee the global convergence in the algorithm. The global convergence is guaranteed using roulette wheel selection algorithm through introducing the optimal maintaining operation, although the convergence is achieved at the cost of slight time consumption. In addition, several tactics for selecting and running the genetic algorithm are presented in this work.

Key words: selection operator; genetic algorithm; computational efficiency; convergence theorem

CLC number: TP301.6　　　**Document code:** A

（责任编辑　史小丽）