

遗传算法的全局收敛性和计算效率分析*

恽为民 席裕庚

(上海交通大学自动化系·上海, 200030)

摘要: 本文应用齐次有限马尔科夫链分析了简单遗传算法、最优保存简单遗传算法和自适应遗传算法的收敛性, 然后对计算效率进行了定性分析, 得到了指导基因操作策略设计的极限分布概率原则。

关键词: 全局收敛性; 计算效率; 简单遗传算法(SGA); 最优保存简单遗传算法(OMSGA); 自适应遗传算法(AGA)

1 引言

遗传算法是一种自适应启发式群体型概率性迭代式全局收敛算法, 来源于进化论和群体遗传学。遗传算法是全局收敛的这一定性结论主要根据 Holland 的模式定理的定性分析^[1], 事实上这一结论普遍受到怀疑和争论。近几年, 在遗传算法全局收敛性的分析方面取得了突破, 运用的数学工具是齐次有限马尔科夫链, Goldberg 和 Segrest^[2]首先使用马尔科夫链分析了遗传算法, Eiben^[3]等用马尔科夫链证明了保留最优个体(elitist)的 GA 的概率性全局收敛, Rudolph^[4]用齐次有限马尔科夫链证明了带有复制、交换、突变操作的标准遗传算法收敛不到全局最优解, 不适合于静态函数优化问题, 建议改变复制策略以达到全局收敛。Qi 和 Palmieri^[5]对浮点数编码的遗传算法进行了严密的数学分析, 用 Markov 链建模, 进行了全局收敛性分析, 但其结果是基于群体数为无穷大这一假设的。

由于以上收敛性结论均建立在计算时间趋于无穷这个条件上, 遗传算法的计算复杂度问题是实际应用更为关心的问题。Buck^[6]和 Muhlenbein^[7]研究了达到全局最优解的遗传算法的时间复杂性问题。

本文应用齐次有限马尔科夫链证明了 SGA 不是全局收敛的, OMSGA 是全局收敛的, 指出 AGA 是全局收敛的, 并基于马尔科夫链对计算效率进行了定性分析。

2 预备知识

令 Φ 为所有长度为 l 的二进制串 s_i 的集合, 串空间 Φ 的维数为 $|\Phi| = 2^l$, 令 Λ 为所有群体大小为 n 的群体 λ_i 的集合, 群体空间的维数为:

$$|\Lambda| = 2^{ln}. \quad (2.1)$$

比例复制时, s_i 的复制概率为:

$$P(s_i) = \frac{f(s_i)}{\sum_{j=1}^n f(s_j)}. \quad (2.2)$$

其中 $f(s_i)$ 是串 s_i 的适合度。串 s_i 和 s_j 之间的海明距离定义为:

* 国家自然科学基金与上海市自然科学基金资助项目。

本文于 1995 年 3 月 31 日收到, 1996 年 4 月 5 日收到修改稿。

$$H(s_i, s_j) = \sum_{k=1}^l |g_{ik} - g_{jk}|. \quad (2.3)$$

其中 g_{ik} 是串 s_i 的第 k 位基因.

经突变使 $H(s_i, s_j) = 0$ 的概率为:

$$P(H(s_i, s_j) = 0) = p_m^{H(s_i, s_j)} (1 - p_m)^{l - H(s_i, s_j)}. \quad (2.4)$$

其中 $p_m \in (0, 1)$ 是突变概率, $P(H(s_i, s_j) = 0) > 0$.

定义 2.1^[8, p2] 设 $X = X_k, k = 1, 2, \dots$, 是定义在概率空间 (Ω, F, P) 上的离散参数的随机过程, 其状态空间 S 为有限集, 如果 X 具有由下列定义的马尔科夫性(或无后效性): 即对任意的非负整数 k , 及任意的状态 $i_0, i_1, \dots, i_{k-1} \in S$, 只要 $P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) > 0$, 总有

$$P(X_{k+1} = i_{k+1} | X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) = P(X_{k+1} = i_{k+1} | x_k = i_k) \quad (2.5)$$

成立, 则称 X 为有限马尔科夫链.

X 在时刻 k 处于状态 i 的条件下, 经 m 步转移, 在时刻 $k + m$ 到达 j 状态的条件概率 $p(X_{k+m} = j | X_k = i)$ 称为 X 的 m 步转移概率, 记为 $p_{ij}(k, k + m)$, 如果对一切 $i, j \in S, p_{ij}(k, k + 1)$ 与时间 k 无关, 则称马尔科夫链是时间齐次的, 简记为 p_{ij} , 以 p_{ij} 作为第 i 行第 j 列排成矩阵 $P = [p_{ij}]$, P 称为转移概率矩阵. 齐次有限马尔科夫链的远期行为由其初始分布和一步转移概率所决定.

定义 2.2^[3] A 是一个方阵,

- 1) 如对所有的 $i, j, a_{ij} > 0$, 记为 $A > 0$, 称 A 为严格正矩阵;
- 2) 如 $A \geq 0$, 若存在 $k \in N$, 使得 $A^k > 0$, 则 A 为正则的;
- 3) 如 $A \geq 0$, 若对所有的 $i, \sum_{j=1}^n a_{ij} = 1$, 则 A 是随机的.

定理 2.1(马尔科夫链的基本极限定理)^[9] 设 P 是正则的齐次马尔科夫链的随机转移矩阵, 则

- 1) 存在一个唯一的概率向量 $\bar{p}^T > 0$, 使得 $\bar{p}^T P = \bar{p}^T$.
- 2) 对于任一初始状态 i (对应的初始概率向量为 e_i^T) 有 $\lim_{k \rightarrow \infty} e_i^T P^k = \bar{p}^T$.
- 3) 有极限概率矩阵 $\lim_{k \rightarrow \infty} P^k = \bar{P}$. 其中 \bar{P} 是 $n \times n$ 随机矩阵, 它的每一行都等于 \bar{p}^T .

基本极限定理说明远期概率与初始状态无关, 该定理是下面收敛性分析的基础.

3 全局收敛性分析

SGA 用齐次有限马尔科夫链来描述时, 状态空间为群体空间 $\Lambda, |\Lambda| = 2^l, \Lambda$ 的元素 λ_i 是一个群体, 包含 n 个串长为 l 的串, $\lambda_i = s_i^j, j \in [1, n]$, 群体空间的概率变化由复制、交换、突变三种基因操作引起, 它们的作用分别用概率矩阵 R, C, M 来描述, 自然地, SGA 的马尔科夫链的概率矩阵 $P = RCM$.

引理 3.1 比例复制操作的概率矩阵 R 是随机的.

证 比例复制操作的作用是按与适合度有关的概率(见 2.2)式)映射到自身和其它状态, 对所有的 $i \in [0, 2^l]$, $\sum_{j=1}^{2^l} r_{ij} = 1$, 所以 R 是随机矩阵.

引理 3.2 交换概率为 $P_c \in [0, 1]$ 的交换操作概率矩阵 C 是随机的.

证 交换操作的作用是按一定的概率把 λ_i 映射到 λ_j , 对所有的 $i \in [0, 2^{ln}]$, $\sum_{j=1}^{2^{ln}} c_{ij} = 1$ 自然地所有的概率 $c_{ij} \geq 0$, 所以 C 是随机矩阵.

引理 3.3 突变概率为 $p_m \in (0, 1)$ 的突变操作的概率矩阵 M 是严格正的随机矩阵.

证 由(2.3)式可得群体状态 λ_i 和 λ_j 之间的海明距离为:

$$H(\lambda_i, \lambda_j) = \sum_{a=1}^n \left(\sum_{b=1}^l |g_{iab} - g_{jib}| \right). \quad (3.1)$$

由(2.4)可推得使 $H(\lambda_i, \lambda_j) = 0$ 的概率为:

$$m_{ij} = p(H(\lambda_i, \lambda_j) = 0) = p_m^{H(\lambda_i, \lambda_j)} (1 - p_m)^{ln - H(\lambda_i, \lambda_j)} > 0. \quad (3.2)$$

所以 M 是严格正矩阵.

定理 3.1 SGA 的转移矩阵 P 是正则的.

证 由引理 3.1 和 3.2, R 和 C 是随机矩阵, 所以 RC 也是随机矩阵, 令 $A = RC$, 因为 M 为严格正矩阵, 所以对于所有的 $i, j \in [1, 2^{ln}]$,

$$p_{ij} = \sum_{k=1}^{2^{ln}} a_{ik} m_{kj} > 0.$$

所以 $P > 0$.

推论 3.1 SGA 构成的马尔科夫链是遍历的.

证 由定理 3.1, SGA 的转移矩阵是正则的, 再由定理 2.1 可知, $\lim_{k \rightarrow \infty} P^k = \bar{P}$, 对于任意状态 i , 有条件概率

$$f_{ii} = P\left(\bigcup_{m=1}^{\infty} X_m = i \mid X_0 = i\right) = 1,$$

状态 i 的平均回转时间

$$\mu_{ii} = \sum_{k=1}^{\infty} k f_{ii}^k < \infty,$$

因此状态 i 是遍历的.

推论 3.1 说明不论群体的初始分布如何, 马尔科夫链中的任何状态均有一大于零的唯一的极限分布, 从任意的状态 i 出发可以在有限时间内到达任意的状态 j . 自然地, $k \rightarrow \infty$, SGA 能遍历状态空间, 但这并不意味着 SGA 能收敛于全局最优解. 为得出最后的全局收敛性结论, 首先定义 GA 的全局收敛概念.

定义 3.1 令 $F_k = \max\{f(s_{ik} | j \in [1, n])\}$ 是时刻 k , 状态 λ_i 时群体中的最大适合度, 令 $F^* = \max\{f(s^j | j \in [1, 2^l])\}$ 是 GA 所求问题的全局最优适合度, 当且仅当下列条件

$$\lim_{k \rightarrow \infty} P(F_k = F^*) = 1 \quad (3.3)$$

成立时, GA 是全局收敛的.

定理 3.2 SGA 不是全局收敛的.

证 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 的概率, 由于状态空间 Λ 中满足 $F_k < F^*$ 的状态数 $2^{ln-l} > 1$, 因此,

$$P(F_k < F^*) \geq P_{ik}.$$

显然为满足 $F_k = F^*$ 的状态的概率为:

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

由定理 3.1 和定理 2.1, 有

$$\lim_{k \rightarrow \infty} P_{ik} > 0.$$

因此,

$$\lim_{k \rightarrow \infty} P(F_k = F^*) \leq 1 - \lim_{k \rightarrow \infty} P_{ik} < 1$$

不满足条件(3.3), 故 SGA 不是全局收敛的.

实验也能发现推论 3.1 和定理 3.2 所说明的现象, 即 SGA 能发现全局最优解, 但不能保证每次都收敛于全局最优解, 即使 $k \rightarrow \infty$. SGA 有全空间搜索的能力, 不能实现全局收敛的原因是 SGA 发现的最优解不能保持.

定义 3.2 设 λ_k^* 是 SGA 马尔科夫链在 k 时刻具有最优适合度 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 称之为最优保存 SGA, 简称为 OMSGA(optimum maintaining simple genetic algorithm).

定理 3.3 OMSGA 是全局收敛的.

证 设 Λ_0 是满足 $F_0 = F^*$ 的状态的集合, 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, 由推论 3.1 可知, 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,

$$\lim_{k \rightarrow \infty} P(F_k = F_0 = F^*) = 1.$$

所以, OMSGA 是全局收敛的.

定义 3.3^[10] 具有比例复制和自适应交换及突变操作的遗传算法, 称之为自适应遗传算法, 简称 AGA(Adaptive Genetic Algorithm), p_c 和 p_m 自适应变化如下:

$$p_c = \begin{cases} k_1(f_{\max} - f')/(f_{\max} - \bar{f}), & f' \geq \bar{f}, \\ k_3, & f' < \bar{f}, \end{cases} \quad (3.4)$$

$$p_m = \begin{cases} k_2(f_{\max} - f)/(f_{\max} - \bar{f}), & f' \geq \bar{f}, \\ k_4, & f' < \bar{f}. \end{cases} \quad (3.5)$$

$$0 < k_1, k_2, k_3, k_4 \leq 1.0. \quad (3.6)$$

其中 f_{\max} 是群体中最大适合度, \bar{f} 是群体的平均适合度, f' 是用于交换的二个串中较大的适合度, f 是突变串的适合度.

命题 3.1 AGA 是全局收敛的.

说明 由(3.4), (3.5)和(3.6)式, $0 \leq p_c \leq 1, 0 \leq p_m \leq 1$, 其中, $f' = f_{\max}$ 时, $p_c = 0$, $f = f_{\max}$ 时, $p_m = 0$, 因此, 任意状态 λ_i 中的最优串能保存, 同时 $P > 0$, 由定理 3.3 可说明 AGA 是全局收敛的.

虽然复制操作转移矩阵实质上是时变的, 但用齐次马尔科夫链描述 SGA 不失一般性, 定理 3.2 证明了 SGA 不是全局收敛的, 这意味着模式定理不能证明 SGA 的全局收敛性, 它说明的只是群体中模式的数量随其适合度而变化的规律.

定理 3.3 说明要实现遗传算法全局收敛的基本思想是把发现的最优解保存下来. 最优

保存可以通过复制、交换和突变操作来实现。最优串复制是一种常用的实现全局收敛的策略^[3]。自适应交换和突变操作^[10]可能也是实现全局收敛的好方法,因为串的交流率和突变率根据其适合度确定,从而能把最优串保存下来,但由于这种策略构成的马尔科夫链是非齐次的,要进行理论上的严密证明需要进一步的努力。

在全局收敛性证明中, $P > 0$ 条件过强,因为 $P \geq 0$ 也可能使 $P^k > 0$, $P > 0$ 的条件主要由 $M > 0$ 引起,那么没有突变操作的 OMSGA 能否全局收敛呢?这是一个有意义但有待解决的问题。

4 基于马尔科夫链的效率分析

运用马尔科夫链才能严密分析遗传算法的时间复杂度,而这依赖于对遗传算法的精确建模。下面运用马尔科夫链的特性和第 3 节得出的结论进行初步的定性分析。

定理 3.3 表明群体的初始分布与全局收敛性无关,但初始分布影响搜索效率,如果初始状态 λ_i 中串完全一样,则 R 和 C 矩阵均不能改变,只有靠正矩阵 M 才能使 $\lambda_i \rightarrow \lambda_j$, $H(\lambda_i, \lambda_j) \simeq \ln p_m$, 若 p_m 小,则状态将长时间在 λ_i 附近徘徊,效率很低,这种现象与“早熟收敛”一致,出现后, R 和 C 均无效, M 效率很低。

假设转移矩阵 P 中的状态如图 1 排列。高适合度的状态排在左边,则左边共有 2^{ln-l} 个状态满足 $F(\lambda_i) = F^*$, $i = 2^{ln-l} - 1, \dots, 2^{ln}$ 。由定理 2.1, 状态的极限概率分布向量为 \bar{p}^T 。为了提高到达全局最优状态的计算效率,自然地希望“适合度高的状态有较高的极限分布概率”,这就是极限分布概率原则,与适合度密切相关的操作矩阵是 R , 适合度高的串复制概率高使得当前状态向适合度高的状态移动,能满足极限概率分布的要求,说明复制操作是加快遗传算法到达全局最优状态的关键,这与模式定理得出的结论是一致的。

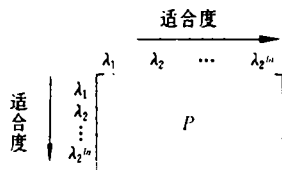


图 1 P 矩阵中状态的排列

根据极限分布概率原则,可以使交换操作和突变操作也与适合度有关,使得高适合度的串的交换和突变概率较低,促使状态向高适合度的状态移动,这个结论与最近 AGA 的研究与实验结果^[10]相符。过强的 R 操作会把状态强有力地吸引到强的局部最优状态,也即通常称的“早熟收敛”,需设计合适的复制操作,以避免过度受适合度影响,及设计合适的突变操作以摆脱局部最优状态。

5 结 论

马尔科夫链是对遗传算法进行建模的有力工具,本文严密证明了 SGA 不是全局收敛的,OMSGA 是全局收敛的,指出 AGA 也是全局收敛的。以上结果的证明主要根据突变操作的概率转移矩阵是严格正矩阵这个条件,实际上,遗传算法中的主要的搜索工具是交换操作,文献[11]初步得出了群体中有效基因不缺失,交换操作的概率转移矩阵也是正则的,但还未能严密证明。更有实用意义的是遗传算法计算复杂度的分析,通过对转移矩阵的分析,本文得到提高遗传算法效率的准则:适合度高的状态需要有高的极限分布概率,该准则有助于设计提高遗传算法计算效率的基因操作策略。

参 考 文 献

- [1] Holland, J. H. . Adaptation in Natural and Artificial Systems. 1st ed. , 1975. 2nd ed. . Cambridge, MA: MIT press. .

1992

- [2] Goldberg, D. E., Segrest, P.. Finite Markov Chain Analysis of Genetic Algorithm. Proc. of the Second Int. Conf. on Genetic Algorithms. 1987, 1—8
- [3] Eiben, A. E., Aarts, E. H., and Van Hee, K. M.. Global Convergence of Genetic Algorithms: An Infinite Markov Chain Analysis. Parallel Problem Solving from Nature. Schwefel, H. P, and Manner, R., Eds. Heidelberg, Berlin: Springer-verlag, 1991, 4—12
- [4] Rudolph, G.. Convergence Properties of Canonical Genetic Algorithms. IEEE Trans. on Neural Networks, 1994, 5 (1): 96—101
- [5] Qi, X., Palmieri, F.. Theoretical Analysis of Evolutionary Algorithms with an Infinite Population Size in Continuous Space, Part I: Basic Properties of selection and Mutation, IEEE Trans. on Neural Networks, 1994, 102—119
- [6] Back, T.. The Interaction of Mutation Rate, Selection and Self-Adaption within Genetic Algorithms. Parallel Problem Solving from Nature, 2, Amsterdam, North Holland, 1992, 84—94
- [7] Muhlenbein, H.. How Genetic Algorithms Really Work. I: Mutation and Hillclimbing. Parallel Problem Solving from Nature, 2, Amsterdam, North Holland, 1992, 15—25
- [8] 施仁杰. 马尔科夫链基础及其应用. 西安: 西安电子科技大学出版社, 1992
- [9] 鲁恩伯杰, D. G. 著, 袁天鑫, 黄午阳译. 社会动态系统引论. 上海: 上海科学技术文献出版社, 1985
- [10] Scrivas, M., Patnaik, L. M.. Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms. IEEE Trans. SMC, 1994, 24(4): 656—666
- [1] 恽为民. 基于遗传的机器人运动规划的研究. 上海交通大学博士论文, 1995

The Analysis of Global Convergence and Computational Efficiency for Genetic Algorithm

YUN Weimin and XI Yugeng

(Department of Automation, Shanghai Jiao Tong University • Shanghai, 200030, PRC)

Abstract: In the paper, the global convergence of simple genetic algorithm (SGA), optimum maintaining SGA (OMSGA) and adaptive genetic algorithm is analyzed using homogeneous finite Markov chain and the computational efficiency is discussed. The limit distribution probability principle is obtained, which is useful to guide the design of gene operation schemes.

Key words: global convergence; computational efficiency; simple genetic algorithm (SGA); optimum maintaining SGA (OMSGA); adaptive genetic algorithm (SGA)

本文作者简介

恽为民 见本刊 1996 年第 3 期第 304 页.

席裕庚 见本刊 1996 年第 3 期第 304 页.