

遗传算法基础理论研究的新近发展

徐宗本 陈志平

(西安交通大学理学院信息科学与系统科学研究所, 西安, 陕西, 710049, 中国)

章祥荪

(中国科学院数学与系统科学研究院应用数学研究所, 北京, 100080, 中国)

摘要 本文综述有关遗传算法收敛性及收敛速度估计的近期研究结果. 在分类概述相关的 Vose-Liepins 模型、Markov 链模型、公理化模型、连续 (积分算子) 模型及收敛速度估计、迭代次数估计与时间复杂性估计的基础上, 指出遗传算法理论研究存在的其它亟待解决的问题.

关键词 遗传算法; 收敛性; 收敛速度分析; 模型; 时间复杂性

MR(1991) 主题分类 68T; 68Q; 65V

1 引言

遗传算法 (Genetic Algorithms, 下称 GAs) 是一类模拟生物进化过程与机制求解问题的自适应人工智能技术. 它的核心思想源于这样的基本认识: 从简单到复杂、从低级到高级的生物进化过程本身是一个自然、并行发生的、稳健的优化过程, 这一优化过程的目标是对环境的适应性, 而生物种群通过“优胜劣汰”及遗传变异来达到进化的目的. 依达尔文的自然选择与孟德尔的遗传变异理论, 生物的进化是通过繁殖、变异、竞争和选择这四种基本形式来实现的. 如果把待解决的问题描述作为对某个目标函数的全局优化, 则 GAs 求解问题的基本作法是: 把待优化的目标函数解释作生物种群对环境的适应性, 把优化变量对应作生物种群的个体, 而由当前种群出发, 利用合适的复制、杂交、变异与选择操作生成新一代种群. 重复这一过程, 直至获得合乎要求的种群或规定的进化时限.

由于具有鲜明的生物背景和对任何函数类 (特别可以无表达式) 可用等突出特点, GAs 自 80 年代中期以来引起人工智能领域的普遍关注, 并被广泛应用于机器学习、人工神经网络训练、程序自动生成、专家系统的知识库维护等一系列超大规模、高度非线性、不连续、极多峰函数的优化. 90 年代以来, 对该类算法的研究日趋成为计算机科学、信息科学与最优化领域研究的热点.

在 GAs 中, 待优化的问题通常被转化作某个合适适应值函数 (fitness function) 的极大化, 而代替作用于原优化变量本身. GAs 通常作用于原问题变量的某个有限长离散编码 (常为定长二进制字符串), 通常称之为个体. 个体全体组成的集合称为个体空间. 一对个体常称之为母

收稿日期: 1999-01-04.

国家 863(306) 及国家自然科学基金资助课题.

体. 母体空间是由所有母体组成的集合. 种群 (population) 规模在迭代过程中一般是固定的, 设其为 N , 则任意 N 个个体组成的集合即为一个种群.

GAs 的具体迭代过程, 或等价地, 由当前种群生成新一代种群的方法通常由一系列遗传操作 (算子) 决定. 这些算子是对自然演化中种群进化机制的类比与模拟. 常见的有选择 (selection)、杂交 (crossover) 和变异 (mutation) 等.

选择算子是种群空间到母体空间的随机映射, 它按照某种准则或概率分布从当前种群中选取那些好的个体组成不同的母体以供生成新的个体. 最常用的选择算子是以个体适应值成比例的所谓比例选择 (proportional selection) 及 ranking-based 选择等.

杂交算子是母体空间到个体空间的随机映射, 它的作用方式是: 随机地确定一个或多个分量位置为杂交点, 由此将一对母体的两个个体分为有限个截断, 再以概率 P_c (称为杂交概率) 交换相应截断得到新的个体. 依杂交点个数的多少, 杂交算子可分为单点杂交、两点杂交和多点杂交等. 多点杂交的极限形式则称为均匀杂交.

变异算子是个体空间到个体空间的随机映射, 其作用方式是: 独立地以概率 P_m (称为变异概率) 改变个体每个分量 (基因) 的取值以产生新的个体.

若用 $X_i(t)$ 表示第 t 代种群的第 i 个个体, 则标准 GAs 的迭代过程可描述为:

- 步 1 置 $k = 0$, 随机产生初始种群 $X(0) = \{X_1(0), X_2(0), \dots, X_N(0)\}$;
- 步 2 独立地从当前种群中选取 N 对母体;
- 步 3 独立地对所选 N 对母体实施杂交生成 N 个中间个体;
- 步 4 独立地对 N 个中间个体进行变异得到 $(k+1)$ 代种群

$$X(k+1) = \{X_1(k+1), X_2(k+1), \dots, X_N(k+1)\};$$

- 步 5 停止或置 $k := k+1$ 并返回步 2.

从以上描述可以看出, 与通常优化技术中的搜索方法不同, GAs 作为一种自适应的随机搜索方法, 其搜索方式是由当前种群所提供的信息, 而不是由单一的方向或结构来决定; 同时, 它将多个个体作为可能的解并考虑搜索空间全局范围内的抽样, 如此导致其能以更大的可能性收敛到全局最优解. 由于这些特性, GAs 能够成功地用于求解众多不同的复杂而困难的优化问题 (包括非数值优化问题). 也正由于如此, 对遗传算法的收敛性理论研究, 要求应用完全不同于传统优化算法的分析与研究方法.

有关遗传算法的已有研究大都集中于算法的实现、改进与应用方面, 而相关的基础理论研究远落后于算法发展. 虽然近年来有关 GAs 的渐近行为分析受到愈来愈广泛的注意, 但已有的研究都是就 GAs 的某一特定实现, 或在某一相对弱的意义下讨论算法的收敛性 (或在某一特定的度量下研究其收敛速度), 均具有相当局限性. 可以说, 到目前为止, 还没有一套完整的理论可以准确、全面地阐明一般 GAs 的收敛性, 从而对其在大量应用中所表现出的全局优化能力作出理论解释; 也还没有找到一个恰当的度量与论证方法精确刻画 GAs 在不同实现下的收敛速度, 从而对 GAs 的各种改进作出统一、公正的评判. 这种一般数学理论基础的缺乏正致命地限制着 GAs 的进一步推广、改进与应用.

本文的目的是, 对已有有关 GAs 收敛性与收敛速度估计的近期结果作分类总结, 以期推动这方面研究的更深入发展. 我们将扼要叙述已有的典型模型与方法, 剖析其优缺点, 并在此基础上, 指出有待进一步研究的问题及其可能的解决途径.

以下, 第二节讨论关于 GAs 收敛性分析的已有模型与结果, 主要介绍 Vose-Liepins 模型、Markov 链模型、公理化模型与连续 (积分算子) 模型; 第三节概述 GAs 在不同实现下收敛速度与复杂性分析的已有结果; 第四节, 我们从更一般角度评述有待进一步研究的 GAs 理论问题。

2 收敛性分析

遗传算法的收敛性通常是指遗传算法所生成的迭代种群 (或其分布) 收敛到某一稳定状态 (或分布), 或其适应值函数的最大或平均值随迭代趋于优化问题的最优值。依据不同的研究方法 & 所应用的数学工具, 已有 GAs 收敛性分析结果可大致分为四类, 即 Vose-Liepins 模型、Markov 链模型、公理化模型和连续 (积分算子) 模型。我们以下分别概述。

2.1 Vose-Liepins 模型

这类模型基于 Vose 和 Liepins(1991) 文 [1] 的工作。其核心思想是: 用两个矩阵算子分别刻画比例选择与组合算子 (即杂交算子与变异算子的复合), 通过研究这两个算子不动点的存在性与稳定性来刻画 GAs 的渐近行为。

设 L 为 GAs(二进制) 编码长度, 则所有可能个体的总数为 $r = 2^L$, 将它们分别表示为 $\{0, 1, \dots, r-1\}$ 。用 $m_{i,j}(k)$ 表示由个体 i, j 通过组合算子的作用获得个体 k 的概率, 定义 r 阶矩阵 $M = (m_{i,j}(0))$ 。引进向量 $\Phi^t \in R^r$, 其第 i 个分量表示个体 i 在第 t 代种群中所占的比例; 向量 $S^t \in R^r$, 其第 i 个分量 S_i^t 表示在第 t 代种群中选择个体 i 进行组合的概率。用 \otimes, \oplus 分别表示逻辑与及异或运算。

对种群规模无限的二进制编码 GAs, Vose 和 Liepins(1991, 参见 [1]) 证明: 个体 k 在下一代种群中的期望比例为

$$E\Phi_k^{t+1} = \sum_{i,j} S_i^t S_j^t m_{i,j}(k), \quad (1)$$

这里 E 为数学期望。进而, $m_{i,j}(k \oplus l) = m_{i+k, j+k}(l)$ 。因此, 只须计算 $m_{i,j}(0)$, 即可得到 $m_{i,j}(k)$ 对任一 k 的值。对于由单点杂交和变异复合而成的组合算子, 容易验证有

$$m_{i,j}(0) = \frac{(1 - P_m)^L}{2} \left\{ \eta^{|i|} (1 - P_c + \frac{P_c}{L-1} \sum_{k=1}^{L-1} \eta^{-\Delta_{i,j,k}}) + \eta^{|j|} (1 - P_c + \frac{P_c}{L-1} \sum_{k=1}^{L-1} \eta^{\Delta_{i,j,k}}) \right\}, \quad (2)$$

其中 $\eta = \frac{P_m}{1-P_m}$, $\Delta_{i,j,k} = |(2^k - 1) \otimes i| - |(2^k - 1) \otimes j|$, 而 $|\cdot|$ 表示相应个体中非零基因的数目。

设 f 为待优化的正值适应值函数, 定义线性算子 F 为第 (i, i) 项取值 $f(i)$ 的 r 阶对角矩阵, 并定义 R^r 上的置换算子 σ_j 为

$$\sigma_j \langle x_0, \dots, x_{r-1} \rangle^T = \langle x_{j \oplus 0}, \dots, x_{j \oplus (r-1)} \rangle^T.$$

引进算子 $\varphi: R^r \rightarrow R^r$ 使满足

$$\varphi(x) = \langle (\sigma_0 x)^T M(\sigma_0 x), \dots, (\sigma_{r-1} x)^T M(\sigma_{r-1} x) \rangle^T.$$

借助于这些算子, Vose 和 Liepins 证明: 下一代种群中各个个体出现的概率为

$$S^{t+1} = \frac{F\varphi(S^t)}{\|F\varphi(S^t)\|_1},$$

或等价地, 记 $\bar{F}u = Fu/\|Fu\|_1$, 则

$$S^{t+1} = \bar{F}\varphi(S^t). \quad (3)$$

公式 (3) 的重要性在于, 说明 GAs 的平均行为可由矩阵 F 和二次非线性算子 φ (从而矩阵 M) 来确定. 而当种群规模趋于无穷时, 关系 $S^{t+1} = \bar{F}\varphi(S^t)$ 则为其极限行为的精确描述 (根据大数定律). 因此, 对于种群规模无限的 GAs, 其收敛性将对应于 $\bar{F}\varphi$ 不动点的稳定性.

显然, 利用 Vose-Liepins 模型研究 GAs 收敛性的核心在于刻画非线性算子 $\bar{F}\varphi$ 的不动点集结构与稳定性. 目前对此尚无一般的结果, 但对 \bar{F} 与 φ 单独作用情形的不动点集与稳定性, Vose 和 Liepins 证明了如下定理

定理 1 a) \bar{F} 的不动点集为 $\{e_1, e_2, \dots, e_r\}$ (其中 e_j 为第 j 个坐标向量), 且 e_j 的吸引域由椭球 $\sum_i (S_i \frac{f(i)}{f(j)})^2 < 1$ 与 R^r 中单位球内的非负点集的交所确定; e_j 是稳定的当且仅当 $f(j)$ 为 f 的最大值;

b) 若 M 为正矩阵, 则当矩阵 $M_* = \{m_{i \oplus j, i}(0)\}$ 的第二个最大特征值小于 0.5 时, φ 的不动点为渐近稳定的.

对于结论 b), 作者猜测在 $0 < P_m < 0.5$ 时, M_* 的第二个最大特征值确实小于 1/2. 利用 Walsh 矩阵的特性及某些组合恒等式, Koehler(1994, 参见 [2]) 通过精确刻画矩阵 M_* 的谱结构肯定了 Vose 与 Liepins 的这一猜测.

采用与文 [2] 相同的论证方法和 Walsh 矩阵在非二进制情形的定义及特性, Bhattacharyya 和 Koehler(1994)(参见 [3]) 将文 [1] 的结果进一步推广到了基数为 $K = 2^v$ (v 为正整数) 的非二进制 GAs (但使用相同的遗传算子). 他们证明: 当 $P_m < 1/K$ 时, M_* 的第二个最大特征值将小于 0.5, 从而 φ 的不动点渐近稳定.

从动力系统的观点, 并利用微分拓扑与算子论工具, Wright 和 Vose(1996)(参见 [4]) 研究了上述 Vose-Liepins 模型的稳定性. 他们证明了使得算子 $\bar{F}\varphi$ 的不动点集有限的适应值函数集合是稠密的和开的. 这样, 对任意适应值函数, 可以认为 $\bar{F}\varphi$ 的不动点集是有限的, 而且其有限性不随适应值函数的小扰动而改变.

Vose-Liepins 模型虽然在种群规模无限的假设下可精确刻画 GAs, 但在有限规模情形下却只能描述 GAs 的平均形态. 为了克服这一缺陷, Nix 和 Vose(1992, 参见 [5]) 结合 Vose-Liepins 模型与 Markov 链描述, 发展了 GAs 的一个精确 Markov 链模型: 设 P_i 是任一规模为 N , 编码长度为 L 的种群, 则 P_i 可唯一地用一个 r 维列向量 $\Phi_i = \langle Z_{0,i}, \dots, Z_{r-1,i} \rangle^T$ 来标识 (称为 P_i 的标识向量), 这里非负整数 $Z_{y,i}$ 表示个体 y 在 P_i 中出现的次数 (因此 $\sum_{y=0}^{r-1} Z_{y,i} = N$). 不难证明所有可能种群的总数为 $\bar{P} = \binom{N+r-1}{r-1}$. 若定义每个可能的种群为一个状态, 则 GAs 可以用一具有 \bar{P} 个状态的有限 Markov 链来描述. 设 \bar{P} 阶矩阵 $Q = (Q_{ij})$ 为相应的概率转移矩阵, 其中 Q_{ij} 表示给定当前种群 P_i , 下一代种群为 P_j 的概率. 显然, 只要能确定 Q_{ij} 的值, 则 GAs 的 Markov 链模型便完全确定.

令 $P_i(y)$ 表示给定 P_i , 在下一代种群中生成个体 y 的概率, 则 Nix 和 Vose 证明: $P_i(y)$ 可由 F 和 φ 如下确定

$$P_i(y) = [\varphi(\frac{F\Phi_i}{\|F\Phi_i\|_1})]_y = [\varphi\bar{F}(\Phi_i)]_y. \quad (4)$$

进而, 若下一代种群为 P_j , 其标识向量为 $\Phi_j = \langle Z_{0,j}, \dots, Z_{r-1,j} \rangle^T$, 则由 P_i 生成 P_j 的概率服

从下列多项式分布

$$Q_{ij} = \frac{N!}{Z_{0,j}! \cdots Z_{r-1,j}!} \prod_{y=0}^{r-1} \{P_i(y)\}^{Z_{y,j}} = N! \prod_{y=0}^{r-1} \frac{\{P_i(y)\}^{Z_{y,j}}}{Z_{y,j}!}. \quad (5)$$

因假设变异概率非零, 上述 Markov 链是遍历的, 故存在稳态分布. Nix 和 Vose 证明: 随着种群规模趋于无穷, 这一稳态分布序列将以概率 1 趋于 $\varphi\bar{F}$ 的不动点, 从而揭示了其与 Vose-Liepins 模型 (3) 的紧密联系.

Nix-Vose 的有限种群模型虽则恰好描述了 GAs 的实际演化过程, 但由于其概率转移矩阵的复杂性, 直接基于该模型分析 GAs 收敛性态是困难的. 而相比之下, Vose-Liepins 的无限种群模型虽然只能描述实际 GAs 演化的平均性态, 但它却精确预报了 GAs 收敛性态随种群规模的变化. 特别地, Vose(1996, 参见 [6]) 最近还证明: 对于大种群的 GAs, 其短期收敛性态是由所选初始种群所位于的 $\varphi\bar{F}$ 之不动点的吸引域决定, 而长期收敛特性则由 $\varphi\bar{F}$ 的具最大吸引域的不动点决定.

上述研究进一步展示了 Vose-Liepins 模型的重要性. 然而, 我们注意到: (1) Vose-Liepins 模型只适用于简单 GAs(换言之, 应用比例选择、单点杂交和单点变异), 还没有推广到适用于更具实用性的其它 GAs 执行策略(如杰出者选择(elitist selection)、锦标赛选择(tournament selection)加多点杂交、均匀杂交等); (2) 对算子 $\varphi\bar{F}$ 不动点性质(结构、稳定性等)还缺少深入系统研究, 已有的结果只考虑了 φ 和 \bar{F} 的分离作用的情形, 还不能形成对 GAs 收敛性态的直接分析; (3) Vose-Liepins 模型不易处理变异、杂交概率随时间变化的情形, 其框架亦很难推广到描述一般非二进制或连续变量情形的 GAs.

2.2 Markov 链模型

由于遗传算法下一代种群的状态通常完全依赖当前种群信息, 而不依赖于以往状态, 故可自然地用 Markov 链描述. 这种方法一直被用于研究不同形式 GAs 的渐近行为. 一些典型的结果如下:

(1) 用 Markov 链描述其动态行为, Horn(1993, 参见 [7]) 分析了不同壁龛算子(niching operators)对 GAs 的随机影响, 但仅就长度为 1 的二进制串这一特殊情形, 并主要从数值模拟的角度进行.

(2) 从更一般的等价类层次表述种群, Mahfoud(1993, 参见 [8]) 详细讨论了当有限 Markov 链用于描述 Boltzmann 锦标选择 GAs 时, 其转移概率的确定过程. 在不考虑杂交与变异情形下, 分析了使该选择过程达到或接近其吸引状态所需的平均迭代次数及平均迁移时间. 该文从更一般层次来刻画种群的思想可用于分析其它更为复杂的 GAs, 而所用分析方法对于我们研究其它选择算子也有一定的借鉴作用. 然而, 该文工作很难推广到所用选择算子与杂交、变异算子联合作用的实用情况中去.

(3) 推广 Vose 和 Liepins(参见 [1]) 及 Nix 和 Vose(参见 [5]) 的工作, Dawid(1994, 参见 [9]) 研究了当每个个体的适应值由一个依赖于整个种群状态的函数来确定时, GAs 的 Markov 链描述问题. 他证明了可用 (4) 与 (5) 的如下推广形式来刻画此时 GAs 的 Markov 链

$$P_y(\Phi_i) = \varphi\left(\frac{F(\Phi_i)\Phi_i}{f^T(\Phi_i)\bar{\Phi}_i}\right)_y,$$

$$Q_{ij} = N! \prod_{y=0}^{r-1} \frac{P_y(\Phi_i)^{NZ_{y,j}}}{(NZ_{y,j})!},$$

这里 $f(\Phi_i) = (f_0(\Phi_i), f_1(\Phi_i), \dots, f_{r-1}(\Phi_i))^T$, $F(\Phi_i) = \text{diag}(f(\Phi_i))$.

易知当 $P_m > 0$ 时, 转移矩阵 Q 为严格正的, 因而上述 Markov 链有唯一的极限分布. 对于 $P_m = 0$, Dawid 就单点杂交和均匀杂交两种情形分别给出了保证上述 Markov 链稳定状态为局部渐近稳定的充要条件.

(4) 基于文 [5] 中的 Markov 链模型, Suzuki(1995, 参见 [10]) 通过估计转移矩阵的特征值, 分析了带有如下修正杰出者选择 GAs 的收敛性与收敛速度: 当前种群中的最佳个体被无条件地保留作为下一代种群中的一个个体. 设 q^t 表示第 t 次迭代时每个包含最优解的种群出现的概率之和, 则在假设杂交概率恒为 1 的前提下, Suzuki 证明了如下结果

定理 2 存在常数 $A, C > 0$ 使得

$$q^t \geq 1 - C|\lambda_*|^t, \quad (6)$$

其中

$$|\lambda_*| = \max_{0 \leq i \leq r-1} \max_{1 \leq j \leq N(i)} |\lambda_{ij}| \leq 1 - AP_m^\delta (1 - P_m)^{L-\delta} < 1, \quad (7)$$

这里 $N = \sum_{i=0}^{r-1} N(i)$, λ_{ij} 为转移矩阵的特征值, δ 为任意两个个体 i, j 之间的 Hamming 距离 $h(i, j)$ 的极大一极小值, 换言之

$$\delta = \max_{0 \leq i \leq r-1} [\min_{0 \leq j \leq r-1, f(i) < f(j)} h(i, j)].$$

因为 $|\lambda_*| < 1$, 故当 $t \rightarrow \infty$ 时 $q^t \rightarrow 1$. 定理 3 从而推出: 对于带修正杰出者选择的 GAs, 包含最优解的种群将以概率 1 出现, 从而具有收敛性.

(5) 用三个随机矩阵 C, M 和 S 分别描述由简单 GAs 中杂交、变异和选择而引起的中间转变, Rudolph(1994, 参见 [11]) 用带有转移矩阵 $P = C \cdot M \cdot S$ 的 Markov 链来描述简单 GAs, 并讨论其用于静态最优化问题的收敛性. 利用有限时齐 Markov 链理论, 他证明了当 $0 < P_m < 1$ 时所述 Markov 链是遍历的. 因此, 若将收敛定义为由 GAs 迭代所产生的最佳适应值序列以概率 1 收敛到优化问题的全局最优值, 则标准 GAs 不收敛! 但其下述变形却是收敛的: 在选择前或选择后总记录最佳个体 (此即杰出者选择策略).

(6) 为了克服上述用 Markov 链描述 GAs 时对所采用编码方法的依赖性, 在将 GAs 按照其搜索过程分为替代抽样和非替代抽样两大类后, Reynolds 和 Gomatam(1996)(参见 [12]) 通过引进以向量为指标的矩阵乘方概念, 详细讨论了上述两类 GAs 在仅有抽样运算、有抽样与比例选择及所有遗传算子联合作用等情况下各算子作用方式的准确定义、及其相应转移概率的精确计算, 由此建立了相应的 Markov 链模型及其收敛性.

(7) 为了考虑各遗传算子的不同组合以及变异概率 (和杂交概率) 时变的情形, Davis 和 Principle(1991)(参见 [13]) 提出了将现有模拟退火算法理论外插到 GAs 的 Markov 链描述中来分析算法的渐近行为的思想. 与单纯的 Markov 链方法相比, 该方法具有能同时考虑时齐与非时齐 GAs 以及较方便地分析带有不同组合方式的遗传算子的优点. 但遗憾的是, 除了一些简单的描述、概率转移阵的确定与数值模拟外, 我们至今仍未能看到有关该类方法的严格理论分析.

(8) 将变异、杂交算子描述作对当前种群中个体(对)的独立随机扰动, 而从当前种群状态到下一种群状态的转移概率由状态空间上满足一定条件的马氏核 (markovian kernel) 来确定, Cerf(1994, 参见 [14]) 应用动力系统随机扰动理论研究了当扰动趋于零时 GAs 的渐近行为, 并讨论了为保证收敛扰动与选择压应满足的条件. 由于有关的假设条件不易检验, 该文结果很难用于对实际 GAs 的分析.

(9) 针对组合优化问题, 在建立适当的邻域结构后, Eiben 等人 (1991, 参见 [15]) 通过含四个参数的函数分别描述选择、组合、估值等运算, 提出了用带参数函数复合表示的抽象 GAs Markov 链模型. 然而, 那里所给出的收敛性结果只适用于满足特定条件的杂交、变异与杰出选择型 GAs. 因为不易找出具体的 GAs 所对应的参数化函数描述, 所得结果也很难直接用于实际.

用 Markov 链模型描述遗传算法虽然有直接、精确的优点, 但由于所采用有限状态 Markov 链理论本身的限制, 与 Vose-Liepins 模型类似, 该模型只能用于描述通常的二进制或特殊的非二进制 GAs. 另外, 这类方法所得收敛性一般是指相应的 Markov 链趋于某一平稳分布, 这与优化中通常所指的收敛性定义不同, 它并不保证 GAs 将一定或以概率 1 收敛到问题的全局最优解. 再则, 相应 Markov 链的状态数 (从而转移矩阵的规模) 通常很大, 这使得具体确定、分析转移阵的性态十分困难 (更不用说具体描述其平稳分布), 因而对于较大规模的 GAs, Markov 链分析只能籍于遍历性考察而得出相应 GAs 收敛性的某些“粗糙”结论. 我们注意到: 截止目前几乎所有文献均只考虑了时齐 Markov 链理论对遗传算法的应用. 所用方法显然不能分析那些变异、杂交概率时变的 GAs, 对此我们必须应用非时齐 Markov 链 (过程) 理论, 而这将使分析变得更加复杂和困难.

2.3 公理化模型

作者 (1998, 参见 [33]) 新近发展了一个既可用于分析时齐又可用于分析非时齐 GAs 的公理化模型. 这一方法的核心思想是: 通过公理化描述 GAs 的选择算子与重组算子 ([33] 中称之为演化算子), 并利用所引进的参量分析 GAs 的收敛性. 值得注意的是, 对于常见的选择算子与重组算子, 所引进的参量能方便地确定, 因而这一模型具有重要的理论意义与应用价值.

设 Ω 表示个体空间 (因而 Ω^N 是种群空间). 适应值函数 f 被假定在测度空间 (Ω, Ψ, μ) 上考虑. 一个集合 $B \in \Psi$ 称为是 f 的满意集, 如果 $f(a) > f(b)$ 对任何 $a \in B, b \in \Omega \setminus B$ 成立 (这一概念是最优解集的直接推广).

公理 1 一个随机映射 $S: \Omega^N \rightarrow \Omega^N$ 是一个 (抽象) 选择算子, 如果

i) 对任何 $X \in \Omega^N, S(X) \subset X$;

ii) 存在正数 β 使

$$P[|S(X)_M| \geq \beta + |X_M|] > 0, \quad (8)$$

这里 P 表示概率, $|D|$ 表示集合 D 的基数, $M = \max\{f(x) : x \in X\}$, 且 $X_M = \{x \in X : f(x) = M\}$.

公理 2 一个随机映射 $E: \Omega^N \rightarrow \Omega^N$ 是一个 (抽象) 演化算子, 如果对任意 $X \in \Omega^N$,

i) $P[X = E(X)] > 0$;

ii) 对任何 $B \in \Psi$, 如果 $X \cap B = \emptyset$, 则 $P[E(X) \cap B \neq \emptyset] > 0$;

iii) 对任何 $B \in \Psi$, 如果 $B \cap X \neq \emptyset$, 则 $P[E(X) \cap B = \emptyset] = 0$.

公理 1 刻画选择算子的两个基本特性: (i) 选择只能在当前种群中进行; (ii) 选择的作用是使种群中最优个体增加. 而公理 2 表述重组 (演化) 算子的如下要求: (i) 任何种群都有被重

组后保持的可能性 (否则, 最优解种群会被破坏); (ii) 对任何满意集, 重组应有能力从任何种群起始到达该满意集; (iii) 如种群演化已经撞入满意集, 则重组应使满意个体不全部丢失.

借助以上公理刻画的选择与演化算子, [33] 定义随机过程

$$X^{(t)} = E(t)S(t)(X^{(t-1)}) \quad (t = 0, 1, 2, \dots)$$

为一抽象模拟演化算法 (Abstract Evolutionary Algorithm-AEA), 其中 $\{E(t)\}$ 和 $\{S(t)\}$ 为一系列演化与选择算子.

为了研究 AEA 的收敛性态, [33] 引进了刻画抽象选择算子与演化算子的下述数量特征
定义 1(选择压, 选择强度) 设 S 为一抽象选择算子.

- (i) 满足使 (8) 式成立的最大整数 β 称之为 S 的选择压 (selection pressure)(记为 S_p);
- (ii) 由下式定义的常数 S_I 称为选择算子 S 的选择强度 (selection intensity)

$$S_I = \inf \{P(|S(X)_M| \geq S_p + |X_M|) : |X_M| = 1, X \in \Omega^N\}.$$

定义 2(保存率, 迁入率, 迁出率) 设 E 为一抽象演化算子, Ψ 为由所有满意集所组成的集族.

- (i) E 的保存率 E_l 定义为

$$E_l = \inf \{P(E(X) = X) : X \in \Omega^N\};$$

- (ii) E 的迁入率 E_e 定义为

$$E_e = \inf \{P(E(X) \cap B \neq \emptyset) | X \in \Omega^N, B \in \Psi, X \cap B = \emptyset\};$$

- (iii) E 的迁出率 E_o^r 定义为

$$E_o^r = \sup \{P(E(X) \cap B = \emptyset) | X \in \Omega^N, B \in \Psi, |X \cap B| = r > 0\}.$$

利用所引进的参数族 $(S_p, S_I, E_l, E_e, E_o^r)$, 可直接建立起抽象演化算法 AEA 的收敛性. 依 [33] 的术语, AEA 称为是次收敛的, 如果对任意满意集 B , $\mu(B) \neq 0$, $\lim_{t \rightarrow \infty} P(X^{(t)} \cap B \neq \emptyset) = 1$; AEA 称为是强收敛的, 如果对任意非零测度满意集 B , $\lim_{t \rightarrow \infty} P(X^{(t)} \subset B) = 1$.

定理 3 在下述条件下, 任何抽象演化算法 AEA 次收敛.

- (i) 选择压 $S_p(t) \geq m > 0$;
- (ii) $\lim_{t \rightarrow \infty} [1 - (1 - E_o^m(t))S_I(t)]/E_e(t) = 0$;
- (iii) $\sum_{t=1}^{\infty} E_e(t) = \infty$ (i.e., $\{E_e(t)\} \notin l^1$).

进而, 如果还有

- (iv) $\lim_{t \rightarrow \infty} S_I(t) = \lim_{t \rightarrow \infty} E_l(t) = 1$,

则 AEA 强收敛.

上述定理条件有甚为直观的解释: (i) 是说选择算子应该有正的选择压 (换句话说, 能保证最优解个体增加), (iv) 是说这一事实必须渐近地以概率 1 成立 ($S_I(t) \rightarrow 1$). 而当 $S_I(t) \rightarrow 1$ 成立时, (ii) 指出 $E_o^m(t)/E_e(t) \rightarrow 0$, 即最终演化的迁出率应是迁入率的高阶无穷小; 而 (iii) 是指演化算子的迁入率不能太小. 显然, 这些条件是对 AEA 收敛性条件的合理刻画.

[33] 通过详细估计常见选择算子与演化算子的选择压、选择强度、保存率、迁入率、迁出率等参数, 导出了一系列具有重要应用价值的 GAs 收敛性结果. 例如, 他们计算出非线性 ranking 选择的选择压 $S_p = p(0 < p < N)$, 而选择强度 $S_I = q^N (q \in (0, 1))$, 而相邻锦标赛选择的选择压 $S_p = s-1$ (s 为锦标赛规模), $S_I = 1$; 由点变异算子构成的演化算子, 其保存率 $E_l = (1 - P_m)^N$, 迁入率 $E_e = 1 - (1 - P_m^L)^N$, 而迁出率 $E_0^m = [1 - (1 - P_m)^L]^m \cdot (1 - P_m^L)^{N-m}$ (P_m : 变异概率, L 为编码长度, N 为种群规模); 而由杂交与变异复合而成的重组算子其 $E_l = (1 - P_m)^N (1 - P_c)^N$, $E_e = [1 - (1 - P_m^L)^N] \lambda^N$, $E_0^m = [1 - (1 - P_c)(1 - P_m)]^m$. 由此可断言: 只要种群规模 N 大于编码长度 L , 变异概率 $P_m(t) \rightarrow 0$, 则由非线性 ranking 选择与变异组合而成的 GAs 强收敛.

以上所述的公理化模型也可用于非遗传算法类的其它模拟演化算法 (如演化策略) 的收敛性分析.

2.3 连续 (积分算子) 模型

大量数值试验表明, 为了有效求解高维连续问题和解决 GAs 实现中的效率与稳健性问题 (参见 [16]), 直接使用原问题的浮点表示而不进行编码转换常有许多优点. 几位作者发展了对于这类连续变量遗传算法收敛性的分析方法.

Peck 和 Dhawan(1995)(参见 [17]) 将 GAs 描述作全局随机搜索方法并论证了按照他们所给的准则, 这一观点相对于 Schema 理论和其它研究方法的恰当性与优点. 依他们的处理方法, GAs 的迭代过程被描述作可行解空间上某些抽样分布的构造与演化. 设标准 GAs 搜索可达的字符串空间为 Π , 其中的字符串记为 A 或 A_i . 假设在第 k 步迭代中, 适应值函数的估值含有噪音, 例如说, 所得结果为一非负随机变量 $y_k(A) = f(A) + \xi_k(A)$, 这里 $\xi_k(A)$ 为一随机变量, 则标准 GAs 可描述作如下的全局随机搜索方法

步 1 选取 Π 上的一个概率分布 P_1 , 令 $k = 1$;

步 2 依 P_k 抽样 N_k 次得串 $A_1^k, \dots, A_{N_k}^k$;

步 3 估计随机变量 $y_k(A_j^k)$ 在 A_j^k 处的值, 这里 $y_k(A) = f_k(A) + \xi_k(A) \geq 0$ (a. e. 1), f_k 是由 f 在 $A_j^i (1 \leq j \leq N_i, 1 \leq i \leq k)$ 处的观察值所构造的非负辅助函数. 若 $\sum_{j=1}^{N_k} y_k(A_j^k) = 0$, 返回步 2 重新进行抽样;

步 4 按照下式确定下一次的抽样分布

$$P_{k+1}(A_i) = \sum_{j'=1}^{N_k} \sum_{j''=1}^{N_k} p_{j'}^k p_{j''}^k \tilde{Q}_k(A_{j'}, A_{j''}, A_i), \quad (9)$$

其中

$$p_j^k = y_k(A_j^k) / \sum_{i=1}^{N_k} y_k(A_i^k). \quad (10)$$

步 5 若满足事先设置的终止准则, 则停止, 否则令 $k = k + 1$, 转步 2.

步 4 中 P_{k+1} 的实现可分为全局和局部两个阶段进行: 全局阶段相应于 GAs 中的选择算子: 依据当前种群中 f 的全局信息由 (10) 式来选择个体 $A_{j'}$ 和 $A_{j''}$; 局部阶段对应于 GAs 中包括杂交与变异的组合算子, 它由依赖于 $A_{j'}$, $A_{j''}$ 而确定的转移概率 $\tilde{Q}_k(A_{j'}, A_{j''}, \cdot)$ 来实现.

步 3 中的辅助函数 f_k 应充分反应 f 的特性.

代替应用字符串 A 及离散搜索空间 Π , 连续 GAs 直接应用原优化问题的可行解 x 与可行域 X , 它亦可描述为与上述完全类似的全局随机搜索方法. 特别地, 如设 M 是实现 Π 到 X

的变换, 则转移概率 $\tilde{Q}_k(A_{j'}, A_{j''}, \cdot)$ 转换到 X 上的抽样分布 $Q_k(M(A_{j'}), M(A_{j''}), \cdot)$, 此时代替 (9) 式, 我们可应用

$$P_{k+1}(dx) = \int_X \int_X R_k(dz') R_k(dz'') Q_k(z', z'', dx), \quad (11)$$

这里 $R_k(\cdot)$ 为 X 上的概率分布 (可看作是 (10) 的推广), 并对应于选择算子, Q_k 对应于组合算子.

对于上述描述, 有关全局随机搜索的已有收敛性理论 (例如参见 [18]) 可直接用于研究 GAs (特别是连续变量 GAs) 的收敛性, 具体地可参见 [17]. 由于是已有理论的套用, 且 [17] 中的结果强加许多不自然的假设, 加之, 对于 R_k 和 Q_k (特别是 R_1 和 Q_1) 的选择及相关参数的设定等问题也没有解决, 该方面的工作还有待进一步改进与完善.

通过研究大样本行为, Qi 和 Palmieri (1994) (参见 [19, 20]) 分别导出了连续 GAs 在使用比例选择、均匀杂交、选择和变异以及三个遗传算子联合作用等情形, 当种群规模趋于无穷时, 种群的概率分布所对应的密度函数应满足的递归公式. 例如, 假设定义于 $D \subset R^m$ 上的适应值函数 $f(x)$ 满足 $0 \leq f_{\min} \leq f(x) \leq f_{\max} < \infty$ 且 f 仅有有限个全局极大值和不连续点. 则在同时使用比例选择及变异算子情形, 他们证明了如下结论

定理 4 设变异算子以相同的条件概率密度函数 $g_{w_k}(\cdot | \cdot)$ ($\sup_{x, z \in D} g_{w_k}(x|z) \leq M < \infty$) 独立地作用于第 k 代种群的每个个体, 则当种群规模 N 趋于无穷时, GAs 的演化过程由具有下述递归密度函数的随机向量序列 $\{x_k\}_{k=0}^{\infty}$ ($x_k \in D$) 来刻画

$$g_{x_{k+1}}(x) = \frac{\int_D g_{x_k}(y) f(y) g_{w_k}(x|y) dy}{\int_D g_{x_k}(y) f(y) dy}. \quad (12)$$

考虑满足下述条件之可加变异干扰 (additive mutation noise)

$$x_{k+1}^i = x_k^i + w_k^i, \quad i = 1, 2, \dots, N,$$

这里 $x_k^1 = (x_k^1, \dots, x_k^{1N})$ 为选择算子作用于 x_k 后所产生的中间种群状态. 若 w_k^i ($1 \leq i \leq N$) 是均值为 0 的独立同分布 m 维随机向量, 且其共同密度函数为 $g_{w_k}(\cdot)$, 则此时可用 $g_{w_k}(x-y)$ 代替 (12) 式中之 $g_{w_k}(x|y)$. 对于这种情形, 他们还进而给出了保证种群平均适应值单调递增的下述充分条件.

定理 5 设目标函数 $f(x)$ 满足 Lipschitz 条件

$$|f(x) - f(y)| \leq L \|x - y\|, \quad \forall x, y \in D,$$

变异密度函数 g_{w_k} 球对称, 且其平均半径定义为

$$\bar{r}(k, x) = \int_D \|y - x\| g_{w_k}(y - x) dy.$$

则使种群平均适应值单调递增 (即 $E[f_{k+1}] \geq E[f_k]$, 这里 $f_k = f(x_k)$) 的充分条件是

$$\int_D f(x) \bar{r}(k, x) g_{x_k}(x) dx \leq \frac{1}{L} \text{var} [f_k].$$

当三个遗传算子联合作用时, 递归密度函数也可类似定理 4 确定 (当然比 (12) 要复杂得多), 详见 [20]. 但对于该完备情形, 作者并未能给出类似于定理 5 的收敛性结果. 注意到, 定理 5 对 f 的假设条件是相当苛刻的, 且该文结果只是在种群规模趋于无穷的条件下得到迭代序列分布的估计, 故只能看作是对 GAs 渐近行为的大样本近似. 换句话说, 我们并不能直接应用这里的结果去研究一般 GAs 的收敛性或指导、改进相应 GAs 的实际执行策略.

象上面两种方法一样, 可用于分析连续 GAs 的框架与方法均不完善, 存在着这样或那样的限制与不足. 目前还没有一个好的方法可用于准确描述连续 GAs 的动态行为, 并在不强加任何严格的或人为的条件下给出相关的收敛性结果.

3 收敛速度估计与复杂性

与收敛性分析紧密相关的另一基本理论问题是 GAs 能以多快的速度收敛. 这方面的研究不仅可从另一侧面来阐明 GAs 的收敛性, 而且对于建立合适的停机准则及恰当的度量标准以全面、客观地评判 GAs 的各种执行策略有重要意义.

然而, 到目前为止, 有关 GAs 收敛速度与复杂性研究仅有少量特殊的结果. 依照这些结果的具体形式, 该方面的工作可归纳为三类估计, 即收敛速度估计、迭代次数估计与时间复杂性估计.

3.1 收敛速度估计

取决于采用何种框架与分析方法来估计 GAs 的收敛速度, 人们通常采用不同的准则. 常见的准则, 例如有: 以种群平均适应值收敛, 以发现最优解的概率及趋于稳态分布的速率等.

对于一类特殊的组合优化问题: 1- 和问题 (即寻找含有最大数目非零字符的字符串问题), Rabinovich 和 Wigderson(1991, 参见 [21]) 分析了带比例选择与均匀杂交的二进制 GAs, 在演化种群平均适应值收敛的意义下之收敛速度. 考虑长度为 L 的二进制串 (于是 1- 和问题的最优值为 L), 设第 k 次迭代种群 P_k 的平均适应值为 $Av(P_k)$, 则他们证明了相应 GAs 的收敛速度为

$$(L - Av(P_k))(1 - \frac{1}{L}) \leq L - Av(P_{k+1}) \leq (L - Av(P_k))(1 - \frac{1}{2L}) \text{ (即线性收敛速度).}$$

以包含最优解的种群出现的概率和趋于 1 为准则并假设杂交概率为 1, Suzuki(1995, 参见 [10]) 利用概率转移矩阵的特征值导出了带有修正杰出选择 GAs 的几何收敛速度. 详见定理 2. 值得注意的是, 这一结果涉及一系列难以准确估计的参数 (如 A, C, φ 等), 因而可进一步精细化.

对于种群规模无限的 GAs, Voget(1996, 参见 [22]) 分析了当孤立、重复地应用单个遗传算子 (例如仅杂交算子) 情形演化种群的分布趋于极限分布的速度. 例如, 设 $E_k (1 \leq k \leq K)$ 为有限集合, $E = \prod_{k=1}^K E_k$ 为搜索空间. 用 $x = (x_i)_{i \in E}$ 表示任一种群, 其中 x_i 表示个体 i 在该种群中的相对比例. 若用 $X(n)$ 表示 n 时刻种群可能状态的随机变量, $x(n)$ 为其具体实现. 则当单独重复使用变异算子时, 他证明 GAs 服从以下速度估计

定理 6 设 $x(1)$ 为初始种群, 则第 n 代种群个体分布 $I^M(n)$ 为

$$P\{I^M(n) = i | X(1) = x(1)\} \\ = \sum_{j \in E} x_j(1) \prod_{k=1}^K \left[\frac{1}{|E_k|} + \left(1_{\{i_k=j_k\}} - \frac{1}{|E_k|} \right) (1 - P_m \frac{|E_k|}{|E_k| - 1})^n \right],$$

这里 i_k 表示个体 i 的第 k 个坐标, 且

$$\lim_{n \rightarrow \infty} P\{I^M(n) = i | X(1) = x(1)\} = \frac{1}{|E|},$$

$$\|P\{I^M(n+1) = i | X(1) = x(1)\} - \frac{1}{|E|}\|_1 \leq \|P\{I^M(n) = i | X(1) = x(1)\} - \frac{1}{|E|}\|_1.$$

显然, 上述结果虽然有助于理解单个遗传操作的渐近行为, 但由于未分析各遗传算子联合作用的实用情况, 较少有实用价值.

基于简单 GAs 的 Markov 链模型, 高勇^[23] 通过考察当前迭代种群的概率分布向量 θ_t 与相应稳态分布向量 π 的全变差距离 (记为 $\|\cdot\|$), 研究了简单 GAs 的收敛速度并证明了如下结果

$$\|\theta_t - \pi\| \leq (1 - (2P_m)^{NL})^t,$$

其中假设 $0 \leq P_m \leq 0.5$. 从上式可看到: 为了达到最快的收敛速度, 一方面最优的变异概率 P_m 应取为 0.5, 而另一方面, 此时对任意的 $t: \|\theta_t - \pi\| \equiv 0$. 这意味着在此种度量下无需进行任何 GAs 迭代. 由此我们认为: 这一度量方法及结果的有效性仍值得探讨.

对于可分函数 $f(x) = \sum_{i=1}^n f_i(x_i)$, 并假设对任一 x_i , GAs 只需一次迭代即可找到最优解且 $P_m \leq 1/n$. Salomon(1997, 参见 [24]) 分析了只使用变异算子的杰出选择型连续 GAs 的收敛速度. 以找到问题最优解的概率 P 为准, 他给出了用问题维数表示的如下估计

$$P \approx 1 - n \left(\frac{n-1}{n} \right)^t, \quad (13)$$

这里 t 为迭代次数. 由于所依赖的假设, 这一结果显然过分特殊而不具一般性.

如上述所示, 关于 GAs 收敛速度的已有结果基本上还只是考虑了二进制情形, 且主要还只是对某些特殊适应值函数类或仅使用某一单一遗传算子情形所作出的, 一般结果还知之甚少.

3.2 迭代次数估计

代替直接考虑收敛速度, 以下两篇论文通过估计在一定准则下 GAs 达到收敛所需的迭代次数 (或界) 来阐明 GAs 的收敛速度.

通过引进 Schema 的可靠性概念, 并建立相应的随机可靠性模型, Chakraborty 和 Dastidar (1993)(参见 [25]) 运用可靠性分析方法研究了简单 GAs 收敛所需的迭代次数. 这一研究的主要思路是: 先计算任一 Schema 的风险函数 (hazard function); 基于对风险函数的特定近似导出 Schema 可靠性的表达; 然后利用所导出的表达并运用可靠性分析方法来估计所需的迭代次数. 更具体地, 他们定义任一 Schema 的可靠性 $R(t)$ 为其直至第 t 代种群仍然存在的概率. 然后依据对风险函数的适当近似证明了

$$R(t) = e^{-\lambda(t\Phi + t(t-1)/2)},$$

这里 λ, Φ 为两个常数. 设 y_i 和 y_e 分别表示种群中所含低于平均适应值 (换言之, 不好的) Schema 的初始与最终比例, 并约定 GAs 的收敛性定义为种群中所含不好 Schema 的比例达到所要求的最终比率 y_e . 则由可靠性分析理论知 GAs 所需迭代次数 T 可由下式确定

$$y_i R(T) = y_e,$$

$$T = 0.5(1 - 2\Phi + \sqrt{(2\Phi - 1)^2 - (8/\lambda)\ln(y_e/y_i)}).$$

由于所运用的可靠性表达是对真正风险函数的近似, 上式所给出的迭代次数估计只能是对真实值的一个近似. 该估计的精度显然也依赖于参数 λ 与 Φ 的选取. 实际中, 确定 λ 与 Φ 的恰当值十分困难, 它不仅依赖于 GAs 的相关参数, 也会受所求解问题性态的影响.

定义 GAs 收敛性为在一定置信水平下访问过所有可能种群状态 (当然包含最优解), Aytug 和 Koehlen(1996)(参见 [26]) 利用 Markov 链中的首次访问时间 (first passage times) 概念, 并依据文 [5] 所建立的 Markov 链模型 (4) 与 (5) 导出了在所述准则下 GAs 所需迭代次数的上下界.

定理 7 设 t^* 为在置信水平 $\delta \in (0, 1)$ 下访问所有可能种群所需的迭代次数, 则 $t_l \leq t^* \leq t_u$, 其中

$$t_l = \left\lceil \max_j \left\{ \frac{\ln(1-\delta)}{\ln \rho(Q - Qe_j e_j^T)} \right\} \right\rceil, \quad (14)$$

$$t_u = \left\lceil \frac{\ln(1-\delta)}{\ln(1 - \min_{i,j} Q_{i,j})} \right\rceil = \left\lceil \frac{\ln(1 - P_m)}{\ln(1 - \min\{P_m^{NL}, (1-\mu)^{NL}\})} \right\rceil, \quad (15)$$

这里 e_j 为第 j 个单位向量, $\rho(\cdot)$ 表示矩阵之谱半径. $\lceil x \rceil$ 表示大于或等于 x 的最小整数.

上述结果适用于一般简单 GAs, 且所得界是用 GAs 的相关参数及其转移矩阵来表示的, 故易于在实际中应用. 但是, 正如文中所指出的, 上述估计并不是精确的, 特别是要求 GAs 以指定概率访问所有可能种群过于苛刻 (事实上我们只要求 GAs 以一定置信水平访问对应于最优解的某些特定个体即可). 另外, 估计式 (14) 与 (15) 并未利用适应值函数与杂交概率等信息, 故定理 7 的结果还有望作进一步改进.

3.3 时间复杂性估计

从另一个角度考察收敛速度, 近来也有文章研究 GAs 的时间复杂性. 这方面已有的结果如下:

(1) 引进适应值比率 (the fitness ratio) 的概念并假设其不变, Ankenbrandt(1991, 参见 [27]) 在导出了关于某一分量取特定值的比率之递推公式之后, 给出了 GAs 达到收敛所需时间的表达式及相应的平均、最坏情形时间复杂性估计.

以非二进制 GAs 为例, 他定义适应值比率为某个分量位置取特定值的那些个体的适应值与所有剩余个体的适应值的比率. 设 $P(t, a) (t = 0, 1, 2, \dots)$ 表示在任一固定分量 (例如 j) 处取值 a 的所有个体在 t 时的比例, 而分量 j 可取之值由字符表 $\{a, b, c, \dots\}$ 确定, 其个数用 $\text{card}(j)$ 表示. 令 f_a 表示在分量 j 处取值 a 的所有个体之平均适应值 (其它符号可类似定义), 则有

$$P(t+1, a) = \frac{f_a P(t, a)}{f_a P(t, a) + f_b P(t, b) + f_c P(t, c) + \dots}.$$

设 r 为此时的适应值比率, 且为常数, 则可由上式得

$$P(t, a) = \frac{r^t P(0, a)}{(1 - P(0, a)) + r^t P(0, a)}.$$

如假定某个分量 j 是最后一个达到收敛的, 则 GAs 收敛所需的时间 t_c 可从上式推出: 省去下标 a 并设 P_f 为分量 j 收敛时 $P(t)$ 应取的值, 则有

$$t_c = \ln \left[\frac{P_f(1 - P(0))}{P(0)(1 - P_f)} \right] / \ln r.$$

从此式可进一步导出其概率平均、最坏情形时 t_c 的值分别为

$$t_c^{av} = \frac{\ln [(N-1)(\text{card}(j)-1)] - 1}{\ln r},$$

$$t_c^{worst} = \ln (N-1)^2 / \ln r.$$

由此, GAs 在平均、最坏情形下的时间复杂性为 $O(\text{evaluate} \cdot N \ln N / \ln r)$, 这里 “evaluate” 表示依赖于可行域的函数估值的复杂性.

上述结果没有直接考虑杂交算子的影响, 且仅适用于比例选择情形. 由于假定适应值比率为常数, 所得结果显然还只是对实际情况的粗糙近似.

(2) 不考虑选择和变异的作用, Asoh 和 Muhlenbein(1994)(参见 [28]) 就简单抽样与均匀杂交这两个特殊情况, 给出了计算 GAs 平均收敛时间的公式. 对于均匀杂交, 他们证明了如下定理

定理 8 设 $r(k)$ 为直到 k 时个体每个分量达到收敛(成熟)的概率且该值对各个分量相同, n 为分量之个数, 则对于均匀杂交的种群, 其平均收敛时间 τ 由下式确定

$$\tau = \sum_{k=1}^{\infty} k(r(k)^n - r(k-1)^n).$$

因上述表达涉及无穷级数求和, 且依赖于其它参数值, 难于应用.

(3) 以种群的平均 Hamming 距离为度量, Louis 和 Rawlins(1993)(参见 [29]) 考虑了 GAs 收敛时间界的估计问题. 作者在忽略变异算子影响的条件下, 仅考虑了选择算子的作用.

设 p_0 表示在初始种群中某一二进制分量所占的比例, 并假定该值对每个分量相同, 则在第 t 次迭代时所有分量取值均固定(即收敛)的概率 $p(t, L)$ 可由下式确定

$$p(t, L) = \left[1 - \frac{6p_0(1-p_0)}{N} \left(1 - \frac{2}{N} \right)^t \right]^L.$$

给定置信水平要求后, 则可由上式得出 GAs 收敛所需时间的上界.

因变异对收敛的影响不可低估, 且上述结果无法描述杂交算子的作用, 该文的结果有待进一步改进.

(4) 对于可分适应值函数, Salomon(1997, 参见 [24]) 在分析只带变异的杰出选择型连续 GAs 时, 也得到了相应的时间复杂性结果

$$t = O \left(\frac{-\ln n}{\ln \left(\frac{n-1}{n} \right)} \right) \approx O(n \ln n),$$

其中后一近似等式当问题规模 n 充分大时成立。

从以上三小节的描述不难看出, 目前已有的关于 GAs 收敛速度与复杂性的结果几乎全部限于二进制情形, 而且或是对特殊的适应值函数、或是对仅使用某一个或某两个遗传算子操作情形的 GAs。所使用的度量准则各不相同, 应该说, 均存在相当的局限性。

4 评注与展望

以上我们概述了已有 GAs 收敛理论的主要结果, 分析了现有研究收敛性理论的不同模型与方法, 并在介绍各文献工作的同时简述了相应分析的优缺点及待改进之处。以下再就 GAs 理论研究存在的其它问题及可能的解决方法谈谈我们的想法。

从大的角度说, 目前尚没有一个好的框架(模型)与方法可用于研究一般 GAs 的收敛性。已介绍的四种模型要么只适应于离散(二进制或特殊的非二进制)GAs, 要么只适用于连续 GAs、要么要求种群规模为无穷。由于遗传算法不是对单点而是对种群的演化, 再加上其本身的随机性与固有的 Markov 特性, 我们认为运用测度值 Markov 过程及鞅收敛理论有望能在统一框架内研究各种形式 GAs 的收敛性, 并建立 GAs 的一般收敛性理论。

类似地, 对于 GAs 收敛速度的研究, 目前已有结果不仅只限于若干非常特殊的 GAs, 而且还没有一个通用的、合理的准则来度量 GAs 的收敛速度。我们感到, 一种可能的解决方法是: 通过修正种群平均 Hamming 距离^[29]使之可描述杂交、变异算子影响, 然后将其与适应值函数的信息结合起来设计收敛速度度量。由于修正的 Hamming 距离将能充分反映在各个遗传算子联合作用下种群的变化(如多样性、是否趋于某一稳定状态等), 而种群适应值变化反映迭代种群是否收敛到全局最优解, 结合这二者预期可给出较合理的收敛速度度量, 并有望建立统一的 GAs 收敛速度分析方法。

要从根本上解决上述 GAs 收敛性所存在的基础问题或许是相当困难的。但我们以为, 目前至少可从解决以下问题入手来扎扎实实推进该方面的研究:

(1) **对非简单 GAs 发展收敛性理论。** 当综合考虑三个遗传算子作用时, 已有研究基本上均只考虑了“比例选择、单点杂交与变异”的简单 GAs 情形, 而大量应用实践已经说明, 应用像杰出者选择、锦标赛选择、均匀杂交、多点杂交及非一致变异等非简单 GAs 操作的遗传算法有较简单 GAs 高得多的计算效率。所以对这些广泛形式的“非简单”GAs 发展收敛性理论有重要意义与迫切性。

(2) **对已有收敛性分析模型的深化研究。** 特别地, 对 Vose-Liepins 模型可考虑通过对 $S^{t+1} = \bar{F}\varphi(S^t)$ 的深入不动点结构分析与稳定性分析来清晰刻画 GAs 的收敛性; 对 Markov 链模型可考虑: (a) 从更精细的层次(例如从个体或基因层次)描述与分析 GAs; (b) 通过对概率转移矩阵的细致谱分析来研究 GAs 的收敛性与收敛速度; (c) 发展非遍历性分析方法; (d) 通过研究平稳分布的高阶矩信息, 获得 GAs 收敛状态的精确刻画; 对公理化模型可考虑各类 GAs 算子复合情形选择压、选择强度、迁入率与迁出率等演化参数的精确确定; 弱化一般收敛性定理; 并最终用于刻画 GAs 收敛性态与 GAs 参数(如种群规模, 杂交概率等)之间的函数关系, 且指导 GAs 参数的最优设置。

(3) **对 GAs 过早收敛现象的理论分析。** 应用 GAs 目的在于寻找所求解问题的全局最优解, 但 GAs 的应用实践却表明: GAs 常收敛到非最优或局部最优状态(即过早收敛现象)。徐宗本等(1997, 参见[31, 32])运用种群多样性概念仔细分析了简单 GAs 的过早收敛特征。问

题是：这一分析对于非简单 GAs 成立吗？特别是，如何从理论上刻画和平衡种群多样性与选择压，使在不发生过早收敛的前提下，GAs 尽快收敛？

(4) 关于 GAs 的加速收敛 (或效率加速) 理论. 除可有效用于传统优化 (搜索) 方法所不能求解的问题外, 现有 GAs 执行策略用于常规问题求解时其效率常不明显优于传统方法. 我们认为: 目前影响 GAs 算法效率的主要原因至少包括: (a) GAs 的随机操作引起算法执行中不可避免的 resampling; (b) 算法过分强化了对种群内部演化机制的模拟, 而忽略了在种群演化中环境对种群的反作用; (c) 广泛使用“杰出者选择”所引起的时间复杂性等. 目前急需研究在 GAs 执行中运用避免 (或利用) resampling 的有效策略、模拟在环境影响下受训种群的快速进化机制及建立非杰出者选择型 GAs 的收敛性理论.

(5) 关于 GAs 的有用性 (utility) 与有效性研究. 基于 Wolpert 和 Macreadg(1995)(参见 [30]) 的 No-Free-Lunch(NFL) 定理, 近年来有人对 GAs 的有用性与有效性提出质疑 (如见 Salomon[24]). 假设所有可能的目标函数服从均匀分布, NFL 定理的基本结论是: 相对于所有可能的目标函数, 任何最优化方法在平均意义下是等同有效的. Salomon 据此认为: GAs 与别的优化方法至多是一样有效的, 而通过对简单 GAs 的分析, 他进而认为: GAs 事实上不会比其它确定性优化算法有效! 果真如此吗? 是否 GAs 根本就无用? 是否 GAs 根本就不可能比其它算法有效? 回答这些问题显然极富挑战性, 而且需要时间, 需要对 GAs 理论本身的透彻理解, 需要对 GAs 执行策略的不断发展. 然而, 我们的看法是: 基于 NFL 定理来扼杀 GAs 明显是错误的. 这是因为, 首先, NFL 定理所基于的“所有可能目标函数类型服从均匀分布”的假设并不符合实际 (例如, 线性函数与非线性函数等概率出现吗?); 其次, GAs 的产生与发展其主要动机 (从而应用对象) 在于求解那些通常优化方法所无法或难以求解的复杂优化问题 (像非数值优化、函数不可精确估值问题的优化、极多峰、多目标问题的优化等), 以平均表现来否定 GAs 的效率显然有失公允; 再则, GAs 是正处发展中的新技术, 应该说只是有了一个基本的框架, 而有关它的搜索机制、基本理论与有效执行策略等还只是正在探索中的问题, 只是通过分析“简单 GAs”而否定一般 GAs 显然是不科学和危险的. 当然, 目前的问题是, 我们急需对 GAs 的本质与基础理论作透彻的研究, 需要将 GAs 研究 (特别是算法设计与应用) 聚焦于复杂优化问题, 而真正发展出对复杂问题有效 (而其它方法不可用) 的 GAs 执行策略.

参考文献

- 1 Vose M D and Liepins G E. Punctuated equilibria in genetic search. *Complex Systems*, 1991, 5: 31-44.
- 2 Koehler G J. A proof of the Vose-Liepins conjecture. *Annals of Mathematics and Artificial Intelligence*, 1994, 10: 409-422.
- 3 Bhattacharyya S and Koehler G J. An analysis of non-binary genetic algorithms with cardinality 2^n . *Complex Systems*, 1994, 8: 227-256.
- 4 Wright A H and Vose M D. Finiteness of the fixed point set for the simple genetic algorithm. *Evolutionary Computation*, 1996, 3: 299-309.
- 5 Nix A E and Vose M D. Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 1992, 5: 79-88.
- 6 Vose M D. Modeling simple genetic algorithms. *Evolutionary Computation*, 1996, 3: 453-472.
- 7 Horn J. Finite Markov chain analysis of genetic algorithms with niching. *Proceedings of the Fifth International Conference on Genetic Algorithms*, San Mateo, Morgan Kaufmann, 1993, 110-117.
- 8 Mahfoud S W. Finite Markov chain models of an alternative selection strategy for the genetic algorithm. *Complex Systems*, 1993, 7: 155-170.

- 9 Dawid H. A Markov chain analysis of genetic algorithms with a state dependent fitness function. *Complex Systems*, 1994, 8: 407-417.
- 10 Suzuki J. A Markov chain analysis on simple genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 1995, 25: 655-659.
- 11 Rudolph G. Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 1994, 5: 96-101.
- 12 Reynolds D and Gomata J. Stochastic modelling of genetic algorithms. *Artificial Intelligence*, 1996, 82: 303-330.
- 13 Davis T E and Principe J C. A simulated annealing like convergence theory for the simple genetic algorithm. *Proceedings of the Fourth International Conference on Genetic Algorithms*, San Mateo, Morgan Kaufmann, 1991, 174-181.
- 14 Cerf R. Asymptotic convergence of a genetic algorithm. *C. R. Acad. Sci. Paris, t.319, Serie I*, 1994, 271-276.
- 15 Eiben A E etc. Global convergence of genetic algorithms: a Markov chain analysis, *Parallel Problem Solving from Nature*, Berlin, Springer-Verlag, 1991, 4-12.
- 16 Goldberg D E. Real-coded genetic algorithms, virtual alphabets, and blocking. *Complex Systems*, 1991, 5: 139-167.
- 17 Peck C C and Dhawan A P. Genetic algorithms as global random search methods: an alternative perspective, *Evolutionary Computation*, 1995, 3: 39-80.
- 18 Zhigljavsky A A. *Theory of global random search*, Kluwer Academic Publishers, 1991.
- 19 Qi X and Palmieri F. Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space part I: basic properties of selection and mutation. *IEEE Transactions on Neural Networks*, 1994, 5: 102-119.
- 20 Qi X and Palmieri F. Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space part II: analysis of the diversification role of crossover. *IEEE Transactions on Neural Networks*, 1994, 5: 120-129.
- 21 Rabinovich Y and Wigderson A. An analysis of a simple genetic algorithm, in *Proceedings of the Fourth International Conference on Genetic Algorithms*, San Mateo, Morgan Kaufmann, 1991, 215-221.
- 22 Voget S. Theoretical analysis of genetic algorithms with infinite population size. *Complex Systems*, 1996, 10: 167-183.
- 23 Gao Y. *Stochastic Model, Theoretical Analyses and Applications of Genetic Algorithms*, Ph.D. dissertation, Xi'an Jiaotong University, Xi'an, 1997 (in chinese).
- 24 Salomon R. Raising theoretical questions about the utility of genetic algorithm. *Evolutionary Programming 4*, Lecture Notes in Computer Science 1213, 1997, 275-284.
- 25 Chakraborty U K and Dastidar D K. Using reliability analysis to estimate the number of generations to convergence in genetic algorithms. *Information Processing Letters*, 1993, 46: 199-209.
- 26 Aytug H and Koehler D G. Stopping criteria for finite length genetic algorithms. *INFORMS Journal on Computing*, 1996, 8: 183-191.
- 27 Ankenbrandt C A. An extension to the theory of convergence and a proof of the time complexity of genetic algorithms. *Foundations of Genetic Algorithms*, San Mateo, Morgan Kaufmann, 1991, 53-68.
- 28 Asoh H and Muhlenbein H. On the mean convergence time of evolutionary algorithms without selection and mutation. *Parallel Problem Solving from Nature*, Berlin, Springer-Verlag, 1994, 88-97.
- 29 Louis S J and Rawlins G J E. Syntactic analysis of convergence in genetic algorithms. *Foundations of Genetic Algorithms 2*, San Mateo, Morgan Kaufmann, 1993, 141-151.
- 30 Wolpert D, Macreadg W. No free lunch theorem for search. Technical Report 95-02-010, Santa Fe Institute, NM, 1995.
- 31 徐宗本, 高勇. 遗传算法过早收敛现象的特征分析及其预防. *中国科学 (E)*, 1996, 26(4): 364-375.
- 32 张讲社, 徐宗本, 梁怡. 遗传算法的整体退火选择及其收敛充要条件. *中国科学 (E)*, 1997, 27(2): 154-164.
- 33 徐宗本, 段启宏, 张文修. 模拟演化计算的一个公理化模型 (待发表).

Theoretical Development on Genetic Algorithms: A Review

Xu Zongben Chen Zhiping

*(Institute of Information Science and System Science, Faculty of Science, Xi'an Jiaotong Univ., Xi'an,
Shaanxi, 710049, P. R. China)*

Zhang Xiangsun

*(Institute of Applied Mathematics, Research Institute of Mathematics and System Science, The Chinese
Academy of Sciences, Beijing, 100080, P. R. China)*

Abstract Recent developments on the convergence and convergent speed analysis of genetic algorithms(GAs) are reviewed in this paper. While the main focus of this review is placed on the uniform account of the existing theoretical models such as the Vose-Liepins model, the Markov chain model, the axiomatization model and the continuous(integral operator) model, theoretical advance on the complexity analysis of GAs has been presented. A series of open problems on GAs analysis is also included.

Key words genetic algorithms; convergence; convergent speed analysis; models; time complexity