DOI:10.16262/j.cnki.1000-8217.1999.02.006

·学科进展·

广义遗传算法的数学结构

董　聪　　郭晓华

(清华大学,智能技术与系统国家重点实验室,北京 100084)

[摘　要]　广义遗传算法以 Morgan 的基因理论及 Eldridge 与 Gould 的间断平衡理论为依据,同时融合了 Mayr 的边缘物种形成理论和 Bertalanffy 一般系统理论的一些思想,这是其在生物学原理上优于经典遗传算法的本质所在。本文对经典遗传算法和模拟退火算法的数学基础进行了系统的分析,指出其效率低的原因所在,在此基础上,重点阐述了广义遗传算法的数学结构,对该算法的全局收敛性进行了证明。

[关键词]　广义遗传算法,基因理论,一般系统理论,数学结构,全局收敛性

20世纪70年代前后,北美和欧洲的一些科学家开始探索用模拟生物进化的方式求解复杂优化问题,其中,以 Holland 提出的遗传算法(genetic algorithm, GA)最具代表性[1]。80年代以来,Hopfield 解决了反馈型人工神经网络(Artificial Neural Network, ANN)的迭代稳定性问题[2],Rumelhart 等人重新发现了 BP 算法,解决了多层前向网络中隐节点的学习问题[3],董聪等人对 BP 算法做了推广[4],揭示了前向网络逼近与泛化的内在机理[5,6],建立了前向网络拓扑结构学习的通用算法[7],解决了前向网络对于离散点集的全局最优逼近问题[8]。ANN 在探索智能形成机制方面所取得的成就给人以启示:研究智能的最好方式是向人类自身学习,在几十亿年的进化过程中,生物体形成了一种优化自身结构以适应环境变化的内在机制。ANN 对 AI 的主要贡献是解决了智能系统如何从环境中自主学习的问题,GA 的新一代支持者则试图揭示学习过程在基因层次上究竟如何实现[9]。

经典遗传算法在生物原理方面以 Darwin 的进化论(1859)和 Mendel 的遗传因子理论(1865)为依据;在算法实现方面,它具备了结构上的隐含并行性、计算原理上的随机性和自适应性。关于随机优化算法可以解决非线性系统的全局最优化问题,自适应方法可以解决机器学习问题,并行算法具有极高的计算效率的信念,使得遗传算法的研究和拓展迅速成为国际学术界和工程界关注的热点[1,8-12]。董聪对经典遗传算法的操作程序进行了系统的剖析[9],纠正了作为经典遗传算法数学基础的 Schema 定理中的一些错误,发现并证明了广义 Schema 定理。董聪指出,算法的随机性并不是保证算法具有解决非线性系统全局优化问题的充分条件,要真正实现全局最优化的功能,必须通过设计合理的算法结构来保证。

广义遗传算法最初是作为一种求解全局优化问题的数学方法提出来的,目的在于解决经典遗传在逻辑结构和计算原理上存在的一些问题,并没有特别强调其非 Darwin 进化论的生物学原理[9]。文献[12]采用广义遗传算法成功地解决了目前世界上单跨最长的铁路、公路两用桥——香港青马大桥传感器群的最优布点设计问题,同样的问题在经典遗传算法的框架下得不出合理的设计方案。无论是模拟计算还是实际应用,广义遗传算法在计算效率方面的优势都很明显,但并不清楚这个早先源于逻辑推理的算法是否为生物系统实际采用。文献[13]对广义遗传算法重新做了诠释,并正式将 Morgan 的基因理论(1920)和 Eldridge 与 Gould 的间断平衡理论(1972)作为其主要的生物学依据。与此同时,广义遗传算法还融合了 Mayr 的边缘物种形成理论

国家自然科学基金(批准号 59505011,59778039)、航空基金、国家重点实验室基金和攀登计划重大项目资助课题.
本文于1998年7月20日收到.

(1963)和 Bertalanffy 一般系统理论(1973)的一些思想。文献[13]标志着广义遗传算法和经典遗传算法在生物学原理上的彻底分离。本文重点阐述了广义遗传算法的数学结构,对该算法的全局收敛性进行了证明。

1 广义遗传算法的数学结构

求解以下非线性优化问题:

$$\max f(X), \quad X=(x_1,x_2,\cdots,x_n)\in \mathbf{K}\subset R^n \tag{1}$$

其中,$\mathbf{K}$ 为紧子集。定义 $S=x_1x_2\cdots x_n$ 为串,S 中的任一固定段为模式(Schema),x_i 为染色体,x_i 中的固定位为基因,$f(X)$ 为适应值函数。从形式上看,串 S 对应于 Morgan 基因理论中的基因型,适应值函数 $f(X)$ 对应于表现型,Schema 对应于基因族或基因组。采用上述符号,广义 Schema 定理可表示为[9]:

$$m(H,k+1)\geq m(H,k)\frac{\bar{f}(H)}{\bar{f}}\left(1-P_c(k)\frac{d(H)}{l-1}\right)^{O_c(H)}\left(1-P_m(k)\frac{d'(H)}{l}\right)^{O_m(H)} \tag{2}$$

其中,$m(H,k)$是第 k 代种群中含有给定模式 H 的串的总数;$\bar{f}(H)$是种群中含有给定模式 H 的串的平均适应值;$\bar{f}$ 是种群的平均适应值;$d(H)$是模式 H 的长度(模式 H 最外侧的两个给定基因间的距离);$d'(H)$是模式 H 所包含的基因总数;l 是串的长度;P_c 是交换操作概率;$O_c(H)$是交换阶次;P_m 是突变操作概率;$O_m(H)$是突变阶次。

经典 Schema 定理的表达形式为[1]:

$$m(H,k+1)\geq m(H,k)\frac{\bar{f}(H)}{\bar{f}}\left(1-P_c(k)\frac{d(H)}{l-1}\right)(1-P_m(k))^{O_m(H)} \tag{3}$$

或

$$m(H,k+1)\geq m(H,k)\frac{\bar{f}(H)}{\bar{f}}\left(1-P_c(k)\frac{d(H)}{l-1}\right)(1-P_m(k)O_m(H)) \tag{4}$$

仅当突变点位于 Schema 域内的基因位时,突变操作才会破坏给定模式。因此,经典 Schema 定理对突变操作破坏给定模式的概率描述并不正确。不难看出,广义 Schema 定理除了纠正经典 Schema 定理中的错误外,还将其适用范围做了延拓。

Darwin 进化论的核心在于自然选择,而自然选择的基本素材是生物界普遍存在着的变异(包括突变)现象。Darwin 认为,可遗传的变异是实现进化的前提。从形式上看,广义 Schema 定理没有说明有利的变异在进化过程中如何通过选择进行累积,变异在其中的作用似乎只是破坏性的,而不是它本应具有的积极作用。另外,受 Darwin 进化论的影响,经典遗传法更强调选择的重要性,而忽略了突变在进化、特别是在定向进化过程中的积极作用。广义 Schema 定理证明了可通过选择使优良个体的繁衍机会增加,从而有利于实现局部最优化;但对交换和突变在确保解空间各点可达性方面的重要作用,却未能提供任何有利的依据。换句话说,无论从经典 Darwin 进化论的角度还是从其他更广泛的角度来看,经典遗传算法在数学结构上均存在缺陷。

与 Holland 以归纳逻辑的方式从 Darwin 的进化论和 Mendel 的遗传因子理论出发建立经典遗传算法的思路不同,广义遗传算法是通过演绎逻辑的方式建立起来的:首先建立一个逻辑上合理的算法框架,然后证明这种框架在数学上是高效的,最后是检验该算法是否为生物系统所采用。

从数学上讲,一种算法要想具备实现全局最优化的功能,它只需满足两个条件:(1)具有实现局部最优化的能力;(2)具有从一个局部最优状态向下一个更好的局部最优状态定向转移的能力。正是在这样一种思想的指导下,我们创立了广义遗传算法。

在生存选择中,广义遗传算法采用了四(n 点交换时实际为 2^{n+1}。为论述简洁,以下均称为"四")分之二择优选择的方式,即在基因交换或基因突变的过程中,允许父辈和子代进行竞争,并让其中的优良者进入下一轮的竞争环境。此举保证了算法的迭代稳定性,并使其具备了实现局部最优化的功能。广义遗传算法借助于以基因型的多点突变操作为主,以基因交换为辅的策略,实现从一个局部最优状态向下一个更好的局部最优状态的定向转移。由于采用了局部最优状态的定向转移,使算法为获得全局最优,其搜索域不必覆盖整个解空间,而只需覆盖感兴趣的有限解空间即可。换句话说,广义遗传算法摈弃了传统随机优化方法普遍采用的遍历搜索策略,转而采用定向演化模式。

2 广义遗传算法全局收敛性证明

设 $f(X)$ 存在 $m(m\geq 1)$ 个取值不同(相同者进行归并)的状态,并将各状态根据其取值大小按升阶排列:$f(X_1)<f(X_2)<\cdots<f(X_i)<\cdots<f(X_j)<\cdots<f(X_m)$。设系统目前正处在状态 X_i,定义:

$$X_i^+=\{X:f(X)>f(X_i)\} \tag{5}$$

将 X_i^+ 所包含的状态数定义为 Ω_i,则 Ω_i 为 i 的单调减函数。

定义 t_{ij} 为系统从状态 X_i 转移到状态 X_j 的概率,则

$$P(X_i \to X_i^+) \sum_{j=i+1}^{m} P(X_i \to X_j) = \sum_{j=i+1}^{m} t_{ij} \geq P(X_i \to X_j) = t_{ij} \qquad (\forall\, i < j \leq m) \tag{6}$$

其中,$P(X)$表示事件(状态)X 发生(出现)的概率。

经典遗传算法实现全局最优功能的前提之一是要求状态 X_i 和状态 X_j 之间相通($X_i \to X_j$)[14]。从式(6)可以看出,状态的单向可达性($X_i \to X_i^+$, $\forall\, i < m$)是广义遗传算法实现全局最优的充分条件。进一步,我们从最大值状态开始,按单向可达性关系逆向绘制一个树状网络,可以推知,只要当前状态位于或可达上述树状网络的任意一个枝杈,则必可保证广义遗传算法收敛于全局最优状态。解空间的遍历搜索对经典遗传算法实现全局最优功能是必要的,但对广义遗传算法则并不必要。事实上,解空间的遍历性是广义遗传算法实现全局最优功能的强充分条件。同样可以看出,选择算子的作用在于保证进化进程的方向性,因此,经典遗传算法将选择算子取为概率算子的做法并无必要。综上所述,与经典遗传算法相比,广义遗传算法实现全局最优化的条件非常宽松。

对于组合优化问题,解空间的状态数是有限的。由于局部极大值点的数目恒少于其状态数,因此,广义遗传算法必可保证在有限步内实现问题的全局最优化。而且从香港青马大桥传感器群的最优布点设计成功,说明广义遗传算法的计算效率达到了大型工程上可实用的程度。

3 经典遗传算法和模拟退火算法的全局优化能力

在经典遗传算法和模拟退火算法全局优化能力的证明中,均使用了有限时齐 Markov 过程这一重要的数学工具。下面对有限时齐 Markov 过程的有关特性及与经典遗传算法和模拟退火算法的关系做简要的介绍和评论。

定义1 有限时齐 Markov 过程是指满足下列条件的一类随机过程:(1)状态集为一个有限集合 $\mathbf{X} = \{X_1, X_2, \cdots, X_m\}$;(2)从状态 X_i 转移到状态 X_j 的转移概率 P_{ij}和时间起点无关。根据以上2点,可定义 $m \times m$ 阶矩阵 $\mathbf{P} = (P_{ij})$为概率转移矩阵。$\mathbf{P}$ 满足以下关系:

$$P_{ij} \geq 0, \quad \sum_{j=1}^{m} P_{ij} = 1, \quad 1 \leq i, j \leq m \tag{7}$$

定义 m 维向量 $\Pi^0 = (\pi_1^0, \pi_2^0, \cdots, \pi_m^0)$,记 $\mu = (\mathbf{P}, \Pi^0)$表示初态为 Π^0,概率转移矩阵为 $\mathbf{P}$ 的有限时齐 Markov 链。则对一切正整数 k 有

$$\Pi^k = \Pi^{k-1}\mathbf{P} = \Pi^0\mathbf{P}^k \tag{8}$$

其中,π_1^0 表示过程 X 初始时刻 $X(0)$处于状态 X_1 的概率。Π^k 是行向量,其第 i 个分量是第 k 步后过程 $X(k)$处于状态 X_i 的概率。可以看出,有限时齐 Markov 过程在任意时刻的状态 Π^k 由其初始分布 Π^0 和一步转移概率矩阵 $\mathbf{P}$ 唯一决定。

定义2 记 $P_{ij}(k)$为时齐 Markov 链的 k 步转移概率。设 $X_i, X_j \in \mathbf{X}$,如果存在某一个正整数 k,使得 $P_{ij}(k) > 0$,则称状态 X_i 可达状态 X_j,记作 $X_i \to X_j$。反之,如果对一切正整数 k,都有 $P_{ij}(k) = 0$,则称状态 X_i 不可达状态 X_j,记作 $X_i \not\to X_j$ 且 $X_j \to X_i$,则称状态 X_i 和状态 X_j 是相通的,记作 $X_i \leftrightarrow X_j$。

定义3 设 $X(0) = X_i, X_i \in \mathbf{X}$,称 $T_{ij} = \min\{k \mid X(0) = X_i, X(k) = X_j\}$为从状态 X_i 出发到状态 X_j 的首达时间,其中,k 为正整数。设 $\pi_i^0 = P(X(0) = X_i) > 0, X_i \in \mathbf{X}$,则

$$f_{ij}(k) = P(T_{ij} = k \mid X(0) = X_i) = P(X(k) = X_j, X(s) \neq X_j, s = 1, 2\cdots, k-1 \mid X(0) = X_i) \tag{9}$$

称为自状态 X_i 出发,经 k 步转移首次到达状态 X_j 的概率,简称首达概率。

定义4 $f_{ij} = \sum_{k=1}^{\infty} f_{ij}(k)$　　(10)

定理1 $f_{ij} > 0$ 的充要条件是 $X_i \to X_j$, $X_i \leftrightarrow X_j$ 的充要条件是$f_{ij} > 0$ 和 $f_{ji} > 0$。

定义5 如 $f_{ii} = 1$,则称状态 X_i 是常返状态,如 $f_{ii} < 1$,则称 X_i 是非常返状态,或滑过状态。

定理2 如 X_j 是常返的,则从状态 X_i 出发,Markov 链将以概率1无穷多次进入状态 X_j;如 X_j 是非常返状态,则从状态 X_i 出发,Markov 链无穷多次进入状态 X_j 的概率是0,或者说 Markov 链只能有限次地进入状态 X_j。

定义6 一个 Markov 链,如果除了整个状态空间构成一个闭集外,不可能再分解出更小的闭集,则称 Markov 链为不可约的。

定理3 不可约有限 Markov 链只有正常返状态。

定义7 如果对 $X_i, X_j \in \mathbf{X}$,存在不依赖于 X_i 的极限 $\lim_{k \to \infty} P_{ij}(k) = P_j > 0$,则称此时齐 Markov 链具

有遍历性。

定义 8 如果一个 Markov 链 $\mu = (\mathbf{P}, \Pi^0)$，正在一个正整数 s，使得 $\Pi^{s+k} = \Pi^s$，$(k = 1, 2, \cdots)$，则称其概率分布向量 Π^s 是平稳的，或把 Π^s 叫做平稳概率分布向量。

定理 4 任一 Markov 链都有一个平稳的概率分布向量。

定理 5 Markov 链是遍历的，其充分必要条件为它有唯一的正的平稳概率分布向量，且该正的平稳概率分布向量就是它的极限分布向量。

经典遗传算法和模拟退火算法全局优化能力的证明方式是类似的：首先构造一个概率转移矩阵 $\mathbf{P}$，证明其所属的时齐 Markov 链是遍历的，由定理 5 知，它有唯一正的平稳概率分布向量，且该平稳概率分布向量就是它的极限分布向量。由于极限分布向量是正的，故算法具有全局优化能力。

事实上，只需极限分布向量中对应于最大(最小)值状态的概率分布分量为正，即可证明相应的算法具有全局优化的能力，再进一步分析，设算法的初始状态为 X_i，目标状态为 X_o，定义 $T_{io} = \min\{k \mid X(0) = X_i, X(k) = X_o\}$，则全局优化算法设计的目标应当是 T_{io} 在某种意义上(如在均值意义上)为最小。可见，经典遗传算法和模拟退火算法在算法设计上存在着目标不明确、功能冗余等问题，这些问题从根本上降低了算法的计算效率。

在进化程序上，广义遗传算法和经典遗传算法有所不同。经典遗传算法的进化程序为：双亲选择→基因交换→基因突变→生存选择→下一代种群；广义遗传算法的进化程序为：双亲选择→基因交换→一家四口→四分之二生存选择→基因突变→一家四口→四分之二生存选择→下一代种群。也就是说，广义遗传算法采用了定向进化的方式，并将选择原则贯彻于整个生命周期。

参 考 文 献

[1] Holland J H. Adaptation in Natural and Artificial Systems. MIT Press, 1992.

[2] Hopfield J J Neural network and physical systems with emergent collective computation abilities. Proc. National Academy Science, U. S. A 1982, **79**:2554—2558.

[3] Rumelhart D E, Hinton G E, Williams R J Learning representations by backpropagation errors. Nature, 1986, **323**(6188):533—536.

[4] 董聪，刘西拉．广义 BP 算法及网络容错性和泛化能力的研究．控制与决策，1998，**13**(2)：120—124.

[5] 董聪，郦正能，夏人伟等．多层前向网络研究进展及若干问题 力学进展，1995，**25**(2)：186—196.

[6] 董聪．多层前向网络的逼近与泛化机制．控制与决策，1998，**13**(增刊).

[7] 董聪．多层前向网络的逼近机理与拓扑结构学习方法．通信学报，1998，**19**(3)：29—34.

[8] 董聪．前向网络全局最优化问题研究．中国科学基金，1997，(1)：23—29.

[9] 董聪．广义遗传算法．大自然探索，1998，**17**(1)：33—37

[10] Krishnakumar K, Goldberg D E. Control system optimization using genetic algorithms. J. Guidance, Control, and Dynamics, 1992, **15**(3): 735—740

[11] Bui T N, Moon B R. Genetic algorithm and graph partitioning. IEEE Trans. on Computers, 1996, **45**(7):841—855.

[12] 董聪，秦权，陈一珉．基于广义遗传算法的青马桥传感器最优布点设计．CCNS'97，南京：人民邮电出版社，1997，479—484.

[13] 董聪，郭晓华．广义遗传算法的逻辑结构及全局收敛性的证明．计算机科学，1998，**25**(5).

[14] Davis T E, Principe J C. A simulated annealing like convergence theory for the simple genetic algorithm. Proc. the 4th Conf on Genetic Algorithm, 1991, 174—181.

THE MATHEMATICAL STRUCTURE OF GENERALIZED GENETIC ALGORITHM

Dong Cong Guo Xiaohua

(*Tsinghua University, The State Key Laboratory of Intelligence Technology and Systems, Beijing* 100084)

Abstract Generalized genetic algorithm is based on some famous modern biologics theories such as the genetic theory by Morgan, the punctuated equilibrium theory by Eldridge and Gould, and the general system theory by Bertalanffy, so it is superior in biologics to classical genetic algorithm. In this paper, a systematic analysis is made on the mathematical principles of the classical genetic algorithm and simulated annealing algorithm for solving global optimization problems of system, and the causes with lower computation efficiency of these algorithms are discovered. On this basis, the mathematical structure of the generalized genetic algorithm is presented, and a proof on global convergence of the present algorithm is made.

Key words generalized genetic algorithm, genetic theory, general system theory, mathematical structure, global convergence