

遗传算法——进化论思想与遗传学原理 在工程优化中的应用

禹慧明 姚汝华 林炜铁

(华南理工大学生物工程系 广州 510641)

摘 要 优胜劣汰是自然界物种进化的法则,近年来,生物科学中的进化论思想与遗传学原理被成功地应用于工程中优化问题的计算,于是产生了一种不同于传统算法的优化算法——遗传算法(GA)。本文介绍遗传算法所采用的进化论思想和遗传学原理,遗传算法的基本操作、算法步骤、不同于传统算法的特点以及遗传算法的发展历史与应用情况等,并对遗传算法对生物科学的可能应用作了简单的展望。

关键词 进化,遗传,优化,遗传算法

众所周知,生物是通过两个基本过程:自然选择和有性繁殖不断进化的。生物通过自然淘汰、变异、遗传进行进化,以适应环境的变化,产生最适个体。1971年,美国密执安大学的心理学、电工学及计算机科学教授 John H. Holland 及其同事、学生将这种进化的思想和遗传学的原理应用于工程优化计算,提出了一种全新的、风格极不同于传统寻优算法的算法——遗传算法(或称基因算法)^[1,2] (Genetic algorithm,简称 GA)。二十多年来,GA吸引了美国、欧洲、日本及中国计算机科学、自动化学科等许多领域众多学者的研究兴趣,取得了丰硕的理论成果和成功的应用。GA是一种仿生学算法,其收敛性没有绝对的保证,然而事实证明 GA 是一种非常有效的算法,具有优越的性能,尤其适合于复杂问题的求解。当我们这些生物学研究工作者阅读了有关 GA 的文献、观摩了 GA 的精英表演之后,深感意外、欣喜,对仿生学这一学科充满了乐观的期待。

1 GA 的思想与算法概述

1.1 GA 的思想

GA 的思想基于达尔文(Darwin)的进化论与孟德尔(Mendel)的遗传学原理。达尔文的进化论认为:每一物种在不断的发展过程中越来越适应环境,只有最适个体将进一步生存、再生,将遗传物质传递给子代,而不适者将被迫灭绝,那些更能适应环境的个体特征被保留下来。这就是适者生存的原理。

孟德尔的遗传学原理认为:遗传是通过细胞中的遗传密码进行的,遗传密码封装于细胞之中,并以基因的形式包含在染色体中。每个基因有特定的位置并控制某个特殊的性质,每个基因产生的个体对环境有一定的适应性。基因的杂交和变异产生更适应环境的子代,通过优胜劣汰的自然选择,对环境适应值高的基因结构被保留下来。GA 在求解一个优化问题时,将操作变量(自变量)表示成“染色体”,在计算机上一般用二进制码串,例如: $19(\text{十进制}) = 1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 10011(\text{二进制})$,选择一群(Population)“染色体”——个体(Individuals),将它们置于问题的“环境”(约束)中,对它们进行交叉、变异、再生等“基因操作”,淘汰对环境适应性差的“染色体”,保留对环境适应值高的“染色体”,反复操作,筛选出最适个体,从而求出问题的解。

1.2 GA 的基因操作

按照 GA 的思想,GA 是通过参数群体的不断进化获得最适个体的,GA 在进行进化时模仿生物进化中遗传过程的基因操作。GA 采用的基因操作主要有交叉(Crossover)、变异(Mutation)与再生(Reproduction)等,分述如下:

1.2.1 交叉: 基因的交叉也叫基因重组,是指两个或两个以上的亲代染色体各自取出自身的部分密码进行交换,组成新的个体的过程。GA 的交叉是按照一定的概率 P_0 。随机地选取部分个体进行交叉, P_0 也叫“交叉率”。图 1 演示了对两个染色体进行交叉操作的过程。[] 所含部分密码为交叉的部分。

Parent 1 1 1 0 0 0 [1 0 1] Crossover offspring 1 1 1 0 0 0 0 0 1
Parent 2 0 0 1 1 1 [0 0 1] → offspring 2 0 0 1 1 1 1 0 1

图 1 交叉算子(操作)

Fig.1 Crossover operator

1.2.2 变异: 变异是指将染色体的某些密码进行突变,在二进制中,就是 0 变 1,1 变 0。在 GA 中采用变异时,按照一定的概率 P_0 对染色体群体中的某些个体的某些密码进行变异。 P_0 称为“变异率”。图 2 演示了将一个染色体进行变异的过程,其中[]所包含的部分为进行变异的部分。变异的作用是产生新的个体,它能有效地防止产生不能进化的单一个体。当所有个体一致时,用交叉产生新个体失效,这时变异是有效的手段。

Parent generation 1 [0] 0 1 1 [1 0 0] 1 Mutation Filial generation 1 1 0 1 1 0 1 1 1

图 2 变异算子(操作)

Fig.2 Mutation operator

1.2.3 再生: 再生就是复制。GA 中的再生是按照一定的概率将群体中的部分个体进行复制。生物学中基因的遗传还有倒位、双倍体等现象,因此在 GA 中也引入了“倒位操作”等基因操作,利用这些能进一步提高 GA 的效率。关于这些操作,可参考有关文献。

1.3 GA 描述

一般的优化问题可表述为:

$$\begin{aligned} \min F &= F(X), F \in R \\ s.t. G(X) &\geq 0, X \in \Omega \subset R^n \end{aligned}$$

其中 $X = [x_1, x_2, \dots, x_n]^T$ 为属于空间区域 Ω 的操作变量(自变量), F 为目标函数, R 表示实数域, R^n 表示 n 维实数空间, $G(X) \geq 0, X \in Q \subset R^n$ 表示对变量的约束,即“环境”。在 GA 中将 X 进行编码,写成一定长度(比如 8 位、16 位等)的二进制编码,采用一定的函数 $fit(X)$,例如 $-F(X)$,描述个体 X 对环境的“适应度”(Fitness)。

1.3.1 GA 的基本步骤^[3]: a. 确定群体大小 N , 交叉率 P_c 、变异率 P_m 。于环境中随机地选取 N 个初始个体(初始化); b. 将个体解码(写成十进制数), 计算每个个体对环境的适应度 $fit(X_i)$, $i = 1, 2, \dots, N$; c. 按照个体的适应度选留适应值高的部分个体; d. 按照交叉率、变异率对所选个体进行交叉、变异、再生等操作, 获取新的个体(子代); e. 计算子代的适应度; f. 将子代加入亲代, 构成新的群体; g. 转入第 3 步, 反复操作, 直到寻优条件满足为止。

1.3.2 计算示例: 考虑优化问题

$$\begin{aligned} \max F &= x^2 \\ s.t. x &\in [0, 31] \end{aligned}$$

取 $N=4$, $P_c=0.6$, $P_m=0.03$, $fit(x_i)=F(x_i)/$

因取的变异率很小,故只需进行交叉操作一次,得到的子代如下:

$K, K=\sum_{j=1}^N F(x_j)$, 取初始化群体如下:

数字串 编号	数字串	x_i	f_i	$f_i/\sum f_i$	被复制 次数
1	01101	13	169	0.14	1
2	11000	24	576	0.49	2
3	01000	8	64	0.06	0
4	10011	19	361	0.31	1

数字串 编号	复制的 数字串	配对 号码	串接 位置	新的 数字串	x_i	f_i
1	01101	2	4	01100	12	144
2	11000	1	4	11001	25	625
3	11000	4	2	11011	27	729
4	10011	3	2	10000	16	256

求得最大值为 $F_{\max}=729$ 。

求得最大值为 $F_{\max}=576$ 。

在上述例子中,只经过一次进化,即将目标函数的值从 576 提高到 729,再经过几次进化即可求得上述优化的最大值。

1.4 GA 的特点

因为利用了生物学中的进化论思想与遗传学原理,GA 是一种与传统寻优算法极不相同的算法,因而有许多不同于传统寻优算法的特点,例如^[4]:GA 不直接和参数打交道,而是处理代表参数 X 的数字编码串 X_1, X_2, \dots, X_N ;GA 同时处理群体(Population)中的所有个体,并行地处理目标函数的局部峰值,而不只是处理个别的个体,这就大大减小了陷入局部最优解的可能性;GA 在寻优过程中操作的是参数的编码,不需用到目标函数的连续性、微分等条件,因而适应面广,能用来解决许多用传统算法不能解决的问题,适合于工程应用。GA 的操作是随机操作,不是确定性的运算,其收敛性由概率理论保证。

2 GA 的发展与应用

2.1 GA 的发展与应用

GA 的早期工作是由 John H. Holland 于 1971 年提出来的^[1,2],Holland^[5]建立了寻优与适应算法的框架,这正式标志着 GA 的诞生,在其后的一段时间内,GA 没有受到充分的关注,近年来 GA 逐步为人们所认识、发展。在 80 年代,GA 进入系统工程的许多领域。Grefenstette 等^[6]将 GA 用于模式识别;Ackley 等将 GA 用于神经网络(NN-Neural Networks)^[7,8]。GA 已广泛应用于机器学习、模式识别、图象处理、工业优化控制、自适应控制、社会科学等方面,并在解决 TSP 问题^[9]、煤气管道的最优控制^[10]、通讯网络连接长度的优化问题^[11]、铁路运输计算优化、化工过程优化^[12]等方面取得了很大的成功。GA 的研究引起了世界各领域学者的兴趣,从 1985 年到 1993 年,每两年举办一届国际性的 GA 学术讨论会,并在美国出版“Evolutionary Computation”国际刊物,在许多相关的学术会议上都有关于 GA 的讨论,在北京举行的 1996 IEEE 系统、人与控制论国际学术会议(IEEE SMC'96)上就有关于 GA 应用的专题讨论。1994 年 1 月,“IEEE Trans. on Neural Networks”组织了有关 GA 的专辑。专辑的编者按指出:“进化思想不仅是生命科学的范畴,进化是一种优化过程,可以在计算机上模拟,并可运用到工程领域”^[13]。GA 已与计算机科学、自动化技术中的许多新理论、新方法相结合,并应用于越来越多的实际问题。完全有理由相信随着计算机科学与计算技术的飞速发展,GA 将有更大的发展。

2.2 GA 在生物科学、生物工程中的应用前景

GA 植根于生物学中的进化论思想与遗传学原理,因而自然有应用于生物科学的潜力。GA 不但可以指导生物科学实验、生产,而且可以用于生物工程中存在的大量的优化、控制等问题。与模糊控制、递阶控制、神经网络等相结合,可用于生物工程中控制问题。例如,生物工程中的发酵过程中有一种间歇补料方式,需要求取最优的补料策略,以获得最佳的经济效益。这种补料控制问题的状态方程是一个离散状态方程^[14]:

$$\begin{cases} X(k+1) = F[X(k), U(k)], \\ y(k) = CX(k), \\ U_{\min} \leq U(k) \leq U_{\max} \end{cases}$$

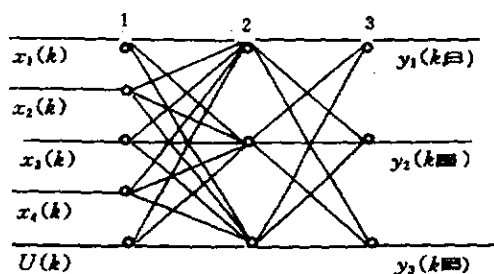


图 3 发酵过程神经网络预估器结构图

Fig. 3 Network predictor structure map in fermentation

1. In put layer, 2. Hid layer, 3. Out put layer

可以采用神经网络(NN)进行预估优化控制,神经网络预估器的结构见图 3。在 NN 中,控制器的训练采用 BP 等算法,而 BP 算法的效率还是不高,因而在实时控制中是有困难的,如采用 GA 训练权系数,将大大提高控制器训练的速度,从而大大改进控制的性能,提高生产效率。

3 结束语

GA 是模仿生物进化与遗传原理设计的工程优化方法,已在工程实际中得到成功的应用,GA 早已突破传统优化问题的范围,用于许多问题的研究与应用。将 GA 应用于生物科学,发展 GA 理论、推广 GA 的应用。GA 还将刺激仿生科学技术的发展。

参 考 文 献

- [1]Holland J H. In: Jacks F L (ed.), *Associative Information Processing*, American Elsevier, NY, 1971, pp. 127~146.
- [2]Holland J H. *Proc. of the NFS Workshop on Learning System Theory and its Applications*, 1973, pp. 43~46.
- [3]Davidor Y. *A Heuristic Strategy for Optimization*, World Scientific Series in Robotic and Automated System, Vol. 1, 1991.
- [4]Caponetto R *et al.* *A Survey Trans Inst. MC*, 1993, 15(3):143~156.
- [5]Holland J H. *University of Michigan Press*, Ann Arbor, 1975.
- [6]Grefenstette J J, Fitzpatrick J M. *Proc. of 1st Int. Conf. on GAs*, 1985, pp. 112~120.
- [7]Ackley D H. *Proc of 1st Int Conf on GAs*, 1985, 121~135.
- [8]Cohon J P *et al.* *Proc. of 2nd Int. Conf. on GAs*, 1987, pp. 148~154.
- [9]Lin Feng - Tse and Kao Cheng - Tan. *IEEE Trans. Systems, Man and Cybernetics*, 1993, 23(6):1752 - 1767.
- [10]Goldberg D E. *Dissertation Abstracts International*, 44(10), 3174B.
- [11]Combs S. Davis L. *Proceedings of the 2th International Conference on Genetic Algorithms*, 1987, pp. 257~260.
- [12]王 强,邵嘉鹤. *控制理论与应用*, 1996, 13(4):477~481.
- [13]Special Issue on Evaluation Computing. *IEEE Trans. Neural Networks*, 1994, 5(1).
- [14]杨一兵,王树青,张泉灵等. *基于神经网络的生化过程预估优化控制*, *生物工程学报*, 1996, 12(1):101~105.
- [15]张晓明,戴冠中,徐乃平. *控制理论与应用*, 1995, 12(3):265~273.

Genetic Algorithms: An Application of Evaluation Theory and Genetics to Optimization Problems in Engineering

Yu Huiming Yao Ruhua Lin Weitie

(Department of Bioengineering, South China University of Technology, Guangzhou 510641)

Abstract The natural rule tells that the superiors survive. Recently, the evaluation theory and principles of genetics of bioscience are employed to the computation of optimization problems in engineering successfully. Thus a computation method for optimization problems, i. e. Genetic Algorithm (GA) which is quite different from the classical ones, has been proposed. In this paper, the evaluation theory and principles of genetics employed to genetic algorithm, the basic operations of genes applied in genetic algorithm, the steps of genetic algorithm, the advantages over the classical computation methods, the development of genetic algorithm and its applications are introduced. A simple prospects for the possible applications of genetic algorithm to bioscience is also given.

Key words Evaluation, heredity, optimization, genetic algorithm