文章编号: 1001-0920(2004)05-0554-03

基于模式定理的遗传算法交叉和变异概率上限

巩敦卫, 孙晓燕
(中国矿业大学 信息与电气工程学院, 江苏 徐州 221008)

摘　要: 基于模式定理的推广形式, 给出含有选择、交叉操作遗传算法一致交叉概率的上限, 以及含有选择、交叉和变异操作遗传算法单点变异和一致变异概率的上限, 分析了含有联赛选择、一致交叉操作遗传算法运行前期和后期对优良模式的影响, 并用 8 位陷阱函数验证了上述结论的正确性. 该结果可用于指导遗传操作与控制参数的设计.

关键词: 遗传算法; 模式定理; 交叉; 变异
中图分类号: TP18　　**文献标识码**: A

Upper limits of crossover and mutation rate of genetic algorithm based on schema theorem

GONG Dun-wei, SUN Xiao-yan
(College of Information and Electronics Engineering, China University of Mining and Technology, Xuzhou 221008, China. Correspondent: GONG Dun-wei, E-mail: dwgong@vip.163.com)

Abstract: The upper limit of uniformity crossover rate of genetic algorithm composed of selection and crossover operators is presented based on the generalization form of schema theorem. The upper limits of single point and uniformity mutation rate of genetic algorithm composed of selection, crossover and mutation operators are given. The effect of genetic algorithm composed of tournament selection, uniformity crossover operator on better schema in prophase and post phase is analyzed. The results are validated through 8-bit trap function. The results can be used to guide the design of genetic operators and control parameters of genetic algorithm.

Key words: genetic algorithm; schema theorem; crossover; mutation

1 引　　言

遗传算法是一种模拟生物进化过程的全局优化搜索方法, 具有简单通用、鲁棒性强、适于并行处理、易于和其他方法相结合以及应用范围广等优点[1]. 目前遗传算法已广泛地应用于工程设计、人工智能和自动控制等领域[2]. 遗传算法在应用之前需要人为设定遗传操作和控制参数, 这些设定具有很大的主观性和经验性, 设定不当会影响遗传算法的性能, 这在一定程度上影响了遗传算法潜能的发挥[3].

最早由 Holland 提出的模式定理给出了模式在进化过程中的变化, 反映了重要建筑块的发现过程. 该定理只说明交叉和变异操作对模式生存的消极影响, 没有给出交叉概率和变异概率的设定方法, 因而虽具有理论意义, 但难于指导实际的遗传算法设计. 随后 Goldberg 等[4]给出了模式定理的推广形式, 基于此给出了单点交叉概率的上限, 并考虑了联赛选择、截断选择和比例选择等对优良模式的影响. 但该文没有考虑其他交叉操作交叉概率上限的确定, 也没有考虑变异操作的作用.

本文基于模式定理的推广形式, 给出含有选择、

收稿日期: 2003-04-29; **修回日期**: 2003-06-20.
基金项目: 国家自然科学基金资助项目(60304016).
作者简介: 巩敦卫(1970—), 男, 江苏铜山人, 副教授, 博士, 从事进化计算、智能控制和变结构控制等研究; 孙晓燕(1978—), 女, 江苏丰县人, 硕士生, 从事协同进化计算、交互式进化计算等研究.

交叉操作遗传算法一致交叉概率的上限及含有选择、交叉和变异操作遗传算法单点变异和一致变异概率的上限, 并用 8 位陷阱函数验证了结论的正确性

2　模式定理的推广形式

Holland 提出的模式定理是遗传算法进化动力学的基本定理. 该定理指出, 在选择、交叉和变异等遗传操作的作用下, 那些低阶、定义长度短、超过种群平均值的模式的数量将随着迭代次数的增加以指数规律增长[5]. 对于比例选择、单点交叉和无变异遗传算法, 模式定理可表示为[4,6]

$$\langle m(H,t+1)\rangle \geq m(H,t)\frac{f(H,t)}{\bar{f}(t)}\left[1-p_c\frac{\delta(H)}{l-1}\right].$$

其中: $\langle m(H,t+1)\rangle$ 表示模式 H 在第 $t+1$ 代所含个体的期望数量, $m(H,t)$ 表示模式 H 在第 t 代所含个体的数量, $f(H,t)$ 为模式 H 在 t 代所含个体的适应度平均值, $\bar{f}(t)$ 为第 t 代进化种群的适应度平均值, p_c 为交叉概率, $\delta(H)$ 为模式 H 的定义长度, l 为进化个体的串长

在上式中, $f(H,t)/\bar{f}(t)$ 由选择操作决定, $\left[1-p_c\frac{\delta(H)}{l-1}\right]$ 由交叉操作决定 上式只考虑比例选择和单点交叉的情况, 对于其他的选择和交叉操作, Goldberg 等[6] 给出如下模式定理推广形式:

$$\langle m(H,t+1)\rangle \geq m(H,t)\varphi P_{sc}.$$

其中: φ 为由选择操作决定的因子; P_{sc} 为交叉生存概率, 由交叉操作决定 对于不同的选择和交叉操作, φ 和 P_{sc} 的取值不同 Goldberg 等[4] 给出了不同选择操作 φ 的表达式

3　基于模式定理的交叉概率上限确定

考虑含有选择、交叉操作遗传算法的两类交叉操作: 单点交叉和一致交叉

3.1　单点交叉[4]

对于单点交叉, 模式 H 的生存概率为 $P_{sc}=1-p_c\frac{\delta(H)}{l-1}$, 此时模式定理可表示为

$$\langle m(H,t+1)\rangle \geq m(H,t)\varphi\left[1-p_c\frac{\delta(H)}{l-1}\right].$$

要使某模式 H 所含个体数随着进化代数的增加而增长, 则要求 $\varphi\left[1-p_c\frac{\delta(H)}{l-1}\right]\geq 1$, 即要求

$$p_c \leq \frac{l-1}{\delta(H)}(1-\varphi^{-1}). \tag{1}$$

由式(1) 知, 可以通过选择适当的交叉概率, 使得某模式所含个体数目经过选择、单点交叉等操作后得到增长

3.2　一致交叉

模式 H 的阶记为 $O(H)$, H 在交叉操作中被破坏的交叉串的取值概率不大于 $\frac{2^{O(H)}-1}{2^l}$, 从而模式 H 的生存概率为 $P_{sc}=1-p_c\frac{2^{O(H)}-1}{2^l}$, 相应地, 模式定理可表示为

$$\langle m(H,t+1)\rangle \geq m(H,t)\varphi\left[1-p_c\frac{2^{O(H)}-1}{2^l}\right].$$

同样, 经过推导可知, 要使某模式 H 所含个体数随着进化代数的增加而增长, 则要求

$$p_c \leq \frac{2^l}{2^{O(H)}-1}(1-\varphi^{-1}). \tag{2}$$

现考虑具体的选择操作对一致交叉概率上限的影响 对于规模为 s 的联赛选择,

$$\varphi=[1-(1-P_t)^s]/P_t,$$

其中 P_t 为优良模式 H 在第 t 代种群中的比例 在遗传算法运行的早期, P_t 很小, $\varphi\approx s$, 为使该模式所包含个体数随着进化代数的增加而增长, 则要求

$$p_c \leq \frac{2^l}{2^{O(H)}-1}(1-s^{-1}).$$

在遗传算法运行的后期, P_t 很大, $\varphi\to 1$, $p_c\to 0$, 要使该模式所包含个体数随着进化代数的增加而增长, 则要求一致交叉概率的上限趋于 0. 也就是说, 由于进化后期种群的个体基本接近, 选择和交叉操作将使优良模式的增长速度变得很慢(这里没有考虑交叉操作对模式增长的积极作用). 对于其他选择方法, 经类似的推导也可得到相应结论

4　基于模式定理的变异概率上限确定

考虑含有选择、交叉和变异操作的遗传算法 为便于讨论, 不妨固定选择和交叉操作而考虑两类变异操作: 单点变异和一致变异

4.1　单点变异

对于单点变异, 变异后模式 H 被破坏的概率为 $O(H)/l$, 若变异概率为 p_m, 则模式 H 的变异生存概率为 $P_{sm}=1-p_m\frac{O(H)}{l}$, 此时模式定理可表示为

$$\langle m(H,t+1)\rangle \geq m(H,t)\varphi P_{sc}\left[1-p_m\frac{O(H)}{l}\right].$$

要使某模式 H 所含个体数随着进化代数的增加而增长, 则要求 $\varphi P_{sc}\left[1-p_m\frac{O(H)}{l}\right]\geq 1$, 即要求

$$p_m \le \frac{l}{O(H)}[1 - (\varphi P_{sc})^{-1}]. \quad (3)$$

由式(3) 知, 可以通过选择适当的变异概率, 使得某模式所含个体数目经过选择、交叉和单点变异等操作后得到增长.

4.2 一致变异

模式 H 在变异操作中被破坏的变异串的取值概率最大为$\frac{2^{O(H)}-1}{2^l}$, 从而模式 H 的变异生存概率为

$$P_{sm} = 1 - p_m \frac{2^{O(H)}-1}{2^l},$$

相应地, 模式定理可表示为

$$\langle m(H, t+1)\rangle \ge m(H, t)\varphi P_{sc}\left[1 - p_m \frac{2^{O(H)}-1}{2^l}\right].$$

经过推导可知, 要使某模式所包含个体数随着进化代数的增加而增长, 则要求

$$p_m \le \frac{2^l}{2^{O(H)}-1}[1 - (\varphi P_{sc})^{-1}]. \quad (4)$$

5 算 例

通过寻找陷阱函数最大值, 验证本文所给一致交叉概率上限及单点变异、一致变异概率上限的正确性. 陷阱函数为[4]

$$f(x) = \begin{cases} 8 - 8x, 0 \le x \le 1; \\ x - 1, 1 < x \le 8. \end{cases} \quad (5)$$

其图形如图 1 所示.

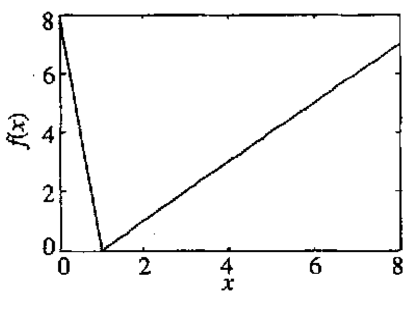

图 1 陷阱函数图形

变量采用 8 位二进制编码, 其最优模式为 00000000, 次优模式为 11111111. 考虑某一较优模式 H = 000000 * * , 则 $O(H)$ = 6, $\delta(H)$ = 5. 算法中, 种群规模为 20, 终止代数为 20, 独立运行 20 次考察该模式所含个体数 20 次平均变化的情况.

首先针对规模 s = 2 的联赛选择, 不进行变异操作, 验证一致交叉概率上限的正确性. 根据式(2) 及 φ 表达式计算每代一致交叉概率理论值, 记为 PC, 实际采用的交叉概率记为 p_c. 模式 H 所含个体数随进化代数的变化如图 2 所示, 其中: 实线为 $p_c \le$ PC 的情况, 虚线为 $p_c >$ PC 的情况. 图示结果进一步验证了本文所给一致交叉概率上限的正确性.

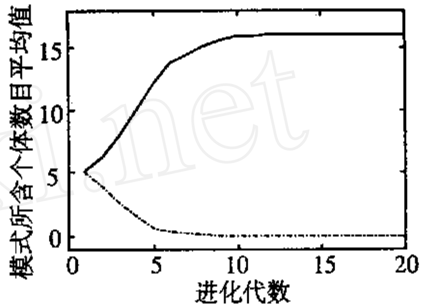

图 2 一致交叉 H 所含个体数目平均值

其次, 固定联赛选择规模 s = 2, 交叉概率 p_c = 0.75, 则一致交叉生存概率 P_{sc} = 0.82, 根据式(3) 和式(4) 计算每代单点变异及一致变异概率理论值, 记为 PM, 实际采用的变异概率记为 p_m. 对于一致交叉、单点变异, 模式 H 所含个体数的变化如图 3 所示; 对于一致交叉、一致变异, 模式 H 所含个体数的变化如图 4 所示. 其中: 实线为 $p_m \le$ PM 的情况, 虚线为 $p_m >$ PM 的情况. 图示结果同样验证了本文所给变异概率上限的正确性.

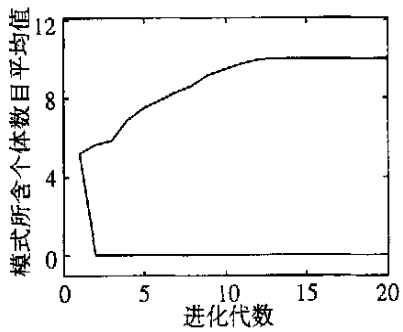

图 3 一致交叉、单点变异, 模式 H 所含个体数目平均值

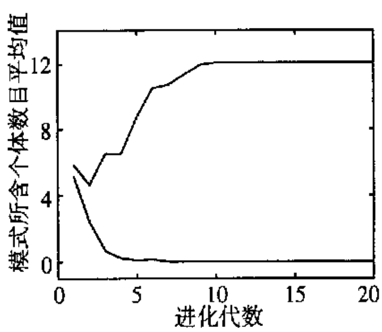

图 4 一致交叉、一致变异, 模式 H 所含个体数目平均值

(下转第 581 页)

个输入上以产生一个新的结果, 如此反复直至达到所期望的输出; 2) 神经网络结构本身决定了它是大规模并行机制, 由于是数据驱动的, 其处理速度比自适应模糊控制器要快得多. 然而, 前一方法存在一个突出的缺点, 即对于MIMO系统其关系矩阵过大, 计算机较难存储; 而后者则弥补了前者的不足.

参考文献(References):

[1] Wang L X. Stable adaptive fuzzy control of nonlinear systems[J]. *IEEE Trans on Fuzzy Syetem*, 1993, 1(2): 146-155.

[2] Sue C Y, Stepanenko Y. Adaptive control of a class of nonlinear systems with fuzzy logic[J]. *IEEE Trans on Fuzzy Systems*, 1994, 2(4): 285-294.

[3] Tong S C, Tang J T, Wang T. Fuzzy adaptive output tracking control of nonlinear systems[J]. *Fuzzy Sets and Systems*, 2000, 111(2): 169-182.

[4] 刘晓霞, 田兆福, 孙金根, 等. 时滞系统的自适应模糊控制器的研究[J]. 信息与控制, 2003, 32(2): 285-288. (Liu X X, Tian Z F, Sun J G, et al. Research on self-adaptive fuzzy controller for tame-delay system [J]. *Information and Control*, 2003, 32(3): 285-288.)

[5] Cen J, Xu D, Sha F B. On sufficient condition for stability independent of delay[J]. *IEEE Trans on Automatic Control*, 1998, 40(9): 1675-1680.

[6] Sun Y J, Hsieh J G, Yang H C. On the stability of uncertain systems with multiple time varying delays [J]. *IEEE Trans on Automatic Control*, 1997, 42(1): 101-105.

(上接第 556 页)

6 结　语

本文基于模式定理的推广形式, 给出了含有选择、交叉操作遗传算法一致交叉概率的上限; 同时给出了含有选择、交叉和变异操作遗传算法单点变异和一致变异概率的上限. 所得结果可用于指导遗传操作与控制参数的设计, 以保证期望模式随进化而增长.

需要说明的是, 本文在交叉和变异概率上限的推导中, 只考虑了交叉和变异操作对模式的破坏作用, 而没有考虑这些操作对模式的积极作用. 如果考虑了这种积极作用, 即使不满足本文所给出的上限要求, 种群中优良模式所含的个体数量也有可能增加. 关于交叉操作与变异操作对模式增长的积极影响以及与交叉概率和变异概率之间的定量关系等问题是进一步要研究的课题.

参考文献(References):

[1] 张文修, 梁怡. 遗传算法的数学基础[M]. 西安: 西安交通大学出版社, 2000. 1-4.

[2] 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. 13-14.

[3] Lobo F G. The parameter-less genetic algorithm: Rational and automated parameter selection for simplified genetic algorithm operation[D]. Lisben: Pela University of Lisben, 2000. 27-30.

[4] Goldberg D E, Sastry K. A practical schema theorem for genetic algorithms design and tuning[R]. Urbana: University of Illinois at Urbana-Champaign, 2001.

[5] Holland J H. *Adaption in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*[M]. Ann Arbor, MI: The University of Michigan Press, 1975.

[6] Goldberg D E, Deb K. A comparative analysis of selection schemes used in genetic algorithms [A]. *Foundations of Genetic Algorithms* [C]. San Mateo, CA: Morgan Kaufmann, 1991. 69-93.