

基于三进制表示的新模式定理

明 亮, 王宇平

(西安电子科技大学 理学院, 陕西 西安 710071)

摘要:一般的模式定理通常使用二进制表示且只讨论了模式存活概率而没有考虑模式新建的概率, 因此它们不能准确地描述模式的变化情况. 本文提出了一种新的模式表示法: 三进制表示法. 利用这种新的表示法, 很容易区分模式的存活和新建, 并分别估计了在均匀杂交算子作用下模式的存活概率和新建概率. 最后对均匀杂交算子给出了在模式的存活和模式的新建共同作用下的模式定理, 这一结果深入和推广了目前模式定理的结果.

关键词: 模式; 模式定理; 杂交; 均匀杂交

中图分类号: TP18 **文献标识码:** A

Schema theorem based on ternary representation

MING Liang, WANG Yu-ping

(Faculty of Science, Xidian University, Xi'an Shaanxi 710071, China)

Abstract: The existing schema theorems usually use the binary representation of schema, and discuss the probability of schema survival without considering the probability of schema construction. Thus they cannot describe the change of schema exactly. A new representation method of schema called ternary representation is proposed. By using this new representation, one can easily distinguish the schema survival and schema construction, and estimate the probability of schema survival and the probability of schema construction for uniform crossover operator, respectively. At last, the schema theorem for uniform crossover operator is proposed considering both schema survival and schema construction. It deepens and generalizes the existing results on schema theorems.

Key words: schema; schema theorem; uniform crossover operator; ternary representation

1 引言(Introduction)

目前, 遗传算法已得到日益广泛的应用, 出现了许多有效的算法^[1~4]. 然而有关遗传算法的理论还很不完善, 其中核心理论是模式定理和收敛性理论. 关于模式定理, 一般是使用二进制表示且只是讨论了模式存活的条件概率, 这实际上只是考虑了杂交算子的破坏作用, 而没有考虑杂交算子的更为积极的新建作用. 若假设参加杂交的两个父代有顺序时, 最近文献[5](第4章)在不同于通常模式存活和新建的意义下讨论了在模式的存活和新建共同作用下的模式定理, 它无法单独分析通常意义下的模式存活所起的作用和模式新建所起的作用. 另外, 在一般的进化算法中, 对参加杂交的两个个体, 并不考虑其顺序. 为了深入研究模式的变化情况, 必须搞清楚模式的存活和新建分别对模式的变化所起的作用. 为此, 本文在不考虑参加杂交的两个个体的顺序的情

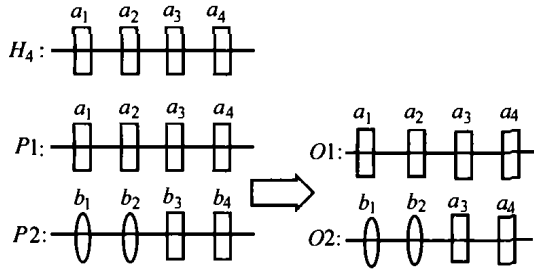
况下, 首先给出了模式的存活和模式的新建等的准确概念, 提出一种新的模式的表示方法——三进制表示法, 该方法能很方便的分别分析在均匀杂交算子的作用下模式的存活概率和模式的新建概率, 从而定量地刻画了模式的存活和模式的新建对模式变化的影响. 最后, 综合考虑杂交算子在模式的存活和新建共同作用下的模式定理, 即考虑无论父代中是否含有属于某个模式的个体, 通过杂交后子代中含有属于该模式的个体的概率. 此研究深入和推广了目前已有模式定理的结果, 对研究能产生较好后代的杂交算子有重要的指导意义.

2 基本概念(Basic concepts)

定义 1 模式的存活.

父代中含有属于模式 H 的个体, 经过遗传算子作用后, 子代中仍含有属于模式 H 的个体, 这种情况称为模式 H 的存活.

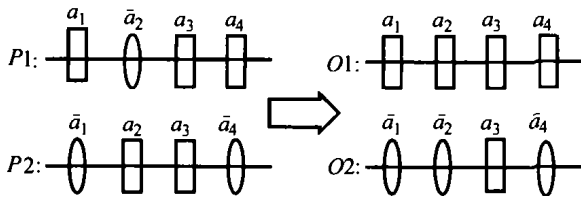
例 1 对于 4 阶模式 H_4 , 在确定位 d_1, d_2, d_3, d_4 上的字符分别为 a_1, a_2, a_3, a_4 , 父代 P_1 在确定位 d_1, d_2, d_3, d_4 上的字符分别为 a_1, a_2, a_3, a_4 , 父代 P_2 在确定位 d_1, d_2, d_3, d_4 上的字符分别为 b_1, b_2, b_3, b_4 , 其中 $b_3 = a_3, b_4 = a_4$. 两父代通过某杂交事件作用后, 子代 O_1 的字符 b_3 和 b_4 均来自父代 P_2 , 但 O_1 仍属于模式 H , 这就是模式 H_4 存活的意义. 如图 1 所示 (为了讨论的简便, 在讨论模式 H_k 时, 只用和 H_k 确定位对应位上的字符来表示该模式).

图 1 模式 H_4 的存活Fig. 1 Survival of schema H_4

定义 2 模式的新建.

父代中没有属于模式 H 的个体, 经过遗传算子作用后, 子代中却含有属于模式 H 的个体, 这种情况称为模式 H 的新建.

例 2 设 H_4 为例 1 中的 4 阶模式, 父代为 P_1 和 P_2 时, 经过某杂交算子作用后, 产生两子代 O_1 和 O_2 , 其中 $O_1 \in H_4$. 显然 O_1 中的字符 a_1 和 a_4 只能来自父代 P_1 , 字符 a_2 只能来自父代 P_2 ; 而字符 a_3 可能来自父代 P_1 , 也可能来自父代 P_2 , 这种情况称为模式的新建. 如图 2 所示.

 a_i : 模式的确定字符; $\bar{a}_i$: 不同于 a_i 的字符; $i = 1 \sim 4$ 图 2 模式 H_4 的新建Fig. 2 Construction of H_4

3 均匀杂交算子的模式定理 (Schema theorem for uniform crossover)

3.1 产生属于模式 H_k 的个体的可行情况 (Situations that will create individual being in H_k)

子代中含有属于模式 H_k 的个体, 能产生这样的个体的父代情况很多, 称每一对这样的父代情况为一个可行情况, 每种可行情况可用一个 k 位 0, 1, 2

三进制串唯一表示, 此三进制串可确定如下: P_1 与 P_2 是两父代个体, 它们产生属于模式 H_k 的个体 O_i . 若 O_i 的与 H_k 的第 r 个确定位 (从左边数) 对应位上的字符只能是从原来的父代 P_i 的对应位上保留下来, 则此三进制串中的第 r 位就取“1”; 若该位上的字符只能来自另外一个个体串 $P_j (j = 3 - i)$ 的对应位, 则此三进制串中的第 r 位就取“2”; 若该位上的字符可能是从个体串 P_i 的对应位中保留下来, 也可能是来自另外一个个体串 $P_j (j = 3 - i)$ 的对应位, 则此三进制串中的第 r 位就取“0”. 若该三进制数代表的非负整数为 j , 则将该可行情况称为第 j 种可行情况, 记做 s_j . 这样每一种可行情况都与一个 k 位的三进制串一一对应. 例如, 图 2 中的可行情况可用三进制串“1201”表示, 又因为该三进制串表示“46”, 因此图 2 表示可行情况 s_{46} .

若某个可行情况对应的两个父代中含有属于模式 H_k 的个体, 称该可行情况为模式 H_k 的存活情况; 若可行情况对应的两个父代中没有属于模式 H_k 的个体, 称该可行情况为模式 H_k 的新建情况.

一般地, 对于模式 H_k , 共有 3^k 种可行情况, 可行情况

$$s_j = x_{k-1}x_{k-2} \cdots x_1x_0,$$

$$j = x_{k-1}3^{k-1} + x_{k-2}3^{k-2} + \cdots + x_13^1 + x_03^0 (j = 0 \sim 3^k - 1),$$

其中: $x_i (i = 0 \sim k - 1)$ 可取 0, 1 或 2; 在 s_j 中, “0”, “1”, “2” 的个数分别用 m_0, m_1, m_2 表示.

根据 x_i 的取值, 所有可行情况可分为如下两个集合 L_1 和 L_2 , L_1 表示所有存活情况的集合, 而 L_2 表示所有新建情况的集合.

$$\begin{cases} L_1 = \{s_j \mid \text{所有 } x_i \in \{0, 1\} \text{ 或所有 } x_i \in \{0, 2\}\}, \\ L_2 = \{s_j \mid \text{必存在 } x_i = 1 \text{ 和 } x_{i'} = 2\}, \\ i = 0 \sim k - 1, i' = 0 \sim k - 1, j = 0 \sim 3^k - 1. \end{cases} \quad (1)$$

这样用三进制表示法便将产生属于模式 H_k 的个体的所有可行情况分为存活情况和新建情况. 为了讨论在概率为 p_0 的均匀杂交算子的破坏和新建的共同作用下的模式定理, 先分别计算模式 H_k 的存活概率 $p_s(H_k, p_0)$ 和模式 H_k 的新建概率 $p_c(H_k, p_0)$, 由于篇幅所限, 下面仅列出计算结果:

$$p_s(H_k, p_0) =$$

$$\left[p_{eq} + p_0 \cdot \frac{1 - p_{eq}}{2} \right]^k + \left[p_{eq} + (1 - p_0) \cdot \frac{1 - p_{eq}}{2} \right]^k - p_{eq}^k, \quad (2)$$

$$p_c(H_k, p_0) = \left(\frac{1+p_{eq}}{2}\right)^k - \left[(1-p_0) \cdot \frac{1-p_{eq}}{2} + p_{eq}\right]^k - \left[p_0 \cdot \frac{1-p_{eq}}{2} + p_{eq}\right]^k + p_{eq}^k. \quad (3)$$

其中 p_{eq} 表示两父代在每一确定位上字符相同的概率, 它可算是两个父代的相似度.

由式(2)和式(3), 可以得到, 经过概率为 p_0 均匀杂交算子作用后, 子代中含有属于模式 H_k 的个体的概率 $p_{s,c}(H_k, p_0)$ 为

$$p_{s,c}(H_k, p_0) = \frac{(1+p_{eq})^k}{2^k}. \quad (4)$$

由上式可以看到, 经过概率为 p_0 均匀杂交算子作用后, 子代中含有属于模式 H_k 的个体的概率与何种均匀杂交算子没有关系, 而只与模式的阶 k 和描述父代相似度的 p_{eq} 有关.

定理 1 (均匀杂交算子的模式定理) 任意两个串长相同的父代经过任意均匀杂交算子作用后, 子代含有属于模式 H_k 的个体的个数的期望值 $E_{s,c}[H_k, p_0]$ 只和模式的阶 k 和描述父代相似度的 p_{eq} 有关.

证 设 B 表示子代中含有属于模式 H_k 的个体的个数, 则 $B \in \{0, 1, 2\}$, 所以有

$$E_{s,c}[H_k, p_0] = \sum_{b \in \{0, 1, 2\}} b \times p(B=b) = p(B=1) + 2p(B=2) = [p(B=1) + p(B=2)] + p(B=2).$$

上式等号右边第一个中括号内式子表示子代中有一个和两个属于模式 H_k 的个体的概率, 也就是不论父代中是否含有属于模式 H_k 的个体, 而子代中会有这样的个体的概率, 它就等于模式的存活概率加上模式的新建概率. 上式等号右边第二项 $p(B=2)$ 表示当两个子代都属于模式 H_k 的概率, $p(B=2)$ 所对应的是存活情况 s_0 发生的概率, 存活情况 s_0 是指两个父代都属于模式 H_k 的可行情况, 也就是两个父代在 k 个确定位完全相同, 那么 $p(B=2) = p_{eq}^k$. 所以有

$$E_{s,c}[H_k, p_0] = p_s(H_k, p_0) + p_c(H_k, p_0) + p_{eq}^k,$$

故由式(4)可得

$$E_{s,c}[H_k, p_0] = \frac{(1+p_{eq})^k}{2^k} + p_{eq}^k.$$

由上式可以看到任何串长相等的两父代经过任意均匀杂交算子作用以后, 子代中存在属于模式的元素的概率及其个数的期望值只关于模式的阶 k 和描述父代相似度的 p_{eq} 有关, 并有

结论 1 模式的阶 k 越大, $p_{s,c}(H_k, p_0)$ 和 $E_{s,c}[H_k, p_0]$ 越小;

结论 2 杂交的父代越相似 (p_{eq} 越大), $p_{s,c}[H_k, p_0]$ 和 $E_{s,c}[H_k, p_0]$ 越大.

4 结束语(Conclusion)

本文从新概念出发计算了一个模式存活概率和新建概率, 详细讨论了均匀杂交算子的综合作用, 重新探讨了均匀杂交算子的模式定理. 本文提出的三进制表示法能计算模式的新建和存活概率, 对设计能产生较好的后代的杂交算子有重要的指导意义.

参考文献(References):

- [1] CNSTILLO P A, ROMERO G. Statistical analysis of the main parameters improved in the design of a genetic algorithm [J]. *IEEE Trans on Systems, Man, and Cybernetics, Part C*, 2002, 32(1): 31-37.
- [2] LEUNG Y M, WANG Yuping. An orthogonal genetic algorithm with quantization for global numerical optimization [J]. *IEEE Trans on Evolutionary Computation*, 2001, 5(1): 41-53.
- [3] LEUNG Y M, WANG Yuping. Multiobjective programming using uniform design and genetic algorithm [J]. *IEEE Trans on Systems, Man, and Cybernetics, Part C*, 2000, 30(3): 293-304.
- [4] KUSHCHU I. Genetic programming and evolutionary generalization [J]. *IEEE Trans on Evolutionary Computation*, 2002, 6(5): 431-442.
- [5] SPEARS W M. *The role of mutation and recombination in evolutionary algorithms* [D]. Virginia: George Mason University, 1998.

作者简介:

明亮 (1977—), 男, 西安电子科技大学博士研究生, 研究方向为进化计算、最优化计算, E-mail: liang-ming1977@yahoo.com.cn;

王宇平 (1961—), 男, 西安电子科技大学应用数学系教授, 博士生导师, 1993 年于西安交通大学获博士学位, 研究领域为进化计算、最优化理论、算法和应用等.