

遗传程序设计的精确模式理论进展^{*})

Advances in the Genetic Programming Exact Schema Theorems

李晓欧^{1,2} 李 淼¹ 梁曼君²

(中国科学院合肥智能机械研究所 合肥230031)¹ (合肥工业大学计算机与信息学院 合肥230009)²

Abstract Schema theorems are descriptions of how the number (or the fraction) of members of the population belonging to a schema varies over time in evolutionary computation. This paper gives full summarization of the advances in Genetic Programming exact schema theorems in recent years, mainly discusses the exact formulation for the expected number of instances of a schema at the next generation in terms of microscopic and macroscopic quantities, then, the effects and the mechanisms of schema creation are clarified.

Keywords Genetic programming, Hyperschema, One-point crossover

1. 引言

从传统意义上来讲,模式定理是用来解释遗传算法是怎样进化的。模式定理可被看成是遗传算法的宏观模型,这就意味着它可以根据当前代测得的宏观量(模式适应度、种群适应度、模式中个体数量等)来确定下一代种群的属性。这些与微观模型形成鲜明对比的宏观量隐含了大量遗传算法自由度信息,由它们可以推导出易于理解和研究的等式。

对传统 GP 模式定理的争论焦点是它们仅提供了在下一代模式 H 实例数量期望值的下界 $E[m(H, t+1)]$, 而不是一个精确值,这使得很难用 GP 模式定理去预测 GP 的未来行为,本文主要介绍的最新 GP 一点交叉模式定理解决了这个问题,并获得 $E[m(H, t+1)]$ 的精确公式。这个模式定理拓展了 GP 理论,阐明了模式创建的内在机制,突破了传统 GP 模式定理只集中在模式的生存与遭破坏的表述上的局限性,并从微观和宏观角度给出了具体的定义。同时,它也很好支持了积木块假设。

2. 定长 GP 模式定理

Riccardo Poli[1998]提出遗传算子(选择、交叉、变异)的作用过程可被看成是伯努利试验,也就是说,新产生的个体或者属于模式 H , 或者不属于模式 H 。因此, $t+1$ 代与模式 H 所能匹配的实例数 $m(H, t+1)$ 就是一个二元随机变量,如果定义 $\alpha(H, t)$ 为每次试验的成功匹配概率(也称之为模式 H 的总体传递概率),则模式定理的精确表达为:

$$E[m(H, t+1)] = M \alpha(H, t) \quad (1)$$

其中, $E[\]$ 表示期望值, M 表示群体规模。

如果 GP 种群的个体都具有固定的规模和形状,在一点交叉,无变异的条件下, $\alpha(H, t)$ 表示如下:

$$\alpha(H, t) = (1 - p_{\infty}) p(H, t) + \frac{p_{\infty}}{N(H)} \sum_{i=0}^{N(H)-1} p(l(H, i), t) p(u(H, i), t) \quad (2)$$

其中, p_{∞} 表示交叉概率, $p(H, t)$ 表示在 t 代模式 H 的选择概率, $N(H)$ 表示模式 H 中结点的总数, $l(H, i)$ 表示将交叉点 i 以上的所有结点用通配符“*” (表示一单个函数或端点) 替代

后所得到的模式, $u(H, i)$ 表示将交叉点 i 以下的所有结点用通配符“*” 替代后所得到的模式, 交叉点 i 随 $N(H)$ 的有效交叉点集变化。图1给出了 $l(H, 1), u(H, 1), l(H, 3), u(H, 3)$ 的具体实例, 已知模式 $H = (* = (+x =))$, 交叉点以数字的形式标注在结点的连接线上。下面给出上述公式的证明:

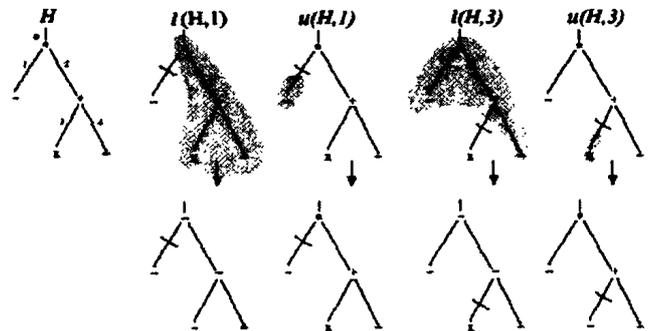


图1 模式实例及隐含的定长积木块

$$\begin{aligned} \frac{E[m(H, t+1)]}{M} &= \alpha(H, t) = (1 - p_{\infty}) p(H, t) + \frac{p_{\infty}}{N(H)} \\ &\sum_{i=0}^{N(H)-1} p(l(H, i), t) p(u(H, i), t) \\ &= (1 - p_{\infty}) p(H, t) + \frac{p_{\infty}}{N(H)} \left(\sum_{i \in B(H)} p(l(H, i), t) p(u(H, i), t) \right. \\ &\left. + (N(H) - \mathcal{L}(H)) p(H, t) p(* * \dots *, t) \right) = p(H, t) \\ &\left(1 - p_{\infty} \frac{\mathcal{L}(H)}{N(H)} \right) + \frac{p_{\infty}}{N(H)} \sum_{i \in B(H)} p(l(H, i), t) p(u(H, i), t) \geq p \\ &(H, t) \left(1 - p_{\infty} \frac{\mathcal{L}(H)}{N(H)} \right) + \frac{p_{\infty}}{N(H)} \sum_{i \in B(H)} (p(H, t))^2 = p(H, t) \left(1 - p_{\infty} \frac{\mathcal{L}(H)}{N(H)} \right) \\ &\left. + p_{\infty} \frac{\mathcal{L}(H)}{N(H)} (p(H, t))^2 \right) \end{aligned}$$

证毕。其中, $B(H)$ 表示模式 H 的交叉点集, $\mathcal{L}(H)$ 表示模式 H 定义长度, $p(* * \dots *, t) = 1, p(l(H, i), t) \geq p(H, t), p(u(H, i), t) \geq p(H, t)$ 。可以看出, 这个精确模式定理是 John Holland 的 GA (Genetic Algorithm) 模式定理, 式(3), 在变异概率为0时的准确表达。

$$E[m(H, t+1)] \geq M p(H, t) (1 - p_{\infty})^{O(H)}$$

^{*}) 国家“863”计划信息技术项目基金资助 (编号863-306-ZD05-01-4)。李晓欧 硕士研究生, 主要研究方向为演化算法、人工智能及其应用。李淼 研究员, 硕士生导师, 主要研究方向为人工智能、专家系统。梁曼君 教授, 硕士生导师, 研究方向为模式识别、人工智能。

$$[1 - p_{so} \frac{\mathcal{L}(H)}{N(H)} (1 - p(H, t))] \quad (3)$$

p_{so} 表示变异概率, $O(H)$ 表示模式中确定位置的个数。

3. GP 超模式定理

为了将式(2)的精确理论适用于任何规模和形状的程序, 我们引入更广义的模式定义。

3.1 超模式

定义 GP 超模式是一个有根结点的树, 它由集合 $F \cup \{=\}$ 中的内结点和集合 $T \cup \{=#\}$ 的叶结点组成, F 和 T 是 GP 运行中用到的函数集和端点集, 符号“=”为一个通配符, 代表一确定结点, 而“#”代表任何有效子树。例如, 对于超模式 $(* \# (+x=))$ 代表如下特征的所有程序: ①根结点为乘积; ②根结点的第一个变量为任何有效子树; ③根结点的第二个变量为相加; ④“+”的第一个变量为 x ; ⑤“+”的第二个变量为端点集中任何有效结点。实际上, 超模式理论更好地概括了以往 Rosca 的变长模式定理(无“=”通配符)及定长模式定理(无“#”通配符), 本质上等价于程序树的集合。

3.2 微观精确 GP 模式定理

由于有了超模式理论, 就可以获得如下对于任意规模和形状的程序种群总体适用的结果。

定理1(微观精确模式定理) 在一点交叉、无变异的条件下, 定长 GP 模式 H 的总体传递概率为:

$$a(H, t) = (1 - p_{so}) p(H, t) + p_{so} \sum_{h_1} \sum_{h_2} \frac{p(h_1, t) p(h_2, t)}{NC(h_1, h_2)} \sum_{i \in C(h_1, h_2)} \delta(h_1 \in L(H, i)) \delta(h_2 \in U(H, i)) \quad (4)$$

其中, $NC(h_1, h_2)$ 表示程序 h_1 和 h_2 之间代表共同区域的树片段结点数量, $C(h_1, h_2)$ 表示共同区域中交叉点的序号集, $\delta(x)$ 表示在 x 为“真”时等于“1”, 否则等于“0”, $L(H, i)$ 表示将交叉点 i 与根结点之间的所有结点用通配符“=”替代后所得到的超模式, 所有与这些结点相连的子树用“#”替代, $U(H, i)$ 表示将交叉点 i 以下的子树用通配符“#”替代后所得到的超模式。

超模式 $L(H, i)$ 和 $U(H, i)$ 是公式(2)中模式 $l(H, i)$ 和 $u(H, i)$ 的概括性定义, 如果交叉操作发生在 $L(H, i)$ 和 $U(H, i)$ 中对应于位置 i 的任何个体间, 所得后代必为模式 H 的实例。

在定理证明之前, 首先介绍一下 $L(H, i)$ 和 $U(H, i)$ 的实例是怎样被创建的。如图2所示的模式 $H = (* = (+x=))$, 模式 $L(H, 1)$ 是通过将根结点用通配符“=”替代, 并将根结点右侧子树用通配符“#”替代后获得的, 形为 $(= = \#)$ 。模式 $U(H, 1)$ 是通过将交叉点以下的子树用通配符“#”替代后获得的, 形为 $(* \# (+x=))$ 。同理, $L(H, 3)$ 和 $U(H, 3)$ 的结果为 $(= \# (=x\#))$ 和 $(* = (+\#=))$ 。

一旦 $L(H, i)$ 和 $U(H, i)$ 的概念建立起来, 就可以很方便地对定理进行证明了:

$$\text{令 } g(h_1, h_2, i, H) = \delta(h_1 \in L(H, i)) \delta(h_2 \in U(H, i))$$

已知双亲程序 h_1 和 h_2 , 以及模式 H 。如果 h_1 和 h_2 在交叉点 i 的操作后获得的子代属于模式 H , 此函数返回“1”, 否则等于“0”。 $p(h_1, h_2, i, t)$ 表示在时刻 t 双亲程序与交叉点的概率分布。 $p(h_1, h_2, i, t)$ 分布下种群的 $g(h_1, h_2, i, H)$ 期望值可简化为:

$$E[g(h_1, h_2, i, H)] =$$

$$\sum_{h_1} \sum_{h_2} \sum_i g(h_1, h_2, i, H) p(h_1, h_2, i, t)$$

由于 $p(h_1, h_2, i, t) = 0$ 对于 $i \notin C(h_1, h_2)$, 则

$$E[g(h_1, h_2, i, H)] = \sum_{h_1} \sum_{h_2} \sum_{i \in C(h_1, h_2)} g(h_1, h_2, i, H) p(h_1, h_2, i, t) \quad (5)$$

$$p(h_1, h_2, i, t) = p(i | h_1, h_2) \cdot p(h_1, t) p(h_2, t)$$

$p(i | h_1, h_2)$ 表示当双亲程序为 h_1 和 h_2 时, 交叉点 i 被选择的条件概率, 而 $p(h_1, t)$ 和 $p(h_2, t)$ 表示双亲程序的选择概率。当一点交叉时

$$p(i | h_1, h_2) = 1 / NC(h_1, h_2)$$

因此

$$p(h_1, h_2, i, t) = \frac{p(h_1, t) p(h_2, t)}{NC(h_1, h_2)} \text{ 证毕。}$$

由于 $g(h_1, h_2, i, H)$ 是双态随机变量, 因此它的期望值代表值为“1”的项所占比例, 这对应了 h_1 和 h_2 的子代匹配模式 H 的比例, 所以 $a(H, t) = E[g(h_1, h_2, i, H)]$ 。

3.3 宏观精确 GP 模式定理

为了将式(4)扩展到宏观的表达形式, 让我们由各种可能的模式阶为0的定长模式作为砌基推导定理, 例如, G_1, G_2, \dots 。这些模式代表了不相关联的程序组, 它们的联合体则表示整个搜索空间。鉴于此, 我们重新得到:

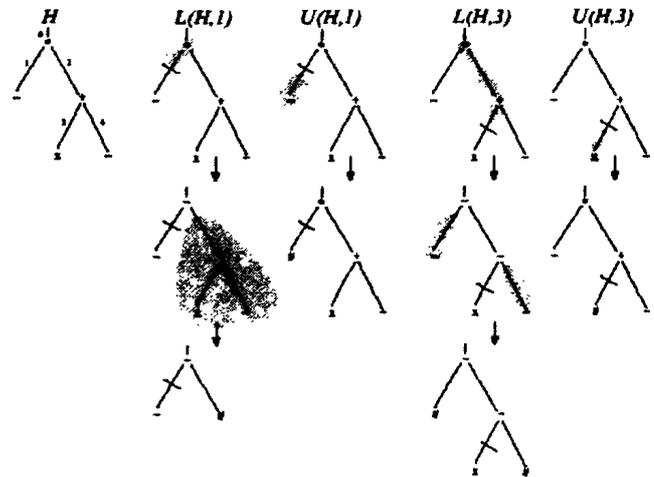


图2 模式实例及隐含的超模式积木块

$$\begin{aligned} \delta(h_1 \in L(H, i)) &= \sum_j \delta(h_1 \in L(H, i) \cap G_j) \\ &= \sum_j \delta(h_1 \in L(H, i)) \delta(h_1 \in G_j) \\ \delta(h_1 \in U(H, i)) &= \sum_k \delta(h_1 \in U(H, i) \cap G_k) \\ &= \sum_k \delta(h_1 \in U(H, i)) \delta(h_1 \in G_k) \\ a(H, t) &= \sum_{h_1} \sum_{h_2} \frac{p(h_1, t) p(h_2, t)}{NC(h_1, h_2)} \sum_{i \in C(h_1, h_2)} \sum_j \delta(h_1 \in L(H, i)) \delta(h_1 \in G_j) \sum_k \delta(h_2 \in U(H, i)) \delta(h_2 \in G_k) \\ &= \sum_j \sum_k \sum_{h_1} \sum_{h_2} \frac{p(h_1, t) p(h_2, t)}{NC(h_1, h_2)} \sum_{i \in C(h_1, h_2)} \delta(h_1 \in L(H, i)) \delta(h_1 \in G_j) \delta(h_2 \in U(H, i)) \delta(h_2 \in G_k) \\ &= \sum_j \sum_k \sum_{h_1 \in G_j} \sum_{h_2 \in G_k} \frac{p(h_1, t) p(h_2, t)}{NC(h_1, h_2)} \sum_{i \in C(h_1, h_2)} \delta(h_1 \in L(H, i)) \delta(h_2 \in U(H, i)) \\ &= \sum_j \sum_k \sum_{h_1 \in G_j} \sum_{h_2 \in G_k} \frac{p(h_1, t) p(h_2, t)}{NC(h_1, h_2)} \sum_{i \in C(G_j, G_k)} \delta(h_1 \in L(H, i)) \delta(h_2 \in U(H, i)) \end{aligned}$$

$$L(H, i) \delta(h_2 \in U(H, i)) = \sum_j \sum_k \frac{1}{NC(G_j, G_k)} \sum_{i \in C(G_j, G_k)} \sum_{h_1 \in G_j} p(h_1, t) \delta(h_1 \in L(H, i)) \sum_{h_2 \in G_k} p(h_2, t) \delta(h_2 \in U(H, i))$$

由此,我们获得如下的定理2。

定理2(宏观精确模式定理) 在一点交叉无变异的条件下,定长 GP 模式 H 的总体传递概率为:

$$\alpha(H, t) = (1 - p_{xo}) p(H, t) + p_{xo} \sum_j \sum_k \frac{1}{NC(G_j, G_k)} \sum_{i \in C(G_j, G_k)} p(L(H, i) \cap G_j, t) p(U(H, i) \cap G_k, t) \quad (6)$$

集合 $L(H, i) \cap G_j$ 和 $U(H, i) \cap G_k$ 或者是定长模式,或者是空集。所以,该模式定理代表了具有低阶(或同阶)模式 H 的总体传递概率。宏观精确模式定理既是对式(2)的广义概括,也是对 Holland 的 GA 模式定理的精确表达。

可以看出,式(6)中 $p(L(H, i) \cap G_j, t) p(U(H, i) \cap G_k, t)$ 项就代表了模式 $L(H, i) \cap G_j$ 和 $U(H, i) \cap G_k$ 的积木块假设。随着 GP 的运行,如果变异率很小(此处为0),基于一点交叉的 GP 群体开始收敛,规模和形状的多样性将会增加,程序包含的结点也越来越多。当和不同形状的程序交叉时, $p(L(H, i) \cap G_j, t)$ 和 $p(U(H, i) \cap G_k, t)$ 的概率将会增加,这正体现了积木块假设所揭示的内容。

结束语 定长精确模式定理清楚地阐明了 GP 模式创建机制,我们可以根据它推断算法收敛与否。同时,它是 John Holland 的 GA 模式定理在变异概率为0时更准确的表达,实现了由 GA 到 GP 模式定理的平滑过渡。通过对模型宏观和微观的分类,从不同角度分析了 GP 运行机理,超模式空间的引入也更好地支持了积木块假设。

参考文献

1 Poli Riccardo. Microscopic and Macroscopic Schema Theories for Genetic Programming and Variable-length Genetic Algorithms with One-Point Crossover, their Use and their Relations with Ear-

lier GP and GA Schema Theories. [Technical Report, CSRP-00-15]. The University of Birminham, UK 2000
 2 Poli Riccardo. Exact Schema Theorems for GP with One-Point and Standard Crossover Operating on Linear Structures and their Application to the Study of the Evolution of Size. [Technical Report, CSRP-00-14]. The University of Birminham, UK 2000
 3 Poli Riccardo. General Schema Theory for Genetic Programming with Subtree-Swapping Crossover. [Technical Report, CSRP-00-16]. The University of Birminham, UK 2000
 4 McPhee Nicholas Freitag. A schema theory analysis of the evolution of size in genetic programming with linear representations. [Technical Report, CSRP-00-22]. The Unive-rsity of Minnesota, USA 2000
 5 McPhee Nicholas Freitag. A schema theory analysis of mutation size biases in genetic programming with linear representations. [Technical Report, CSRP-00-24]. The Unive-rsity of Minnesota, USA 2000
 6 Poli Riccardo. Why the Schema Theorem is Correct also in the Presence of Stochastic Effects. In: Proc. of the Congress on Evolutionary Computation(CEC2000), San Diego, USA. 2000. 487~492
 7 Poli Riccardo, Langdon William B. Schema Theory for Genetic Programming with One-point Crossover and Point Mutation. Evolutionary Computation, 1998, 6(3): 231~252
 8 Poli Riccardo, Langdon William B. A Review of Theoretical and Experimental Results on Schemata in Genetic Programming. In: Proc. of the First European Workshop on Genetic Programming, Paris. 1998. 1~15
 9 Poli Riccardo. Hyperschema Theory for GP with One-Point Crossover, Building Blocks, and Some New Results in GA Theory. In: Genetic Programming, Proc. of EuroGP'2000, Edinburgh. 2000. 163~180
 10 Poli Riccardo, Langdon William B. A New Schema Theory for Genetic Programming with One-point Crossover and Point Mutation. In: Genetic Programming 1997, Proc. of the Seventh Intl. Conf. USA. 1997. 18~25

(上接第154页)

表3 识别率的排名

分类器	第一名	第二名	第三名	第四名	第五名
本方法	8	2	0	0	1
线性	0	3	4	0	4
二次	0	4	2	2	3
1-近邻	2	0	1	6	2
5-近邻	1	2	4	3	1

然而,对于数据集“青光眼”,该方法表现出了最低的性能,其原因在于当训练样本数很少时,MDL 标准倾向于选择简单判别规则而不是最优规则。

结束语 本文提出了基于勒让德多项式和 MDL 标准的构造非线性判别函数的方法,对很多问题该方法表现出了比其它方法好的性能。然而,当只有很少训练样本时,该方法不能构造一个好的规则。采用更有效的特征选择是可以进一步

深入研究的课题。

参考文献

1 Rissanen J. A Universal Prior for The Integers and Estimation by Minimum Discription Length. Ann. of Statistics, 1983, 11: 416~431
 2 Stone M. Cross-Validation Choice and Assessment of Statistical Predictions. J. Roy. Statist. Soc., 1974, 36: 111~147
 3 Japanese Technical Committee for Optical Character Recognition. ETL3 Character Database. Electrotechnical Laboratory, 1993
 4 Murphy P M, Aha D W. UCI Repository of Machine Learning Databases. [Machine Readable Data Repository]. University of California, Department of Information and Computer Science, Irvine, California, 1991
 5 边肇祺,张学工等编著. 模式识别(第二版). 清华大学出版社, 2000