

基于克隆选择算法的面向程序路径 测试数据生成方法

郑钧泽^{1,2}, 徐晓峰¹, 郭东辉^{1,3}

- (1. 厦门大学 物理系, 福建 厦门 361005;
2. 沈阳师范大学 软件学院, 辽宁 沈阳 110006;
3. 厦门大学 电子工程系, 福建 厦门 361005)

摘要: 面向程序路径的测试数据自动生成技术是软件测试自动化的关键技术之一。文中结合程序分支函数叠加法和克隆选择算法的全局搜索性, 提出一种基于克隆选择算法的面向程序路径测试数据生成方法。希望能够借助克隆选择算法具有多样性、记忆性、可实现快速全局优化搜索的优点, 设计一种新的面向程序路径的测试数据自动生成方法。对算法的原理和实现做了详细描述, 并将其与传统的基于遗传算法、模拟退火算法来实现软件测试数据自动生成方法进行实验对比, 证实了该方法能较快地生成指定路径的测试数据。

关键词: 克隆选择算法; 分支函数叠加法; 测试数据生成

中图分类号: TP311.5

文献标识码: A

文章编号: 1673-629X(2009)08-0008-03

A Path-Oriented Test Data Generation Approach Based on Clonal Selection Algorithm

ZHENG Junze^{1,2}, XU Xiaofeng¹, GUO Donghui^{1,3}

(1. Department of Physics, Xiamen University, Xiamen 361005, China;

2. School of Software, Shenyang Normal University, Shenyang 110006, China;

3. Department of Electronic Engineering, Xiamen University, Xiamen 361005, China)

Abstract: Path-oriented test data generation approach is one of key technologies in automatic software testing. Combining with function minimization methods and the search ability of clonal selection algorithm, a path-oriented test data generation approach based on clonal selection algorithm is presented. It designs a new automatic generating method for testing data based on path-oriented, which can be supported from the advantages such as diversity, remembrance and so on, by clonal selection algorithm. Describe the principle and realization of the approach in detail. By comparing with both genetic algorithm and simulated annealing algorithm, this test data generation based on clonal selection algorithm is proved to be more efficient in finding optimal solution.

Key words: clonal selection algorithm; function minimization method; test data generation

0 引言

面向程序路径的测试数据自动生成技术是软件测试自动化的关键技术之一, 它是指在软件测试过程中针对特定待测试的程序路径产生具体测试数据的方法^[1], 目前主要有: 随机法、符号执行法、启发式算法等几种^[2]。

对于较复杂的程序或测试策略, 随机法几乎不可能以合理的成本生成足够的测试数据; 符号执行法在复杂的代数运算、处理数组和指针引用时存在着困难, 也可能面临组合爆炸问题^[3]; 利用启发式算法进行测试数据生成是目前比较实用的方法, 其目前可分为: 禁忌搜索^[4]、模拟退火^[5]、遗传算法^[6-9]等几种, 但这些算法普遍存在着易于陷入局部极小值、全局搜索速度慢等缺点^[10]。因此笔者希望能够借助克隆选择算法^[11]具有多样性、记忆性、可实现快速全局优化搜索的优点, 设计一种新的面向程序路径的测试数据自动

收稿日期: 2008-11-26; 修回日期: 2009-03-01
基金项目: 国家自然科学基金(60753001); 福建省自然科学基金计划资助项目(A0410007); 国家人事部留学人员创业基金
作者简介: 郑钧泽(1985-), 男, 福建福州人, 研究领域为软件工程; 徐晓峰, 博士, CCF 会员, 研究方向为软件测试自动化; 郭东辉, 教授, 博士生导师, CCF 高级会员, 研究方向为人工智能、网络通信、集成电路设计。

生成方法。

文中首先介绍克隆选择算法进行测试数据生成的基本原理及需要解决的关键问题, 然后对该方法的实现进行详细描述, 最后将其与基于遗传算法、模拟退火算法的测试数据生成方法进行实验比对。

1 克隆选择的测试数据自动生成

克隆选择算法是模仿自然免疫系统功能的一种智能算法, 算法中抗原对应于待求解的问题, 抗体对应于问题的一个解。经过抗体的克隆、变异、选择, 在候选解的周围寻找更优解, 并引入消亡和补充机制保存优势抗体, 保持了抗体群的多样性, 使得克隆算法能快速、有效地搜索全局优化。在用克隆选择的测试数据自动生成时, 把被测程序的输入变量编码为抗体, 符合待测路径的测试数据比作抗原, 基本原理如图 1 所示。首先必须解决的问题有: 抗体编码与解码, 亲和度计算和算法终止准则。

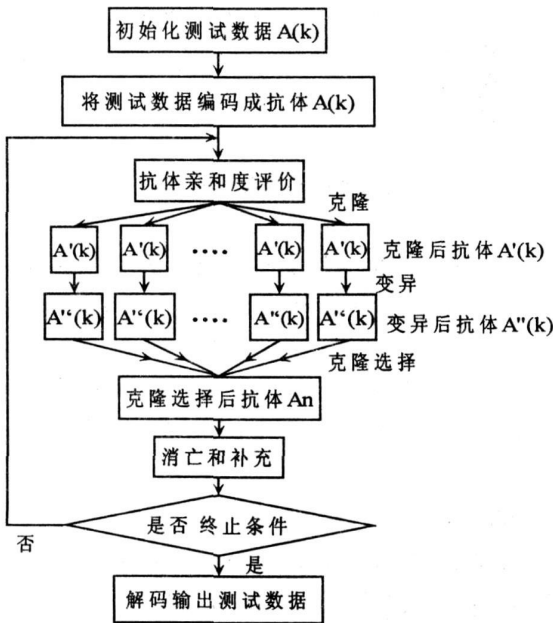


图 1 克隆算法进行测试数据自动生成原理图

1.1 抗体编码与解码

将多个输入参数采用多参数级联编码^[6]作为抗体, 即首先将每个参数 $x_i (i < n)$ 单独编码为长度为 L_i 二进制位串, 然后将所有的位串连接起来, 形成一个个体, 如图 2 所示。在编码时, 各个输出参数的位数可根据输入参数的类型进行设置。抗体解码时按照级联编码实施抗体编码的逆映射。

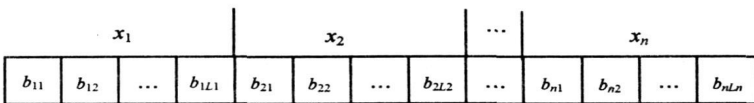


图 2 抗体编码

1.2 亲和度的计算

亲和度是抗体优劣的一种量化反映, 它的构造直接影响问题求解的效率。对于面向路径的测试数据生成, 我们的目标是构造一个能很好地评价所生成的测试数据的优劣, 并能引导克隆算法最终找到目标参数值的函数。笔者利用“分支函数叠加法^[1]”作为亲和度评价函数的构造方法。假定被测程序有 n 个输入参数 x_1, x_2, \dots, x_n , 待测路径上有 m 个分支点, 在每个分支点前插入一个分支函数, $\varphi_1 = f_1(x_1, x_2, \dots, x_n), \varphi_2 = f_2(x_1, x_2, \dots, x_n), \dots, \varphi_m = f_m(x_1, x_2, \dots, x_n)$, 在程序结束处插入分支函数叠加形式的评价函数表达式:

$$F = \varphi_1 + \varphi_2 + \dots + \varphi_m \quad (1)$$

其中, 分支函数值 φ_i 计算采用 Tracey 计算方法, 如表 1^[12] 所示, 表中 k 为一个较小正数。当一组测试数据驱动被测单元时, 分支函数的取值反映了当前这组测试数据下的被测单元的实际执行路径与待测路径的偏离程度。

表 1 分支函数值

谓词(Predicate)	分支函数 φ (Object Function)
Boolean	If true then 0 else k
$a = b$	If $\text{abs}(a - b) = 0$ then 0 else $\text{abs}(a - b) + k$
$a \neq b$	If $\text{abs}(a - b) \neq 0$ then 0 else k
$a < b$	If $a - b < 0$ then 0 else $(a - b) + k$
$a \leq b$	If $a - b \leq 0$ then 0 else $(a - b) + k$
$a > b$	If $a - b < 0$ then 0 else $(b - a) + k$
$a \geq b$	If $a - b \geq 0$ then 0 else $(b - a) + k$
$\neg a$	Negation is moved inwards and propagated over a
a and b	$\varphi(a) + \varphi(b)$
a or b	$\min(\varphi(a), \varphi(b))$

1.3 算法终止准则

利用克隆选择算法进行测试数据生成的终止条件是:

- (1) 亲和度 $F = 0$, 即测试数据满足指定的测试路径。
- (2) 迭代次数超过上限, 此时认为此路径为不可达路径。

2 基于克隆选择的测试数据自动生成

基于以上的描述, 对基于克隆选择算法的测试数据自动生成方法进行详细描述:

(1) 初始化: 随机在各变量的输入范围内产生一组随机数, 编码成抗体群 A , 规模为 N 个抗体, 每个抗体的编码长度为 L_i 。

(2) 亲和度评价: 结合代码插装技术, 在待测路径

所经过的每个分支点前插入一个分支函数, 由(1)式计算抗体群 A 中各抗体的分支函数值 $F(A_i)$ ($i = 1, 2, \dots, N_{i-1}$), 则抗体 A_i 和抗原的亲密度 $f(A_i) = F_{\max} - F(A_i)$, 其中 F_{\max} 为抗体群 A 中分支函数的最大值, 将亲密度由大到小按降序排列, 得到 $A = \{A_1, A_2, \dots, A_N\}$ 且 $f(A_i) \geq f(A_{i+1}), i = 1, 2, \dots, N_{i-1}$ 。

(3) 克隆: 将第 i 个抗体 A_i 按公式 $q_i = \text{Int}(\frac{\beta V}{i})$ 克隆 q_i 个, 得到种群为 $N_c = \sum_{i=1}^N \text{Int}(\frac{\beta V}{i})$ 的克隆抗体群为:

$$A' = \{A_{11}, A_{12}, \dots, A_{q1}; A_{21}, A_{22}, \dots, A_{2q2}; \dots, A_{N1}, A_{N2}, \dots, A_{NqN}\}$$

其中 β 为克隆系数, 用来控制克隆的规模, Int 为取整函数。这样, 克隆使得亲密度越高的抗体克隆的规模就越大, 在很大程度上保证高亲密度抗体中的优秀基因得到更好的保存和发展。

(4) 变异: 对 A' 中的抗体按照突变概率 P_m 进行基本位突变操作, 形成新的抗体群 A'' 。

(5) 评价和选择: 计算抗体群 A'' 的亲密度, 从每一个子群中选出一个最优抗体组成新的抗体群 A_n , 以获得更多的全局最优解和局部最优解。

(6) 消亡和补充: 模拟生物克隆选择中 5% ~ 10% 的 B 细胞自然消亡的过程。随机产生规模为 N_r 的抗体群 A_r , 选出 N_s ($N_s = (0.05 \sim 0.1)N$) 个亲密度最高的抗体替换 A_n 中亲密度最低的 N_s 个抗体, 形成规模为 N 抗体群 A , 这样使算法具有记忆功能, 并实现了种群的多样化。

(7) 检查是否符合终止条件, 若符合, 则终止; 否则转到(2), 进入下一次迭代。

3 算法对比与分析

选择软件测试文献中使用最广的三角形问题、二元一次方程求解问题^[12]进行实验。三角形问题使用三个整数作为输入, 表示三角形的三条边。程序输出三角形的类型: 等边、直角、等腰、不等边或非三角形。二元一次方程求解问题使用三个整数作为方程 $ax^2 + bx + c = 0$ 的三个参数 a, b, c , 输出为二元一次方程的根。

为了验证基于克隆选择的测试数据生成的效果, 文中将其与传统基于遗传算法、模拟退火算法实现测试数据生成的方法进行以下实验作比较:

- (1) 生成等边三角形的测试数据;
- (2) 生成直角三角形的测试数据;
- (3) 二元一次方程存在两个相等实根的测试数

据。

克隆选择算法、遗传算法、模拟退火均采用多参数级联编码, 实验中每个输入参数均编码成 8 位, 即每个输入参数范围设定为 0 ~ 255。其他基本参数设置如下: 克隆选择算法抗体群 $N = 40, N_r = 1/2N, \beta = 0.6, P_m = 0.15$; 为了对比客观, 保证遗传算法、模拟退火算法与克隆算法有相同的计算量, 遗传算法设置为 100 (约为 $N_c + N_r$), 交叉率为 0.8, 变异率 $1/24^{1/3}$; 模拟退火的方法设置种群为 50, 初始温度为 500, 退火方式为 $t = 1/\lg(1 + T)$, 实验中三种算法各生成 100 次数据, 记录首次出现的解的迭代次数, 取算术平均值。结果如图 3 所示。

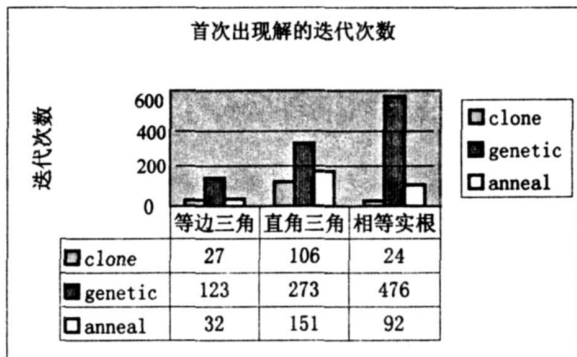


图 3 克隆选择算法与遗传算法、模拟退火算法迭代次数对比

图 3 可以看出克隆选择算法在生成测试数据的收敛速度上优于遗传算法和模拟退火算法, 这是由于克隆选择算法在生成测试过程中, 克隆、变异的过程相当于大范围的邻域的搜索, 同时引入消亡和补充保持了种群的多样性, 使其具有较高的测试数据生成效率。

4 结束语

将测试程序的输入编码成抗体融合克隆选择算法全局搜索性, 结合分支函数叠加法, 提出一种新的面向路径软件测试数据自动生成方法, 并与传统的遗传算法、模拟退火算法测试数据生成方法进行了实验对比, 证实该方法能较快地生成指定路径的测试数据。下一步工作可以通过改进亲密度函数使之能更有效地引导迭代, 提高算法效率。

参考文献:

- [1] Korel B. Automated software test data generation[J]. IEEE Trans on Software Eng, 1990, 16: 870- 879.
- [2] Pargas, Harrold R P, Jean M, et al. Test- data generation using genetic algorithms[J]. Journal of Software Testing Verification and Reliability, 1999, 9(4): 263- 282.

(下转第 14 页)

MCMMBCR 算法在选择路由时将节点的移动速度作为一个参数,所以该算法所选择的路由不仅能够均衡使用节点的电源,而且具有较高的稳定性,从而网络生存期相对于 CMMBCR 算法来讲也 longer。

4 结束语

文中提出了一个具有移动特征的 ad hoc 网络电源感知路由协议 MCMMBCR。该算法在选择路由时不仅考虑到节点的电源状态,而且将节点的移动速度作为一个参数来综合考虑。仿真结果表明,该算法可以降低数据传输的时间延迟,带来较稳定的路由,同时能够延长网络的生存期。下一步的工作应该考虑到节点的差异性,即各个节点的处理能力和通信能力不同的情况下,对算法性能带来的影响。

参考文献:

[1] Ramanathan R, Redi J. A Brief Overview of Ad Hoc Networks: Challenges and Directions [J]. IEEE Communications Magazine: 50th Anniversary Communication Issue, 2002(5): 20–23.

[2] Feeney L. An energy – consumption model for performance analysis of routing protocols for mobile ad hoc networks [J]. ACM J. of Mobile Networks and Applications, 2001, 3(6): 239–249.

[3] Cartigny J, Simplot D, Stojmenovic I. Localized minimum – energy broadcasting in ad – hoc networks [C]// Proceedings of

INFOCOM 2003: Twenty – Second Annual Joint Conference of the IEEE Computer and Communications Societies. San Francisco [s. n.], 2003: 2210–2217.

- [4] 余根坚, 郑宝玉. 一个基于图邻接矩阵模型的多跳 Ad Hoc 网络路由协议及网络性能分析 [J]. 计算机科学, 2006, 33(11): 25–28.
- [5] 张德跃, 杨峰, 展中华, 等. 传感器网络的一种能量感知分簇路由算法 [J]. 计算机技术与发展, 2007, 17(11): 67–69.
- [6] 冯运仿. 一种提高 Ad Hoc 网络节点能量效率的路由协议 [J]. 网络与信息, 2008, 22(1): 38–39.
- [7] 彭革刚, 杨疆湖, 高传善. 移动自组网中基于路径稳定性的 QoS 路由协议 [J]. 计算机研究与发展, 2004, 41(6): 916–922.
- [8] 李桓, 黄传河. 移动 Ad hoc 网络路由协议的一种节能策略 [J]. 微电子学与计算机, 2004, 21(5): 44–47.
- [9] Singh S, Raghavendra C S. PAMAS – Power Aware Multi – Access protocol with Signalling for Ad hoc Networks [J]. ACM Comm. Review, 1998, 28(3): 5–26.
- [10] Singh S, Woo M, Raghavendra C S. Power – Aware Routing in Mobile Ad hoc networks [C]// Proceedings of MobiCom'98. Dallas, Texas: [s. n.], 1998: 181–190.
- [11] Toh C K. Maximum Battery Life Routing to Support Ubiquitous Mobile Computing in Wireless Ad Hoc Networks [J]. IEEE Communications Magazine, 2001(1): 138–147.
- [12] 许力, 郑宝玉. MANET 环境下基于能量保护的路由策略及其研究进展 [J]. 电子与信息学报, 2005, 27(5): 827–834.

(上接第 10 页)

[3] 薛云志, 陈伟, 王永吉, 等. 一种基于 MessyGA 的结构测试数据自动生成方法 [J]. 软件学报, 2006, 17(8): 1688–1697.

[4] Eugenia D, Javier T, Raquel B. Automated software testing using a metaheuristic technique based on Tabu Search [C]// In 18th IEEE International Conference on Automated Software Engineering. Montreal, Canada: [s. n.], 2003: 310–313.

[5] Tracey N, Clark J, Mander K. Automated program flaw finding using simulated annealing [C]// In Proceeding of ACM SIGSOFT international symposium on Software testing and analysis. Florida, USA: [s. n.], 1998: 73–81.

[6] 英伟, 谢军, 奚红宇, 等. 遗传算法在软件测试数据生成中的应用 [J]. 北京航空航天大学学报, 1998, 24(4): 434–437.

[7] Lin J C, Yeh P L. Using Genetic Algorithms for Test Case Generation in Path Testing [C]// In Proceedings of the 9th Asian Test Symposium. Taipei, Taiwan: [s. n.], 2000: 241–247.

[8] Pei M, Goodman E D, Gao Z, et al. Automated Software Test

Data Generation Using A Genetic Algorithm [R]. GARAGE of Michigan State University, 1994.

- [9] 伦立军, 丁雪梅, 李英梅. 基于遗传算法的测试数据生成研究 [J]. 计算机工程, 2005, 31(23): 82–84.
- [10] Gao Haichang, Feng Boqin, Zhu Li. A kind of SAA GA hybrid meta – heuristic algorithm for the automatic test data generation [C]// In Proceedings of 2005 International Conference on Neural Networks and Brain. Beijing, China: [s. n.], 2005: 111–114.
- [11] 杜海峰. 免疫克隆计算与人工免疫网络研究与应用 [R]. 西安: 西安电子科技大学, 2003.
- [12] Tracey N, Clark J, Mander K, et al. An Automated Framework for Structural Test Data Generation [C]// In Proceedings of 13th IEEE International Conference on Automated Software Engineering. Honolulu, Hawaii, USA: [s. n.], 1998: 285–288.
- [13] Jones B F, Sthamer H H, Eyres D E. Automatic structural testing using genetic algorithms [J]. Software Eng, 1996(9): 299–306.