

## 基于改进鱼群算法的路径测试数据生成

王培崇<sup>1,2\*</sup>, 钱旭<sup>1</sup>

(1. 中国矿业大学 机电与信息工程学院, 北京 100083; 2. 石家庄经济学院 信息工程学院, 石家庄 050031)

(\* 通信作者电子邮箱 wpeichong@126.com)

**摘要:**针对自动化软件测试中测试数据自动化生成的问题,提出了一种基于人工鱼群算法的解空间搜索方案。为了提高人工鱼群算法的求解能力,在鱼群算法中引入混沌搜索机制。人工鱼群算法在每次迭代之后,将针对当前全局最优解进行局部混沌搜索,同时淘汰掉部分劣质个体;随后,根据种群的最佳个体收缩解空间搜索区域,并在该空间内随机产生部分新个体。最后,通过在两种三角形判定程序上的实验证明,该算法收敛速度快,求解精度高。

**关键词:**先进计算;人工鱼群算法;混沌搜索;软件测试;数据生成

**中图分类号:** 文献标志码:A

### Path test data generation based on improved artificial fish swarm algorithm

WANG Peichong<sup>1,2\*</sup>, QIAN Xu<sup>1</sup>

(1. School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing 100083, China;

2. School of Information Engineering, Shijiazhuang University of Economics, Shijiazhuang Hebei 050031, China)

**Abstract:** To solve the path test data generation automatically in software testing, a new scheme on searching solution space based on Artificial Fish Swarm (AFS) algorithm was proposed. To improve the ability of original AFS, chaotic searching was introduced to reform AFS' local searching ability and precision of solution. Once AFS finished an iteration process, chaos algorithm was executed with global best solution. At the same time, some partial individuals with bad state were washed out. Then, according to the optimization individual contracting the searching space, some new individuals were generated randomly. Two kinds of triangle program were tested and the results show that the improved AFS has faster convergence and higher calculation accuracy.

**Key words:** advanced computation; Artificial Fish Swarm Algorithm (AFSA); chaotic searching; software testing; data generating

## 0 引言

软件测试<sup>[1-3]</sup>是软件质量保证中重要的环节。其中,面向软件结构的白盒测试是软件测试中主要测试方法之一。在白盒测试中,所设计测试用例的质量直接影响软件测试的效果。随着软件模块规模的不断增加,面向结构的测试用例的组合将会出现海量增长,成为了一个 NP 难问题<sup>[4-5]</sup>。针对此问题,相关的研究成果更多倾向于使用启发式群体智能算法进行求解。文献[6]提出了基于启发式贪婪搜索框架的测试用例约简算法;文献[7]提出了基于改进差分演化算法的测试数据自动生成机制;文献[8]通过研究遗传算法机制,并提出了一种基于改进遗传算法的面向路径测试的测试数据生成方法;文献[9]在基本遗传算法的基础上,通过引入“位权重”和“邻居”的机制,设计了能够产生多条路径上测试数据的改进遗传算法求解机制。上述算法中,差分演化算法运算效率较高,但是求解离散优化问题需要额外转换;而遗传算法则存在较易收敛于局部最优的问题。针对此问题,本文提出了一种混沌<sup>[10-11]</sup>人工鱼群<sup>[12-14]</sup>算法,充分将鱼群算法的快速性和混沌搜索的精确性结合,并应用该算法成功求解了路径测试中的数据生成问题。

## 1 路径测试的模型

设存在程序 P,其输入变量为  $I = (x_1, x_2, \dots, x_n)$ ,  $D_{x_i}$  为  $x_i$  的定义域,则程序 P 的输入域空间表示为:  $D = D_{x_1} \times D_{x_2} \times \dots \times D_{x_n} = \{ (x_1, x_2, \dots, x_n) \mid \forall x_i \in D_{x_i} \}$ 。P 的某一路径为  $path = \langle n_{k_1}, n_{k_2}, \dots, n_{k_p} \rangle$ , 则用例生成问题的求解目标即寻找一组输入  $I \in D$ , 使得生成用例沿路径  $path$  执行。表示为  $test\_case = (path, I_0, E, S)$ ,  $path$  为待生成路径,  $I_0$  为初始测试用例,  $E$  为评价策略,  $S$  是空间搜索机制。

路径测试中,决定程序执行转向的是判定节点,将判定节点的分支谓词定义为  $e_1(x) \text{ op } e_2(x)$ , 其中  $e_1(x)$  和  $e_2(x)$  为谓词表达式,  $op$  是关系操作符。将该表达式等价转换为  $f(x)$   $rel$  0 的形式,则分支函数与分支谓词的对应关系如表 1 所示。

表 1 分支函数表达式

分支谓词	分支函数表达式	rel 符号
$e_1(x) > e_2(x)$	$f(x) = e_2(x) - e_1(x)$	( < )
$e_1(x) \geq e_2(x)$	$f(x) = e_2(x) - e_1(x)$	( ≤ )
$e_1(x) < e_2(x)$	$f(x) = e_1(x) - e_2(x)$	( < )
$e_1(x) \leq e_2(x)$	$f(x) = e_1(x) - e_2(x)$	( ≤ )
$e_1(x) = e_2(x)$	$f(x) = \text{abs}(e_1(x) - e_2(x))$	( = )
$e_1(x) \neq e_2(x)$	$f(x) = -\text{abs}(e_1(x) - e_2(x))$	( ≤ )

收稿日期:2012-10-11;修回日期:2012-11-29。

基金项目:教育部高校博士点建设基金资助项目(20110023110002);石家庄经济学院博士科研基金资助项目(2010)。

作者简介:王培崇(1973-),男,河北辛集人,副教授,博士,主要研究方向:智能信息处理、软件质量保证;钱旭(1962-),男,江苏南京人,教授,博士生导师,博士,主要研究方向:智能信息处理。

在测试数据驱动下,分支函数可以描述被测单元的实际执行路径对于选定路径的覆盖程度。该函数的构造是为了解决程序直接执行过程中的分支冲突,评价当前的测试效果<sup>[1]</sup>。

1)  $f(x) > 0$ , 表示分支谓词为 false, 表明在分支节点处实际执行路径与指定路径存在冲突,  $f(x)$  值越大, 测试数据与目标数据相差越大。

2)  $f(x) < 0$ , 表示分支谓词为 true, 表明当前测试数据能够覆盖指定路径上的该分支节点, 即为目标测试数据。

## 2 基于鱼群算法的测试数据生成

### 2.1 基本鱼群算法

主要参数:  $visual$ , 鱼的视野范围;  $\lambda$ , 拥挤因子 ( $0 < \lambda < 1$ );  $step$ , 移动步长;  $try\_number$ , 尝试次数。

设个体鱼所处位置的食物浓度为  $f(X)$ , 即求解目标函数。当前鱼个体的状态设为  $X = (X_1, X_2, \dots, X_n)$ , 以求解最小化问题  $\min f(X)$  为例, 鱼群算法主要算子如下:

1) 自由游动: 当前鱼个体在自己视野范围  $visual$  内随机游动。

2) 觅食算子: 在视野范围  $visual$  之内随机选择一新状态  $X_j$ , 如果  $f(X_j) < f(X_i)$ , 当前鱼向该方向前进一步; 否则, 选择一个新的  $X_j$  进行尝试。尝试  $try\_number$  次后仍然不能移动, 则随机移动一步。

3) 聚群算子: 在当前鱼个体的视野范围  $visual$  内搜索聚集鱼群的中心位置  $X_c$ , 并探测附近的同伴个数  $s$ 。如果该中心位置不拥挤 ( $s/N < \lambda$ ), 并且  $f(X_c) < f(X_i)$  则当前鱼向该方向前进一步, 否则执行觅食算子。

4) 追尾算子: 当前鱼个体在其视野范围  $visual$  内搜索最优鱼个体  $X_{min}$ 。设  $X_{min}$  视野领域内的伙伴数为  $s'$ , 如果  $f(X_i) > f(X_{min})$  且  $s'/N < \lambda$ , 表明该位置食物较多且不拥挤, 则当前鱼向该位置前进一步, 否则执行觅食算子。

更新公告板: 设置一个公告板, 用于记录鱼群内历史最佳鱼的状态  $X_{gbest}$ 。各人工鱼个体每次迭代完毕后检查自身状态, 如果  $f(X_i) < f(X_{gbest})$ , 修改  $X_{gbest}$  为  $X_i$ 。

### 2.2 改进的混沌人工鱼群算法

鱼群算法收敛速度较快, 但是由于种群多样性降低, 算法容易收敛于局部最优。混沌搜索机制可以很好地避免算法陷入局部最优, 但是单纯的混沌搜索技术需要较大的迭代才能获得最优解, 且对初始值敏感。

通过对两个算法的分析可知, 如果将 AFSA 的全局搜索结果作为混沌搜索的初始值, 势必会降低混沌搜索的迭代次数, 改变其对初始值敏感的缺点, 提高求解精度。同时, 为了增加种群的多样性, 在鱼群算法后期, 根据当前最佳鱼个体的位置动态收缩搜索区域, 在该区域内随机产生多个新个体, 替换掉适应度较低的部分个体。

混沌搜索算法求解问题时, 首先需要执行 Logistic 映射, 将混沌空间映射到待优化问题的解空间上, 映射函数为:  $x_{k+1} = u \cdot x_k(1 - x_k)$ ,  $0 \leq x_k \leq 1$ ; 然后针对该空间利用混沌动态特性实现求解。当  $u = 4$  并且  $x_0 \notin \{0, 0.25, 0.5, 0.75, 1\}$  时, Logistic 映射序列呈现混沌动态特征。混沌搜索算法的进化公式为:  $mx_i^{k+1} = 4mx_i^k(1 - mx_i^k)$ ,  $i = 1, 2, \dots, n$ ,  $mx_i^k$  为第  $j$  个混沌变量,  $k$  为迭代次数, 当  $mx_i^0 \in (0, 1)$  并且  $mx_i^0 \notin \{0.25, 0.5, 0.75\}$  则  $mx_i^k$  将会在  $(0, 1)$  内分布。混沌搜索算法 (Chaos Local Search Algorithm, CLSA) 见算法 1。

算法 1 混沌搜索算法 (CLSA)。

步骤 1 迭代变量  $k = 0$ 。设决策变量为  $x_i^k$  ( $i = 1, 2, \dots,$

$n$ ), 执行式(1)的映射变化;

$$mx_i^k = (x_i^k - x_{\min,i}) / (x_i^k - x_{\max,i}) \quad (1)$$

$x_{\min,i}$  和  $x_{\max,i}$  分别表示第  $i$  维变量的下界和上界。

步骤 2 执行公式  $mx_i^{k+1} = 4mx_i^k(1 - mx_i^k)$  ( $i = 1, 2, \dots, n$ ),

$n$ ) 计算得到下一次迭代的变量;

步骤 3 执行式(2)将混沌变量转换为决策变量;

$$x_i^{k+1} = x_{\min,i} + mx_i^{k+1}(x_{\max,i} - x_{\min,i}); i = 1, 2, \dots, n \quad (2)$$

步骤 4 根据决策变量  $x_i^{k+1}$  ( $i = 1, 2, \dots, n$ ) 对新解进行性能评价;

步骤 5 设初始解为  $X^0 = [x_1^{(0)}, \dots, x_n^{(0)}]$ , 若新解优于初始解或者迭代次数已经达到最大值, 则将新解输出, 否则返回步骤 2。

算法 2 混沌人工鱼群算法 (CLAFSA)。

步骤 1 在解空间内, 随机生成含有  $n$  个个体鱼的鱼群;

步骤 2 执行原始鱼群算法 AFSA 进行解空间搜索, 设获得最佳个体鱼为  $X_{gbest}$ ;

步骤 3 对于搜索完毕鱼群保留其中 30% 的较优秀的鱼个体, 淘汰其余 70% 个体;

步骤 4 对当前最佳个体  $X_{gbest}$  执行混沌搜索算法 (CLSA), 并将 CLSA 的求解结果替换  $X_{gbest}$ ;

步骤 5 若算法迭代次数满足或  $X_{gbest}$  在限定次数内基本不变化, 则输出  $X_{gbest}$ , 结束程序; 否则继续执行如下步骤;

步骤 6 按下述公式收缩搜索区域。

$$x_{\min,i} = \max\{x_{\min,i}, x_{g,i} - r(x_{\max,i} - x_{\min,i})\}; 0 < r < 1 \quad (3)$$

$$x_{\max,i} = \min\{x_{\max,i}, x_{g,i} + r(x_{\max,i} - x_{\min,i})\}; 0 < r < 1 \quad (4)$$

$x_{g,i}$  表示当前最佳  $X_{gbest}$  的第  $i$  维变量的值。

步骤 7 在收缩后的空间内随机产生 70% 的新个体, 然后返回步骤 2。算法流程如图 1 所示。

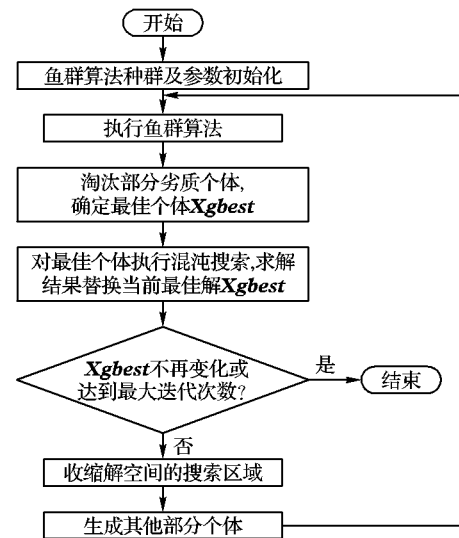


图1 CLAFSA 的程序流程

### 2.3 初始鱼群的构造

设种群空间范围为  $[x_{\min}, x_{\max}]$ , 种群定义为  $X(0) = \{X_1(0), X_2(0), \dots, X_n(0)\}$ , 代表提取的影响分支执行的变量, 对每一个变量随机赋实数值, 种群中的第  $i$  个个体标记为  $X_i(0) = \{x_{i1}(0), x_{i2}(0), \dots, x_{id}(0)\}$ 。

### 2.4 适应度函数的设置

根据第 2.2 节测试模型的分析可知, 测试数据的生成问题其实是寻找一个输入  $\{x_1, x_2, \dots, x_n\}$ , 使得该输入在程序执行时能够覆盖路径  $path$ 。假设该路径分支数为  $m$ , 第  $j$  个分

支函数标记为 $f_j(X_i(t))$ ,则将鱼群算法中的适应度函数记为

$$F(X_i(t)) = \sum_{j=1}^n \mu(f_j(X_i(t))).$$

$f_j(X_i) \leq 0$ 时,  $u(f_j(X_i)) = 0$ , 否则  $u(f_j(X_i)) = f_j(X_i)$ 。  
当适应度函数为0时,认为鱼群中个体鱼的状态  $X_i(t)$  能够全覆盖该位置所有分支。

### 2.5 择优处理

设最优个体为  $X_{gbest}$ , 最佳适应度  $gbest = F(X_{gbest}(t))$ , 算法对于优秀个体的选择操作方法如下:

$$\begin{aligned} &\text{If } F(X_i(t)) < F(X_{gbest}(t)) \\ &\quad X_{gbest}(t) = X_i(t) \end{aligned}$$

## 3 实验结果与对比

第2章通过研究混沌搜索算法和鱼群算法的原理,将鱼群算法与混沌搜索算法进行融合,设计了一个改进的混沌鱼群搜索算法,并提出了应用该算法求解路径测试数据生成的机制。为了验证该算法的有效性,本章将利用该算法求解等边三角形、直角三角形两种典型判定程序的路径测试测试数据生成,并将其求解结果与 PSO、AFSA 等进行对比。

### 3.1 等边三角形实验

为了与文献[7]中提供的算法进行比较,同样将本文算法 CLAFSA 的种群规模设置为 100, 120, 140, 160, 180, 200, 220, 240, 260, 280 等。针对每一个群体规模,CLAFSA 均执行 10 次,然后计算 10 次运行中寻找到最佳解的平均时间和平均迭代次数。图2、图3展示基于等边三角形实验对比情况。

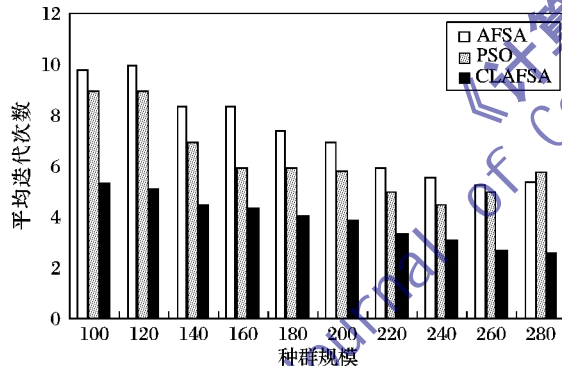


图2 等边三角形实验中平均迭代次数的比较

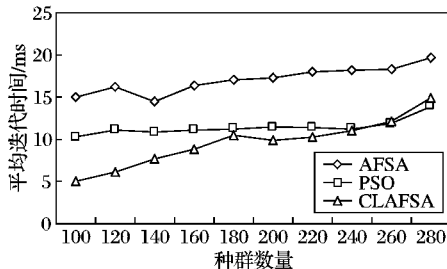


图3 等边三角形实验中平均迭代时间的比较

### 3.2 直角三角形实验

与等边三角形的实验设置相同,每一个算法执行 10 次,然后计算 10 次运行中寻找到最佳解的平均时间和平均迭代次数。图4~5显示的是基于直角三角形实验对比情况。

从上述两个实验的结果可以看出,由于 CLAFSA 算法引入了混沌局部搜索,其收敛速度明显较 AFSA 和 PSO 快,同时有效克服了 AFSA 求解精度低的缺点,不失为一种解决路径测试数据自动化生成的有效算法。

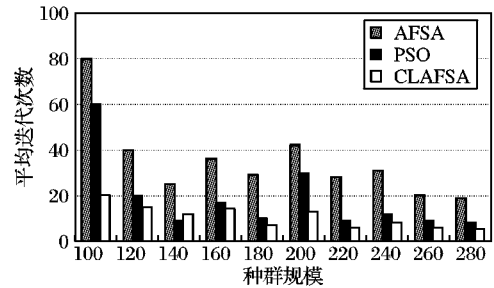


图4 直角三角形实验中平均迭代次数的比较

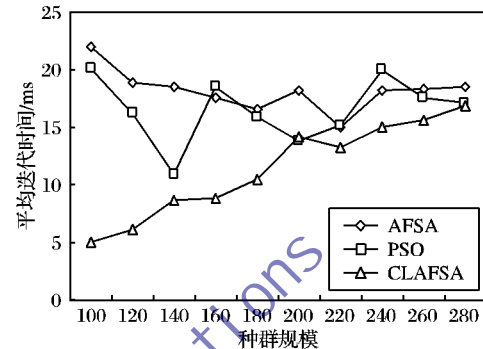


图5 直角三角形实验中平均迭代时间的比较

## 4 结语

测试用例自动化生成技术是软件质量保证中的一个重要问题。为了解决该问题,本文提出了应用人工鱼群算法解决该问题的思路,并对人工鱼群算法进行了相应的改进,通过引进混沌搜索技术提高了算法的求解精度,并有效避免了算法陷入局部最优。通过在直角三角形和等边三角形两个程序上的实验,证明了该算法在求解路径测试用例自动生成中的有效性。进一步的研究将从理论上分析该算法的收敛必然性,同时推广该算法用于解决其他相关问题。

### 参考文献:

- [1] 宫云战. 软件测试[M]. 北京: 国防工业出版社, 2006.
- [2] 杜庆峰. 高级软件测试技术[M]. 北京: 清华大学出版社, 2011.
- [3] SOMMERVILLE I. 软件工程[M]. 程成, 译. 北京: 机械工业出版社, 2012.
- [4] 孙家泽, 王曙燕. 基于改进 DPSO 的组合测试数据生成算法[J]. 计算机工程, 2012, 38(7): 40-45.
- [5] 聂鹏, 耿技, 秦志光. 软件测试用例自动生成算法研究综述[J]. 计算机应用研究, 2012, 29(2): 401-405.
- [6] 顾庆, 唐宝, 陈道蓄. 一种面向测试需求部分覆盖的测试用例集约简技术[J]. 计算机学报, 2011, 34(5): 879-887.
- [7] 黄小斌, 王希武, 常东升, 等. 改进的差分演化算法在测试数据生成中的应用[J]. 计算机应用, 2009, 29(6): 1722-1724.
- [8] 曹洋, 胡春华, 陈少波, 等. 一种路径测试数据自动生成方法[J]. 计算机工程, 2011, 37(9): 25-28.
- [9] 王元珍, 顾鹏. 产生多条路径上测试用例的改进遗传算法[J]. 计算机工程, 2006, 32(13): 196-198.
- [10] 王凌, 刘波. 微粒群优化与调度算法[M]. 北京: 清华大学出版社, 2008.
- [11] 邱劲, 王平. 基于混沌映射的压缩图像加密算法[J]. 计算机科学, 2012, 39(6): 44-46.
- [12] 彭勇, 唐国磊, 薛志春. 基于改进人工鱼群算法的梯级水库群优化调度[J]. 系统工程理论与实践, 2012, 31(6): 1118-1125.
- [13] PENG YONG. An improved artificial fish swarm algorithm for optimal operation of cascade reservoirs[J]. Journal of Computers, 2011, 6(4): 740-746.
- [14] 王培崇, 钱旭, 雷凤君. 一种新的小生境人工鱼群聚类算法[J]. 计算机应用, 2012, 32(8): 2189-2192.