

基于 K 均值 PSOABC 的测试用例自动生成方法

贾冀婷

(西安邮电大学 计算机学院, 陕西 西安 710121)

摘 要:软件测试中测试用例自动生成技术对于确保软件质量与降低开发成本都是非常重要的。文中基于 K 均值聚类算法与粒子群算法和人工蜂群算法相结合的混合算法,提出了一种测试用例自动生成方法,并且对此方法进行了仿真实验。实验结果表明,与基本的粒子群算法、遗传算法的测试用例自动生成方法相比较,基于文中改进算法的测试用例自动生成方法具有测试用例自动生成效率高、收敛能力强等优点。

关键词:软件测试;测试用例;粒子群优化算法;人工蜂群算法; K 均值聚类算法

中图分类号: TP311

文献标识码: A

文章编号: 1673-629X(2015)06-0012-04

doi:10.3969/j.issn.1673-629X.2015.06.003

Automatic Testcase Generation Method Based on PSOABC and K -means Clustering Algorithm

JIA Ji-ting

(College of Computer, Xi'an University of Posts and Telecommunications,
Xi'an 710121, China)

Abstract: To improve the automation ability of testcase generation in software testing is very important to guarantee the quality of software and reduce the cost of software. In this paper, propose an automatic testcase generation method based on particle swarm optimization, artificial bee colony algorithm and K -means clustering algorithm, and carry out the simulation experiments. The results show that the improved algorithm's efficiency is better and convergence ability is stronger than other algorithms such as particle swarm optimization and genetic algorithm in the automation ability of testcase generation.

Key words: software testing; testcase; particle swarm optimization; artificial bee colony algorithm; K -means clustering algorithm

0 引言

随着计算机技术的飞速发展,计算机软件出现的错误种类也越来越多,随之带来的软件安全隐患也越来越多。软件测试技术是确保软件开发质量与提高软件可靠性的关键。但软件测试需要投入大量的人力、物力以及时间,因此软件测试中的自动化程度对于确保软件质量与降低开发成本都是非常重要的。其中,测试用例自动生成技术又是提高软件测试自动化程度的关键。

粒子群优化算法和人工蜂群算法都是基于群体智能的全局随机优化算法,并且都具有控制参数少、易于实现、计算简单等优点,但粒子群优化算法在寻找最优解的过程中存在容易陷入局部最优的缺点,即算法没有较强的搜索能力,而人工蜂群算法中的搜索算子则

侧重于提高搜索能力。

因此,文中将这两种算法相结合,提出了一种新的混合算法 PSOABC。 K 均值聚类算法是根据函数准则进行分类的聚类算法,具有容易理解、简单易实现、收敛速度快等优点,但是它存在对初始聚类中心选取敏感、容易收敛到局部最优解等缺点。因此,文中提出了基于 PSOABC 与 K 均值聚类算法相结合的改进算法的测试用例自动生成算法。并通过仿真实验与基本的粒子群优化算法、遗传算法的测试用例自动生成方法进行比较。

实验结果表明,基于文中改进的 PSOABC 与 K 均值聚类算法的测试用例自动生成方法具有测试用例自动生成效率高、收敛能力强等优点。

收稿日期: 2014-07-15

修回日期: 2014-10-22

网络出版时间: 2015-05-06

基金项目: 国家自然科学基金资助项目(61050003); 陕西省科技攻关计划项目(2009K08-26); 陕西省教育自然科学基金项目(09JK727)

作者简介: 贾冀婷(1978-),女,硕士,工程师,研究方向为软件测试、智能优化算法。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150506.1627.018.html>

1 粒子群优化算法与人工蜂群算法相结合的混合算法 PSOABC

1.1 粒子群优化算法

粒子群优化算法^[1-4]是一种模拟鸟群觅食过程中群体行为的群体智能算法,该算法通过种群中粒子之间的相互协作与信息共享来实现最优解的搜寻。粒子群优化算法具有控制参数少、计算简单、易于实现等优点。

粒子群优化算法的具体步骤为:首先随机生成初始种群,并为种群中每个粒子都确定一个适应值。在寻找最优解的过程中,种群中的每个粒子都会追随两个极值,即个体极值与全局极值。所谓个体极值是粒子自身迄今为止找到的最优解;而全局极值是整个种群迄今为止找到的最优解。最终,种群中的粒子经过迭代搜寻得到最优解。

粒子群优化算法的数学表示为:设在一个 n 维的空间中,有一个种群包含 m 个粒子,即 $x = \{x_1, \dots, x_i, \dots, x_m\}$, 第 i 个粒子位置表示为 $x_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}^T$, 其速度表示为 $v_i = \{v_{i1}, v_{i2}, \dots, v_{in}\}^T$ 。第 i 个粒子的个体极值表示为 $P_i = \{P_{i1}, P_{i2}, \dots, P_{in}\}^T$, 种群的全局极值表示为 $P_g = \{P_{g1}, P_{g2}, \dots, P_{gn}\}^T$, 通过粒子群算法的基本思想,第 i 个粒子可由公式(1)、(2)改变其速度和位置:

$$v_{id}^{(t+1)} = v_{id}^{(t)} + c_1 r_1 (p_{id}^{(t)} - x_{id}^{(t)}) + c_2 r_2 (p_{gd}^{(t)} - x_{id}^{(t)}) \quad (1)$$

$$x_{id}^{(t+1)} = x_{id}^{(t)} + v_{id}^{(t+1)} \quad (2)$$

其中, $d \in \{1, 2, \dots, n\}$; $i \in \{1, 2, \dots, m\}$; t 表示当前的进化代数; r_1, r_2 是分布在 $[0, 1]$ 之间的随机数; c_1 和 c_2 是学习因子。

1.2 人工蜂群算法

人工蜂群算法^[5-9]是对自然界中蜜蜂采蜜过程中所表现出的智能行为进行模拟。人工蜂群算法具有控制参数少、易于实现、计算简单等优点。

人工蜂群算法将蜜蜂根据不同的分工分为采蜜蜂、观察蜂和侦察蜂三种,每个蜜源的位置代表一个可能解,蜜源的花蜜量对应着每个可能解的适应值。算法的具体步骤为:

首先,随机产生初始种群,即 N 个初始解。每个解 $x_i (i = 1, 2, \dots, N)$ 表示一个 D 维的向量, D 为优化参数的个数。接着,三种蜜蜂分别开始对所有的初始解进行循环搜索。采蜜蜂记住自身历史的最优解,在附近邻域进行搜索,其搜索公式为:

$$x_{ij}' = x_{ij} + \varphi(x_{ij} - x_{kj}) \quad (3)$$

其中, $k \in \{1, 2, \dots, N\}$, $j \in \{1, 2, \dots, D\}$, 且 $k \neq i$; φ 是 $[-1, 1]$ 之间的随机数。

观察蜂选择某个蜜源的概率 q_i 的计算公式为:

$$q_i = \frac{f_i}{\sum_{n=1}^N f_n} \quad (4)$$

其中, f_i 表示第 i 个解的适应值; N 表示蜜源数量。

如果一个蜜源经过限定的循环次数之后没有发生变化,则该蜜源处的采蜜蜂变为侦察蜂,该蜜源位置将会被侦察蜂找到的随机新位置替换。假设被放弃的位置为 x_i , 则侦察蜂替换 x_i 的公式如下:

$$x_{ij}' = x_{\min}^j + \text{rand}(0, 1)(x_{\max}^j - x_{\min}^j) \quad (5)$$

1.3 粒子群优化算法与人工蜂群算法相结合的混合算法 PSOABC

粒子群优化算法在寻找最优解的过程中存在容易陷入局部最优的缺陷,即算法没有较强的搜索能力,而人工蜂群算法中的搜索算子则侧重于提高搜索能力。因此,文中将这两种算法相结合,提出一种新的混合优化算法 PSOABC,用来提高算法的寻优能力。为了进一步提高算法的搜索能力,PSOABC 将初始种群随机分为两组,其中一组中的个体按照粒子群优化算法进行寻优,而且另一组中的个体按照人工蜂群算法进行寻优。

在寻优过程中,每隔一定的迭代次数对两组种群产生的最优个体进行比较,找到相对更优的个体,使两组避免陷入局部最优。具体的 PSOABC 算法的步骤如下:

(1) 设置参数,初始化种群;

(2) 将种群随机分为两组,第一组中的个体按照粒子群算法进行寻优,第二组中的个体按照人工蜂群算法进行寻优;

(3) 当迭代次数达到指定次数时,对两组产生的最优个体进行比较,将较优个体的值赋给公式(1)中的 p_g , 并从第一组中随机选取值赋给公式(3)中 x_k ;

(4) 两组分别通过公式(1)~(3)产生新解,并计算其适应值;

(5) 第一组更新 p_i 和 p_g ;

(6) 第二组中如果新解的适应值优于 x_i , 则替换 x_i , 否则保持不变;计算 x_i 的适应值,并根据公式(4)计算概率 q_i ;观察蜂根据 q_i 选择蜜源,并根据公式(3)产生新解,如果新解的适应值优于 x_i , 则替换 x_i , 否则保持不变;

(7) 判断是否有需要放弃的解,若有,则按照公式(5)产生新解;

(8) 如果当前群体的最优个体满足精度要求或进化已达到最大迭代次数,则终止算法,否则跳转至(3)。

2 K 均值聚类算法

K 均值聚类算法^[10-14]是聚类分析最常用的方法之一,是解决聚类问题的一种基于划分的经典算法。该算法具有容易理解、简单实现、收敛速度快等优点。

K 均值聚类算法的基本思想是首先从 n 个数据中随机选择 k 个作为初始聚类中心,然后计算每个数据到各聚类中心的距离,根据最近邻原则,将所有数据划分到离它最近的那个中心所代表的聚类中,接着分别计算新生成的各聚类中数据的均值作为各聚类新的中心,比较新的聚类中心和上一次得到的聚类中心,如果新的聚类中心没有发生变化,则算法收敛,输出结果;如果新的聚类中心和上一次的聚类中心相比较发生变化,则要根据新的聚类中心对所有数据进行重新划分。直到满足算法的收敛条件为止。

设每个粒子 $x_i = (c_{i1}, c_{i2}, \dots, c_{ik})$ 代表 k 个聚类的中心点,其中 c_{ij} 代表第 i 个粒子的第 j 类中心点坐标向量。通常采用的适应度函数是:

$$f(x_i) = \frac{1}{\sum_{j=1}^k \sum_{x_i \in c_j} \text{dist}(x_i, c_{ij})} \quad (6)$$

其中, $\text{dist}(x_i, c_{ij})$ 表示一个数据元素 x_i 和一个聚类中心 c_{ij} 之间的欧氏距离。

3 基于 PSOABC 与 K 均值聚类算法的测试用例自动生成方法

3.1 适应值函数构造

文中通过对分支函数进行叠加来构造适应值函数。分支函数是分支谓词到实值的一个映射,能量化地描述在测试数据的驱动下,被测单元的实际执行路径对选定路径的覆盖程度。分支函数能真实地反映分支谓词的真假情况。文中的适应值函数构造如下:

设待测路径上有 m 个分支点, n 个参数,则 m 个分支函数分别为: $f_1 = f_1(x_1, x_2, \dots, x_n)$, $f_2 = f_2(x_1, x_2, \dots, x_n)$, \dots , $f_m = f_m(x_1, x_2, \dots, x_n)$; 而该路径的分支函数为:

$$F = C - (F(f_1) + F(f_2) + \dots + F(f_m)) \quad (7)$$

其中, $F(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$; C 为一个较大的整数。

3.2 基于 PSOABC 与 K 均值聚类算法的测试用例自动生成算法

K 均值聚类算法具有收敛速度快、局部搜索能力强等优点,但聚类结果容易受到初始聚类中心的影响,如果初始聚类中心选取不当算法极易陷入局部最优。为了避免 K 均值聚类算法陷入局部最优,文中提出了基于改进的 PSOABC 与 K 均值聚类算法的测试用例生成算法,该算法将改进的 PSOABC 算法进行演化搜索

得到全局最优的 K 个聚类中心,然后以此为初始聚类中心来进行 K 均值优化,从而获得理想的聚类划分。

文中基于改进的 PSOABC 与 K 均值聚类算法的测试用例自动生成算法的实现过程为:在每次进行迭代的过程中,利用改进的 PSOABC 算法搜寻出多个聚类中心,再将每一个 PSOABC 算法得到的聚类中心进行一次 K 均值优化,将以上这两种算法交替进行,直到聚类结束。文中提出的算法,一方面,通过 K 均值优化加快了算法的收敛速度;另一方面,通过改进的 PSOABC 算法的融入使算法对初始聚类中心的依赖性降低,提高了算法的鲁棒性,从而获得了较为理想的聚类划分。

算法的具体步骤描述如下:

(1) 设置参数,初始化种群;

(2) 利用改进的 PSOABC 算法搜索出多个新的聚类中心;

(3) 将得到的新的聚类中心进行一次 K 均值优化,按照最近邻原则进行聚类划分,并求出每一类新的聚类中心,并用新的聚类中心更新 PSOABC 算法的解;

(4) 如果当前群体的最优个体满足精度要求或进化已达到最大迭代次数,则终止算法,否则跳转至(2);

(5) 输出最后的聚类中心,并得到其适应值。

4 实验及结果分析

文中将一个判断三角形类型的程序作为被检测程序,分别使用基于改进的 PSOABC 与 K 均值聚类算法、基本的粒子群算法、遗传算法等测试用例自动生成算法自动生成测试用例判断三角形为例进行实验,给出实验数据并对并分析。

表 1 是文中提出的算法与基本的粒子群算法、遗传算法在数据数目的大小从 10 到 50 时,分别生成最优解的迭代次数以及相应的运行时间。仿真实验中分别对每个数据数目的大小各运行 10 次。表 1 记录了找到最优解迭代次数的平均值以及生成最优解所消耗时间的平均值。

表 1 文中算法与基本的粒子群算法、遗传算法生成三角形测试数据的迭代次数和时间

数据数目	找到最优解迭代的次数			找到最优解迭代的时间/ms		
	文中算法	粒子群算法	遗传算法	文中算法	粒子群算法	遗传算法
10	19.2	56.1	263.3	16.8	52.1	323.5
20	15.7	50.9	239.7	17.8	49.3	286.5
30	13.5	45.7	152.9	17.9	50.2	265.4
40	8.1	37.4	161.5	14.7	47.3	268.3
50	9.3	41.3	116.5	13.3	44.5	227.7

以平均迭代次数和平均运行时间作为评价指标,文中基于改进的 PSOABC 与 K 均值聚类算法与基本的

粒子群优化算法、遗传算法的实验数据进行比较。可明显看出,文中改进算法的各项评价指标都比基本的粒子群算法、遗传算法的要更优。因此,基于改进的 PSOABC 与 K 均值聚类算法生成测试用例的效果明显比较好。

5 结束语

实验结果表明,以平均迭代次数和平均运行时间作为评价指标,文中基于改进的 PSOABC 与 K 均值聚类算法与基本的粒子群优化算法、遗传算法的实验数据进行比较,可明显看出改进算法的各项评价指标都比基本的粒子群算法、遗传算法的要更优。由此可见,基于改进的 PSOABC 与 K 均值聚类算法的测试用例自动生成方法具有测试用例自动生成效率高、收敛能力强等优点,且有一定的实用性。

参考文献:

- [1] Eberhart R C, Kennedy J. A new optimizer using particle swarm theory[C]//Proc of the sixth international symposium on micro machine and human science. Nizgoya, Japan: [s. n.],1995:39-43.
- [2] Kennedy J, Eberhart R C. Particle swarm optimization[C]//Proc of IEEE conf on neural network. Perth:IEEE,1995:1942-1948.
- [3] Kennedy J, Eberhart R C. Swarm intelligence[M]. San Francisco, CA: Morgan Kaufmann Publishers Inc, 2001.
- [4] 郭长友. 一种自适应惯性权重的粒子群优化算法[J]. 计算机应用与软件, 2011, 28(6): 289-292.
- [5] Karaboga D, Basturk B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony algorithm[J]. Journal of Global Optimization, 2007, 39(3): 459-471.
- [6] Karaboga D. An idea based on honey bee swarm for numerical optimization[D]. Ercyes; Ercyes University, 2005.
- [7] Karaboga D, Basturk B. A comparative study of artificial bee colony algorithm[J]. Applied Mathematics and Computation, 2009, 214(1): 108-132.
- [8] 刘 路, 王太勇. 基于人工蜂群算法的支持向量机优化[J]. 天津大学学报, 2011, 44(9): 803-809.
- [9] 毕晓君, 王艳娇. 改进人工蜂群算法[J]. 哈尔滨工程大学学报, 2012, 33(1): 117-123.
- [10] 谢秀华, 李陶深. 一种基于改进 PSO 的 K-means 优化聚类算法[J]. 计算机技术与发展, 2014, 24(2): 34-38.
- [11] 杨 韬, 邵良杉. 采用改进的 k 均值聚类分析策略的粒子群算法[J]. 计算机工程与应用, 2009, 45(12): 52-54.
- [12] 欧陈委. K-均值聚类算法的研究与改进[D]. 长沙: 长沙理工大学, 2011.
- [13] 周爱武, 陈宝楼, 王 琰. K-Means 算法的研究与改进[J]. 计算机技术与发展, 2012, 22(10): 101-104.
- [14] 傅景广, 许 刚, 王裕国. 基于遗传算法的聚类分析[J]. 计算机工程, 2004, 30(4): 122-124.
- [15] 罗 辛, 欧阳元新, 熊 璋, 等. 通过相似度支持度优化基于 K 近邻的协同过滤算法[J]. 计算机学报, 2010, 33(8): 1437-1445.
- [16] Anand D, Bharadwaj K K. Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities[J]. Expert Systems with Applications, 2011, 38(5): 5101-5109.
- [17] Zhou K, Yang S H, Zha H. Functional matrix factorizations for cold-start recommendation[C]//Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval. [s. l.]; ACM, 2011: 315-324.
- [18] Cergani E, Miettinen P. Discovering relations using matrix factorization methods[C]//Proceedings of the 22nd ACM international conference on information & knowledge management. [s. l.]; ACM, 2013: 1549-1552.
- [19] Zhang Y, Zhu X, Shen Q. A recommendation model based on collaborative filtering and factorization machines for social networks[C]//Proc of 5th IEEE international conference on broadband network & multimedia technology. [s. l.]; IEEE, 2013: 110-114.
- [20] Yu K, Zhu S, Lafferty J, et al. Fast nonparametric matrix factorization for large-scale collaborative filtering[C]//Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval. [s. l.]; ACM, 2009: 211-218.
- [21] Takács G, Pilászy I, Németh B, et al. Scalable collaborative filtering approaches for large recommender systems[J]. Journal of Machine Learning Research, 2009, 10(3): 623-656.
- [22] 李 改, 李 磊. 基于矩阵分解的协同过滤算法[J]. 计算机工程与应用, 2011, 47(30): 4-7.
- [23] Takács G, Pilászy I, Németh B, et al. Investigation of various matrix factorization methods for large recommender systems[C]//Proc of IEEE international conference on data mining workshops. [s. l.]; IEEE, 2008: 553-562.
- [24] 杨 阳, 向 阳, 熊 磊. 基于矩阵分解与用户近邻模型的协同过滤推荐算法[J]. 计算机应用, 2012, 32(2): 395-398.

(上接第5页)