

## 基于自适应 $t$ 分布混合变异的人工萤火虫算法

杜晓昕\*, 张剑飞, 孙 明

(齐齐哈尔大学 计算机与控制工程学院, 黑龙江 齐齐哈尔 161006)

(\*通信作者电子邮箱 xiaoxin\_du@163.com)

**摘 要:**针对人工萤火虫(AGSO)算法中存在一些漫无目的随机运动的萤火虫及一些萤火虫在非全局极值点出现严重聚集时,收敛速度降低,甚至陷入局部极值的问题,提出一种基于自适应  $t$  分布混合变异的人工萤火虫算法。用自适应  $t$  分布变异和最优调教变异来增强种群的多样性,限制算法陷入局部最优;定义了变异控制因子对变异的运行进行控制,结合历史状态信息给出了自适应  $t$  分布混合变异描述。该变异方法能使算法同时提高全局探索能力和局部开发能力。通过典型函数算例和实际应用算例实验结果表明,该算法是可行有效的,比传统算法具有较快的寻优速度和较高的寻优精度。

**关键词:**人工萤火虫算法;自适应  $t$  分布变异;最优调教变异;变异控制因子;全局探索能力;局部开发能力

**中图分类号:**TP183 **文献标志码:**A

### Artificial glowworm swarm optimization algorithm based on adaptive $t$ distribution mixed mutation

DU Xiaoxin\*, ZHANG Jianfei, SUN Ming

(College of Computer and Control Engineering, Qiqihar University, Qiqihar Heilongjiang 161006, China)

**Abstract:** The convergence speed of Artificial Glowworm Swarm Optimization (AGSO) algorithm declines, even falls into local minimums, when some glowworms gather in non whole extreme points or some glowworms wander around aimlessly. Concerning this problem, an AGSO algorithm based on adaptive  $t$  distribution mixed mutation was proposed. Adaptive  $t$  distribution mutation and optimization adjustment mutation was introduced into the AGSO algorithm to improve the diversity of glowworm swarm, and prevent the AGSO algorithm from falling into local minimums. Mutation control factor was defined. Combining history status information, the description of adaptive  $t$  distribution mixed mutation was given. The mutation method could enhance ability of global exploration and local development. The emulation results of representative test functions and many application examples show that the proposed algorithm is reliable and efficient. Meanwhile, this algorithm is better than tradition algorithm in terms of speed and precision for seeking the optimum.

**Key words:** Artificial Glowworm Swarm Optimization (AGSO) algorithm; adaptive  $t$  distribution mutation; optimization adjustment mutation; mutation control factor; global exploration ability; local development ability

## 0 引言

人工萤火虫(Artificial Glowworm Swarm Optimization, AGSO)算法是印度学者 Krishnanad 等于 2005 年提出的模拟自然界萤火虫求偶和觅食行为的群体智能优化算法<sup>[1]</sup>。该算法实现简单,全局搜索能力强,具有很强的搜索自适应性和通用性等优点。目前人工萤火虫算法已应用于多模态函数优化<sup>[2]</sup>、背包问题<sup>[3]</sup>、集群机器人模拟<sup>[4]</sup>、聚类分析<sup>[5]</sup>和数值优化<sup>[6]</sup>等各类组合优化问题中,取得了较好效果。但随着人工萤火虫算法应用的不断拓广,人们发现算法中存在一些漫无目的随机运动的萤火虫或一些萤火虫在非全局极值点出现严重聚集时,收敛速度将大大减慢,甚至也会陷入局部极值,搜索性能劣化。如果把变异操作加入到人工萤火虫算法中,增强了种群的多样性,在一定程度上能避免算法陷入局部最优。

文献[2]的变异方法,需要人工设定变异因子值,变异操作缺乏自适应性,此外变异没有很好利用当前萤火虫已搜索到的状态信息。文献[7]提出的基于自适应高斯变异的人工

萤火虫(Adaptive Gauss Mutation-Artificial Glowworm Swarm Optimization, AGM-AGSO)算法,其变异方法充分利用公告板中的历史最优萤火虫状态信息和高斯变异的优点,变异具有自适应性,在一定程度上提高了收敛速度和求解精度,但由文献[8]可知,柯西变异的全局探索能力较强,而高斯变异的局部开发能力较强,因此文献[7]变异方法的性能优势在于提高了算法进化后期的局部开发能力,并没有提高算法进化初期的全局探索能力。由文献[9]可知  $t$  分布变异结合了柯西变异和高斯变异的两者优点,使得算法在进化初期既具有良好的全局探索能力,又在进化后期具有较优的局部开发能力。

本文采用的变异操作为自适应混合变异,变异充分利用公告板中的历史最优萤火虫状态信息,对非最优萤火虫进行自适应  $t$  分布变异,对最优萤火虫进行最优调教变异,这种混合变异方法既可以保证算法进化初期全局探索能力较强,有效地保持种群的多样性,又可以保证局部开发能力较强,使进化后期的收敛速度加快,整个变异操作无需人工参与具有良好的自适应性。

收稿日期:2013-01-22;修回日期:2013-02-24。

基金项目:国家自然科学基金资助项目(61100103);黑龙江省教育厅科学技术研究项目(12531758)。

作者简介:杜晓昕(1983-),女,江苏徐州人,讲师,硕士,主要研究方向:人工智能、图像处理;张剑飞(1974-),男,黑龙江齐齐哈尔人,副教授,博士研究生,主要研究方向:数据挖掘;孙明(1979-),男,山东烟台人,副教授,博士,主要研究方向:混沌神经网络。

## 1 人工萤火虫算法

AGSO 在求解空间内随机生成  $n$  只萤火虫, 每只萤火虫有自身的决策域, 所有萤火虫的荧光素初始值都设为相同的  $l_0$ 。第  $i$  只萤火虫在  $q$  时刻的位置为  $x_i(q) = [x_i^{(1)}(q), x_i^{(2)}(q), x_i^{(3)}(q), \dots, x_i^{(n)}(q)]$ , 第  $i$  只萤火虫在  $q$  时刻的位置的第  $p$  个分量为  $x_i^{(p)}(q)$ , 适应度函数为  $f(x)$ , 第  $i$  只萤火虫在  $q$  时刻的荧光素浓度为  $l_i(q)$ 。萤火虫向其邻居移动是按照一定的概率。对于萤火虫  $i$  和萤火虫  $j$ , 如果萤火虫  $j$  在萤火虫  $i$  的决策域内, 并萤火虫  $j$  的荧光素值比萤火虫  $i$  的高, 则萤火虫  $j$  为萤火虫  $i$  的邻居。萤火虫的所在的位置越好, 其荧光素值也越高, 这也说明目标函数越优。 $n$  只萤火虫经过不断的运动, 较多的萤火虫最终聚集在荧光素值较高的萤火虫周围。算法执行过程中每次迭代执行四个阶段, 它们分别为: 荧光素更新阶段、移动概率更新阶段、位置更新阶段和邻域范围更新阶段。

荧光素更新阶段 更新规则如式(1):

$$l_i(t) = (1 - \rho)l_i(t-1) + \gamma f(x_i(t)) \quad (1)$$

其中:  $\rho$  为荧光素挥发系数,  $\gamma$  为应适度提取比例。

移动概率更新阶段  $P_{ij}(t)$  表示第  $i$  只萤火虫  $t$  时刻向其邻居集合中第  $j$  只萤火虫移动的概率如式(2):

$$P_{ij}(t) = \frac{l_j(t) - l_i(t)}{\sum_{k \in N_i(t)} l_k(t) - l_j(t)} \quad (2)$$

其中:  $j \in N_i(t)$ ,  $N_i(t) = \{j: \|x_j(t) - x_i(t)\| < r_d^i(t); l_i(t) < l_j(t)\}$  表示第  $i$  只萤火虫  $t$  时刻的邻居集合,  $\|x\|$  表示  $x$  的范数。

位置更新阶段  $P_{ij}(t)$  根据轮盘法选择移动方向, 则  $x_i(t+1)$  计算公式如式(3):

$$x_i(t+1) = x_i(t) + s \left( \frac{x_j(t) - x_i(t)}{\|x_j(t) - x_i(t)\|} \right) \quad (3)$$

其中  $s$  为移动步长。

邻居范围更新阶段 更新规则如式(4):

$$r_d^i(t+1) = \min\{r_s, \max\{0, r_d^i(t) + \beta(n_t - |N_i(t)|)\}\} \quad (4)$$

其中:  $r_s$  为萤火虫感知范围,  $r_d^i(t)$  为第  $i$  只萤火虫  $t$  时刻的动态决策范围,  $\beta$  为邻域变化率,  $n_t$  为邻居阈值。

## 2 基于自适应 $t$ 分布混合变异的 AGSO 算法

本文对 AGSO 进行了改进, 提出了一种基于自适应  $t$  分布混合变异的人工萤火虫 (Adaptive  $t$  Distribution Mixed Mutation-Artificial Glowworm Swarm Optimization, ATM-AGSO) 算法。

### 2.1 自适应 $t$ 分布变异

$t$  分布又称学生分布, 含有参数自由度  $n$ , 当  $t(n \rightarrow \infty) \rightarrow N(0, 1)$ , 一般  $n \geq 30$  两者偏离可以忽略; 当  $t(n=1) = C(0, 1)$ , 其中:  $N(0, 1)$  为高斯分布,  $C(0, 1)$  为柯西分布。即标准高斯分布和柯西分布是  $t$  分布的两个边界特例分布<sup>[9]</sup>。

定义1 变异控制因子  $\eta$ 。 $\eta$  定义见式(5), 其表达变异发挥作用的大小程度:

$$\eta = 1 - \frac{i}{D-1}; i = 0, 1, 2, \dots, D-1 \quad (5)$$

其中  $D$  为算法的最大迭代次数。 $\eta$  在第一次迭代时值为 1, 在最后一次迭代时值为 0, 中间迭代时其值为 1 到 0 之间的递

减数值, 递减数值为一个等差数列, 公差为  $-1/(D-1)$ 。 $\eta$  值为 1 时表示变异发挥作用最大,  $\eta$  值为 0 时表示变异发挥作用最小。

定义2 自适应  $t$  分布变异。对人工萤火虫的状态  $X_i$  进行自适应  $t$  分布变异, 定义如下:

$$X_i^t = X_i + X_i \cdot \eta \cdot t(D) \quad (6)$$

其中:  $X_i^t$  为变异后的萤火虫个体状态,  $X_i$  为第  $i$  只萤火虫的状态,  $\eta$  为变异控制因子,  $t(D)$  为以算法的最大迭代次数  $D$  为参数自由度的  $t$  分布。式(6)中在  $X_i$  的基础上增加了  $t$  分布型随机干扰项  $\eta \cdot t(D)$ , 充分利用当前种群的信息干扰, 因此式(6)能使萤火虫跳出局部极值点的束缚收敛于全局极值点, 同时也提高了收敛速度。在变异控制因子  $\eta$  的作用下, 随着算法进化变异发挥的作用也逐渐减小, 直至为零。

自适应  $t$  分布变异使用算法的迭代次数  $D$  作为  $t$  分布的自由度参数。算法在运行初期,  $D$  值较小,  $t$  分布变异近似于柯西分布变异, 具有良好的全局探索性; 算法在运行后期,  $D$  值较大,  $t$  分布变异近似于高斯分布变异, 具有较优的局部开发性; 算法运行中期,  $t$  分布变异介于柯西分布变异和高斯变异之间。

### 2.2 最优调教变异

对于当前最优萤火虫, 它已经是趋向于最优解, 因此对它的变异的扰动能力应该较小, 而并非很强, 因此对最优萤火虫的变异选择扰动能力适中的高斯变异。本文对最优萤火虫进行最优调教变异。

定义3 最优调教变异。在最优萤火虫个体加上一个服从高斯分布的随机扰动项, 最优调教变异定义如下:

$$X_{\text{best}}^G = X_{\text{best}} + X_{\text{best}} \cdot \eta \cdot N(0, 1) \quad (7)$$

其中:  $X_{\text{best}}^G$  为变异后的萤火虫个体状态,  $X_{\text{best}}$  为当前最优萤火虫,  $\eta$  为变异控制因子,  $N(0, 1)$  为服从均值为 0 方差为 1 的高斯分布随机变量。

### 2.3 ATM-AGSO 自适应混合变异描述

在 ATM-AGSO 算法进化过程中, 当萤火虫陷入局部极值时, 将当前萤火虫群中最差萤火虫用公告板中的最优萤火虫取代得到了新的萤火虫群状态, 对萤火虫群中除最优萤火虫之外的非最优萤火虫进行自适应  $t$  分布变异, 对最优萤火虫进行最优调教变异, 对变异后的最优萤火虫和公告板上的最优萤火虫比较, 取二者最优的替换公告板。这样就增加了萤火虫的状态的多样性, 有利于萤火虫跳出局部极值点进行全局搜索, 同时也提高了搜索速度。在算法迭代过程中, 当公告板在连续 2 次迭代过程中没有改变或者变化很小时, 此时看作萤火虫寻优停滞, 则启动自适应  $t$  分布变异和最优调教变异, 综上这种变异方法称为自适应  $t$  分布混合变异。萤火虫群中, 只有少数几个最优萤火虫, 绝大部分为非最优萤火虫, 因此对非最优萤火虫进行的  $t$  分布变异在整个变异操作中起主要作用。

### 2.4 ATM-AGSO 算法步骤

ATM-AGSO 算法创新在于引入  $t$  分布混合变异, 该变异方法具有同时提高全局探索能力和局部开发能力的优点, 具有很强的自适应性。

ATM-AGSO 算法设计如下。

步骤1 参数  $\rho, \gamma, \beta, s, l_0, n_t, m, D, p$  初始化。

步骤2 公告板初始化, 对萤火虫目标函数值进行初始化。

步骤3 根据式(1)更新萤火虫的荧光素值。

步骤4 萤火虫运动阶段, 计算每只萤火虫的邻居萤火

虫集合,基于轮盘法对移动方向的萤火虫进行选择,根据式(3)完成位置更新,根据式(4)完成对决策域半径的更新。

步骤 5 比较当前萤火虫目标函数值和公告板中的值,取最优的赋给公告板。

步骤 6 若公告板在连续 2 次迭代过程中没有改变或者变化很小( $< p$ ),则转向步骤 7;否则转向步骤 8。

步骤 7 将当前萤火虫群中最差萤火虫用公告板中的最优萤火虫取代得到了新的萤火虫群状态。对新的萤火虫群中非最优萤火虫按式(6)进行自适应  $t$  分布变异;对最优萤火虫按式(7)进行最优调教变异,对变异后的最优萤火虫和公告板上的最优萤火虫比较,取二者最优的替换公告板。

步骤 8 终止条件判断:是否已达到的最大迭代次数  $D$ ,若不满足,转向步骤 3,进行下一代萤火虫群优化过程;否则转向步骤 9。

步骤 9 算法终止:输出公告板中萤火虫中的位置和目标函数值。

ATM-AGSO 算法流程如图 1 所示。

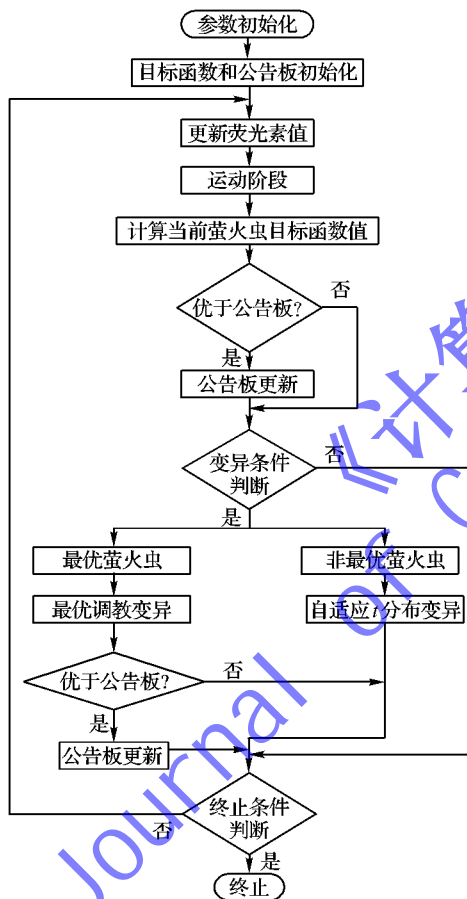


图 1 ATM-AGSO 算法流程

### 3 算例实验仿真及结果性能分析

选取文献[1]提出的基本人工萤火虫(AGSO)算法和文献[7]提出的基于自适应高斯变异的人工萤火虫(AGM-AGSO)算法与本文提出的基于自适应  $t$  分布混合变异的人工萤火虫(ATM-AGSO)算法进行性能对比,仿真实验在 Intel Core i3, 4GB 内存, Windows 7 操作系统的计算机上,采用 Matlab7.0 编程实现。

#### 3.1 算例仿真分析 1

为验证本文提出的 ATM-AGSO 算法性能,选取 4 个典型多峰函数<sup>[10-11]</sup>作为测试算例。

算例 1 SchafferF6 函数,具有强烈振荡的二维多峰函数,在(0,0)处,有最小值 0。

$$f_1 = 0.5 + \frac{\sin^2 \sqrt{x_1^2 + x_2^2} - 0.5}{[1.0 + 0.001 * (x_1^2 + x_2^2)]^2};$$

$$-100 \leq x_1, x_2 \leq 100$$

算例 2 Griewank 函数,高维多峰函数,在(0, ..., 0)处,有最小值 0。

$$f_2 = \frac{1}{4000} \sum_{i=1}^{30} x_i^2 - \prod_{i=1}^5 \cos \frac{x_i}{\sqrt{i}} + 1; -600 \leq x_i \leq 600$$

算例 3 Rastrigin 函数,高维多峰函数,在(0, ..., 0)处,有最小值 0。

$$f_3 = \sum_{i=1}^{30} (x_i^2 - 10 \cos 2\pi x_i + 10); -5.12 \leq x_i \leq 5.12$$

算例 4 Ackley 函数,高维多峰函数,在(0, ..., 0)处,有最小值 0。

$$f_4 = \left[ -20 \exp \left( -0.2 \sqrt{\frac{1}{30} \sum_{i=1}^{30} x_i^2} \right) + 20 \right] - \left[ \exp \left( \frac{1}{30} \sum_{i=1}^{30} \cos 2\pi x_i \right) - \exp(1) \right];$$

$$-30 \leq x_i \leq 30$$

对上述函数测试, AGSO、AGM-AGSO 和 ATM-AGSO 参数设置如表 1。

表 1 实验参数设置

参数	数值	参数	数值
萤火虫规模 $n$	100	邻域阈值 $n_t$	5
最大迭代次数 $D$	180	初始荧光素值 $l_0$	5
控制荧光素值 $\rho$	0.4	移动步长 $s$	0.3
衡量函数值 $\gamma$	0.6	参数 $p$	$10^{-4}$
控制邻居变化范围 $\beta$	0.08		

图 2~5 为对上述 4 个函数每个算法各运行 50 次求得最优值平均值的常用对数值随迭代次数变化的进化曲线。综合的实验数据分析中不难看出:对于函数  $f_1$ 、 $f_2$ 、 $f_3$  和  $f_4$ , ATM-AGSO 在进化初期的收敛速度要明显快于 AGM-AGSO,在进化后期二者的收敛速度相近,整体收敛速度 ATM-AGSO 更快,AGSO 的收敛速度要远远低于 ATM-AGSO 和 AGM-AGSO。求解精度 ATM-AGSO 比 AGM-AGSO 对于函数  $f_1$  提高了 2 个数量级,对于函数  $f_2$  提高了 2 个数量级,对于函数  $f_3$  提高了 3 个数量级,对于函数  $f_4$  提高了 1 个数量级;ATM-AGSO 比 AGSO 对于函数  $f_1$  提高了 12 个数量级,对于函数  $f_2$  提高了 10 个数量级,对于函数  $f_3$  提高了 6 个数量级,对于函数  $f_4$  提高了 10 个数量级。故可知 ATM-AGSO 算法在求解精度、收敛速度和鲁棒性方面显示出一定的优势。

#### 3.2 算例仿真分析 2

为了进一步验证 ATM-AGSO 的有效性,采用 4 个经典的实际应用算例。每个算例均通过大量实验数据测试分析如下。

算例 1 采用文献[12]中的最小推力滚珠导轨优化为算例。该算例要求在外载荷一定时,推力最小,即运动最灵敏。优化数学模型描述如下:

$$\min T(\alpha, \beta, \varepsilon, p_0) = \frac{2k(K\varepsilon)^{1/2}}{\left[ \left( \frac{p_0 + p \cos \alpha}{\sin \beta} + \frac{p \sin \alpha}{2 \cos \beta} \right) \frac{1}{z} \right]^{1/2} \sin \beta}$$

$$\left[ 2p_0 z + p \left( \frac{\cos \alpha}{\sin \beta} + \frac{\sin \alpha}{\cos \beta} \right) \right]$$



s. t.  $30^{\circ} \leq \alpha \leq 90^{\circ}, 30^{\circ} \leq \beta \leq 60^{\circ}$   
 $0.3 \leq \varepsilon \leq 1, 40 \leq p_0 \leq 80$

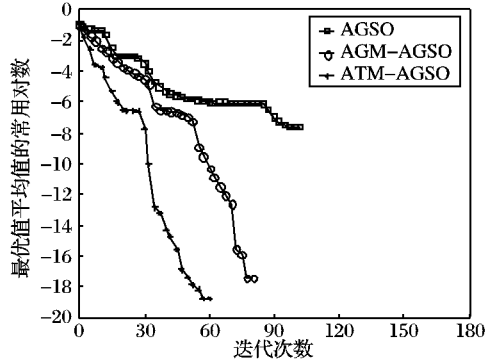


图2 SchafferF6 函数曲线寻优曲线对比

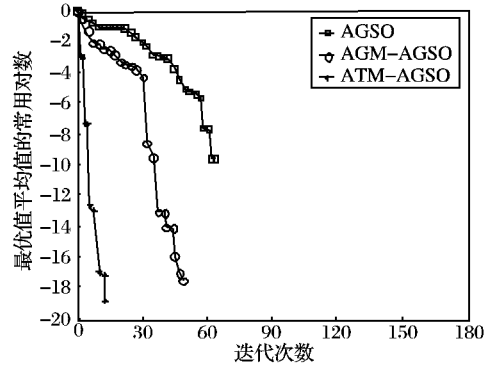


图3 Griewank 函数曲线寻优曲线对比

算例2 采用文献[13]中的神经网络权值优化为算例,该算例属于高维优化问题。数学模型为九点模式分类问题。

算例3 采用文献[14]中对非线性系统惯性时延模型参数估计,模型形式为:  $\frac{y(s)}{u(s)} = \frac{K}{Ts + 1} e^{-\tau s}$ , 其中待估计参数是比例系数  $K$ 、惯性系数  $T$  和时延系数  $\tau$ 。该算例测试存在噪声的情况。

算例4 采用文献[15]中对40Gb/s的光纤通信系统二阶

偏振色散(Polarization Mode Dispersion, PMD)补偿。

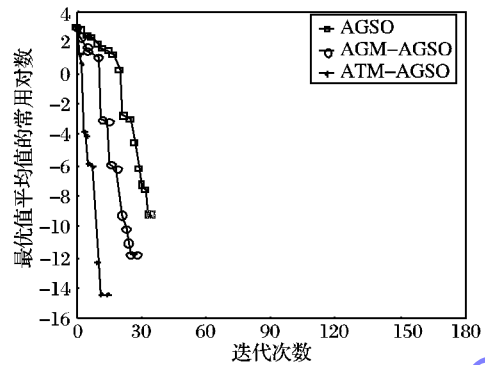


图4 Rastrigin 函数曲线寻优曲线对比

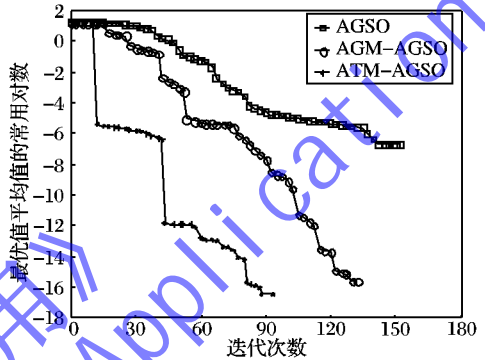


图5 Ackley 函数曲线寻优曲线对比

测试参数如下:算例1 采用文献[12]中的选定参数:  $p = 50 \text{ N}$ ,  $k = 0.01 \text{ mm}$ ,  $K = 0.5 \text{ N/mm}^2$ ,  $z = 4$ ;算例2 算法收敛约定为当误差小于0.56,优化次数为56次;算例3 中参数的真实值为  $K = 10$ ,  $T = 5$  和  $\tau = 9$ ,其他参数设置见文献[14];算例4 测试以光纤链路偏振度(Degree Of Polarization, DOP)作为适应度函数,理论上完全进行二阶 PMD 补偿后 DOP 值应为1。每个算例中 AGSO、AGM-AGSO 和 ATM-AGSO 三种算法的参数设置见表1,结果如表2所示。

表2 三种算法优化结果对比

算法	算例1					算例2		算例3			算例4
	$\alpha$	$\beta$	$\varepsilon$	$P_0$	优化结果	收敛率/%	耗时/s	$K$	$T$	$\tau$	DOP 均值
AGSO	36.2806	50.0007	0.4280	43.0060	0.7325	60	2.306	10.0016	5.0098	8.0705	0.873
AGM-AGSO	30.0000	58.8868	0.3000	40.9889	0.6825	100	1.290	10.0002	5.0028	8.9090	0.969
ATM-AGSO	30.0000	59.9906	0.3000	40.0020	0.6825	100	1.066	10.0001	5.0010	8.9960	0.982

由表2可以看出,在算例1中 ATM-AGSO 和 AGM-AGSO 的优化结果精度相近,但优化参数 ATM-AGSO 比 AGM-AGSO 略好;在算例2中 ATM-AGSO 和 AGM-AGSO 的收敛率相同,但 ATM-AGSO 比 AGM-AGSO 的收敛速度快;在算例3中 ATM-AGSO 比 AGM-AGSO 的参数估计值更接近真实值;在算例4中 ATM-AGSO 的优化补偿值 DOP 的平均值最接近理论值。而 AGSO 在四个算例中的性能都远远低于 ATM-AGSO 和 AGM-AGSO。故 ATM-AGSO 算法比 AGSO 和 AGM-AGSO 更稳定,精度更好,优化结果相当理想。

#### 4 结语

本文针对 AGSO 算法的不足,为了同时提高算法的全局探索能力和局部开发能力,结合高斯变异和柯西变异的各自优点,提出了一种基于自适应  $t$  分布混合变异的人工萤火虫算法,利用不同的函数对 AGSO、AGM-AGSO 和 ATM-AGSO 三个算法进行仿真对比实验性能分析,经过多个算例和大量的实验数据测试结果表明 ATM-AGSO 算法能有效摆脱局部极

值的束缚,同时也能加快搜索速度,较 AGSO 和 AGM-AGSO 求解精度更高,算法更稳定。下一步研究方向是将算法有效地应用到实际工程中。

#### 参考文献:

- [1] KRISHNANAND K N, GHOSE D. Theoretical foundations for rendezvous of glowworm-inspired agent swarms at multiple locations [J]. Robotics and Autonomous Systems, 2008, 56(7): 549 - 569.
- [2] 李咏梅,周永权,韦军.用于函数优化的层次结构萤火虫群算法[J].应用科学学报,2012,30(4):393 - 396.
- [3] GONG Q Q, ZHOU Y Q, YANG Y. Artificial glowworm swarm optimization algorithm for solving 0-1 knapsack problem [J]. Advanced Materials Research, 2011, 144: 166 - 171.
- [4] KRISHNANAND K N, GHOSE D. Chasing multiple mobile signal sources: a glowworm swarm optimization approach [C]// Proceedings of the Third Indian International Conference on Artificial Intelligence. Piscataway: IEEE Press, 2007: 1308 - 1327.

(下转第 1972 页)

4) RS-KMCE 算法采用了随机取样策略,选取具有多样性和正确性的基聚类子集融合,融合结果最优。该算法与 BA-KMCE 相比,增加了基聚类正确性的检测,因此其选择性聚类融合结果较优。对于 2 个仿真数据库和 Segmentation 数据库,RS-KMCE 的改善性能较为显著。

5) 参数  $por$  的取值依赖于具体的数据集。但是总体上来看,参数  $por$  较小时,基聚类个数较少造成一定信息损失,融合结果较差。随着参数  $por$  的增大,聚类融合性能不断改进。当参数  $por$  接近 1 时,聚类融合的性能下降,这是由于选择性聚类融合退化成普通聚类融合所导致的。图 3 所示为两组仿真数据随着参数  $por$  在  $[0,1]$  取值的变化而得到的聚类融合结果示意图。

6) RS-KMCE 对所有数据库的聚类性能均有提高,但是提高的程度依赖于具体的数据库。仿真数据的性能改进较为明显,而 UCI 数据库是真实数据,数据分布复杂,因此其改善不如仿真数据库明显。对于 Wine 数据库 BA-KMCE 和 RS-KMCE 的聚类融合性能相当,而对于 Segmentation,RS-KMCE 比 BA-KMCE 的性能高出 3%。

#### 4 结语

本文提出了基于随机取样的选择性聚类融合 RS-KMCE。其中,随机取样策略可以避免算法陷入局部极小;而且依据多样性和正确性的综合评价,有利于算法快速收敛到同时具有多样性和正确性的聚类融合。通过实验证明 RS-KMCE 的性能优于 KMC、KMCE 和 BA-KMCE。本文所提出的选择性聚类融合采用了  $K$ -means 算法作为基聚类算法,但是本文所提出的选择性聚类融合不受基聚类算法类型的限制。未来将研究其他类型基聚类算法的选择性聚类融合。而且在式(4)中,参数  $\lambda$  用于调整正确性和多样性在综合评判中的重要程度,该参数的取值依赖于具体的数据集。未来将深入研究该参数的取值与具体数据集之间的关系,得出参数  $\lambda$  取值与数据集之间的理论指导。

#### 参考文献:

- [1] STREHL A, GHOSH J. Cluster ensembles - a knowledge reuse framework for combining multiple partitions [J]. *Journal of Machine Learning Research*, 2002, 3: 583-617.
- [2] FERN X Z, BRODLEY E C. Cluster ensembles for high dimensional data clustering: an empirical study, #CS06-30-02[R]. Corvallis, USA: Oregon State University, 2004.
- [3] WANG T. CA-Tree: a hierarchical structure for efficient and scalable coassociation-based cluster ensembles [J]. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, 2011, 41(3): 686-698.
- [4] IAM-ON N, BOONGOEN T, GARRETT S, *et al.* A link-based approach to the cluster ensemble problem [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(12): 2396-2409.
- [5] FRED A L N, JAIN A K. Combining multiple clusterings using evidence accumulation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(6): 835-850.
- [6] GIONIS A, MANNILA H, TSAPARAS P. Clustering aggregation [C]// *ICDE '05: Proceeding of 2005 IEEE International Conference on Data Engineering*. Piscataway: IEEE, 2005: 341-352.
- [7] KUNCHEVA L I, VETROV D P. Evaluation of stability of  $k$ -means cluster ensembles with respect to random initialization [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(11): 1798-1808.
- [8] FISCHER B, BUHMANN J M. Bagging for path-based clustering [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(11): 1411-1415.
- [9] MINAEI-BIDGOLI B, TOPCHY A, PUNCH W. A comparison of resampling methods for clustering ensembles [C]// *Proceeding of 2004 International Conference on Machine Learning: Models, Technologies, and Applications*. Las Vegas: CSREA Press, 2004: 939-945.
- [10] FERN X Z, BRODLEY E C. Random projection for high dimensional data clustering: a cluster ensemble approach [C]// *ICML 2003: Proceedings of the 20th International Conference on Machine Learning*. Washington, DC: AAAI Press, 2003: 186-193.
- [11] TOPCHY A, JAIN A K, PUNCH W. Clustering ensembles: models of consensus and weak partitions [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(12): 1866-1881.
- [12] HONG Y, KWONG S, WANG H L, *et al.* Resampling-based selective clustering ensembles [J]. *Pattern Recognition Letters*, 2009, 30(3): 298-305.
- [13] ZHOU Z H, TAN W. Clusterer ensemble [J]. *Knowledge-Based Systems*, 2006, 19(1): 77-83.
- [14] TUMER K, AGOGINO A K. Ensemble clustering with voting active clusters [J]. *Pattern Recognition Letters*, 2008, 29(14): 1947-1953.
- [15] JIA J H, XIAO X, LIU B X, *et al.* Bagging-based spectral clustering ensemble selection [J]. *Pattern Recognition Letters*, 2011, 32(10): 1456-1467.
- [5] HUANG Z X, ZHOU Y Q. Using glowworm swarm optimization algorithm for clustering analysis [J]. *Journal of Convergence Information Technology*, 2011, 6(2): 78-85.
- [6] YANG Y, ZHOU Y Q. Glowworm swarm optimization algorithm for solving numerical integral [C]// *Communications in Computer and Information Science*. Berlin: Springer, 2011, 134: 389-394.
- [7] 莫愿斌, 刘付永, 张宇楠. 带高斯变异的人工萤火虫优化算法 [J]. *计算机应用研究*, 2013, 30(1): 121-123.
- [8] LAN K-T, LAN C-H. Notes on the distinction of Gaussian and Cauchy mutations [C]// *Proceedings of the Eighth International Conference on Intelligent Systems Design and Applications*. Piscataway: IEEE Press, 2008: 272-277.
- [9] 周方俊, 王向军, 张民. 基于  $t$  分布变异的进化规划 [J]. *电子学报*, 2008, 36(4): 121-123.
- [10] 刘佳昆, 周永权. 具有主从结构的并行人工萤火虫群优化算法 [J]. *计算机工程与应用*, 2012, 48(14): 33-37.
- [11] 刘洪霞, 周永权. 一种基于模式搜索算子的人工萤火虫优化算法 [J]. *小型微型计算机系统*, 2011, 32(10): 2130-2133.
- [12] 张梅凤, 邵斌, 甘勇, 等. 基于变异算子与模拟退火混合的人工鱼群优化算法 [J]. *电子学报*, 2006, 34(8): 1381-1385.
- [13] 杨爱红, 魏振方. 基于混合策略机制的人工鱼群算法 [J]. *中南林业科技大学学报*, 2011, 31(10): 193-197.
- [14] 曲良东, 何登旭, 黄勇. 一种新型的启发式人工鱼群算法 [J]. *计算工程*, 2011, 37(17): 140-142.
- [15] 许恒迎, 孙伟斌, 张霞, 等. 自适应视野和步长的局部领域人工鱼群算法 [J]. *计算机工程与设计*, 2012, 33(7): 2815-2821.

(上接第 1925 页)