

引用格式:齐欣,靳雁霞,张晋瑞,等. 大数据环境中简化粒子群算法的改进研究[J]. 微电子学与计算机,2020,37(2):25-29.
[QI X,JIN Y X,ZHANG J R,et al. Improvement of simplified particle swarm optimization in big data environment[J]. Microe-
lectronics & Computer,2020,37(2):25-29.]

大数据环境中简化粒子群算法的改进研究

齐 欣,靳雁霞,张晋瑞,程琦甫

(中北大学 大数据学院,山西 太原 030051)

摘 要:针对大数据存在的高维、强约束和多目标等复杂优化问题,本文提出一种改进的群智能优化算法——狮群简化粒子群算法(LSA-SPSO).该算法将狮群算法的分组思想融入简化粒子群优化算法中,将粒子分为三组寻优,每组使用不同的学习因子和学习维度向量,以此帮助种群执行不同的搜索机制,从而增强了种群的多样性.此外,引入种群育种,有利于粒子跳出局部最优位置,提高了算法的全局搜索性能.仿真实验表明,本文提出的改进算法有效改善了传统群智能算法中存在的不足,可以更好的应用到大数据中.

关键词:大数据;简化粒子群;狮群算法;分组;学习因子;学习维度向量;种群育种

中图分类号: TP18;TP391.41

文献标识码: A

文章编号: 1000-7180(2020)02-0025-05

Improvement of simplified particle swarm optimization in big data environment

QI Xin,JIN Yan-xia,ZHANG Jin-rui,CHENG Qi-fu

(College of Bigdata,North University of China, Taiyuan 030051, China)

Abstract: Aiming at the complex optimization problems such as high dimension, strong constraints and multi-objective in big data, an improved swarm intelligence optimization algorithm, Lions Simplified Particle Swarm Optimization(LSA-SPSO), is proposed in this paper. The algorithm integrates the grouping idea of Lions algorithm into the simplified particle swarm optimization algorithm, and divides the particles into three groups for optimization. Each group uses different learning factors and dimension vectors to help the population perform different search mechanisms, thus enhancing the diversity of the population. In addition, introducing population breeding can help the particles to jump out of the local optimum position and improve the global search performance of the algorithm. The simulation results show that the improved algorithm proposed in this paper effectively improves the shortcomings of traditional swarm intelligence algorithm, and can be better applied to big data.

Key words: big data; simplified particle swarm optimization; lions optimization; grouping; learning factor; learning dimension vector; population breeding

1 引言

粒子群优化算法(Particle Swarm Optimization,PSO)是由是由 Kennedy 和 Eberhart^[1]于 1995 年提出的一种群智能算法,其启发来源于鸟群的行

为规律,用来解决优化问题.由于 PSO 算法具有可
调参数少、结构简单、易于实现等优点,在各类优化
问题中得到了广泛的应用,如多目标优化,生物信号
识别,决策调度,机器人实时路径规划等领域^[2-3].

大量研究表明,基本 PSO 算法存在早熟收敛、

难以跳出局部最优值等缺点,近年来对粒子群优化算法的改进陆续被提出.研究者虽然在改进粒子群优化算法的研究中取得了一定的成绩,改善了粒子的寻优能力,但是相关算法的改进策略具有一定的复杂性^[4-7].本文提出了狮群简化粒子群算法(LSA-SPSO).将狮群算法的分组思想融入简化粒子群结构中,根据适应值高低及狮群比例将每代粒子分为三组,即雄狮、雌狮和幼狮.每组使用不同的学习因子和学习维度向量,帮助种群执行不同的搜索机制,从而增强了种群的多样性.此外,借鉴达尔文适者生存的竞争法则,引入种群育种,删除每次迭代中适应值最低的粒子,再加入新粒子,有利于粒子跳出局部最优位置,提高了算法的全局搜索性能.通过对标准测试函数进行仿真实验,结果表明,与其他改进算法相比,本文提出的狮群简化粒子群算法(LSA-SPSO)收敛速度更快,求解精度更高.

2 相关工作

2.1 粒子群优化算法(PSO)

在 D 维搜索空间中,种群粒子数为 m ,将每个粒子视为优化问题的一个可行解,由预先设定的适应度函数计算粒子的适应值.种群经逐代搜索,最后找到全局最优解.粒子 i 的当前位置表示为向量 $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$,当前速度表示为向量 $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$. $p_i = (p_{i1}, p_{i2}, \dots, p_{id}, \dots, p_{iD})$ 是粒子自身经历的最优位置, $p_g = (p_{g1}, p_{g2}, \dots, p_{gd}, \dots, p_{gD})$ 是种群经历的最优位置.粒子的速度和位置的更新公式如下:

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

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

式中,下标 $i = 1, 2, \dots, m$; 下标 $d = 1, 2, \dots, D$; t 是当前迭代次数; ω 是惯性权重; c_1 和 c_2 是学习因子,也称加速度系数; $r_1, r_2 \sim U[0, 1]$ 的独立随机数.为了防止粒子在迭代过程中离开搜索空间,限定速度取值范围和位置取值范围分别是 $[v_{\min}, v_{\max}]$, $[x_{\min}, x_{\max}]$.

2.2 简化粒子群优化算法(PSPO)

文献[8]提出一种简化粒子群优化算法,在进化方程中去掉了速度项.位置更新如式(3)所示:

$$x_{id}^{t+1} = \omega x_{id}^t + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \quad (3)$$

式(3)粒子位置项由三部分组成:第一部分表示粒子当前位置对粒子位置更新的影响;第二部分为

“认知”部分,表示粒子自身经历的最优位置对粒子位置更新的影响;第三部分为“社会”部分,表示种群经历的最优位置对粒子位置更新的影响.

2.3 狮群算法(LSA)

狮群算法^[9]的思想是将每个食物源的位置视为优化问题的一个可行解,通过适应度函数判断解的质量.根据狮群的群体特点分组,采用不同的公式迭代进化最后找到全局最优解.

狮群的群体特点描述为:狮子一般群居生活,一个狮群由雌狮、幼狮以及雄狮组成.其中雄狮的地位最高数量最少,雌狮的地位次之数量最多,幼狮的地位最低数量较多.雄狮是狮群的核心,雌狮相对稳定,幼狮存活率较低,大多数活不过两岁,一方面取决于食物情况,同时还面临着被雄狮驱赶出狮群的问题.受此启发将狮群算法的分组思想融入简化粒子群优化算法中,提出一种改进的群智能优化算法.

3 狮群简化粒子群算法(LSA-SPSO)

3.1 LSA-SPSO 的参数更新策略

3.1.1 惯性权重动态调整策略^[5]

$$\omega_t = \left(\omega_{\max} - (\omega_{\max} - \omega_{\min}) * \frac{t}{T} \right) * rand \quad (4)$$

式中, t 是当前迭代次数; T 是最大迭代次数; $\omega_{\max}$ 、 $\omega_{\min}$ 分别是惯性权重的最大值、最小值; $rand$ 是 $[0, 1]$ 之间的随机数.随着迭代次数的增加,惯性权重呈上下波动但整体下降的趋势,平衡了种群局部搜索与全局搜索的性能,避免粒子陷入早熟收敛.

3.1.2 学习因子

学习因子,也称加速度因子,不同的搜索机制选择不同学习因子.根据影响粒子飞行方向程度的强弱,定义 c_h 、 c_m 、 c_l 分别是强学习因子、中学习因子、弱学习因子,计算公式如式(5):

$$\left. \begin{aligned} c_h &= \frac{0.2}{T} * t + 1.1 \\ c_m &= 0.2 * rand * \cos(t) + 0.7 \\ c_l &= -\frac{0.2}{T} * t + 0.3 \end{aligned} \right\} \quad (5)$$

式中, t 是当前迭代次数; T 是最大迭代次数; $rand$ 是 $[0, 1]$ 之间的随机数.不同的学习因子在迭代过程中值的变化情况如图 1 所示.

从图 1 可以看出,在给定范围内强学习因子趋于线性递增,而弱学习因子趋于线性递减,中学习因子呈较稳定的上下波动趋势.通过调整学习因子,可以调整粒子的飞行轨迹,有助于尽快找到全局最优位置.

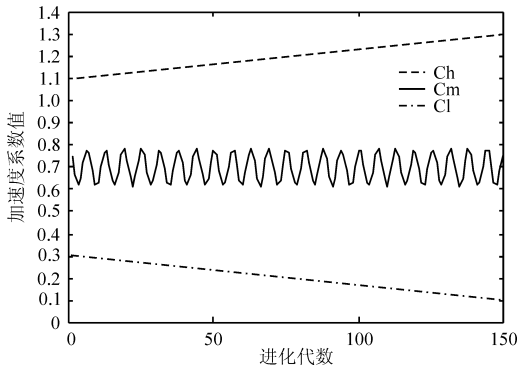


图1 学习因子变化曲线图

3.1.3 学习维度向量

学习维度向量 l_{di} 表示粒子从相应的向量中选择学习哪一维, d 表示变量个数. 具体来说, 用“yes”或者“no”分别表示粒子从向量中选择或者不选择相应的维度作为它的学习目标. 记录“yes”的个数, 将个数多的定义为高学习维度向量 l_{dh} , 表示粒子更多的向该向量学习; 个数少的为低学习维度向量 l_{dl} , 表示粒子较少的向该向量学习. 对于一个四维向量 p_{id} , 若 $l_{di} = [\text{yes}, \text{no}, \text{yes}, \text{no}]$, 表示粒子将从 p_{id} 的第一、第三维度上学习.

3.2 LSA-SPSO 的位置更新策略

在 LSA-SPSO 算法迭代中, 每个粒子按照其适应值的高低, 根据升序规则排序, 再根据狮群比例将种群分为三组, 把适应值最高的一部分粒子定义为雄狮, 适应值最低的一部分粒子定义为幼狮, 剩余的粒子定义为雌狮.

3.2.1 幼狮

在狮群中, 生存能力最低的幼狮会被雄狮驱赶出狮群, 即在每次迭代中删除种群适应值最低的粒子. 剩余的幼狮粒子普遍位置较差、适应值较低, 它们可能会朝着错误的方向移动. 对于这些粒子, 应放弃当前的个体信息, 从全局最优中获得更多的信息. 幼狮位置更新如下式.

$$x_{id}^{t+1} = \omega_i x_{id}^t + c_l r_1 l_{dl} (p_{id} - x_{id}) + c_h r_2 l_{dh} (p_{gd} - x_{id}) \quad (6)$$

式中, t 是当前迭代次数; ω_i 是动态变化的惯性权重; r_1, r_2 是 $[0, 1]$ 之间的随机数; c_l, c_h 是不同的学习因子; l_{dl}, l_{dh} 是不同的学习维度向量.

3.2.2 雌狮

雌狮相对稳定. 在 SPSO 算法中位置项包括三部分: 带惯性权重项、向粒子自身经历的最优位置学习项、向种群经历的最优位置学习项. 定义雌狮中粒子的最优位置为最优雌狮, 雌狮位置项拓展为四部

分, 第四部分是“次社会”部分, 表示向最优雌狮学习项. 加入最优雌狮, 有利于粒子获取更多的信息来更新自身位置. 粒子将从最优雌狮, 即次社会学习部分学到更多的信息. 雌狮位置更新如式(7)所示.

$$x_{id}^{t+1} = \omega_i x_{id}^t + c_l r_1 l_{dl} (p_{id} - x_{id}) + c_m r_2 l_{dm} (p_{gd} - x_{id}) + c_h r_3 l_{dh} (p_{ld} - x_{id}) \quad (7)$$

式中, r_3 是 $[0, 1]$ 之间的随机数; c_h, c_m, c_l 是不同的学习因子; l_{dh}, l_{dm}, l_{dl} 是不同的学习维度向量; p_{id}, p_{gd}, p_{ld} 分别表示粒子第 d 维的个体最优位置、全局最优位置、雌狮最优位置.

3.2.3 雄狮

雄狮是整个狮群的核心, 通常在连续迭代中这部分粒子的适应值已经得到了相当大的改善. 表示粒子朝着一个好的方向移动, 并且不管社会信息如何它都保持着当前的趋势, 也就是说粒子将更多地关注个体信息. 雄狮位置更新如式(8).

$$x_{id}^{t+1} = \omega_i x_{id}^t + c_h r_1 l_{dh} (p_{id} - x_{id}) + c_l r_2 l_{dl} (p_{gd} - x_{id}) \quad (8)$$

式中, c_h, c_l 是不同的学习因子; l_{dh}, l_{dl} 是不同的学习维度向量.

3.3 种群育种

(1) 在每次迭代中删除种群适应值最低的粒子, 该粒子的当前位置较差, 不利于尽快找到全局最优位置. 为了提高算法的搜索效率, 将它删除.

(2) 在下一迭代前, 向种群添加一个新粒子, 新粒子的初始位置是由个体最优值更新最频繁的两个粒子的位置组合而决定的. 添加新粒子后, 每个粒子按照其适应值的高低, 根据升序规则重新排序, 而后进入下一次迭代, 有利于粒子跳出局部最优位置, 提高了算法的全局搜索性能.

3.4 LSA-SPSO 算法流程

(1) 随机初始化种群中每个粒子的位置, 设定相应的参数.

(2) 根据适应度函数计算每个粒子的适应值, 并按照升序规则排序.

(3) 根据粒子的适应值, 更新粒子的个体最优值和种群的全局最优值.

(4) 根据狮群比例将粒子群分为三组, 根据式(4)计算惯性权重, 根据式(5)选择学习因子, 初始化学习维度向量.

① 幼狮: 粒子的位置按式(6)更新.

② 雌狮: 选出最优雌狮, 粒子的位置按式(7)更新.

③ 雄狮: 粒子的位置按式(8)更新.

(5)种群育种

(6)达到事先预定的最大迭代次数后,退出循环,同时输出全局最优值. 否则返回步骤 2).

4 仿真实验及分析

4.1 测试函数

为了验证本文改进算法的性能,将 LSA-SPSO 算法与 PSO,SPSO 和 LSA 算法进行测试比较. 在仿真实验中采用 4 个典型基准测试函数进行对比分析. 其中 Sphere 函数是单峰函数,主要用于检验算法的收敛速度和求解精度,Griewank 函数、Rastrigin 函数、Ackley 函数是多峰函数,主要用于检验算法的全局搜索能力.

(1)Sphere 函数表达式及搜索范围:

$$f_1(x) = \sum_{i=1}^n x_i^2 \quad (-100 \leq x_i \leq 100)$$

在 $x = \{0, \cdots 0\}$ 处取得全局最优值 $f_1(x) = 0$.

(2)Griewank 函数表达式及搜索范围:

$$f_2(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$$

$(-600 \leq x_i \leq 600)$
在 $x = \{0, \cdots 0\}$ 处取得全局最优值 $f_2(x) = 0$.

(3)Rastrigin 函数表达式及搜索范围:

$$f_3(x) = \sum_{i=1}^n (x_i^2 - 10\cos(2\pi x_i) + 10)$$

$(-30 \leq x_i \leq 30)$
在 $x = \{0, \cdots 0\}$ 处取得全局最优值 $f_3(x) = 0$.

(4)Ackley 函数表达式及搜索范围:

$$f_4(x) = -20\exp\left(-0.2 \times \sqrt{\frac{1}{30} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{30} \sum_{i=1}^n \cos 2\pi x_i\right) + 20 + e$$

$(-32 \leq x_i \leq 32)$
在 $x = \{0, \cdots 0\}$ 处取得全局最优值 $f_4(x) = 0$.

4.2 参数设置

在实验中,测试函数的维数 $D=30$,最大迭代次数 $\text{genmax}=150$,种群粒子数 $m=30$. 对于 PSO 算法和 SPSO 算法, $w=0.8, c_1=c_2=2$. 对于 LSA 算法和 LSA-SPSO 算法, $w_{\max}=0.9, w_{\min}=0.4$,种群按 1:6:3 的狮群比例分为三组,其中雄狮个数为 3,雌狮个数为 18,幼狮个数为 9. l_{dh} 中 yes 个数等于

$0.8 * D, l_{\text{dh}}$ 中 yes 个数等于 $0.1 * D$. 当进化次数大于最大迭代次数 genmax 时,停止迭代并输出实验结果.

4.3 实验测试结果

针对每个测试函数各个算法在指定收敛精度 10^{-5} 下独立运行 20 次并求其平均迭代次数. 各个算法平均迭代次数的比较见表 1. 其中,成功率=达到精度的实验次数÷总实验次数;期望迭代次数=种群粒子数×(平均迭代次数÷成功率);“∞”则表示次数是无穷大.

表 1 在目标精度下 20 次独立实验的迭代次数

	实验代码	平均值	最小值	最大值	成功率	期望迭代次数
f_1	PSO	150	150	150	0	∞
	SPSO	43	26	47	1	1 290
	LSA	14	12	17	1	420
	LSA-SPSO	13	6	27	1	390
f_2	PSO	150	150	150	0	∞
	SPSO	45	45	54	1	1 350
	LSA	15	12	19	1	450
	LSA-SPSO	12	6	24	1	360
f_3	PSO	150	150	150	0	∞
	SPSO	48	45	72	1	1 440
	LSA	22	20	34	1	660
	LSA-SPSO	19	18	23	1	570
f_4	PSO	150	150	150	0	∞
	SPSO	76	33	150	0.8	2 850
	LSA	26	20	38	1	780
	LSA-SPSO	22	19	35	1	660

由表 1 可以看出,对于单峰函数 Sphere 函数,PSO 算法没有达到目标精度,SPSO 算法达到目标精度的平均迭代次数为 43,LSA 和 LSA-SPSO 算法平均迭代次数为 14 和 13,其中,LSA-SPSO 算法的收敛速度更快,求解精度更高. 对于多峰函数 Griewank 函数,PSO 算法明显陷入局部最优点难以跳出,SPSO 算法平均迭代次数较高为 45,算法跳出局部最优点的能力不强,可以看出 LSA 算法和 LSA-SPSO 算法平均迭代次数为 15 和 12,表明 LSA-SPSO 算法可以更有效的跳出局部最优点,使粒子在其他区域继续寻优. 对于多峰函数 Rastrigin 函数,PSO 算法很难在大量的局部最优点在正弦拐点的情况下找到全局最优的位置,SPSO 算法收敛速度较慢,LSA 算法和 LSA-SPSO 算法分别是 22

和19,说明两者都有较好的平衡局部搜索与全局搜索的能力,可以快速收敛到目标精度,相较之下LSA-SPSO更能展现出更优的全局搜索性能.对于复杂多峰函数Ackley函数,存在许多局部最小点,PSO算法无法达到目标精度,SPSO算法稳定性较差达到目标精度成功率为80%,LSA算法和LSA-SPSO算法的平均迭代次数为26和22,都达到了目标精度且成功率为100%,均有效解决了粒子陷入局部最优的问题,提高了种群的全局搜索能力.总体来说LSA-SPSO算法平均迭代次数最低,均在12~22次之间,说明改进的LSA-SPSO算法能够更快的收敛到目标精度,具有更好的实用性.

不论是在寻找单峰函数还是多峰函数的最优值问题上,实验表明基本PSO算法由于全局搜索能力不强,陷入局部最优点导致算法过早停滞.而SPSO算法收敛速度较慢,在解决复杂多峰函数时稳定性不强,不利于搜索全局最优值.而LSA算法、LSA-SPSO算法在50次迭代后,4个测试函数均可以达到全局最优值或者达到最优值附近可接受的范围. LSA-SPSO算法曲线最先达到收敛,说明与其他算法相比,改进LSA-SPSO算法可以有效避免粒子陷入局部最优点,有更好的收敛速度和更高的求解精度,全局搜索性能更优.

5 结束语

本文提出了一种狮群简化粒子群算法(LSA-SPSO).一方面,将狮群算法的分组思想融入简化粒子群优化算法中,将粒子分为三组寻优,每组使用不同的学习因子和目标维度向量,使种群在迭代过程中有所波动但相对稳定,从而增强了种群的多样性.另一方面,借鉴达尔文适者生存的竞争法则,引入种群育种,提高了算法的全局搜索性能.仿真实验结果表明:相比于PSO、SPSO、和LSA算法,LSA-SPSO算法的寻优性能得到了明显改善,收敛速度和求解精度都得到提高.提出的改进算法可以更好地应用

到不同群智能算法中,提高数据处理能力的同时将有效地解决大数据复杂的优化问题,在数据挖掘技术方面产生重要的影响.

参考文献:

- [1] 王金永. 基于粒子群优化算法的聚类分析研究[D]. 青岛:青岛理工大学,2015.
- [2] 王宏杰,师彦文. 结合初始中心优化和特征加权的K-Means聚类算法[J]. 计算机科学,2017,44(11A):467-469.
- [3] 李武,赵娇燕,严太山. 基于平均差异度优选初始聚类中心的改进K-均值聚类算法[J]. 控制与决策,2017,32(4):759-762.
- [4] 陈朝威,常冬霞. 基于密度差分的自动聚类算法[J]. 软件学报,2018,29(4):43-52.
- [5] KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proceedings of ICNN95-International Conference on Neural Networks. IEEE, 2002.
- [6] 李斌,王劲松,黄玮. 一种大数据环境下的新聚类算法[J]. 计算机科学,2015,42(12):247-250.
- [7] 许明杰,蔚承建,沈航. 基于Spark的并行K-means算法研究[J]. 微电子学与计算机,2018,35(5):101-105.
- [8] 胡旺,李志蜀. 一种更简化而高效的粒子群优化算法[J]. 软件学报,2007,18(4):861-868.
- [9] PERAM T, VEERAMACHANENI K, MOHAN C K. Fitness-distance-ratio based particle swarm optimization [C]// Swarm Intelligence Symposium. IEEE, 2012.

作者简介:

齐欣 女,(1994-),硕士研究生. 研究方向为智能优化算法. E-mail:448463832@qq.com.

靳雁霞 女,(1973-),博士,副教授. 研究方向为智能优化算法、虚拟现实技术.

张晋瑞 男,(1994-),硕士研究生. 研究方向为虚拟仿真.

程琦甫 男,(1994-),硕士研究生. 研究方向为虚拟仿真.