

多目标优化算法 NSGA-II 的改进

刘旭红 刘玉树 张国英 阎光伟

(北京理工大学计算机科学与工程系, 北京 100081)

摘要 该文提出了 NSGA-II 算法的一种改进算法—INSGA。在引入算术交叉算子的同时,提出并引入累积排序适应度赋值策略。实验表明,INSGA 具有更高的收敛速度和更好的种群多样性。

关键词 多目标进化算法 Pareto 前端 NSGA-II 算法

文章编号 1002-8331-(2005)15-0073-03 文献标识码 A 中图分类号 TP301

Improvement of Multi-objective Optimization Algorithm NSGA-II

Liu Xuhong Liu Yushu Zhang Guoying Yan Guangwei

(Dept. of Computer Science and Engineering, Beijing Institute of Technology, Beijing 100081)

Abstract : An improved version of NSGA-II, INSGA, is proposed. INSGA adopts arithmetic crossover operator. Meanwhile, a new accumulated rank fitness assignment strategy is proposed. As experiments demonstrate, the results of INSGA have higher convergence speed and better population diversity than those of NSGA.

Keywords : multi-objective optimization, Pareto front, NSGA-II

1 引言

由于多目标进化算法可以在一次运行中得到多个 Pareto 优化解,近年来,在多目标优化领域已经成为一个研究热点,出现了许多优秀的算法,取得了较好的效果^[1]。其中非支配排序算法 NSGA-II (Non-dominated Sorting Genetic Algorithm II) 是具有代表性的算法^[2,3]。NSGA-II 是 NSGA^[4] 算法的改进,在 NSGA 的基础上加上了精英策略、密度值估计策略和快速非支配排序策略,在很大程度上改善了 NSGA 的缺点。实验证明 NSGA-II 的结果优于有代表性的其他几种算法^[2]。但 NSGA-II 采用的 SBX (Simulated Binary Crossover)^[5] 交叉算子搜索性能相对较弱,所使用的个体 Pareto 排序值方法有时不能很好地反映个体周围的密度信息,从而在一定程度上限制了算法的搜索性能,使得 NSGA-II 在种群的多样性保持和收敛速度方面尚不能令人满意。

在 NSGA-II 算法的基础上,针对 NSGA-II 存在的问题,该文提出一种改进的 NSGA-II 算法——INSGA 算法 (Improved Non-dominated Sorting Genetic Algorithm),在引入算术交叉算子的同时,提出并引入累积排序适应度赋值策略。最后,在收敛速度和种群多样性保持方面进行了实验验证。

2 INSGA 算法

2.1 交叉算子

NSGA-II 中采用 SBX 交叉算子。SBX 算子模拟二进制交叉算子的过程,对实数编码的父个体进行交叉操作,即对于给定的随机交叉点,交换两个父个体位于交叉点两侧的部分。

该文将算术交叉算子^[6,7]引入 NSGA-II。设 X_i^t 和 X_j^t 分别为第 t 代两个个体交叉点处对应的决策变量的真实值编码,则交叉后两个个体的相应的决策变量值为:

$$X_i^{t+1} = aX_i^t + bX_j^t \quad (1)$$

其中 a 和 b 为 $[-0.5, 1.5]$ 上均匀分布的随机数,且 $a+b=1$ 。将 a 和 b 不仅仅限于 $[0, 1]$ 区间,可以保证该交叉算子的搜索区

域覆盖 X_i^t 和 X_j^t 的所有邻域,且二者之间的区域搜索几率较大。显然,该算术交叉算子比 SBX 具有更好的全局搜索能力,能更好地保持种群的多样性。

2.2 累积排序适应度赋值策略

NSGA-II 采用的 Pareto 排序策略是:当前群体中不被任何其他个体支配的个体是非支配个体,其 Pareto 排序值为 1,全部非支配个体的集合是第一级非支配个体集;从当前群体中将这个体去掉,新产生的非支配个体的 Pareto 排序值为 2,组成的集合为第二级非支配个体集。依次类推,直到所有的个体的 Pareto 排序值确定为止。以 $r(y, t)$ 表示的 t 代中的个体 y 的 Pareto 排序值。

这种赋值方法的一个缺点是,个体的 Pareto 排序值有时不能很好反映个体周围的密度信息,如图 1 所示,尽管个体 a 周围种群的密度大于个体 b 周围的种群密度,但它们的 Pareto 排序值都为 2。尽管 NSGA-II 算法中有密度信息估计的部分,但所采用的密度信息的估计仅限于同一级非支配个体集中^[2],对于图 1 中的个体 a 和 b 仍然具有同样的机会繁殖后代。

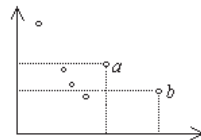


图1 个体周围种群密度的影响

该文提出的累积排序适应度赋值策略同时考虑个体的 Pareto 排序值和密度信息。首先,类似于 NSGA-II 对所有的个体进行 Pareto 排序,得到每一个个体的 Pareto 排序值。设在第 t 代种群中支配个体 y 的个体集为 y_1, y_2, \dots, y_n , 则个体 y 累积排序值定义为支配个体 y 的所有个体的 Pareto 排序值的和,如式 (2) 所示:

$$\text{rank}(y, t) = 1 + \sum_{i=1}^n r(y_i, t) \quad (2)$$

由此,对于图1中的个体 a 和 b 的累积排序值分别为4和2。采用累积排序适应度赋值策略并没有增加算法的复杂度,只需对NSGA-II中的快速非支配排序算法做少许改动即可,如下所示,其中黑体对应的两行是改进后的算法。

fast-non-dominated-sort(P)

for each $p \in P$

$S_p = \phi$ $n_p = 0$ $p_{rank} = 1$;

for each $q \in P$

if ($p < q$) then $S_p = S_p \cup \{q\}$;

else if $q < p$ then $n_p = n_p + 1$;

if $n_p = 0$ then

$p_r = 1$;

$F_i = F_i \cup \{p\}$;

$i = i + 1$;

while $F_i \neq \phi$

$F_{i+1} = \phi$;

for each $p \in F_i$

for each $q \in S_p$

$n_q = n_q - 1$ $q_{rank} = q_{rank} + p_r$;

if $n_q = 0$ then

$q_r = i + 1$;

$F_{i+1} = F_{i+1} \cup \{q\}$;

$i = i + 1$;

其中,每一个个体 p 对应四个数据结构:(1)个体集 S_p 记录 p 支配的所有个体;(2) n_p 记录优于 p 的个体数目;(3) p_r 记录 p 的Pareto排序值;(4) p_{rank} 记录 p 的累积排序值。

F_i 记录第 i 级非支配个体集。

首先,找到群体中所有 $n_i = 0$ 的个体,将其存入第一级非支配个体集 F_1 ,这些个体的Pareto排序值和累积排序值都为1。然后,对于 $F_i (i \geq 1)$ 中的每个个体 p ,考察它所支配的个体集 S_p ,将 S_p 中的每个个体 q 的支配它的个体数量 n_q 减1,并计算个体 q 的累积排序值 q_{rank} 。如果支配 q 的个体数量 $n_q = 0$,则将 q 存入下一级非支配集 F_{i+1} 。直到所有的个体排序完毕,算法中止。

2.3 INSGA 算法过程

随机产生一个规模为 N 的初始种群 P_0 ,将种群中的所有个体快速非支配排序。采用选择、交叉遗传算子产生一个规模为 N 的子代种群 Q_0 。其中,选择算子主要根据累积排序值评价个体的优劣,选择累积排序值小的个体参与繁殖。将 P_0 和 Q_0 合并为一个规模为 $2N$ 的种群 R_0 ,对 R_0 进行非支配排序得到非支配个体集(F_1, F_2, \dots),选择前 $i (i$ 满足式(3))个非支配集和

F_{i+1} 的前 $N - \sum_{j=1}^i |F_j|$ 个个体组成 P_{i+1} 。

$$\sum_{j=1}^i |F_j| \leq N, \text{ 且 } \sum_{j=1}^{i+1} |F_j| > N \quad (3)$$

再由 P_i 经选择、交叉产生 Q_i ,将 P_i 和 Q_i 合并为 R_i 。重复上面的循环,直到满足停止条件。从第一代开始主流程如下:

$R_i = P_i \cup Q_i$;

(F_1, F_2, \dots) = fast-non-dominated-sort(R_i);

$P_{i+1} = \phi$ $i = 1$

while $|P_{i+1}| + |F_i| \leq N$

crowding-distance-assignment(F_i);

$P_{i+1} = P_{i+1} \cup F_i$;

$i = i + 1$;

sort(F_i, F_{i+1});

$P_{i+1} = P_{i+1} \cup F_i$, 其中 $1 \leq i \leq N - |P_{i+1}|$;

$Q_{i+1} = \text{make-new-pop}(P_{i+1})$;

$t = t + 1$ $i = i + 1$;

几点说明:

(1)子程序crowding-distance-assignment和sort分别对某一个非支配个体集进行拥挤距离赋值和拥挤距离排序,具体过程可以参考文献[2],在此不再赘述;

(2)子程序make-new-pop中的选择算子使用联赛选择,个体的优劣主要由个体的累积排序适应度值确定,累积排序适应度值小的个体优。对于具有相同累积排序适应度值的个体,再进行拥挤距离的比较,拥挤距离小的个体优。

3 实验结果

该文采用INSGA对文献[2]所列举的两个典型的测试函数进行了计算,并与由NSGA-II计算得到的结果进行比较(两个测试函数均为最小化问题)。

$$F_1 = \begin{cases} f_1 = 1 - \exp[-\sum_{i=1}^3 (x_i - \frac{1}{\sqrt{3}})^2] \\ f_2 = 1 - \exp[-\sum_{i=1}^3 (x_i + \frac{1}{\sqrt{3}})^2] \end{cases}$$

其中 $-4 \leq x_1, x_2, x_3 \leq 4$ 。

F_1 的最优解为 $x_1 = x_2 = x_3 \in [-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}]$ 。

$$F_2 = \begin{cases} f_1(x) = \sum_{i=1}^{n-1} [-10 \exp(-0.2 \sqrt{x_i^2 + x_{i+1}^2})] \\ f_2(x) = \sum_{i=1}^n |x_i|^{0.8} + 5 \sin(x_i)^3 \end{cases}$$

其中 $-5 \leq x_1, x_2, x_3 \leq 5$ 。

对于两个测试问题,两种算法所采用的相同的设置为:实数编码方式,种群规模为100,联赛选择,联赛规模为6,交叉概率为0.8,进化代数50代。

图2是实验结果。(a)是NSGA-II对于函数 F_1 的运行结果,(b)是INSGA对于 F_1 的运行结果,(c)是NSGA-II对于函数 F_2 的运行结果,(d)是INSGA对于 F_2 的运行结果。由实验可以看出,采用INSGA得到的Pareto曲线分布更加均匀,且解的精确度更高。

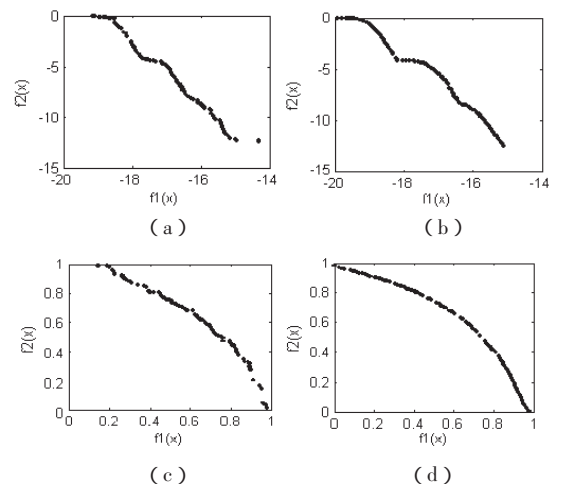


图2 NSGA-II和INSGA的实验结果比较

4 结论

该文使用算术交叉算子代替 NSGA-II 算法中的 SBX 交叉算子,提高了算法的搜索性能;同时,提出了累积排序适应度赋值策略,更好地维持了种群的多样性。在两个典型的多目标优化测试函数上进行了实验,对照结果可以证明 INSGA 比 NSGA-II 具有更快的收敛速度和更好的种群多样性。

(收稿日期:2004年8月)

参考文献

- 1.谢涛,陈火旺,康立山.多目标优化的演化算法[J].计算机学报,2003;26(8):997~1003
- 2.Kalyanmoy Deb,Amrit Pratap,T Meyerivian.A fast and elitist multi-

(上接19页)

较好地解决加速收敛和停滞早熟的矛盾,对算法性能的改善比较明显。算法的时间复杂度分析如下:算法初始化的复杂度为 $O(NK)$;L0层包含 R 个蚂蚁构造解和计算公式(1)(2),其复杂度为 $O(R2NKn)$;L1层局部搜索的复杂度为 $O(2LNKn)$;L2层信息素更新的复杂度为 $O(NK)$ 。因此,在最坏情况下整个算法的复杂度为 $O(M_{\text{converg}}RNKn)$ 。 M_{converg} 是达到收敛条件的演化代数。可以看到,算法的时间复杂度受收敛情况影响很大,且随问题规模的扩大,算法的复杂度呈线性上升。

5 总结

该文在分析蚁群优化算法的多 Agent 结构的基础上,提出了一种自适应蚁群优化聚类算法。该算法分为三个层次,每一层均由多个 Agent(即蚂蚁)组成。L0层是解构造层,L1层是可行解改进层(算法采用局部搜索),L2层是信息素更新层。更新后的信息素矩阵作为蚂蚁一次演化结果的反馈,为下一轮搜索提供启发信息。算法选取解构造过程中的变异概率 p 以及信息素残留度 ρ 作为自适应参数,在蚁群演化过程中自动调节其值,达到加快收敛和全局搜索的最佳效果。

实验结果及分析验证了该算法的有效性。与 GA 和 SA 聚类算法相比较,该算法确实表现出更好的性能,聚类效果和运行效率均优于上述两种方法。与成熟的聚类算法相比较,在运行效率和处理大规模聚类问题上有待进一步提高。

(收稿日期:2005年3月)

参考文献

- 1.J Bilmes,A Vahdat,W Hsu.Empirical observations of probabilistic heuristics for the clustering problem[R].Technical Report TR-97-018, International Computer Science Institute,University of California,Berkeley,CA,1997
- 2.Sung C S,Jin H W.A tabu-search-based heuristic for clustering[J].Pattern Recognition,2000;33:949~958
- 3.Murthy C A,Chowdhury N.In search of optimal clusters using genetic algorithms[J].Pattern Recognition Letters,1996;17:825~832
- 4.Brown D E,Huntley C L.A practical application of simulated annealing[J].Pattern Recognition,1992;25(4):401~412
- 5.J L Deneubourg,S Goss,N Franks et al.The dynamics of collective sorting:Robot-like ants and ant-like robots[C].In J A Meyer,S Wilson (C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

objective genetic algorithm:NSGA-II[J].IEEE Transactions on Evolutionary Computation,2002;6(2):182~197

- 3.Kalyanmoy Deb.Multi-objective optimization using evolutionary algorithms[M].New York:John Wiley & Sons Press,2001:210~302
- 4.Deb K,Agrawal R B.Simulated binary crossover for continuous search space[J].Complex Systems,1995;9:115
- 5.N Srinivas,K Deb.Multiobjective function optimization using nondominated sorting genetic algorithms[J].Evol Comput,1995;3(3):221~248
- 6.朱学军,陈彤,薛量等.多个体参与交叉的 Pareto 多目标遗传算法[J].电子学报,2001;29(1):106~109
- 7.陈文平,康立山.基于多父体杂交的多目标演化优化算法[J].计算机工程与应用,2003;39(10):79~82
- eds.Proceedings of the First International Conference on Simulation of Adaptive Behavior:From Animals to Animates, Cambridge,MA:MIT Press,1991:356~363
- 6.E Lumbert,B Faieta.Diversity and adaption in populations of clustering ants[C].In J-A Meyer,S W Wilson eds.Proceeding of the Third International Conference on Simulation of Adaptive Behavior:From Animals to animates, Cambridge,MA:MIT Press/Bradford Books,1994;3:501~508
- 7.N Monmarché.On data clustering with artificial ants[C].In A Freitas ed.Data Mining with Evolutionary Algorithms,Research Directions-papers from the AAAI Workshop Menlo Park,CA:AAAI press,1999:23~26
- 8.吴斌,郑毅,傅伟鹏等.一种基于群体智能的客户行为分析算法[J].计算机学报,2003;26(8):913~918
- 9.吴斌,傅伟鹏,郑毅等.一种基于群体智能的 Web 文档聚类算法[J].计算机研究与发展,2002;39(11):1429~1435
- 10.Dorigo M,Maniezzo V,Colomi A.Ant System:Optimization by a Colony of Cooperating Agents[J].IEEE Trans on System,Man and Cybernetics,1996;26(1):29~41
- 11.Dorigo M,Bonabeau E,Theraulaz G.Ant algorithm and stigmergy[J].Future Generation Computer Systems,2000;16:851~871
- 12.Yang Xin-Bin,Sun Jing-Gao,Huang Dao.A new clustering method based on ant colony algorithm[C].In Proceeding of 4th World Congress on Intelligent Control and Automation,Shanghai,P R China,2002:2222~2226
- 13.Shelokar P S,Jayaraman V,K Kulkarni B D.An ant colony approach for clustering[J].Analytica Chimica Acta,2004;509:187~195
- 14.Nicolas Labroche,Nicolas Monmarché,Gilles Venturini.AntClust:Ant clustering and web usage mining[C].In GECCO 2003,LNCS 2723,Berlin Heidelberg:Springer-Verlag,2003:25~36
- 15.H Azzag,N Monmarché,M Slimane et al.A clustering algorithm based on the ants self-assembly behavior[C].In ECAL 2003,LNAI 2801,Berlin Heidelberg:Springer-Verlag,2003:564~571
- 16.M Milano,A Roli.MAGMA:A Multiagent Architecture for Metaheuristic[J].IEEE Trans on System,Man and Cybernetics,2004;33(2):925~941
- 17.Sergios Theodoridis,Konstantinos Koutroumbas.Pattern Recognition[M].Second Edition,Elsevier Science Academic Press,USA,2003
- 18.朱庆保,杨志军.基于变异和动态信息素更新的蚁群优化算法[J].软件学报,2004;15(2):185~192