基于多种群差分进化的多目标优化算法

宋 通 庄 毅

(南京航空航天大学计算机科学与技术学院 南京 210016)

摘 要 针对差分进化算法(Differential Evolution Algorithm,DE)求解多目标优化问题时易陷入局部最优的问题,设计了一种双向搜索机制,它通过对相反进化方向产生的两个子代个体进行评价,来增强DE算法的局部搜索能力;设计了多种群机制,它可令各子群独立进化一定次数再执行全局进化,以完成子群间进化信息的交流,这一方面降低了算法陷入局部最优的风险,另一方面增强了Pareto解集的多样性,使Pareto前沿面的解集分布更为均匀。实验结果表明,相比于NSGA-II等同类算法,所提方法在搜索Pareto最优解时效率更高,并且Pareto最优解集的精度及分布程度比前者更好。

关键词 差分进化,多目标优化,多种群

中图法分类号 TP301.6 **文献标识码** A

A Kind of Multi-objective Optimization Algorithm Based on Differential Evolution with Multi-population Mechanism

SONG Tong ZHUANG Yi

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract In order to avoid the situation of falling into local optimum in solving the multi-objective optimization problem(MOP) with differential evolution algorithm(DE), we designed a bidirectional search mechanism which can improve the ability of local search of the DE. We also designed a multi-population mechanism for DE, which can reduce the risk of local optimum, and make the Pareto fronts more evenly distributed. Experimental results shows that, compared with similar algorithms such as NSGA-II, the proposed method is more efficient, while the precision and distribution of Pareto optimal solution set is better than the former.

Keywords Differential evolution, Multi-objective optimization, Multi-population

1 引言

许多科学研究和工程应用领域涉及到多目标优化问题(Multi-objective Optimization Problem,MOP),诸如飞行器航迹规划、武器目标分配问题等。由于多个目标函数往往存在相互冲突,一个目标性能的提升通常会导致其他目标性能的降低,因此,多目标优化问题通常会产生一个解集,解集外不存在比解集内更好的解,称该解集为Pareto最优解集。进化算法作为一类启发式搜索算法,可同时搜索多个Pareto最优解,已被成功地应用于多目标优化领域。进化多目标优化(Evolutionary Multi-objective Optimization,EMO)[1]已成为近年来新的研究热点。典型的进化多目标优化算法有NSGA-II[12]、SPEA2[2]等。

Storn和Price最早提出差分进化算法(Differential Evolution,DE)[13]用以解决连续优化问题。DE的主要特性在于进化过程中的自适应寻优机制,其具有收敛迅速、自适应性强、可靠性高等优点。文献[3]将差分进化算法用于多目标优化(Differential Evolution Multi-objective Optimization,DEMO),验证了该方法的可行性与高效性,但由于DE本身存在解质量不高的缺陷,因此直接将其应用于多目标优化并不能得到高质量的全局最优解集。文献[4]将双种群搜索机制引入DE,并将其应用于求解约束多目标优化问题。文献[5]用混沌序列构造DE的初始种群以进行多目标优化,增强了DE的搜索能力,但由于DE本身收敛速度很快,因此在Pareto解集多样性及解集分布的广泛性方面,还有待提升。

本文提出一种基于多种群差分进化的多目标优化算法(Multi-Population DEMO,MP-DEMO),将多种群机制引入DEMO,避免了种群过早收敛,并可使Pareto最优解集分布更加均匀。KUR[6]和ZDT[7]测试函数的实验表明,MP-DEMO算法能够快速收敛到Pareto前沿面,并很好地保证了解集的均匀性。

2 相关知识回顾

2.1 多目标优化问题

不失一般性,以最小化问题为例,一个具有n个决策变量、m个目标函数变量的多目标优化问题用数学方法可简单

到稿日期:2011-10-11 返修日期:2012-03-03 本文受航空科学基金(2010ZC13012)资助。

宋 通(1987-),男,硕士生,主要研究方向为多目标优化,E-mail:icloudy@live.cn;庄 毅(1956-),女,教授,博士生导师,主要研究方向为网络安全、分布与并行计算等。

表示为[8]：

$$\begin{cases}\min\ y=F(x)=(f_1(x),f_2(x),\cdots,f_m(x))^{\mathrm{T}}\\ \text{s.t. } g_i(x)\leqslant 0,i=1,2,\cdots,q\\ \quad h_j(x)=0,j=1,2,\cdots,p\end{cases}\tag{1}$$

式中，$x=(x_1,\cdots,x_n)\in X$ 为 n 维决策变量，$X\subset R^n$ 为 n 维决策空间；$y=(y_1,\cdots,y_m)\in Y$ 为各目标函数值构成的 m 维目标变量，$Y\subset R^m$ 为 m 维目标空间。$F(x)$ 定义了 m 个从决策空间到目标空间的目标函数映射；$g_i(x)\leqslant 0(i=1,2,\cdots,q)$ 针对决策变量 x，定义了 q 个不等式约束；$h_j(x)=0(j=1,2,\cdots,p)$ 则定义了 p 个等式约束。

实际应用中，不存在可以同时满足 m 个优化目标的解。因此，在缺乏用户经验及偏好信息的情况下，求解多目标优化问题必须兼顾所有目标函数，为用户提供 Pareto 最优解集。参考文献[11]，以最小化为例，下面给出 Pareto 最优解集的定义。

定义 1(支配)　解 $x^{(1)}$ 支配 $x^{(2)}$，记作 $x^{(1)}\preceq x^{(2)}$，当且仅当条件(1)和条件(2)同时满足：

(1)$x^{(1)}$ 的所有目标分量函数值都不比 $x^{(2)}$ 差：$\forall i=1,2,\cdots,m, f_i(x^{(1)})\leqslant f_i(x^{(2)})$；

(2)$x^{(1)}$ 至少有一个目标分量函数值比 $x^{(2)}$ 严格更优：$\exists f_i\in F$，使 $f_i(x^{(1)})<f_i(x^{(2)})$。

如果解 $x^{(1)}$ 不支配 $x^{(2)}$，则记作 $x^{(1)}\npreceq x^{(2)}$。对于有限解集，可以将所有的解进行互相比对，最终找到一个解集，其中任意两个解互不支配。对于不属于此解集的其他解，总是可被此解集中的某个解支配。换句话说，该解集中的解比解集外的要好。将此解集称为非支配解集。

定义 2(非支配解集)　非支配解集 $X'\subseteq X$，对于 $\forall x^{(1)}\in X'$，不存在 $x^{(2)}\in X$ 使得 $x^{(2)}\preceq x^{(1)}$ 成立。

当集合 X 为整个决策空间时，非支配解集 X' 称为 Pareto 最优解集。Pareto 最优解集通常处于目标空间可行域的边缘。

2.2　差分进化算法

DE 算法最早由 Storn 和 Price 提出，具有简单、可靠、高效等特点，尤其适合解决连续优化问题。与传统进化算法相比，DE 同样也基于种群进化寻优机制。其不同之处在于，DE 采用特殊的变异算子对个体进化方向进行干预扰动，同时采用一对一的选择机制产生子代个体。因此，变异和选择操作是控制 DE 进化过程的关键。

对于种群 P，父代个体 $p_i\in P$，其临时子代 p_i' 由以下变异算子产生，如式(2)所示[3]：

$$p_i'=\gamma\cdot p_{best}+(1-\gamma)p_i+F\cdot\sum_{k=1}^{K}(p_{i_a^k}-p_{i_b^k})\tag{2}$$

式中，$\gamma\in[0,1]$ 代表最佳个体对变异方向的影响程度权重，p_{best} 是父代最佳个体，F 是扰动向量权重，K 是扰动向量的个数，$p_{i_a^k}$ 和 $p_{i_b^k}$ 分别是从父代随机选取的除 p_{best} 之外的其他个体，k 为进化迭代次数，i 标识种群中某一个体，i_a^k 和 i_b^k 分别代表针对个体 i 选取的其他个体的标号。γ、K、F 与算法性能相关，其深入分析参见文献[9]。

从式(2)可以看出，DE 的变异算子由两部分构成，一般将 $\gamma\cdot p_{best}+(1-\gamma)p_i$ 称为差分向量，其后称为扰动向量。差分向量利用种群最佳个体信息，引导其他个体向最佳个体进化；扰动向量产生随机变异，为 DE 提供了自适应特性。

DE 在算法结构中包含了隐性的自适应过程，针对每一个候选解，根据父代个体质量调整进化方向。在优化过程前期，个体间较大的差异度导致扰动向量权重增加，搜索范围趋于延伸，可有效增加搜索范围。优化后期个体趋于相似，扰动向量减小，种群趋向于收敛。

DE 的选择算子采用贪心准则，临时子代优于父代则保留，否则淘汰[3]：

$$p_i^{(k+1)}=\begin{cases}p^{(k)'}, & \text{if } (p_i^{(k)'})>f(p_i^{(k)})\\ p^{(k)}, & \text{if } (p_i^{(k)'})\leqslant f(p_i^{(k)})\end{cases}\tag{3}$$

式中，k 为进化迭代次数，$p_i^{(k+1)}$ 为新产生的子代个体，$p_i^{(k)}$ 为父代个体，$p_i^{(k)'}$ 为临时子代个体，f 为适应值评价函数。

3　基于多种群差分进化的多目标优化算法

3.1　算法思想描述

DE 具有子代产生机制简单、收敛迅速、实现简便等特性，这些优点使得 DE 被广泛应用于各领域的优化问题[10]。尽管如此，DE 仍存在一些不足，最显著的缺陷在于其所得解的质量不高。多数情况下，DE 能够迅速收敛至全局最优解附近，却不能精确地收敛于该最优解；其次，DE 使用的选择算子虽然可以保证算法的快速收敛，却增加了陷入局部最优的可能[14]。

DE 的变异算子由差分向量和扰动向量构成。用 DE 求解单目标优化问题时，差分向量被定义为介于最佳个体及选定个体之间的向量。但是在多目标优化领域，进化算法的目的是找出 Pareto 解集，而非单一的最佳个体，这与单目标优化情况存在很大的差异。

本文设计了双向搜索机制以改善 DE 的搜索能力。根据双向随机优化理论[9](Bidirectional Random Optimization，BRO)在进行搜索寻优的过程中同时进行正向搜索和反向搜索。BRO 理论证明了，如果正向搜索所得的解不能达到优化期望，那么反向搜索则会以更高的概率使所得的解达到优化。基于该理论，DE 在扰动进化方向的过程中可以同时进行正向搜索和反向搜索，从而增强 DE 的局部搜索能力，加速收敛过程。同时，考虑到多目标优化的需求，寻优过程应尽量保证解集多样性以使 Pareto 解集分布更为均匀。借鉴蛙跳算法(Shuffled Frog Leaping，SFL)思想，本文引入多种群机制以提高进化过程中种群的多样性。通过多种群机制，一方面可保证各种群不会受其他种群进化的影响，从而保证了解的多样性；另一方面，各种群之间定期的信息交互可避免所得解陷入局部最优。

3.2　应用实例分析

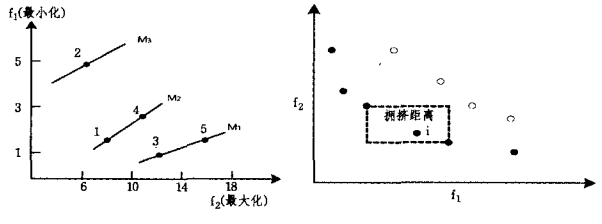
本文设计的多目标优化算法具有较强的实际应用意义。文献[16]利用多目标优化的方式构建了多基地多无人机协同侦察问题的数学模型，并利用传统进化算法对该模型进行求解，取得了较好的效果。由此可见，在某些需要频繁变更决策者、迅速产生决策的场合，本文算法的多目标特性可辅助不同的决策者做出与其期望目标相符合的策略；同时，由于该算法采用差分进化作为对解空间搜索寻优的方式，因此与传统进化算法相比，本文算法在求解效率方面表现更好[3]。

3.3　变异算子设计

本文利用非支配集执行多目标情况下最佳个体的选定操作。对个体 p_i 执行变异操作时，需检查该个体是否被支配，

如果该个体被支配，则可以确定该个体的支配集 D_i，对于 $\forall p_j \in D_i$，$p_i \preceq p_j$。最佳解 p_{best} 从 D_i 中随机选取。定义在 p_{best} 和 p_i 之间的向量为变异算子的差分向量。如果 p_i 已是非支配解，则令 $p_{best}=p_i$，在这种情况下，只有扰动向量起作用。变异算子中的扰动向量由该种群中随机选取的个体对构成。与单目标 DE 的主要区别在于，变异过程中的最佳个体总是随机变化的。

多目标优化一般采用非支配层(Non-dominated Fronts)等级和拥挤距离(Crowding Distance)对个体适应值进行评价[11]。其中，非支配层等级用来对不同非支配层内的个体的适应值进行评价；拥挤距离则用于对同一非支配层内不同个体的适应值进行评价。利用 NSGA-II((Non-dominated Sorting Genetic Algorithm II))中的非支配集排序算法[11]，可以将种群按支配关系划分为若干非支配层 $M=\{M_1, M_2, \cdots, M_r\}$。设种群通过文献[11]的非支配集排序算法已被划分为 r 个非支配层。其中，对于 $\forall M_i \in M$，$M_i \cup M_{i+1} \cdots \cup M_r$ 构成种群的一个子集，M_i 即为该子集的非支配解集。同一非支配层内的个体互不支配；不同非支配层间，$\forall p_s \in M_i$，$\forall p_t \in M_j$，当且仅当 $i<j$ 时，有 $p_s \preceq p_t$。对于不同非支配层来说，M_1 是当前种群非支配解集，即 Pareto 最优解集，其内任何个体都不被种群中其他个体支配，适应值最高；M_2 中的所有个体被 M_1 中的个体支配，适应值次之。依次类推，M_{i+1} 中的所有个体被 $M_1 \cup \cdots \cup M_i$ 中的个体支配，M_r 中的个体适应值最低。考虑到多样性，要求同一非支配层内的个体分布尽可能广泛。参考文献[11]，对同一非支配层内的不同个体，采用 NSGA-II 中的拥挤距离分配算法对其适应值进行评价。个体的拥挤距离越大，与同一非支配层内的其他最近个体差距越大，由此导致该非支配层的个体分布更广泛，有利于保持非支配解集的多样性，使个体适应值越高。综上所述，个体处于不同非支配层时，非支配层等级序号越小，其个体适应值越高；个体处于同一非支配层时，拥挤距离越大，其适应值越高。非支配层与拥挤距离的相关示例如图 1 所示[11]。



(a)非支配层等级　　(b)拥挤距离

图 1　非支配层等级与拥挤距离示例

在图 1(a)的示例中，种群个体经非支配集排序算法最终可确定 3 个非支配层，同一非支配层内个体互不支配，不同非支配层间个体存在支配关系。如个体 3、5 互不支配，个体 1、4 互不支配；个体 3 或 5 可以支配种群内其他个体，个体 1、4 可支配除个体 3、5 以外种群内其他个体，依次类推。图 1(b)虚线框表示经过拥挤距离算法分配的个体 i 的拥挤距离，直观上看，对于同一非支配层内的个体，拥挤距离越大，Pareto 最优解集分布越广泛。

在差分向量、扰动向量及个体适应值评价的基础上，本文设计了双向搜索机制，定义了多目标 DE 的变异算子，以改进 DE 的搜索能力。所谓双向搜索，是指在差分向量上叠加扰动向量时，同时在两个方向进行搜索的方法。如果朝某方向进化不能提高个体的适应值，那么朝相反方向进化则有更高的概率来增加个体适应值。其原理如图 2 所示。

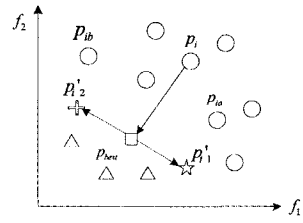


图 2　引入双向搜索机制的差分进化变异算子

在图 2 中，圆形个体及三角形个体共同构成了父代种群。其中，三角形个体集合表示支配 p_i 的解集 D_i；实线箭头表示个体 p_i 进行变异的差分向量，其指向的方形为临时个体，表示只有差分向量作用时 p_i 的进化结果；虚线箭头表示扰动向量，星形及十字形表示在差分向量的基础上进行个体进化方向扰动所产生的临时个体。通过差分向量与扰动向量的叠加，最终可以产生两个 p_i 的临时子代个体 p'_{i1} 和 p'_{i2}，其分别代表两个相反的扰动向量所能达成的进化目标。当变异算子产生的临时子代 p'_{i1} 适应值比 p_i 差时，将扰动向量以相反的方向进行探索，产生的临时子代 p'_{i2} 具有更高的概率，可以取得更好的适应值。因此，双向搜索机制可以增强 DE 的局部搜索能力，从而降低 DE 陷入局部最优的风险。式(4)给出变异算子产生方法：

$$p_i'=\begin{cases} C+E, & fit(C+E)>fit(p_i) \\ C-E, & fit(C+E)<fit(p_i) \\ P_i, & fit(C+E)=fit(p_i) \end{cases} \tag{4}$$

式中，$C=\gamma \cdot p_{best}+(1-\gamma)p_i$ 为式(2)中的差分向量，$E=F \cdot \sum_{k=1}^{K}(p_{i_a^k}-p_{i_b^k})$ 为式(2)中的扰动向量，fit 为指定个体的适应值。

3.4　多种群机制

借助 SFL(Shuffled Frog Leaping)[15]算法思想，本文为 DE 算法设计了多种群机制。SFL 的进化种群由若干蛙群组成，不同蛙群独立进化、互不干扰。进化时，蛙群内个体存在信息交互，而蛙群间个体互相独立。进化一定程度后，各蛙群重新混合为整体进行信息交互。SFL 算法具有高效的计算性能和良好的全局寻优能力，本文在 SFL 算法的基础上，为 DE 设计了多种群机制，其包括 3 个主要步骤：分割、进化和重组。

(1)将 DE 种群随机划分为 s 个子群，设置子群迭代上界为 $k_{shuffle}$；

(2)各子群独立使用 DE 引导搜索解空间，进行子群进化；

(3)各子群每进化 $k_{shuffle}$ 代时，将所有子群混合打乱并重新划分子群。

多种群机制中，各子群分别执行进化过程，不受其他种群进化过程的干扰，可在一定程度上保证 Pareto 解集的多样性；另一方面，子群进化过程中，通过定期混合打乱的方式进行进化信息交换，可大幅降低算法陷入局部最优的概率。

3.5　MP-DEMO 算法描述

多目标优化包含两种目的：收敛至 Pareto 最优解，并使 Pareto 解集分布更均匀。本文采用双向搜索机制避免 DE 陷入局部最优，可使种群收敛至全局 Pareto 最优；引入多种群机制维持种群多样性，使 Pareto 最优解集分布更为均匀。一

般而言，多目标进化算法的终止条件被设定为种群中所有解都互不支配，或迭代次数达到预设的上限。本文提出的基于多种群差分进化的多目标优化算法(MP-DEMO)描述如下。

Step1 随机生成大小为 N_{pop} 的初始种群 $P_{initial}$，设置子群迭代界限 $k_{shuffle}$；

Step2 将种群随机划分为 s 个子群，其分别包含 t 个个体，使得 $N_{pop}=s\times t$；

Step3 对于每个子群，令 $k=1$，当 $k<k_{shuffle}$ 时执行以下步骤：

Step3.1 按支配关系确定子群个体 p_i 在子群中的支配集 D_i；

Step3.2 当 $D_i\neq\varnothing$ 时，从 D_i 中选取随机个体作为最佳个体 p_{best}，否则令 $p_{best}=p_i$，构造差分向量 $C=\gamma\cdot p_{best}+(1-\gamma)p_i$；

Step3.3 从子群中随机选取 K 对个体，构造扰动向量 $E=F\cdot\sum_{k=1}^{K}(p_{i_a}^k-p_{i_b}^k)$；

Step3.4 生成临时子代 $p_i^{1'}=C+E$，$p_i^{2'}=C-E$，与原子群共同构成临时子群；

Step3.5 利用非支配集排序算法，确定临时子群中个体所处的非支配层，并使用拥挤距离分配算法为种群内个体计算拥挤距离；

Step3.6 根据式(4)确定 p_i 子代，构成子代子群，令 $k=k+1$。

Step4 将所有子群混合，检查是否满足终止条件，不满足转 Step2 继续迭代；

Step5 满足终止条件，算法终止，现有种群即所得 Pareto 最优解集。

4 实验结果及分析

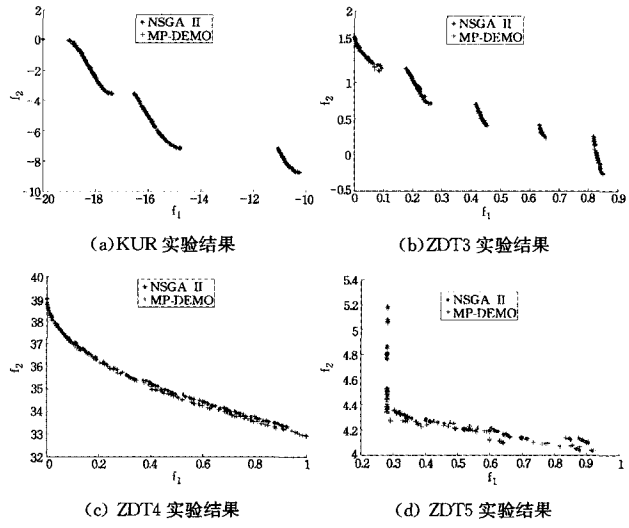
为了验证本文算法的可行性，选取文献[8]使用的 KUR 和 ZDT 共 4 个测试函数进行实验。其中，KUR 问题较简单，不可扩展，其决策变量数目为 3；ZDT3 问题决策变量为 30；ZDT4 和 ZDT5 决策变量为 10。测试函数均为双目标优化，函数优化目标均为最小化。测试函数及定义如表 1 所列[8]。

表 1 测试函数及定义

函数	定义
KUR	$f_1(x)=\sum_{i=1}^{2}[-10\exp(-0.2\sqrt{x_i^2+x_{i+1}^2})]$ $f_2(x)=\sum_{i=1}^{3}[\lvert x_i\rvert^{0.8}+5\sin(x_i^3)]$ $-5\leqslant x_i\leqslant 5, i=1,2,3$
ZDT3	$f_1(x)=x_1$ $f_2(x)=g\cdot h$ $g=1+9\cdot\sum_{i=2}^{m}x_i/(m-1)$ $h=1-\sqrt{f_1/g}-(f_1/g)\sin(10\pi f_1)$
ZDT4	$f_1(x)=x_1$ $f_2(x)=g\cdot h$ $g=1+10\cdot(m-1)+\sum_{i=2}^{m}(x_i^2-10\cdot\cos(4\pi x_i))$ $h=1-\sqrt{f_1/g}$
ZDT5	$f_1(x)=1-\exp(-4x_1)\cdot\sin^6(6\pi x_1)$ $f_2(x)=g\cdot h$ $g=1+9\cdot((\sum_{i=2}^{m}x_i)/(m-1))^{0.25}$ $h=1-(f_1/g)^2$

实验中，种群大小设置为 $N=100$，变异率 $p_m=0.3$，最大进化次数 $G_{max}=250$；差分进化参数设为 $K=2$，$\gamma=0.7$，$F=0.5$。为了方便对比，种群参数及差分进化参数与文献[10]的相同。

在图 3 中，星号代表 NSGA II 所得解，十字符号代表本文算法所得解。由于所有目标函数均需最小化，因此靠近原点的解效果更好。从图 3 可以看到，受益于双向搜索机制及多种群进化机制，本文算法比 NSGA-II 收敛性稍好，同时解的分布也更加均匀。



(a)KUR 实验结果 (b)ZDT3 实验结果

(c) ZDT4 实验结果 (d) ZDT5 实验结果

图 3 测试函数算法求解效果对比图

结束语 本文设计的双向搜索机制改善了 DE 的搜索性能，采用多种群机制优化 Pareto 解集分布，提升了算法收敛性及 Pareto 解集分布的均匀性。双向搜索机制的优点在于，同时朝前、后两个方向进行探索，提升了单次变异操作的搜索广度；多种群机制借鉴 SFL 算法思想，保留了进化过程解的多样性，可避免算法陷入局部最优。经典测试函数的数值仿真结果表明，本文算法在解集的逼近性及均匀性方面都优于 NSGA2-II 算法，因此提出的基于多种群差分进化的多目标优化算法(MP-DEMO)在求解多目标优化问题方面有一定的优势；同时，在某些需要频繁变更决策者、迅速产生决策的场合，本文算法也具有较强的实际应用意义。

参 考 文 献

[1] Coello A, Coello C, David A, et al. Lamont. Evolutionary Algorithms for Solving Multi-Objective Problems[M]. Boston: Academic Publishers, 2002: 594-604

[2] Zitzler E, Thiele L, Laumanns M, et al. Performance assessment of Multi-objective optimizers: an analysis and review[J]. IEEE Trans. Evol. Comput., 2003, 7: 117-132

[3] Babu B, Jehan M M L. Differential evolution for Multi-objective optimization: Congress on Evolutionary Computation (CEC 2003) [C]// Canberra. Australia: IEEE, 2003: 2696-2703

[4] 孟红云，张小华，刘三阳. 用于约束多目标优化问题的双群体差分进化算法[J]. 计算机学报，2008，31(2)：228-235

[5] 牛大鹏，王福利，何大阔，等. 多目标混沌差分进化算法[J]. 控制与决策，2009，24(3)：361-370

[6] Deb K. Multi-Objective genetic algorithms: Problem difficulties and construction of test problems[J]. Evolutionary Computation, 1997, 7(3): 205-230

[7] Zitzler E, Deb K, Thiele L. Comparison of multi-objective evolutionary algorithms: Empirical results[J]. IEEE Trans. Evol. Comput., 2000, 8(2): 173-195

[8] 公茂果，焦李成，杨咚咚，等. 进化多目标优化算法研究[J]. 软件学报，2009，12(2)：271-289

[9] Noman N, Iba H. Accelerating differential evolution using an a-

daptive local search[J]. IEEE Trans, Evolut. Comput., 2008, 12:107-125

[10] Neri F, Tirronen V. Recent advances in differential evolution: a survey and experimental analysis[J]. Artif Intell Rev, 2010, 33: 61-106

[11] Deb K. Multi-objective optimization[M]. Berlin: Springer, 2005: 273-316

[12] Deb K, Pratap A, Agarwal S, et al. A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197

[13] Storn R, Price K. Differential Evolution-A Simple and Efficient Heuristic for global Optimization over Continuous Spaces[J]. Journal of Global Optimization, 1997, 11(4): 341-359

[14] Brest J, Greiner S. Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems[J]. IEEE Transactions on Evolutionary Computation, 2006, 10(6): 646-657

[15] Fang C, Wang L. An effective shuffled frog-leaping algorithm for resource-constrained project scheduling problem[J]. Computers & Operations Research, 2011, 39(5): 890-901

[16] 田菁, 沈林成. 多基地多无人机协同侦察问题研究[J]. 航空学报, 2007, 28(4): 913-921

(上接第 198 页)

4.3 实验参数设置

实验一共采用了两种限制数，分别为 100、200。参数 $M=100, m=2, n=3, p=20, q=18$。且在计算 U 值时，每次对矩阵值加 0.5 进行数据处理，因为计算 U 值的公式中出现了 $\log S_{ij}$ 和 $\log(1-S_{ij})$，而在矩阵 S_{ij} 值为 0 或者是 1 时，log 运算会出错。本文实验采用的是十指交叉验证，实验次数为 100 次，结果为 100 次的平均值。

4.4 算法收敛性分析

本算法是基于结合限制的 K-means 算法的改进，通过聚类稳定性来寻找较优的分配次序，这种分配次序的修改一般不影响其收敛性。而 CKM 等一系列基于划分的限制 K-means 算法已经被文献[6]证明不收敛，本算法亦不收敛，为此，规定了最大迭代次数以保证算法的终止。

4.5 实验结果

采用了 UAILA+CKM 与 UALA+CKM，CKM 及 K-means 进行比较，100 和 200 限制数的实验结果分别如表 2 和表 3 所列。

表 2 算法准确率：限制数 100

Data Set	K-means	CKM	UAUL+CKM	UAILA+CKM
Iris	84.93%	89.78%	89.09%	**90.05%**
Wine	64.83%	71.20%	71.28%	**71.33%**
Lung Cancer	52.80%	58.77%	**61.33%**	61.00%
Soybean(small)	83.97%	92.73%	94.83%	**96.53%**
Zoo	84.68%	88.61%	86.25%	**86.93%**

表 3 算法准确率：限制数 200

Data Set	K-means	CKM	UAUL+CKM	UAILA+CKM
Iris	84.93%	89.95%	91.15%	**91.15%**
Wine	64.83%	71.49%	**71.65%**	71.40%
Lung Cancer	52.80%	58.67%	58.83%	**58.83%**
Soybean(small)	83.97%	93.33%	93.90%	**97.50%**
Zoo	84.68%	92.03%	89.64%	**91.22%**

从表 2 和表 3 可以看出，所有 CKM 算法的效果都比传统 K-means 算法的好，且 UAILA 的效果基本都不低于 UALA。

对于数据集 Iris，限制数为 100 时，UALA 准确率低于 CKM 算法，而 UAILA 的准确率都高于 UALA 算法和 CKM 算法。限制数为 200 时，UAILA 和 UALA 算法准确率都高于 CKM 算法，且 UAILA 也不低于 UALA 算法。

对于数据集 Wine、Lung Cancer 和 Soybean，限制数为 100 和 200，UAILA 和 UALA 准确率都高于 CKM 算法，除了数据集 Wine 在限制数为 200 和数据集 Lung Cancer 在限制数为 100 时，UAILA 比 UALA 低一点，其它都高于 UALA。而且数据集 Soybean 的准确率有大幅度的提高，限制数为 100 时，数据集 Soybean 的准确率提高了 1.7%；限制数为 200 时，准确率提高了 3.6%。

对于数据集 Zoo，限制数为 100 和 200 时，UAILA 的准确率都要高于 UALA，但这两种算法的准确率都低于 CKM 算法。这说明 UAILA 虽然比 UALA 更有效地提高了准确率，但还是有缺陷。

我们可以看到，具有样本少、数据高维性等特点的数据集 Lung Cancer 在限制数为 200 时的准确率都低于限制数为 100 时的准确率。因此对于高维数据的数据集的限制聚类，还有待进一步的研究。

从上面的实验结果可以知道，迭代排序比一次排序的聚类效果好，这也说明 CKM 算法中数据对象稳定性会随着限制的加入而改变。

结束语 虽然 CKM 算法较传统的 K-means 在准确率上得到了一定的提高，但是 CKM 算法具有分配次序的敏感性，而 UALA 对于学习一个好的分配次序起了一定作用。分配次序随着聚类过程中限制的逐渐加入而随之变化，但 UALA 采用无监督思想进行一次性确定来分配次序。针对 UALA 存在的问题，本文在 UALA+CKM 的基础上，采用迭代思想，利用 CKM 算法逐步确定稳定性，提出了一种基于分配次序聚类不稳定性迭代学习算法。上述实验结果表明，该算法有效地改善了聚类效果。

参 考 文 献

[1] Wagstaff K, Cardie C, Rogers S, et al. Constrained K-means clustering with background knowledge[C]// Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001). 2001: 577-584

[2] Hong Yi, Kwong S. Learning Assignment Order of Instances for Constrained K-means Clustering Algorithm[J]. IEEE Systems, Man and Cybernetics Society, 2009, 39(2): 568-574

[3] Jain A. Data clustering: 50 years beyond K-means[J]. Pattern Recognition Letters, 2010, 31: 651-666

[4] Jain A K, Murty M N, Flynn P J. Data Clustering: A Review [J]. ACM Computing Surveys, 1999, 31(3): 265-323

[5] 肖宇, 于剑. 基于近邻传播算法的半监督聚类[J]. 软件学报, 2008, 19(11): 2803-2813

[6] 何振峰, 熊范纶. 结合限制的分隔模型及 K-Means 算法[J]. 软件学报, 2005, 16(5): 799-809

[7] Wagstaff K, Cardie C. Clustering with Instance-level Constraints [C]// Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000). 2000: 1103-1110