

基于 IFI 与 FUA 的 Pareto 遗传算法

李少波, 杨观赐

(贵州大学 CAD/CIMS 工程技术中心, 贵阳 550003)

摘要: 在适应值快速辨识算法和基于聚类排挤的外部种群快速替换算法的基础上, 提出了搜索 Pareto 最优解集的快速遗传算法。在该算法中, IFI 算法实现个体适应值的快速辨识, FUA 维持种群多样性和 Pareto 最优解集的均匀分布性。采用 FPGA 算法对多种多目标 0/1 背包问题进行仿真优化, FPGA 算法能够以较少的计算成本搜索到高精度、分布均匀、高质量的 Pareto 非劣解集, 收敛速度和收敛准确性均优于强度 Pareto 进化算法(SPEA)。

关键词: 快速遗传算法; Pareto 最优性; 适应值快速辨识算法; 快速替换算法

Pareto Genetic Algorithm Based on IFI and FUA

LI Shao-bo, YANG Guan-ci

(Institute of CAD/CIMS Engineering, Guizhou University, Guiyang 550003)

【Abstract】 This paper proposes a fast Pareto genetic algorithm for searching pareto optimal solution set. It is based on a new approach for fast evaluation of fitness of individuals and a clustering based external population update scheme for maintaining population diversity and even distribution of Pareto solutions. Experiments on a set of multi-objective knapsack optimization problems shows that FPGA can obtain high-quality, well distributed non-dominated Pareto solutions with less computational efforts compared to other state-of-art algorithms, it has advantages in its convergence speed and quality over the state-of-the-art SPEA algorithm.

【Key words】 fast Pareto genetic algorithm(FPGA); Pareto optimality; fitness fast identify algorithm; fast update algorithm(FUA)

1 概述

在高维决策空间中, 搜索 Pareto 最优解集是一项十分复杂的任务, 传统的多目标优化方法通过评价函数将多个目标函数转化为单目标函数, 然后采用单目标优化技术, 但存在以下缺点:

(1) 构造单目标评价函数需要决策者事先提供深奥的偏好知识, 而在某些情况下无法提供这种偏好信息;

(2) 大多数单目标优化技术均是基于局部优化的搜索算法, 虽然能够找到单目标优化问题的局部或全局最优解, 但是不能并行地搜索到在 Pareto 最优解曲面上的全部或多个均匀分布的最优解, 不能满足多目标决策的灵活要求和某些动态变化的环境要求;

(3) 当需要可选的 Pareto 最优解时, 需要反复构造评价函数并运行搜索算法。

遗传算法通过遗传复制实现个体结构重组, 通过选择操作引导种群进化, 能够在单一种群中并行地处理一组解的集合, 文献[1,2]认为多目标搜索与优化是遗传算法最合适的应用领域之一。Zitzler^[3]提出了强度Pareto进化算法(SPEA), 并将其成功地应用于多目标 0/1 背包问题。该算法采用两个种群, 种群 P (规模为 N) 用于遗传复制, 外部种群 P' (规模为 N') 存储已找到的 Pareto 非劣解, 根据两个种群中个体的支配关系确定个体的 Pareto 适应值, 应用聚类分析技术维持种群多样性。SPEA 是真正意义上的 Pareto 最优解搜索算法, 然而, 适应值计算和聚类分析需要 $O(m(N+N')^3)$ 级的计算开销。文献[4]提出了一种有别于聚类算法的邻近排挤算法, 删除多余最优解, 使 Pareto 最优解分布均匀, 但仅仅是“均值和方差在一定程度上都优于算法 SPEA”; 文献[5]提出了一种基于正交设计的多目标演化算法以求解多目标优化问题, 但其实质上是小生境演化和分裂的迭代过程。其他实现小生境技术的

算法采用适应度共享技术^[6]来维持种群多样性, 这需要目标问题的先验知识来解决共享半径参数问题, 实现起来比较困难。

把遗传算法应用到多目标优化问题时有 3 个主要问题需要解决:

(1) 为了引导搜索向 Pareto 最优解集的方向进行, 如何有效地进行适应值赋值和选择。

(2) 为了防止早熟获得分布性和扩展性好的最优解集, 如何维持群体多样性。

(3) 如何在进化过程中防止获得的 Pareto 最优解丢失。

本文研究 Pareto 非劣解集的快速辨识和适应值赋值与种群多样性维持技术, 提出的算法将实现 $O(m(N+N')^2)$ 级的适应值计算时间复杂度和 $O(NN' \log N)$ 级的小生态算法时间复杂度。

笔者研究了保种群的多样性的可持续性进化的算法模型^[7]、Pareto 非劣解集的快速辨识和适应值赋值与种群多样性维持技术, 提出的算法将实现 $O(m(N+N')^2)$ 级的适应值计算时间复杂度和 $O(NN' \log N)$ 级的小生态算法时间复杂度。

2 Pareto 最优性概念

多目标优化在本质上是向量优化, 要想界定其解的概念, 必须先界定向量集的序关系。考虑最大化问题, 对任意的决策向量 x_1, x_2 , 当且仅当 $f(x_1) > f(x_2)$, 称 x_1 支配 x_2 , 记为 $x_1 \succ x_2$; 当且仅当 $f(x_1) \geq f(x_2)$, 称 x_1 弱支配 x_2 , 记为 $x_1 \succ = x_2$; 当且仅当

基金项目: 国家自然科学基金资助项目(50575047)

作者简介: 李少波(1973-), 男, 教授、博士, 主研方向: 智能系统, 进化算法; 杨观赐, 硕士研究生

收稿日期: 2007-03-13 **E-mail:** lishaobo@gzu.edu.cn

$\neg f(x_1) \geq f(x_2) \wedge \neg f(x_2) \geq f(x_1)$, 称 x_1 与 x_2 无关, 记为 $x_1 \sim x_2$ 。对任意的 x , 如果存在另外的 x' , 使得 $f_i(x) \leq f_i(x')$, ($i=1,2,\dots,m$), 其中, 至少有一个不等式严格成立, 称 x 为 Pareto 最优解^[1], 又称为 Pareto 非支配解、非劣解或有效解。对于一个多目标优化问题, 存在一组或无穷多个 Pareto 最优解, 全部 Pareto 最优解称为 Pareto 最优解集, 相应的目标向量构成 Pareto 最优面 (多维目标空间的超平面), 在不考虑偏好信息时, Pareto 最优解之间无优劣之分。

3 Pareto 遗传算法

3.1 算法结构

基于适应值快速辨识与外部种群替换的 Pareto 遗传算法 (fast pareto genetic algorithm, FPGA) 采用 2 个种群, 规模为 N 的种群 P 用于交叉和变异操作, 规模为 N' 的外部种群 P' 用于存储已找到的 Pareto 非劣解。两种种群均参与回放二联赛选择, 以实现定向进化和精英保留策略。对多目标 0/1 背包问题, 每个候选解采用长度为 m 位的二进制编码表示, 每个二进制位直接表示变量 $x_{ij} | i \in [1, m]$ 的值。FPGA 的流程如图 1, 形式化描述如下:

- (1) 用 Gauss 随机发生器随机初始化种群 P 和建立空外部种群 P' ;
- (2) 确定 P 的非劣个体 j ;
- (3) 用基于聚类排挤技术的快速替换算法 FUA 更新 P' ;
- (4) 用非劣个体快速分层算法 IFI 确定种群 $P+P'$ 的个体适应值;
- (5) 采用回放二联赛从 $P+P'$ 中选择 N 个个体充满交配池;
- (6) 应用交叉和变异算子生成种群 P ;
- (7) 如果达到规定进化代数, 结束, 否则转(2)。

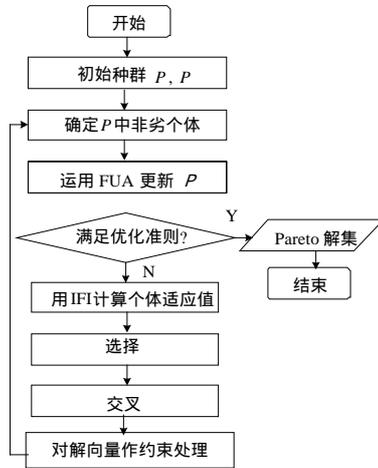


图 1 FPEA 的流程

本文采用 Gauss 随机发生器随机初始化分布均匀的种群 P , 对产生的解向量进行相似性分析, 尽可能产生比较多的模式, 为种群的进化与收敛创造基本条件; 在步骤(2)中, 确定 P 的非劣个体即是找出当前的 Pareto 最优曲线或者最优面; 在步骤(3)中, 非劣个体依次进入外部种群 P' , 在这个过程中, 运用 FUA 更新 P' 。

3.2 基于聚类排挤的快速替换算法(FUA)

共享技术^[6]的实现需要解决共享半径参数问题, 需要目标问题先验知识, 但这种先验知识对于很多工程问题是难以得到的, 同时, 这样获得的优化结果往往带有一定程度的相对性和主观性, 适应值共享方法的成功应用受到了制约; 聚类分析技术算法复杂, 计算成本高。FPGA 提出并实现基于聚

类排挤小生态技术的 FUA, 使外部种群中存储的 Pareto 非劣解集均匀地逼近 Pareto 最优面, FUA 的流程见图 2。伪代码如下:

```

for( $P$  中的每一个非劣个体  $j$ )
 $S = \Phi, flag = 1;$ 
for ( $k \in P' \wedge flag = 1$ )
if ( $j >= k$ )  $S = S \cup \{k\}, next k;$ 
else if ( $j < k \parallel j = k$ )  $flag = 0;$ 
if ( $flag = 0$ ) next  $j;$ 
else if ( $S \neq \Phi$ )  $P' = P' - S \cup \{j\}, next j;$ 
else if ( $|P'| < N'$ )  $P' = P' \cup \{j\};$ 
else 求出  $j$  与  $P'$  中所有个体的距离; 对所求距离进行快速排序;
依排序结果求出最相似的 2 个个体;
用  $j$  替换  $P'$  中最邻近的 2 个个体中; 离  $j$  较近的个体; next  $j;$ 

```

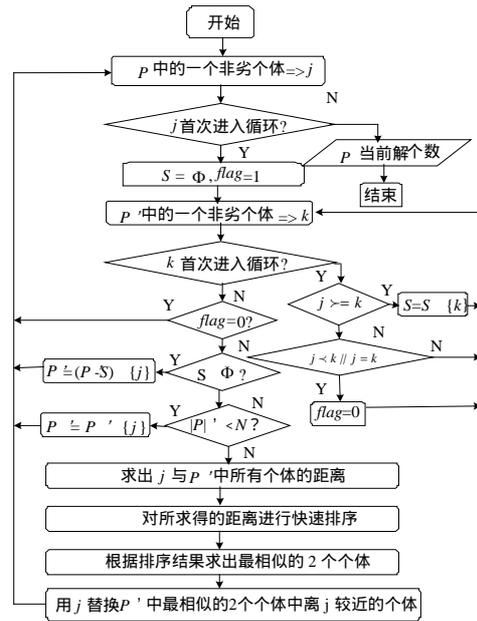


图 2 FUA 流程

FUA 的外层 for 语句遍历种群 P , 循环次数为 N 。内层 for 语句遍历种群 P' , 循环次数为 N' 。语句 $P' = P' - S \cup \{j\}$ 在最坏情况下循环次数也为 N' 。而最深层 else 中完成相似性替换的过程中仅有 $O(N' \log N')$ 的时间复杂度; FUA 在最坏情况下的时间复杂度为 $O(NN' \log N')$, 在最好情况下的时间复杂度仅为 $O(N)$ 数量级, 低于共享技术和聚类分析的计算成本。本算法的初衷是维持种群的多样性增大个体间的差异, 因此, 算法考虑的是目标空间中非劣解前沿的分布情况, 个体间距离采用欧几里德距离测度。

3.3 适应值快速辨识算法(IFI)

个体的适应值用 $P+P'$ 中非劣个体的层次序号表示, 第 k 层非劣个体的适应值为 k 。对任意的个体 $i \in P+P'$, 用 n_i 表示优于 i 的个体数量, 用 S_i 表示弱劣于 i 的个体子集, 用 F_k 表示第 k 层非劣个体子集。分层适应值快速计算过程伪代码描述如下:

```

初始化  $k=1;$ 
for each  $i \in P+P'$ 
 $S_i = \Phi, n_i = 0;$ 
for each  $j \in P+P'$ 
if ( $j \leq i$ ) then  $S_i = S_i \cup \{j\};$ 
else if ( $j > i$ ) then  $n_i = n_i + 1;$ 
next  $j;$ 

```

```

if ( $n_i = 0$ ) then  $F_k = F_k \cup \{i\}$ ;
next  $i$ ;
while ( $F_k \neq \Phi$ )
 $H = \Phi$ ;
for each  $i \in F_k$ 
for each  $j \in S_i$ 
 $n_i = n_i - 1$ ;
if ( $n_i = 0$ ) then  $H = H \cup \{j\}$ ;
next  $j$ ;
next  $i$ ;
 $k = k + 1$ ;
 $F_k = H$ ;

```

外层 for 语句确定 $N+N'$ 个 n_i 值和 $N+N'$ 个子集 S_i ，并确定第 1 层非劣解，完成该任务需要 $O(m(N+N')^2)$ 级的比较计算。while 语句确定余下的非劣解的分层。最坏情况下，每层仅包含一个非劣解，那么，确定 $N+N'-1$ 个分层共需要 $O((N+N')^2)$ 级的计算开销。因此，在最坏情况下，适应值辨识过程的总计算开销为 $O(m(N+N')^2) + O((N+N')^2)$ 或 $O(m(N+N')^2)$ 数量级。采用该适应值辨识过程，每一层有多个个体，它们被认为具有相同的竞争能力，那么在进化选择时，不容易出现某一个体适应值特别高而抑制适应值比它低的个体的情况，在一定程度上潜在维系种群的多样性，这一点在对比实验中将得到验证。

4 仿真对比实验

4.1 实验方案

考虑 9 个不同的多目标优化问题，即考虑背包数量分别为 $n=2, 3, 4$ ，物品数量分别为 $m=250, 500$ 和 750 的多目标 0/1 背包问题。背包 i 的容量规定为 c_i ，物品 j 相对于包 i 的重量为 w_{ij} ，效益值为 p_{ij} ， w_{ij} 与 p_{ij} 均初始化为区间 $[10, 100]$ 上的随机整数，包的容量公式为

$$c_i = 0.5 \sum_{j=1}^m w_{ij}$$

对所有的测试问题，种群规模依背包数量和物品数量取不同的值，见表 1。变异概率和交叉概率(一点交叉)分别取 0.01 和 0.8 的值。

由于个体编码的随机性及交叉与变异的随机性，可能产生不满足约束的解，因此本文采用简单的贪婪修复机制进行约束处理。对物品 j ，求 n 个效益/重量比的最大值 q_j ，令

$$q_j = \max_{i=1}^n \{p_{ij} / w_{ij}\}$$

m 种物品共有 m 个 q_j ， $j=1, \dots, m$ 值，对这 m 个 q_j 值按非降次序排序。

对不满足约束条件的编码按 q_j 值的非降次序依次将非 0 的二进制位清 0，即从背包中删除相应的物品，直至满足约束条件为止。这样的约束处理机制可确保在满足容量约束的同时，尽可能减少背包总效益值的损失。

4.2 实验结果与分析

采用运行结束时外部种群中非重复的 Pareto 非劣解的相对质量来测度和比较算法的性能。定义相对质量的公式为

$$R_{FPGA, SPEA} = \frac{|\{i \in P'_{FPGA}, \exists j \in P'_{SPEA} : i > j\}|}{|P'_{FPGA}|}$$

$$R_{SPEA, FPGA} = \frac{|\{j \in P'_{SPEA}, \exists i \in P'_{FPGA} : j > i\}|}{|P'_{SPEA}|}$$

对每个随机生成的 0/1 背包问题运行 FPGA 和 SPEA 各 40 次。每次独立运行，FPGA 和 SPEA 都在同一随机初始种群的基础上开始搜索，并经历 10 000 代的进化搜索。40 次运

行结束后，相对质量性能指标的算术平均值如表 1 所示。

表 1 种群规模参数和相对质量性能指标值

n	m	N'	N	$R_{FPGA, SPEA}$	$R_{SPEA, FPGA}$
2	250	30	120	0.834 97	0.047 50
	500	40	160	0.958 93	0.021 87
	750	50	200	0.959 69	0.018 00
3	250	40	160	0.996 63	0.000 63
	500	50	200	1.000 00	0
	750	75	225	1.000 00	0
4	250	50	200	1.000 00	0
	500	60	240	1.000 00	0
	750	70	280	1.000 00	0

对所有测试的 0/1 背包问题，FPGA 搜索到的 Pareto 非劣解的质量一致地优于 SPEA。

对背包数为 2、物品数为 750 的背包问题，随机采样一次运行结果，进化过程不同阶段外部种群非重复 Pareto 非劣解在目标空间的分布如图 3 所示。图 3 中的数据清晰地表明，随着进化代数的增加，外部种群中存储的 Pareto 非劣解均匀地逼近真实的 Pareto 最优面，且 FPGA 比 SPEA 表现了更均匀和更快的逼近趋势。

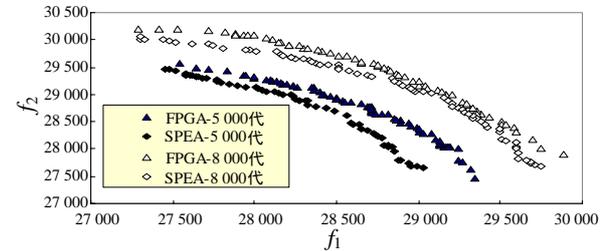


图 3 外部种群中 Pareto 非劣解的进化趋势与分布

从图 3 可以看出，外部种群中存储的 Pareto 非劣解近似地分布于目标空间中以坐标原点为球心的超球面上，某个解离原点的距离越远，解质量越高。

可采用某个层面上全体非劣解到原点的平均距离测度该层非劣解的平均质量，并根据平均距离的增长趋势衡量不同算法的收敛速度。对背包数分别为 2~4，物品数为 750 的背包问题，运行 FPGA 和 SPEA 各为 40 次，平均距离的算术平均值随进化代数变化的趋势如图 4 所示。

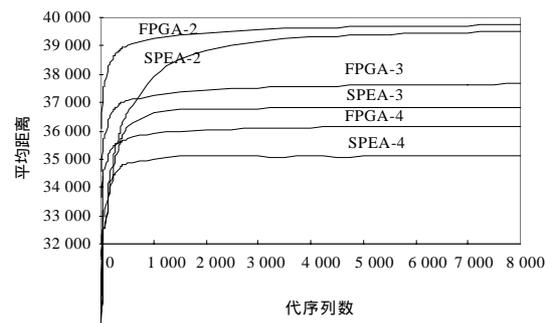


图 4 Pareto 非劣解到目标空间原点的平均距离增长趋势

图 4 中与 FPGA 对应的各趋势曲线具有更陡峭的上升斜率，这表明在搜索的初始阶段，FPGA 具有明显的收敛速度优势。当趋势曲线进入平缓上升阶段后，FPGA 各趋势曲线均分布于相应的 SPEA 趋势曲线上方，这表明在搜索的后期阶段，FPGA 可生成精度更高的 Pareto 非劣解集。因此，在收敛速度和收敛准确性 2 个方面，FPGA 一致地优于 SPEA。

文献[8,9]认为遗传漂移是导致种群收敛到单一个体和早

(下转第 192 页)