

## 基于信息熵的改进 PESA 算法

王堃<sup>1</sup>, 王琳琳<sup>1</sup>, 刘艳<sup>2</sup>, 张玉华<sup>1</sup>, 吴蒙<sup>1</sup>

(1. 南京邮电大学 宽带无线通信与传感网技术教育部重点实验室, 江苏 南京 210003;

2. 南京工业大学 自动化与电气工程学院, 江苏 南京 210009)

**摘要:**针对 PESA 算法所需的计算运算量、计算难度及运算时间都随着解集数量的增加而急剧增加的问题, 将熵值度量指标引入到 PESA 算法中, 提出了基于信息熵的 PESA 算法(C-PESA, comentropy-based PESA)。该算法根据信息熵指标在量化度量 Pareto 解集的分布特性, 判断种群进化是否到达成熟阶段, 本算法迭代 1 300 次时即到达成熟阶段, 从而尽早结束了算法复杂的优化过程, 在一定程度上简化了 PESA 算法的时间复杂度。仿真结果表明, 随着进化种群数量的增长, C-PESA 算法的计算量只是呈现线性增加, 算法的计算时间缩短接近 4 倍, 进化计算效率得到提高。

**关键词:**进化计算; PESA 算法; 多目标优化; 信息熵

中图分类号: TP18

文献标识码: A

文章编号: 1000-436X(2013)11-0033-09

## Improved PESA algorithm based on comentropy

WANG Kun<sup>1</sup>, WANG Lin-lin<sup>1</sup>, LIU Yan<sup>2</sup>, ZHANG Yu-hua<sup>1</sup>, WU Meng<sup>1</sup>

(1. Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education,

Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

2. College of Automation and Electrical Engineering, Nanjing University of Technology, Nanjing 210009, China)

**Abstract:** Aiming at the issue that the computational effort the complexity and the running time of PESA algorithm are increasing rapidly with the growth of the solutions set number, a comentropy-based PESA algorithm (C-PESA) by merging the entropy value metric into PESA algorithm was proposed. According to the distributed characteristic of the entropy value metric over the Pareto solution set, the proposed algorithm could determine whether the population has developed to the mature stage, which is reached when the number iterations is 1 300 in C-PESA. Thereby, the optimization process can be finished as soon as possible, and in a certain extent, the time complexity of PESA was simplified. Simulation results show that the computational effort of C-PESA increases linearly with the rising number of solutions. Meanwhile, the computation time is improved almost four times, and the evolutionary computation efficiency is also enhanced.

**Key words:** evolutionary computation; PESA algorithm; multi-objective optimization; comentropy

收稿日期: 2012-08-06; 修回日期: 2013-04-23

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目(2011CB302903); 国家自然科学基金资助项目(61100199, 61100213, 61170276, 61201160); 教育部高等学校博士学科点专项科研基金资助项目(20113223120007); 江苏省高校自然科学基金资助项目(10KJA510035, 12KJD510007); 南京邮电大学宽带无线通信与传感网技术教育部重点实验室开放研究基金资助项目(NYKL201107)

**Foundation Items:** The National Basic Research Program of China (973 Program) (2011CB302903); The National Natural Science Foundation of China (61100199, 61100213, 61170276, 61201160); The Specialized Fund for the Doctoral Program of Higher Education (20113223120007); The Program of Natural Science for Universities of Jiangsu Province(10KJA510035, 12KJD510007); Open Research Fund of Key Lab of Broadband Wireless Comm and Netw Technology (NUPT) of Ministry of Education (NYKL201107)

## 1 引言

最近 20 年来,进化算法已被成功应用于多目标优化领域,演变为进化多目标优化(EMO, evolutionary multi-objective optimization)<sup>[1,2]</sup>。从第一代进化多目标优化算法的典型代表矢量评价遗传算法(VEGA, vector-evaluated genetic algorithms)<sup>[3]</sup>,以及第二代进化多目标优化算法中的 PESA(pareto envelope-based selection algorithm)<sup>[4]</sup>,到当前基于多种新型进化机制进行多目标优化算法,都是通过不同的方法结合进化计算来更有效地求解多维多目标优化问题。

这其中,PESA 算法充分利用了精英保留策略<sup>[4]</sup>,设置了内、外部种群,寻找并删除外部种群中拥挤系数最大的个体。PESA 算法的经典之处在于,采用了网格方法(hyper-grid)将个体空间划分成若干个格子(hyper-box),这样每个个体就与某个网格相互关联,从而用选择网格代替个体选择(individual-based selection),利用网格法记录解集信息。网格法保持种群的多样性,为判断解点(solution)是否存在提供了方便,在一定程度上提高了算法的效率。然而针对 PESA 算法,其时间复杂度可以表示为  $O(N \times N^*)$ <sup>[5]</sup>,这里  $N$  和  $N^*$  分别为进化种群和外部种群的大小。根据 Big O notation<sup>[6]</sup>可以看出,随着解集数量的增加,PESA 算法进行进化计算所需的运算量、计算难度及运算时间都将急剧增加,这将给进化计算的进行带来极大麻烦。

作为一种同样采用网格法记录进化种群信息的度量指标,信息熵度量指标能够量化地评估多目标进化过程中解点的分布情况。在多目标进化算法中,非支配解点均匀分布于 Pareto 前端(pareto front),随着种群逐步进化到达成熟阶段,进化解集逐渐覆盖了整个前端。在此过程中,由于进化种群呈现增长状态,导致了信息熵数值的不断增长。因此,在信息熵饱和现象出现之前,分布情况在一定程度上显著提高。也就是说,进化种群接近所有可行域且非支配,从而沿着 Pareto 前端尽可能均匀地分布。当提及分布饱和性时,也就是进化种群达到成熟阶段,此时信息熵数值保持不变<sup>[7]</sup>。因此,通过对熵值持续计算,可以清晰地观察到解集的进化发展情况和成熟阶段。于是可以获知进化种群是否已经达到成熟,进而决定是否停止优化演算进程。尤为重要的是,信息熵度量指标在计算时间上具有极大优势。

随着解点数量的增长,信息熵值度量指标的计算量呈现线性增加。信息熵指标的计算难度增加不多,计算时间和计算量则随着解集数目的增长幅度均不大。

基于此,作者提出基于信息熵的改进 PESA(C-PESA, comentropy-based PESA),将信息熵度量指标引入到 PESA 算法中,利用信息熵指标在量化度量 Pareto 解集的分布特性,即均匀性、多样性和收敛性方面的突出表现,及时考量 PESA 算法进化后种群个体的进化程度。通过信息熵值指标的变化情况,观察进化解集从逐步进化发展过程到达成熟阶段的过程<sup>[9]</sup>,由此获知种群进化是否到达成熟阶段,以便尽早结束 PESA 算法的优化过程。当种群进化到种群分布均匀多样之后,及时结束 PESA 算法,避免大量不必要的复杂进化过程,这在一定程度上简化算法的时间复杂度,提高 PESA 进化算法的效率和性能。

## 2 相关工作

一般来说,(MOP, multi-objective optimization problem)<sup>[5]</sup>即为同时处理多个目标函数的最优化问题。在多目标优化问题中,由于有多个目标函数,因此存在如下问题:某个解可以使得其中某个目标实现最优,但可能同时使得其他目标最差,因此需要在多目标中进行折衷,寻找折衷解的集合,该集合称为 Pareto 最优解集(Pareto-optimal set)<sup>[8]</sup>。当现有解集中不再存在解的时候,解集被认为最优(optimal),也被称为 Pareto 最前端(Pareto-front)。解决一个多目标优化问题就是找到最佳的 Pareto 优化前端(Pareto-optimal front)。

而进化多目标优化算法就旨在解决一系列多目标优化问题。解集中解点的分布可以代表进化种群的多样性<sup>[9]</sup>,也就是说,在目标空间中解点分布范围越大,进化种群多样性就越强。总的来说,多目标进化算法的实质就是找到一个理想解集,其中每个解点都紧靠最优解前端且均匀分布在最优平面上<sup>[10]</sup>。

目前,进化多目标优化的算法主要经历 3 个阶段:第一阶段主要采用个体选择方法和种群多样性保持策略。第二代进化多目标优化算法主要以精英保留机制为特征。目前,也就是第三阶段,一些新的进化机制也被引入进化多目标优化领域。

在多目标优化领域引入新型进化范例是当前

进化多目标研究的趋势。因此，作者引入信息熵理论<sup>[11]</sup>结合进化计算来解决多目标优化问题。

信息熵理论在进化计算中也有很多应用。Shannon 引入了信息熵来度量任意进化过程的信息量，以度量一个解集是如何扩散和均匀分布的<sup>[12]</sup>。Tan 提出了 Parzen window 密度估计来估计可能性密度的非参数方法<sup>[13]</sup>。其基本思想就是充分利用熵值指标，提供对每个解点沿着已发现 Pareto 前端的信息内容的评估。Farhang-Mehr A 提出的熵值中，引入网格法来尽可能地度量相邻的影响信息<sup>[14]</sup>。而后定义了影响函数来描述进化种群中的影响信息，求解所有解点之间影响函数数值的总和，即为目标密度函数<sup>[7]</sup>，而后得到信息熵指标的数值。

Corne 等人对 PESA 做了一些改进，提出了基于区域选择 (region-based selection) 概念的 PESA-II 算法<sup>[15]</sup>。Mikkel 提出了一种 PESA 改进算法使得 PESA 的运行时间复杂度达到  $O(GN \log^{M-1} A \log \log A)$ <sup>[16]</sup>，其中， $G$  代表进化代数， $N$  代表种群大小。鉴于上述两篇文献与本文的接近性，本文将在第 5 节进一步给出详细的对比论述。

因为信息熵指标的核心准则是依据大量的相邻信息来考量解集的分布，从各方面来说，信息熵度量指标具有极大优势。本文将信息熵度量指标引入到第二代进化多目标优化算法的经典代表 PESA 算法中，通过量化评估多目标进化过程中解点的分布情况，反映进化解集的进化发展情况和成熟阶段，由此获知种群进化是否到达成熟阶段，及时结束 PESA 算法，以此来简化算法的时间复杂度。此外，信息熵度量指标在多样性<sup>[14]</sup>度量方面表现突出，它充分考虑了进化种群中所有解点的位置，量化地反映解集中的信息传递，因而信息熵指标可以准确度量一个 Pareto 解集的分布性和收敛性。尤为重要的是，随着解点数量的增长，信息熵值度量指标的计算量呈现线性增加。计算难度将增加不多，计算时间和计算量也只是缓慢增长。

### 3 C-PESA 算法

#### 3.1 PESA 算法简介

作为第二代进化多目标优化算法的经典代表，PESA 算法引入了精英保留策略<sup>[4]</sup>。首先设置了一个内部种群和一个外部种群，开始进化后，将内部种群的支配个体加入外部种群中。每新加入一个个体到外部种群中，就要同时淘汰外部种群中的一

个个体。即在外种群中寻找拥挤系数最大的个体，并将其删除，如果多个个体的拥挤系数相同，则随机地删除一个。为了保持种群的多样性，采用了网格方法 (hyper-grid) 将个体空间划分成若干个格子 (hyper-box)。这样每个个体就与某个网格相关联。PESA 算法用选择网格代替个体选择，在一定程度上提高了算法的效率。

PESA 算法的具体步骤如图 1 所示。

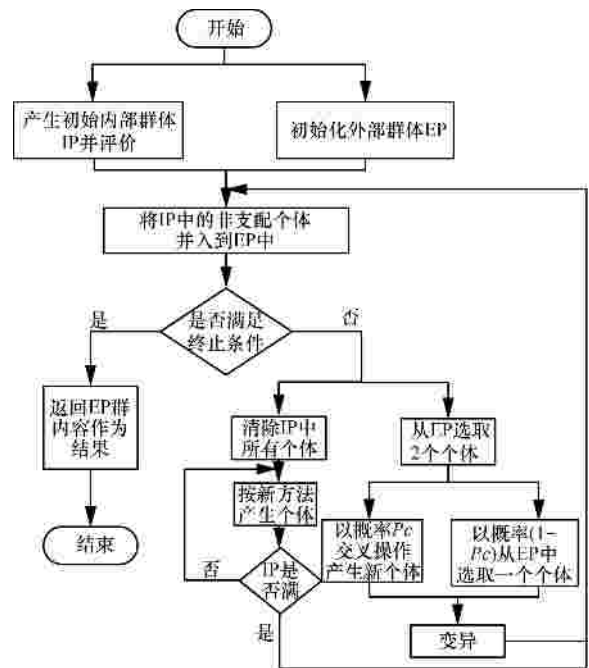


图 1 PESA 算法流程

PESA 算法用选择网格代替个体选择策略保持了种群的多样性，与基于个体选择的 MOEA<sup>[17]</sup>相比，在一定程度上提高了算法的性能。但是算法步骤中欠缺对时间复杂度的优化步骤，所以作者提出 C-PESA 算法，避免 PESA 算法因为进化种群数量增加而带来的复杂进化过程，在一定程度上减少了算法的计算量。

#### 3.2 C-PESA 算法流程图

考虑到 PESA 算法和信息熵度量指标均采用了网格方法作为核心手段，作者将熵值度量指标引入 PESA 算法中，提出了改进的 C-PESA 算法。通过持续的信息熵值指标计算情况，观察进化解集从逐步进化发展过程到达成熟阶段的过程。也就是说，根据信息熵指标的变化情况判断种群进化是否已经到达成熟阶段。依据得到的种群个体的进化程度，尽早结束 PESA 算法的优化过程。具体的 C-PESA 算法基本流程如图 2 所示。

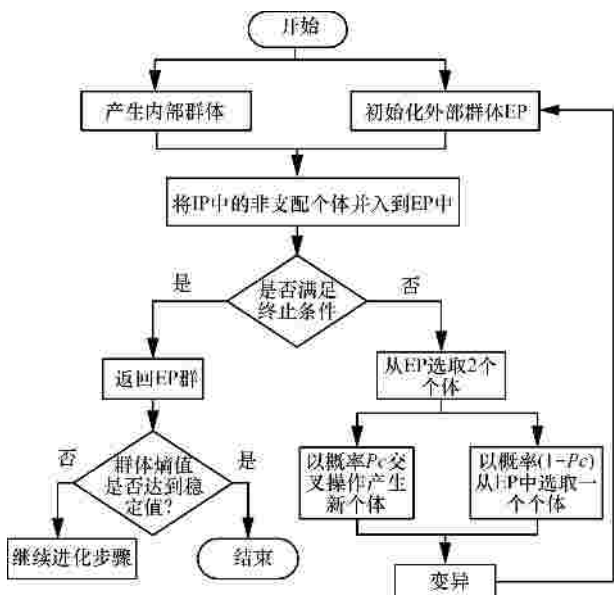


图 2 C-PESA 算法流程

首先，度量每一次返回的 EP 进化种群的信息熵数值，根据信息熵指标的变化情况，判断 PESA 算法的进化种群是否在逐步发展乃至已经达到成熟阶段。其次，在进化计算中，非支配解点均匀分布于 Pareto 前端。随着种群逐步进化达到成熟，进化解集逐渐覆盖了整个前端。在进化种群处于增长状态时，信息熵值在不断增长，分布情况在一定程度上显著提高。当信息熵达到饱和现象出现之前，进化种群接近所有可行域且非支配，并且沿着 Pareto 前端尽可能均匀地分布。最后，当进化种群达到成熟阶段时，也就是达到分布饱和性时，此时信息熵数值保持不变，依据这个特性，当求得的信息熵数值在一定范围内稳定不变时，说明 PESA 的优化效果已经达到，优化计算过程可以从这里停止。这样就可以避免大量不必要的复杂进化过程，在一定程度上简化了算法的时间复杂度。

### 3.3 信息熵计算规则

在优化过程中，笔者得到一系列多目标进化算法的解集。运用信息熵指标度量多样性依据如下步骤进行。

**Step1** 分割目标空间为合适大小的网格。

当提到一个有着  $r$  个目标的多目标进化问题，需要创建一个  $2r$  的边界，其下边界标记为  $lb_k$ ，上边界标记为  $ub_k(k=1,2,\dots,r)$ 。一般说来，对于有  $r$  个目标的多目标优化问题，标记为  $(lb_1, lb_2, \dots, lb_r)$  和  $(ub_1, ub_2, \dots, ub_r)$ 。此外，一个网格可以被分割成若干个小区域，每一个都被称作超立方体 (HC, hyper-cubes)。详细的划分是依据进化过程中的种群数

量和作者所需要优化的目标。然后，用  $r^i$  表示 HC，这里  $i=(i_1, i_2, \dots, i_r)$ ， $d$  是一个常数且比 2 大， $d$  表示每个维度上的分割次数。鉴于  $d$  的常数值是主观选择，所以应该注意如何选择一个合适的数值。此选择原则是保证每一个网格尺寸不大于一个特定区域，在此特定区域中，所有的解点都可以被认为是相同解点，或者解释为此决策制定不会对其中解点产生影响。因此，每一个  $r^i$  的边界可以描述为

$$rub_{k,j} = [lb_k + (i_k/d)(ub_k - lb_k)]w_k \quad (1)$$

$$rlb_{k,i} = [lb_k + ((i_k - 1)/d)(ub_k - lb_k)]w_k \quad (2)$$

其中， $w_k$  是第  $k$  维上每一个小区域的宽度，可表示为

$$w_k = range_k / d \quad (3)$$

其中， $range_k$  代表第  $k$  维的宽度。

如果是  $m$  个目标的优化问题，可以创建一个有着  $a_1 a_2 \dots a_m$  个网格的超网格平面(这里  $a_1, a_2, \dots, a_m$  表示每一维度上网格的数量)。对于  $m$  个目标问题，应该注意到超网格的数量  $a_1 a_2 \dots a_m$  应该少于进化种群数量。也就是说， $a_1 a_2 \dots a_m < N$  ( $N$  为进化种群数)。

**Step2** 计算目标空间的影响函数。目标空间的影响函数定义为

$$O(l_i, y): R^? \rightarrow R$$

其中， $l_i, y$  是第  $i$  个目标点和第  $y$  个目标点之间的欧几里得距离 (Euclidean distance)， $R$  表示距离值。标准的影响函数可以表示为多种不同的形式，如抛物线、方波或高斯函数等。影响函数是根据问题需求做出合适选择。这里，笔者选择高斯影响函数，影响函数形式如下：

$$W(r) = \frac{1}{s \sqrt{2\pi} e^{-r/\sqrt{2}s}} \quad (4)$$

其中， $s$  是分布程度的标准偏差， $r$  表示欧几里得距离。笔者选择第  $i$  个目标并且计算此目标和解集中其他点之间的影响函数。在这个步骤中，如何选择一个合理的参数  $s$  尤为重要。如果  $s$  过小，影响函数的变化将极为剧烈。因此个体的密度值就不能正确反映其解集的分布情况。然而，当  $s$  值太大时，会使得影响函数变化平稳<sup>[18]</sup>。

因此，作者引入  $3s$  规则(three-sigma rule)<sup>[18]</sup>来进行影响函数的计算，通过  $3s$  规则得到一个合理参数值  $s$ 。具体地说，设定  $3s$  等于第  $k$  维的宽度。 $3s$  规则阐述了几乎所有值的标准偏差控制在  $3s$  范围内，因此，超过  $3s$  误差范围的点应该被舍弃。

利用  $3s$  规则计算所得的  $s$  值是相当精确且合理的。

### Step3 计算密度函数。

可行目标空间每一个解点的密度公式被定义为所有解点的影响函数总和。也就是说，目标密度是自身和其他所有解点之间影响函数数值的总和，如式(5)所示。遍历进化种群中的所有解点，在通过解点目标距离进行影响函数计算之后可得到每个目标的密度值。

当提到  $m$  个目标问题时，密度函数可以被定义为

$$D(y) = \sum_{i=1}^n W(l_{i \rightarrow y}) = \sum_{i=1}^{a_1} \sum_{i=1}^{a_2} \dots \sum_{i=1}^{a_m} W(l(\langle i_1, i_2, \dots, i_m \rangle, y)) \quad (5)$$

其中， $W(l(\langle i_1, i_2, \dots, i_m \rangle, y))$  是第  $a_m$  维度上第  $i$  个目标点和第  $y$  个目标点之间的欧几里得距离。

对密度函数的公式，作者进行了简化。一般来说，选择  $b_k = [(a_k)^{1/q}]$ 。这里  $q$  是一个整数， $q$  不小于 2。假定  $b_k = [(a_k)^{1/2}]$ ，对于一个解点，其对应超网格为  $\langle u_1, u_2, \dots, u_m \rangle$ 。假定  $c_k = \min\{u_k + b_k, a_k\}$ ，这里  $u_k$  为前面定义的网格标识， $d_k = \max\{u_k - b_k, 1\}$ 。因此，最简单的密度函数应该被描述为

$$D(y) = \sum_{i=1}^n W(l_{i \rightarrow y}) = \sum_{i=d_1}^{c_1} \sum_{i=d_2}^{c_2} \dots \sum_{i=d_m}^{c_m} W(l(\langle i_1, i_2, \dots, i_m \rangle, y)) \quad (6)$$

### Step4 计算整个进化解集的熵值。

根据式(7)计算信息熵值。因为在 Shannon 定义的信息熵概念中，熵值可能性总和为 1，定义标准化密度  $r_{ij}$  为<sup>[12]</sup>

$$r_{ij} = \frac{D_{ij}}{\sum_{k_1=1}^{a_1} \sum_{k_2=1}^{a_2} D_{k_1 k_2}} \quad (7)$$

其中， $D_{ij}$  即  $D_{ij}(y)$ ，是 Step3 提及的密度函数， $i$  和  $j$  为网格位置索引， $k_1, k_2$  表示二维维度上的位置索引。

对标准化密度进行上述定义后，可得出

$$\sum_{k_1=1}^{a_1} \sum_{k_2=1}^{a_2} r_{k_1 k_2} \ln(r_{k_1 k_2}) = 1, r_{k_1 k_2} > 0; \forall k_1 k_2 \quad (8)$$

因此，对于一个二维目标空间，信息熵度量指标可被定义为

$$H = - \sum_{k_1=1}^{a_1} \sum_{k_2=1}^{a_2} r_{k_1 k_2} \ln(r_{k_1 k_2}) \quad (9)$$

对于一个  $m$  维目标空间，可行域被划分成  $a_1 a_2 \dots a_m$  个网格，信息熵被定义为

$$H = - \sum_{k_1=1}^{a_1} \sum_{k_2=1}^{a_2} \dots \sum_{k_m=1}^{a_m} r_{k_1 k_2 \dots k_m} \ln(r_{k_1 k_2 \dots k_m}) \quad (10)$$

其中， $k_1, k_2, \dots, k_m$  表示  $m$  维维度上的位置索引。

由于进化种群呈现增长状态，导致了信息熵数值的不断增长。因此，在信息熵饱和现象出现之前，分布情况在一定程度上显著提高。也就是说，进化种群接近所有可行域且非支配，从而沿着 Pareto 前端尽可能均匀地分布。当提及分布饱和性时，也就是进化种群达到成熟阶段，此时信息熵数值保持不变。当信息熵值不再变化时，可以获知进化种群已经达到成熟。从而信息熵数值不变的时刻，即可以停止优化演算进程。

## 4 仿真分析

为了考察信息熵指标对种群进化过程的精确刻画情况，即 C-PESA 进化算法的时间复杂度是否减少，作者进行了一系列实验。实验仿真环境在 jmetal 框架<sup>[19]</sup>下测试 C-PESA 算法对几类经典测试问题的解决情况。实验中选取的经典测试问题、维数、变量域、目标函数等相关情况如表 1 所示，C-PESA 算法的交叉概率、交叉分布指数、突变概率、突变分布指数等相关参数<sup>[5]</sup>选取如表 2 所示。

实验 1 用来测试信息熵指标对种群进化过程的刻画情况，测试 C-PESA 算法对解决经典问题 ConstrEx<sup>[19]</sup>、Kursawe<sup>[20]</sup>、ZDT1<sup>[21]</sup>后进化种群的信息熵数值和进化种群的 Pareto 分布情况，通过对比证明信息熵度量指标的有效性。实验 2 测试将信息熵、GD(generational distance)<sup>[22]</sup>、HV(hypervolume)<sup>[23]</sup>、IGD(inverted generational distance)<sup>[24]</sup>等不同度量指标添加进 PESA 算法后，随着种群数量的增加所需的计算时间，测试问题为 DTLZ1<sup>[25]</sup>。实验 3 测试 C-PESA 算法的时间复杂度。

在仿真过程中，每一类实验共进行 30 次，其数据均为 30 次实验的平均值。

实验 1 测试信息熵指标对种群进化过程的刻画情况。

本实验测试了 PESA 算法对测试问题 ConstrEx、Kursawe、ZDT1 的解决情况。测试问题详细情况如表 1 所示。图 3 记录了测试问题 ConstrEx 的实验结果，图 4 和图 5 分别为 Kursawe 和 ZDT1。图 3(a)、图 4(a)和图 5(a)记录了熵值随着迭代次数变化的情况，图 3(b)、图 4(b)、图 5(b)记录了与 C-PESA 算法相对应的 Pareto 分布情况。

表 1 实验中用到的测试问题

测试问题	维数	变量域	目标函数
ConstrEx	2	$x_1 \in [0.1, 1.0]$ $x_2 \in [0, 5]$	$f_1(x) = x_1$ $f_2(x) = (1 + x_2) / x_1$
ZDT1	30	$x_i \in [0, 1]$	$f_1(x) = x_1, f_2(x) = g(x) [1 - \sqrt{x_1 / g(x)}]$ $g(x) = 1 + 9(\sum_{i=2}^n x_i) / (n - 1)$
Kursawe	3	$x_i \in [-5, 5]$	$f_1(x) = \sum_{i=1}^{n-1} (-10e^{-(0.2)\sqrt{\frac{1}{n-1} \sum_{k=1}^n x_k^2}})$ $f_2(x) = \sum_{i=1}^n ( x_i ^a + 5 \sin(x_i^b))$
DTLZ1	$k +  x_k  - 1$	$x_k \in [0, 1]$	$f_1(x) = \frac{1}{2} x_1 x_2 \prod_{k=1}^n (1 + g(x_k))$ $f_2(x) = \frac{1}{2} x_1 x_2 \prod_{k=1}^n (1 - x_{k-1}) (1 + g(x_k))$ ⋮ $f_{k-1}(x) = \frac{1}{2} x_i (1 - x_2) (1 + g(x_k))$ $f_k(x) = \frac{1}{2} (1 - x_1) (1 + g(x_k))$ where $g(x_k) = 100 \left[  x_k  + \sum_{x_i \in x_k} ((x_i - 0.5)^2 - \cos(20p(x_i - 0.5))) \right]$

表 2 C-PESA 参数设置

C-PESA 参数	参数值
交叉分布指数(distribution index for crossover)	15
交叉概率(crossover probability)	0.8
突变分布指数(distribution index for mutation)	20
突变概率(mutation probability)	$1/n(n$ 是输入数量)

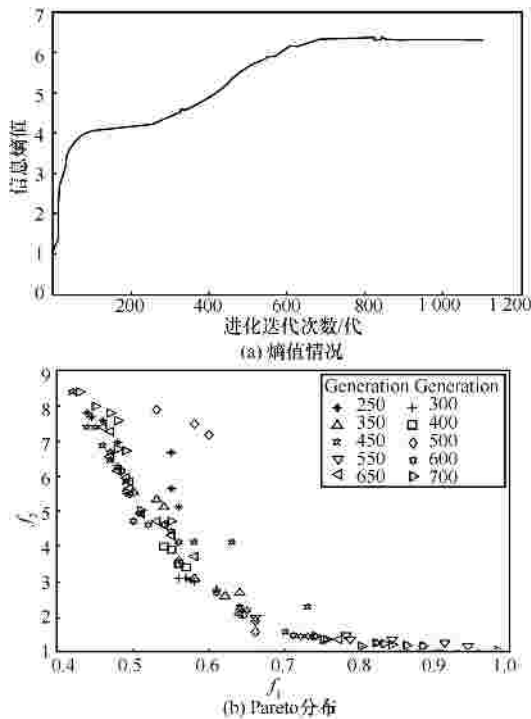


图 3 C-PESA 测试 ConstrEx

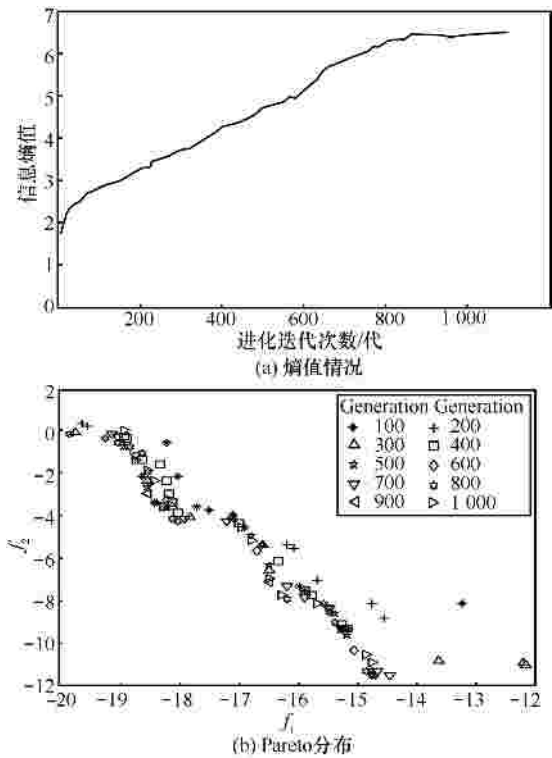


图 4 C-PESA 测试 Kursawe

在测试问题 ConstrEx 中，从图 3(a)可以看出，当进化迭代次数达到 700 代左右时，熵值已经达到一个定值，此后保持稳定；对照图 3(b)，从 400~600 代，Pareto front 的情况显著提高；这时可看出从进化计算开始到达 700 代之前，进化解集逐渐覆盖了

整个前端。此时，信息熵值在不断增长，分布情况在一定程度上显著提高，进化种群沿着 Pareto 前端尽可能均匀地分布。而当进化种群达到 700 代时，此时信息熵数值在一定范围内稳定不变时，说明 C-PESA 的优化效果已经达到，C-PESA 算法可以从这里结束。

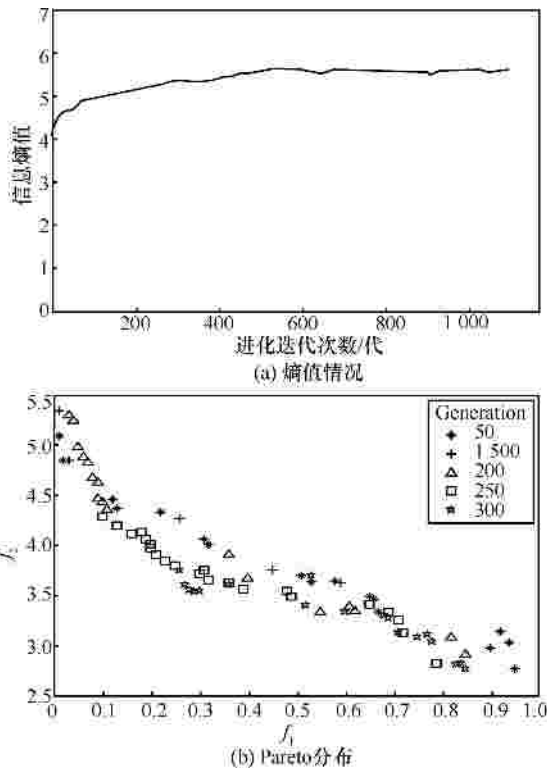


图 5 C-PESA 测试 ZDT1

同理，在测试问题 Kursawe 和 ZDT1 中，图 4(a) 中的 800 代和图 5(a) 中的 250 代为熵值的稳定值，即 PESA 算法的结束时刻。图 4(b) 中的 200~700 代和图 5(b) 中的 50~200 代为 Pareto front 显著提高的时期。

通过这个实验可以发现，熵值可以作为多目标进化算法的一个度量指标，用来度量其多样性、收

敛性等特性，并反映进化计算算法的进化过程，从而判断进化算法是否已经达到预期效果。

实验 2 测试不同度量指标的计算时间。

度量种群多样性的经典指标还有 GD、HV 和 IGD，其详细说明如表 3 所示。

实验测试问题为 DTLZ1，进化种群数量为 100，所用算法为 PESA 算法。

图 6 分别记录了信息熵、GD、HV 和 IGD 度量种群多样性所需要的时间。由图 6 可知，GD、HV、IGD 度量方法随着种群数量增长，其计算时间都呈现急剧增加趋势。而熵值度量指标的计算量增长缓慢，并且更加接近线性增长，这将大大提高度量效率。

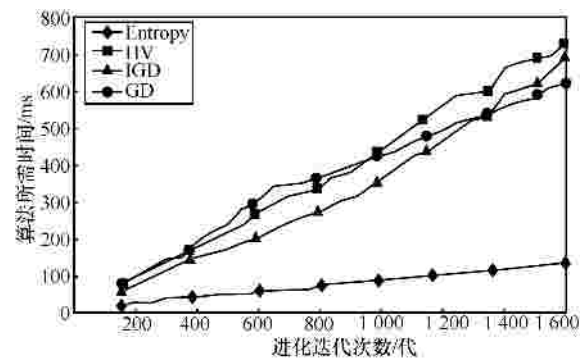


图 6 不同度量指标的计算时间对比

实验 3 信息熵引入到 PESA 算法后的进化算法的时间复杂度。

本实验测试 C-PESA 算法的时间复杂度。实验测试问题为 ConstrEx，进化种群数量为 100。记录实验数据如图 7 所示。图 7 中虚线为进化终止时的迭代次数。

图 7 中可以看出当进化迭代次数达到 1300 代时，种群进化已经到达成熟阶段，算法的优化过程可以就此结束。这样在一定程度上简化了算法的时间复杂度。此实验说明，随着进化种群的增加，

表 3 实验中的相关度量指标

名称	表达式	参数说明
GD (generational distance)	$GD = A / N_{\text{known}} \sqrt{\sum_{i=1}^{N_{\text{known}}} d_i^2}$	$N_{\text{known}}$ : 进化所得前端解集的解点数量。 $d_i$ : 进化解集和真实 Pareto 前端的距离
HV (hypervolume)	$HV(A) = \int_{(0,L,D)}^{(1,L,1)} \partial_A(z) dz$	$A$ 是目标向量 $Z$ 是目标空间 $\partial$ 是 attainment 方程 <sup>[23]</sup>
IGD (inverted generational distance)	$IGD(p, p_i) = \frac{\sum_{x \in P_i} (d_{\min}(x, p))^2}{ P_i }$	$P_i$ 是 Pareto 最优解前端解集, $P$ 是非支配解集, $d_{\min}(x, p)$ 是非支配解集中的解点 $x$ 到 Pareto 前端的最短欧几里得距离

C-PESA 算法所需的计算时间只是接近线性增加，计算量增加不大。通过持续的信息熵值指标计算情况可以观察到进化解集从逐步进化发展过程到达成熟阶段的过程。

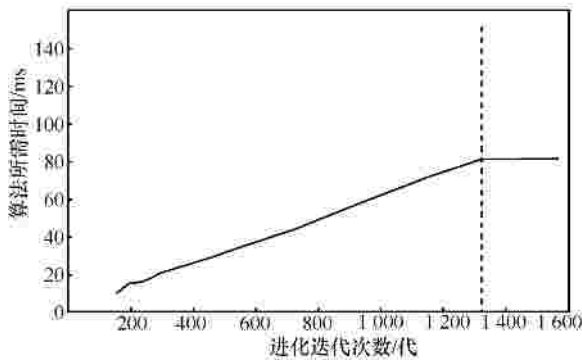


图 7 C-PESA 算法的进化时间

### 5 讨论

本节将对文献[15,16]中 2 个与本文工作最为接近的 PESA 改进算法进行对比论述。

在文献[15]中,Corne 等人提出了基于区域选择的 PESA- 算法。相比其他基于个体选择的进化算法, PESA-II 算法使用网格选择,使得算法的效率得到一定程度的提高。但该算法用网格选择代替个体选择,在一定程度上提高了算法的效率。但是, PESA- 算法进行进化计算所需的运算量、计算难度及运算时间都将随着进化种群的数量而急剧增加,这将给进化计算的进行带来极大麻烦。与前者相比,本文提出的 C-PESA 算法仍然采用基于网格选择的精英保留策略,同时利用信息熵指标反映进化种群从开始进化到达成熟的过程,根据信息熵值的不变性,获知进化种群已经成熟、种群已经分布均匀多样等相关信息,从而及时地结束 PESA 算法的计算过程。信息熵指标随着进化种群数量的显著增加,时间复杂度接近线性增长,即  $O(N)$ 。相比于 PESA- 算法的  $O(N \times N^*)$ ,本文提出的算法由于引入信息熵,使得 C-PESA 的时间复杂度得到一定的优化。

在文献[16]中,Mikkil 提出的新算法使得 PESA 算法运行时间复杂度减少到  $O(GM \log^{M-1} A \log \log A)$ 。但是这些改善主要着眼于优化进化算法步骤或者算子,侧重于考虑如何提高非支配分类的效率等。与前者相比,本文提出的 C-PESA 算法利用熵值度量,使运算更简单便捷,它精确地反映了种群进化

信息,根据其变化情况考量 PESA 算法的进化后种群个体的进化程度,判断种群进化是否到达成熟阶段,从而可以尽早结束 PESA 算法的优化过程。同时,随着解集数量的增加,熵值指标的计算量只是接近线性增加,这在一定程度上简化了算法的时间复杂度。

### 6 结束语

本文对一些研究者提出的信息熵度量指标进行了深入研究,鉴于 PESA 算法和信息熵度量指标都采用了网格方法,提出将熵值度量指标引入 PESA 算法中。利用信息熵指标对进化过程精确反应,从而判断进化种群被优化的程度。仿真结果表明,信息熵指标可以精确刻画种群进化过程,在计算时间上优于现有的一些度量方法,并且极大地优化了 PESA 进化算法的时间复杂度。

### 附录：3.3 节算法的符号描述

本文 3.3 节算法的符号描述如表 4 所示。

表 4 文中 3.3 节算法的符号描述

符号	描述
$lb_k(k=1,2,\dots,r)$	上边界
$ub_k(k=1,2,\dots,r)$	下边界
$r^j, j=(i_1, i_2, \dots, i_r)$	超立方体 HC
$D$	每个维度上的分割次数
$w_k$	第 $k$ 维上每一个小区域的宽度
$range_k$	第 $k$ 维的宽度
$a_1, a_2, \dots, a_m$	每一维度上网格的数量
$O(l_{i,y})$	目标空间的影响函数
$l_{i,y}$	第 $i$ 个目标点和第 $y$ 个目标点之间的欧几里得距离
$S$	分布程度的标准偏差
$R$	欧几里得距离
$D()()$	密度函数
$W(l(<i_1, i_2, \dots, i_m>, y))$	第 $a_m$ 维度上第 $i$ 个目标点和第 $y$ 个目标点之间的欧几里得距离
$\langle u_1, u_2, \dots, u_m \rangle$	超网格
$u_k$	网格标识
$?_{ij}$	标准化密度
$i, j$	网格位置索引
$k_1, k_2, \dots, k_m$	$m$ 维维度上的位置索引

### 参考文献：

[1] SALEM F A, PETER J F. Diversity management in evolutionary

- many-objective optimization[J]. IEEE Trans on Evolutionary Computation, 2011, 15(2):183-195.
- [2] OLIVER S, ADRIANA L, CARLOS A. On the influence of the number of objectives on the hardness of a multiobjective optimization problem[J]. IEEE Transactions on Evolutionary Computation, 2011, 15(4):444-455.
- [3] SCHAFFER J D. Multiple objective optimization with vector evaluated genetic algorithms[A]. Proceedings of the International Conference on Genetic Algorithms and Their Applications[C]. Pittsburgh, PA, USA, 1985. 93-100.
- [4] CORNE D W, KNOWLES J D, OATES M J. The Pareto-envelope based selection algorithm for multi-objective optimization[A]. Proceedings of the International Conference on Parallel Problem Solving from Nature[C]. Paris, France, 2000. 839-848.
- [5] 公茂果, 焦季成, 杨咚咚等. 进化多目标优化算法研究[J]. 软件学报, 2009, 20(2):271-289.
- GONG M G, JIAO L C, YANG D D, *et al.* Research on evolutionary multi-objective optimization algorithms[J]. Journal of Software, 2009, 20(2):271-289.
- [6] [http://en.wikipedia.org/wiki/Big\\_O\\_notation](http://en.wikipedia.org/wiki/Big_O_notation)[EB/OL].
- [7] DAVID R W, ANDREA A, JOHN A C. Evolutionary improvement of programs[J]. IEEE Transactions on Evolutionary Computation, 2011, 15(4):515-538.
- [8] ABIDO M A. Two-level of nondominated solutions approach to multiobjective particle swarm optimization[A]. Proceedings of Genetic and Evolutionary Computation Conference[C]. London, 2007. 728-733.
- [9] RODRIGO P, LEANDRO N, DE C, *et al.* Neural network ensembles: immune-inspired approaches to the diversity of components[J]. Natural Computing, 2010, 9(3):625-653.
- [10] LOURDES A, JUAN J M. Diversity through multiculturalism: assessing migrant choice policies in an island model[J]. IEEE Transactions on Evolutionary Computation, 2011, 15(4):543-560.
- [11] SAXENA D K, DEB K. Non-Linear dimensionality reduction procedure for certain large-dimensional multi-objective optimization problems: employing coreentropy and a novel maximum variance unfolding[A]. Proceedings of the 4th International Conference on Evolutionary Multi-Criterion Optimization[C]. Sendai, Japan, 2007. 772-787.
- [12] SHANNON C E. A mathematical theory of communication[J]. Bell System Technical Journal, 1948, 27:379-429.
- [13] TAN K C, GOH C K, MAMUN A A, *et al.* An evolutionary artificial immune system for multi-objective optimization[J]. European Journal of Operational Research, 2008, 187:371-392.
- [14] ALI F M, SHAPOUR A. On the entropy of multi-objective design optimization solution set[A]. Proceedings of Design Engineering Technical Conferences and Computer and Information Engineering Conference[C]. Chicago, USA, 2002.829-838.
- [15] CORNE D W, JERRAM N R, KNOWLES J D, *et al.* PESA-II: region-based selection in evolutionary multi-objective optimization[A]. Proceedings of the Genetic and Evolutionary Computation Conference[C]. San Francisco, USA, 2001. 283-290.
- [16] MIKKEL T J. Reducing the run-time complexity of multiobjective EAs: the NSGA-II and other algorithms[J]. IEEE Transactions on Evolutionary Computation, 2003, 7(5):503-515.
- [17] ZHANG Q F, LI H. MOEA/D: a multiobjective evolutionary algorithm based on decomposition[J]. IEEE Transactions on Evolutionary Computation, 2007, 11(6):712-731.
- [18] [http://en.wikipedia.org/wiki/Standard\\_deviation](http://en.wikipedia.org/wiki/Standard_deviation)[EB/OL].
- [19] <http://www.aboutus.org/Jmetal.com>[EB/OL].
- [20] KURSAWE F. A variant of evolution strategies for vector optimization[A]. Proceedings of International Conference on Parallel Problem Solving from Nature[C]. London, UK, 1991. 193-197.
- [21] ZITZLER E, DEB K, THIELE L. Comparison of multi-objective evolutionary algorithms: empirical results[J]. Evolutionary Computation, 2000, 8(2):173-195.
- [22] VAN VELDHUIZEN D A, LAMONT G B. Evolutionary computation and convergence to a pareto front[A]. Late Breaking Papers at the Genetic Programming 1998 Conference[C]. California, USA, 1998. 221-228.
- [23] CAO Y. Matlab central file exchange: hypervolume indicator[EB/OL]. <http://www.mathworks.fr/matlabcentral/fileexchange/>, 2008.
- [24] VELDHUIZEN D A, VAN LAMONT G B. Multiobjective Evolutionary Algorithm Research: A History and Analysis[R]. Technique Report, 1998.
- [25] DEB K, AGRAWAL R B. Simulated binary crossover for continuous search space[J]. Complex Systems, 1994, 1(9):115-148.

#### 作者简介：



王堃 (1981-), 男, 江苏南京人, 博士, 南京邮电大学副教授、硕士生导师, 主要研究方向为无线通信与信息安全、普适计算、物联网与传感网技术、人工智能与数据挖掘等。



王琳琳 (1990-), 女, 江苏连云港人, 南京邮电大学硕士生, 主要研究方向为人工智能与数据挖掘。



刘艳 (1982-), 女, 江苏南京人, 南京工业大学讲师, 主要研究方向为物联网与传感网技术。

张玉华 (1990-), 男, 江苏南通人, 南京邮电大学硕士生, 主要研究方向为人工智能与数据挖掘等。

吴蒙 (1963-), 男, 上海人, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为无线通信与信息安全等。