

Multi-Objective Evolutionary of Distribution Algorithm Using Kernel Density Estimation Model

Na Luo^{1,2} and Feng Qian^{1,2}

¹⁾ School of Information Science and Engineering, East China University of Science and Technology, Shanghai, P.R.China
(naluo@ecust.edu.cn)

²⁾ Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University
of Science and Technology, Shanghai, P.R.China (fqian@ecust.edu.cn)

Abstract—Estimation of Distribution Algorithm (EDA) is a kind of new evolutionary algorithm which updates and samples from probabilistic model in evolutionary computation. Recently it is used to solve multi-objective problems. The key is how to construct probability model suitable for real distribution and how to keep diversity of solutions. In this paper a new multi-objective evolutionary of distribution algorithm using kernel density estimation model is presented. It used kernel density estimation method to obtain probability density of samples and generate new population with stochastic universal sampling method. In order to get pareto front of multi-objective problems, fitness sharing method is used. 5 bi-objective test problems are selected to test the performance of the new algorithm. The results show that multi-objective evolutionary of distribution algorithm using kernel density estimation model has better suitable performance for test problems comparing with non-dominated sorting genetic algorithm II, multi-objective particle swarm optimization and multi-objective estimation of distribution algorithm.

Keywords—Multi-Objective Evolutionary Optimization, Kernel Density Estimation, Non-Dominated Solutions, Pareto Front.

基于核密度估计的多目标分布估计算法

罗娜^{1,2} 钱锋^{1,2}

¹⁾ 华东理工大学信息科学与工程学院, 上海, 中国

²⁾ 华东理工大学化工过程先进控制和优化技术教育部重点实验室, 上海, 中国

摘 要 分布估计算法是一类新的进化算法, 它通过概率模型对每一代得到的优良解进行学习, 引导算法进行搜索, 加速最优解的获得。用该方法解决多目标问题是近年来一个研究的热点问题, 其关键在于概率模型的选择和如何保持多目标问题解的分散性。本文给出了一种采用核密度估计的多目标分布估计算法。该算法采用核密度估计算法使得到的样本分布接近于实际分布, 通过随机遍历抽样方法生成新种群, 多目标问题的处理采用适应值共享得到 PARETO 前沿。与 NSGA-II、多目标粒子群算法、多目标分布估计算法对 5 个测试函数的优化结果表明, 基于核密度估计的多目标分布估计算法对多个测试函数具有更好的适应性。

关键词 多目标进化算法, 核密度估计, 非劣解, PARETO 前沿

1. 引言

在科学与工程实践中, 很多决策问题都是多目标优化问题。多目标问题的各个目标之间可能是相互竞争的, 从而无法实现各目标同时达到最优。与单目标问题相比, 多

目标优化问题可以表述为:

$$y_i = f_i(x), \quad i = 1, \dots, D$$

$$\text{Minimize } y = f(x) = (f_1(x), \dots, f_D(x))$$

$$\text{subject to } e(x) = (e_1(x), \dots, e_J(x)) \geq 0$$

$$x \in S$$

其中, 目标 y 是向量 x 的函数, 向量 x 包括 P 个决策变量 (x_1, \dots, x_P) , 决策变量服从 J 个约束: $e_j(x) \geq 0, \quad j = 1, \dots, J$; S 为决策变量可行解域。

国家 973 项目(2009CB320603), 国家自然科学基金项目名称(20876044), 国家 863 计划重点项目(2008AA042902), 上海市科技攻关项目(09DZ1120400), 上海市重点学科建设项目资助(Project number: B504)

多目标问题的解不是单个解或一组连续的解，而是一组或几组连续解的集合。若 $x^* \in S$ ，且在 S 中不存在比 x^* 更优的解，则称 x^* 是多目标问题的 Pareto 最优解。对于一个给定的多目标优化问题 $f(x)$ ，所有 Pareto 最优解构成 Pareto 最优解集，其所有 Pareto 最优解所对应的目标向量构成多目标问题的 Pareto 前沿(Pareto Front)，多目标优化算法的设计目标就是寻找尽可能逼近 Pareto 前沿、且均匀分布的非支配解集。

与传统的基于梯度的优化方法需要将多目标问题化为单目标问题求解、每次求解只有一个 Pareto 解不同，进化计算领域的优化算法直接将多目标作为向量进行求解，一次可以寻得多个 Pareto 解，从而加快了寻优的效率。Deb 等人提出的 NSGA-II[1]是一种被广泛认可并在很多领域得到成功应用的求解多目标优化问题的遗传算法。NSGA-II 基于 Pareto 最优，根据个体之间的优劣性评价个体的好坏，采用快速非支配排序法对父代种群与其产生的子代种群的合集中的个体进行秩(rank)划分，秩越低个体越优；然后计算同秩集合的拥挤度(distance)，由拥挤程度衡量个体的多样性，最终根据个体的秩和拥挤度作为个体胜出的标准。此外，多目标粒子群算法[2]也相继提出并解决了一些实际的多目标优化问题。

与其他进化计算领域的优化算法相比，分布估计算法(Estimation Algorithm of Distribution, EDA)是一种基于概率模型的优化算法，也是进化计算领域的一个研究热点。与遗传算法不同，分布估计算法是一种利用概率模型对问题空间中有希望出现最优解的区域进行建模，然后对概率模型进行随机采样产生新的种群进而引导算法进行搜索。分布估计算法用概率模型取代遗传算法的交叉、变异操作，能够避免在连锁问题中遗传算法将解引导到局部最优的缺点。对于多目标问题的优化，研究者们相继提出了多目标 PBIL 算法[3]、多目标贝叶斯算法[4]、多目标 MIDEA 算法[5]、多目标 parzen 窗分布估计算法[6]、多目标 UMDA[7]等多目标分布估计算法。

无论是对于单目标问题还是多目标问题，选择合适的概率模型都是分布估计算法的核心问题。对于连续域的分布估计算法，高斯模型成为概率模型的首选。但实际样本并不一定满足高斯概率分布，同时样本的分布经常是没有任何先验概率的，因而采用基于假定概率模型的分布估计算法进行优化存在一定的缺陷。为克服这个缺点，Bin Li 等[8]提出了采用混合高斯函数来表征实际样本的概率分布，Shigeyoshi[9]、Dingnan[10]等提出了基于直方图模型的分布估计算法。结合多目标问题的特点，本文提出了一种基于核密度估计的多目标分布估计算法，采用核密度估计算法使样本分布接近于实际分布，通过直接抽样生成新

种群，采用适应值共享处理多目标问题得到 PARETO 前沿。与 NSGA-II、多目标粒子群算法、多目标 UMDA 相比，基于核密度估计的多目标分布估计算法具有更好的适应性。

2. 分布估计算法及核密度估计

2.1 分布估计算法

分布估计算法的基本思想就是使用概率的方法描述和表示每一代群体的分布情况，根据概率分布生成新的个体，保持整体向最优方向的演化。在分布估计算法中，一个优化问题中每个自变量 x_i 被看成一个随机变量，所有的随机变量构成一个随机向量 $x = (x_1, x_2, \dots, x_i, \dots, x_l)$ 。个体是随机向量的一个取值，而群体就是对应于该随机向量的一个分布。随机向量的分布是群体性能的一个指标，利用这个指标可以紧凑和整体地表示该群体。分布中包含了随机变量之间的概率依赖关系，这种关系也是一种基因之间的关系，学习随机变量的分布就等于是学习基因之间的关系。在一个概率分布上的采样过程可以生成更有价值的群体和个体。因此，分布估计算法利用每一代的个体，从中学习随机向量的分布，然后在学习到的分布的基础上再生成下一代新个体，如此循环。

对于连续优化问题，分布估计算法分为两种思路：一种是把连续随机变量离散化，然后使用离散分布估计算法。离散的方法可以简化计算和减少存储要求，这在很多情况下是一个好的选择。离散化的问题主要是由此而带来的误差太大。另一种思路是直接对连续随机变量建模。为简化问题，通常使用的概率模型是高斯模型。这样，建模的主要任务就是计算模型的均值和方差(或协方差矩阵)。使用高斯模型有一些优点：容易处理和实现，参数少。另外，在很多实际情况下变量的分布是高斯分布，或近似高斯分布。在变量独立假设下，连续的分布估计算法中的样本分布模型可以用公式(1)表示

$$P(x_1, x_2, \dots, x_l) = P(x_1)P(x_2) \cdots P(x_l) \quad (1)$$

假设 $P(x_i)$ 服从正态分布：

$$P(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2\sigma_i^2}(x_i - \mu_i)^2} \quad (2)$$

简记为 $P(x_i) \sim N(\mu_i, \sigma_i^2)$ 。当给定一组样本时，使用最大似然估计方法估计分布的均值 μ_i 和方差 σ_i^2 。

$$\hat{\mu}_i = \frac{1}{n} \sum_{k=1}^n x_{ik} \quad (3)$$

$$\hat{\sigma}_i^2 = \frac{1}{n} \sum_{k=1}^n (x_{ik} - \hat{\mu}_i)^2 \quad (4)$$

其中， x_{ik} 是变量 x_i 的第 k 个样本， $\hat{\mu}_i$ 和 $\hat{\sigma}_i^2$ 分别为 μ_i 和

σ_i^2 的估计。

在一个给定的正态分布上的采样并不困难。一个连续独立随机变量的分布估计算法如下：

步骤 1：初始化群体，并对每一个个体估值；

步骤 2：从群体中选择 m 个最优个体；

步骤 3：根据选择的个体估计随机向量的概率分布；

步骤 4：根据上一步估计的分布，采样新一代个体，并对每一新个体估值；

步骤 5：如果某准则满足，则算法停止；否则，返回步骤 2。

由以上算法可以看出，分布估计算法的核心是概率密度分布模型的生成及表示。对概率密度分布模型的估计可以分为参数法、半参数法和非参数法。参数法是根据经验假定样本的分布符合某一类型的概率模型（如符合高斯分布），求得假设模型参数的方法；半参数法是 30 年代才发展起来的一种统计模型，通过既有参数分量，又含有非参数分量的模型描述实际问题；非参数法是采用无参数的直方图等方法来表示概率模型的方法。分布估计算法根据采用概率分布估计函数可以分为 PBIL, cGA, UMDA, MIMIC, COMIT, BMDA, ECGA, BOA, FDA, IDEA, HEDA 等。这些算法或采用固定的概率模型，如 BOA 采用贝叶斯网络进行概率密度估计，或采用非参数的概率模型，如 HEDA 采用直方图进行概率密度估计。由于变量之间依赖关系未知，通过主成分分析、概率主成分分析分子、因子分析、混合因子分析等方法降低变量之间的耦合，简化概率模型的建立也在分布估计算法中取得了一定的效果。

核密度估计概率分布是一种非参数估计方法，采用核密度方法，无需假设数据的分布情况，同时核密度估计能够反映样本的分布情况，广泛应用于模式识别领域。本文采用核密度估计算法对样本的概率模型进行估计。

2.2 核密度估计

与文献[7]所用的直方图法估计密度分布相比，核函数方法固定单元格体积，求落入单元格内的样本数，并以此来估计密度。采用核函数方法，避免了直方图法对于固定维数的单元格，其个数随数据向量的维数呈指数增长的缺点。

核密度估计假设样本的概率密度分布函数是一个光滑的连续函数，落在某一区间的概率可以通过核函数来表示。设 x_1, x_2, \dots, x_n 是从一维总体 x 中抽出的独立同分布的数据， X 具有未知密度 $f(x), x \in R$ ，如果存在全直线上有界函数 $K(u) \geq 0$ 且满足以下条件：

$$(1) \int_{-\infty}^{+\infty} |k(u)| du < +\infty \quad (2) \lim_{|u| \rightarrow \infty} uk(u) = 0 \quad (3) k(-u) = k(u) \quad (4)$$

$$\int_{-\infty}^{+\infty} k(u) du = 1$$

则： $f(x)$ 的密度核估计为：

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-x_i}{h_n}\right) \quad (5)$$

式中， $k(u)$ 为窗或核函数； h_n 为与 n 有关的正的光滑参数，称其为带宽或光滑参数或窗宽。

核函数的选择可以有多种：如 Parzen 窗(uniform)、三角(Trangle)、Epanechnikov、四次(Quartic)、三权(Tri-weight)、高斯(gauss)、余弦(Cosinus)、指数(Exponent)等。核函数的选择取决于根据距离分配各个样本点对密度贡献的不同。通常选择什么核函数不是密度估计中最关键的因素，因为选用任何核函数都能保证密度估计具有稳定相合性。最重要的是带宽对估计分布的光滑程度影响很大，自然地如何选择带宽将成了最重要的问题。核函数的密度估计之所以能受到欢迎，是因为它在带宽选择上能从数学的角度进行论证带宽最优原则。并且在独立同分布的情况下，核估计量具有逐点渐进无偏性和一致渐进无偏性、均方相合性、强相合性、一致强相合性等。

对样本分别采用直方图、核密度对概率分布进行估计的结果见图 1。

控制核密度宽度的参数，根据样本的数量选择，该参数太大，则得到的分布模型分辨率低，分布密度平均化；该参数太小，则分布模型的统计变动大。

核密度估计的带宽选择问题，一般指的是全局带宽问题，包括 3 种最优原则：(1)极小化均方误差；(2)极小化积分均方误差；(3)极小化渐进积分均方误差。一般 h_n 随着 n 增大而减小，当 $n \rightarrow \infty$ 时， $h_n \rightarrow 0$ ，但 h_n 取得太小则随机干扰增大，降低了估计的准确性； h_n 取得太大则估计曲线太光滑，无法体现估计样本的差别。

3. 多目标核密度分布估计算法

核密度在估计样本的概率分布时，对多模态函数，有自适应聚类效果，可以在算法进化过程中寻找到多个局部最优点，这样，在多个局部最优点间寻找到全局最优点的机会增加，从而容易收敛到最优解。算法(Multi-objective kernel density estimation of distribution algorithm ,MKEDA)如下所示：

算法. MKEDA

(1) 随机产生 M 个个体作为初始群体 $D_l, l=1$

(2) 计算 M 个个体的适应值，如果符合终止条件，算法结束，否则继续进行；

(3) 筛选群体中的非劣解放入精英集。如果精英集的个数小于优势群体个数 N ，选择所有精英集作为优势群体 D_l^s ；否则计算其中每个个体的适应度，按与适应度值成比

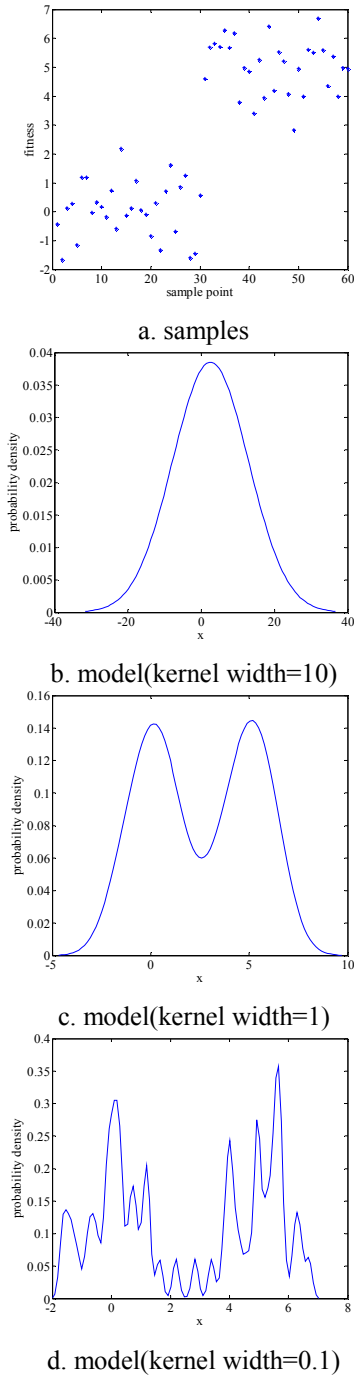


图 1. 样本的概率图模型

Fig. 1. Probabilistic model of individuals

例的概率方法随机选择 $N < M$ 个个体作为优势群体 D_t^s ;

(4) 由优势群体 D_t^s 采用核密度估计方法建立概率密度模型;

(5) 从概率向量 $p(x)$ 中采用随机遍历抽样 M 次, 得到新一代的群体, 返回(2)。

3.1 核密度估计函数宽度的选取

本文算法的核函数采用高斯核函数。核密度估计函数

的宽度对核密度估计的作用非常重要, 对于基于核密度的分布估计算法, 在群体进化过程中, 每次建模的数据均不同, 需要动态地选取核密度的宽度。核密度宽度的选取方法有以下几种:

■ 确定的核宽度控制。不使用搜索过程中的任何反馈信息, 根据确定的规则修改确定核宽度的值。如使用宽度跟代数成线性关系或非线性的的方案, 不定期地使用该规则。运行代数越大时, 核密度宽度越小。

$$w(t) = \frac{(w_i - w_e)(T_{\max} - t)}{T_{\max}} + w_e \quad (6)$$

$$w(t) = \left(\frac{t-1}{T_{\max}-1} \right)^\lambda (w_i - w_e) + w_e \quad (7)$$

其中, w_i 表示初始宽度, w_e 表示最小宽度, T_{\max} 表示最大代数, t 表示当前进化代数。

■ 适应的核宽度控制。利用搜索过程中的反馈信息来决定改变策略核宽度的值。如采用选择个体作为样本, 求样本和其 k 个近邻之间的平均距离作为核宽度。

$$w(t) = \frac{1}{n} \sum_{j=1}^n \frac{1}{k} \sum_{i=j-k/2}^{j+k/2} x_i \quad (8)$$

3.2 抽样方法的选取

通过随机变量抽样从核密度分布中产生新种群。一种简单的减少抽样误差的抽样方法是 Baker 的随机遍历抽样, 该方法首先提出来时用于遗传算法中的比例选择操作。将该抽样方法扩展用于离散概率分布模型的采样。伪代码如图 2 所示。

```

1 % S: sampling individual number, n: number of variables, H:
number of bins
2 % p[j][h]: probability density of bin h of variable xj
3 % l[j][h]: left edge position of bin h of variable xj
4 % v[S][n]: array for sampled vectors
5 for j=1:n
6     ptr=rand(1,1);
7     sum1=0;
8     k=1;
9     xh=randperm(S);
10    for h=1:H
11        expected=p(j,h)*S; %
12        sum1=sum1+expected;
13        while sum1>ptr
14            ptr=ptr+1;
15            v(xh(k),j)=l(j,h)+(l(j,h+1)-l(j,h))*rand(1,1);
16            k=k+1;
17        end
18    end
19 end

```

图 2. E-SUS 的伪代码

Fig. 2. Pseudo matlab code of the E-SUS

3.3 多目标问题的处理

与 NSGA-II[1]的方法相同, 多目标核密度分布估计算

法采用适应值共享方法选择下一代建立概率模型所用到的非劣解。在可行解目标空间中选取其中非劣解个体作为“精英集”(elitism); 通过小生境(niche)技术给精英集中的非劣解个体分配适应度值, 聚集程度越大的个体适应度越小, 精英集中第 i 个个体的适应度为:

$$F(i) = \frac{1}{\sum_{j=P} s(d(i, j))} \quad (9)$$

其中 P 为精英集, $s(d(i, j))$ 为精英集中第 i 个个体和第 j 个个体适应度分享函数:

$$s(d(i, j)) = \begin{cases} 1 - \left(\frac{d(i, j)}{\sigma_{share}} \right)^\alpha, & \text{if } d(i, j) < \sigma_{share} \\ 0, & \text{其它} \end{cases} \quad (10)$$

其中 $d(i, j)$ 为第 i 个个体和第 j 个的距离, 可以采用决

策变量空间距离 $d(i, j) = \|x_i - x_j\|$, 也可采用目标变量空间距离 $d(i, j) = \|y_i - y_j\|$, σ_{share} 为 niche 半径。

采用与适应度值成比例的轮盘赌概率方法选取精英集中个体作为下一代建模需要的群体。选出每次迭代后群体中非劣解, 加入精英集, 并删除其中的劣解, 形成新的精英集; 若精英集中非劣解个体个数超过精英集的容量, 则删除其中适应度最小的部分个体; 迭代多次之后的精英集便是算法所得非劣解集。

4. 实验仿真

多目标优化问题的测试函数的设计比单目标优化问题的设计要复杂得多, Schaffer、K.Ded 等人[1]针对多目标演化算法的任务和困难设计了一系列测试函数。这些函数往

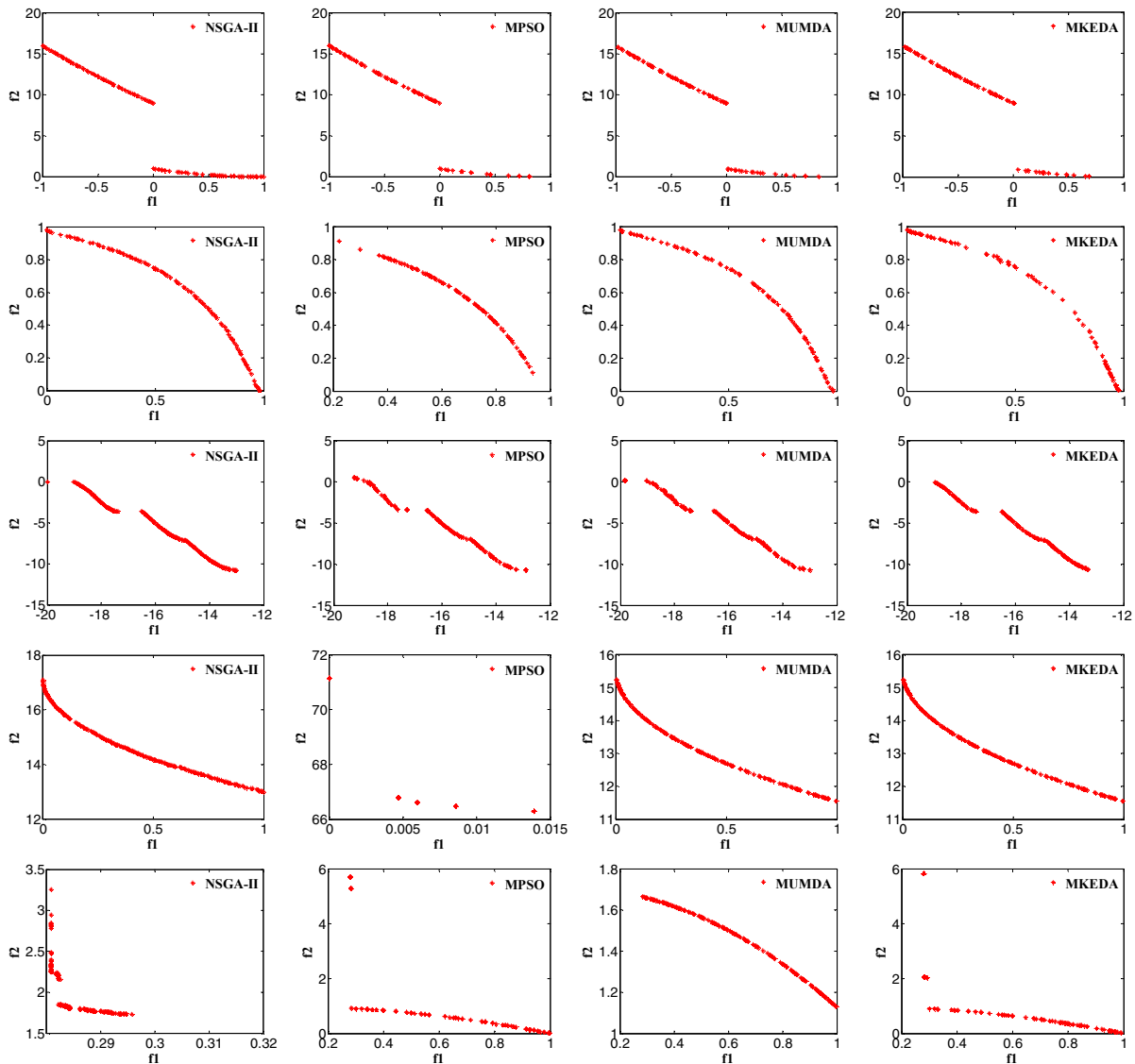


图 3. 算法对 Schaffer, FON, KUR, ZDT4 和 ZDT6 问题寻优的非劣解比较

Fig.3. Obtained non-dominated solutions on Schaffer, FON, KUR, ZDT4 and ZDT6 problems

往是具有多峰性、欺骗性的函数，而且其 Pareto 最优解集也通常含有孤立最优点、凸的、非凸的、非均匀的区域。这些特点都会增加算法完成任务的难度。本文采用 5 个多目标问题优化问题(Schaffer, FON, KUR, ZDT4 和 ZDT6)进行数值实验并比较多目标核密度分布估计算法与 NSGA-II、MPSO、MUMDA 的寻优性能。

对 Schaffer, FON, KUR, ZDT4 和 ZDT6 等问题的仿真结果如图 3 所示，图中从左到右的图像依次为算法 NSGAII, MPSO, MUMDA 和 MKEDA 所求得的非劣解集。对于 Schaffer 问题，取种群规模为 100，运行 50 代后四种算法都能收敛到 Pareto Front 上寻找到 PARETO 最优解，但 NSGAII 的均匀性要优化其他三种算法。对于 FON 问题，取种群规模为 100，运行 100 代，四种算法中，MPSO 的寻优效果较差，这可能是参数设置不合适的缘故。对于 KUR 这样的 PARETO 解不连续的问题，取种群规模为 200，运行 200 代，MKEDA 的结果更好，解的分布均匀分布在了所有片段上。对于问题 ZDT4，这是一个较难解决的问题，取种群规模为 200，运行 200 代，四种算法都未能求得收敛到 ParetoFront 上的解，但是 KMEDA 所求结果收敛性好且在有效的目标空间中均匀分布，其效果与 MUMDA 相同。对问题 ZDT6，取种群规模为 200，运行 200 代，优化结果表明了 MKEDA 的优越性，该算法较其他三种算法能够找到更为好的结果。

从 Schaffer, FON, KUR, ZDT4 和 ZDT6 等问题的寻优结果来看，MKEDA 算法参数设置少，寻优效果比较稳定，对所测试函数均有较好的寻优效果，NSGAII 和 MPSO 算法的寻优效果与参数的设置关系很大，寻优效果不稳定，MUMDA 算法在某些问题上的寻优效果比较差，适应性不如 MKEDA。

5. 结论

为使分布估计算法中的概率密度模型接近于样本实际分布，在求解多目标问题方面，本文给出了一种采用核密度估计的多目标分布估计算法。该方法采用核密度估计算法估计样本的概率密度函数，通过随机遍历抽样方法生成新种群，采用适应值共享得到多目标问题的 PARETO 前沿。同时本文将该算法与 NSGA-II、多目标粒子群算法、多目标分布估计算法对 3 个测试函数进行了性能测试。优化结果表明，基于核密度估计的多目标分布估计算法对多个测试函数具有更好的适应性。

参考文献

[1] K Deb,A Pratap,S Agarwal et al. A fast and elitist muhiobjective

genetic algorithm: NSGA-II. Trans on Evolutionary Computation, vol.6, no.2, pp. 182-197,2002.

- [2] K.E.Parsopoulos, M.N.Varhatis. Particle swarm optimization method in multiobjective problems. In: Proc. ACM Symp on Applied Computing, Madrid, Spain, 2002, pp. 603-607.
- [3] Sujin Bureerat and Krit Sriworamas. Population-Based Incremental Learning for Multiobjective Optimisation. A. Saad et al. (Eds.): Soft Computing in Industrial Applications, ASC 39, 2007, pp. 223-232.
- [4] Laumanns, M. and Ocenasek, J.. Bayesian Optimization Algorithms for Multi-objective Optimization[C]. In Merele Guerv6s, J. J. el al., editors. Proceedings of Parallel Problem Solving from Nature VII, pp. 298-307, 2002.
- [5] Thierens, D. and Bosmann, P. A. N., Multi-Objective Mixture based Iterated Density Estimation Evolutionary Algorithm[C]. In Spector, L. et al., editors, Proceedings of Generic and Evolutionary Computation Conference. pages: 663-670, 2001.
- [6] Costa. M. and Minisci, E.. MOPED: A Multi-objective Parzen based Estimation of Distribution Algorithm for Continuous Problems[C]. In Fonseca C. M. et al.. editors, Proceedings of the Second Internaitional Conference on Evolutionory Multi-Critrion Optimization, pp. 282-294, 2003.
- [7] P. Larrañaga, R. Etxeberria, J.A. Lozano, J.M. Peña, Optimization in continuous domain by learning and simulation of Gaussian networks, in: The Proceeding of the 2000 Genetic and Evolutionary Computation Conference Workshop Program, Las Vegas, Nevada, pp. 201-204, 2000.
- [8] Bin Li, Run-tian Zhong, Xian-ji Wang, Zhen-quan Zhuang. Continuous Optimization based-on Boosting Gaussian Mixture Model[C]. in Proceeding of the 18th International Conference on Pattern Recognition (ICPR 2006), Hong Kong, China,vol.1, pp.1192-1195,2006.
- [9] Shigeyoshi Tsutsui, Martin Pelikan, and David E. Goldberg. Evolutionary Algorithm using Marginal Histogram Models in Continuous Domain. In Proc. of the 2001 Genetic and Evolutionary Computation Conference Workshop Program,2001.
- [10] Ding N, Zhou S D, Sun Z Q. Histogram-Based Estimation of Distribution Algorithm: A Competent Method for Continuous Optimization Histogram-based estimation of distribution algorithm: a competent method for continuous optimization[J]. Journal of computer science and technology, vol. 23, no. 1, pp. 35-43.,2008.