分布估计算法中差分采样策略的研究

答辩人：董兵
导　师：周爱民　研究员

华东师范大学
计算机科学软件工程学院

2017年5月

華東師範大學 EAST CHINA NORMAL UNIVERSITY

研究目的和意义

- 传统的分布估计算法往往需要考虑选择什么样的模型来进行采样
- 通过结合差分演化来进一步提升分布估计算法的采样
- 提出的差分采样策略，对于单目标分布估计算法和多目标分布估计算法性能都有提升

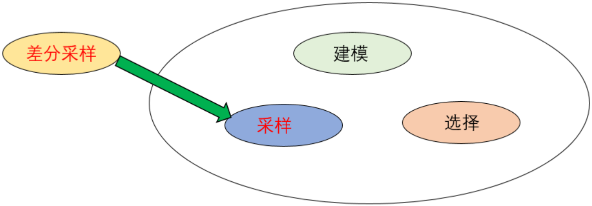


图 1: 差分采样策略

主要工作

- 受差分演化启发，提出一种基于差分演化的采样策略，即差分采样策略。
- 对于单目标优化问题，利用基于特征向量的差分演化去改进采样，通过expensive Local Search进一步改进解集质量，提出**基于差分采样的单目标分布估计算法**。
- 对于多目标优化问题，利用差分采样策略去改进RM-MEDA的采样，提出**基于差分采样的多目标分布估计算法**。

演化算法是一种基于种群的启发式优化算法。演化算法适用于解决多种问题，几乎不需要任何的假设前提条件。

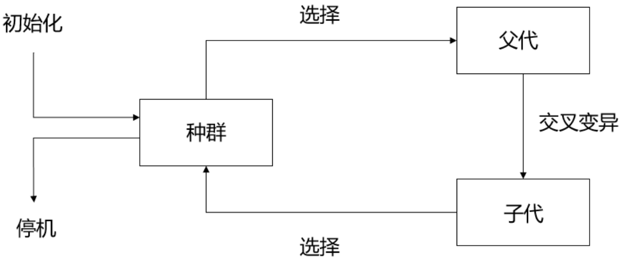


图 2: 演化算法流程图

分布估计算法

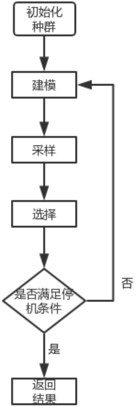


图 3: 分布估计算法流程

- 分布估计算法主要有三个主要步骤：建模、采样、选择
- 传统的分布估计算法通过从建立的概率模型中采样产生新的个体

差分演化

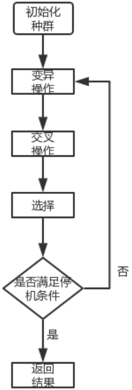


图 4: 差分演化流程

- 变异操作：产生变异向量
  常用的变异策略：
  - DE/rand/1/bin
  - DE/best/1/bin
  - DE/rand/2/bin
  - DE/best/2/bin
  - DE/current-to-best/1/bin
- 交叉操作：利用交叉算子结合变异向量和目标向量来产生试验向量。
- 选择：利用一对一的竞争机制从目标向量和试验向量中挑选个体。

单目标优化问题几乎出现在科学和工程应用的各个领域中。本文研究的单目标优化问题针对的是连续空间的全局优化问题，即是求目标函数的最小值或者最大值。
对于全局优化问题在本文做出以下定义：

$$\begin{aligned} &min \quad f(x) \\ &s.t. x \in [a_i, b_i]^n \end{aligned} \tag{1}$$

其中$x = (x_1, x_2, \cdots, x_n)^T \in R^n$ 是决策变量向量，$[a_i, b_i]^n$是搜索空间区域，$f : R^n \rightarrow R$ 则是目标函数。

主要思路

如何改进分布估计算法的采样

主要思路

? ~~如何改进分布估计算法的采样~~

✓ 利用差分演化去改进采样

主要思路

? ~~如何改进分布估计算法的采样~~

✓ 利用差分演化去改进采样

? 如何进一步提升采样效率

主要思路

? ~~如何改进分布估计算法的采样~~

✓ 利用差分演化去改进采样

? ~~如何进一步提升采样效率~~

✓ 利用DE-EIG来进行采样，避免种群信息丢失

主要思路

- ? ~~如何改进分布估计算法的采样~~
- ✓ 利用差分演化去改进采样
- ? ~~如何进一步提升采样效率~~
- ✓ 利用DE-EIG来进行采样，避免种群信息丢失
- ? 如何进一步优化单目标分布估计算法

主要思路

- ? ~~如何改进分布估计算法的采样~~
- ✓ 利用差分演化去改进采样
- ? ~~如何进一步提升采样效率~~
- ✓ 利用DE-EIG来进行采样，避免种群信息丢失
- ? ~~如何进一步优化单目标分布估计算法~~
- ✓ 利用expensive Local Search进一步提炼解集质量

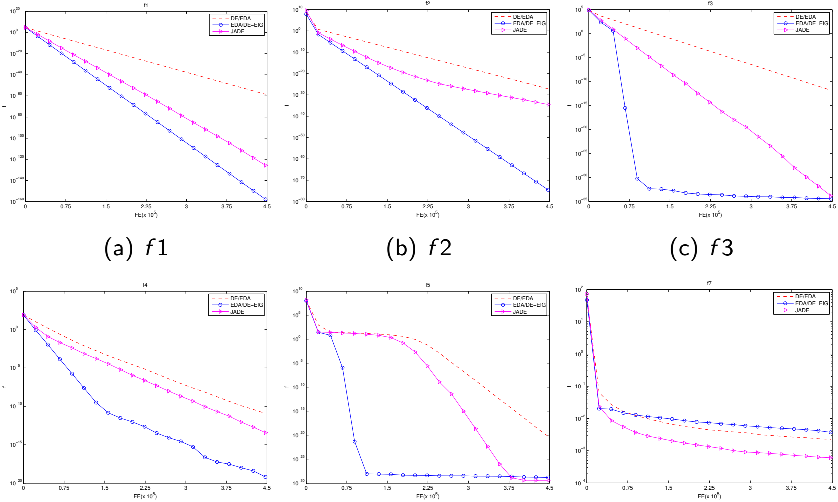
基于差分采样的单目标分布估计算法

初始化种群$Pop(t) = \{x_1, x_2, x_3, \cdots, x_N\}$ (N是种群的大小)
while *not terminate* **do**
  构建概率模型:
  $p(x) = \prod_{i=1}^{n} \mathcal{N}(x_i; \mu_i, \sigma_i)$
  根据下面的流程产生新的实验向量$u_{i,G}$:
  if $rand() < CRP$ **then**
    根据DE-EIG采样得到$u_{i,G}$
  else
    根据概率模型$p(x)$采样得到$u_{i,G}$
  end
  if $f(u_{i,G}) < f(x_{i,G})$ **then**
    $x_{i,G+1} = u_{i,G}$
  else
    $x_{i,G+1} = x_{i,G}$
  end
  if $Coverage(\theta, G, G_e)$ **then**
    执行expensive LS
  end
  $t = t + 1$
end

图 5: EDA/DE-EIG算法

- 在基于DE/EDA算法框架的基础上，利用DE-EIG进行采样，同时结合expensive LS进一步提高解集质量，提出基于差分采样的单目标分布估计算法。
- 为了增加种群的多样性，通过利用随机参数来设置*CRP*，可以提高种群的多样性和算法的鲁棒性。

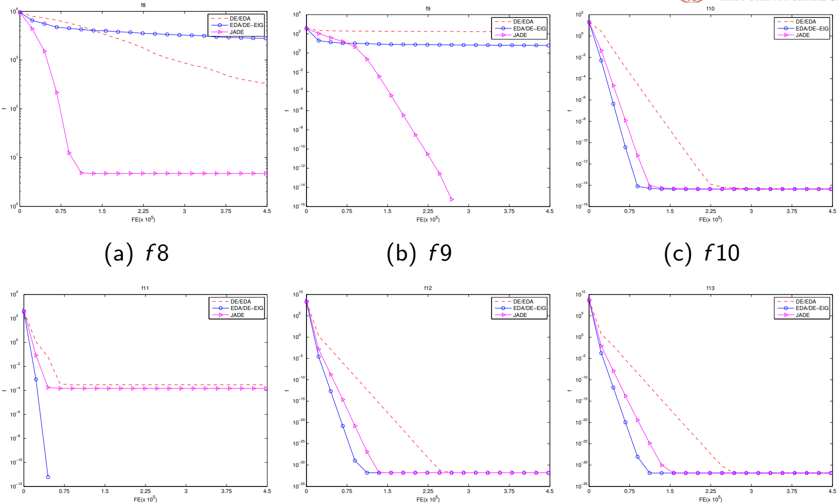
华东师范大学 EAST CHINA NORMAL UNIVERSITY



(a) $f1$　(b) $f2$　(c) $f3$

(d) $f4$　(e) $f5$　(f) $f7$

图 6: 在$f1-f13$中除了$f6$的12个测试题上目标函数平均值的折线图



(a) $f8$ (b) $f9$ (c) $f10$

(d) $f11$ (e) $f12$ (f) $f13$

图 7: 在$f1-f13$中除了$f6$的12个测试题上目标函数平均值的折线图

实验结果分析

- 三个算法在$f6$测试题上都收敛得很早，因此不用折线图展示。
- 和DE/EDA相比，EDA/DE-EIG具有较大的领先，在**10**个测试题上都具有较大幅度优势（除了$f7, f8$）。
- 和JADE相比，在$f1, f2, f3, f4, f11$ **5**个测试题上在下降趋势和最终结果都有明显优势，在$f10, f12, f13$ **3**个测试题上，下降趋势更佳，对于$f5$虽然最终结果不如JADE，但是下降趋势表现得更好，在$f7, f8, f9$ 3个测试题上表现得不如JADE。

	EDA/DE-EIG	JADE	DE/EDA
$f1$	**1.54e-159 ± 5.11e-159**	$3.90e-127 \pm 2.74e-126(+)$	$1.39e-59 \pm 2.58e-59(+)$
$f2$	**1.02e-75 ± 7.46e-76**	$2.60e-35 \pm 1.64e-34(+)$	$5.15e-28 \pm 4.68e-28(+)$
$f3$	**4.01e-35 ± 8.47e-35**	$7.79e-35 \pm 2.51e-34(\sim)$	$1.23e-12 \pm 1.20e-12(+)$
$f4$	**5.01e-20 ± 3.06e-19**	$3.15e-14 \pm 6.42e-14(+)$	$9.90e-12 \pm 2.69e-11(+)$
$f5$	$1.46e-29 \pm 2.62e-29$	**3.85e-30 ± 9.58e-30(−)**	$3.37e-21 \pm 8.66e-21(+)$
$f6$	**0.00e+00 ± 0.00e+00**	**0.00e+00 ± 0.00e+00(∼)**	**0.00e+00 ± 0.00e+00(∼)**
$f7$	$3.60e-03 \pm 1.00e-03$	**6.01e-04 ± 2.23e-04(−)**	$2.20e-03 \pm 5.59e-04(-)$
$f8$	$2.79e+03 \pm 5.02e+02$	**4.74e+00 ± 2.34e+01(−)**	$1.82e+03 \pm 6.72e+02(-)$
$f9$	$6.23e+00 \pm 2.21e+00$	**0.00e+00 ± 0.00e+00(−)**	$1.54e+02 \pm 1.96e+01(+)$
$f10$	**4.44e-15 ± 0.00e+00**	**4.44e-15 ± 0.00e+00(∼)**	**4.44e-15 ± 0.00e+00(∼)**
$f11$	**0.00e+00 ± 0.00e+00**	$1.48e-04 \pm 1.05e-03(\sim)$	$2.96e-04 \pm 1.46e-03(\sim)$
$f12$	**1.57e-32 ± 5.53e-48**	**1.57e-32 ± 5.53e-48(∼)**	**1.57e-32 ± 5.53e-48(∼)**
$f13$	**1.35e-32 ± 1.11e-47**	**1.35e-32 ± 1.11e-47(∼)**	**1.35e-32 ± 1.11e-47(∼)**
		3(+)6(∼)4(−)	6(+)5(∼)2(−)

图 8: 基于差分采样的单目标分布估计算法

- 红色的表示表现最佳的是EDA/DE-EIG
- 蓝色的表示三个算法皆表现最佳
- 黄色的表示JADE结果最好

现实生活中的问题往往具有多个目标，而每个目标之间的最优值往往是冲突的。在本文中，假设多目标问题中的每个问题都是最小化问题，则做出以下定义：

$$\begin{aligned} \min \quad & F(x) = (f_1(x), \cdots, f_m(x)) \\ \text{s.t} \quad & x \in \Omega \end{aligned} \tag{2}$$

其中，$x = (x_1, \cdots, x_n)^T \in R^n$ 是决策变量向量，$\Omega = \Pi_{i=1}^{n}[a_i, b_i] \subset R^n$ 表示可能的搜索空间区域，$f_i : R^n \to R, i = 1, \cdots, m$ 是一个连续的目标函数，$F(x)$则是相应的目标函数向量。

在一般条件下，根据Karush-Kuhn-Tucker条件可以推导出：连续多目标问题在决策空间中的Pareto解集是一个连续分段的（m-1）维的流形体（m是目标数）。对于一个成功的多目标演化算法（multiobjective evolutionary algorithm, MOEA）来说，独立的个体应该是在决策空间中分散在Pareto解集附近，如图 9 所示。

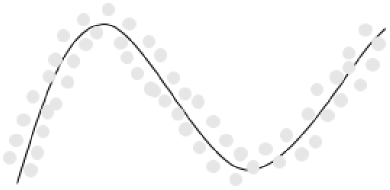


图 9: 连续多目标问题决策空间中个体的分布情况

RM-MEDA算法

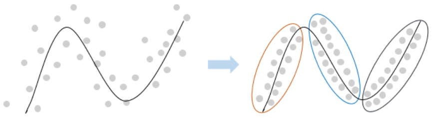


图 10: LPCA划分聚类过程

- 根据连续多目标问题以上的特性，基于规律模型的多目标分布算法（RM-MEDA）算法被提出用于解决连续多目标问题。
- 在每一次迭代中，通过局部主成分分析（LPCA）在决策空间中的区域建立概率分布模型，然后通过使用拉丁实验设计采样得到新的子代种群。

存在的问题

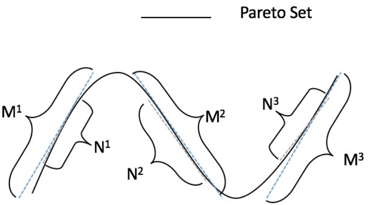


图 11: 通过缩放比例来覆盖Pareto解集

- 在每一个聚类之中，N^k用来表示聚类中的Pareto解集，M^k则用来覆盖每个聚类中的Pareto解集。

? 如何设置一个合适的缩放比例

存在的问题

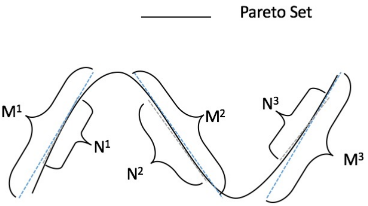


图 12: 通过缩放比例来覆盖Pareto解集

- 在每一个聚类之中，N^k用来表示聚类中的Pareto解集，M^k则用来覆盖每个聚类中的Pareto解集。

? ~~如何设置一个合适的缩放比例~~

✓ 通过差分采样策略来进行采样

存在的问题

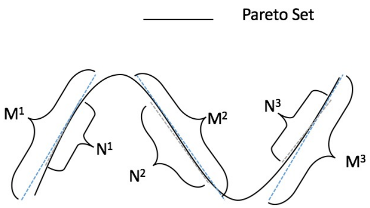


图 13: 通过缩放比例来覆盖Pareto解集

- 在每一个聚类之中，N^k用来表示聚类中的Pareto解集，M^k则用来覆盖每个聚类中的Pareto解集。

? ~~如何设置一个合适的缩放比例~~

? 如何提升种群的多样性

✓ 通过差分采样策略来进行采样

存在的问题

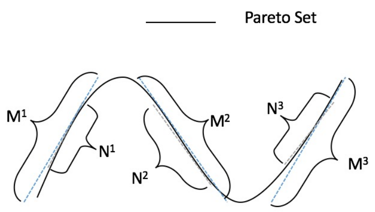


图 14: 通过缩放比例来覆盖Pareto解集

- 在每一个聚类之中，N^k用来表示聚类中的Pareto解集，M^k则用来覆盖每个聚类中的Pareto解集。

? ~~如何设置一个合适的缩放比例~~

? ~~如何提升种群的多样性~~

✓ 通过差分采样策略来进行采样

✓ 提出一种新型的变异策略来进一步提升种群多样性

差分采样策略

1 对于每个给定的聚类求得相应的协方差矩阵C并进行分解操作：

$$C = EDE^T$$

E是协方差矩阵C的特征向量矩阵，D是由特征值组成的对角矩阵。

2 对于聚类中每一个个体x，将其映射到隐空间中：

$$y = x \cdot R.$$

R是特征向量矩阵E中前$(m-1)$个主要成分。

3 在隐空间中对于种群个体进行变异操作：

$$y' = y_{r_1} + rand \cdot (y_{r_2} - y_{r_3}) + F \cdot (y_{r_2} - y_{r_3})$$

4 将y'映射到原始的决策空间

$$x' = y' \cdot R^T.$$

5 返回产生的新的个体

$$x'' = x' + \varepsilon'$$

ε'是一个服从分布$\mathcal{N}(0, \sigma_\tau I)$的高斯噪音（$\tau \in \{1, 2, \cdots, K\}$是一个随机产生的整数）

图 15: 差分采样策略

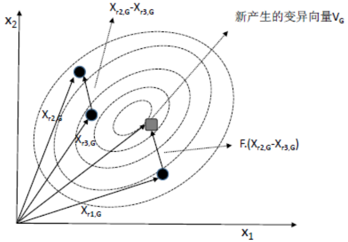


图 16: 变异策略

- 通过特征向量将种群转换到隐空间中，在隐空间中利用新型的变异策略完成变异操作，在将种群映射到原始空间中。
- 利用差分采样策略来取代分布估计算法中原有的建模采样。

基于差分采样的多目标分布计算法

```
初始化一个随机种群Pop(t)，并且设置t为0。
while not terminate do
    建模：通过建立概率模型δ来描述种群中个体的分布情况。
    采样：根据概率模型将种群划分为不同的聚类C_i。对于每一个聚类，分别应
    用DES来产生新的候选集合Q_i，最终生成集合Q = ∪_i Q_i。
    选择：从Q ⋃ Pop(t)中选择N个个体来构建新的种群Pop(t + 1)。
    t = t + 1
end
```

图 17: DES-RM-MEDA

- 利用差分采样策略来改进RM-MEDA的的采样，有效提高算法性能

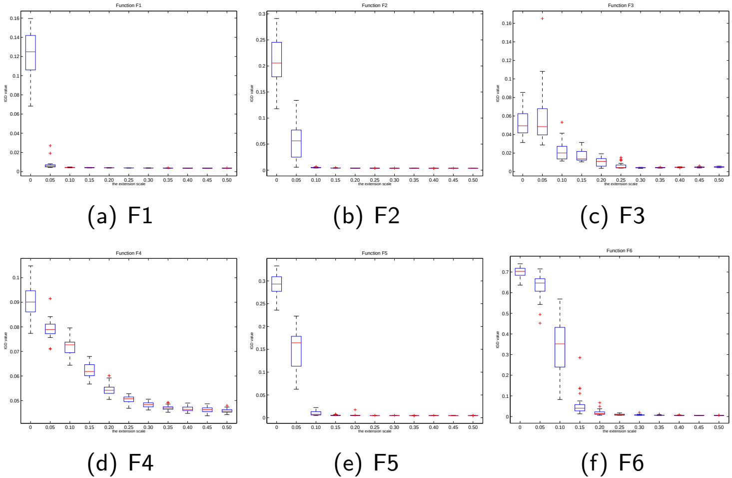
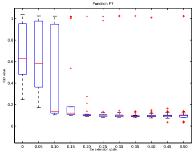
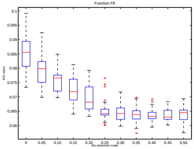


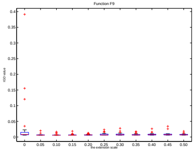
图 18: 不同缩放比例下，RM-MEDA的IGD指标的箱线图



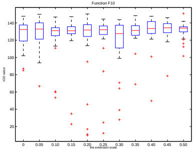
(a) F7



(b) F8



(c) F9

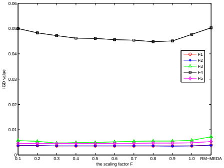


(d) F10

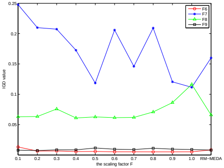
图 19: 不同缩放比例下，RM-MEDA的IGD指标的箱线图

- 如果不设置缩放比例，那么RM-MEDA表现则不是很好
- 缩放比例设置的越大，则RM-MEDA表现得也更加优秀
- 对于设置较大的缩放比例也可能会造成性能的不稳定
- 总的来说，如何在实践中设置一个最佳的缩放比例还是比较困难的。

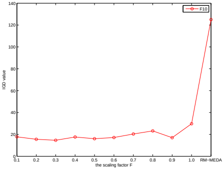
缩放因子F的敏感性



(a)



(b)



(c)

图 20: 不同缩放因子下，DES-RM-MEDA的性能比较

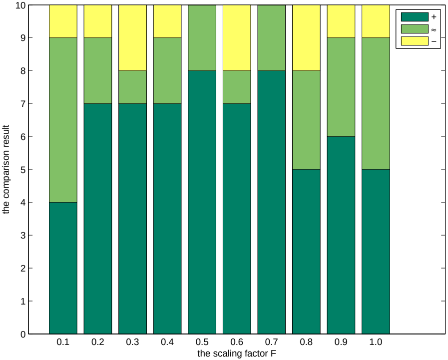
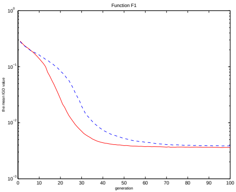
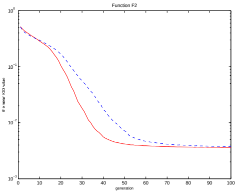


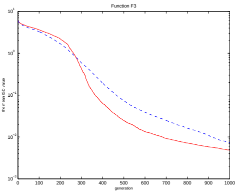
图 21: Wilcoxon’s rank test层叠柱状图



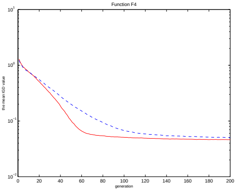
(a) F1



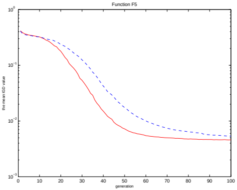
(b) F2



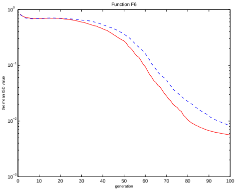
(c) F3



(d) F4

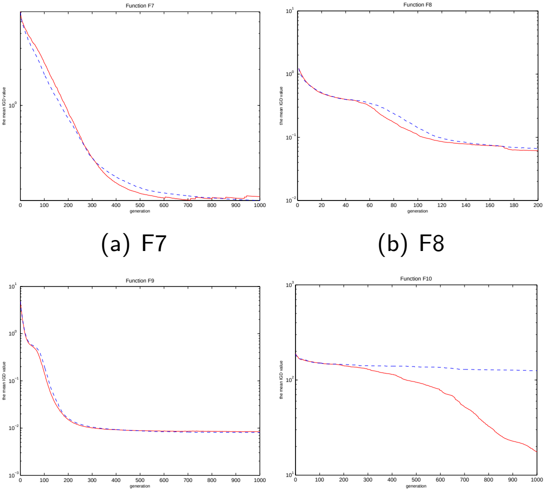


(e) F5



(f) F6

图 22: IGD指标均值趋势图。实线表示DES-RM-MEDA，虚线表示RM-MEDA。



(a) F7 (b) F8

(c) F9 (d) F10

图 23: IGD指标均值趋势图

- 两个目标的测试题，DES-RM-MEDEA在测试题$F1, F2, F3, F6, F10$ 5个测试题上，无论是下降趋势还是最终结果都具有优势。对于$F9$,两个算法的表现几乎一致。在$F7$上，其表现在后阶段表现不太稳定。
- 三个目标的测试题，DES-RM-MEDA在下降趋势和最终结果都处于领先。

- 本文针对分布估计算法中的采样，提出了差分采样策略，对于单目标优化问题和多目标优化问题分别提出了基于差分采样的单目标分布估计算法和基于差分采样的多目标分布估计算法。通过综合的实验分析，差分采样策略对于提高分布估计算法性能具有重大意义。

当然本文的工作还有一些不足之处，因此做以下几点展望：

- 将差分采样策略应用于其它多目标优化算法
- 进一步优化EDA/DE-EIG算法
- 对于DE/EDA算法中，分布估计算法和差分演化的资源分配是一个值得继续探索的话题

Thanks!

- B. Dong, A. Zhou, and G. Zhang, A Hybrid Estimation of Distribution Algorithm with Differential Evolution for Global Optimization, 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016.
- B. Dong, A. Zhou, and G. Zhang, Sampling in Latent Space for a Multiobjective Estimation of Distribution Algorithm, 2016 IEEE Congress on Evolutionary Computation (CEC), 2016.