

基于岛屿群体模型的多目标演化算法研究

赵凤强¹ 徐毅² 李广强³

(大连民族学院机电信息工程学院 大连 116600)¹ (大连理工大学计算机科学与工程系 大连 116023)²
(大连海事大学信息科学技术学院 大连 116026)³

摘要 近年来,基于 Pareto 最优概念的多目标演化算法成为演化计算的研究热点,并已在工程领域中得到了广泛应用。在多目标演化算法 NSGA-II 基础上,给出了并行非劣分层多目标演化算法(PNSMEA)。该算法引入了粗粒度岛屿模型,整个群体被划分成若干个子群体,每个子群体单独演化计算。子群体在进化过程中,每隔一定进化代数交换集合中的个体,以保证各子群体中个体的多样性,提高多目标问题非劣最优域搜索的广度。引入算术交叉算子,以克服 NSGA-II 中 SBX(Simulated Binary Crossover)交叉算子搜索能力较弱的缺点。试验结果表明,PNSMEA 算法不仅可改善 NSGA-II 算法的搜索孤立区域困难和早收敛的问题,而且所获得的 Pareto 解集具有更好的分布性。

关键词 多目标演化算法,NSGA-II,Pareto 解集,岛屿模型

中图分类号 TP301 **文献标识码** A

Research on Multi-objective Evolutionary Algorithm Based on Island Model

ZHAO Feng-qiang¹ XU Yi² LI Guang-qiang³

(College of Electromechanical & Information Engineering, Dalian Nationalities University, Dalian 116600, China)¹

(Department of Computer Science and Engineering, Dalian University of Technology, Dalian 116023, China)²

(College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China)³

Abstract Recently the reaserch on multi-objective evolutionary algorithms based on Pareto optimization concept has become a research hotspot. And it has been widely applied in engineering fields. This paper presented a parallel nondominated sorting genetic multi-objective evolutionary algorithm(PNSMEA) based on NSGA-II. PNSMEA adopes island model and the population is divided into several sub-populations that evolve separately. The sub-populations migrate good individules each other at intervals of some generations, which can keep individules' diversity and broad the search domain of each sub-population. PNSMEA adopes arithmetic crossover operator to overcome the weak search capability of SBX operator used by NSGA-II. The test results show that PNSMEA can not only improve the premature problem as well as the search capability in the isolated regions of NSGA-II but also contribute to obtaining the Pareto solution sets with better distribution.

Keywords Multi-object evolutionary algorithm, NSGA-II, Pareto solution set, Island model

近年来,多目标优化问题求解已成为演化计算的一个重要研究方向,而基于 Pareto 最优概念的多目标演化算法则是当前演化计算的研究热点。从 1985 年至今已出现多种优秀的多目标演化算法,例如多目标遗传算法(Multi-Objective Genetic Algorithm, MOGA)^[1]、小生境 Pareto 遗传算法(Niched-Pareto Genetic Algorithm, NPGA)^[2]、改进的非支配排序算法(Non-dominated Sorting Genetic Algorithm II, NSGA-II)^[3]、改进的强 Pareto 演化算法(Strength Pareto Evolutionary Algorithm, SPEA2)^[4]、Pareto 存档演化策略(Pareto Archived Evolution Strategy, PAES)^[5]、改进的多目标 Messy 遗传算法(Multi-Objective Messy Genetic Algorithm, MOMGA-II)^[6]等。NSGA-II 则是其中较具代表性的算法,试验证明其结果优于其它几种算法^[3]。但 NSGA-II 也存在着不足

之处,例如搜索性能相对较弱、搜索孤立区域存在一定困难等。

针对 NSGA-II 算法存在的不足,本文引入了并行算法中的粗粒度岛屿模型,通过岛屿模型与 NSGA-II 的结合,旨在解决 NSGA-II 算法中搜索孤立区域困难;引入了算术交叉算子,意在提高算法的搜索性能。

1 NSGA-II 算法

1.1 多目标优化问题定义

一个多目标问题可定义如下^[7],其中决策向量 $x \in R^m$,目标向量 $y \in R^n$: $\min y = f(x) = (f_1(x), f_2(x), \dots, f_n(x))$ 。

定义 1(优劣性, Dominance/ Inferiority) 目标向量解 $u = (u_1, u_2, \dots, u_n)$ 优于 $v = (v_1, v_2, \dots, v_n)$, 记为 $u < v$, 当且仅

到稿日期:2010-01-21 返修日期:2010-04-06 本文受国家自然科学基金(60674061)资助。

赵凤强(1972-),男,博士,讲师,主要研究方向为智能系统与优化、故障诊断及噪声控制, E-mail: fqzhao2002@yahoo.com.cn;徐毅(1975-),男,博士生,主要研究方向为智能系统与优化;李广强(1973-),男,副教授,主要研究方向为智能控制等。

当 $\forall i \in \{1, 2, \dots, n\}, u_i \leq v_i \wedge \exists i \in \{1, 2, \dots, n\}, u_i < v_i$.

定义 2(非劣最优解, Pareto Optimal) 决策变量 $x_u \in R^n$ 称为多目标问题的非劣最优解, 当且仅当不存在决策变量 $x_v \in R^n$ 使得相应的目标向量 $v = f(x_v \in R^n) = (v_1, v_2, \dots, v_n)$ 优于 $u = f(x_u \in R^n) = (u_1, u_2, \dots, u_n)$ 。由所有非劣最优解组成的集合称为多目标优化问题的最优解集, 也称为可接受解集或有效解集, 记为 P_{true} 。相应非劣最优解的目标向量称为非支配的(non-dominated), 由所有非支配的目标向量构成多目标问题的非劣最优目标域(Pareto Front), 记为 PF_{true} 。 P_{true} 与 PF_{true} 由具体的多目标优化问题中各目标函数所决定, 是一个确定的不变集合。

1.2 NSGA-II 算法简介

NSGA-II 是 Deb 等人于 2002 年提出的一种多目标演化算法。它是在 NSGA 算法基础上发展起来的, 其特点是: 采用非支配排序方法将种群个体分解为多层的 Pareto 解集, 同层间的个体通过计算拥挤距离判断优劣, 采用精英选择策略。算法的计算复杂度为 $O(MN^2)$ (M 代表目标数量, N 代表种群大小)。NSGA-II 由于精巧的排序方法和优良的性能, 已经成为一种“艺术级”的标准多目标优化算法^[8]。试验结果也证明其计算性能优于其它几种多目标演化算法^[3]。NSGA-II 算法伪代码如图 1 所示^[7]。

```

Initialize Population
Generate random population - size M
Evaluate Objective Values
Assign Rank(level)Based on Pareto Dominance - "sort"
Generate Child Population
  Binary Tournament Selection
  Recombination and Mutation
For i=1 to Number of Generations
  With Parent and Child Population
  Assign Rank(level)Based on Pareto - "sort"
  Generate sets of nondominated fronts
  Loop(inside) by adding solutions to next generation
  starting from the "first" front until M individuals found
  determine crowding distance between points on each front
  Select points(elitist)on the lower front with lower rank
  and are outside a crowding distance
  Create next generation
  Binary Tournament Selection
  Recombination and Mutation
  Increment generation index
End of Loop
  
```

图 1 NSGA-II 算法伪代码

2 并行非劣分层多目标演化算法(PNSMEA)

NSGA-II 也存在着几点不足之处: ①采用模拟二进制交叉算子(Simulated Binary Crossover, SBX), 搜索性能相对较弱, 使得 NSGA-II 算法在收敛速度方面尚不能令人满意^[9]。②搜索孤立区域存在一定困难^[10]。针对 NSGA-II 算法存在的不足, 较多学者采用了并行遗传算法中的岛屿模型与 NSGA-II 相结合的方式来提高算法性能^[11-13]。这些方法采用的基本方式是在不同的岛屿上预先设置不同的搜索方向和搜寻区域, 每个岛屿搜索一个目标。由于每个岛屿只专注于搜索一个目标, 因而可使整个算法具有更好的搜索能力和更快的收敛速度。试验结果也表明了这一点。但这种方式的缺点是预先要知道算例的 Pareto 最优解形状, 当算例的 Pareto 最优解形状未知的时候, 算法很难预先定义好每个岛屿的搜索方向和区域^[11]。

针对上述存在的问题, 本文提出了一个并行非劣分层多目标演化算法(Parallel Non-dominated Sorting Genetic Algorithm), 算法将并行演化算法中的岛屿模型与 NSGA-II 相结

合, 每个岛屿维护一个单独进化的子种群, 每个子种群都对整个目标区域进行搜索, 而不是搜索不同方向和区域, 岛屿模型中每一个子群体都以不同的杂交率和变异率进行演化, 每隔一定代数子群体之间相互交换最优解, 各个子群体之间优选个体的相互迁移保证了各子群体中个体的多样性, 达到了不同控制参数下选择压力之间的有效平衡; 每个子种群的交叉概率和变异率不相同, 且有不同的搜索空间, 杂交率和变异率小的子群体可维持整个群体的稳定性, 而杂交率和变异率大一些的子群体能够搜索到更好的解, 通过改变每个岛屿模型种群的交叉概率和变异率使得整个群体稳定快速地向 Pareto 非劣最优目标域逼近。

虽然同每个岛屿只搜索一个目标相比, 本文的方式降低了搜索效率, 但却更为通用, 算法并不需要知道算例的 Pareto 最优解形状, 也无需预先设定每个岛屿的搜索区域和方向。算法通过控制各个岛屿不同的杂交率和变异率以及最优个体的相互迁徙来保证群体中个体的多样性, 从而提高算法的搜索性能。

2.1 其它操作算子的改进

NSGA-II 采用 SBX 交叉算子。SBX 算子模拟二进制交叉算子的过程, 对实数编码的父个体进行交叉操作, 即对于给定的随机交叉点, 交换两个父个体位于交叉点两侧的部分。试验证明这种算法搜索性能相对较弱^[9]。为了改善算法的搜索性能, 本文采用多父体算术交叉算子, 其已被证明对采用实数编码的进化算法非常有效, 能够克服 SBX(Simulated Binary Crossover)交叉算子搜索能力较弱的缺点^[14]。多父体算术交叉算子描述如下: 进行交叉操作时, 从种群中随机取 t ($t < N$) 个个体, 设分别为 $X_1^t \sim X_t^t$, 则子代个体 X^c 计算方法如下: $X^c = \sum_{k=1}^t r_k \times X_k^t$, 其中 $r_k \in [0, 1]$ 是随机数, 且 $\sum_{k=1}^t r_k = 1$ 。

NSGA-II 采用多项式变异算子, 本文则采用均匀变异和高斯变异相结合的变异算子。在算法的前期采用均匀变异有利于开扩新的搜索空间, 算法后期采用高斯变异则有利于局部搜索。

2.2 PNSMEA 算法实现描述

本文提出的 PNSMEA 算法是在 NSGA-II 算法的基础上增加了并行算法中的岛屿模型。在算法实现过程中, 整个群体被划分成若干子群体, 每个子群体称为一个岛屿, 所有岛屿上的子群体构成了整个群体。每一个子群体都采用基本 NSGA-II 算法并以不同的杂交率和变异率进行演化, 每隔一定代数交换子群体中的精英个体。算法实现时采用的是伪并行遗传算法, 即所有的操作都是在一个处理机上串行执行。每个子种群在一个独立的线程中进化计算, 各个子群体周期地将精英个体发送给主线程。主线程接受子群体的精英个体, 然后根据预先设定的迁徙模型发送给其它子群体, 线程之间的数据交换采用 Windows 系统提供管道技术实现。PNSMEA 算法实现模型如图 2 所示。

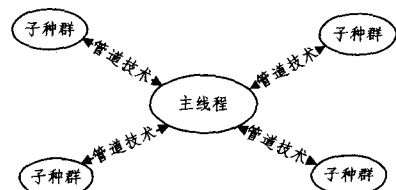


图 2 PNSMEA 算法实现模型

每个子种群的执行步骤如下。

参数: N (群体大小), T (运行最大代数), G_n (迁徙个体数目), G_r (迁徙频率)。

Step1 产生初始种群, 计算每一层个体的拥挤距离。

Step2 对种群进行分层。

Step3 联赛选择操作。

Step4 交叉、变异操作, 产生子代。

Step5 联合亲代和子代种群产生新一代种群。

Step6 从新一代种群中选出精英个体发送到主线程, 由主线程发送给其它子群体。

Step7 接收主线程发送来的个体, 运用联赛选择决定新一代种群哪些个体被替换。

Step8 $t=t+1$, 转 Step2。

3 仿真实例

3.1 测试函数

多目标求解过程中, Pareto 最优阵面的某些特征(例如凸或非凸的、离散的、非均匀分布的等)会不利于算法的求解。针对具有不同最优阵面的 Pareto 求解问题, Zitzler 等人设计出了多个相应的测试函数。本文选取其中有代表性的 3 个^[15]作为本文算法的测试函数, 其具体描述如表 1 所列。

表 1 测试函数

测试函数	n 变量边界	目标函数	最优解	注解
ZDT1	30 [0,1]	$f_1(x) = x_1$ $f_2(x) = g(x)[1 - \sqrt{x_1/g(x)}]$ $g(x) = 1 + 9(\sum_{i=2}^n x_i)/(n-1)$	$x_{i1} \in [0,1]$ $x_i = 0, i=2, \dots, n$	有凸面
ZDT2	30 [0,1]	$f_1(x) = x_1$ $f_2(x) = g(x)[1 - (x_1/g(x))^2]$ $g(x) = 1 + 9(\sum_{i=2}^n x_i)/(n-1)$	$x_{i1} \in [0,1]$ $x_i = 0, i=2, \dots, n$	无凸面
ZDT3	30 [0,1]	$f_1(x) = x_1$ $f_2(x) = g(x)[1 - \sqrt{x_1/g(x)} - \frac{x_1}{g(x)} \sin(10\pi x_1)]$ $g(x) = 1 + 9(\sum_{i=2}^n x_i)/(n-1)$	$x_{i1} \in [0,1]$ $x_i = 0, i=2, \dots, n$	有凸面, 不连接

3.2 算法测试度量值选取

本文的 PNSMEA 算法旨在解决 NSGA-II 算法搜索孤立区域困难和收敛速度较慢的问题, 因而本文主要从算法的收敛性和解的分布性来度量算法性能。采用 Zitzler 的 hyper-volume 度量值^[16]计算 Pareto 解的收敛性, 采用 Deb 的 Δ 度量值^[3]计算 Pareto 解的分布性指标。由于本文的 3 个测试函数都是最小化问题, 根据 hyper-volume 和 Δ 定义可知, hyper-volume 和 Δ 值越小表明算法越优。此外度量值还包括算法执行的时间。

3.3 参数设置

对于不同的算法和测试函数, 每次模拟运行都使用下面的参数。NSGA-II: 运行代数 $Gen_{max} = 750$, 群体规模 $N = 100$, 交叉概率 $P_c = 0.9$, 变异概率 $P_m = 1/n$ (n 为决策变量维数), 模拟二进制交叉参数 $\eta_c = 20$, 多项式变异参数 $\eta_m = 20$, $\beta = 1$ 。PNSMEA: 岛屿个数 = 3, 子种群运行代数 $Gen_{max} = 250$, 子群体规模 $N = 100$, 各个子群的交叉概率 $P_c = 0.7, 0.8, 0.9$, 各子种群的变异率 $P_m = 0.01, 0.02, 0.03$, 迁徙周期 = 10, 迁徙规模 = 50, 迁徙模型采用的是环状拓扑模型, 即子种群的

个体迁徙只发生在相邻的子种群上。

3.4 测试条件

PNSMEA 算法和 NSGA-II 算法的实现采用 VC++6.0 语言编译, 操作系统为 Windows2000, 硬件运行环境: P4 2.0 GHz, 内存 512M。

3.5 算法测试结果

对于 3 个测试问题, 每种算法各自运行了 5 次。表 2 为两种算法的度量值比较。图 3—图 5 是两种算法对 3 个测试函数所求得的最佳未被支配解集在目标空间的表示。从表 1 试验结果可以看出, 在 3 个测试函数中, 对于本文的度量值, PNSMEA 都要优于 NSGA-II 算法。对于 ZDT1 的凸 Pareto 的非劣最优目标区域、ZDT2 的凹 Pareto 的非劣最优目标区域、ZDT3 的离散 Pareto 区域, 从图 3—图 5 也可看出, PNSMEA 算法在解的多样性和解沿着 Pareto 非劣最优目标区域的分布性上, 要比 NSGA-II 算法更优一些。

表 2 测试结果

测试函数	演化算法	Hyper-volume		Δ		时间(s)	
		均值	方差	均值	方差	均值	方差
ZDT1	PNSMEA	0.3651315490	0.000147	0.478868	0.00441	41.109	2.212
	NSGA-II	0.404883011	0.000191	0.59709	0.00645	95.937	3.346
ZDT2	PNSMEA	0.7303524184	0.000103	0.44783	0.00520	42.094	1.355
	NSGA-II	0.7763868883	0.000164	0.60341	0.00732	87.187	3.443
ZDT3	PNSMEA	0.2129300540	0.000131	0.28201	0.00217	49.281	1.621
	NSGA-II	0.2480549939	0.000223	1.08581	0.00356	77.922	2.782

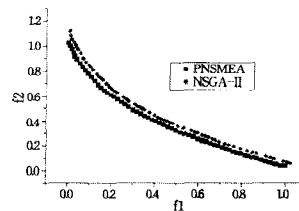


图 3 ZDT1 解集

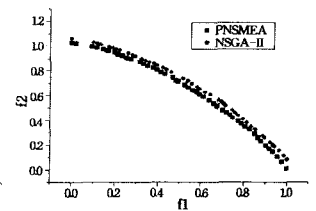


图 4 ZDT2 解集

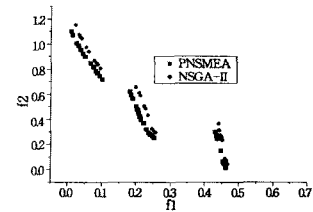


图 5 ZDT3 解集

在多目标演化算法中, 种群大小的选择是一个重要的问题, 它直接影响了多目标演化算法收敛到 Pareto 最优阵面的能力。小的种群不能够保证个体的多样性, 然而无限增加群体的大小并不能在性能上有所改进, 因为大群体意味着更大的计算开销。在 PNSMEA 算法中, 采用并行机制很好地解决了种群大小和计算开销这两者之间的矛盾, 既保证了最终 Pareto 最优解集较好的分布性和延展性, 也减少了算法的计算开销。

结束语 本文在 NSGA-II 算法基础上, 通过引入岛屿群体模型, 给出了并行非劣分层多目标演化算法 (PNSMEA), 并对某些算子做了改进。通过测试函数试验结果的比较表明, PNSMEA 算法在求解 Pareto 解集的收敛性和分布性方面, 要优于 NSGA-II 算法。应该说明, 这是有限测试函数的试验结果。

(下转第 205 页)

于所有的其他分类方法。应该说,较之其他几种数据集,对 NCI 数据集的分类更具有挑战性。NCI 数据集有 9 个类别且总共只有 60 个样本,每类样本的样本数只有 2~9 个。在这样一个小样本数据集上,基于复杂数学模型分类方法均不可避免地存在“过拟合”现象。虽然有研究表明 SVM 分类方法很适合微阵列数据分类处理,但在多类别、超小样本数据集上,本文的 PC 分类法表现出了一定的优异性。

结束语 本文提出了一种基于主曲线的微阵列数据分类方法,通过计算测试样本与各类样本主曲线的方差距离来确定其类别。实验结果表明,在处理高维小样本数据时,本文的方法较之其他几种主流分类方法有一定的优势。微阵列数据是典型的高维、小样本数据,其高维特性可以通过基因选择方法来进行降维处理,但其小样本特性却是我们在分类器设计时必须面对的问题,所以,本文的分类方法是一个值得继续深入研究的课题。

参 考 文 献

[1] Pochet N, De Smet F, Suykens J A, et al. Systematic benchmarking of microarray data classification; assessing the role of non-linearity and dimensionality reduction[J]. *Bioinformatics*, 2004, 20(17):3185-3195

[2] Piatetsky-Shapiro G, Tamayo P. Microarray data mining; facing the challenges[J]. *ACM SIGKDD Explorations Newsl etter*,

2003, 5(2):1-5

[3] Li J, Ng S, Wong L. Bioinformatics adventures in database research[C]//*Proceedings of the international conference on database theory (ICDT)*. 2002:31-46

[4] Cho S-B. Exploring features and classifiers to classify gene expression profiles of acute leukemia[J]. *Int. J. Pattern Recogn. Artif. Intell.*, 2002, 16(7):1-13

[5] Eisen M B, Brown B O. DNA arrays for analysis of gene expression[J]. *Methods Enzymol*, 1999, 303:179-205

[6] Mallick B K, Ghosh D, Ghosh M. Bayesian classification of tumors using gene expression data[J]. *Journal of the Royal Statistical Society*, 2005, B 67:219-232

[7] Quinlan J R. *C4. 5; Programs for Machine Learning*[M]. San Mateo, California: Morgan Kaufmann Publishers, 1993

[8] A novel ensemble of classifiers for microarray data classification

[9] Hastie T, Stuetzle W. Principal curves[J]. *J. A. Stat. A*, 1988, 84:502-516

[10] Hastie T. Principal curves and surfaces[D]. Stanford University, 1984

[11] Kegil B, Krzyzak A, Linder T, et al. Learning and Design of Principal Curves[J]. *IEEE Trans. Pattern. Anal. Mach. Intell.*, 2000, 22(3):281-297

[12] Wang Haonan, Lee Thomas C M. Extraction of Curvilinear Features from Noisy Point Patterns using Principal Curves[J]. *Pattern Recognition Letters*, 2008, 29:2078-2084

(上接第 192 页)

参 考 文 献

[1] Fonseca C M, Fleming P J. Genetic algorithms for multiobjective optimization; formulation, discussion and generation[A]//*Proceedings of the 5th International Conference on Genetic Algorithms*[C]. San Mateo, California, 1993:416-423

[2] Horn J, Nafpliotis N, Goldberg D E. A Niche Pareto Genetic Algorithm for Multiobjective Optimization[A]//*IEEE World Congress on Computational Intelligence*[C]. Piscataway, New Jersey, 1994:82-87

[3] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm; NSGA-II[J]. *IEEE Transactions on Evolutionary Computation*, 2002, 6(2):182-197

[4] Zitzler E, Laumanns M, Thiele L. SPEA2: Improving the Strength Pareto Evolutionary Algorithm[R]. Switzerland, May 2001

[5] Knowles J D, Corne D W. Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy[J]. *Evolutionary Computation*, 2000, 8(2):149-172

[6] Van Veldhuizen D A, Lamont G B. Multiobjective Optimization with Messy Genetic Algorithms[A]//*Proceedings of the 2000 ACM Symposium on Applied Computing*[C]. Villa Olmo, Como, Italy: ACM, 2000:470-476

[7] Coello Coello C A, Van Veldhuizen D A, Lamont G B. *Evolutionary Algorithms For Solving Multi-objective Problems*[M]. New York: Kluwer Academic/Plenum, 2002

[8] Coello Coello C A. Evolutionary multi-objective optimization; a

historical view of the field[J]. *Computational Intelligence Magazine*, IEEE, 2006, 1(1):28-36

[9] 刘旭红, 刘玉树, 张国英, 等. 多目标优化算法 NSGA-II 的改进[J]. *计算机工程与应用*, 2005(15):73-75

[10] Ciekki Coello C A. List of References on Evolutionary Multiobjective Optimization[EB/OL]. <http://www.lania.mx/~ccoella/EMOObib.html>

[11] Branke J, Schmeck H, Deb K, et al. Parallelizing Multi-objective Evolutionary Algorithms; Cone Separation[A]//*Congress on Evolutionary Computation(CEC'2004)*[C]. IEEE, 2004

[12] Deb K, Zope P, Jain A. Distributed computing of pareto-optimal solutions with evolutionary algorithms[J]. *Evolutionary Multi-Criterion Optimization*, 2003, 2632:534-549

[13] Ishibuchi H, Narukawa K. Spatial Implementation of Evolutionary Multiobjective Algorithms with Partial Lamarckian Repair for Multiobjective Knapsack Problems[A]//*Proceedings of the Fifth International Conference on Hybrid Intelligent Systems*[C]. IEEE, 2005

[14] 周育人, 李元香, 王勇. 一种有效的实数编码遗传算法[J]. *武汉大学学报:理学版*, 2003, 49(1):39-43

[15] Zitzler E, Deb K, Thiele L. Comparison of multiobjective evolutionary algorithms; Empirical results[J]. *Evol. Comput*, 2000, 8(2):173-195

[16] Zitzler E, Thiele L. Multiobjective evolutionary algorithms; A comparative case study and the strength pareto approach[J]. *IEEE Transactions on Evolutionary Computation*, 1999, 3(4):257-271