

# 基于 NSGA-II 的进化算法学习研究

王宇, 张闻强, 赵文君

(河南工业大学粮食信息处理与控制教育部重点实验室 河南 郑州 450001)

【摘要】多目标优化问题又称多标准优化问题,进化算法作为解决多目标问题被广泛的应用,在实际应用中经常会遇到众多多目标问题如车间调度、轮船调度等。为使踏入进化计算领域的初学者提供一个简要的学习样例,降低学习的难度。简要介绍多目标进化算法 NSGA-II,叙述其改进并对进化算法在多目标优化中应用的几个测试问题结果进行分析。使研究人员对 NSGA-II 以及其他进化算法有一定理解。在进一步的深度算法学习,解决现实问题上埋下理论基础。

【关键词】多目标优化; 进化算法学习; 精英策略; 拥挤度; NSGA-II

## 1. 第一代进化算法

起初进化算法的灵感来自于 1989 年 Goldberg 给出的建议:引入非支配排序与小生境技术。后代学者扩充内容,进行完善,主要步骤为先确定非支配解,然后对每个解分配一个很大的虚拟适应度值,同时这些非支配解通过这个虚拟适应度值的共享来保持种群的多样性<sup>[1]</sup>。然后这些非支配解作为第一批个体暂时储存,从余下的种群中确定第二批非支配个体,它们被分配一个比第一批解中最小适应度值相比更小的虚拟适应度值,继续储存。依次生成余下各批次个体,该过程循环执行直到整个种群被划分为若干等级为止。以 NSGA 进化算法为例,其主要特点为:

1.1 非支配排序的高计算复杂度。目前所使用的非支配排序算法其计算复杂度为  $O(MN^3)$ ( $M$  为目标对象,  $N$  为种群大小)。可以看出,NSGA 计算量为指数级。无论是高目标还是大种群,计算花费都较大。

1.2 缺少精英主义。最近的研究结果证实精英策略可以加快遗传算法的性能,而且也可以防止在每一代进化的过程中最优解的丢失。

1.3 指定共享参数。为了保持解的多样性,希望在每一代种群生成中进行信息的共享。设置一个共享的规范参数(QSR),其帮助在父代个体交配生成子代个体时能够有效的遗传优良信息。起初为静态设置,在发展的过程中,慢慢的引入自适应的动态操作<sup>[2]</sup>。

## 2. 第二代进化算法

从 20 世纪末开始,精英机制被证明可以有效的提升算法的收敛性,以此为标志,加入精英保留策略的多目标优化算法被称为第二代进化算法<sup>[3]</sup>。已经广泛应用的 SPEA-II,PAES-II 等算法都是基于精英保留机制流行起来的。尤其是在 2002 年,Deb 等人对算法 NSGA 的进行改进,提出第二代进化算法 NSGA-II<sup>[4]</sup>。经过多年发展,它已经成为是最优秀的进化多目标优化算法之一。其改善主要是针对如上所述的三个方面:引入快速非支配排序方法,不仅有效降低了计算复杂度,而且通过将父代与子代种群进行结合,下一代种群从双倍种群集合中择优选择,一定程度上表现出解的广泛性,增加精英策略,保证优良的个体在每一代进化过程中能够保留下来,并参加每一代的交叉变异,有效的提高了收敛速度,最重要的是加入了拥挤

度比较算子,个体属性增加拥挤度参数,通过运算克服了人为指定共享指标的缺陷,而且通过每一代的运算可以自适应的更改,其作为种群中个体间的比较标准,使得准 Pareto 域中的个体能均匀地扩展到整个 Pareto 域,保证了种群的多样性<sup>[1]</sup>。

基于 Pareto 分级的快速非支配排序方法将计算复杂度由  $O(MN^3)$  降到  $O(MN^2)$ <sup>[5]</sup>。引入精英策略与拥挤度后子代与父代通过竞争,不仅保护了优良的个体能够持续生存,而且提高种群解的分布性,提高了整体进化水平。NSGA-II 是第二代进化多目标优化的主要经典算法之一。

## 3. 算法测试

算法的提出是为了解决实际问题。进化算法灵感来自与达尔文的进化论思想,基于数学模型模拟生物进化中的种群,变异,交配等过程,是一种迭代算法。但在现实问题构建数学模型之前,要对算法本身进行性能测试<sup>[6]</sup>。本文所用的测试函数为当今广泛使用的 3 个 ZDT 问题,3 个 DTLZ 问题。为了接下来做 NSGA-II 算法对比的时候更加便利,使用一个集成的可以更方便的评价算法的平台是非常重要的,JMetal 是一款基于 Java 语言开发的针对多目标优化算法集成包。其结构清晰,算法丰富,囊括当今大部分进化算法。CPU:酷睿 4 核 I5-3.30Hz,内存 4G。为了加深对算法的理解,设置不同的 4 种参数,具体请参照表 1。

表 1 参数设置

参数项	NSGA-IIa	NSGA-IIb	NSGA-IIc	NSGA-IId
种群	100	100	100	100
交叉率	1.0	0.8	0.5	0.2
变异率	1/n	1/n	1/n	1/n
交叉分布指数	15	15	15	15
变异分布指数	20	20	20	20

检验进化计算的好坏通常以其收敛到 Pareto 前沿面的速度,局部搜索,全局搜索,解的广泛性等为标准。在数学模型上为了直观的分析算法,引入多目标优化器的性能评估指数:EPSILON,HV,IGD,SPREAD<sup>[7]</sup>。分别对比均值与标准差,该界面效果表 2 所示,中位数与间距范围(IQR),效果图如表 3,在此我们以 HV 为例。

表2 HV. Mean and standard deviation

	NSGA-IIa	NSGA-IIb	NSGA-IIc	NSGA-IId
ZDT1	6.64e-01 <sub>2.7e-04</sub>	6.63e-01 <sub>3.5e-04</sub>	6.62e-01 <sub>5.2e-04</sub>	6.56e-01 <sub>1.2e-03</sub>
ZDT2	3.27e-01 <sub>3.1e-04</sub>	3.26e-01 <sub>3.7e-04</sub>	3.24e-01 <sub>6.3e-04</sub>	3.16e-01 <sub>6.7e-03</sub>
ZDT4	6.56e-01 <sub>3.5e-03</sub>	6.53e-01 <sub>1.5e-02</sub>	6.47e-01 <sub>1.9e-02</sub>	6.12e-01 <sub>3.3e-02</sub>
DTLZ1	6.36e-01 <sub>2.3e-01</sub>	6.17e-01 <sub>2.3e-01</sub>	5.04e-01 <sub>2.5e-01</sub>	2.11e-02 <sub>6.2e-02</sub>
DTLZ2	4.40e-01 <sub>4.5e-03</sub>	4.40e-01 <sub>5.5e-03</sub>	4.38e-01 <sub>5.8e-03</sub>	4.32e-01 <sub>6.8e-03</sub>
DTLZ6	3.37e-01 <sub>5.7e-02</sub>	2.33e-01 <sub>6.0e-02</sub>	6.98e-02 <sub>3.4e-02</sub>	5.53e-04 <sub>1.9e-03</sub>

表3 HV. Mean and standard deviation

	NSGA-IIa	NSGA-IIb	NSGA-IIc	NSGA-IId
ZDT1	3.55e-01 <sub>2.8e-02</sub>	3.97e-01 <sub>3.6e-02</sub>	4.76e-01 <sub>3.6e-02</sub>	5.29e-01 <sub>4.0e-02</sub>
ZDT2	3.49e-01 <sub>2.7e-02</sub>	4.02e-01 <sub>3.2e-02</sub>	4.88e-01 <sub>3.6e-02</sub>	5.53e-01 <sub>5.0e-02</sub>
ZDT4	3.42e-01 <sub>3.0e-02</sub>	4.34e-01 <sub>5.7e-02</sub>	5.69e-01 <sub>6.3e-02</sub>	7.90e-01 <sub>1.4e-01</sub>
DTLZ1	8.70e-01 <sub>1.7e-01</sub>	9.30e-01 <sub>2.0e-01</sub>	1.00e-00 <sub>1.9e-01</sub>	1.12e-00 <sub>1.0e-01</sub>
DTLZ2	6.99e-01 <sub>5.0e-03</sub>	7.08e-01 <sub>4.3e-02</sub>	7.22e-01 <sub>4.6e-02</sub>	7.57e-01 <sub>5.2e-02</sub>
DTLZ6	7.49e-01 <sub>6.3e-02</sub>	8.06e-01 <sub>5.0e-02</sub>	8.26e-01 <sub>5.3e-02</sub>	8.47e-01 <sub>5.2e-02</sub>

在大数据的处理中,通常会有异常数据存在,我们称之为离散点。去对整体数据结果分析往往起到反作用,对结果会带来不良影响。为此我们要找出异常数据,分析其产生的原因,常常成为发现问题进而改进决策的契机。箱形图为我们提供了识别异常值的一个标准<sup>[8]</sup>。在一个样本、种群或概率分布中,我们用中位数来代表其中的一个数值或个体,它将种群集合划分为相等的上下两部分。它在所有标志值中所处的位置确定的全体单位标志值的代表值,不受分布数列的极大或极小值影响,从而在一定程度上提高了中位数对分布数列的代表性。同样的对EPSILON,HV,IGD,SPREAD指数进行分析,我们以IGD为例,如图1。

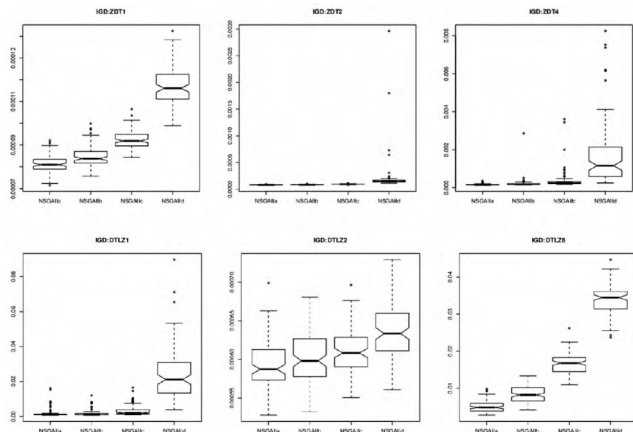


图1 Problems.IGD.Boxplot

每个时期都受着当代思想发展的制约,在第二代进化算法研发至今,越来越多的问题凸显:如何将复杂的现实问题建立数学模型,如何面对高维多目标问题,如何进一步降低计算复杂度以及一直存在争议的停止标准。在接下来的发展中学者们仍待研究<sup>[5]</sup>。

#### 4. 结束语

本文就多目标算法NSGA-II的基本框架进行概述,其目的

主要为刚踏入多目标进化领域的研究人员提供一个不仅学习NSGA-II 还有其他多目标进化算法的学习引导。起初,简单介绍了第一代与第二代多目标优化领域的经典进化算法,其中重点分析两代之间的改进方法,引入了现今流行的优化思想策略。中间,作者以学习NSGA-II 为例,从算法思想,参数设置,运行以及结果分析上对进化算法进行一个清楚全面的解析。最后,就现如今进化算法的瓶颈待解决的问题进行简单的阐述,并对以后可能要发展的方向进行展望<sup>[9]</sup>。

进化算法NSGA-II 作为经典的多目标优化算法之一,在很多领域得到应用,是众多刚接触多目标研究领域的科研人员学习的第一个算法。对于不清楚多目标优化领域的科研人员来说,算法思想,参数如何设置,结果如何进行分析让很多人步履艰难。为此,本文用大幅度的篇幅介绍NSGA-II 算法,使研究人员对初步的学习阶段可以很快的对多目标进化算法领域有一个清楚的认识。本文在进化算法的学习初起到这很好的引导作用,但学习算法的目的是为了可以在其基础上结合现实目标问题进行改进,解决现实问题,故研究人员以后的道路上需要深入学习。

#### 参考文献:

- [1]公茂果,焦李成,杨咚咚,马文萍.进化多目标优化算法研究[J].软件学报,2009,02: 271-289.
- [2]杨善学.解决多目标优化问题的几种进化算法[D].西安电子科技大学,2007.
- [3]谭艳艳.几种改进的分解类多目标进化算法及其应用[D].西安电子科技大学,2013.
- [4]Kalyanmoy Deb, Sameer Agarwal, T. Meyarivan, AmritPratap. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II [J]. IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION,2002,2 (6): 182-197.
- [5]肖晓伟,肖迪,林锦国,肖玉峰.多目标优化问题研究概述[J].计算机应用研究,2011,03: 806-827.
- [6]许婧祺.多目标优化算法研究综述[J].科技信息,2010,32: 115-116.
- [7]韩丽霞.求解多目标优化问题的新遗传算法[J].计算机科学,2003,01: 64-66,95.
- [8]王怀亮.箱须图在识别统计数据异常值中的作用及R语言实现[J].商业经济,2011,05: 64-65.
- [9]PINED M.O, 张智海译[M].调度:原理、算法和系统.北京:清华大学出版社,2007: 1-7.

#### 作者简介:

王宇(1990~),男(汉族),河南驻马店,河南工业大学信息学院研究生,在读硕士,研究方向:多目标优化、算法设计与分析。张闻强(1975~),男(汉族),河南新乡,河南工业大学信息学院副教授,博士,硕士生导师,研究方向:算法分析,多目标优化。赵文君(1993~),女(汉族),河南新乡,河南工业大学信息学院研究生,在读硕士,研究方向:算法分析。