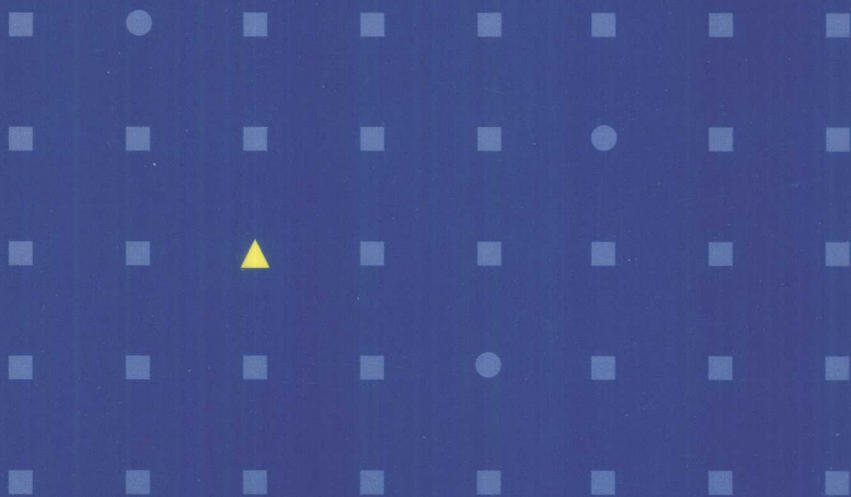


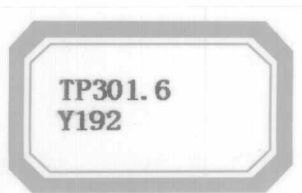
进化算法的模式、 涌现与困难性研究

Evolutionary Algorithms:
Schema, Emergence and Hardness

杨海军 李建武 李敏强 著



科学出版社



郑州大学 *04010778341Z*

进化算法的模式、涌现 与困难性研究

(Evolutionary Algorithms: Schema, Emergence and Hardness)

杨海军 李建武 李敏强 著



科学出版社
北京

TP301.6
Y192

内 容 简 介

本书旨在系统地介绍进化算法的模式、涌现与困难性等若干问题的理论研究和典型应用,共分为7章内容。

首先,建立了进化计算的统一框架和进化算子的统一表示。其次,将建筑块的思想推广到整个进化计算领域,得到了准确的模式公式、模式的形式不变性和变长度的模式公式。证明了满足一定条件的有限群体遗传算法构成 Devaney 意义下的混沌。提出了一种可以直接测试适应值曲面特征的排序统计分析方法;分析了遗传算法适应值曲面的复杂程度,提出了基于随机游走模型的适应值曲面关联维数测试方法。最后,提出了一种改进的子群遗传算法,并将其应用于多模态函数的优化求解。

本书可以作为管理科学和信息技术专业的研究生教材,亦可作为有关科研人员和工程技术人员的参考用书。

图书在版编目(CIP)数据

进化算法的模式、涌现与困难性研究/杨海军,李建武,李敏强著. —北京:科学出版社,2012

ISBN 978-7-03-033175-5

I. ①进… II. ①杨…②李…③李… III. ①计算机算法—研究 IV. ①TP301.6

中国版本图书馆 CIP 数据核字(2011)第 277286 号

责任编辑:马 跃/责任校对:田 贺
责任印制:张克忠/责任设计:蓝正设计

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

新科印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2012 年 2 月第 一 版 开本:720×1000 1/16

2012 年 2 月第一次印刷 印张:13 1/2

字数:260 000

定价:55.00 元

(如有印装质量问题,我社负责调换)

前 言

进化计算(evolutionary computation, EC)作为一个新兴的交叉学科领域,自20世纪60年代以来,无论是在理论探索方面还是在应用拓展方面,均得到了全面快速的发展,已经成为系统科学、计算机科学、人工智能等领域中的一个重要分支。众多专家运用随机过程、动力系统、统计学等理论方法,以进化计算的群体进化过程为对象,深刻分析了进化过程和进化算子的性质和行为,得出了一些新颖的结论,对指导高效进化算法的创新设计和广泛应用提供了很好的理论基础。同时,针对在经济和社会等工程技术和管理工作中的重大决策任务,以及科学研究领域中大量存在的复杂计算任务,进化计算无论是在面向这些任务的复杂系统建模与模拟方面,还是在复杂决策分析和优化问题的求解方面,均提供了一套崭新的研究思想和解决方案。与传统理论和方法相比,进化计算的群体进化模型和简单进化算子使得它具有很好的系统建模能力和空间搜索能力。特别是用于复杂优化问题求解的进化算法(evolutionary algorithms)有着强大的全局最优解搜索能力、问题域的独立性、信息处理的隐并行性(implicit parallelism)、应用的鲁棒性和操作的简明性,成为一种具有良好普适性和规模化的优化方法。

本书旨在系统地介绍进化算法的若干理论研究和典型应用,包括我们最近完成的一些研究成果:首先,建立了进化计算(包括遗传算法、遗传规划、进化策略、进化规划等各类进化算法)的统一框架和进化算子的统一表示。其次,将建筑块的思想推广到整个进化计算领域,得到了准确的模式理论公式、模式的形式不变性和变长度进化计算中的模式理论。将进化算法表示为离散拓扑空间上的转移自映射的一个复合函数,以符号动力系统为工具,证明了满足一定条件的有限群体的遗传算法构成了 Devaney 意义下的混沌。同时,阐述了遗传算法适应值曲面(fitness landscape)的概念,从图论的角度对遗传算法适应值曲面进行了分析,基于适应值曲面上的随机游走相关函数,详细推导了相关长度计算公式。讨论了 NFL(no free lunch)定理,并分析了遗传算法困难度研究的意义。阐述了遗传算法一欺骗问题(GA-deceptive)中的各种定义及定理,描述了模式欺骗对遗传算法困难度的影响。针对基因关联问题进行了统计分析,定义了基因关联方差及基因关联相关系数,提出了一种可以直接测试适应值曲面特征的排序统计分析方法。利用分形理论分析了遗传算法适应值曲面的复杂程度。最后,分析了进化计算中的并行算法,描述了子群遗传算法(cohort genetic algorithms, CGA)的并行性,

提出了一种改进的子群遗传算法，并应用于多模态函数的优化求解。

本书的研究工作得到了国家自然科学基金面上项目(项目批准号: 70171002, 70571057, 70771002, 71171010)、国家杰出青年科学基金项目(项目批准号: 70925005)和航空基金(项目批准号: ZG200951076)的资助。有关成果系多年来课题组全体成员努力工作和协同攻关的结果。同时,也得到了 D. E. Goldberg、J. H. Holland 等教授和国内众多同行专家学者的指导和帮助,谨对此表达诚挚的谢意。

本书由李敏强统筹规划,由李敏强和杨海军负责全书的整理和校阅。其中,第 1、7 章由李敏强和杨海军撰写,第 2、3、6 章由杨海军和李敏强撰写,第 4、5 章由李建武和李敏强撰写。在本书的研究和整理工作中,得到了中国科学院数学与系统科学研究院汪寿阳教授、南京大学盛昭瀚教授、天津大学寇纪淞教授等的热心指导和鼓励,谨对此表示衷心的感谢。

进化计算和进化算法是一个正处于快速发展中的科学领域,其理论和应用方面均存在大量问题并尚待进一步的深入研究。由于著者学识水平和可获得资料的限制,本书的撰写难免存在不妥之处,敬请同行专家和诸位读者批评指正。

著者

2011 年 10 月于天津大学北洋园

目 录

前言

第 1 章 进化计算概述	1
1.1 进化计算的发展历程	1
1.2 进化计算的主要特点	7
1.3 进化计算的理论研究	9
1.4 进化计算的应用现状	13
1.5 主要研究内容	16
第 2 章 进化算法的模式理论	18
2.1 进化算子的统一表示	18
2.2 模式理论及相关定义	20
2.3 建筑块假设和适度模式	22
2.4 基于排序选择下的准确模式理论	25
2.5 依概率选择下的准确模式理论	32
2.6 模式的形式不变性及准确模式理论	34
2.7 变长度进化计算中的模式理论	44
2.8 基于准确模式理论下的建筑块理论	48
2.9 本章小结	54
第 3 章 进化算法中的涌现及混沌	55
3.1 进化计算中的涌现	55
3.2 进化计算中的混沌现象	56
3.3 有限群体遗传算法的动力性	59
3.4 本章小结	67
第 4 章 遗传算法适应值曲面和困难度影响因素	68
4.1 适应值曲面概念的起源与发展状况	68
4.2 遗传算法适应值曲面的定义及相关结构分析	70
4.3 随机游走模型的时间序列分析	78
4.4 模式适应值曲面与动态适应值曲面	82
4.5 NFL 定理及遗传算法困难度分析的意义	89

4.6	遗传算法的模式欺骗性分析	95
4.7	基因关联分析与测试	103
4.8	影响遗传算法困难度的其他因素	122
4.9	本章小结	126
第5章	遗传算法困难度测试	127
5.1	几种常见的遗传算法困难度测试方法	127
5.2	遗传算法适应值曲面上的排序统计分析	131
5.3	遗传算法适应值曲面的分形分析	136
5.4	基于模式适应值曲面的基因关联测试	139
5.5	测试实数编码遗传算法的困难度	142
5.6	基于进化动力统计分析的遗传算子性能测试	146
5.7	困难问题的构造及遗传算法改进策略	151
5.8	本章小结	163
	本章附录:遗传算法性能测试函数	165
第6章	并行进化算法和子群遗传算法	169
6.1	并行进化算法	169
6.2	子群遗传算法及其在皇家大道函数上的应用	178
6.3	子群遗传算法在多模态问题上的应用	182
6.4	本章小结	187
第7章	总结与展望	188
7.1	总结	188
7.2	展望	190
	参考文献	192

第 1 章 进化计算概述

20 世纪 60 年代以来,进化计算作为一个新兴的交叉学科,从无到有,由弱到强,现在已经成为系统科学、计算机科学、人工智能等领域中的一个重要分支^[1]。在工程技术、社会经济以及科学计算领域中存在着大量的复杂性计算问题,由于这些问题表现出的高度非线性、不可导甚至不连续性,复杂度非常高,传统的优化算法在这些问题面前已显得无能为力,而进化计算的出现为解决这类问题开辟了一条崭新的途径。在过去的 50 多年中,以遗传算法(genetic algorithms, GA)、进化规划(evolutionary programming, EP)、进化策略(evolution strategies, ES)、遗传规划(genetic programming, GP)为代表的启发式随机搜索算法,在达尔文(C. R. Darwin)生物进化思想和后达尔文主义(Neo-Darwinian)进化思想的启迪下,在遗传学特别是现代分子生物学的启发下^[2,3],在理论和应用两个方面都取得了长足的发展。

进化计算的快速发展是与其广阔的应用前景紧密相关的。随着经济、科学和社会的不断向前发展,现实中面临的重大问题越来越多,其复杂程度也越来越高,对求解算法的要求也就随之提高了。为了提高算法的性能,必须不断加强对进化计算工作机理的认识,从根本上解决这一问题。

1.1 进化计算的发展历程

进化计算的起源可以追溯到 20 世纪 50 年代中期, R. M. Friedberg 首先将进化思想与计算机相结合^[1,4],用于寻找一个可以计算给定函数的输入、输出问题的程序,这一工作可以认为是遗传规划的起点。据目前可查文献记载, H. J. Bremermann 是第一个将进化思想用于计算机并解决数值优化问题的学者^[5]。他们的开拓性工作为 20 世纪 60 年代中期遗传算法、进化规划和进化策略发展的起点, 20 世纪 80 年代末期的遗传规划也间接地受到了影响。虽然遗传算法、进化规划和进化策略这三种进化计算中的主流算法是在同时期发展起来的,并具有相同的生物学背景,面对的数值优化问题也极其相似,但在它们近 30 年的早期发展中,彼此间是独立发展的。直到 1990 年 10 月,在多特蒙德召开的一次会议才将进化计算领域不同研究分支的学者聚集在一起,使不同的分支相互认知、融合,从而形成进化计算。由于遗传规划求解问题的具体思想与遗传算法非常相似,再加上其出现的时期比较晚,大多将之并入遗传算法中加以考虑。本书

考虑到遗传规划在实际应用中所表现出的特点及其理论上的特性,故将蓬勃发展的遗传规划作为进化计算中的一个独立分支加以考虑。下面分别介绍一下进化计算中四个主要分支的发展历程。

1.1.1 遗传算法的发展历程

J. H. Holland 被公认为是遗传算法的创立者^[6]。20 世纪 60 年代初期, Holland 在研究生物学、人类社会及控制领域中动态系统的适应性问题时,采用简单的模拟机制在计算机上模拟复杂的适应性现象。在实验中发现,要想得到稳定的适应性系统(robust adaptive systems),一个具有进化能力的系统是关键。在这一时期,遗传算法中的进化操作已经出现了,并提出了利用群体进化的思想,指出新的群体应当基于当前群体的有效性来产生。尽管他当时没有给出实现这些思想的具体技术,但确实引进了群体、适应值、复制、变异、交叉等基本概念。遗传算法作为一个专有名词首次出现在 1967 年 J. D. Bagley 的博士论文中^[7],当时该算法主要是为了优化游戏中的参数设定,但遗传算法的基本机制已经出现。同时期, Holland 的许多学生的工作在遗传算法的形成过程中起了非常重大的作用,除了 Bagley 外,还有 D. J. Cavicchio 等,这些 Michigan 大学的学生最后大多成为了该领域的著名学者。

遗传算法和其他任何一门新兴学科一样,在饱受怀疑和反对的目光中走过了它的最初阶段,到了 20 世纪 70 年代中期,关于遗传算法的研究才缓缓展开。1975 年, Holland 出版了遗传算法领域中的第一本专著 *Adaptation in Natural and Artificial Systems*,这是遗传算法领域中一个里程碑式的著作。该书对遗传算法中的基本理论及方法进行了系统的阐述,提出了著名的模式理论(schemata theorem)及隐并行性,用以解释遗传算法的运行机理,在明确了进化算子的前提下,将算法应用于系统适应性研究、自动控制和机器学习等多个领域^[8]。

1976 年在 Michigan 大学召开了第一次关于遗传算法的讨论会,虽然这是一次仅有 20 多人参加的会议,但今天看来,其意义却是不可低估的,它使遗传算法的研究走出了封闭的状态。

遗传算法在最初并不是作为数值优化工具出现的,但其在优化问题上的成功,却成了遗传算法快速发展的主要原因, Cavicchio^[9]和 de Jong^[10]在这一方面做出了突出的贡献。前者使用整数编码的遗传算法进行模式识别的研究;后者在其博士论文中首次将遗传算法用于函数优化问题,并进一步研究了遗传算法的理论问题和参数优化问题,利用五个基本函数建立了测试平台,并给出了关于进化算子的经验分析, de Jong 的这一研究结果在今天的遗传算法领域仍具有普遍的指导意义。

除了 Michigan 大学之外, Pittsburgh 大学是另一个研究遗传算法的中心。在众多学者的不断努力下, 遗传算法不断向前发展。1985 年, 在 Pittsburgh 召开了第一届遗传算法国际会议(International Conference on Genetic Algorithms, ICGA), 该会议每两年举行一次, 到 1997 年为止, 该会议一共举办了七届。1999 年 ICGA 与遗传规划的会议(by Genetic Programming Society)合并, 改为每年举行一次的国际会议(Genetic and Evolutionary Computation Conference, GECCO), GECCO 是一个涵盖进化计算各个领域的综合性国际会议。

1989 年, 遗传算法领域中的另一本专著 *Genetic Algorithms in Search, Optimization and Machine Learning* 问世了, 这是遗传算法领域中另一个里程碑式的著作。作者 D. E. Goldberg 在全面研究遗传算法的基础上, 对当时的工作进行了全面细致的总结和分析。该书的最大特点在于理论与应用的结合, 不但适用于广大的研究人员, 而且还是一本很好的教材^[11]。

1991 年 ISGA(International Society for Genetic Algorithms)举办了一个新的国际会议(Foundations of Genetic Algorithms, FOGA), 该国际会议在遗传算法领域具有更大的学术影响力, 该会议每两年举行一次, 已经成为引领遗传算法研究方向的一个国际会议。FOGA 会议论文集收录的文章主要以遗传算法理论性研究为主。

1994 年 6 月, IEEE 神经网络委员会(IEEE Neural Network Council)召开了第一届进化计算国际会议(International Conference on Evolutionary Computation, ICEC), 该会议每年举办一次。1999 年, 该会议与进化规划的会议(Evolutionary Programming Conference)合并为一个新的国际会议 CEC(Congress on Evolutionary Computation), 该会议由 IEEE 神经网络委员会和遗传规划团体(the Evolutionary Programming Society)等几家团体合办, 其会议的范围非常广泛, 几乎包括了进化计算的所有领域。最近, 与遗传算法相关的重要会议很多, 不再一一列举。

在欧洲, 从 1990 年开始也每隔一年举办一次类似的会议, 即 PPSN(Parallel Problem Solving from Nature)会议。以遗传算法理论基础为核心的学术会议 FOGA 也从 1990 年起每隔一年举办一次。

美国 MIT 出版社从 1993 年开始出版 *Evolutionary Computation* 和 *Adaptive Behavior* 杂志。世界上第一本关于人工智能研究的杂志 *AI Trends* 于 1993 年更名为 *Critical Technology Trends*, 并在更名启示中讲到: “遗传算法、自适应系统、细胞自动机、混沌理论和人工智能一样, 都是对今后 10 年的计算机技术有重大影响的关键技术。”

随着 Internet 技术的发展和普及应用, 遗传算法的有关研究单位建立了大量的专题遗传算法网站, 其中最为著名的是由美国海军人工智能应用研究中心建立的 GA _ Archives 检索网站(<http://www.aic.nrl.navy.mil/galist/>), 它包

括了世界范围内的开展遗传算法和进化计算研究的大学和机构,历年来的可公开发表的论文和报告,有关国际会议的消息,典型应用案例和程序(源代码),等等。

这些众多的研究机构和频繁的国际学术活动集中反映了遗传算法的学术意义和应用价值。目前,遗传算法已成为一个多学科、多领域的重要研究方向。

1.1.2 进化规划的发展历程

进化规划的思想是 L. J. Fogel 在 1960 年首先提出来的。L. J. Fogel 在有限状态机(finite-state machines)的研究中发现,早期的神经网络方法和启发式算法无法满足系统不断提高性能的要求,他希望通过系统的自适应机制来达到这一要求,正是这一基本思想导致了进化规划的产生。所谓状态机就是将周围的环境情况用有限字符集中的符号组成的符号序列来表示,根据目前已知的符号序列得到新的字符序列,通过适应值函数的比较以便获得更准确的预测结果。1966 年, L. J. Fogel、A. J. Owens 和 M. J. Walsh 合著了 *Artificial Intelligence Through Simulated Evolution* 一书,该书系统地阐述了进化规划的基本思想,在进化规划的过程发展中具有重要的历史意义^[12]。早期的进化规划在问题预测和自动控制领域获得了一定的成功,这种成功也没有改变进化规划不被接受的命运。除了 L. J. Fogel 外, B. E. Lutter、R. C. Huntsinger 和 D. W. Dearholt 等也为进化规划的早期发展做出了重要的贡献。Lutter 和 Huntsinger 将自动状态机的成果应用于工程领域,取得了较好的效果;Dearholt 则使用进化规划的思想进行实际问题的解决^[1]。

在 20 世纪 80 年代中后期,随着应用范围的扩大,进化规划进入了一个新的发展阶段。在这一发展阶段, D. B. Fogel 做出了重要贡献。从 20 世纪 80 年代中后期到 20 世纪 90 年代, D. B. Fogel 就使用了进化规划解决 TSP(traveling salesperson problem)问题,改善了求解性能;他将进化规划应用于连续参数优化问题领域,扩大了进化规划的适用范围,增强了进化规划的影响力。在此期间,进化规划不断扩展其应用领域,取得了相当丰富的成果。进化规划作为神经网络中的学习算法,提高了神经网络的性能;进化规划中的自适应机制,不但提高了其自身的性能,而且被广泛应用于进化计算的各个方向^[13~15]。

1992 年, D. B. Fogel 和 W. Atmar 在美国举办了进化规划的第一届会议。L. J. Fogel 和 D. B. Fogel 应用进化规划算法成功地解决了许多实际问题,例如,应用进化规划和神经网络相结合进行胸部癌症的检测^[16]。

1.1.3 进化策略的发展历程

I. Rechenberg、H. P. Schwefel 和 P. Bienert 是进化策略公认的发明者^[1]。1964年,他们还在柏林技术大学上学的时候,在设计一种能够评测风洞中可弯曲的三维物体最小变形的设备过程中发明了该算法。与其他研究工作一样,开始阶段,他们希望通过传统的数值算法来解决这一问题,但结果失败了。传统的数值算法无法跳出局部最优解,算法很快收敛。Rechenberg 首先想到了将进化思想应用到算法的设计中,他将进化论中的生物选择、重组、变异应用到参数优化中^[17],可以认为进化策略的思想已经形成,虽然这只是进化策略的一个初始形态,但它涵盖了进化策略的基本思想。Rechenberg 的基本思想如下:首先随机产生一个离散的初始个体,其次将一个满足二项分布的变异作用在初始个体上,这样得到一个新的个体,从这两个个体中选择适应值较大的个体作为下一代的父个体,这就完成了一次循环。这种算法应该称为 $(1+1)$ 进化策略。

Schwefel 首先将 Rechenberg 的算法付诸实施^[18]。Schwefel 的实验非常成功,该算法避开了局部最优解的陷阱,获得了全局最优解。这一成功对于进化策略而言同样具有里程碑式的意义。Bienert 的贡献在于将进化策略思想应用到实际具有工业意义的具体问题上^[19],该问题是传统算法所不能解决的,Bienert 的硕士论文成功地解决了这一问题,这一工作连接了进化策略的理论与实践,其意义重大。

随后 Rechenberg^[20]进一步改进了进化策略,在解决 n 维凸函数的优化中,给出了算法的收敛性分析,并在此基础上给出了著名的“ $\frac{1}{5}$ 成功原则”,即通过外在改变二项分布的方差的经验值。Rechenberg 又给出了多群体下的进化策略算法,即 $(\mu+1)$ 进化策略,群体内个体间的重组操作也开始被使用。Rechenberg 的这项工作对于进化策略的进一步发展起到了决定性的作用,群体概念和重组算子的引入从理论上完善了进化策略,为其在实际应用中取得成功奠定了理论基础。但是这一算法由于变异算子缺乏自适应机制而显得先天不足,并未得到广泛的推广和应用。

Schwefel^[21]在 $(\mu+1)$ 进化策略的基础上迈出了重要的一步,将其推广到 $(\mu+\lambda)$ 进化策略和 (μ,λ) 进化策略中,这两种形式的进化策略成为后来广泛使用的两种形式。进化规划中使用的自适应机制在进化策略中也被使用。进化策略的研究中心在德国,主要以 Schwefel 在多特蒙德领导的研究团队和 Rechenberg 在柏林领导的研究团队为主。

随着并行思想的深入,进化策略在 20 世纪 90 年代引入了并行概念并取得了

一定的成功。这一时期正是进化计算各个分支相互融合的时期,进化策略研究团体发挥了重要的作用。

在进化策略的理论研究方面,H. G. Beyer 做出了重要的贡献。Beyer^[22,23]首先建立了进化策略的数学模型,并以微分几何为工具研究了非球面适应值空间的性质,研究了 (μ, λ) 等多种进化策略的收敛性质以及算法的动力学性质,为进化策略的下一步发展打下了基础。

1990年10月,第一届PPSN会议在德国举行,该会议每两年举办一次,是欧洲关于进化计算方面重要的国际会议之一。

1.1.4 遗传规划的发展历程

遗传规划是进化计算中最年轻的分支。1989年,J. R. Koza发表了著名的论文*Hierarchical genetic algorithms operating on populations of computer programs*,标志着遗传规划的诞生;1992年和1994年J. R. Koza先后出版了两部遗传规划的专著,这是遗传规划走向成熟的标志^[24,25]。

J. R. Koza是Holland的学生,他对遗传算法的深入研究是发明遗传规划的一个先决条件。在遗传算法中采用的是定长的编码方式,这一方式限制了遗传算法的应用范围,对于层次化的问题,采用定长的编码方式显然不合适,遗传算法很难解决层次化的问题,这种要求促使具有变长编码方式的遗传算法出现。计算机程序是一种典型的具有层次化结构的对象,传统的遗传算法编码方式难以处理这一对象。遗传规划采用长度和大小均可变化的树型结构进行编码,这种编码方式具有灵活性,而且易于实现,所以遗传规划的编码方式可以很好地解决这一矛盾。遗传规划的搜索过程就是在计算机程序构成的搜索空间内,通过进化算子的作用,寻找具有高适应值的计算机程序的过程。

LISP语言是遗传规划采用的基本计算机语言。LISP语言是1958年由J. McCarthy等发明的一种计算机语言。该语言具有结构简单、易于操作的特点。该语言的数据结构只有两种,即节点(atom)和目录(list)。所谓节点就是符号和数字组成的集合中的任意一个元素,如在集合 $\{A, B, C, 4, 5, 6\}$ 中,任意一个元素均可以构成LISP语言中的一个节点;所谓目录就是由小括号构成的嵌套结构,如 $(A (B (C C) 6) 6)$ 。遗传规划采用的是解析树表示,LISP语言非常适合这种树型结构的表示,现在LISP语言在遗传规划中仍普遍使用。在LISP语言被广泛应用的同时,遗传规划也可以用C和C++语言来解决问题。

J. R. Koza认为应用遗传规划解决实际问题的五个主要准备阶段如下^[25]:
①给定一个终止符号集合;②基本函数集合;③适应值函数的度量;④参数设定;⑤运行结果的选定方式和运行终止条件的设定。

J. R. Koza 发明了遗传规划, 并应用它解决了许多具体问题。在应用领域迅速扩展的同时, 遗传规划的理论研究也在快速展开, 在这一方面, R. Poli 等工作取得了显著的成果^[26,27]。针对遗传规划本身的特点, 在借鉴遗传算法中模式理论的基础上, 进一步给出了基于不同交叉和变异情况下的准确模式公式。另外, 建筑块假设 (building blocks hypothesis) 在遗传规划中也同样受到了一些质疑^[28]。

1.1.5 进化计算各个分支的融合发展

遗传算法、进化策略和进化规划是在不同的地区, 由不同领域内的研究人员各自独立地提出来的。其早期面对的问题也各不相同, 彼此之间缺乏必要的学术交流。进化策略和进化规划领域的研究人员首先建立了联系, 通过学术会议两个领域的研究人员将这种联系保持了下来。随后遗传算法和遗传规划也加入了进化计算这个大家庭中。虽然不同的分支各有其特点, 但它们具有相同的仿生基础——生物进化思想。1990 年, J. C. Bezdek 在多特蒙德的会议上首次使用了进化算法一词, 现在进化算法和进化计算是通用的。

随着进化计算各个分支的相互融合, 不同分支的国际会议或是相互合并, 或是将其他分支的研究方向纳入自己的会议范围内, 进一步加强了彼此间的联系。例如, 进化规划中的自适应机制已经被遗传算法和遗传规划所借鉴, 随着时间的推移, 这种相互融合的趋势必将进一步加强。

1993 年, MIT 出版社开始出版发行刊物 *Evolutionary Computation*, 这一刊物成为进化计算领域中重要的学术杂志。该期刊面向进化计算领域中理论和实际应用问题。1997 年, IEEE 出版了另一本重要的进化计算刊物 *IEEE Transactions on Evolutionary Computation*。这些国际学术期刊的问世, 促进了进化计算学术界的相互交流, 使进化计算的各个分支真正融合为一体。

1.2 进化计算的主要特点

进化计算是一种基于自然选择机制下的全局性随机搜索算法。它从随机产生的初始群体开始, 经过进化算子的不断作用, 在迭代过程中逐步改进当前群体中个体的适应值, 直到算法得到问题的最优解或满意解为止, 在总结进化计算各个分支特点的基础上, 给出进化计算的主要特点。

1.2.1 遗传算法与遗传规划的主要特点

遗传算法具有以下主要特点:

(1)遗传算法对优化对象限制很少。由于遗传算法在选定编码方式后,处理的是参数的编码集而不是函数本身,所以对优化函数的性质没有具体的要求。这与传统优化方法对优化函数在连续性、可导方面的要求不同。

(2)遗传算法的搜索初始点是初始群体,而不是单一的初始点,这种方式有利于搜索过程以较大的概率跳出局部极值点,提高了算法搜索到问题全局最优解的能力^[29]。

(3)遗传算法具有显著的隐并行性^[8]。从模式的角度考虑遗传算法的求解过程,算法在处理群体中个体的同时,也在对该个体所在的模式进行处理,具有并行处理的性质。

(4)遗传算法构造简单、易于实现,对具体的运行环境没有过多的限制,可以在不同的软硬件平台上实现。

(5)遗传算法具有典型的动力学特征。遗传算子在算法运行过程中会产生涌现现象,遗传算法在求解一些问题时会表现出混沌及分形特征。

遗传规划除了具有遗传算法所具有的上述五个特点外还具有如下特点:

(1)遗传规划采用动态编码方式,可以解决层次化问题,扩大了进化计算的研究范围。

(2)遗传规划的输出结果是动态可变的。

1.2.2 进化规划的主要特点

进化规划具有以下主要特点:

(1)进化规划是一种采用自适应机制的进化算法。在进化规划中步长向量受进化过程的控制自发地进行调解。

(2)进化规划的搜索初始点也是初始群体,而不是单一的初始点,算法采用自顶向下和自底向上的方式进行搜索。

(3)进化规划中不采用重组和交叉算子。进化规划将群体内的每一个个体看做一个物种,物种间的杂交是不存在的,因此算法不使用重组和交叉算子,这在进化计算中是一个特例^[1,30]。

(4)进化规划构造简单、易于实现,可以与其他启发式算法结合使用,也可以作为其他算法的一个学习工具,如在神经网络中的应用。

(5)进化规划具有典型的动力学特征。随着时间的推移,进化规划在运行过程中,其群体表现出动力学特征。

1.2.3 进化策略的主要特点

进化策略具有以下主要特点:

- (1) 进化策略是首先采用自适应机制的进化算法^[1,31]。
- (2) 进化策略的搜索初始点最初使用的是单一的初始点, 这种方式的进化策略已基本不再使用; 现在普遍使用的进化策略算法也是采用初始群体作为搜索的初始点。
- (3) 进化策略构造简单, 易于实现, 可以采用嵌套方式来应用这一算法。
- (4) 进化策略具有典型的动力学特征。特别是在 One-Max (1 + 1) 问题上已经有了肯定的结论。

1.2.4 进化计算的特点总结

综合 1.2.1~1.2.3 节的内容, 下面给出进化计算的主要特征:

- (1) 进化计算具有宽广的适用范围。由于进化计算采用灵活的编码方式(二进制编码、实数编码和字符集编码等), 算法处理的是编码集合, 因此对优化对象没有过多限制。
- (2) 进化计算的搜索初始点是初始群体, 而不是单一的初始点, 这种方式有利于搜索过程以较大的概率跳出局部极值点, 提高了算法搜索到问题全局最优解的能力。这是进化计算区别于一般启发式算法的主要特点之一。
- (3) 进化计算构造简单, 易于实现, 相互之间可以结合使用, 也可以与其他启发式算法结合使用。
- (4) 进化计算具有内在的并行计算特征。
- (5) 进化计算具有典型的动力学特征。
- (6) 进化计算是一种启发式随机搜索算法。

1.3 进化计算的理论研究

进化计算的理论研究主要集中在算法的收敛性研究和运行机理研究两个方面。在这两个方面, 进化计算的各个分支各有其特点。

1.3.1 遗传算法和遗传规划的理论研究

因为遗传算法的理论研究基本涵盖了遗传规划的理论研究, 所以本节将二者统一考虑。遗传算法在收敛性上的研究方法主要分为随机模型方法和解析方法。

应用随机模型方法对遗传算法的收敛性进行研究, 是因为遗传算法的进化过程可以表示为有限状态下的马尔可夫链(Markov chain)。这种表示的前提是遗传算法的下一代群体中个体的存在概率仅与上一代群体的存在概率相关, 与其他无

关。遗传算法所要解决的问题，一般可以表示为有限空间上的求解过程，在这种情况下，算法必然收敛。对于无限群体的遗传算法而言，Eiben 等证明了在精英保留策略下，算法依概率 1 收敛于问题的全局最优解^[32]。马尔可夫链的结果被广泛应用于遗传算法的收敛性证明中，并取得了丰富的成果。Goldberg 和 Segrest^[33]首先将遗传算法进行了有限马尔可夫过程处理，研究了其收敛性。Fogel、Rudolph、Nix 和 Vose^[34~36]等分别使用该工具在不同的时期对遗传算法的收敛性进行了深入的研究。以上的成果大多是在概率意义下基于算法遍历性的研究，徐宗本^[37]等将遗传算法的进化过程等同于鞅序列，应用随机理论中关于鞅的有关结论，证明了一大类遗传算法在采取非精英保留策略下算法仍依概率 1 收敛于问题的全局最优解。

应用解析方法研究遗传算法的收敛性，当首推 Holland 提出的模式定理，他证明了在遗传算子的作用下，群体中模式的收敛性。这是一个不等式的表示，而且在模式生存上也仅能看到与上一代模式的关系，Goldberg、Stephens 等推动了模式理论不断向前发展^[11,38,39]；在遗传规划方面，Poli 将模式理论做了进一步推广^[26~28,40]。从一代到多代，从近似到准确，模式定理不断向前发展。建筑块假设和模式定理统称为遗传算法中的模式理论。在遗传规划领域，模式理论的指导意义同样重要。

模式理论在研究算法收敛性的同时，揭示了算法的运行机理。在算法运行机理的研究上，Liepins、Vose、Nix 和 Wright 等分别使用统计学、代数学和解析工具对遗传算法的动力性从断续平衡到混沌现象^[36,41~45]进行了深入研究。

虽然这两方面的理论研究都很重要，但在揭示遗传算法的本质性上仍有很长的路要走，因此在算法的具体应用和参数调控上发挥的作用仍然有限。

随着遗传算法的基础理论研究即收敛性研究和收敛速度研究的进展，及 NFL 定理^[46,47]的提出，遗传算法的困难度分析已被提上日程。Wolpert 和 Macready^[46]的 NFL 定理指出：对于任意的表现度量，当对所有可能的目标函数做平均时，所有搜索算法的表现是完全一样的。因此，对于一个特定的问题，研究遗传算法优化该问题的困难程度，即遗传算法困难度分析，具有重要的意义。

Bethke^[48]最初利用 Walsh 变换对遗传算法一困难问题与遗传算法容易问题进行了描述。Bethke 指出，如果一个函数的 Walsh 分解中高阶分割对应了较重要(数值较大)的 Walsh 系数，则该函数难于用遗传算法求解，为遗传算法一困难问题(GA-hardness)。然而在实际中并不能通过 Walsh 模式变换来判定一个函数是否是遗传算法一困难问题，因为计算 Walsh 系数需要对编码空间中的每个点计算适应值。

遗传算法一欺骗问题往往是遗传算法一困难问题。欺骗问题是指在遗传算法的搜索过程中，存在某些低阶模式难以重组为期望的高阶模式，或者更为严重的