

管理、决策与信息系统丛书



2

李敏强 寇纪淞 林 丹 李书全 著

遗传算法的 基本理论与应用



科学出版社
Science Press

管理、决策与信息系统丛书

遗传算法的基本理论与应用

李敏强 寇纪淞 林丹 李书全 著

科学出版社

2002

内 容 简 介

本书旨在系统地介绍遗传算法的理论、应用和发展,共包括9个章节的内容。

首先,本书讲述了遗传算法的起源、历程和主要研究方向,介绍了遗传算法的基本原理。其次,讨论了遗传算法的一般收敛性理论,遗传算法的马尔可夫链模型和收敛性分析,遗传算法的随机泛函分析。还介绍了遗传算法的模式理论,特别是遗传算法的模式欺骗性理论,以及欺骗问题的实验分析。并详细讨论了微观遗传策略——遗传算子的分析与设计,以及微观遗传策略中的参数设置和适应性微观遗传策略的设计。讨论了宏观遗传策略——遗传算法结构分析与设计。接下来介绍了遗传算法在知识获取中的应用,特别是概念学习和特征提取的遗传算法方法。讨论了遗传规划的原理、方法和收敛性分析,及其在典型问题中的应用。最后,介绍了遗传算法的发展——进化计算的原理与方法,给出了进化算法的一般框架和收敛性分析,讨论了NFL定理的意义,以及浮点实数编码的遗传算法在求解约束优化问题中的应用。附录中给出了一组典型的性能测试函数。

本书可以作为信息技术和管理科学专业的研究生教材,亦可供有关科研人员和工程技术人员阅读参考。

图书在版编目(CIP)数据

遗传算法的基本理论与应用/李敏强等著,一北京:科学出版社,2002.3
(管理、决策与信息系统丛书/汪寿阳,杨晓光主编)

ISBN 7-03-009960-5

I. 遗… II. 李… III. ①遗传-算法 基础理论 ②遗传-算法-应用
IV. TP18

中国版本图书馆 CIP 数据核字(2001)第 098477 号

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 蕾 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2002年3月第 一 版 开本:BS(720×1000)

2002年3月第一次印刷 印张:27 1/4

印数:1—2 000 字数:506 000

定价:45.00 元

(如有印装质量问题,我社负责调换(环伟))

前 言

20 世纪 40 年代以来,科学家不断努力从生物学中寻求用于计算科学和人工系统的新思想、新方法。很多学者对关于从生物进化和遗传的机理中开发出适合于现实世界复杂适应系统研究的计算技术——自然进化系统的计算模型,以及模拟进化过程的算法进行了长期的开拓性的探索和研究,John H. Holland 教授及其学生首先提出的遗传算法就是一个重要的发展方向。

从 1989 年起,我们开始接触到了遗传算法的一些研究成果。1994 年,寇纪淞教授在美国访问研究期间,有幸到美国密歇根大学拜访了 John H. Holland 教授。Holland 教授热情接待,并将最新出版的 *Adaptation in Natural and Artificial Systems* 第二版签名相赠。同时,介绍了他自己和多位知名的遗传算法专家的研究情况,使我们对遗传算法的研究和应用领域有了更深刻的认识。

按照生物学上可进化性的概念,遗传算法所追求的也是当前群体产生更好个体的能力,即遗传算法的可进化性或称群体可进化性。遗传算法的理论和方法研究也围绕着这一目标展开。比如,如何更好地模拟复杂系统的适应性过程和进化行为?在优化问题求解中怎样才能具备全局收敛性?算法的搜索效率如何评价?算法的设计与参数控制的理论基础是什么?遗传算法与其他算法的如何结合?等等。关于遗传算法第 1 个方面的研究主要是由 Holland 的团队和 Santa Fe Institute 的 EVCA 研究组等少数单位开展的。大部分专家和研究机构则主要集中于遗传算法作为优化问题通用求解算法的一系列问题,20 世纪 80 年代以后产生了大量的研究成果。

与传统的启发式优化搜索算法相比,遗传算法的主要本质特征在于群体搜索策略和简单的遗传算子。群体搜索使遗传算法得以突破邻域搜索的限制,可以实现整个解空间上的分布式信息探索、采集和继承;遗传算子仅仅利用适应值度量作为运算指标进行染色体的随机操作,降低了一般启发式算法在搜索过程中对人机交互的依赖。这样就使得遗传算法获得了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,应用的鲁棒性,操作的简明性,成为一种具有良好普适性和可规模化的优化方法。

本书旨在系统地介绍遗传算法的理论、应用和发展,包括了我们最近完成的一些研究成果。首先,概括性地介绍了遗传算法的起源、历程和主要研究方向,详细

描述了遗传算法的基本原理。其次,重点研究了遗传算法的随机理论与分析,遗传算法的模式理论与分析。探讨了微观遗传策略——遗传算子的分析与设计,宏观遗传策略——遗传算法结构分析与设计。然后,讨论了遗传算法在知识获取领域的应用,介绍了遗传规划的基本原理、方法和典型应用实例。最后,从发展的角度,介绍了进化算法的原理与方法,讨论了 NFL 定理的意义。

本书的研究工作得到了国家自然科学基金、天津市自然科学基金、教育部博士点研究基金等的资助,有关成果系多年来课题组全体成员努力工作和精诚合作的结果。同时,David E. Goldberg、Kenneth De Jong、John Koza、Stepanie Forrest、David B. Fogel、Thomas Bäck、Zbigniew Michalewicz 等诸位教授和专家,以及 Erick Cantú-Paz 博士和 H.-G. Beyer 博士提供了大量资料,与 George L. Rudolph 博士的网络交流也对我们的研究工作提供了很多有益的帮助,谨在此表达诚挚的谢意!

本书由寇纪淞、李敏强统筹规划,由李敏强、林丹、李书全整理和校阅。其中,第 1、2、4、5、6、7 章由李敏强撰写,第 3 章第 1 节、第 9 章由林丹撰写,第 3 章第 6 节、第 8 章第 1~3 节由李书全撰写,第 3 章第 2~5 节、第 8 章第 4 节由戴晓晖撰写,第 8 章第 5 节由李建武撰写。参加本书研究工作的还有马丰宁博士、宋庆伟博士、田军博士等。对于他们的辛勤工作和贡献,表示衷心的感谢。

在本书的研究和整理工作中,得到汪寿阳先生多方面的指导和鼓励,并有幸列入本套丛书,同时,本书的编辑出版也得到杨晓光先生的直接帮助,在此一并表示感谢。

遗传算法是一个处于快速发展中的科学分支,其理论和应用均还有大量问题尚待进一步深入研究。由于作者学识水平和可获得资料的限制,本书不妥之处在所难免,敬请同行专家和诸位读者批评指正。

著 者

2002 年 3 月于天津大学

目 录

前 言

第一章 概述	(1)
1.1 从生物进化到进化计算	(1)
1.2 遗传算法的特征与发展	(4)
1.3 遗传算法理论研究	(8)
1.4 遗传算法的应用	(13)
本章附录:遗传算法的基本术语	(15)
第二章 遗传算法的基本原理	(17)
2.1 复杂系统的适应过程	(18)
2.2 遗传算法的基本描述	(26)
2.3 遗传算法的模式理论	(47)
2.4 遗传算法与其他搜索技术的比较	(59)
2.5 遗传算法计算实例	(66)
第三章 遗传算法的随机理论与分析	(75)
3.1 遗传算法的一般收敛性理论	(75)
3.2 遗传算法的马尔可夫链模型	(82)
3.3 齐次遗传算法收敛性分析	(94)
3.4 遗传算法的收敛速率分析	(98)
3.5 广义退火遗传算法的收敛性	(102)
3.6 遗传算法的随机泛函分析	(115)
第四章 遗传算法的模式理论与分析	(120)
4.1 遗传算法的模式收敛性分析	(120)
4.2 遗传算法模式欺骗问题分析	(133)
4.3 遗传算法模式欺骗性的充分条件	(148)
4.4 模式欺骗问题的实验分析	(158)
第五章 遗传算子的分析与设计	(163)
5.1 群体的分析与设计	(163)
5.2 选择算子的性质分析	(175)

5.3	交叉算子的性质分析	(185)
5.4	变异算子的性质分析	(199)
5.5	微观遗传策略中的参数设置	(204)
5.6	适应性微观遗传策略的设计	(208)
第六章	遗传算法结构分析与设计	(216)
6.1	适应函数的复杂性分析	(218)
6.2	单纯多群体遗传算法	(230)
6.3	协同多群体遗传算法	(233)
6.4	遗传算法与位爬山算法的结合	(244)
6.5	组合优化问题的混合遗传算法	(247)
6.6	并行遗传算法	(253)
第七章	遗传算法与知识获取	(262)
7.1	基于数据库的知识发现	(262)
7.2	CS-1 系统与 LS-1 系统	(268)
7.3	描述性概念的学习方法	(270)
7.4	概念学习的模型表示	(274)
7.5	CNF 范式规则学习的 GA 方法	(283)
7.6	DNF 范式规则学习的 GA 方法	(289)
7.7	概念学习中的特征提取	(297)
7.8	基于 GA 的特征选择	(301)
7.9	基于 GA 的特征变换	(306)
7.10	基于 GA 的一种自动聚类方法	(310)
第八章	遗传规划及其应用	(316)
8.1	遗传规划的基本原理与方法	(316)
8.2	遗传规划的收敛性分析	(324)
8.3	遗传规划/遗传算法在复合地基承载力计算中的应用	(331)
8.4	遗传规划在混沌时间序列预测中的应用	(336)
8.5	基于遗传规划的确定性模式分类器	(343)
第九章	进化算法的原理与方法	(349)
9.1	进化算法的一般框架	(349)
9.2	进化算法的自适应机制	(360)
9.3	基于实数编码的遗传算法的收敛性	(364)
9.4	进化规划和进化策略的收敛性	(370)
9.5	不采用精英个体保留策略的进化算法的收敛性	(373)

9.6 关于 NFL 定理的推导与讨论·····	(374)
9.7 进化规划和进化策略中三种变异算子的性质和比较 ·····	(379)
9.8 用 FGA 求解约束优化问题·····	(387)
附录 GA 性能测试函数 ·····	(399)
主要参考文献·····	(404)

第一章 概 述

辩证唯物主义认为世界是物质的,物质是运动的,运动是有规律的,而规律是可以被认识和掌握的。通过认识和掌握事物的规律,可以较好地解决人们面临的各种问题。人类在认识和改造自然的过程中,创造了数学、物理、化学、生物等学科,从不同侧面研究自然界,并取得了显著效果。由于客观事物的复杂性以及人们对事物认识的不断深化,人们发现仅靠某一门学科很难比较完美地解决实际工作中遇到的问题,因而出现了交叉学科。

20 世纪 40 年代以来,科学家不断努力从生物学中寻求用于计算科学和人工系统的新思想、新方法,比如早期的自动机理论就试图采用类似神经元的元素建造一种新型的思维机器。很多学者对关于从生物进化和遗传的机理中发展出适合于现实世界复杂适应系统(complex adaptive systems)研究的计算技术——自然进化系统的计算模型(computational models of natural evolutionary systems)和模拟进化算法(simulated evolutionary algorithm)等,进行了开拓性的长期探索和研究^[1~7],Holland 及其学生首先提出的遗传算法(genetic algorithms,简称 GAs 或 GA)就是一个重要的发展方向^[3~7]。

1.1 从生物进化到进化计算

按照达尔文(C. Darwin)的进化论^[8,9],地球上的每一物种从诞生开始就进入了漫长的进化历程。生物种群从低级、简单的类型逐渐发展成为高级、复杂的类型。各种生物要生存下去就必须进行生存斗争,包括同一种群内部的斗争、不同种群之间的斗争,以及生物与自然界无机环境之间的斗争。具有较强生存能力的生物个体容易存活下来,并有较多的机会产生后代;具有较低生存能力的个体则被淘汰,或者产生后代的机会越来越少,直至消亡。达尔文把这一过程和现象叫做“自然选择,适者生存”。

生物学家对自然界生物的进化机理进行了系统的研究,在如此“短暂”的时间里,生物界由最简单的单细胞生物,发展到现在的纷繁复杂、种群繁多的高级生物,充分证明了自然界的“自然选择,适者生存”的思想的实际意义,与之有关的关于生物进化的研究结论,已得到广泛的接受和应用。

按照孟德尔和摩根(G. Mendel, T. Morgan)的遗传学理论^[8,9],遗传物质是作为一种指令密码封装在每个细胞中,并以基因的形式排列在染色体上,每个基因有特殊的位置并控制生物的某些特性。不同的基因组合产生的个体对环境的适应性不一样,通过基因杂交和突变可以产生对环境适应性强的后代。经过优胜劣汰的自然选择,适应值高的基因结构就得以保存下来,从而逐渐形成了经典的遗传学染色体理论,揭示了遗传和变异的基本规律。现代遗传学则对基因的本质、功能、结构、突变和调控进行了深入探讨,开辟了遗传工程研究的新领域。在一定的环境影响下,生物物种通过自然选择、基因交换和变异等过程进行繁殖生长,构成了生物的整体进化过程。

遗传物质是细胞核中染色体上的有效基因^[8,9],其中包含了大量的遗传信息。染色体上携带着关于生物性状的物质元素,生物体所表现出来的外在特征是对其染色体构成的一种体现。生物的进化本质体现在染色体的改变和改进上,生物体自身形态的变换是染色体结构变化的表现形式。

基因组合的特异性决定了生物体的多样性,基因结构的稳定性保证了生物物种的稳定性,而基因的杂交和变异使生物进化成为可能。生物的遗传是通过父代向子代传递基因来实现的,而这种遗传信息的改变决定了生物体的变异。

生物进化过程的发生需要四个基本条件:①存在由多个生物个体组成的种群;②生物个体之间存在着差异,或群体具有多样性;③生物能够自我繁殖;④不同个体具有不同的环境生存能力,具有优良基因结构的个体繁殖能力强,反之则弱。

生物群体的进化机制可以分为三种基本形式^[1~10]:

(1)自然选择:控制生物体群体行为的发展方向,能够适应环境变化的生物个体具有更高的生存能力,使得它们在种群中的数量不断增加,同时该生物个体所具有的染色体性状特征在自然选择中得以保留。

(2)杂交:通过杂交随机组合来自父代染色体上的遗传物质,产生不同于它们父代的染色体。生物进化过程不需要记忆,它所产生的能很好适应自然环境的信息都包含在当前生物体所携带的染色体的基因库(gene pool)中,并且可以很好地由子代个体继承下来。

(3)突变:随机改变父代个体的染色体上的基因结构,产生具有新染色体的子代个体。变异是一种不可逆过程,具有突发性、间断性和不可预测性,对于保证群体的多样性具有不可替代的作用。

另外,生物进化是一个开放的过程,自然界对进化中的生物群体提供及时的反馈信息,或称为外界对生物的评价。评价反映了生物的生存价值和机会。在基于相同环境下的生存竞争中,生存价值低的个体被淘汰了,生存下来的个体则具有较高的生存价值。由此形成了生物进化的外部动力机制。

大多数高级生物体是以自然选择和有性生殖这两种基本过程实现进化发展的。自然选择决定了生物群体中哪些个体能够存活并繁殖,有性生殖保证了生物体后代基因中的杂交和重组,从而使得群体的进化比其他方式更加快速而有效。

自然界的生物进化是一个不断循环的过程。在这一过程中,生物群体也就不断地完善和发展。可见,生物进化过程本质上是一种优化过程,在计算科学上具有直接的借鉴意义^[1,10]。在计算机技术迅猛发展的时代,生物进化过程不仅可以在计算机上模拟实现^[1~3],而且还可以模拟进化过程,创立新的优化计算方法,并应用到复杂工程领域之中,这就是 GA 等一类模拟自然进化的计算方法的思想源泉^[1~7,11]。以生物进化过程为基础,计算科学学者提出了各种模拟形式的计算方法。

尽管到目前为止进化计算的发展不过 20 余年,但其思想可以追踪到 20 世纪 50 年代^[1~3,11~14]。一般认为,进化计算(evolutionary computation, EC)包括三个组成部分^[12~14]:①由美国密歇根大学 John H. Holland 教授提出的遗传算法^[1,3],②由美国科学家 Lawrence J. Fogel 等人提出的进化规划(evolutionary programming, 简称 EP)^[2,13],③由德国科学家 Ingo Rechenberg 和 Hans-Paul Schwefel 提出的进化策略(evolutionary strategies, 简称 ES 或 ESs)^[13,15]。他们用不同的进化控制模式模拟了生物进化过程,从而形成了三种具有普遍影响的模拟进化的优化计算方法。这三种方法也统称为进化算法(evolutionary algorithms, 简称 EA 或 EAs)。

进化算法是一种基于自然选择和遗传变异等生物进化机制的全局性概率搜索算法。与基于导数的解析方法和其他启发式搜索方法(如爬山方法,模拟退火方法, Monte Carlo 方法)一样,进化算法在形式上也是一种迭代方法。它从选定的初始解出发,通过不断迭代逐步改进当前解,直至最后搜索到最优解或满意解。在进化计算中,迭代计算过程采用了模拟生物体的进化机制,从一组解(群体)出发,采用类似于自然选择和有性繁殖的方式,在继承原有优良基因的基础上,生成具有更好性能指标的下一代解的群体。

优化问题采用进化计算求解的一般过程包括以下步骤:

- 1) 随机给定一组初始解;
- 2) 评价当前这组解的性能;
- 3) 根据 2) 的评价结果,从当前解中选择一定数量的解作为基因操作的对象;
- 4) 对所选择的解进行基因操作(杂交或称为交叉、突变或称为变异),得到一组新的解;
- 5) 返回到 2),对该组新的解进行评价;
- 6) 若当前解满足要求或进化过程达到一定的代数,计算结束,否则转向 3) 继

续进行。

进化算法是一种随机化搜索方法,在初始解生成以及选择、交叉与变异等遗传操作过程中,均采用了随机处理方法。与其他搜索技术(如梯度搜索技术、随机搜索技术、启发式搜索技术和枚举技术等)相比,进化算法具有以下特点^[4,7,13,14]:

1)进化算法的搜索过程是从一群初始点开始搜索,而不是从单一的初始点开始搜索,这种机制意味着搜索过程可以有效地跳出局部极值点。特别是当采用有效的保证群体多样性的措施时,进化算法可以很好地将局部搜索和全局搜索协调起来,既可以完成极值点邻域内解的求精,也可以在整个问题空间实施探索,得到问题全局最优解的概率大大提高了。

2)进化算法在搜索过程中使用的是基于目标函数值的评价信息,而不是传统方法主要采用的目标函数的导数信息或待求解问题领域内知识。进化算法的这一特点使其成为具有良好普适性和可规模化(scalability)的优化方法。

3)进化算法具有显著的隐式并行性(implicit parallelism)。进化算法虽然在每一代只对有限解个体进行操作,但处理的信息量为群体规模的高次方。

4)进化算法在形式上简单明了,不仅便于与其他方法相结合,而且非常适合于大规模并行计算机运算,因此可以有效地用于解决复杂的适应性系统模拟和优化问题。

5)进化算法具有很强的鲁棒性(robustness),即在存在噪声的情况下,对同一问题的进化算法的多次求解中得到的结果是相似的。进化算法的鲁棒性在大量的应用实例中得到了充分的验证。

1.2 遗传算法的特征与发展

Holland 的早期工作主要集中于生物学、社会学、控制工程、人工智能等领域中的一类动态系统的适应性问题(adaptation),其中适应性概念描述了在环境中表现出较好行为和性能的系统结构的渐进改变过程,简称系统的适应过程^[1,2,7]。Holland 认为,通过简单的模拟机制可以描述复杂的适应性现象。因此,Holland 试图建立适应过程的一般描述模型,并在计算机上开展模拟试验研究,分析自然系统或者人工系统对环境变化的适应性现象,其中遗传算法仅仅是一种具体的算法形式。

Bremermann, De Jong 等人则注重将遗传算法应用于参数优化问题^[4,16],极大地促进了遗传算法的应用。所以,遗传算法既是一种自然进化系统的计算模型,也是一种通用的(general purpose)求解优化问题的适应性搜索方法。目前,人们最关注和普遍使用的遗传算法是其后一种性质。

从整体上来讲,遗传算法是进化算法中产生最早、影响最大、应用也比较广泛

的一个研究方向和领域,它不仅包含了进化算法的基本形式和全部优点,同时还具备若干独特的性能^[3~7,12,15,17]:

1)在求解问题时,遗传算法首先要选择编码方式,它直接处理的对象是参数的编码集而不是问题参数本身,搜索过程既不受优化函数连续性的约束,也没有优化函数导数必须存在的要求。通过优良染色体基因的重组,遗传算法可以有效地处理传统上非常复杂的优化函数求解问题。

2)若遗传算法在每一代对群体规模为 n 的个体进行操作,实际上处理了大约 $O(n^3)$ 个模式,具有很高的并行性,因而具有显著的搜索效率。

3)在所求解问题为非连续、多峰以及有噪声的情况下,能够以很大的概率收敛到最优解或满意解,因而具有较好的全局最优解求解能力。

4)对函数的性态无要求,针对某一问题的遗传算法经简单修改即可适应于其他问题,或者加入特定问题的领域知识,或者与已有算法相结合,能够较好地解决一类复杂问题,因而具有较好的普适性和易扩充性。

5)遗传算法的基本思想简单,运行方式和实现步骤规范,便于具体使用。

鉴于遗传算法具有上述特征,一经提出即在理论上引起了高度重视,并在实际工程技术和经济管理领域得到了广泛地应用,产生了大量的成功案例。

1962年,John Holland在 Outline for a Logic Theory of Adaptive Systems 一文中^[1],提出了所谓监控程序(supervisory programs)的概念,即利用群体进化模拟适应性系统的思想。他注意到在建立智能机器的研究中,不仅可以完成单个生物体的适应性改进,而且通过一个种群的许多代的进化也可以取得非常好的适应性效果。为了获得一个好的学习方法,仅靠单个策略的改进是不够的,采用多策略的群体繁殖往往能产生显著的学习效果。尽管他当时没有给出实现这些思想的具体技术,但却引进了群体、适应值、选择、变异、交叉等基本概念。1966年,Fogel^[2]等人也提出了类似的思想,但其重点是放在变异算子而不是采用交叉算子。1967年,Holland的学生 J. D. Bagley 通过对跳棋游戏参数的研究,在其博士论文中首次提出“遗传算法”一词^[4,204]。

Holland以二进制字符集 $\{0,1\}$ 构成的代码串表示实际问题的描述结构或参数,称为“染色体”(chromosome)。对这些“染色体”进行变换,利用“染色体”中所包含的信息决定新一代“染色体”,并最终得到问题的解。这种方法对所要解决的问题类型几乎没有任何限制,所需要的信息只是每个染色体的评价价值。这种使用简单编码和选择机制的算法能够解决相当复杂的问题,并且解决实际问题时不需要该领域的专门知识。通过对这些简单的染色体进行迭代处理,从这些染色体中发现并保存好的染色体,进而逐步发现问题的最优解,这些思想就是遗传算法理论的雏形。

同时,Fraser 采用计算机模拟自然遗传系统,1962 年提出了和现在的遗传算法十分相似的概念与思想^[18]。但是,Fraser 和其他一些学者并未认识到自然遗传方法可以转化为人工遗传算法。

在 20 世纪 60 年代中期至 70 年代末期,基于自然进化的思想遭到怀疑和反对。Holland 及其数位博士坚持了这一方向的研究。在 Holland 发表论文后的十余年中,从事遗传算法研究的论文开始慢慢出现。大多数研究都集中在美国 Michigan 大学的 Holland 及其学生当中。因此,遗传算法大多数著名学者都曾经是 Michigan 大学的学生,如:David E. Goldberg、Kenneth A. De Jong、John R. Koza、Stepanie Forrest 等。1975 年,Holland 出版了专著《自然与人工系统中的适应性行为》(Adaptation in Natural and Artificial Systems)^[3],该书系统地阐述了遗传算法的基本理论和方法,提出了对遗传算法的理论发展极为重要的模式理论(schema theory),其中首次确认了选择、交叉和变异等遗传算子,以及遗传算法的隐并行性,并将遗传算法应用于适应性系统模拟、函数优化、机器学习、自动控制等领域。

另外,Daniel J. Cavicchio 的博士论文中探讨了一组实验^[11],将基于整数编码的遗传算法应用于模式识别问题,研究了保持群体差异性的选择策略。De Jong 在其博士论文研究中首次把遗传算法用于函数优化问题^[16],并对遗传算法的机理与参数设计问题进行了较为系统地研究。De Jong 深入全面地研究了模式定理和遗传算子的行为,将其与自己大量实验工作相结合,建立了著名的五函数测试平台。通过实验,他给出如下结论:①初始群体容量越大,离线性能越好,但在线性能的初始值较差。②变异可以降低某些基因的丢失机会,提高变异概率能避免成熟前收敛,但却降低在线性能。③交叉概率越大,群体中新结构的产生越快,当交叉概率等于 0.6 时,在线性能与离线性能都较好。

1975 年之后,遗传算法作为函数优化器(function optimizers)不但在各个领域得到广泛应用,而且还丰富和发展了若干遗传算法的基本理论。1980 年,Bethke 对函数优化 GA 进行了研究^[19],包括应用研究和数学分析。Smith 在 1980 年首次提出使用变长位串的概念^[20],这在某种程度上为以后的遗传规划奠定了基础。Goldberg^[4]、Davis^[21]、Grefenstette^[22]、Bauer^[23]、Srinivas 和 Patnaik^[24]等大批研究人员对遗传算法理论的基本框架和遗传算子进行了构建和改进,并将遗传算法分别应用于工程设计、自动控制、经济金融、博弈问题、机器学习等诸多领域之中。

1989 年,David Goldberg 出版了 Genetic Algorithms in Search, Optimization and Machine Learning 一书^[4],这是第一本遗传算法教科书,它是对当时关于遗传算法领域研究工作的全面而系统的总结,因而也成为引用最多的参考书之一。与 Holland 的著作侧重于适应性系统的进化数学分析不同,本书将遗传算法的基本原

理与范围广泛的应用实例相结合,并给出了大量可以使用的应用程序。1991年,Davis编辑出版了 Handbook of Genetic Algorithms^[21],其中包括了 GA 在工程技术和社会生活中的大量应用实例。

John R. Koza 将遗传算法用于处理不定长树形字符串或一组程序,提出了遗传规划(genetic programming,简称 GP)的概念^[25,26]。树状表示方法是 Koza 教授于 1989 年首次提出的,这种表示方法的主要特点之一就是染色体结构是动态变化的层次结构,它受环境影响而改变,因而对问题的表示更加自然。该方法是一种与领域无关的自适应搜索解空间的有效算法。通过增加染色体结构的复杂性,它拓展了传统遗传算法的应用范围。Koza 教授认为不同领域中许多看起来不相同的问题都可看成是寻找一定的计算机程序问题,即许多不同领域的问题都可形式化为程序归纳问题,而遗传规划提供了实现程序归纳的方法,如公式、规划(plan)、控制策略、计算程序、模型(Model)、决策树、对策策略(game-playing strategy)、转换函数、数学表达式等都称之为计算机程序。1992 年,Koza 教授出版了第一本遗传规划专著 Genetic Programming 1^[25],两年之后又出版了第二本关于遗传规划的专著^[26]。Koza 教授虽然尚未建立遗传规划的完整理论体系,但他通过大量的实验说明了遗传规划能够成功地解决一类复杂问题,为基于符号表示的函数学习问题增添了一个强有力的工具。

随着遗传算法研究和应用的不断深入与扩展,1985 年,在美国召开了第一届遗传算法国际会议,即 ICGA(International Conference on Genetic Algorithm)。这次会议是遗传算法发展的重要里程碑,此会以后每隔一年举行一次。从 1999 年起,ICGA 和 GP(Genetic Programming Society)的系列会议合并为每年一次的遗传和进化国际会议(Genetic and Evolutionary Computation Conference, GECCO)。

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中心的学术会议 FO-GA(Foundation of Genetic Algorithm)也从 1990 年起每隔一年举办一次。

1994 年 1 月,IEEE 神经网络委员会(IEEE Neural Network Council)出版了第一本“进化计算”专集;1994 年 6 月,IEEE 神经网络委员会召开第一届“进化计算”国际学术会议(IEEE ICEC),以后每年召开一次。1997 年,该委员会创办了 IEEE Transactions on Evolutionary Computation 杂志(<http://www.ewh.ieee.org/tc/nnc/pubs/tec/>),David B. Fogel 任主编。从 1999 年开始,IEEE ICEC 与 EP 的年会合并为进化计算国际会议(Congress on Evolutionary Computation, CEC),每年召开一次。

美国 MIT 出版社从 1993 年开始出版 Evolutionary Computation(<http://mit-press.mit.edu/journal-home>,De Jong 主编)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,并在更名启事中讲到——“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,都是对今后十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 GA 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,有关国际会议消息,典型应用案例和程序(源代码),等等。

这些众多的研究单位和频繁的国际学术活动中反应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。

1.3 遗传算法理论研究

与传统的启发式优化搜索算法(爬山方法、模拟退火法、Monte Carlo 方法等)相比,遗传算法(以及广义上的进化算法)的主要本质特征在于群体搜索策略和简单的遗传算子。群体搜索使遗传算法得以突破邻域搜索的限制,可以实现整个解空间上的分布式信息采集和探索;遗传算子仅仅利用适应值度量作为运算指标进行随机操作,降低了一般启发式算法在搜索过程中对人机交互的依赖。

按照生物学上可进化性(evolvability)的概念,遗传算法所追求的也是当前群体产生比现有个体更好个体的能力,即遗传算法的可进化性或称群体可进化性。因此,遗传算法的理论和方法研究也就围绕着这一目标展开。关于下面五个问题的回答,就成为 GA 理论研究的主要方向:

- 1) 遗传算法如何更好地模拟复杂系统的适应性过程和进化行为?
- 2) 遗传算法在优化问题求解中怎样才能具备全局收敛性?
- 3) 遗传算法的搜索效率如何评价?
- 4) 遗传策略的设计与参数控制的理论基础是什么?
- 5) 遗传算法与其他算法的如何结合?

其中,遗传策略包括 GA 流程设计、群体设定、群体初始化、GA 算子、终止条件等,广义上的遗传策略还包括遗传算法与其他算法结合形成的混合算法。

需要指出的是,关于遗传算法第 1) 个问题的研究主要是 Holland 的团队和 Santa Fe Institute 的 EVCA 研究组(Evolving Cellular Automata, <http://www.santafe.edu/projects/evca/>)等少数单位。大部分研究机构的研究主要集中于遗传算法和进化算法作为优化问题通用求解算法的一系列问题。

1.3.1 遗传算法的基础理论研究

在优化理论中,采用迭代算法求解一个特定问题,若该算法的搜索过程所产生的解或函数的序列的极限值是该问题的全局最优解,则该算法是收敛的。

遗传算法的基础理论主要以收敛性分析为主,即群体收敛到优化问题的全局最优解的概率。从整体上讲,可以分为基于随机过程的收敛性研究和基于模式理论的收敛性分析,我们将前者称为遗传算法的随机模型理论(stochastic modeling),后者称为遗传算法的进化动力学理论(evolution dynamics)。

1. 随机模型理论

对于有限的编码空间和有限的群体,遗传算法的搜索过程可以表示为离散时间的马尔可夫链模型(Markov chain model),从而可以采用已有的随机过程理论进行严密分析。遗传算法满足有限马尔可夫链(finite Markov chain)的基本特征,具有齐次性,存在极限概率分布。由于编码空间的有限性,标准遗传算法可以搜索到空间上的任何一个点。在采用精英保留策略下,遗传算法可以以概率 1 收敛于问题的全局最优解。

1987 年,Goldberg 和 Segrest^[58]运用有限马尔可夫链理论对遗传算法进行了收敛性分析,Eiben 等人证明了一类抽象遗传算法在 Elitist 选择情况下的概率收敛情况^[59],Rudolph 用齐次有限马尔可夫链证明了带有选择、交叉和变异操作的标准遗传算法收敛不到全局最优解,但是如果让每一代群体中的最佳个体不参加交叉与变异操作而直接保留到子代,那么遗传算法是收敛的^[60]。Qi 和 Palmieri 对浮点数编码的遗传算法,在基于连续空间中群体规模为无穷大这一假设下进行了严密的数学分析^[61]。Fogel 和 Suzuki 从进化计算的角度对 GA 收敛问题进行了研究^[63~65]。李书全等采用泛函分析理论证明了 GA 的收敛性^[66]。

Vose, Nix, Liepins 等采用统计动力学方法,分析了无穷群体下 GA 的搜索轨迹和不动点收敛性^[7]。Cerf 等采用摄动理论和马尔可夫链,对遗传算法的渐近收敛特性进行分析,得出了遗传算法运行及全局收敛性的一般性结论^[62]。

随机收敛性分析主要是在群体无穷大和进化代数趋于无限的条件之上,研究遗传算法的极限行为。事实上遗传算法的计算复杂度问题是实际应用中更为关心的问题,Bäck^[118]和 Mühlenbein^[176]等研究了到达全局最优解的遗传算法的时间复杂性问题,恽为民^[57]等基于马尔可夫链对此进行了进一步的分析。

但是,正是由于遗传算法的搜索过程的统计抽象描述,使得随机模型的收敛性分析远离了遗传算法的设计与应用。随机模型理论不仅适应于遗传算法,也适用于进化策略、进化规划、遗传规划,以及其他类随机化搜索算法。因此,尽管随机模