

王小平 曹立明 著

遗传算法

——理论、应用

与
软件
实现



西安交通大学出版社

529
TP18
W37
王小平 曹立明 著

遗传算法

——理论、应用

与
软件
实现

本书附录可从本馆主页 <http://lib.szu.edu.cn/> 上由“馆藏检索”该书详细信息后下载，也可到视听部复制



A0960840

西安交通大学出版社

· 西安 ·

内 容 提 要

遗传算法是一种借鉴生物界自然选择和进化机制发展起来的高度并行、随机、自适应搜索算法。由于其具有健壮性,特别适合于处理传统搜索算法解决不好的复杂的和非线性问题。以遗传算法为核心的进化算法已与模糊系统理论、人工神经网络等一起成为计算智能研究中的热点,受到许多学科的共同关注。

本书全面系统地介绍了遗传算法的基本理论,重点介绍了遗传算法的经典应用和国内外的新发展。全书共分 11 章。第 1 章概述了遗传算法的产生与发展、基本思想、基本操作以及应用情况;第 2 章介绍了基本遗传算法;第 3 章论述了遗传算法的数学基础;第 4 章分析了遗传算法的多种改进方法;第 5 章初步介绍了进化计算理论体系;第 6 章介绍了遗传算法应用于数值优化问题;第 7 章介绍了遗传算法应用于组合优化问题;第 8 章介绍了遗传算法应用于机器学习;第 9 章讨论了遗传算法在智能控制中的应用;第 10 章讨论了遗传算法与人工生命研究的相关问题;第 11 章介绍了遗传算法在图像处理、模式识别中的应用。

本书可供从事计算智能、自动控制、图像处理与模式识别、系统工程、优化设计、高性能计算等专业领域的研究人员阅读,也可供研究生及高年级大学生参考。

图书在版编目(CIP)数据

遗传算法——理论、应用与软件实现/王小平,曹立明
著. —西安:西安交通大学出版社,2002.1
ISBN 7-5605-1448-0

I. 遗… II. ①王… ②曹… III. 遗传-算法
IV. TP18

中国版本图书馆 CIP 数据核字(2001)第 079562 号

*

西安交通大学出版社出版发行
(西安市兴庆南路 25 号 邮政编码:710049 电话:(029)2668315)
陕西宝石兰印务有限责任公司印装
各地新华书店经销

*

开本:787 mm×1 092mm 1/16 印张:22.125 字数:537 千字
2002 年 1 月第 1 版 2002 年 1 月第 1 次印刷
印数:0 001~3 000 定价:40.00 元(平)
48.00 元(精)

发行科电话:(029)2668357,2667874

前 言

19 世纪中叶, 查尔斯·达尔文(Charles Darwin)在总结前人进化思想的基础上, 用大量的科学事实证明了生物进化过程在总体上表现为: 从低级到高级, 从简单到复杂, 从不完善形式到完善形式, 从单一适应到多种适应, 从低的有序性到高的有序性, 以及沿着物种数目日益增多的方向发展进化。达尔文认为, 生物进化的动力和机制在于自然选择。自然选择是用变异作材料, 通过生存斗争实现的。凡是具有适应环境的有利变异的个体, 在生存斗争中将有更多机会生存和繁殖后代, 而适应性较差的个体将被淘汰, 因此, 生物进化便是“物竞天择, 适者生存”的过程。达尔文提出进化思想之后的一个半世纪以来, 现代生物学将遗传和进化列为生物学的主要研究对象, 一直在发展和完善达尔文的进化学说。

科学技术从蒸汽机时代到计算机时代、信息社会的演化过程中, 几乎每项发明和技术革新都可以视为规模不等的搜索试验工程。搜索的起点总是多个已经部分成功的过去答案, 改变和选择的标准是是否具备类似自然的环境适应性, 而舍弃的不是一个而是多个答案。莱特(Wright)兄弟发明的飞机便是很好的技术进化范例, 在他们之前设计飞行物并不是新鲜的事物, 但他们在前人失败(也包含局部成功)的基础上, 经过对飞机结构无数次局部的改变、组合和经历实际考验的比较和选择过程, 飞机终于能够像类人猿由爬行发展到直立行走一样“飞”了起来。莱特兄弟当时既没有从达尔文进化论和后来的孟德尔(G. Mendel)遗传学说中寻找灵感和线索的可能, 也不具备现代的计算手段, 但他们显然从基于朴素的、类似自然选择的人工选择方式搜索更好答案的实践中得到了好处, 实现了人类飞行的梦想。

人工智能是人类发展到计算机时代的又一个梦想。然而近 40 多年来, 人工智能的研究经历了多次成功与失败的起伏, 计算复杂性的困难和认知手段的贫乏并没有因为电子计算机技术本身的突飞猛进得到解决。传统的人工智能基于物理符号进行逻辑推理和启发性搜索求解, 虽然取得了巨大的成功, 但一般只能适合于较为狭窄的应用领域。物理符号计算假设认为: 任何达到人类智能水平的智能系统至少在本质上可以看作一个符号处理系统; 知识工程的观点认为: “知识就是力量”, 智能行为决定于知识的数量和质量; 连接主义假设认为: 智能产生于众多并行分布的神经元的相互作用; 进化学派认为: 智能行为是控制和自然环境综合作用下进化的结果。目前的问题在于, 是否有必要或者有可能在常识知识的学习和获取自动化、知识表达方式普适性、搜索求解高效率和全局化、智能体活化于环境等多个方面获得统一的方法论, 很多学者对解决这些问题的一些新的理论和方法持失望或怀疑的态度, 原因也在于此。毋庸讳言, 人工智能又一次陷入低潮时期, 但也可以认为是一个新的力量积蓄时期。人工智能是否也能有朝一日真的“飞”起来, 我们虽然不能根据新的理论所获得的证据简单地推测, 但对此, 正像相信空气动力学助推航空航天事业的发展一样深信不疑。

1975 年, 美国密歇根(Unoversity of Michigan)大学的心理学教授、电子工程学与计算机科学教授 John. H. Holland 和他的同事与学生共同研究了具有开创意义的遗传算法理论和方法。遗传算法是一种借鉴生物界自然选择和进化机制发展起来的高度并行、随机、自适应搜索算法。简单而言, 它使用了群体搜索技术, 将种群代表一组问题解, 通过对当前种群施加选择、交

叉和变异等一系列遗传操作,从而产生新一代的种群,并逐步使种群进化到包含近似最优解的状态。由于其思想简单、易于实现以及表现出来的健壮性,遗传算法赢得了许多应用领域,特别是近年来在问题求解、优化和搜索、机器学习、智能控制、模式识别和人工生命等领域取得了许多令人鼓舞的成就。以遗传算法为核心的进化算法,已与模糊系统理论、人工神经网络等一起成为计算智能研究中的热点,受到广泛的关注。

本书全面系统地介绍了遗传算法的基本理论和方法,重点介绍了遗传算法的经典应用和国内外的新发展。全书共分 11 章。第 1 章概述了遗传算法的产生与发展、基本思想、基本操作以及应用情况。第 2~5 章构成遗传算法理论和实现技术的基本框架,其中第 2 章介绍了基本遗传算法;第 3 章论述了遗传算法的数学基础;第 4 章分析了遗传算法的多种改进方法;第 5 章初步介绍了进化计算理论体系,并着重介绍了遗传程序设计问题。第 6~11 章按遗传算法不同的应用领域,分别介绍了遗传算法的许多应用问题,其中第 6 章介绍了遗传算法应用于数值优化问题;第 7 章介绍了遗传算法应用于组合优化问题;第 8 章介绍了遗传算法应用于机器学习;第 9 章讨论了遗传算法在智能控制中的应用;第 10 章讨论了遗传算法与人工生命研究的相关问题;第 11 章介绍了遗传算法在图像处理、模式识别中的应用。最后,在本书的附录部分提供了有关遗传算法及进化计算的国际学术组织及其活动情况的信息,以及用 C 语言编写的基本遗传算法实现和几个应用问题的源程序,以便读者参考。此外,本书附有一张 CD-ROM 光盘,其中除实例源程序代码及可执行程序外,还包括大量国内、外相关研究的最新参考文献及一些遗传算法自由软件的程序,为读者进一步学习和研究遗传算法提供了方便。

本书是作者近年来潜心学习和研究国内外遗传算法理论、方法和应用成果的一个总结。目的是为读者提供有关遗传算法的全面了解。本书写作时,考虑到遗传算法是一种新的计算智能方法,因此,对遗传算法、进化计算的生物学背景以及应用问题的背景,都作了必要的诠释,并尽可能多地包括了一些新的见解、思想以及应用领域,当然也包括了作者自己的许多研究工作和心得。本书由王小平和曹立明合作编著,由王小平撰写全部初稿,曹立明负责修正,并由两人统一定稿。

在本书的写作过程中,得到了同济大学计算机系教授、博士生导师施鸿宝的热心指导和鼓励;西安交通大学刘路放博士审阅了书稿,并提出了许多宝贵意见,特在此向他们表示衷心的感谢。本书的完成得到了铁道部科技发展项目(编号 B99127-01484)和国家自然科学基金项目(批准号 70171061)的资助。另外,在写作过程中,参考了大量的相关书籍和文献,也向这些作者致以诚挚的谢意!本书的第一作者于 1996~1998 年作为访问学者在日本国东京大学工学部学习研究期间,岛田庄平教授和冈野靖彦教授给予了热情的指导并提供了方便的研究条件;同济大学计算机系领导给予了作者许多工作和生活上的帮助和关心;本书的出版得到了西安交通大学出版社的支持,谨在此一并表示由衷的感谢。

由于本书涉及的研究范围很广,近年的研究成果如雨后春笋,层出不穷。限于作者的学识水平,书中疏漏及不当之处在所难免,感谢读者批评指正。

作 者

2000 年 12 月于上海

目 录

前言

第 1 章 遗传算法简介

1.1 遗传算法的产生与发展	(1)
1.2 遗传算法概要	(4)
1.3 遗传算法的基本操作	(10)
1.4 遗传算法的应用情况	(14)
参考文献	(16)

第 2 章 基本遗传算法

2.1 简单函数优化的实例	(18)
2.2 遗传基因型	(21)
2.3 适应度函数及其尺度变换	(25)
2.4 遗传操作	(28)
2.5 算法设计与实现	(38)
参考文献	(50)

第 3 章 遗传算法的数学基础

3.1 模式定理	(51)
3.2 Walsh 模式变换	(55)
3.3 非均匀 Walsh 模式变换	(61)
3.4 欺骗问题	(61)
3.5 遗传算法动态分析	(65)
参考文献	(66)

第 4 章 遗传算法的改进

4.1 分层遗传算法	(68)
4.2 CHC 算法	(69)
4.3 messy GA	(71)
4.4 自适应遗传算法	(73)
4.5 基于小生境技术的遗传算法	(74)
4.6 混合遗传算法	(76)
4.7 并行遗传算法	(79)
参考文献	(85)

第5章 进化计算初步

5.1 进化计算理论的基本框架	(87)
5.2 进化策略	(89)
5.3 进化规划	(91)
5.4 遗传程序设计	(92)
参考文献	(102)

第6章 遗传算法与数值优化

6.1 最优化问题	(104)
6.2 只含上、下限约束的最优化问题	(105)
6.3 优化的约束问题解决	(108)
6.4 多目标优化问题	(115)
参考文献	(122)

第7章 遗传算法与组合最优化

7.1 巡回旅行商问题	(123)
7.2 作业调度问题	(130)
7.3 背包问题	(136)
参考文献	(140)

第8章 遗传算法与机器学习

8.1 基于遗传算法的机器学习	(143)
8.2 密歇根方法	(144)
8.3 匹茨堡方法	(150)
8.4 学习与进化之间的交互	(150)
参考文献	(156)

第9章 遗传算法与智能控制

9.1 线性系统辨识	(158)
9.2 非线性系统辨识	(165)
9.3 控制系统设计自动化和高性能实现	(168)
9.4 基于遗传算法优化的模糊控制	(173)
9.5 遗传优化神经网络控制	(189)
9.6 混合软计算技术	(202)
参考文献	(205)

第10章 遗传算法与人工生命

10.1 人工生命概述	(210)
10.2 基于遗传算法的人工生命模型设计	(214)

10.3 蚁群协作觅食模拟模型·····	(219)
10.4 免疫系统模型·····	(222)
10.5 进化硬件问题·····	(225)
10.6 进化对策论·····	(229)
参考文献·····	(232)

第 11 章 遗传算法与图像处理、模式识别

11.1 图像歪斜校准·····	(235)
11.2 图像分割·····	(236)
11.3 图像基元识别与提取·····	(247)
参考文献·····	(251)

附录

I 有关遗传算法及进化计算的国际学术组织及其活动情况·····	(254)
I.1 遗传算法、进化计算相关的学术会议·····	(254)
I.2 知名的学术研究机构·····	(255)
I.3 Internet 上的相关资源·····	(258)
I.4 自由软件·····	(259)
II 源程序清单·····	(260)
II.1 基本遗传算法源程序·····	(260)
II.2 基本遗传学习分类系统源程序·····	(272)
II.3 遗传优化神经网络源程序·····	(294)
II.4 遗传识别提取基元源程序·····	(304)
II.5 基于遗传算法的人工生命模拟源程序·····	(315)
III 名词术语中英文对照·····	(339)

第 1 章 遗传算法简介

遗传算法(Genetic Algorithms, GA)研究的历史比较短,20 世纪 60 年代末期到 70 年代初期,主要由美国 Michigan 大学的 John Holland 与其同事、学生们研究形成了一个较完整的理论和方法,从试图解释自然系统中生物的复杂适应过程入手,模拟生物进化的机制来构造人工系统的模型。随后经过 20 余年的发展,取得了丰硕的应用成果和理论研究的进展,特别是近年来世界范围形成的进化计算热潮,计算智能已作为人工智能研究的一个重要方向,以及后来的人工生命研究兴起,使遗传算法受到广泛的关注。从 1985 年在美国卡耐基·梅隆大学召开的第一届国际遗传算法会议(International Conference on Genetic Algorithms: ICGA'85),到 1997 年 5 月 IEEE 的 Transactions on Evolutionary Computation 创刊,遗传算法作为具有系统优化、适应和学习的高性能计算和建模方法的研究渐趋成熟。本章在介绍遗传算法的产生和发展历史之后,概述了遗传算法的基本理论和应用情况。

1.1 遗传算法的产生与发展

早在 20 世纪 50 年代和 60 年代,就有少数几个计算机科学家独立地进行了所谓的“人工进化系统”研究,其出发点是进化的思想可以发展成为许多工程问题的优化工具。早期的研究形成了遗传算法的雏形,如大多数系统都遵行“适者生存”的仿自然法则,有些系统采用了基于种群(population)的设计方案,并且加入了自然选择和变异操作,还有一些系统对生物染色体编码进行了抽象处理,应用二进制编码。60 年代初期,柏林工业大学的 I. Rechenberg 和 H. P. Schwefel 等在进行风洞实验时,由于设计中描述物体形状的参数难以用传统方法进行优化,因而利用生物变异的思想来随机改变参数值,并获得了较好的结果。随后,他们对这种方法进行了深入的研究,形成了进化计算的另一个分支——进化策略(Evolutionary Strategy, ES),如今 ES 和 GA 已呈融合之势。也是在 20 世纪 60 年代, L. J. Fogel 等人在设计有限态自动机(Finite State Machine, FSM)时提出了进化规划(Evolutionary Programming, EP),他们借用进化的思想对一组 FSM 进行进化,以获得较好的 FSM。他们将此方法应用到数据诊断、模式识别和分类及控制系统的设计等问题中,取得了较好的结果。后来又借助进化策略方法发展了进化规划,并用于数值优化及神经网络的训练等问题中。

由于缺乏一种通用的编码方案,人们只能依赖变异而非交叉来产生新的基因结构,早期的算法收效甚微。20 世纪 60 年代中期, John Holland 在 A. S. Fraser 和 H. J. Bremermann 等人工作的基础上提出了位串编码技术。这种编码既适用于变异操作,又适用于交叉(即杂交)操作,并且强调将交叉作为主要的遗传操作。随后, Holland 将该算法用于自然和人工系统的自适应行为的研究中,并于 1975 年出版了其开创性著作“Adaptation in Natural and Artificial Sys-

tems”。以后, Holland 等人将该算法加以推广, 应用到优化及机器学习等问题中, 并正式定名为遗传算法。遗传算法的通用编码技术和简单有效的遗传操作为其广泛、成功地应用奠定了基础。Holland 早期有关遗传算法的许多概念一直沿用至今, 可见 Holland 对遗传算法的贡献之大。他认为遗传算法本质上是适应算法, 应用最多的是系统最优化的研究。

Holland 早期的工作集中在所谓的认知系统 CS 1 (Cognitive System 1) 的研究, 借助最优化的方法获取学习的规则, 遗传算法是他考虑的途径之一。于是他将基于遗传的机器学习 (Genetic-based Machine Learning, GBML) 方法发展成为 CS 1 的分类系统 (Classifier System) 学习方法, 奠定了遗传算法重要思想的基础。遗传算法适用于最优化问题, 归功于 Holland 的学生 De Jong, 而 Grefenstette 开发了第一个遗传算法软件——称为 GENESIS, 为遗传算法的普及推广起了重要作用。对遗传算法研究影响力最大的专著, 要属于 1989 年美国伊利诺大学的 Goldberg 所著的“Genetic Algorithms in Search, Optimization, and Machine Learning”。这本书对于遗传算法理论及其多领域的应用展开了较为全面的分析和例证。1992 年, Michalewicz 出版了另一本很有影响力的著作“Genetic Algorithms + Data Structures = Evolution Programs”, 对遗传算法应用于最优化问题起到了推波助澜的作用, 1994 年该书又再版发行。

20 世纪 70 年代以来, 关于遗传算法的博士论文, 比较有代表性的有 A. D. Bethke 的“作为函数优化器的遗传算法”(密歇根大学, 1980 年)、De Jong 的“一类遗传自适应系统的行为分析”(密歇根大学, 1975 年)、T. E. Davis 的“从模拟退火收敛理论向简单遗传算法的外推”(佛罗里达大学, 1991 年)。表 1.1 列出了遗传算法理论的经典研究成果。

表 1.1 遗传算法理论的经典研究成果

年份	贡献者	内容
1962	Holland	程序漫游元胞计算机自适应系统框架
1968	Holland	模式定理的建立
1971	Hollstein	具有交配和选择规则的二维函数优化
1972	Bosworth, Foo, Zeigler	提出具有复杂变异、类似于遗传算法的基因操作
1972	Frantz	位置非线性和倒位操作研究
1973	Holland	遗传算法中试验的最优配置和双臂强盗问题
1973	Martin	类似遗传算法的概率算法理论
1975	De Jong	用于 5 个测试函数的研究基本遗传算法基准参数
1975	Holland	出版了开创性著作《Adaptation in Natural and Artificial Systems》
1981	Bethke	应用 Walsh 函数分析模式
1981	Brindle	研究遗传算法中的选择和支配问题
1983	Pettit, Swigger	遗传算法应用于非稳定问题的粗略研究
1983	Wetzel	用遗传算法解决旅行商问题(TSP)
1984	Mauldin	基本遗传算法中用启发知识维持遗传多样性
1985	Baker	试验基于排序的选择方法
1985	Booker	建议采用部分匹配计分、分享操作和交配限制法

续表 1.1

年份	贡献者	内容
1985	Goldberg, Lingle	TSP 问题中采用部分匹配交叉
1985	Grefenstette, Fitzpatrick	对含噪声的函数进行测试
1985	Schaffer	多种群遗传算法解决多目标优化问题
1986	Goldberg	最优种群大小估计
1986	Grefenstette	元级遗传算法控制的遗传算法
1987	Baker	选择中随机误差的减少方法
1987	Goldberg	复制和交叉时最小欺骗问题(MDP)
1987	Goldberg, Richardson	借助分享函数的小生境和物种归纳法
1987	Goldberg, Segrest	复制和交叉的有限马尔可夫链
1987	Goldberg, Smith	双倍染色体遗传算法应用于非稳定函数优化
1987	Oliver, Smith, Holland	排列重组算子的模拟和分析
1987	Schaffer, Morishima	串编码自适应交叉试验
1987	Whitley	子孙测试应用于遗传算法的选择操作

20 余年来,遗传算法的应用无论是用来解决实际问题还是建模,其范围不断扩展,这主要依赖于遗传算法本身的逐渐成熟。近年来,许多冠以“遗传算法”的研究与 Holland 最初提出的算法已少有雷同之处,不同的遗传基因表达方式,不同的交叉和变异算子,特殊算子的引用,以及不同的再生和选择方法,但这些改进方法产生的灵感都来自大自然的生物进化,可以归为一个“算法簇”。人们用进化计算(Evolutionary Computation)来包容这样的遗传“算法簇”。它基本划分为四个分支:遗传算法(GA)、进化规划(EP)、进化策略(ES)和遗传程序设计(GP)。遗传算法研究热潮的兴起,人工智能再次成为人们关注的焦点。有些学者甚至提出,进化计算是人工智能的未来。其观点是,虽然我们不能设计人工智能(即用机器代替人的自然智能),但我们可以利用进化通过计算获得智能。目前,进化计算与人工神经网络、模糊系统理论一起已形成一个新的研究方向——计算智能(computational intelligence)。人工智能已从传统的基于符号处理的符号主义,向以神经网络为代表的连接主义和以进化计算为代表的进化主义方向发展。

应该说,20 世纪 80 年代中期以来是遗传算法和进化计算的蓬勃发展期。以遗传算法、进化计算为主题的多个国际会议在世界各地定期召开。1985 年,在美国卡耐基·梅隆大学召开的第一届国际遗传算法会议 ICGA'85,以后该会议每隔一年举行一次。1997 年夏季在美国密歇根大学召开了 ICGA'97。现在与之平行进行的国际会议很多,其中 International Conference on Evolutionary Programming 和 IEEE International Conference on Evolutionary Computation 也分别召开了 6 届和 4 届。此外,每年夏季在美国斯坦福大学召开有关遗传程序设计的国际会议(The Annual Conference of Genetic Programming)。有关遗传算法基础理论的学术活动也很活跃,第一届遗传算法与分类系统研讨会(The First Workshop on the Foundations of Genetic Algorithms and Classifier Systems, FOGA/CS)1990 年在美国印第安那大学召开,以后

每隔两年召开一次,从会议论文中选编的论文集:“Foundations of Genetic Algorithms I”和“Foundations of Genetic Algorithms II”两卷已由 MIT 出版社出版发行。在欧洲“Parallel Problem Solving from Nature:FPSN”为题的国际会议从 1990 年开始在德国举行以来,在比利时和德国隔年轮流举行。同样值得一提的是,国际互联网上也有多种相关的 mailing list, Usenet 上还有专门的新闻组 comp. ai. genetic。由于进化计算应用广泛,一些杂志及国际会议论文集中都有这方面的文章,现在还出版了两种关于进化计算的新杂志“Evolutionary Computation”和“IEEE Transactions on Evolutionary Computation”,一些国际性期刊也竞相出版这方面的专刊。另外,日本新的计算机发展规划 RWC 计划(Real World Computing Program)也把遗传算法、进化计算作为其主要支撑技术之一,用来进行信息的集成、学习及组织等。

1980 年以来,人们越来越清楚地意识到传统人工智能方法的局限性,而且随着计算机速度的提高及并行计算机的普及,遗传算法和进化计算对机器速度的要求已不再是制约其发展的因素。德国 Dortmund 大学 1993 年末的一份研究报告表明,根据不完全统计,进化算法已在 16 个大领域、250 多个小领域中获得了应用。遗传算法在机器学习、过程控制、经济预测、工程优化等领域取得的成功,已引起了数学、物理学、化学、生物学、计算机科学、社会科学、经济学及工程应用等领域专家的极大兴趣。某些学者研究了进化计算的突现行为(emergent behavior)后声称,进化计算与混沌理论、分形几何将成为人们研究非线性现象和复杂系统的新的三大方法,并将与神经网络一起成为人们研究认知过程的重要工具。20 世纪 90 年代以后,人们比较重视遗传算法的一些基本问题,De Jong 称为“重访基本的假设”,这方面的研究内容主要有①表示和形态发生学;②拉马克算子等的引入;③非随机配对和物种形成;④分散的、高度并行的模型;⑤自适应系统;⑥共同进化系统。同时由于遗传算法在应用研究方面的长处主要得益于其求解的有效性、现有仿真环境下易于实现、可扩充性和易于与其他方法相结合,可以预料在不远的将来,随着理论研究的不断深入和应用领域的不断拓广,遗传算法和进化计算将取得长足的发展。

我国有关遗传算法、进化计算的研究,从 20 世纪 90 年代以来一直处于不断上升的时期,特别是近年来,遗传算法、进化计算的应用在许多领域取得了令人瞩目的成果。据不完全统计,1997~1999 年三年间发表在国内二级以上学术刊物上有关遗传算法、进化计算的文章接近 200 篇左右,该类研究获得不同渠道的经费资助比例也在逐年上升。武汉大学刘勇、康立山等于 1995 年出版了《非数值并行计算(第 2 册)——遗传算法》;陈国良、王煦法等于 1996 年出版《遗传算法及其应用》;潘正军、康立山等于 1998 年出版了《演化计算》;周明、孙树栋于 1999 年出版了《遗传算法原理及其应用》。国内有关遗传算法的 BBS 电子公告牌有国家智能中心曙光站 bbs. neic. ac. cn、北京大学阳光创意站 bbs. pku. edu. cn、清华大学水木清华站 bbs. net. tsinghua. edu. cn、西安交通大学兵马俑站 bbs. xanet. edu. cn 等。

本书附录 I 介绍了有关遗传算法及进化计算的国内、外学术组织及其活动情况。

1.2 遗传算法概要

1.2.1 生物进化理论和遗传学的基本知识

在介绍遗传算法之前,有必要了解有关的生物进化理论和遗传学的基本知识。

我们知道,生命的基本特征包括生长、繁殖、新陈代谢和遗传与变异。生命是进化的产物,

现代的生物是在长期进化过程中发展起来的。达尔文(1858年)用自然选择(natural selection)来解释物种的起源和生物的进化,其自然选择学说包括以下三个方面:

(1) **遗传(heredity)** 这是生物的普遍特征,“种瓜得瓜,种豆得豆”,亲代把生物信息交给子代,子代按照所得信息而发育、分化,因而子代总是和亲代具有相同或相似的性状。生物有了这个特征,物种才能稳定存在。

(2) **变异(variation)** 亲代和子代之间以及子代的不同个体之间总有些差异,这种现象,称为变异。变异是随机发生的,变异的选择和积累是生命多样性的根源。

(3) **生存斗争和适者生存** 自然选择来自繁殖过剩和生存斗争。由于弱肉强食的生存斗争不断地进行,其结果是适者生存,具有适应性变异的个体被保留下来,不具有适应性变异的个体被淘汰,通过一代代的生存环境的选择作用,物种变异被定向着一个方向积累,于是性状逐渐和原先的祖先种不同,演变为新的物种。这种自然选择过程是一个长期的、缓慢的、连续的过程。

达尔文的进化理论是生物学史上的一个重要里程碑,它解释了自然选择作用下生物的渐变式进化。1866年孟德尔发表了“植物杂交实验”的论文,他提出的遗传学的两个基本规律——分离律和自由组合律,奠定了现代遗传学的基础。随着细胞学的发展,染色体、减数分裂和受精过程相继被发现,Water S. Sutton发现染色体的行为与基因的遗传因子行为是平行的,因此提出遗传因子是位于染色体上的。美国遗传学家摩尔根(T. H. Morgan)进一步确立了染色体的遗传学说,认为遗传形状是由基因决定的,染色体的变化必然在遗传形状上有所反映。生物的形状往往不是简单地决定于单个基因,而是不同基因相互作用的结果,基因表达要求一定的环境条件,同一基因型在不同的环境条件下可以产生不同的表现型。20世纪20年代以来,随着遗传学的发展,一些科学家用统计生物学和种群遗传学的成就重新解释达尔文的自然选择理论,他们通过精确地研究种群基因频率由一代到下一代的变化,来阐述自然选择是如何起作用的,形成现代综合进化论(synthetic theory of evolution)。种群遗传学是以种群为单位而不是以个体为单位的遗传学,是研究种群中基因的组成及其变化的生物学。在一定地域中,一个物种的全体成员构成一个种群(population),种群的主要特征是种群内的雌雄个体能够通过有性生殖实现基因的交流。生物的进化实际上是种群的进化,个体总是要消亡,但种群则是继续保留,每一代个体基因型的改变会影响种群基因库(gene pool)的组成。而种群基因库组成的变化就是这一种群的进化,没有所谓的生存斗争问题,单是个体繁殖机会的差异也能造成后代遗传组成的改变,自然选择也能够进行。综合进化论对达尔文式的进化给予了新的更加精确的解释。

生物进化非常复杂,现有的进化理论所不能解释的问题比已经解释的问题还要多。除了达尔文的渐变进化外,人们又提出了很多新的非达尔文式进化理论,如木村资生的分子进化中性理论(neutral theory of molecular evolution)、Goldschmidt的跳跃进化(saltation)、N. Eldredge的间断平衡进化(Punctuated Equilibrium Evolution)等。随着生物学的前沿领域——生物物理、分子生物学和生物化学的发展,关于生物进化的理论仍在发展之中,但以自然选择为核心的进化理论比其他学说的影响广泛而深远,它仍然是各种生物进化理论的一个重要基础。

遗传算法模拟的是怎样的生物进化模型呢?假设对相当于自然界中的一群人的一个种群进行操作,第一步的选择是以现实世界中的优胜劣汰现象为背景的;第二步的重组交叉则相当于人类的结婚和生育;第三步的变异则与自然界中偶然发生的变异是一致的。人类偶然出现

的返祖现象便是一种变异。由于包含着对模式的操作,遗传算法不断地产生出更加优良的个体,正如人类向前进化一样。所采用的遗传操作都与生物尤其是人类的进化过程相对应。如果我们再仔细分析遗传算法的操作对象种群,实际上它对应的是一群人,而不是整个人类。一群人随着时间的推移而不断地进化,并具备越来越多的优良品质。然而,由于他们的生长、演变、环境和原始祖先的局限性,经过相当一段时间后,他们将逐渐进化到某些特征相对优势的状态(例如中国人都是黄皮肤、黑眼睛以及特有的文化和社会传统习惯),我们定义这种状态为平衡态。当一个种群进化到这种状态,这种种群的特性就不再有很大的变化了。一个简单的遗传算法,从初始代开始,并且各项参数都设定,也会达到平衡态。此时种群中的优良个体仅包含了某些类的优良模式,因为该遗传算法的设置特性参数使得这些优良模式的各个串位未能得到平等的竞争机会。

现实世界中许多民族,每个民族都有各自的优缺点。历史上民族之间通过多种形式的交流(包括战争、移民等),打破了各个民族的平衡态,从而推动他们达到更高层次的平衡态,使整个人类向前进化。现实生活中的例子可以在生物实验室中找到,为了改良动、植物品种,常常采用杂交、嫁接等措施,即是为了这个目的。

既然遗传算法效法基于自然选择的生物进化,是一种模仿生物进化过程的随机方法。下面先给出几个生物学的基本概念与术语,这对于理解遗传算法是非常重要的。

染色体(chromosome) 生物细胞中含有的一种微小的丝状化合物。它是遗传物质的主要载体,由多个遗传因子——基因组成。

脱氧核糖核酸(DNA) 控制并决定生物遗传性状的染色体主要是由一种叫做脱氧核糖核酸(deoxyribonucleic acid 简称 DNA)的物质构成。DNA 在染色体中有规则地排列着,它是个大分子的有机聚合物,其基本结构单位是核苷酸。每个核苷酸有四种称为碱基的环状有机化合物中的一种、一分子戊糖和磷酸分子组成。许多核苷酸通过磷酸二酯键相结合形成一条长长的链状结构,两个链状结构再通过碱基间的氢键有规律地扭合在一起,相互卷曲起来形成一种双螺旋结构。

核糖核酸(RNA) 低等生物中含有一种核糖核酸(ribonucleic acid, 简称 RNA)的物质,它的作用和结构与 DNA 类似。

遗传因子(gene) DNA 或 RNA 长链结构中占有一定位置的基本遗传单位,也称为基因。生物的基因数量根据物种的不同多少不一,小的病毒只含有几个基因,而高等动、植物的基因却数以万计。一个基因或多个基因决定了组成蛋白质的 20 种氨基酸的组成比例及其排列顺序。

遗传子型(genotype) 遗传因子组合的模型叫遗传子型。它是性状染色体的内部表现,又称基因型。一个细胞核中所有染色体所携带的遗传信息的全体称为一个基因组(genome)。

表现型(phenotype) 由染色体决定性状的外部表现,或者说,根据遗传子型形成的个体,称为表现型。

基因座(locus) 遗传基因在染色体中所占据的位置。同一基因座可能有的全部基因称为等位基因(allele)。

个体(individual) 指染色体带有特征的实体。

种群(population) 染色体带有特征的个体的集合称为种群。该集合内个体数称为群体的大小。有时个体的集合也称为个体群。

进化(evolution) 生物在其延续生存的过程中,逐渐适应其生存环境,使得其品质不断得到改良,这种生命现象称为进化。生物的进化是以种群的形式进行的。

适应度(fitness) 在研究自然界中生物的遗传和进化现象时,生物学家使用适应度这个术语来度量某个物种对于生存环境的适应程度。对生存环境适应程度较高的物种将获得更多的繁殖机会,而对生存环境适应程度较低的物种,其繁殖机会就会相对较少,甚至逐渐灭绝。

选择(selection) 指决定以一定的概率从种群中选择若干个体的操作。一般而言,选择的过程是一种基于适应度的优胜劣汰的过程。

复制(reproduction) 细胞在分裂时,遗传物质 DNA 通过复制而转移到新产生的细胞中,新的细胞就继承了旧细胞的基因。

交叉(crossover) 有性生殖生物在繁殖下一代时两个同源染色体之间通过交叉而重组,亦即在两个染色体的某一相同位置处 DNA 被切断,其前后两串分别交叉组合形成两个新的染色体。这个过程又称基因重组 recombination,俗称“杂交”。

变异(mutation) 在细胞进行复制时可能以很小的概率产生某些复制差错,从而使 DNA 发生某种变异,产生出新的染色体,这些新的染色体表现出新的性状。

编码(coding) DNA 中遗传信息在一个长链上按一定的模式排列,也即进行了遗传编码。遗传编码可以看作从表现型到遗传子型的映射。

解码(decoding) 从遗传子型到表现型的映射。

1.2.2 遗传算法的基本思想

现在,我们引用上面的术语来更好地描述遗传算法的基本思想。遗传算法是从代表问题可能潜在解集的一个种群(population)开始的,而一个种群则由经过基因(gene)编码(coding)的一定数目的个体(individual)组成。每个个体实际上是染色体(chromosome)带有特征的实体。染色体作为遗传物质的主要载体,即多个基因的集合,其内部表现(即基因型)是某种基因组合,它决定了个体的形状的外部表现,如黑头发的特征是由染色体中控制这一特征的某种基因组合决定的。因此,在一开始需要实现从表现型到基因型的映射即编码工作。由于仿照基因编码的工作很复杂,我们往往进行简化,如二进制编码。初代种群产生之后,按照适者生存和优胜劣汰的原理,逐代(generation)演化产生出越来越好的近似解。在每一代,根据问题域中个体的适应度(fitness)大小挑选(selection)个体,并借助于自然遗传学的遗传算子(genetic operators)进行组合交叉(crossover)和变异(mutation),产生出代表新的解集的种群。这个过程将导致种群像自然进化一样的后生代种群比前代更加适应于环境,末代种群中的最优个体经过解码(decoding),可以作为问题近似最优解。

遗传算法采纳了自然进化模型,如选择、交叉、变异、迁移、局域与邻域等。图 1.1 表示了基本遗传算法的过程。计算开始时,一定数目 N 个个体(父个体 1、父个体 2、父个体 3、父个体 4……)即种群随机地初始化,并计算每个个体的适应度函数,第一代也即初始代就产生了。如果不满足优化准则,开始产生新一代的计算。为了产生下一代,按照适应度选择个体,父代要求基因重组(交叉)而产生子代。所有的子代按一定概率变异。然后子代的适应度又被重新计算,子代被插入到种群中将父代取而代之,构成新一代(子个体 1、子个体 2、子个体 3、子个体 4……)。这一过程循环执行,直到满足优化准则为止。

尽管这样单一种群的遗传算法很强大,可以很好地解决相当广泛的问题。但采用多种群即有子种群的算法往往会获得更好的结果。每个子种群像单种群遗传算法一样独立地演算若

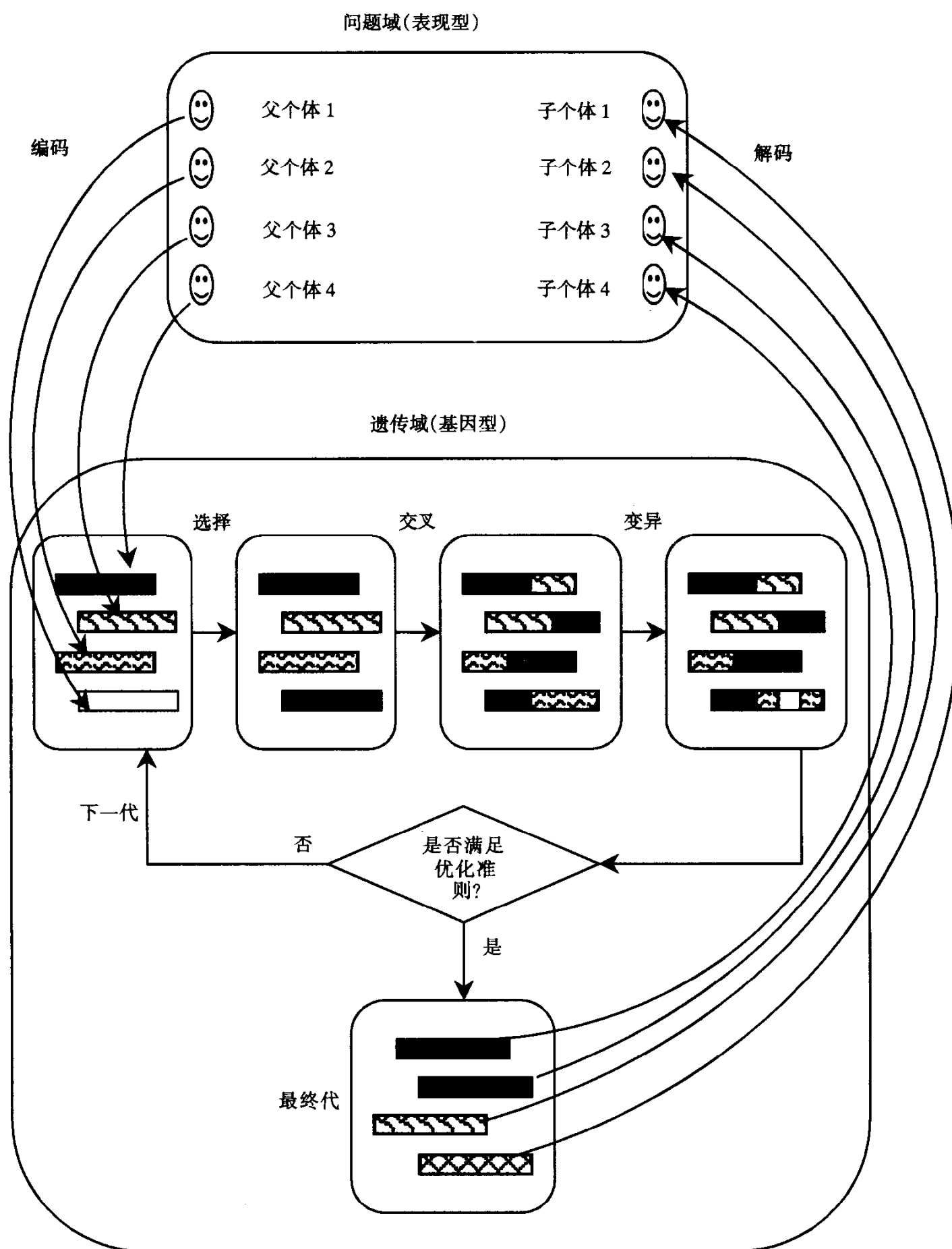


图 1.1 遗传算法的过程

千代后,在子种群之间进行个体交换。这种多种群遗传算法更加贴近于自然中种族的进化,称为并行遗传算法(Paralleling Genetic Algorithm, PGA),此算法在第4章中将详细介绍。

1.2.3 遗传算法的特点

我们知道,传统的优化方法主要有三种:枚举法、启发式算法和搜索算法:

(1) **枚举法** 枚举出可行解集合内的所有可行解,以求出精确最优解。对于连续函数,该方法要求先对其进行离散化处理,这样就可能因离散处理而永远达不到最优解。此外,当枚举空间比较大时,该方法的求解效率比较低,有时甚至在目前先进计算工具上无法求解。

(2) **启发式算法** 寻求一种能产生可行解的启发式规则,以找到一个最优解或近似最优解。该方法的求解效率比较高,但对每一个需求解的问题必须找出其特有的启发式规则,这个启发式规则一般无通用性,不适合于其他问题。

(3) **搜索算法** 寻求一种搜索算法,该算法在可行解集合的一个子集内进行搜索操作,以找到问题的最优解或者近似最优解。该方法虽然保证不了一定能够得到问题的最优解,但若适当地利用一些启发知识,就可在近似解的质量和效率上达到一种较好的平衡。

随着问题种类的不同以及问题规模的扩大,要寻求一种能以有限的代价来解决搜索和优化的通用方法,遗传算法正是为我们提供的一个有效的途径,它不同于传统的搜索和优化方法。主要区别在于:

① **自组织、自适应和自学习性(智能性)**。应用遗传算法求解问题时,在编码方案、适应度函数及遗传算子确定后,算法将利用进化过程中获得的信息自行组织搜索。由于基于自然的选择策略为“适者生存,不适应者被淘汰”,因而适应度大的个体具有较高的生存概率。通常,适应度大的个体具有更适应环境的基因结构,再通过基因重组和基因突变等遗传操作,就可能产生更适应环境的后代。进化算法的这种自组织、自适应特征,使它同时具有能根据环境变化来自动发现环境的特性和规律的能力。自然选择消除了算法设计过程中的一个最大障碍,即需要事先描述问题的全部特点,并要说明针对问题的不同特点算法应采取的措施。因此,利用遗传算法的方法,我们可以解决那些复杂的非结构化问题。

② **遗传算法的本质并行性**。遗传算法按并行方式搜索一个种群数目的点,而不是单点。它的并行性表现在两个方面,一是遗传算法是内在并行的(inherent parallelism),即遗传算法本身非常适合大规模并行。最简单的并行方式是让几百甚至数千台计算机各自进行独立种群的演化计算,运行过程中甚至不进行任何通信(独立的种群之间若有少量的通信一般会带来更好的结果),等到运算结束时才通信比较,选取最佳个体。这种并行处理方式对并行系统结构没有什么限制和要求,可以说,遗传算法适合在目前所有的并行机或分布式系统上进行并行处理,而且对并行效率没有太大影响。二是遗传算法的内含并行性(implicit parallelism)。由于遗传算法采用种群的方式组织搜索,因而可同时搜索解空间内的多个区域,并相互交流信息。使用这种搜索方式,虽然每次只执行与种群规模 n 成比例的计算,但实质上已进行了大约 $O(n^3)$ 次有效搜索,这就使遗传算法能以较少的计算获得较大的收益。

③ **遗传算法不需求导或其他辅助知识,而只需要影响搜索方向的目标函数和相应的适应度函数**。

④ **遗传算法强调概率转换规则,而不是确定的转换规则**。

⑤ **遗传算法可以更加直接地应用**。

⑥ **遗传算法对给定问题,可以产生许多的潜在解,最终选择可以由使用者确定(在某些特殊情况下,如多目标优化问题不止一个解存在,有一组 pareto 最优解。这种遗传算法对于确认可替代解集而言是特别合适的)。**