

船舶与海洋工程系列
CHUANBO YU HAIYANG GONGCHENG XILIE



遗传算法理论及其 在船舶横摇运动控制中的应用

● 金鸿章 王科俊 何 琳◎著



哈尔滨工程大学出版社
Harbin Engineering University Press

船舶与海洋工程系列
CHUANBO YU HAIYANG GONGCHENG XILIE



遗传算法理论及其 在船舶横摇运动控制中的应用

● 金鸿章 王科俊 何 琳◎著

 哈尔滨工程大学出版社
Harbin Engineering University Press

内 容 简 介

遗传算法是模拟生物自然进化的过程和机制发展起来的智能优化算法,在许多行业和多个学科领域得到广泛应用和关注。本书重点阐述遗传算法的收敛分析等理论成果和其在船舶横摇运动控制中的应用。全书共分9章,第1章给出了遗传算法的基本原理及其改进方法;第2章~第4章研究了遗传算法的收敛性定义、收敛性分析和收敛速度估计方法;第5章探讨了遗传算法的种群多样性;第6章给出了模糊遗传算法;第7章提出了综合改进的遗传算法;第8章和第9章介绍了遗传算法在船舶横摇运动建模和控制中的应用成果。

本书可以作为研究遗传算法的参考书,可供智能科学、控制科学、计算机科学、系统科学和管理科学等领域的人员阅读,也可供高年级本科生和研究生参考。

图书在版编目(CIP)数据

遗传算法理论及其在船舶横摇运动控制中的应用/金鸿章,
王科俊,何琳著. —哈尔滨:哈尔滨工程大学出版社,2007
ISBN 978-7-81073-649-7

I. 遗… II. ①金…②王…③何… III. 遗传-算法-
应用-船舶减摇装置 IV. U664.7

中国版本图书馆 CIP 数据核字(2007)第 129148 号

出版发行 哈尔滨工程大学出版社
社 址 哈尔滨市南岗区东大直街 124 号
邮政编码 150001
发行电话 0451-82519328
传 真 0451-82519699
经 销 新华书店
印 刷 黑龙江省教育厅印刷厂
开 本 787mm×1 092mm 1/16
印 张 10
字 数 212 千字
版 次 2007 年 10 月第 2 版
印 次 2007 年 10 月第 2 次印刷
定 价 20.00 元
<http://press.hrbeu.edu.cn>
E-mail: heupress@hrbeu.edu.cn



前 言

自从 1975 年美国 Michigan 大学的 John Holland 教授正式提出遗传算法以来,经过几十年的发展,取得了许多理论进步和丰硕的应用成果。自 20 世纪 90 年代中期以来在世界范围内已形成了进化计算的研究热潮,作为智能科学的重要研究方向和人工生命研究的重要工具,遗传算法的研究备受关注。

作为一种通用的优化算法,遗传算法具有编码和遗传操作简单,优化不受限制条件的约束等特点。其并行性和全局搜索能力使其在几乎所有需要使用优化算法的领域都得到了成功应用。20 世纪 90 年代以后,遗传算法被广泛应用于各种复杂系统的自适应控制以及复杂的优化问题中。

在本书中,主要介绍两方面的内容。在第 1 章至第 7 章中,介绍了遗传算法的一些理论问题,以及我们在遗传算法研究中的一些成果。主要介绍遗传算法的基本原理和它的改进方法;研究了遗传算法收敛性定义和统一数学描述,收敛速度,最优保留遗传算法及其收敛性;研究了遗传算法种群多样性的问题;把模糊理论用于遗传算法的研究中,进行了模糊遗传算法研究;还对综合改进遗传算法进行了研究。在第 8 章和第 9 章中,介绍了遗传算法的应用问题。主要介绍遗传算法在舰船横摇运动的模糊建模中的应用;遗传算法在舰船减摇鳍控制器优化设计中的应用。虽然书中以舰船为对象,针对舰船横摇运动建模的不确定性和非线性问题,进行了舰船横摇运动的模糊建模和减摇鳍控制器优化设计,但是这种方法也可以应用到其他具有不确定性和非线性系统的建模和控制系统设计中去。

在第 1 章中,我们在介绍遗传算法的生物学基础上,进一步介绍遗传算法的基本原理和它的改进方法。

在第 2 章中,对遗传算法收敛性定义和统一数学描述进行了研究。在遗传算法的理论研究中,遗传算法收敛性分析占有很重要的地位,全面而准确的收敛性定义和判据是进行遗传算法收敛性分析的基础。本章从种群个体适应度和状态分布两个角度给出不同层次上的遗传算法收敛性定义,同时给出遗传算法两种统一的清晰 Markov 链模型,并在这两种模型基础上,从分析简单遗传算法不收敛的原因入手,直观地给出了使简单遗传算法收敛的方法,并用实例验证这些方法的有效性,使目前遗传算法在这方面的研究统一化和系统化。

在第 3 章中,对遗传算法收敛性和收敛速度进行了研究。针对遗传算法收敛性的结论是在时间趋于无穷的前提下得到的问题,本章从遗传算法的状态转

移概率矩阵出发,把简单遗传算法及其各种变形归纳为三种抽象数学模型,给出其对应的收敛性和收敛速度问题的数学描述,期望建立遗传算法收敛性研究的统一基础。并以此为基础,利用有限 Markov 链分别从随机矩阵和矩阵特征值两个角度分析上述三种类型遗传算法的收敛性和它们的收敛速度问题。

第 4 章是对最优保留遗传算法及其收敛性分析研究。本章分析了最优保留遗传算法的运行机制及其收敛的本质原因和充分条件;给出最优保留遗传算法更一般的规范化定义及两种实现方式,并分析了它们的收敛性;由此得到影响最优保留遗传算法全局收敛的关键因子。在此基础上,提出一种变形的最优保留遗传算法,并证明了其整体收敛性。

在第 5 章中,进行了遗传算法种群多样性的研究。本章分析遗传算法过早收敛现象的表现及产生原因,指出它与种群多样性变化的关系。提出新的种群多样性度量方法及其定量计算方法,为进一步理解遗传算子对种群多样性的影响,通过实验研究它们在遗传算法中的作用及其相互影响。

模糊遗传算法问题在第 6 章中进行了研究。基于模糊逻辑的遗传算法是当前遗传算法发展的一个新方向。本章在阐述模糊遗传算法的基本设计思想基础上,根据第 5 章对遗传算法种群多样性的研究结果给出一种新的模糊遗传算法。

在第 7 章中,提出了综合改进遗传算法。以提高遗传算法搜索效率和鲁棒性为目的,从遗传算法的结构、基因操作和参数设置向自适应、自组织形式发展,并进行以系统综合为主要实现方式的遗传算法综合改进算法的研究是遗传算法应用研究的主要方向。

对具有强不确定性和严重非线性对象的控制一直是控制理论研究的一个重要研究内容。在第 8 章和第 9 章中,我们以舰船为对象,研究遗传算法在舰船横摇运动的模糊建模中的应用,研究遗传算法在舰船减摇鳍控制器优化设计中的应用。

尽管已提出了船舶小角度横摇运动的线性数学模型,但它在大幅度横摇时,非线性比较严重,且水动力参数有严重的不确定性。第 8 章在给出模糊建模的结构和辨识方法的基础上,用模糊神经网络来映射 T-S 模型的结构建立舰船横摇的模糊模型,用第 6 章给出的模糊遗传算法辨识模型的结构和参数,用仿真来证明这种建模方法的有效性。

遗传算法作为一种新的优化技术,在控制系统的优化设计中有重要的应用。针对舰船减摇鳍控制系统的特点,第 9 章提出了适于减摇鳍控制器优化设计的改进的遗传算法。在线性和非线性情况下,对不同海况下的减摇鳍控制器的 PID 参数进行优化,并与传统的单纯形法优化的结果进行对比,结果显示了遗传算法在复杂系统参数优化中的有效性及优越性。

算法理论分析通常应该伴随在算法的设计过程之中。但对于遗传算法而

言,由于它具有太广泛的应用领域和遗传操作间的非线性作用及其所带有的随机性和不确定性等因素,遗传算法本身构成了一个复杂系统,这给遗传算法的理论分析带来了很大困难。正因为如此,目前遗传算法理论方面的研究工作还相当有限,结果也不太深入,但在遗传算法的应用研究方面却取得了一些令人鼓舞的成果。

本书的编著过程中,参考了大量的相关文献。作者借本书出版之机会,向与课题相关的同仁和学生们,向参考和引用过他们文献的作者致以真诚的感谢!

著 者



目 录

第 1 章 遗传算法的基本原理及其改进方法	1
1.1 生物进化与遗传	1
1.2 遗传算法研究的发展与现状	2
1.3 遗传算法的基本原理	3
1.4 遗传算法的理论研究	7
1.5 遗传算法的改进研究	8
1.6 综合改进方法的研究	19
第 2 章 遗传算法收敛性定义和统一数学描述	23
2.1 数学基础	23
2.2 遗传算法的收敛性定义	24
2.3 遗传算法的统一数学描述	26
2.4 概率模型	31
第 3 章 遗传算法收敛性和收敛速度的研究	33
3.1 数学基础	33
3.2 GA 收敛性和收敛速度问题的数学描述	34
3.3 基于随机矩阵的 GA 收敛性分析和收敛速度估计	34
3.4 基于特征值的 GA 收敛性分析和收敛速度估计	37
第 4 章 最优保留遗传算法及其收敛性分析	50
4.1 最优保留与最优保留遗传算法	50
4.2 全局收敛的 EGA 及其收敛本质	50
4.3 全局收敛 EGA 的实现方法及其收敛性分析	51
4.4 一种新形式的 EGA	54
4.5 一种自适应 GA 的收敛性及时齐性分析	56
第 5 章 遗传算法种群多样性的研究	60
5.1 过早收敛现象的表征及产生原因	60
5.2 种群多样性的定义和度量	61
5.3 种群多样性的含义	67
5.4 遗传算子对种群多样性的影响	68
第 6 章 模糊遗传算法	78
6.1 模糊遗传算法的基本设计思想	78
6.2 基于种群多样性的模糊遗传算法	82
第 7 章 综合改进遗传算法	86

7.1	现有 GA 改进方法及其在 TMRGA 中的应用	87
7.2	TMRGA 的描述	88
7.3	仿真结果及 TMRGA 性能分析	92
第 8 章	遗传算法在船舶横摇运动的模糊建模中的应用	100
8.1	船舶横摇运动的传统数学模型	100
8.2	模糊遗传算法在 T-S 模型辨识中的应用	102
8.3	船舶横摇模糊建模	110
第 9 章	遗传算法在船舶减摇鳍控制器优化设计中的应用	115
9.1	船舶减摇鳍控制系统构成	115
9.2	减摇鳍控制规律	118
9.3	适用于减摇鳍 PID 控制器优化设计的改进的遗传算法	119
9.4	线性情况下基于遗传算法的减摇鳍系统 PID 控制器的优化设计	123
9.5	非线性情况下基于遗传算法的减摇鳍 PID 控制器的优化设计	129
参考文献		141



第 1 章

遗传算法的基本原理及其改进方法

自从 1975 年美国 Michigan 大学的 John Holland 教授正式提出遗传算法以来,经过几十年的发展,取得了许多理论进步和丰硕的应用成果。20 世纪 90 年代中期以来在世界范围内已形成了进化计算的研究热潮,作为人工智能的重要研究方向和人工生命研究的重要工具,遗传算法的研究备受关注。作为一种通用的优化算法,遗传算法具有编码和遗传操作简单,优化不受限制条件的约束的特点,其并行性和全局搜索能力使其在几乎需要使用优化算法的领域得到了成功应用。本章将在介绍遗传算法的生物学基础上,介绍遗传算法的基本原理和它的改进方法。

1.1 生物进化与遗传

遗传算法是受生物进化学说和遗传学说的启发而发展起来的。因此在介绍遗传算法之前,有必要介绍有关生物进化理论和遗传学说的基本知识。

我们知道,地球上的生物都是在长期进化过程中发展起来的,生命的基本特征包括生长、繁殖、新陈代谢和遗传与变异。达尔文(1858 年)用自然选择来解释物种的起源和生物的进化,其学说的主要内容:地球上的生物具有很强的繁殖能力,能够产生许多后代。生物的不断繁殖使后代的数目大量增加,而在自然界中生物赖以生存的资源是有限的,因此,为了生存,生物就需要竞争。只有那些对环境适应能力强的生物才能在竞争中生存下来,即适者生存,不适者消亡,这是自然选择的结果。生物在繁殖过程中,通过遗传,使物种保持相似。与此同时,由于变异,物种会产生差别,甚至形成新的物种。

达尔文的进化论解释了自然选择作用下生物的渐变式进化。孟德尔(1866 年)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。经过 100 多年的发展,人们提出了许多生物进化和遗传理论,如,染色体的遗传学说,种群遗传学说,现代综合进化论等。

遗传算法就是借用生物进化和遗传的规律,通过繁殖、遗传、变异、竞争,实现优胜劣汰的随机优化方法。为了便于理解遗传算法,有必要给出几个生物学的基本概念和术语。

染色体(chromosome):生物细胞中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):染色体中决定生物遗传性质的物质。DNA 是有规则的双螺旋结构,由两条平行的脱氧核苷酸长链盘旋而成。每个核苷酸由四种称为碱基的环状有机化合物组成。DNA 中的两条链上的碱基通过氢链连接起来,形成碱基对。

基因(gene):有遗传效应的 DNA 片段。基因是控制生物遗传的物质单元,每个基因含有

成百上千个脱氧核苷酸。它们在染色体上呈线性排列,排列的顺序代表了遗传信息。

基因型(genotype):基因组合的模型。它是性状染色体的内部表现。一个细胞核中的所有染色体所携带的遗传信息的全体成为一个基因组(genome)。

表现型(phenotype):基因形成的个体,是由染色体性状决定的外部表现。

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

个体(individual):染色体带有特征的实体。

种群(population):个体的集合。该集合内个体数成为种群的大小。

进化(evolution):生物在其延续生存的过程中,逐渐适应其生存环境,使其品质不断得到改良的现象。

适应度(fitness):生物学家用来度量某个物种对生存环境适应程度的量。适应度大的物种繁殖机会多,生存的可能性大;适应度小的物种繁殖机会少,生存的可能性小,甚至逐渐灭绝。

选择(selection):以一定的概率从种群中选择若干个体的操作。

复制(reproduction):在细胞分裂时,将遗传物质 DNA 转移到新产生的细胞中的操作。

交叉(crossover):有性生殖生物在繁殖下一代时,两个同源染色体发生基因交换的操作。这一操作又称基因重组,俗称“杂交”。

变异(mutation):包括基因突变和染色体变异。基因突变是指基因分子结构的改变,包括 DNA 碱基对的增添、缺失或改变;染色体变异是指染色体在结构上或数目上的变化,其中数目的变化对新染色体的产生起着很大的作用。

编码(coding):DNA 中遗传信息在一个长链上按一定模式排列,即进行了遗传编码,实质是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 遗传算法研究的发展与现状

遗传算法是 20 世纪 60 ~ 70 年代由美国 Michigan 大学的 Holland 教授及其学生和同事发展起来的。20 世纪 60 年代初, Holland 教授开始认识到生物的自然遗传现象与人工自适应行为的相似性,他认为不仅要研究自适应的系统,还要研究与之相关的环境,因此他提出在研究和设计人工自适应系统时,可以借鉴生物自然遗传的基本原理,模仿生物自然遗传的基本方法。1967 年,他的学生 Bageley 在博士论文中首次提出“遗传算法(Genetic Algorithms)”一词。此后, Holland 指导学生完成了多篇博士论文。到 20 世纪 70 年代初, Holland 提出了“模式定理(Schema Theorem)”(一般认为是“遗传算法的基本定理”),从而奠定了遗传算法研究的理论基础。

在遗传算法研究的历史上 1975 年是十分重要的一年。这一年, Holland 出版了他的著名专著《Adaptation in Nature and Artificial System》,该书是作者十几年间许多思想及其实现的结晶,它详细阐述了遗传算法的理论,并为其奠定了数学基础,发展了一整套模拟生物自适应系统的理论。同年, De Jong 完成了对遗传算法研究具有指导意义的博士论文“An Analysis of a Class of Genetic Adaptive System”,他基于模式定理做了大量严格的计算实验,给出了明确的结论;并在此基础上给出了著名的 De Jong 五函数测试平台,定义了性能评价标准,并以

函数优化为例,对遗传算法的六种方案的性能及机理进行了详细的实验和分析,他的工作成为后继者的范例并为遗传算法以后的广泛应用奠定了坚实的基础。

20世纪80年代以后,遗传算法被广泛应用于各种复杂系统的自适应控制以及复杂的优化问题中。Goldberg在遗传算法的研究中起着继往开来的作用,他在1983年的博士论文中第一次把遗传算法用于实际的工程系统——煤气管道工程的优化。从此,引起了各学科学者对遗传算法应用的广泛兴趣,使遗传算法的理论研究更为深入,应用研究成果更为丰富。随着遗传算法理论的进一步完善,遗传算法的应用领域的不断拓展,越来越受到国际学术界的重视。1985年,在美国召开了第一届关于遗传算法及其应用的国际会议,并且成立了国际遗传算法学会,此会以后每两年举行一次。1989年,Goldberg出版了《Genetic Algorithms in Search, Optimization and Machine Learning》,总结了遗传算法研究的主要成果,对遗传算法及其应用作了全面而系统的论述。一般认为,这一时期的遗传算法从古典阶段发展到了现代阶段,该书则奠定了现代遗传算法的基础。1991年,Davis编辑出版了《Handbook of GAs》,其中包括了遗传算法在工程技术和社会生活大量的应用实例。

除了GA本身的发展外,另一方面,由于GA的开放性和可扩展性,使得它很容易与专家系统(Expert System)、模糊逻辑(Fuzzy Logic)、神经网络(Neural Network)以及模拟退火(Simulated Annealing,简称SA)、混沌理论(Chaos)等智能计算方法相互渗透和结合,这方面的研究对开拓21世纪新的智能计算技术具有重要的意义。GA方面的文章也不断在《Artificial Intelligence》,《Machine Learning》等杂志上发表。《Machine Learning》杂志已有数期GA专刊,人工智能和机器学习的国际会议也相继为GA开辟了专刊。1994年在美国奥兰多召开的IEEE全球计算智能大会上,进化计算同模糊逻辑、神经网络一样,被列为会议的三项主题之一,而且将它们三者统称为计算智能,成为研究智能控制的重要基础。GA与模拟退火、神经网络、确定退火(Deterministic Annealing)、弹性网(Elastic Net)等一起被称作物理计算(Physical Computation,或按自然法则计算),因为它们都是模拟某些自然规律的结果。Solving Parallel Problem from Nature主要是围绕遗传算法、进化策略、遗传规划和模拟退火的国际会议,从1991年开始每两年举办一次。

近十几年来,国内许多学科及专业的学者也开始研究和应用遗传算法,并发表了数篇较有影响的综述性文章,对国内学者进一步研究及应用遗传算法起到了积极作用。近几年来,无论是理论研究,还是应用研究,国内的学者都取得了许多成果,使我国在遗传算法研究上处于与国际同一水平线上。

1.3 遗传算法的基本原理

遗传算法的内涵哲理启迪于自然界生物从低级、简单,到高级、复杂的漫长而绝妙的进化过程,借鉴于达尔文的物竞天择,适者生存的自然遗传机理,其本质是一种求解问题的高效并行全局搜索方法。它能在搜索过程中自动获取和积累有关搜索空间的信息,并自适应地控制搜索过程以求得最优解。

遗传算法实际上是模拟由个体组成群体的整体学习过程,其中每个个体表示给定问题搜索空间的一个解点。它从任一初始化的群体出发,通过随机选择(使种群中优秀的个体有更多的机会传到下一代)、交叉(体现了自然界中种群个体间的信息交换)和变异(在种群中引入新的变种确保种群中信息的多样性)等遗传操作,使群体一代一代地进化到搜索空间

中越来越好的区域,直至抵达最优点。本节将简单介绍 GA 的基本实现方法和步骤,及三种主要遗传算子的特点。

1.3.1 遗传算法的基本结构

GA 是 Holland 教授于 20 世纪 60 年代末基于进化及遗传理论提出来的优化搜索方法,是模拟自然界生物进化过程的计算模型。简单 GA 解决问题的基本步骤如下:

1. 将问题的解表示成编码串(生物学术语称为“染色体”),每一码串代表问题的一个可行解;
2. 随机产生一定数量的初始码串群 pop_0 , 该种群就是问题可行解的一个集合;
3. 将初始码串群置于问题的“环境”当中,并给出种群中每一个体码串适应问题环境的适应值(评价);
4. 根据码串个体适应度的高低对初始种群 pop_0 (或 pop_k) 执行选择操作(selection), 随机选取父本种群 F_k , 优良的个体被大量复制, 而劣质的个体复制得少, 甚至被淘汰掉;
5. 由父本种群 F_k 以交叉概率 p_c 经交叉(crossover) 产生种群 C_k ;
6. 对种群 C_k 以变异概率 p_m 执行变异(mutation) 操作得新的种群 $pop_{(k+1)}$ 。

这样反复执行第 3 步到第 6 步,使码串群体一代代不断进化,最后搜索到最适应问题环境的个体,求得问题的最优解。

简单 GA 的操作过程如图 1-1 所示。

The procedure of simple genetic algorithm

```

Begin
  t = 0
  initialize the population P(t)
  evaluate the population P(t)
  While (not termination-condition) do
    t = t + 1
    select P(t) from P(t-1)
    recombine P(t)
    evaluate P(t)
  End
End

```

图 1-1 遗传算法的基本流程

从 GA 的基本步骤可看出,GA 有三种基本操作:

1. 选择(selection or reproduction)

依据码串的适应值确定该个体在下一代复制的数目。种群中第 i 个个体期望被复制的数目:

$$n_r = M \cdot f(s_i) / \sum_{i=1}^M f(s_i) \quad (1-1)$$

其中, M 为种群规模, $f(s_i)$ 为个体 s_i 的适应度值, 且 $f(s_i)$ 为正值。为执行复制操作, 应先计算当前种群中各个体 s_i 的适应度值 $f(s_i)$, 并计算出种群总的适应度值

$$F = \sum_{i=1}^M f(s_i)$$

及每一个体串被选中的概率

$$p_s(s_i) = f(s_i)/F \quad (1-2)$$

与相应的个体串 s_i 的累积概率

$$q_i = \sum_{j=1}^i p_s(s_j)$$

典型的复制过程是执行回放式随机采样(轮盘赌) M 次,每次选中一个个体组成新的群体,具体方法如下:

1) 每次在区间 $[0,1]$ 上产生一个随机浮点数 r ;

2) 如果 $q_{i-1} \leq r \leq q_i$, 则第 i 个个体被复制一次。适应度大的个体被选中的概率大,相应地被复制的次数就多;相反,适应度小的个体被选中的概率就小,复制的次数也少,甚至得不到复制,被淘汰掉。选择(或复制)是遗传算法的关键,它体现了自然界中适者生存的思想。

2. 交叉(crossover)

按一定的概率从种群中随机选出一定数量的个体并随机组对,然后从某一随机选定的位置开始交换两个个体串的某些位,其中概率 P_c 称为交叉概率,交叉概率给出了期望参与交叉的个体数量 $n_c = P_c \cdot M$,其中 M 为种群规模。交叉的具体方法如图 1-2 所示。

$\cdot f\bar{a}$	交叉	$\bar{z}q\bar{a}$
$A = (a_1 a_2 a_3 a_4 : a_5 a_6 a_7 \cdots a_l)$		$A = (a_1 a_2 a_3 a_4 : b_5 b_6 b_7 \cdots b_l)$
$B = (b_1 b_2 b_3 b_4 : b_5 b_6 b_7 \cdots b_l)$		$B = (b_1 b_2 b_3 b_4 : a_5 a_6 a_7 \cdots a_l)$

图 1-2 单点交叉

交叉体现了自然界中信息交换的思想,是 GA 主要搜索算子,也是 GA 区别于其他搜索算法重要特征。

3. 变异(mutation)

按一定的概率 P_m 对群体中某些个体的某些位进行变异,对二进制编码而言,即码值由 1 变为 0,或由 0 变为 1。其中概率 P_m 称为变异概率, B_m 给出了期望变异的个体码值位数

$$B_m = P_m \cdot l \cdot M \quad (1-3)$$

其中 M 为种群规模, l 为个体串长度。变异的具体操作如图 1-3 所示。变异算子模仿了自然界的基因突变现象,保证了算法能搜索到问题解空间的每一点,从而使 GA 具有全局寻优能力。

$\cdot f\bar{a}$
$A = (a_1 0_2 a_3 a_4 a_5 1_6 a_7 \cdots a_l)$
子代
$B = (a_1 1_2 a_3 a_4 a_5 0_6 a_7 \cdots a_l)$
◇ ◇

图 1-3 变异操作

◇—变异位置

1.3.2 三种基本遗传操作的数学描述

设个体的编码字母表 $A = \{0, 1, \dots, \alpha - 1\}$, 个体串的长度为 l , 则个体空间表示为 $S = A^l = \{0, 1, \dots, \alpha - 1\}^l$, 且 $|S| = \alpha^l$. 设在遗传过程中, 种群规模固定为 M , 则种群空间和父本空间分别表示为

$$S^M = \{(s_1, s_2, \dots, s_M) \mid s_i \in S, i = 1, 2, \dots, M\},$$

$$S^2 = \{(s_1, s_2) \mid s_1, s_2 \in S\}$$

由上述观点及 GA 的基本思想, 选择、交叉及变异三种遗传操作可看作是在空间 S^M, S^2, S 间的随机映射, 它们是对自然界中一些遗传机制类似的、理想的抽象。由此可给出上述三种遗传操作的严格的概率定义。

1) 选择 $T_s: S^M \rightarrow S^2$ 表示一个随机映射, 代表算法允许父代种群中个体进行繁殖下一代的可能性;

根据自然选择规律, 常以与个体适应度成比例的方式来定义选择概率。即对任何给定的种群 $X \in S^M$, 在 T_s 的作用下选取 $(s_1, s_2) \in S^2$ 为父本的概率为。

$$P\{T_s(X) = (s_1, s_2)\} = \frac{f(s_1)}{\sum_{s_i \in X} f(s_i)} \cdot \frac{f(s_2)}{\sum_{s_j \in X} f(s_j)} \quad (1-4)$$

其中 $1 \leq i \leq M, 1 \leq j \leq M$ 。

2) 交叉 $T_c: S^2 \rightarrow S$ 它作用于两个个体而产生一个新的个体, 是对生物种群有性繁殖的抽象。对于给定的父本 $s_i, i = 1, 2$, 以概率 P_c 产生个体 s 的概率为

$$P\{T_c[(s_1, s_2)] = s\} = \begin{cases} \frac{k \cdot P_c}{l}, & \text{当 } y \neq s_1 \\ (1 - P_c) + \frac{k \cdot P_c}{l}, & \text{当 } y = s_1 \end{cases} \quad (1-5)$$

其中 $k = k(s_1, s_2)$ 是由 s_1, s_2 交叉可成功产生个体的交叉点位置个数, y 表示交叉产生的个体, 是中间变量。

3) 变异 $T_m: S \rightarrow S$ 对于给定的个体 $s_i \in S$ 和变异概率 P_m , T_m 对 s_i 的作用在于独立地以概率 P_m 对 s_i 的各分量进行改变, 其意义在于模拟生物进化中基因的突变。这样, 变异算子定义为

$$P\{T_m(s_i) = s\} = P_m^{|s_i - s|} (1 - P_m)^{l - |s_i - s|} \quad (1-6)$$

其中 $|s_i - s|$ 为个体 s_i 与 s 间的 Hamming 距离。

1.3.3 遗传算法的特点

遗传算法具有很强的鲁棒性, 这是因为比起普通的优化搜索方法, 它采用了许多独特的方法和技术, 归纳起来, 主要有以下几个方面。

1. 遗传算法的处理对象不是参数本身, 而是对参数集进行了编码的个体。此编码操作, 使得遗传算法可直接对结构对象进行操作。这一特点使得遗传算法具有广泛的应用领域。

2. 遗传算法采用同时处理群体中多个个体的方法, 即同时对搜索空间中的多个解进行评估。更形象地说, 遗传算法是并行的爬多个峰。这一特点使遗传算法具有较好的全局搜索性能, 减少了陷入局部最优的风险。同时, 这使遗传算法本身也十分易于并行化。

3. 在标准遗传算法中,基本上不用搜索空间的知识或其他辅助信息,而仅用适应度函数值来评估个体,并在此基础上进行遗传操作。需要着重指出的是,遗传算法的适应度函数不仅不受连续可微的约束,而且其定义域可以任意设定。对适应度函数的唯一要求是,对于输入可计算出加以比较的正的输出。遗传算法的这一特点使它的应用范围大大扩展。

4. 遗传算法不是采用确定性规则,而是采用概率的变迁规则来指导它的搜索方向。

上述的这些具有特色的技术和方法使得遗传算法使用简单,鲁棒性强,易于并行化,从而应用范围甚广。

1.4 遗传算法的理论研究

算法理论分析通常应该伴随在算法的设计过程之中。但对于 GA 而言,由于它具有太广泛的应用领域和遗传操作间的非线性作用及其所带有的随机性和不确定性等因素,GA 本身构成了一个复杂系统,这给 GA 的理论分析带来了很大困难。正因为如此,目前 GA 理论方面的研究工作还相当有限,研究结果也不太深入。本小节主要介绍与本书研究有密切关系的 GA 两部分理论研究现状。

1.4.1 收敛性分析和收敛速度估计

1975 年 Holland 在其开创性著作中给出了 GA 的模式定理,从模式操作的角度分析 GA 的性能特点,奠定了 GA 的理论基础。模式定理是目前影响最广,研究最多的一种 GA 理论,虽然它可在一定程度上估计 GA 在迭代过程中模式的期望变化,但它无法对种群的构成、收敛性、收敛速度及适应度的分布等做出适当的估计,尚不能从根本上证明 GA 的收敛性,GA 的全局收敛性问题一直倍受关注。

目前有关 GA 收敛性的研究都是在对算法加某些限制条件后而得到 GA 的修正形式的基础上进行的,主要从以下几个方面进行了研究。

1. 目前在分析 GA 收敛性时,较常用的是把 GA 整个群体看作一种状态分布,在一定条件下把 GA 的运行过程看成一个随时间演变的有限的马尔可夫链。关于 GA 收敛性的大部分结果均是基于这种随机过程数学模型基础上的。Rodolph 运用齐次有限 Markov 链证明了标准 GA 是非全局收敛的,不适合于静态函数优化,建议改变复制、变异策略以达到全局收敛。Eiben 利用无限 Markov 链证明了具有最优保留个体策略的 GA 概率性全局收敛。恽为民等运用齐次有限 Markov 链证明简单 GA 不是全局收敛的,最优保留简单 GA 是全局收敛的。Qi 在种群规模为无穷大的假设下,对浮点数编码的 GA 进行了全局收敛性分析^[20,21]。张讲社、徐宗本等提出了一种结合模拟退火的改进 GA,并给出改进算法全局收敛的充要条件。徐宗本、高勇等分析了 GA 过早收敛的起因与特征,提出了一种可预防和克服过早收敛的改进 GA,并证明了改进算法依概率收敛到全局最优状态。

2. 采用新的数学模型和数学工具,从另一个角度看待 GA 的运行过程,是目前比较新颖的研究方向。N.J. Radcliffe 提出了一种形态(forma)分析法,该方法把搜索空间按照某种等价关系划分为等价类,以便进行分析。A. Prugel-Bennett 等则采用统计力学的方法对 GA 进行分析。Qi X F 和 Palmieri F 利用条件概率与边缘分布对 GA 进行了研究,给出了 GA 演变公式,但不是 Markov 过程,没有给出最终的演变结果。Yao L 和 Sethares W A 利用随机变数的均值性质分析了 GA 的收敛性问题。王丽薇、洪家荣从知识表示与算符相结合的角度考察 GA,提

出了一个 GA 收敛的充分条件。

3. 直接考虑 GA 的收敛性比较困难,需要寻找恰当的数学工具和数学模型,需要较深的数学知识。一些学者根据现有算法和 GA 的相似性(或构造与 GA 等价的算法),利用现有算法的理论基础稍作改变来进行 GA 的收敛性分析。梁艳春、周春光提出了一种与标准 GA 等价的 GA,从理论上探讨了 GA 的收敛条件和收敛速度问题。张铃、张钊分析了 GA 与 SA 算法的相似性,提出了 SMA 算法,将以前发展的一套 SA 算法的理论移植到 GA 中来,为 GA 算法的精度、可信度和计算复杂性的定量分析提供了有力的理论工具。另外,分析 GA 与常规算法结合而成的新算法的收敛性对 GA 收敛性的研究也是有帮助的。Lin W, Delgado-Frias J G 分析了牛顿 - 拉夫逊遗传算法的收敛性。徐洪泽证明了带有爬山算子 GA 的收敛性。

GA 收敛速度的研究具有非常重要的理论及实际意义。它能够从理论上对各类修正算法的优劣提供评判标准,并且,可对给出的实用改进 GA 进行理论上的指导。目前探讨这一问题的文章很少,也不深入,而且大部分结果仅为纯数学表达形式,未与 GA 具体的遗传控制参数相联系,缺乏实际意义,离利用 GA 收敛速度估计结果来分析和评判算法优劣还有一定距离。Suzuki 对简单 GA 的收敛速度估计进行了初步探讨。彭宏、王兴华给出了无变异条件下 GA 的收敛速度估计。Rosenthal 研究了 Markov 链的收敛速度,可用于分析到达全局最优解 GA 的时间复杂性。

1.4.2 控制参数和遗传算子对 GA 性能的影响

目前这方面的研究较少,而且所得到的一些参数设置与 GA 性能的关系中仅考虑了某一个参数的影响。Coli M 和 Palazzari P 提出了一种根据不同问题寻找最佳编码方案,建立有意义的积木块来提高 GA 收敛速度的方法。Kiasimir Dr, Kolarov 分析了选择在整个 GA 运行中起的重要作用及种群大小、选择压力 and 不同选择方案对 GA 收敛性和收敛速度的影响。Syswerda G 通过实验研究了三种不同遗传操作对 GA 性能的影响,得到的一些结论均基于实验结果统计值,缺乏理论依据。Yee Leung, Yong Gao 定义了种群多样度的概念,建立了 GA 早熟与种群大小、变异率及其他种群特性的关系。张铃、张钊讨论了种群多样性与适应度函数调整对 GA 收敛性的影响。但总的来说,这方面研究的深度和广度都还不够,应在综合已有的研究成果基础上,对 GA 各算子、参数对 GA 性能的影响做全面分析。这包括定义新的性能评价指标和寻找有效的分析方法,以准确评价和分析各遗传算子的作用及参数设置对 GA 的影响。

1.5 遗传算法的改进研究

尽管遗传算法得到了广泛的应用,但它远非十分完善,还存在许多急待改进的缺点和不足,主要表现在以下几个方面:① 算法存在的不成熟收敛问题(Premature Convergence)。尽管与其他方法相比,遗传算法能很好地解决这一问题,但仍不能令人十分满意;② 算法存在收敛速度低的问题,特别是在处理高维数且复杂程度较高的问题时,这个问题尤为突出。这两个问题是遗传算法应用中的一对主要矛盾,如何在保证遗传算法全局收敛的同时兼顾遗传算法的搜索效率是遗传算法理论和应用研究中的一个重点和难点;算法自身参数的选取存在困难,如群体规模的大小、初始群体的选择、交叉变异概率的确定等等都无法用一个量化的关系式来描述,需要在研究和应用中进一步探索其规律性,减少在参数选择过程中的盲目

性和随机性。

许多学者从不同侧面对 GA 进行了改进研究。其中包括围绕基本遗传算法本身所做的改进及将 GA 与其他智能方法融合以提高 GA 性能的研究。本节将从这两方面对 GA 的改进研究加以介绍。对于后者,将重点介绍与本书有密切联系的自适应遗传算法(Adaptive Genetic Algorithm,简称 AGA)、模拟退火遗传算法(Simulated Genetic Algorithm,简称 SAGA)和模糊遗传算法(Fuzzy Genetic Algorithm,简称 FGA)的发展情况。

1.5.1 基本 GA 的改进研究

这方面的研究开展得较早,成果也较丰富。下面以简单 GA 的执行步骤为主线分别加以说明。

1. GA 控制参数的选择

GA 中需要选择的参数有:种群规模 M 、编码长度 l 、代沟 G 、交叉率 P_c 和变异率 P_m 。各参数的不同选取对 GA 的性能影响很大。为了选择合适的 GA 参数,许多学者进行了大量的研究。主要结果见表 1-1。

表 1-1 GA 控制参数的选择

序号	名称	研究者	主要特点
1	最优参数范围	Schaffer	GA 实际应用经验的总结; $N: 20 \sim 30$, $P_c: 0.75 \sim 0.95$, $P_m: 0.005 \sim 0.01$, $G: 0.3 \sim 1.0$
2	自主优化	Grefenstette	利用上层 GA 优化下层 GA 的参数
3	μ GA	Krishnakumar	小群体方法,克服 N 大造成的计算费时
4	自适应 N	Baker, Arabas	限制当前种群产生后代数,克服过早收敛 对个体赋予生存寿命来调整,保持种群多样性
5	自适应 P_c, P_m	Strinivas Li, Lucasius Whitely Forgarty	P_c, P_m 随父串适应度自适应变化 利用种群多样性函数调整 P_c, P_m 保持理想的 EER P_m 与父串间 Hamming 距离成反比 P_m 随遗传代数按指数下降

其中 4,5 的自适应参数设置方法是研究的热点。它考虑的是如何使 GA 在求解某一问题的不同阶段自动地调整某些参数,如种群规模、实施概率等。目前许多学者认识到这些参数需要随着遗传进程而自适应变化,这种有自组织性能的 GA 具有更高的鲁棒性、全局最优性和效率。在 1.5.3 节中将进一步介绍 AGA 的发展。

2. 编码方法

编码是 GA 应用时首先要面对的问题,它关系到 GA 能否对所处理的问题进行合理且有效的描述。不同的编码方案对遗传算法的性能会产生不同的影响,通过适当的编码方法,可以缩小编码长度、减少劣质个体,提高算法性能。常见的编码方法如表 1-2 所示。