

协同演化算法 及其在数据挖掘 中的应用

董红斌 贺志 著



中国水利水电出版社

www.waterpub.com.cn

协同演化算法及其在数据挖掘中的应用

董红斌 贺 志 著



中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

演化算法是一种模拟生物演化过程与机制求解优化问题及搜索问题的一类自组织、自适应人工智能技术。协同演化算法是针对传统演化算法的不足而兴起的,通过构造两个或多个种群,建立它们之间的竞争或合作关系,多个种群相互作用来提高各自性能,适应复杂系统的动态演化环境,以达到种群优化的目标。协同演化算法是近年来已成为计算机智能研究的一个热点,已广泛应用于电子工程、模式识别、交通运输规划、经济管理和工程设计优化等领域。

本书从协同演化理论和演化算法相结合的角度出发,着重介绍了协同演化算法及其在数据挖掘的应用方面的研究工作。全书共分为 10 章,内容包括绪论、函数优化、约束优化、多目标优化、兴趣度量优化、数据集的优化和规则形式的简化及演化聚类算法等。

本书可作为计算机科学、信息科学、人工智能自动化技术及数据挖掘等相关专业的研究生、教师和科技工作者的参考书。

图书在版编目(CIP)数据

协同演化算法及其在数据挖掘中的应用 / 董红斌, 贺志著. —北京: 中国水利水电出版社, 2008

ISBN 978-7-5084-5688-1

I. 协… II. ①董…②贺… III. 算法—应用—数据采集
IV. TP274

中国版本图书馆 CIP 数据核字 (2008) 第 094083 号

书 名	协同演化算法及其在数据挖掘中的应用
作 者	董红斌 贺 志 著
出版 发行	中国水利水电出版社 (北京市三里河路 6 号 100044) 网址: www.waterpub.com.cn E-mail: mchannel@263.net (万水) sales@waterpub.com.cn 电话: (010) 63202266 (总机)、68367658 (营销中心)、82562819 (万水)
经 售	全国各地新华书店和相关出版物销售网点
排 版	北京万水电子信息有限公司
印 刷	北京市天竺颖华印刷厂
规 格	184mm×260mm 16 开本 12 印张 262 千字
版 次	2008 年 7 月第 1 版 2008 年 7 月第 1 次印刷
印 数	0001—3000 册
定 价	26.00 元

凡购买我社图书, 如有缺页、倒页、脱页的, 本社营销中心负责调换
版权所有·侵权必究

前 言

半个世纪以来,人们从不同的角度对生物系统及其行为特征进行了模拟,产生了一些对现代科技发展有重大影响的新兴学科。从 20 世纪 40 年代起,生物模拟就构成了计算科学的一个组成部分。演化算法是一种模拟生物演化过程与机制求解优化问题与搜索问题的一类自组织、自适应人工智能技术。这类技术的核心思想源于这样的基本知识:生物进化过程本身就是一个自然的、并行发生的、稳健的优化过程。这一过程的目标在于使生命体达到适应环境的最佳结构与效果,生物种群通过“优胜劣汰”及遗传变异来达到演化的目的。按照达尔文的自然选择学说和孟德尔的遗传变异理论,生物演化是通过繁殖、变异、重组和选择这四种基本形式来实现的。演化算法就是基于这种思想发展起来的一类随机搜索技术。它们是模拟由个体组成的群体的学习过程,其中每个个体表示给定问题搜索空间的一点。演化算法从一个初始的群体出发,通过随机选择、变异和重组过程,使群体演化到搜索空间中越来越好的区域。选择过程是群体中适应性好的个体比适应性差的个体有更多的复制机会,交叉算子将父辈信息结合在一起并将它们遗传到子代个体,变异保证在群体中产生新的个体。演化算法这种以生物智能或自然现象为基础的随机搜索算法具有比数学规划方法更大的优越性,使得演化算法已成为人工智能领域的研究热点。

协同演化算法是近年来计算智能研究的一个热点,它是针对演化算法的不足而兴起的,通过构造两个或多个种群,建立它们之间的竞争或合作关系,多个种群通过相互作用来提高各自性能,适应复杂系统的动态演化环境,以达到种群优化的目的。协同演化算法与一般的演化算法的根本差别在于它的演化过程,在协同演化中,一个个体的适应度的计算是在与其他个体的交互过程中进行的,依赖于不同的问题,交互伙伴可以是同一种群的个体或不同种群的个体。目前,协同演化算法的研究已受到国内外学者的广泛关注,国际上已经开展了一系列协同演化算法的研究工作。

数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的数据集中识别有效的、新颖的、潜在有用的,以及最终可理解模式的非平凡过程。在从数据中发现知识的过程中,由于数据集的规模较大、知识的表现形式各异,传统的搜索算法总是存在着效率或者性能上的不足。这为遗传算法在数据挖掘中的应用提供了广阔的空间。

本书在深入研究协同演化算法的基础上,从协同演化理论和演化算法相结合的角度出发,结合函数优化、约束优化、多目标优化、聚类及关联规则等问题进行了较为深入的研究。从提出的算法特点上来看,第 3、4、5 章是数值优化方法,均

属于连续函数的优化。其中，第 3 章基于博弈论提出了一种求解函数优化问题的混合策略协同演化算法，第 4、5 章则是在第 3 章的基础上针对不同类型的数值优化问题，提出了两种改进的混合策略演化算法。第 6、7、8 章主要解决关联规则质量太低的问题，分别从优化兴趣度量、优化数据集和简化规则形式这三个方面解决这个问题。第 9 章是结合聚类问题展开工作，将演化算法应用于聚类中，给出了一种混合策略演化聚类算法。本书的特点是算法理论与应用并重，期望能为相关领域的研究者提供可供借鉴的研究方法和思路。

本书的内容是作者在北京交通大学计算机与信息学院攻读博士学位期间的研究工作的总结。特别感谢北京交通大学黄厚宽教授和田盛丰教授多年来的悉心培养和指导。感谢清华大学石纯一教授、北京交通大学何军博士、于剑教授、哈尔滨工程大学印桂生教授、侯薇、丁泽进和周成的支持和帮助。

本书的出版得到了国家自然科学基金 (60443003)、黑龙江省自然科学基金 (F200605) 和黑龙江省教育厅海外学人科研资助项目 (1153h21) 的支持，在此表示特别的谢意。

由于作者水平有限，书中不妥之处在所难免，欢迎广大读者对本书提出宝贵的批评、意见与建议。

作者电子邮箱是：donghongbinbjtu@gmail.com, hezhidragon@126.com。

作者

2008 年 3 月

目 录

第 1 章 绪论	1
1.1 演化算法概述	1
1.1.1 演化算法的起源	1
1.1.2 演化算法的应用领域	4
1.1.3 演化算法的发展趋势	5
1.2 演化算法的基本原理	5
1.2.1 生物学基础	5
1.2.2 演化算法与最优化	6
1.2.3 演化算法的一般框架	7
1.2.4 没有免费午餐定理	8
1.3 协同演化算法研究现状	8
1.3.1 协同进化论	8
1.3.2 演化博弈论	10
1.3.3 协同演化算法设计	14
1.3.4 协同演化算法的应用	17
第 2 章 数据挖掘和关联规则	19
2.1 数据挖掘	19
2.1.1 数据挖掘的定义和过程	19
2.1.2 挖掘的数据形式	20
2.1.3 数据挖掘的功能	21
2.2 关联分析	22
2.2.1 关联规则的定义	22
2.2.2 关联规则挖掘的问题	23
2.3 聚类分析	24
2.3.1 聚类分析的意义	24
2.3.2 聚类分析的概念	25
2.3.3 聚类分析的算法	26
2.4 遗传算法在规则发现中的应用	27
2.4.1 个体表示	28
2.4.2 算子	29
2.4.3 种群初始化	31

2.4.4	适应度	31
2.4.5	分类规则发现和关联规则发现的不同	32
第 3 章	基于混合策略的协同演化算法	34
3.1	引言	34
3.2	混合策略协同演化规划的基本思想	35
3.2.1	概率分布函数的性能	35
3.2.2	混合策略协同演化规划框架	37
3.3	求解函数优化问题的 MSCEP 算法	38
3.4	实验结果和分析	41
3.4.1	实验结果	41
3.4.2	实验分析	49
3.5	小结	49
第 4 章	求解约束优化问题的混合策略演化算法	55
4.1	引言	55
4.2	求解约束优化问题的基本思想	56
4.2.1	约束优化问题	56
4.2.2	约束比较规则	57
4.3	求解约束优化问题的 CMSEP 算法	57
4.4	实验结果和分析	59
4.4.1	实验结果	59
4.4.2	结果分析	63
4.5	小结	64
第 5 章	求解多目标优化问题的混合策略演化算法	69
5.1	引言	69
5.2	相关工作	69
5.2.1	关键技术	69
5.2.2	研究现状	71
5.3	求解多目标优化问题的基本思想	72
5.3.1	多目标优化问题	72
5.3.2	强度 Pareto 优化的基本思想	72
5.4	混合策略的 Pareto 演化规划	74
5.5	实验结果和分析	74

5.5.1	测试函数	74
5.5.2	结果分析	76
5.6	小结	76
第 6 章	兴趣度量的优化	78
6.1	关联规则的兴趣度量	78
6.1.1	兴趣度量的目的	78
6.1.2	兴趣度量的分类	79
6.2	发现度量约束下的规则	85
6.3	发现基于残差的最优相关规则	86
6.3.1	残差分析	87
6.3.2	互信息量	89
6.3.3	用遗传算法发现优化相关规则 GADCR	90
6.4	实验	93
6.4.1	人工数据集上的实验结果	93
6.4.2	真实数据集上的实验结果	96
6.5	小结	97
第 7 章	数据集的优化	98
7.1	离散化简介	98
7.1.1	有监督和无监督离散化	100
7.1.2	一元与多元离散化	103
7.1.3	MVD	104
7.2	一种基于聚类的无监督多元离散化方法 EMVD-BDC	105
7.2.1	动机	105
7.2.2	EMVD-BDC	107
7.3	MVD 的一种优化算法 OMVD	111
7.3.1	最大支持度差异 (MSD)	112
7.3.2	用遗传算法优化 MSD 列表	112
7.4	实验	113
7.4.1	人工数据集上的实验	113
7.4.2	真实数据集上的实验	115
7.5	小结	122

第 8 章 规则形式的简化	126
8.1 最优关联规则的简介	126
8.1.1 问题的提出和定义	126
8.1.2 发现最优置信度规则的经典算法: FOCR	126
8.1.3 其他算法	127
8.2 发现最优 interest 规则	128
8.2.1 问题的定义	128
8.2.2 发现最优 interest 规则的算法 FOIR	129
8.3 发现最优相对密度规则	134
8.3.1 密度度量	134
8.3.2 相对密度度量	135
8.3.3 最优相对密度规则的发现算法 FORDAR	136
8.4 实验	139
8.4.1 人工数据集上的实验	139
8.4.2 真实数据集上的实验	142
8.5 小结	147
第 9 章 基于混合策略的演化聚类算法	149
9.1 引言	149
9.2 模糊 C 均值聚类	150
9.2.1 模糊 C 均值聚类算法	150
9.2.2 模糊权和有效性函数	150
9.3 混合策略演化聚类算法	152
9.4 实验结果和分析	153
9.4.1 人工数据集和真实数据集	153
9.4.2 有效性指标实验分析	155
9.4.3 算法性能分析	156
9.4.4 参数选择	161
9.5 小结	161
第 10 章 总结	163
参考文献	166

第 1 章 绪论

1.1 演化算法概述

半个世纪以来,人们从不同的角度对生物系统及其行为特征进行了模拟,产生了一些对现代科技发展有重大影响的新兴学科。演化算法是一种模拟生物演化过程与机制求解优化问题及搜索问题的一类自组织、自适应人工智能技术。这类技术的核心思想源于这样的基本知识:生物进化过程本身就是一个自然的、并行发生的、稳健的优化过程。这一过程的目标在于使生命体达到适应环境的最佳结构与效果,生物种群通过“优胜劣汰”及遗传变异来达到演化的目的。按照达尔文的自然选择学说和孟德尔的遗传变异理论,生物演化是通过繁殖、变异、重组和选择这 4 种基本形式来实现的。演化算法就是基于这种思想发展起来的一类随机搜索技术。它们是模拟由个体组成的群体的学习过程,其中每个个体表示给定问题搜索空间的一点。演化算法从一个初始的群体出发,通过随机选择、变异和重组过程,使群体演化到搜索空间中越来越好的区域。选择过程是群体中适应性好的个体比适应性差的个体有更多的复制机会,交叉算子将父辈信息结合在一起并将它们遗传到子代个体,变异保证在群体中产生新的个体。演化算法这种以生物智能或自然现象为基础的随机搜索算法具有比数学规划方法更大的优越性,使得演化算法已经成为人工智能领域的研究热点 [1-8]。

1.1.1 演化算法的起源

生物演化是鲁棒性最强、也是最有效的一种来自自然的问题求解方法。演化算法 (Evolutionary Algorithm, EA) 则是受生物演化启发而提出的模拟进化机制。由于对生物演化本质的不同认识,形成了模拟生物演化的不同学派。演化算法最初有三大分支,即遗传算法、演化策略和演化规划。20 世纪 90 年代初,在遗传算法的基础上又发展了一个新的分支 - 遗传程序设计。虽然它们在算法实现方面具有一些差别,但它们具有一个共同的特点,即都是借助生物演化的思想和原理来解决问题 [4-8]。

1. 遗传算法

遗传算法 (Genetic Algorithm, GA) 是由达尔文进化理论引发的一种计算模型,是模拟生物体的适应性和演化能力,用于设计自动演化的计算机程序。早在 20 世纪 40 年代,就有学者开始研究如何利用计算机进行生物模拟技术。进入 20 世纪 60 年代后,美国 Michigan 大学的 Holland 教授及其学生们受到这种生物模拟技术的启发,创造一种基于生物遗传和演化(进化)机制的适合于复杂系统优化计算的自适应随机优化技术 - 遗传算法 [6]。在遗传算法中,问题的最优解的搜索是通过对位串结构的染色体个体进行遗传操作而得到的。每个染色体是问题的候选解。通过复制、交叉和变异等遗传操作来持续地产生新个体,从而使个体评价函数的质量得以提高 [6]。70 年代初, Holland 教授提出了遗传算法的基本定理 - 模式定

理,从而奠定了遗传算法的理论基础,模式定理揭示出群体中的优良个体的样本数呈指数级增长的规律。1975年,他出版了第一本系统论述遗传算法和人工自适应系统的专著—《自然系统和人工系统的自适应》。20世纪80年代,Holland教授实现了第一个基于遗传算法的机器学习系统—分类器系统,开创了遗传算法机器学习的新概念[10]。

虽然已有多种不同的算法形式,但对人们面临的各种现实问题而言,目前这些算法形式是远远不够的。因此,结合具体的实际问题,研究采用合理有效的策略或模拟生物演化的新理论来设计或生成新颖有效的算法,是遗传算法研究的重要方向之一。

2. 演化策略

20世纪60年代初,德国柏林工业大学的学生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的方法进行优化,因而他们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,他们进一步对这种方法进行了深入的研究与发展,形成了演化计算的另一个分支—演化策略(Evolution Strategy,ES)[5,11]。

演化策略不同于遗传算法,它直接在解空间上进行操作,强调的是演化搜索方向和步长的自适应调节。早期演化策略的种群只包含一个个体,而且只使用变异操作。在每一演化代变异后的个体与其父代个体进行比较再选择两者之优。它使用的变异算子是基于正态分布的变异操作。这一演化策略称为(1+1)演化策略或二元演化策略。(1+1)演化策略有很多弊端,如有时收敛不到全局最优解、效率较低等。它的改进是增加种群内个体的数量,从而演化为($\mu+1$)演化策略。此时种群内含有 μ 个个体,随机地选取一个个体进行变异,然后取代种群中最差的个体。($\mu+1$)演化策略与(1+1)演化策略的相似之处是每次只产生一个后代。为了进一步提高效率,后来又发展为($\mu+\lambda$)演化策略和(μ,λ)演化策略,并且引进了重组操作,即由两个个体按类似杂交的方式生成一个个体。($\mu+\lambda$)演化策略采用变异和重组操作,根据种群内的 μ 个个体产生 λ 个个体,然后将这($\mu+\lambda$)个个体进行比较,从中选取 μ 个最优者,将他们保留到子代群体中。(μ,λ)演化策略则是在新产生的 λ ($\lambda > \mu$)个个体中选取 μ 个最优者。

演化策略主要适合求解数值优化问题,近年来,由于遗传算法也采用浮点数编码技术来求解数值优化问题,演化策略与遗传算法之间的相互渗透已使两者之间没有了明显的界限。

3. 演化规划

演化规划(Evolutionary Programming,EP)的方法最初是由美国科学家L.J.Fogel等在20世纪60年代提出的。他们在人工智能的研究中发现,智能行为就是具有预测其所处环境的状态、并按给定目标作出适当响应的能力。在研究中,他们将模拟环境描述成有限字符集中的符号组成的序列,于是问题转化为怎样根据当前观察到的符号序列作出响应,以获得最大收益。这里收益的计算是按环境中将要出现的下一个符号及预先定义好的收益目标来确定。演化规划中常用有限状态自动机(Finite State Machine,FSM)来表示这样的策略,这样,问题就变成如何设计一个有效的FSM。L.J.Fogel等借用演化的思想对一群FSM进行演化,以获得较好的FSM。他们将这种方法应用到数据诊断、模式识别、分类和控制系统的设计等

问题中,取得了较好的结果 [4,8]。20 世纪 90 年代, D.B.Fogel 借助演化策略方法对演化规划进行了发展,并在数值优化及神经网络的训练等问题中获得成功。这样,演化规划就演变成为一种优化搜索算法,并在很多实际领域得到应用 [12,13]。

4. 遗传程序设计

计算机出现以来,计算机科学的一个重要目标就是让计算机自动进行程序设计,即只要明确告诉计算机要解决的问题而不需要告诉它如何去做,遗传程序设计 (Genetic Programming, GP) 就是在该领域的一种尝试。1992 年, Koza 将遗传算法应用于计算机程序的优化设计及自动生成,提出了遗传程序设计的概念,并成功地将遗传程序设计方法应用于人工智能、机器学习、符号处理等方面 [7]。遗传程序设计采用遗传算法的基本思想,使用一种更为灵活的表示方式—分层结构来表示解空间。这些分层结构的叶节点是问题的原始变量,中间节点则是组合这些原始变量的函数。遗传程序设计即是使用一些遗传操作动态地改变这些结构以获得解决问题的可行的计算机程序。由于遗传程序设计采用一种更自然的方式,使得它的应用领域非常广泛。

虽然演化算法的几个分支有差异,但其基本原理都是采用生物演化的思想。近年来,它们的差异越来越小,并互相渗透,可以纳入到演化算法的统一框架。

5. 演化算法的特性

为了求解各种优化问题,人们已经发展了各种各样的优化算法,这些算法都有自己的长处,都有自己的使用范围,也都有自身的弱点。同其它算法相比演化算法主要特性如下 [1,3,17,18]:

(1) 自适应性。演化算法的智能性包含自组织、自适应、自学习。应用演化算法求解问题时,在确定了编码方案、适应值函数和遗传算子后,算法将利用演化过程中获得的信息自适应搜索。

(2) 并行性。演化算法的性质决定了其本质的高度并行性,表现在两个方面:一是演化计算形式上简单明了,算法本身非常适合大规模并行计算,最简单的并行方式是让各处理机各自独立地进行种群的演化,在演化过程中各处理机之间可以进行协调合作,最后从所有处理机的所有种群中选择一个或多个最佳解;二是演化计算的隐式并行性,由于演化计算采用种群的方式组织搜索,从而它可以同时搜索解空间内的多个不同区域。

(3) 全局优化。演化算法的搜索过程是从一群初始点开始搜索,这种机制意味着搜索过程可以有效地跳出局部极值点。当采用有效的群体多样性时,算法可以将局部搜索和全局搜索协调起来,使得找到问题最优解的概率大大提高。

(4) 随机性。演化算法中的选择、交叉和变异操作都带有一定的随机性,而不是确定的精确规则。它采用随机方法进行最优解搜索,选择体现了向最优解逼近的原则,交叉体现了最优解的产生原则,变异体现了全局最优解的复盖原则。采用概率的变迁规则来指导搜索方向。

(5) 普适性。由于演化算法使用基于目标函数的评价信息,并不需要问题导数等与问题直接相关的信息,只需适应值和串编码等通用信息,故几乎可处理任何问题。这一特点使其成为

具有良好通用性和可规模化的优化算法。

(6) 鲁棒性。演化算法具有很强的鲁棒性,即在噪声存在的情况下,对同一问题的 EA 多次求解中得到的结果是相近的。

1.1.2 演化算法的应用领域

随着国民经济、科学技术和社会的不断发展,人们在社会生活中遇到的问题日趋复杂,迫切要求解决问题的理想方法。演化计算提供了一种求解复杂系统优化问题的通用框架,对问题的种类有很强的鲁棒性,已广泛应用于众多学科。应用范围包括生产调度问题、生物工程、机器学习、网络设计、交通运输、自动控制和工程设计等众多领域。主要应用领域可以概括如下 [1, 3, 17, 18]:

(1) 函数优化。函数优化是演化算法的经典应用领域,也是对演化计算进行性能评价的常用算例。人们构造了许许多多的各种形式的测试函数,有连续函数也有离散函数,有凸函数也有凹函数,有低维函数也有高维函数,有确定性函数也有随机性函数,有单峰函数也有多峰函数等。对于高度非线性、多目标的函数优化问题,用演化算法可以很方便地得到它们的解,而传统的方法往往难以奏效。

(2) 组合优化。随着问题规模的不断扩大,组合优化问题的搜索空间急剧扩大,要想用枚举法或精确法求出其精确最优解是不可能的。对于这类复杂问题,更应把精力放在寻求其满意解上,而演化计算则是寻求这种满意解的最佳工具之一。实践证明,演化算法对于组合优化中的 NP 完全问题非常有效。例如,演化算法已经在求解旅行商问题、装箱问题和 n 皇后问题等方面得到了成功的应用。

(3) 机器学习。学习能力是高级自适应系统所应具备的能力之一。基于演化计算的机器学习,特别是分类器系统,在许多领域得到了应用。例如,演化计算被用于模糊控制规则的学习,利用演化算法学习隶属度函数,从而更好地改进了模糊系统的性能。基于演化计算的机器学习可用于调整人工神经网络的连接权,也可用于神经网络结构的集成和模式分类。

(4) 图像处理和模式识别。图像处理和模式识别是计算机视觉中的一个重要研究领域。在图像处理过程中,如扫描、特征提取、图像分割等不可避免地会产生一些误差,这些误差会影响到图像处理和识别的效果。如何使这些误差最小是使计算机视觉达到实用化的重要要求。演化算法在图像处理中的参数优化方面已经取得了良好的应用结果。

(5) 生产调度问题。生产调度问题在很多情况下所建立起来的数学模型难于精确求解,而目前在现实生产中也主要是靠一些经验来进行调度。现在演化算法已成为解决调度问题的有效工具。在单件生产车间调度、流水线生产车间调度、生产规划、任务分配等方面演化算法都得到了有效的应用。

(6) 自动控制领域。在自动控制领域中有很多与优化相关的问题需要求解,演化算法已在其中得到了初步的应用,并显示出了良好的效果。例如, Karr 等人用 EA 设计自适应模糊逻辑控制器,取得了显著的效果。另外, EA 在故障诊断和机器人规划中的应用也取得了成功。

(7) 人工智能应用。EA 在人工智能的应用领域包括数据挖掘与知识获取、模式识别、人工神经网络集成和多 Agent 系统等。

(8) 社会与经济领域。EA 在早期就应用于囚徒两难问题的分析,近年来将 EA 应用于交通运输、电力系统、金融和社会经济机制设计已经成为 EA 的应用热点。

1.1.3 演化算法的发展趋势

演化算法是一种以模拟生物进化过程与机制求解问题的自适应智能优化算法,这种智能优化算法在工程领域的成功应用,激励人们从更广泛的生物或自然现象寻求启发以构造新的算法,来解决工程中广泛存在的复杂问题。

有关演化算法的收敛性研究是过去十几年间演化算法基础研究的核心,已有一系列重要工作。与收敛性分析紧密相关的另一基本理论问题是演化算法能以多快的速度收敛。这些问题包括有关收敛速度估计、迭代次数估计和时间复杂性估计。概括的说,目前已有的关于演化算法收敛速度与复杂性的结果基本限于二进制,而且或是对特殊的适应度函数,或是对使用某一个或两个演化算子操作情形的算法,仍不具有普遍性。关于演化算法有效性和能解性研究,目前尚无理论结果。

演化算法的设计是演化算法研究的另一个核心内容,从产生到此后的长时期内,演化算法的设计一直是人们所关注的主要方面。该方面的研究主要围绕两个永恒的主题:一是扩大可应用领域;二是使之更加有效。前者旨在设计和发现有效的算法执行策略以解决过去不能解决或不能有效解决的问题,而后者在于修正和改进已有算法,使之更加有效或具有所期望的某种特别性态。从方法论上说,算法研究主要从更宏观、更本质的角度模拟自然进化原理与机制,模拟生物智能的生成过程,并用以求解问题,进而融合数学、物理、化学、生物学、计算机技术等其他学科的原理与方法,使所设计的算法执行策略更加有效。演化算法进一步研究工作有如下几个方面: ① 对已有算法收敛性分析模型的深化研究; ② 对非标准演化算法发展收敛性理论; ③ 对演化算法过早收敛现象的理论分析; ④ 关于演化算法有用性与有效性研究; ⑤ 关于算法的效率加速理论和策略的研究; ⑥ 开发基于仿生与数理分析驱动相结合的新型执行策略; ⑦ 各种演化算法执行策略与其他相关算法比较研究; ⑧ 深化专门领域的算法执行策略研究 [1-3]。

1.2 演化算法的基本原理

1.2.1 生物学基础

生物的发展是一个漫长的演化过程。生物在繁衍生存过程中,逐渐适应于其生存环境,使得其品质不断得到改良,这种生命现象称为进化(或称演化(Evolution))。生物的演化是以群体的形式共同进行的,这样的群体称为种群(Population),组成种群的单个生物体称为个

体 (Individual)。每一个生物个体对其自身的生存环境都有不同的适应能力, 这种能力称为个体的适应度 (Fitness) [2]。19 世纪 50 年代, 英国生物学家达尔文通过对世界各地生物的考察和人工选择的实验提出了生物进化论。达尔文进化论的中心思想是自然选择学说, 生物演化主要因为有遗传、变异和选择三个原因。遗传是生物所共有的特性, 正是这种遗传性, 使得生物能够把它的特性、性状传给后代, 在后代中保持相似。变异是一切生物所具有的共同特征, 是生物个体之间相互区别的基础。生物的变异性为生物的进化和发展创造了条件。选择决定了生物演化的方向。选择分为人工选择和自然选择。人工选择是在人为环境下, 把对人有利的生物特性保留下来, 淘汰不利的生物特性。自然选择是指生物在自然界的生存环境中适者生存、劣者淘汰的过程。生物就是在遗传、变异和选择三者综合作用下, 不断地向前发展和演化。生物进化论揭示了生物长期自然选择的进化规律, 自然选择过程蕴涵着一种搜索和优化的先进思想。科学家们从中得到启迪, 将这种思想用于工程和科学领域, 发展了多种进化学习算法, 为许多传统优化方法难以解决的优化问题提供了新的途径。

1.2.2 演化算法与最优化

所谓最优化问题, 就是在满足一定的约束条件下, 寻找一组参数值, 以使某些最优性度量得到满足, 即使系统的某些性能指标达到最大或最小 [18]。最优化问题根据其目标函数、约束函数的性质以及优化变量的取值等可以分为许多类型, 每一种类型的最优化问题根据其性质的不同都有其特定的求解方法。不失一般性, 设所考虑的最优化问题为

$$\begin{cases} \min f(X) \\ \text{s.t. } X \in S = \{X \mid g_i(X) \leq 0, \quad j = 1, \dots, m\} \end{cases}, \quad (1.1)$$

其中, $f(X)$ 为目标函数, $g_i(X)$ 为约束函数, S 为约束域, X 为 m 维优化变量。通常最大化问题很容易转化为最小化问题, 对于 $g_i(X) \geq 0$ 的约束和等式约束也可转化为 $-g_i(X) \leq 0$ 的约束, 所以式 (1.1) 所描述的最优化问题不失一般性。

当 $g_i(X) \leq 0 (j = 1, \dots, m)$ 所限制的约束空间为整个 n 维欧氏空间, 即 R^n 时, 上述最优化问题为无约束优化问题, 即

$$\begin{cases} \min f(X) \\ \text{s.t. } X \in S \subset R^n \end{cases}, \quad (1.2)$$

最优化问题的应用涉及工业、社会、经济、管理等各个领域, 很多实际问题进行数学建模后可抽象为一个数值函数的优化问题。对于这样一类优化问题, 通常按照搜索机制的不同可以将优化算法分为三类 [1, 3, 17, 18]:

(1) 经典算法。如牛顿法、单纯形法、共轭方向法、最速下降法和罚函数法等, 这类传统的算法在实际中得到了充分的应用, 并且仍然在应用数学、计算机科学及工程应用中发挥着重

```

begin
    确定编码的形式和搜索空间, 选择遗传算子和参数值;
    设置代数  $t = 0$ ;
    随机生成初始群体;
     $P(0) = \{(x_{0,1}, a_{0,1}), (x_{0,2}, a_{0,2}), \dots, (x_{0,u}, a_{0,u})\}$ ;
    计算每个个体的适应度值  $\Phi(X_{0,i})$ ;
    while (终止条件不满足) do
         $t = t + 1$ ;
        对  $P(t-1)$  进行重组操作生成群体  $P'(t)$ ;
        对  $P'(t)$  进行变异操作生成群体  $P''(t)$ , 计算其中每个个体的
        适应度值  $\Phi(x_{t,i}), i = 1, 2, \dots, \lambda$ ;
        对  $(Q \cup P''(t))$  进行选择操作生成群体
         $P(t) = \{(x_{t,1}, a_{t,1}), (x_{t,2}, a_{t,2}), \dots, (x_{t,u}, a_{t,u})\}$ ,
        其中  $Q$  代表  $P(t-1)$  的某个子集或空集;
    end
end

```

图 1.1 演化算法的一般框架

要的作用。但这些传统的优化方法存在一些缺点, 它们往往只能求出优化问题的局部最优解, 因为这些算法总是求得最近的一个极值点, 只有能够给出很好的初始点才有可能得出所需要的全局最优解。

(2) 智能算法。这类方法的思想主要来自于自然界, 如神经网络、模拟退火、禁忌搜索和演化算法等。这类算法一般都是非确定性算法, 它们对目标函数要求低、容易实现、稳定性好, 适合于大规模、非线性、多峰多态函数、含离散变量的全局优化问题, 在求解速度和质量上远远超过经典算法, 因而是一类高效近似算法。

(3) 混合算法。将智能算法与经典算法相结合来构造一些高效的新的混合算法也是目前解决优化问题的一个方向, 如将演化算法与传统的、基于问题知识的启发式搜索技术相结合构成基于演化算法的混合搜索算法, 综合了演化算法的全局搜索性和启发式方法的局部优化快速收敛性的优点, 这种混合策略可以从根本上提高演化算法的计算性能。

1.2.3 演化算法的一般框架

下面给出求解优化问题的一般框架 [3]。设适应度函数为 $\Phi: S \rightarrow R$, A 表示包含附加的搜索状态信息 (A 通常是需要自适应调整的策略参数, A 可以为空) 的适当的表示空间。在演化算法中, 搜索群体中的任一个体 (x, a) 均是笛卡儿集 $S \times A$ 中的一个元素, 其

中, $x \in S, a \in A$ 。个体 (x, a) 的适应值由优化变量 x 的适应值 $\Phi(x)$ 给出。 $\mu (\mu \geq 1)$ 表示父代群体的大小, $\lambda (\lambda \geq 1)$ 表示子代群体的大小。将演化过程中第 t 代的群体记为 $P(t) = \{(x_{t,1}, a_{t,1}), (x_{t,2}, a_{t,2}), \dots, (x_{t,u}, a_{t,u})\}$, 则用演化算法求解问题 (1.1) 的过程如图 1.1 所示。

1.2.4 没有免费午餐定理

在最优化理论研究中, 最值得一提的是 Wolpert 和 Macready 于 1997 年发表的题为 No Free Lunch Theorems for Optimization 的论文, 提出并论证了所谓的没有免费午餐定理 (No Free Lunch Theorems), 简称 NFL 定理 [21]。

NFL 定理简单表述为: 对于所有可能的问题, 任意给定两个算法 A 、 A' , 如果 A 在某些问题上表现比 A' 好 (差), 那么 A 在其他问题上的表现就一定比 A' 差 (好), 也就是说, 任意两个算法 A 、 A' 对所有问题的平均表现度量是完全一样的。

自从 NFL 定理提出以来, 有关定理本身及其相关结论的争论在学术界一直持续不断, 因为 NFL 定理本身涉及了优化算法最基本的问题, 而且其结论多少有点出人意料。NFL 定理的主要价值在于它对研究与应用优化算法时的启示作用 [3, 21]。虽然 NFL 定理是在许多假设条件下得出的, 但它仍然在很大程度上反映出了优化算法的本质。当面对一个大的而且形式多样的适应值函数类时, 就必须考虑算法间所表现出的 NFL 效应, 即若算法 A 在某些函数上的表现超出算法 A' , 则在这类的其他适应值函数上, 算法 A' 表现得比 A 要好。因此, 对于整个函数类, 不存在万能的最佳算法, 所有算法在整个函数类上的平均表现度量是一样的。

1.3 协同演化算法研究现状

演化作为从生命现象中抽取的重要的自适应机制, 已为人们所普遍认识和广泛应用, 然而现有的演化模型存在一个共同的不足是未能很好的反映出这样一个普遍存在的事实: 多数情况下, 整个系统复杂的自适应演化过程事实上是一个系统与局部相互作用的协同演化过程, 也就是说它是大规模协同动力学系统。如何反映演化的多样性、多层次性、自适应性和自组织过程, 则是有待解决的问题, 这也正是解决演化计算机理的关键所在。演化算法以生物进化过程作为自适应全局优化搜索过程的模拟对象, 进化论的中心思想是自然选择学说, 生物演化主要因为有遗传、变异和选择三个原因, 而演化博弈论的生物学模型同样涉及生物进化过程尤其是生物种群的协同演化过程, 因此, 将演化算法、生物进化论和博弈论结合起来进行研究就成为演化算法研究的一个可行途径。

1.3.1 协同进化论

近 40 年来, 生物学上兴起了另一个重要理论—协同演化论, 已成为生态学的基础理论。