



● 普通高等院校信息与计算科学专业系列丛书

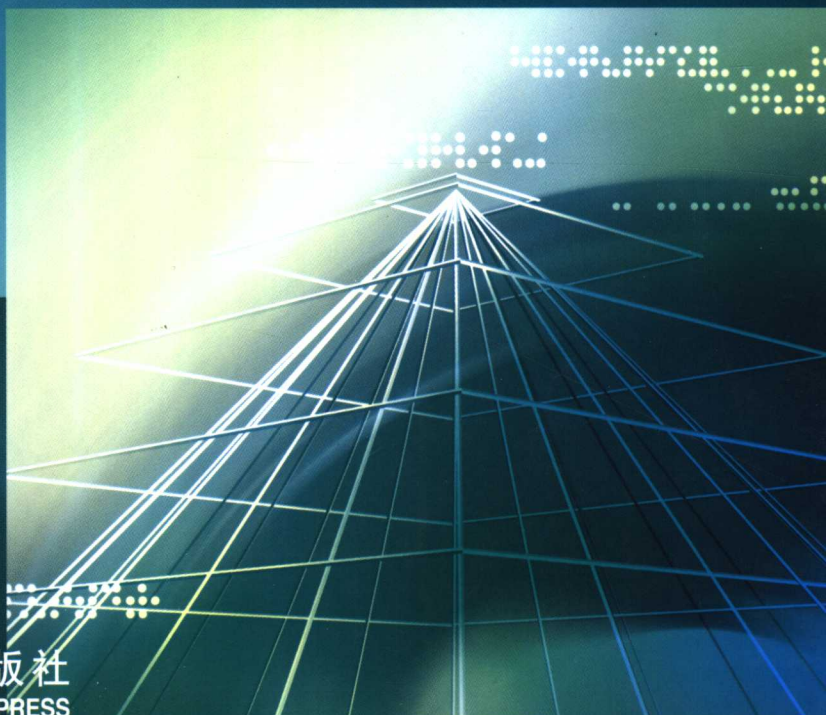


教育科学“十五”国家规划课题研究成果

计算智能 (第一册)

——模拟进化计算

徐宗本 编著



高等教育出版社
HIGHER EDUCATION PRESS

83
6
4
5

普通高等院校信息与计算科学专业系列丛书

教育科学“十五”国家规划课题研究成果

计算智能

(第一册)

——模拟进化计算

徐宗本 编著

高等教育出版社

内容提要

计算智能是以模型(计算模型、数学模型)为基础、以分布并行计算为特征的模拟人的智能求解问题的理论与方法。本书系统地讲述了计算智能的基本理论与基本方法。全书分三部分:第一部分从模拟智能生成过程的观点讲述模拟进化计算理论;第二部分从模拟智能结构的观点讲述神经网络理论;第三部分从模拟智能行为的观点讲述模糊逻辑与模糊推理。全书突出基础(特别是数学基础),强调背景(特别是生物与工程背景),着眼研究与发展。

本书可供信息与计算科学专业、数学与应用数学专业的本科生作为专业课或选修课教材使用,也可作为运筹与控制、信息科学、计算机科学、系统科学等专业的研究生教材,亦可供从事计算智能研究与应用的教师与研究人员参考。

图书在版编目(CIP)数据

计算智能(第一册)——模拟进化计算/徐宗本编著.

—北京:高等教育出版社,2004.2

ISBN 7-04-013839-5

I. 计... II. 徐... III. 人工智能-神经网络-计算 IV. TP183

中国版本图书馆CIP数据核字(2003)第116464号

出版发行	高等教育出版社	购书热线	010-64054588
社 址	北京市西城区德外大街4号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-82028899		http://www.hep.com.cn
经 销	新华书店北京发行所		
印 刷	天津新华印刷一厂		
开 本	787×960 1/16	版 次	2004年2月第1版
印 张	9.75	印 次	2004年2月第1次印刷
字 数	170 000	定 价	12.80元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

信息与计算科学专业系列教材编委会

顾问 李大潜 刘应明
主任 徐宗本
副主任 王国俊 马富明 胡德焜
委员 (以姓氏笔画为序)
 韦志辉 叶中行 白峰杉 羊丹平 孙文瑜
 吕 涛 阮晓青 陈发来 沈世鎰 陈 刚
 张志让 吴 微 柳重堪 凌永祥 徐 刚
 徐树方 黄象鼎 雍炯敏
秘书 李水根 王 瑜

总序

根据教育部 1998 年颁布的普通高等院校专业目录,“信息与计算科学”专业被列为数学类下的一个新专业(它覆盖原有的计算数学及其应用软件、信息科学与运筹控制等专业)。这一新专业的设置很好地适应了新世纪以信息技术为核心的全球经济发展格局下的数学人才培养与专业发展的需要。然而,作为一个新专业,对其专业内涵、专业规范、教学内容与课程体系等有一个自然的认识与探索过程。教育部数学与统计学教学指导委员会数学类专业教学指导分委员会(下称教指委)经过过去两年艰苦细致的工作,对这些问题现在已有了比较明确的指导意见,发表了《关于信息与计算科学专业办学现状与专业建设相关问题的调查报告》及《信息与计算科学专业教学规范》(讨论稿)(见《大学数学》第 19 卷 1 期(2003))。为此,全国高等学校教学研究中心在承担全国教育科学“十五”国家级规划课题——“21 世纪中国高等教育人才培养体系的创新与实践”研究工作的基础上,根据教指委所颁布的新的教学规范,组织国内各高校的专家教授,进行其子项目课题“21 世纪中国高等学校信息与计算科学专业教学内容与课程体系的创新与实践”的研究与探索。为推动本专业的教材建设,该项目课题小组与高等教育出版社联合成立了“信息与计算科学专业系列教材编委会”,邀请有多年教学和科研经验的教师编写系列教材,由高等教育出版社独家出版,并冠以教育科学“十五”国家规划课题研究成果。

按照新的《信息与计算科学专业教学规范》(讨论稿),信息与计算科学专业是以信息技术和计算技术的数学基础为研究对象的理科类专业。其目标是培养学生具有良好的数学基础和数学思维能力,掌握信息与计算科学基础理论、方法与技能,受到科学研究的训练,能解决信息技术和科学与工程计算中的实际问题的高级专门人才。毕业生能在科技、教育、信息产业、经济与金融等部门从事研究、教学、应用开发和管理工作的,能继续攻读研究生学位。根据这一专业目标定位和落实“强基础、宽口径、重实际、有侧重、创特色”的办学指导思想,我们认为,本专业在数学基础、计算机基础、专业基础方面应该得到加强,而各学校在这三个基础方面可大体一致,但专业课(含选修课)允许各校自主选择、体现各自特点。考虑到已有大量比较成熟的数学基础与计算机基础课程教材,本次教材编写主要侧重于专业基础课与专业课(含选修课)方面。

信息与计算科学,就其范畴与研究内容而言,是数学、计算机科学和信息工程等学科的交叉,已远远超出数学学科的范畴。但作为数学学科下的一个理科专业,信息与计算科学专业则主要研究信息技术的核心基础与运用现代计算工具高效求解科学与工程问题的数学理论与方法(或更简明地说,研究定向于信息技术与计算技术的数学基础),这一专业定位明显地与计算机科学、信息工程专业有区别。基于这一定位,信息与计算科学专业可包括信息科学与科学计算(计算数学)两个大的方向。科学计算方向在我国已有长期的办学经验,通常被划分为偏微分方程数值解、最优化理论与方法、数值逼近与数值代数、计算基础等学科子方向。然而,对于信息科学,它到底应该怎样划分学科子方向?应该怎样设置专业与专业基础课?所有这些都仍是正在探索的问题。

任何技术都可以认为是延伸与扩展人的某种功能的方式与方法,所以信息技术可以认为是扩展人的信息器官功能的技术。人的信息器官主要包括感觉器官、传导器官(传导神经网络)、思维器官和效应器官四大类型,其功能则主要是信息获取、信息传输、信息处理和信息应用(控制),因而感测技术、通信技术、智能技术与控制技术通常被认为是最基本的信息技术(常称之为信息技术的四基元),其他信息技术可认为是这四种基本技术的高阶逻辑综合或分解衍生。所以可以把信息科学理解为是“有关信息获取、信息传输、信息处理与信息控制基础的科学”。从这个意义上,我们认为:信息处理(包括图像处理、信号分析等)、信息编码与信息安全、计算智能(人工智能、模式识别等)、自动控制等可构成信息科学的主要学科子方向。这一认识也是教指委设置信息与计算科学专业信息科学方向课程的基本依据。

本系列教材正是基于以上认识,为落实新的《信息与计算科学专业教学规范》(讨论稿)而组织编写的。我们相信,该系列教材的出版对缓解本专业教材的紧缺局面,对推动信息与计算科学专业的快速与健康发展会大有裨益。

从长远的角度看,为适应不同类型院校和不同层次要求的课程需求,本系列教材编委会还将不断组织教材的修订和编写新的教材,从而使本专业的教学用书做到逐步充实、完善和多样化。我们诚恳希望采用本系列教材的教师、同学们及广大读者对书中存在的问题及时指正并提出修改意见和建议。

信息与计算科学专业系列教材编委会

2003年8月31日

前 言

信息技术被广泛认为是 21 世纪主导全球经济发展格局、引导社会发展进程的核心技术。信息技术的物质基础是光电子、微电子、传感器等,其载体是计算机,其核心是通讯技术,而其理论基础与支撑则是数学。数学技术不仅提供信息表示与编码的方式和语言,而且也提供信息处理(如转换、压缩、提取、加密等)、信息加工与信息利用的直接基础与核心算法。对于很多高、新信息技术而言,它们的本质是数学技术。

信息技术当今的发展以数字化、网络化、智能化为特征。所谓数字化是指信息的表示、储存、传输与处理以数字(特别是离散数字)为基准;网络化是指承载信息,乃至处理、加工、应用信息的主渠道是网络,或者说,是分布式并行信息处理系统;智能化则是指在信息处理的方式上溶入或模仿人的智能。让机器“听”懂人类语言、“看”清文字图像、与人“说”话,让信息处理系统像人那样具有综合、优化、联想、辨识、学习等能力,所有这些都是智能化信息技术所努力追求的目标。虽然说,目前的信息技术还远未达到上述所期望的那样高度智能化水平,但在过去的几十年间,人们的确取得了大批令人振奋并极大推动了信息处理智能化进程的成果,计算智能(computational intelligence)正是这些突出成果中的一个典型代表。

计算智能是借助现代计算工具模拟人的智能求解问题(或处理信息)的理论与方法,它是人工智能(artificial intelligence)的深化与发展。如果说人工智能是以知识库(更确切些,专家规则库)为基础、以顺序离散符号推理为特征的,计算智能则是以模型(计算模型、数学模型)为基础,以分布、并行计算为特征。前者强调规则的作用与形成,而后者强调模型的建立与构成;前者依赖专家个人知识,而后者强调自组织、自学习与自适应。

模拟人的智能通常基于不同的观点与角度,例如,从模拟智能生成过程的观点,从模拟智能产生与作用所赖以存在的结构角度,从智能的表现行为角度等等。在所有这些方面,计算智能在近 20 多年的发展中都取得了巨大成就。本书编写目的是在上述诸方面的每一方面选择一个典型,以向读者系统介绍计算智能的基本理论、基本原理与基本方法。在智能生成过程模拟中,我们选择介绍了模拟进化计算;在智能结构模拟中,我们选择讲述了人工神经网络理论;在智能

行为模拟中,我们则选择介绍了模糊逻辑与模糊推理。之所以如此选择,一是作者多年在这方面工作,二是它们既能代表计算智能过去 20 多年间发展的最突出成果,又能反映计算智能的当前研究热点。

国内外已有多本著作专门论述上述之一或全部内容,但很可惜,它们或者内容过于宽泛,或者过于偏重工程应用,还没有一本能适用于为数学系学生讲授的著作。鉴于近年来投身于信息科学研究的数学工作者越来越多,而且计算智能已被列为信息与计算科学专业的专业课之一,编著一套适宜于对数学系学生讲授而且能够引导青年数学工作者尽快进入计算智能领域的教材变得十分必要和紧迫,这即是编著本书的直接动机与原由。因此,本书的编著原则是:(1) 突出基础(特别是数学基础)。尽可能以严密的理论和形式化的定义、定理、算法与模型的形式提出和表述问题,避免空泛;(2) 强调背景。在注重形式化的数学表述基础上,突出介绍所述理论与方法的生物背景与工程应用背景,力求使读者对方法的“源头”和“流向”有明确的了解,对理论的价值与应用有直接的认识;(3) 着眼于研究与发展。在内容的取舍上既考虑它们自身的理论与应用价值,也考虑它们的可研究性与可发展性,侧重介绍了那些既有重要理论与应用价值,又有广阔研究空间与发展前景的新理论与新方法。在每一册,我们也专章介绍与之相关的发展动态与研究进展,特别提出具体的研究问题。所有这些旨在使读者能尽可能快地直接进入计算智能研究的相关前沿。

为了满足各高校不同层次的教学需要,特别是既可将本书作为一门专业课讲授,亦可将本书的某些部分内容作为选修课讲授,我们将整体内容分册独立出版。

本书除可供信息与计算科学专业、数学与应用数学专业的本科生作为专业课或选修课教材使用外,也可作为计算数学、信息科学、运筹与控制、计算机科学、系统科学等专业的研究生教材使用,亦可供从事计算智能研究与应用的各学科研究工作者参考。

编 著 者

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人给予严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话：(010) 58581897/58581698/58581879/58581877

传 真：(010) 82086060

E - mail：dd@hep.com.cn 或 chenrong@hep.com.cn

通信地址：北京市西城区德外大街4号

高等教育出版社法律事务部

邮 编：100011

购书请拨打电话：(010)64014089 64054601 64054588

策划编辑 王 瑜
责任编辑 王 瑜
封面设计 王凌波
责任绘图 郝 林
版式设计 陆瑞红
责任校对 杨凤玲
责任印制 韩 刚

目 录

第 I 章 概 论 1

- § 1.1 模拟进化计算技术 1
 - 1.1.1 什么是模拟进化计算技术? 1
 - 1.1.2 模拟进化计算技术所处理的基本问题 2
- § 1.2 模拟进化计算的生物学基础 4
 - 1.2.1 遗传变异理论 4
 - 1.2.2 进化论 5
 - 1.2.3 遗传与进化的系统观 5
 - 1.2.4 免疫学原理 6
- § 1.3 模拟进化计算的一般框架 8
- § 1.4 典型例子 10
 - 1.4.1 遗传算法(GA) 10
 - 1.4.2 免疫算法(IA) 11
 - 1.4.3 演化策略(ES) 12
- § 1.5 模拟进化计算的本质优点与适用领域 13
- 习题一 15

第 2 章 模拟进化算法的基本要素与数学描述 17

- § 2.1 模拟进化算法的形式化模型 17
- § 2.2 编码格式 19
- § 2.3 适应度度量 27
- § 2.4 选择算子 28
 - 2.4.1 比例型 29
 - 2.4.2 排序型 31
 - 2.4.3 非单调型 31

§ 2.5 繁殖算子 33
2.5.1 变异算子 34
2.5.2 交叉算子 36
§ 2.6 进化参数 38
习题二 40

第 3 章 模拟进化计算的典型执行策略 41

§ 3.1 杰出者记录与“父子混合”选择策略 41
§ 3.2 适应值共享策略 43
§ 3.3 并行实现策略 45
3.3.1 基于群体分组的并行策略 46
3.3.2 基于空间分解的并行策略 48
§ 3.4 混合策略 50
§ 3.5 自适应策略 55
习题三 60

第 4 章 遗传算法的搜索机理 61

§ 4.1 种群增长方程 61
§ 4.2 交叉算子的搜索可达域 65
§ 4.3 变异算子的搜索可达域 68
§ 4.4 选择算子的搜索能力与搜索速度 72
§ 4.5 遗传算法的搜索机制 74
习题四 75

第 5 章 遗传算法的收敛性理论 76

§ 5.1 种群序列的收敛性定义及性质 76
§ 5.2 遗传算法的马氏链分析 78
5.2.1 马氏链的定义及相关性质 78
5.2.2 标准遗传算法的马氏链分析 85
5.2.3 杰出者选择遗传算法的概率收敛性 87
§ 5.3 遗传算法的公理化分析 88

5.3.1	进化算子的特征数	89
5.3.2	抽象模拟进化算法的概率收敛性定理	92
5.3.3	应用举例	96
§ 5.4	遗传算法的鞅分析	98
5.4.1	条件期望与鞅	98
5.4.2	遗传算法的几乎必然收敛性	99
习题五		109

第 6 章 模拟进化计算的新近发展 111

§ 6.1	蚁群算法	111
§ 6.2	粒子群优化	114
§ 6.3	差分演化算法	116
§ 6.4	人口迁移算法	118
§ 6.5	基于思维进化的机器学习	120
习题六		123

第 7 章 评注与展望 124

§ 7.1	有关理论基础研究	124
§ 7.2	有关算法设计	129
§ 7.3	有关模拟进化计算的应用	132

参考文献 134

第 I 章 概 论

本章从整体框架上讲述模拟进化计算 (simulated evolutionary computation, 简称 SEC) 的基本原理、生物基础、算法框架、本质优点与适用领域, 目的在于使读者对模拟进化计算技术有一个整体了解, 以增强对以后各章节内容重要性及关联性的认识。

§ 1.1 模拟进化计算技术

本节从宏观、定性的意义上定义什么是模拟进化计算技术, 阐述它所处理的基本问题与求解问题的基本原理。

1.1.1 什么是模拟进化计算技术?

模拟进化计算 (SEC) 技术是模拟自然界生物进化过程与机制求解优化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 (从简单到复杂, 从低级向高级) 本身是一个自然的、并行发生的、稳健的优化过程。这一优化过程的目的在于使生命体达到适应环境的最佳结构与效果, 而生物种群通过“优胜劣汰”及遗传变异来达到进化 (优化) 的目的。曾经主宰地球的恐龙由于庞大的身躯跟不上环境的变迁而灭绝; 长颈鹿为了觅食而长长了脖子; 老鼠的机敏是为了生存而挣扎的结果; 青蛙的存活则得益于其两栖的能力; 人类解放出有力的双手, 得益于类人猿求生的努力, 而正是这一对不再用于行走的双手, 使人类成了这个世界的主宰。所有这些无不显示出自然进化的神奇力量。然而, 自然进化遵循着一种奇妙的规律, 这一规律截然不同于人类自己发明的任何解析方法。我

们知道,人体 $36.8\text{ }^{\circ}\text{C}$ 的恒定体温是至关重要的, $0.5\text{ }^{\circ}\text{C}$ 的偏差足以使人产生病态.那么在 $0\text{ }^{\circ}\text{C} \sim 100\text{ }^{\circ}\text{C}$ 这样大的范围内,人体是如何求出如此精确的最佳体温呢?这绝不是诸如梯度下降法或模拟退火算法所能算出的,而我们也相信目前的解析方法远远无法完整描述体温、人体结构与环境间错综复杂的关系.自然进化的规律是达尔文(Darwin)发现的,它的核心思想是“物竞天择、适者生存”.它对人类的作用与影响可以说不亚于牛顿(Newton)的万有引力学说,而奇怪的是这一理论时至今日才被应用于工程实践.

SEC 是一种更为宏观意义下的仿生优化算法,它模仿的是一切生命与智能的生成与进化过程.它不仅模拟达尔文“优胜劣汰、适者生存”的进化原理激励好的结构,而且也通过模拟孟德尔等人的遗传变异理论在优化过程中保持已有的结构,同时寻找更好的结构.在以下 § 1.2,我们将进一步介绍有关遗传变异理论的要义,以使读者对 SEC 的生物学基础有一个较完整的认识.

1.1.2 模拟进化计算技术所处理的基本问题

从抽象的意义上说,SEC 所处理的问题可描述作

$$\max_{X \in \Omega} f(X), \quad (1.1)$$

这里 Ω 是某个 n 维空间(例如 \mathbf{R}^n)上的一个子集, $f: \Omega \rightarrow \mathbf{R}^1$ 是一个可能无法解析表达的函数(例如, X 取为人体的体温, $\Omega = [0\text{ }^{\circ}\text{C}, 100\text{ }^{\circ}\text{C}]$, f 为人体在给定环境下的适应程度).(1.1)自然也可理解为常义的极大化问题,它特别包含各类数学规划、组合优化、随机优化等优化模型.

例 1.1 (囚犯困境问题) 囚犯困境(prisoner dilemma)是一个古老的游戏问题.假定有两个囚犯共同参与了某项犯罪且被关在相互隔离的牢房里.每个囚犯都被要求供出犯罪事实,而依据他们各自的表现,审问人员将对他们进行惩罚或给出不等程度的奖励.每个囚犯或者选择背叛另一方,或选择与对方攻守同盟(即合作).如果只有一个囚犯选择背叛,该囚犯就会被奖励,另一个囚犯被惩罚;如果两个都选择背叛,则两个都被监禁并受到折磨;如果两个都选择合作,则两个都会受到中等的奖励.例如,对各囚犯的奖分以表 1.1 决定,则囚犯困境问题是指:对每一个囚犯而言,如何选择背叛还是合作,以使自己能获得最高的奖励.

表 1.1 囚犯困境游戏奖分表(p_i 为囚犯 i 的奖分)

囚犯 1	囚犯 2	p_1	p_2	评 注
背叛	背叛	1	1	相互背叛受惩罚
背叛	合作	5	0	背叛受奖励
合作	背叛	0	5	受奖诱发背叛
合作	合作	3	3	相互合作受奖

这是一个 SEC 所能解决的典型问题(文献[31, 60]). 在这个问题中, $\Omega = \{\text{合作, 背叛}\}$, 而目标函数 f 是不能解析表达的(事实上, 即使是对它的估值, 也有赖于对方(另一囚犯)的表现才能决定).

例 1.2 (旅行商问题) 旅行商问题(traveling salesman problem, 简称 TSP)是指这样一个组合优化问题: 某个商人欲到 n 个城市 c_1, c_2, \dots, c_n 推销商品, 希望选择一条路线使得商人走遍每个城市(仅能访问一次)后回到起点且所走路程最短. 如令城市 c_i 与 c_j 的距离为 d_{ij} , 用 x_{ij} 表示商人是否以顺序 i 访问城市 c_j (更具体地, $x_{ij} = 1$ 表示商人以顺序 i 访问城市 c_j , 否则 $x_{ij} = 0$), 则 TSP 问题可以描述作如下优化问题:

$$\min_{x_{ij} \in \{0, 1\}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n d_{ij} x_{ij},$$

$$\text{s. t. } \begin{cases} \sum_{j=1}^n x_{ij} = 1, i = 1, 2, \dots, n, \\ \sum_{i=1}^n x_{ij} = 1, j = 1, 2, \dots, n, \\ \sum_{i=1}^n \sum_{j=1}^n x_{ij} = n. \end{cases}$$

这也是一个 SEC 作为 benchmark 问题求解的典型问题. 在这个问题中,

$$\Omega = \left\{ (x_{ij})_{n \times n} \in \{0, 1\}^{n \times n} \mid \sum_{j=1}^n x_{ij} = 1, i = 1, 2, \dots, n; \right. \\ \left. \sum_{i=1}^n x_{ij} = 1, j = 1, 2, \dots, n; \sum_{i=1}^n \sum_{j=1}^n x_{ij} = n \right\};$$

目标函数 f 是已知的, 且可定义为

$$f(x_{ij}) = - \sum_{i=1}^n \sum_{j=1, j \neq i}^n d_{ij} x_{ij}.$$

例 1.3 (随机函数的优化问题) 考虑如下带随机噪声的优化问题

$$\max_{x \in \Omega} f(x) = - \sum_{i=1}^{30} i x_i^4 - \text{random}[0, 1),$$

其中 $\text{random}[0, 1)$ 表示取值在 $[0, 1)$ 内的随机数, 而

$$\Omega = \{x = (x_1, x_2, \dots, x_{30}) \mid |x_i| \leq 1.28\}.$$

作为求解(1.1)的一类仿生优化算法, SEC 的基本原理可概括为: (1) 将问题(1.1)的目标函数 f (已知或未知) 对应作某生物群体在给定环境下的适应性度量; (2) 将(1.1)中的优化变量 X 对应作生物群体中的个体; (3) 将 SEC 过程对应作相应生物群体在给定环境下的演化进程. 所以, 为了理解和发展 SEC, 我们必须对自然进化、遗传变异及相关生物学基础有一个清晰的了解.

§ 1.2 模拟进化计算的生物学基础

我们首先概述有关遗传变异的生物学基础,然后概述自然进化论以及遗传与进化的系统观点,最后再介绍必要的免疫学知识,所有这些构成 SEC 的核心生物学基础.

1.2.1 遗传变异理论

世间的生物从其亲代继承特征或性状,这种生命现象称之为遗传(heredity),研究这种生命现象与机理的科学即为遗传学(genetics).由于有遗传作用,自然界才有稳定的物种,人们种瓜得瓜,种豆得豆,鱼之所以至今还仍然会在水中遨游,鸟仍然还会在天空中飞翔都是这个缘由.

构成生物的基本结构与功能单位是细胞(cell).细胞中含有一种微小的丝状化合物称为染色体(chromosome),生物的所有遗传信息都包含在这个复杂而又微小的染色体中.遗传信息是由基因(gene)组成的,生物的各种性状由其相应的基因所决定,基因是遗传的基本单位.细胞通过分裂具有自我复制的能力,在细胞分裂过程中,其遗传基因也同时被复制到下一代,从而其性状也被下一代所继承.经过生物学家研究,现在人们已经明白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)的物质所构成.除此之外,染色体中还含有大量蛋白质. 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.每个核苷酸由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所组成.许多核苷酸通过磷酸二酯键相结合形成一个长长的链状结构,两个链状结构再通过碱基间的氢键互补成对地组合在一起,相互卷曲起来形成一种双螺旋结构(参见图 1.1).另外,低等生物中还含有一种叫作核糖核酸(ribonucleic acid,简记 RNA)的物质,它的作用与 DNA 类似.基因就是 DNA 或 RNA 长链中占有一定位置的基本遗传单位.生物的基因数量根据物种的不同也多少不一,小的病毒只含有几个基因,而高等动植物的基因却以数万计(例如人的基因在 3~5 万个之间).DNA 中,遗传信息在一条长链上按一定的模式排列,亦即进行了遗传编码.一个基因或多个基因决定了蛋白质的 20 种氨基酸的组成比例及其排列顺序.遗传基因在染色体中所占据的位置称为基因座(locus),同一基因座可能有的全部基因称为等位基因(allele).一种生物所特有的基因及其构成形式称为该生物的基因型(genotype),而该生物在环境中所呈现出的相应性状被称为该生物的表现型(phenotype).一个细胞核中所有染色体所携带的遗传信息(即 DNA)的全体称