

进化计算与粗糙集 研究及应用

邱玉霞 编著



冶金工业出版社

<http://www.cnmp.com.cn>

进化计算与粗糙集 研究及应用

邱玉霞 编著

北 京
冶金工业出版社
2009

内 容 简 介

本书以进化计算与粗糙集为研究目标,内容包括进化计算的数列模型及其在收敛性分析中的应用,基于种群信息熵的思维进化算法自适应搜索策略研究,基于位编码可分辨矩阵的规则提取方法的研究,基于思维进化算法和粗糙集的图像处理方法的研究等。

本书可作为计算机专业师生教学参考用书,也可供计算机相关专业研究人员、从业人员阅读。

图书在版编目(CIP)数据

进化计算与粗糙集研究及应用/邱玉霞编著.

—北京:冶金工业出版社,2009.5

ISBN 978-7-5024-4921-6

I. 进… II. 邱… III. ①算法理论—研究
②数值计算—研究 IV. 0241

中国版本图书馆 CIP 数据核字(2009)第 057634 号

出 版 人 曹胜利

地 址 北京北河沿大街嵩祝院北巷 39 号,邮编 100009

电 话 (010)64027926 电子信箱 postmaster@cnmip.com.cn

责任编辑 贾 玲 李培禄 美术编辑 张媛媛 版式设计 葛新霞

责任校对 卿文春 责任印制 李玉山

ISBN 978-7-5024-4921-6

北京百善印刷厂印刷;冶金工业出版社发行;各地新华书店经销

2009 年 5 月第 1 版,2009 年 5 月第 1 次印刷

850mm×1168mm 1/32;5.5 印张;150 千字;164 页;1-2000 册

20.00 元

冶金工业出版社发行部 电话:(010)64044283 传真:(010)64027893

冶金书店 地址:北京东四西大街 46 号(100711) 电话:(010)65289081

(本书如有印装质量问题,本社发行部负责退换)

前 言

计算智能方法往往具有自学习、自组织、自适应的特征和简单、通用、鲁棒性强、适于并行处理的优点，在并行搜索、联想记忆、模式识别、知识自动获取等方面得到了广泛的应用。进化计算与粗糙集理论是计算智能的两个热点研究方向，是目前信息科学、自动化科学、计算机科学的交叉和前沿研究领域。思维进化算法是模拟人类思维进化过程的一种新型进化计算方法。本书分别以进化计算与粗糙集为研究目标，研究内容分为四大部分，一是进化计算的数列模型及其在收敛性分析中的应用；二是基于种群信息熵的思维进化算法自适应搜索策略研究；三是基于位编码可分辨矩阵的决策规则获取算法的研究；四是基于思维进化算法与粗糙集的图像处理方法的研究。

具体内容包括：

(1) 从进化机制出发，研究了进化算法的种群进化的特点并定义了种群适值函数，进而建立了进化的数列模型，并分析了在该模型下几种典型进化算法的收敛特性；

(2) 研究了思维进化算法的进化机理，引入信息论中信息熵的概念，提出了基于种群进化熵的思维进化算法，并在进化计算的统一框架下，证明了该算法在数列意义下是收敛的，数值优化实验表明该算法具有良好的

性能;

(3) 研究了粗糙集理论的可分辨矩阵,提出了基于位编码可分辨矩阵规则获取策略,并将其成功应用于水泥窑炉运行操作的决策规则获取;

(4) 提出了基于粗糙集理论和思维进化算法的图像分析新方法,将基于种群进化熵的思维进化算法应用于数字图像分割的最佳阈值寻优;应用粗糙集理论中决策表建立形状分类机制,并应用基于位编码可分辨矩阵的规则获取算法,提取决策规则;将上述方法综合应用于染色体畸变分析系统的设计中。

本书的创新性成果包括:

(1) 建立了进化计算方法的数列模型。把复杂的随机过程映射成为种群适值序列,从该序列的性质来分析种群的进化过程,从而便于用数学方法分析种群的进化过程,为进化计算理论研究提供了一种新方法。

(2) 应用数列模型分析了几种典型进化算法的收敛性。给出了基于种群适值链的进化算法的收敛条件,使用区间套等定理证明了进化计算方法的全收敛性。

(3) 提出了基于种群进化熵的思维进化算法。将信息论中信息熵的思想引入思维进化算法的进化操作设计,改进了思维进化算法的趋同操作,算法可根据种群进化信息估计种群进化熵,实现搜索区域自适应调整,提高了搜索效率。

(4) 提出了基于位编码可分辨矩阵的规则获取策略。首先分析了可分辨矩阵求取属性值约简的可能性及合理性,进而提出了基于位编码可分辨矩阵规则获取算法。该算法以位编码可分辨矩阵为基础,实现属性和属

性值约简，并将其应用于水泥窑炉运行操作的决策规则约简。

(5) 将基于种群进化熵的思维进化算法和粗糙集理论分别应用于图像处理的图像分割和形状分类中。设计了基于思维进化算法与粗糙集理论的染色体畸变分析系统。

在本书的编写过程中，参阅了国内外一些进化计算与粗糙集研究方面的资料和文献，吸收和借鉴了国内外学者的有关研究成果，在此向他们表示衷心的感谢。

由于水平有限，书中不妥之处，恳请读者批评指正。

作 者

2009 年 2 月

目 录

1 绪论	1
1.1 本书背景及意义	1
1.2 进化计算与思维进化算法	2
1.2.1 进化计算理论	3
1.2.2 思维进化算法 (MEA)	6
1.3 粗糙集理论	9
1.3.1 粗糙集基本概念	10
1.3.2 粗糙集研究现状	11
1.4 本书的主要内容及结构安排	17
1.4.1 主要研究内容	17
1.4.2 本书的结构安排	18
参考文献	19
2 进化计算的数列模型及收敛性分析	30
2.1 概述	30
2.2 进化计算	31
2.2.1 进化计算的主要特点	32
2.2.2 进化计算的主要分支	33
2.2.3 进化算法的进化机制	38
2.2.4 进化算法的统一性描述	41
2.3 进化计算的种群适值链及收敛性分析	44

2.3.1 种群适值链	44
2.3.2 基于种群适值链的收敛性分析	46
2.3.3 仿真分析	53
2.4 本章小结	55
参考文献	55
 3 基于种群进化熵的思维进化算法	58
3.1 概述	58
3.2 思维进化算法	58
3.2.1 思维进化算法的提出	59
3.2.2 基本概念	61
3.2.3 算法特点	63
3.2.4 算法构架	64
3.2.5 研究现状与前景	68
3.3 基于种群进化熵的思维进化算法 (PEMEA)	73
3.3.1 进化熵	73
3.3.2 基于种群进化熵的搜索策略	75
3.3.3 算法收敛性分析	77
3.3.4 数值优化实验	81
3.4 本章小结	92
参考文献	93
 4 基于位编码可分辨矩阵的规则提取方法研究	97
4.1 概述	97
4.2 粗糙集理论基本概念	97
4.2.1 知识与知识表达系统 ^[1,6]	97
4.2.2 粗糙集合基本概念 ^[1,6]	100
4.2.3 知识的约简	102

4.2.4 决策表与决策规则	106
4.3 基于位编码可分辨矩阵的规则获取的算法	109
4.3.1 可分辨矩阵与决策表属性约简 ^[5]	109
4.3.2 可分辨矩阵与决策表属性值约简	110
4.3.3 基于位编码可分辨矩阵的决策表约简	113
4.4 本章小结	120
参考文献	121
5 PEMEA 与粗糙集理论在图像处理中的应用	122
5.1 概述	122
5.1.1 图像数字化	123
5.1.2 图像变换	123
5.1.3 图像增强	123
5.1.4 图像恢复	123
5.1.5 图像分割	123
5.1.6 图像分析与描述	124
5.1.7 图像数据压缩	124
5.1.8 图像重建	124
5.2 基于思维进化算法的数字图像分割方法研究	125
5.2.1 图像分割概述	125
5.2.2 基于思维进化算法的图像分割方法	130
5.2.3 仿真实验	133
5.3 基于粗糙集理论的形状分析方法	135
5.3.1 形状分析概述	136
5.3.2 基于粗糙集理论的形状分析方法	141
5.3.3 仿真实验	143
5.4 染色体畸变分析系统设计	148
5.4.1 染色体及核型	149
5.4.2 染色体畸变分析系统	149

5.5 本章小结	159
参考文献	160
6 结论与展望	162
6.1 结论	162
6.2 展望	164

1 绪 论

1.1 本书背景及意义

1956 年 John McCarthy 在 Dartmouth 大学召开的会议上提出了“人工智能”这一术语,从此一门新兴的学科便正式诞生了。至今人工智能已发展成了一门涉及多个学科的交叉学科,广泛应用于复杂系统的智能控制、信息安全、工程设计、图像处理、调度规划、优化理论和人工生命等领域^[1~3]。

智能是指个体经过合理的思维,通过有目的的行为,有效地解决问题和适应环境的综合性能力。人工智能(Artificial Intelligent, AI)就是研究如何用人工的方法和技术,模仿、延伸和扩展人类智能,以及如何提高机器的智能水平,使机器成为具有感知、推理、决策的智能机器系统^[1~3]。

人工智能是一门交叉学科,它与脑科学(Brain Sciences)、认知科学(Cognitive Sciences)共同构成智能科学,探索人类智慧的奥秘与规律,研究智能行为的基本理论和实现技术及在机器中复现人类智能。世界上许多科学家和研究人员分别从不同角度探索对人类智能的模拟方法,并逐步形成了符号主义(Symbolism)、连结主义(Connectionism)和行为主义(Behaviorism)三大流派的理论与方法。

传统的人工智能是符号主义,它以 Newell 和 Simon 提出的物理符号系统假设和有限合理性原理为基础,认为物理符号系统是智能行为充分和必要的条件。物理符号系统由一组符号实体组成,它们都是物理模式,可在符号结构的实体中作为组分出现。该系统可以进行建立、修改、复制、删除等操作,以生成其他符号结构。Feigenbaum^[4]研究成功的第一个专家系统及其提出的知识工程(Knowledge Engineering)的概念成为人工智能走向工程应用

的典型代表。

连接主义与行为主义间的差距相对较小,两者都源于模拟自然生命现象,是信息科学与生命科学相互交叉、相互渗透和相互促进的产物。因此可将人工智能方法分为两大类,即符号智能和计算智能(或智能计算)。符号智能是以知识为基础,通过推理进行问题求解,也即传统的人工智能。计算智能是以数据为基础,通过训练建立联系,进行问题求解。

计算智能是以生物进化的观点认识和模拟智能^[1]。按照这一观点,智能是在生物的遗传、变异、生长以及外部环境的自然选择中产生的。在用进废退、优胜劣汰的过程中,适应度高的(头脑)结构被保存下来,智能水平也随之提高。因此说计算智能就是基于结构演化的智能。计算智能以连接主义的思想为主,形成了人工神经网络、进化计算、人工免疫算法、蚁群算法、模糊集、粗糙集、多智能体(Multi-agent)系统等众多的理论和研究领域。这些方法通常具有以下共同的要素:自适应的结构、随机产生的或指定的初始状态、适应度的评测函数、修改结构的操作、控制过程的参数。计算智能的这些方法往往具有自学习、自组织、自适应的特征和简单、通用、鲁棒性强、适于并行处理的优点。在并行搜索、联想记忆、模式识别、知识自动获取等方面得到了广泛的应用。

20世纪50年代以后一段时间,符号智能体系取得了巨大的成功,但80年代中期以来,这种经典人工智能的发展由辉煌转入相对停滞,而计算智能在神经网络的带动下异军突起。与生命科学、系统科学密切联系是计算智能的突出特点,正是由于这个特点,不仅计算机科学家,而且众多其他学科的学者也加入到计算智能的研究中来,极大地促进了它的发展^[5]。

1.2 进化计算与思维进化算法

计算智能以连接主义的思想为主,形成了人工神经网络、进化计算、人工免疫算法、蚁群算法、模糊集等众多的理论和研究领

域。各种理论与方法虽然所强调的内容不同,但是它们都建立在人或生物体的生物学基础上,通过模拟生物体系统来完成一定的智能任务。进化计算^[6] (Evolutionary Computation, EC) 这一术语是在 20 世纪 90 年代初被提出的,是一种模拟生物遗传、进化的随机寻优技术。其主要分支有遗传算法 (Genetic Algorithms, GA)、遗传规划 (Genetic Programming, GP)、进化策略 (Evolution Strategies, ES) 和进化规划 (Evolutionary Programming, EP)。

进化计算学科的出现,促进了不同分支之间的交流,可以取长补短。各个分支都有自己的优缺点,研究它们的优越性,并融合成新的进化算法,可以促进进化计算更加广泛的应用^[7]。思维进化算法 (Mind Evolutionary Algorithm, MEA) 是一种新型的进化算法^[8],其思想来源于模仿人类进化过程中思维的进化。它继承了进化主义的“群体”和“进化”的概念,改进了群体优化过程中的搜索策略和进化计算的进化操作算子。

1.2.1 进化计算理论

生物是解决问题的能手。正是在自然界的启示下,一些学者希望通过模拟自然界的生物进化过程来实际问题,于是诞生了进化计算这一学科。

进化计算 (EC) 这一术语是在 20 世纪 90 年代初被提出的,是一种模拟生物遗传、进化的随机寻优技术,由遗传算法 (GA)、遗传规划 (GP)、进化策略 (ES) 和进化规划 (EP) 等分支组成,其他的诸如 DNA 计算和分子计算,也开始应用在实际问题中。上述进化计算的四个主要分支是由不同学者提出的,基本上是独立发展,没有交流。各个分支都有自己的优缺点,研究它们的优越性,并融合成新的进化算法,可以促进进化计算更广泛地应用。

进化计算是一种通用的问题求解方法,它采用某种编码技术来表示各种复杂的结构,并通过遗传操作模拟进化过程,最终获得性能指标高的个体,即问题的解。

1.2.1.1 进化计算理论的基本思想

达尔文的进化论所揭示的进化机制在本质上是一种鲁棒搜索和优化过程。进化了的动植物种群在细胞、器官、个体和群体这多个不同层次上都表现出被优化了的复杂行为。生物物种在进化过程所解决的各种问题具有混沌、偶然、暂态和非线性相互作用等特点,具有这样特点的问题正是传统优化方法所难以解决的。而进化过程就可以用来解决那些启发式解不存在或者由启发式解得不到满意结果的问题。

在进化理论中,目前为人们所普遍接受的是所谓的“新达尔文模型”^[6]。这种理论认为生命的历史能够用种群和物种中发生的一系列的物理过程来解释。这一系列的物理过程包括了再生变异竞争和选择。新达尔文模型理论可以总结为:个体是选择的主要作用对象;遗传变异在很大程度上是一种偶然现象,随机过程在进化中起到了十分重要的作用;基因型的变化在很大程度上是基因重组的结果,仅在相当小的程度上是变异的结果;在渐进的进化过程中可以伴随有表型的变异;并非所有的表型变化都是自然选择的结果;进化是自适应性和多样性的变化过程,而不仅仅是基因编码的变化;选择是概率性的,不是确定性的。这些生物进化论上的观点就构成了模拟进化研究的理论基础。

1.2.1.2 进化计算的主要内容

进化计算的研究内容相当广泛,反映了多学科交叉的特点,目前,进化计算的研究主要集中在以下几个方面^[7]:

(1) 进化计算的理论研究。由于进化计算缺乏统一、完整的理论体系,为推动进化计算研究的发展,迫切需要宏观的理论指导。

(2) 新的计算模型。目前实现的计算模型只是生物进化的很小一部分,仍有许多进化模型可以发掘。一些新的计算模型在解决某些问题时取得了较好的效果,如免疫系统模型、协同进化模

型及思维进化算法等。

(3)应用研究。如进化优化、训练人工神经网络及图像处理和模式识别等。

(4)并行和分布式演化计算。

(5)进化机器学习。

(6)进化计算内涵扩充。进化计算已从模拟生物的进化过程扩充到模拟大自然的进化过程(Problem solving from nature, 也称 Natural computation)。它不仅采用“仿生”策略, 也通过模拟“拟物”进化过程来进行问题求解。

1.2.1.3 进化计算的研究现状

20 世纪 80 年代开始, 美国 Illinois 大学的 D. E. Goldberg 在他领导的 Illinois 遗传算法实验室(Illinois Genetic Algorithms Laboratory, IlliGAL)中对遗传算法的基本理论进行了广泛的研究, 取得了一系列的成果^[9~15]。早在 1989 年, 他就出版了目前被认为是遗传算法最经典、最全面的教科书^[16]。

1992 年 H. Kargupta 将 Shannon 的信息熵^[18]引入遗传算法, 研究了遗传算法群体的多样性, 他借鉴了通信领域中的信息冗余性, 研究了在不损失进化群体的多样性的情况下如何提高冗余性^[18]。

关于遗传算法的控制参数的优化和优化效率的性能评估的研究也引起一些学者的重视, 遗传算法的控制参数是指群体规模、杂交率、变异率以及其他一些遗传算法的参数。目前, 还没有一种通用的方法, 可以根据要解决的问题自动地设置高效的控制参数, 只能根据经验和多次运行遗传算法的结果进行人工设置和修正。J. J. Grefenstette^[19]利用遗传算法来优化遗传算法的控制(MetaGA), 通过 K. A. DeJong 提出的优化效率的性能评价指标“在线性能”与“离线性能”来构造适应度, 获得了一些有益的结论。

Holland 的标准遗传算法并没有提供如何解决约束优化问题, 在应用遗传算法解决工程中的复杂优化问题时, 对于约束的处理是一个至关重要的问题。Michalewicz 对此作了坚持不懈的研究,

提出了几种较有效的方法^[20, 21]。对于约束的处理,应用最多的方法是惩罚函数法^[22, 23]。

武汉大学软件工程国家重点实验室在国内的研究中起步较早,进化计算(亦称为“演化计算”)在他们的并行计算研究室是一个重要的研究方向,并已经出版了专著^[24, 25];有许多硕士、博士研究生围绕进化计算选题^[26~29]。另外,关于进化计算方面的著作不断面世^[6, 30, 31]。

太原理工大学在承担国家 863 项目(863-306ZT06-6)和国家自然科学基金项目(60374029)的研究过程中提出的思维进化算法,已成为进化计算中的一个重要研究领域,并应用在优化计算^[32, 33]、图像处理^[34, 35]、系统建模^[36]和智能控制^[37~44]等领域。

1.2.2 思维进化算法(MEA)

大自然是我们解决各种问题时获得灵感的源泉。几百年来,将生物界所提供的答案应用于实际问题求解已被证明是一个成功的方法。众所周知,自然界所提供的答案是经过漫长的自适应过程——演化过程而获得的结果。除了演化过程的最终结果,我们也可以利用这一过程本身去解决一些较为复杂的问题^[45]。这样,我们不必非常明确地描述问题的全部特征,只需要根据自然法则来产生新的更好解。进化计算正是基于这种思想而发展起来的一种通用的问题求解方法。更由于它的不断发展及其在一些应用领域内取得的成功,进化计算表现出了良好的应用前景。据德国 Dortmund 大学 1993 年末的一份研究资料报道,根据不完全统计,进化计算方法已在 16 个大领域、250 个小领域中获得了应用^[46]。

随着对进化计算理论研究的深入,人们已不满足于原有的计算模型,一些新的计算模型在解决某些问题时取得了较好的效果,如免疫系统模型和协同进化模型等。思维进化算法是一种新型的进化计算模型,其思想来源于模仿人类思维进化的过程,即模仿人类思维中趋同、异化两种思维模式交互作用,推动思维进

步的过程。它继承了进化主义的“群体”和“进化”的概念；以“群体寻优”代替“个体寻优”。

1.2.2.1 基本思想

达尔文将生存斗争的适者生存,不适者淘汰的过程叫做自然选择。自然选择的过程表明,竞争是生物进化的内在动力。而生物个体通过遗传,将成功者的基因传到下一代,而下一代在继承父代的基因基础上,又发展了自身。这样,遗传和变异成为生物进化的内在因素。这就是我们通常所指的自然进化过程。

自然进化过程只是自然界中“智能”现象的一种,实际上,生物体还有很多可以挖掘的“智能”。如生物的免疫系统具有抗原识别、细胞分化、记忆和自我调节功能^[47],从而实现自身的免疫,抑制异物的入侵。诸如此类的还有人工神经网络、模糊系统等。

人类是自然界最具智慧的生物,自从地球上出现人类以后,世界发生了前所未有的变化。同其他的生命体相比,人类最重要的工具和优势是人类的大脑,体现在人类具有区别于其他普通生命体的思维能力。

思维即应用有用信息的意识活动,是人类智能的体现,是人脑对现实事物间接的和概括的加工形式,它以内隐的或外显的动作或言语形式表现出来。思维有十分复杂的脑机制,它在脑内对客观事件的关系与联系进行多层加工,揭露事物的内在联系和本质特征,是认识的高级形式。计算机模拟人脑对信息进行编码和译码、加工、存储和提取的驱动过程的基本形式包括分析、综合、抽象、概括、比较、系统化和具体化等过程,它们既是思维的基本过程,也是智能操作的基本形式。

思维方式不是一成不变的,而是在不断的“进化”之中。人类出生时除了具有最基本的反应以外,具有很少的知识和经验,可以说只有很少或没有思维。而随着年龄的增长,通过不断的学习,积累越来越多的知识和经验,认知体系不断完善,这一过程称之为学习过程。同时,人脑还使人类具有创造的能力,这种创造