

# 软计算及其应用

温显斌 张 桦  
张 颖 权金娟 © 编著



科学出版社

[www.sciencep.com](http://www.sciencep.com)

# 软计算及其应用

温显斌 张 桦 编著  
张 颖 权金娟

科学出版社

北 京

## 内 容 简 介

软计算方法是国际上最新发展起来的数学优化方法,它在国民经济的各个领域都有广泛的应用。本书较系统地介绍了软计算及其应用方法,包括模拟退火算法、人工神经网络计算方法、遗传算法、支持向量机和模糊计算等。书中从结构上对软计算方法进行了统一地描述,并注重叙述各内容之间的相互融合,特别注意讲述这些软计算方法的实际应用,并给出了其应用实例。

本书取材新颖,反映了当前国际先进的软计算技术,并兼顾课堂教学、自学的特点。叙述深入浅出,易读易懂,可作为高等院校相关专业的研究生、本科生的教材和参考书,也可供有关学科的教师及工程技术人员参考。

### 图书在版编目(CIP)数据

---

软计算及其应用/温显斌等编著. —北京:科学出版社,2009

ISBN 978-7-03-023427-8

I. 软… II. 温… III. 电子计算机-计算方法 IV. TP301.6

中国版本图书馆 CIP 数据核字(2008)第 181895 号

---

责任编辑:余江 潘继敏 / 责任校对:朱光光

责任印制:张克忠 / 封面设计:耕者设计工作室

**科学出版社** 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

**中国科学院印刷厂** 印刷

科学出版社发行 各地新华书店经销

\*

2009 年 2 月第 一 版 开本: B5(720×1000)

2009 年 2 月第一次印刷 印张: 12 1/2

印数: 1—2 500 字数: 233 000

**定价: 31.00 元**

(如有印装质量问题,我社负责调换(科印))

## 前 言

信息科学被广泛认为是 21 世纪主导全球经济发展格局、引导社会发展进程的学科,围绕信息科学的发展而形成的信息产业也将成为 21 世纪的重要经济活动。它们包括通信技术、计算机技术、自动化技术、软件技术、感测技术、电子与微电子技术、新材料技术、生物技术等,其理论基础与支撑是数学。数学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,它们的本质还是数学。随着人们对信息科学与信息技术研究的不断深入,对信息科学中相关的数学工具的掌握和运用能力也需要不断的提高。

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、传输与处理都以数字(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融入或模仿人的智能,让机器“听”懂人类的语言、“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地推动了信息处理智能化进程的成就,软计算(Soft Computing)正是这些突出成果中的一个典型代表,它是由多个学科相互交叉和渗透的结果,得益于模糊数学、人工智能、自动控制理论等许多学科,包括通过对人类模糊思维方式的模拟而产生的模糊计算方法,通过对自然界中生物进化机制的模拟而产生的进化计算方法以及通过对动物脑神经的模拟而产生的神经计算方法等,它已逐渐成为现代数学的重要分支之一,并且正处于快速发展的阶段。

虽然软计算方法相对而言还是一门新学科,其理论、方法还有待进一步地发展和完善,但它所具有的先进性、实用性、有效性使我们相信,软计算方法不仅是科学研究的一个重要课题,而且会成为每一个工程师及科学家必须掌握的基础知识。希望本书的出版能为先进计算技术的研究和推广起到积极的作用。

当前,软计算方法的研究成果比较分散。为了推动高校及科研院所研究人员对软计算方法的了解,让初涉优化领域的人花费较少的时间和精力认识并掌握该优化技术,我们试编了本书,将软计算方法所涉及的模糊计算、遗传算法、神经计算等进行有机集成,在保证各类算法的结构统一性和系统完整性的基础上,从基本概

念、理论基础、算法流程、应用示例入手,对软计算方法进行系统的阐述,并通过介绍各类算法的实际应用及本课题组的研究成果为大家提供参考。

本书部分内容自2000年开始在天津理工大学相关专业的研究生课程中讲授。全书自成体系,共6章。第1章为绪论;第2章为模拟退火算法,介绍了模拟退火算法的产生、机理、流程及其实现技术和应用;第3章为人工神经网络,介绍了神经网络的基本概念、学习方法,针对典型网络模型,介绍了主要算法和训练步骤;第4章为遗传算法的产生、机理、流程及其实现技术和应用;第5章为支持向量机;第6章为模糊计算中的基础理论、基本方法及模糊系统建模的相关知识。

感谢科学出版社的大力帮助,感谢参与研究工作的有关师生的大力支持。此外,在本书的编写过程中,我们参阅和引用了许多国内外同行们的学术论文和著作,编者在此表示深深的感谢。

本书所述课题得到了国家自然科学基金(60872064,60375003)、天津市自然科学基金(08JCYBJC12300,08JCYBJC12200)的资助。

本书可作为高等院校相关专业的研究生、本科生的教材,也可供有关学科的研究人员参考。由于编者水平有限,不足之处,欢迎读者批评指正。

编 者

2008年9月

# 目 录

## 前言

<b>第 1 章 绪论</b> .....	1
1.1 软计算与人工智能的关系 .....	1
1.2 软计算科学的主要分支 .....	2
1.2.1 人工神经网络 .....	2
1.2.2 遗传算法 .....	4
1.2.3 模糊逻辑 .....	5
1.3 软计算的特性 .....	7
1.4 软计算研究的主要问题 .....	8
1.4.1 学习 .....	8
1.4.2 搜索 .....	10
1.4.3 推理 .....	12
<b>第 2 章 模拟退火算法</b> .....	14
2.1 概述.....	14
2.1.1 物理退火过程 .....	14
2.1.2 Metropolis 算法 .....	15
2.1.3 模拟退火算法 .....	18
2.2 模拟退火算法的收敛性分析.....	20
2.2.1 模拟退火算法的 Markov 链描述 .....	20
2.2.2 模拟退火算法的收敛性 .....	21
2.3 模拟退火算法的设计.....	25
2.3.1 初始温度 $t_0$ .....	26
2.3.2 终止温度 $t_c$ .....	26
2.3.3 Markov 链长 $L_k$ .....	27
2.3.4 控制参数的更新函数 $T(t)$ .....	28
2.4 模拟退火算法的应用.....	29
2.4.1 模拟退火算法应用的一般要求 .....	29
2.4.2 典型组合优化问题的模拟退火算法 .....	31
参考文献 .....	33

<b>第 3 章 神经网络</b> .....	35
3.1 神经网络的基本概念 .....	35
3.1.1 生物神经元模型 .....	35
3.1.2 人工神经元模型 .....	38
3.1.3 神经网络模型 .....	41
3.2 神经网络的学习方法 .....	43
3.2.1 学习机理 .....	43
3.2.2 学习方法 .....	44
3.2.3 学习规则 .....	45
3.3 前向式神经网络与算法 .....	48
3.3.1 感知器及算法 .....	48
3.3.2 BP 网络与误差反向传播算法 .....	51
3.4 反馈网络模型及其主要算法 .....	57
3.4.1 Hopfield 网络与算法 .....	57
3.4.2 Boltzmann 机网络和学习方法 .....	60
3.4.3 自组织特征映射网络和算法 .....	64
3.5 神经网络的系统设计 .....	66
3.5.1 神经网络的适用范围 .....	67
3.5.2 神经网络的设计过程和需求分析 .....	68
3.5.3 神经网络的性能评价 .....	70
3.5.4 输入数据的预处理 .....	73
3.6 神经网络的应用 .....	74
3.6.1 基于神经网络的优化计算 .....	74
3.6.2 图像边缘检测 .....	77
参考文献 .....	79
<b>第 4 章 遗传算法</b> .....	83
4.1 遗传算法的概念 .....	83
4.1.1 遗传算法的生物遗传学基础 .....	83
4.1.2 遗传算法的一般结构 .....	84
4.1.3 遗传算法的特点 .....	86
4.2 标准遗传算法的基本设计 .....	88
4.2.1 编码 .....	88
4.2.2 适应度函数 .....	90
4.2.3 遗传算法的基本操作 .....	93
4.2.4 遗传算法的终止控制设计 .....	97

---

4.3	遗传算法的模式理论	97
4.3.1	模式概念	97
4.3.2	模式定理	99
4.3.3	遗传算法有效处理的模式数量	100
4.4	遗传算法的理论与分析	100
4.4.1	遗传算法的一般收敛性理论	100
4.4.2	遗传算法的 Markov 链模型	101
4.4.3	遗传算法的收敛速度分析	103
4.4.4	遗传算法结构分析与设计	106
4.5	遗传算法的发展	107
4.5.1	改进遗传算法的一般结构	107
4.5.2	编码问题	108
4.5.3	遗传运算	110
4.5.4	控制参数	113
4.5.5	混合遗传算法	113
4.6	遗传算法的应用	114
4.6.1	巡回旅行商问题	114
4.6.2	进化神经网络	121
4.6.3	基于遗传算法的分类器系统	123
	参考文献	125
<b>第 5 章</b>	<b>支持向量机</b>	<b>128</b>
5.1	基本原理	128
5.1.1	统计学习概述	128
5.1.2	支持向量机	130
5.2	支持向量机用于多类问题	133
5.3	支持向量机用于回归	135
5.3.1	$\epsilon$ 不敏感损失回归	136
5.3.2	核岭回归	140
5.4	支持向量机的算法	141
5.5	贝叶斯方法与高斯过程	143
5.5.1	贝叶斯方法	143
5.5.2	高斯过程	145
5.6	支持向量机的应用	145
5.6.1	文本分类	145
5.6.2	图像识别	147

---

5.6.3 手写数字识别 .....	149
参考文献 .....	150
<b>第6章 模糊计算</b> .....	<b>152</b>
6.1 模糊系统概述 .....	152
6.1.1 传统数学与模糊数学 .....	152
6.1.2 不相容原理 .....	154
6.2 模糊集合与隶属度函数 .....	154
6.2.1 模糊集合与隶属度函数 .....	154
6.2.2 模糊集合的运算 .....	158
6.3 模糊关系与模糊矩阵 .....	161
6.3.1 普通关系 .....	162
6.3.2 模糊关系 .....	163
6.3.3 模糊关系的合成 .....	163
6.3.4 模糊矩阵 .....	165
6.3.5 模糊蕴含关系 .....	166
6.4 模糊逻辑与模糊推理 .....	167
6.4.1 模糊逻辑 .....	167
6.4.2 语言变量 .....	168
6.4.3 模糊推理 .....	169
6.5 模糊系统的建模 .....	180
6.5.1 模糊系统模型 .....	180
6.5.2 模糊系统模型的建立方法 .....	181
6.6 模糊系统与其他软计算的混合计算 .....	184
6.6.1 模糊神经网络 .....	184
6.6.2 模糊推理与遗传算法的结合 .....	186
参考文献 .....	188

# 第 1 章 绪 论

软计算(Soft Computing)作为一种创建计算智能系统的新颖方法,正在引起人们的关注。目前已经认识到,复杂的实际问题需要智能系统对各种不同来源的知识、技术和方法进行组合。人们期望这些智能系统在特定领域拥有像人类一样的专门知识,在变化的环境中能够调节自身将学习做得更好,并对怎样做出决策和采取行动进行解释。在解决实际计算问题时,协同地而不是互斥地采用几种计算技术通常具有优越性,所产生的系统被称为互补的混合智能系统。设计这类智能系统的精髓就是神经-模糊计算,其中神经元网络负责识别模式和按变化的环境进行自适应调节,模糊推理系统包含对人类知识进行推理和决策。这两种互补方法,连同一些非求导优化技术的集成,产生了一门新颖的学科,即软计算科学。

## 1.1 软计算与人工智能的关系

软计算是正在发展起来的一种计算方法,它与人脑相对应,具有在不确定及不精确环境中进行推理和学习的卓越能力。

软计算是利用非精确裕度和非确定裕度来获得易控性、鲁棒性以及低成本的方法集合。它不是一种单一的方法,而是由若干种计算方法构成,包括神经网络、模糊逻辑及一些非导数优化方法,如遗传算法和模拟退火。它们分别提供不同方面的能力,其中模糊逻辑主要处理非精确性和进行近似推理,神经网络使系统获得学习和适应的能力,遗传算法等进化算法则提供进行随机搜索和优化的能力。其中的每一种方法都有其长处,将这些技术紧密集成就形成了软计算的核心;通过协作,可以保证使计算有效利用人类知识、处理不精确及不确定的情况,对未知或变化的环境进行学习和调节以提高性能。针对学习和自适应,软计算需要强化计算。在这个意义下,软计算与计算智能有共同的特征。

通常软计算中不进行很多符号运算,这就是我们将它作为区别于传统人工智能的新学科的原因。人类在推理和作结论时通常使用自然语言。传统人工智能致力于以语言或符号规则的形式来表达和模拟人类的智能行为。在行为可以存储在符号结构化数据库的假设下,传统人工智能是以符号运算为基础的,这就是所谓的物理符号系统假设。在某些很窄的问题领域中,如果存在明显的知识,那么符号系统为人类专家建模提供了很好的基础。也许最成功的传统人工智能成果就是基于知识的系统或专家系统。

传统人工智能的文献反映了有关智能系统的早期工作。许多人工智能先驱按照自己的哲学对人工智能做出了定义,以下列出了一些有代表性的人工智能定义,以及专家系统定义:

- ① 人工智能是对存在于环境、感知和行为中代理体的研究。
- ② 人工智能是使计算机做聪明事情的艺术。
- ③ 人工智能是一种编程风格,为了完成目标,它按照规则对数据进行操作。
- ④ 人工智能是为诸如计算机等机器提供的活动,以显示出在人类看来是智能的行为。
- ⑤ 专家系统是计算机程序,它利用专家知识,力图在某个狭窄的问题领域获得高水平的性能。
- ⑥ 专家系统是对人类专家的笨拙模仿。

这些定义勾勒出了一个明显的人工智能框架,尽管由于概念框架快速变化,这些定义可能是很暂时的。读者可能会疑问,“难道人工智能已经过时了吗?”

称软计算的成分为“现代智能部分”不可避免地依赖于个人的判断。的确,目前许多关于现代人工智能的书籍中描写了神经网络以及可能的其他软计算组成,这说明人工智能领域是在稳步地发展,人工智能与软计算的区分边界并不明显。显然,下一代人工智能方法将更复杂,关于人工智能边界划分的讨论超出了本书的范围。

实际中,由于知识的获取和表示是复杂而艰巨的任务,符号运算限制了传统人工智能理论可以应用的领域。更多的注意力已经转向从生物界获得灵感的方法,例如,脑建模、进化算法和免疫建模,它们模仿了负责产生自然智能的生物机制。这些方法与传统人工智能方法有某种正交关系,通常它们可以弥补符号机制的缺点。

人工智能研究的长期目标是创建和理解机器智能。从这个角度看来,软计算与人工智能享有共同的最终目标。从某种意义上讲,软计算是在人工智能对控制论(研究人类和机器中的信息与控制)影响下的进化。

## 1.2 软计算科学的主要分支

### 1.2.1 人工神经网络

神经网络的研究可以追溯到 1890 年 James 的《心理学》,该书认为大脑皮层每一点的活力产生于其他点势能释放的综合效能,即其他点的兴奋次数、强度和所接受的能量。

1943 年,心理学家 McCulloch 与数学家 Pitts 对神经元进行了形式化的研究,

提出了神经元的数学模型(MP模型)。随后,1944年由Hebb提出了改变神经元连接强度的Hebb规则,即Hebb学习规则,这个规则至今仍是人工神经网络学习算法的一个基本原则。1957年,Rosenblatt首次引进感知器概念以图模拟生物的感知学习能力。1962年,Widrow提出自适应线性元素,它是用于自适应系统的连续取值的线性网络。这是人工神经网络发展的第一阶段。由于人工神经网络的思想完全不同于以顺序离散符号推理为基本特征的人工智能,因而引起了许多人的兴趣和争论。

1969年,人工智能创始人当中的Minsky和Paper发表了《认知论》一书,指出单隐含层的感知器神经网络模型无法解决最为经典的“异或(XOR)”问题。由于Minsky在人工智能领域中的巨大影响,加之数值计算机的突飞猛进和人工神经网络自身的历史性限制,人工神经网络理论的发展在20世纪70年代处于一个低潮。这是人工神经网络发展的第二个阶段。而难能可贵的是,仍有不少学者致力于这一研究,如Webos在1974年提出BP学习理论。美国Boston大学的S. Grossberg提出自适应共振理论的研究等,从而为人工神经网络的进一步发展奠定了必要的理论基础。事实上,70年代末到80年代初,人工智能在模拟视觉和听觉上遇到了极大的困难,而且计算机的集成度也有不可逾越的线路微型化的物理极限。人们需要从新的角度去考虑智能计算的发展道路。

1982年,美国加州工学院物理学家J. J. Hopfield提出了Hopfield神经网络系统,提出了“能量函数”的概念,并将一个组合优化问题与动力系统的稳定吸引子相联系,在旅行商问题(TSP)的计算上取得了进展,开拓了人工神经网络用于联想记忆和优化计算的新途径。这时,人工神经网络理论、技术和应用的发展进入了第三个阶段,并于20世纪80年代末和90年代初掀起了一个不小的热潮。另外,McClelland和Rumelhart于1988年利用多层反馈学习算法成功地解决了单层感知器不能解决的“异或(XOR)”问题。

当然,人工神经网络根本无法同人体神经系统相比,而且人体神经系统的结构、功能、行为规律至今仍不是很清楚。人体神经系统可分为“中枢神经系统”和“周围神经系统”。前者包括“脑”和“脊髓”两部分;后者包括“体干神经系统”和“自主神经系统”。脑可分为前脑(含大脑、丘脑及丘脑下部),中脑(含上丘、下丘)和后脑(含延髓、脑桥及小脑),分别管理人体的各种信息、各种器官活动及感觉,同时具有各种信息的存储、提取、分析与综合的功能,使人具备知觉、辨别、学习和思维的行为特征。脊髓位于脊柱内,是脑和周围神经系统的信息传递通道。体干神经系统遍及人体外表各部位,其中传入神经是感觉神经,而传出神经可控制模纹肌运动。自主神经系统是按不同生理功能独立管理人体内脏器官的系统,不受意志的指挥。人的神经系统由这些不同功能的子系统组成,是一个具有自学习、自适应和自组织特征的复杂巨系统。

人工神经网络仅是神经元联结形式的数学抽象,即有向图。目前人工神经网络的研究主要集中在神经网络的基本理论(包括人工神经网络模型、机制、性能、能力及相关算法),神经网络智能信息处理应用系统(包括信号处理、模式识别、优化控制及机器智能),神经网络计算机实现等方面。

### 1.2.2 遗传算法

遗传算法(Genetic Algorithm,GA)是进化计算的基本分支之一;进化计算还包括进化规则、进化策略,以及20世纪90年代初期在遗传算法基础上发展起来的遗传程序设计。遗传算法是模拟生命演化与进化的过程,它认为生命的自然进化过程本质上是一个学习和优化的过程,这个过程的目的是使生命体达到适应环境的最佳结构与效果。自然进化中“物竞天择,适者生存”的规律是达尔文发现的。遗传算法正是在客观意义上模拟达尔文“优胜劣汰、适者生存”的原理以激励良态的结构,又通过模拟孟德尔遗传变异理论推动良态结构的进化,其数学本质仍然是一种随机的优化与搜索算法。

种瓜得瓜,种豆得豆,这种从亲代继承特征的生命现象称为遗传。

遗传信息由基因组成,生物的各种性状特征由其相同的基因决定。承载基因的物质载体是细胞中一种称为染色体的微小丝状化合物。而控制并决定生物遗传性状的染色体主要由脱氧核糖核酸(DNA)构成,低等生物中染色体则由核糖核酸(RNA)构成。DNA在染色体中有规则地排列成双螺旋结构。基因就是DNA或RNA长链中占有一定位置的基本遗传单位。遗传基因在染色体中所占据的位置称为基因座。同一基因座可能有的全部基因称为等位基因,而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。

个体的遗传在有性繁殖中,两个同源染色体之间通过交叉而重组,也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连接而形成两个新的染色体。在细胞分裂时,遗传物质DNA通过复制而转移到新产生的细胞中,同时也可能因小概率事件使某些基因座的基础产生变异而生成新的染色体。个体,即单个生物,对其生存环境都有一个适应能力,称为个体的适应度,而不同的个体可能有不同的适应度。通常,生物的进化是以群体的形式共同进行的,通过群体中不同适应度的个体的繁殖使有利于生存环境的基因逐渐增多,从而产生越来越适应环境的物种。这是设计遗传仿生算法的生物学基础。

遗传算法的操作对象是一组可行解组成的群体,其中每一个个体表示一个编码字符串,使算法表现出良好的并行性和全局优化性。目标函数解释为个体编码的适应值,从而只需利用目标的取值信息而无须利用梯度等高价值信息,使算法适用于大规模、高度非线性的不连续多峰函数的优化以及无解析表达式的目标函数的优化,具有较强的通用性。

“遗传算法”这一术语首次出现于 J. D. Bagley 1967 年的论文中,并讨论了遗传算法在自动博弈中的应用。同一时期,从生物学层面注意生物遗传模拟的,还有 Rosenberg、Weinberg 等。第一个将遗传算法用于函数优化的是 Hollstien,他在 1971 年的论文《计算机控制系统中的人工遗传自适应方法》中提出了遗传算法用于数字反馈控制的方法,讨论了二变量函数的优化问题。1975 年是遗传算法研究史具有里程碑意义的一年,在这一年, J. H. Holland 出版了历史性的著名专著《自然系统和人工系统的适配》,阐述了遗传算法的基本理论和方法,提出了著名的模式理论。同年, K. A. De Jong 在其论文《遗传自适应系统的行为分析》中,将 Holland 模式理论与计算实验结合起来,为算法的实现和应用打下了坚实的基础。通过 20 年的努力,终于在 20 世纪 80 年代迎来了遗传算法的兴盛发展时期。

20 世纪 80 年代,奠定遗传算法的理论基础主要有两个:一个是 Holland 早期的 Schema 理论;另一个是 Goldberg 早期的种群马氏链模型。Schema 理论属于遗传算法的几何理论。其方法比较直观易懂,本书是从这一角度介绍遗传算法的思想和技术。这一理论提出了一个至今尚未解决的建筑块假设,在这一假设成立与否的问题上,目前的工作主要集中在欺骗函数的研究上。所谓欺骗函数是指对遗传算法进行误导,使其收敛到非全局最优解状态的函数,研究欺骗函数的主要方法是沃尔什(Walsh)变换。遗传算法的马氏链模型主要包括种群马氏链模型、Vose 模型和 Cerf 扰动马氏链模型。早期的研究主要是通过马氏链的极限理论来分析遗传算法的收敛性,近年来又有徐宗本教授采用软方法与迭代法得到了比较系统的结果。

在算法实现和技术应用上,近年来在编码方法、选择机制、杂交机制、变异机制的设计方面都有重要的进展。尤其是在遗传算法的执行策略改进方向上,提出了遗传算法与模拟退火算法相结合、遗传算法与局部优化算法相结合的改进方法。值得进一步探讨研究的技术还有并行遗传算法、共存演化遗传算法和混沌遗传算法等。

遗传算法和人工神经网络同样具有学习、分类和优化计算的功能,但二者的区别与互补是明显的。神经网络的学习包含了两个优化过程,一个是网络连接权重的优化学习,另一个是网络拓扑结构的优化学习。前者主要是基于梯度下降的误差反向传播律法,即 BP 算法;该算法无法学习网络的拓扑结构。遗传算法用于人工神经网络,原则上可用于连接权重的学习、网络拓扑结构的学习和网络学习规则的学习。近年来,基于演化计算的神经网络系统,即演化式神经网络系统,已成为一种自适应学习模型而受到关注和研究。

### 1.2.3 模糊逻辑

逻辑是研究关于推理过程规律的科学,而数理逻辑则是用数学方法,即符号体

系的形式化方法,来研究推理规律的一个数学分支,通常称为数学基础。早期的逻辑代数研究起始于 Leibniz(1646—1716),将命题形式表达为符号公式,建立了科学史上最早的逻辑演算。到 1847 年,自学成才的数学家 Boole 发表了标志 Boole 代数成熟的名著《逻辑的数学分析》,奠定了二值逻辑的基础。后来,耶拿(Jana)大学的数学教授 Gottlob Frege(1848—1925),又在逻辑演算中引进量词、变量和命题函数等概念,使数理逻辑具备了完整的表达能力。在 20 世纪 30 年代以后,渐渐形成了以证明论、模型论、递归函数论和公理集合论为核心的现代经典数理逻辑,其共同基础为二值逻辑的命题演算和谓词演算,即“四论两演算”。

在二值逻辑中,一个可以分辨真假的句子称为命题,一个命题的真或假叫做命题的真值。命题为真,其真值为 1;命题为假,其真值为 0。因此,命题的逻辑值域为  $\{0,1\}$ ,即命题非真即假,非假即真,二者必居其一。但是,在现实生活中,不能以绝对的真或假来分辨的句子比比皆是。例如,“今天天气有点凉”,这类句子具有模糊性,称之为模糊命题。用经典的二值逻辑处理模糊命题,会遇到难以克服的矛盾。因此,对于模糊命题,仅用 1 和 0 两个逻辑值是不够的,必须在 1 和 0 之间采用多个逻辑值来表示不同的真值。从而出现了多值逻辑,其中特别称逻辑值域为  $[0,1]$  闭区间的连续值逻辑为模糊逻辑。

模糊逻辑的集合论基础是模糊集,又称 Fuzzy 集。自 1955 年 L. A. Zadeh 提出 Fuzzy 集概念以来,关于 Fuzzy 系统的研究便迅速发展起来。1979 年,捷克学者 Pavelka 发表了以“On Fuzzy Logic”为题的三篇文章,为 Fuzzy 命题演算提供了一种比较完整的理论框架,这是 Fuzzy 逻辑方面的奠基性工作。随后不久,出现了关于 Fuzzy 逻辑的热烈讨论,也出现了一些命名中含“Fuzzy Logic”的专著。

作为经典逻辑的推广,Fuzzy 逻辑的历史演变大致在两个不同的层次上进行。在第一个层次上,把二值逻辑中的赋值从值域  $\{0,1\}$  扩充为连续值域  $[0,1]$ 。这时,Fuzzy 逻辑常与 Fuzzy 集的隶属度相关联,并可用于对后者的研究。在第二个层次上,是把公理系统中的公式赋予某种真度,同时推理规则也被程序化了,但是这种推广并不是经典逻辑的简单重复。这方面早期的工作有如 Pavelka 的 Fuzzy 逻辑系统,Cerla 在抽象格上的 Fuzzy 逻辑系统等。随着 Fuzzy 逻辑理论研究的逐步深入发展,Fuzzy 逻辑的应用研究与技术也取得了充分的成功,1974 年,英国科学家 Mamdani 和 Assilian 首次提出模糊逻辑控制理论,并成功地应用于热电厂的蒸汽机控制;1976 年,Mamdani 又将该理论成功地应用于水泥旋转窑的控制。从此,模糊逻辑控制理论与技术作为优化与控制的一门新分支得以快速的发展。现在,各种模糊推理集成块与模糊控制器、模糊控制系统已经成为技术市场上的商品;反过来,技术市场和需求又推动了模糊控制理论与技术的发展。

Fuzzy 推理是 Fuzzy 逻辑的表现形式和 Fuzzy 控制技术的理论基础。而 Fuzzy 语言是 Fuzzy 推理的形式化工具,也是人的某些思维方式的形式化载体。

对于推理模型, Zadeh 于 1973 年首次提出解决这个问题的 CRI(Compositionnal Rule of Inference)方案,随后,王国俊教授于 1999 年提出了另一种基于逻辑含义考虑的推理方法,即 3I 方案,改善了 Fuzzy 推理与 Fuzzy 逻辑的结合问题。

### 1.3 软计算的特性

软计算工作特性可以归纳如下:

① 人类的专门知识。软计算以模糊推理规则形式及传统的知识表示形式来使用人类的专门知识,解决实际问题。

② 受感于生物的计算模型。受生物神经网络的激励,软计算中大量使用人工神经网络来处理有关感知、模式识别、非线性回归与分类的问题。

③ 新的优化技术。软计算中应用了来源于不同思想的新颖的最优化方法,包括遗传算法(受进化与选择过程的启发)、模拟退火(受热力学启发)、随机搜索方法以及下山单纯型法。这些优化方法不需要目标函数的梯度向量,因此在解决复杂优化问题时有更大的灵活性。

④ 数值计算。与符号人工智能不同,软计算主要依赖于数值计算。如何在软计算中引入符号技术,将是这一领域中一个活跃的研究方向。

⑤ 新的应用领域。由于其数值计算特性,软计算可以找到比传统人工智能方法更广阔的新的应用领域。这些应用领域大多需要强化计算,包括自适应信号处理、自适应控制、非线性系统辨识、非线性回归和模式识别。

⑥ 无模型学习。神经网络和模糊推理系统能够只利用目标系统的采样数据来创建模型。对目标系统的深入了解可以有助于建立初始的模型结构,但并不是缺它不可的。

⑦ 强化计算。神经-模糊和计算不需要很多待求解问题的背景认识,而主要依赖于大量快速的运算从数据集中寻找规则或规律。这是计算智能领域的普遍特征。

⑧ 容错性。神经网络和模糊推理系统都有好的容错性。从神经网络中删除一个神经元,或是从模糊推理系统中去掉一条规则,并不会破坏整个系统。相反,由于具有并行和冗余的结构,系统可以继续工作,尽管性能质量会有所下降。

⑨ 目标驱动特性。软计算是目标驱动的,即只要从长远看,我们是在向目标移动,至于从当前状态到最终解走的是一条什么路径则是无关紧要的。这点对于诸如遗传算法、模拟退火和随机搜索方法等非导数优化方案都是成立的。特定的领域知识有助于降低计算量和搜索时间,但并不是必需的。

⑩ 实际应用。大多数实际问题的规模很大,而且不可避免地存在不确定性,这就妨碍了待求解问题的传统方法的应用。软计算是一种集成方法,对于各子任

务可以采用特定的不同技术,以便对实际问题找到满意解。

软计算科学与技术正在快速的发展,各种新技术和新应用不断涌现。通过全世界各个不同科学学者的共同努力,可以说软计算已经建立了坚实的基础,进一步发展的动力是为即将到来的美好明天创建高度自动化的智能机器。

## 1.4 软计算研究的主要问题

### 1.4.1 学习

学习是一个有特定目的的知识获取过程,并且通过这一过程逐渐形成、修改新的知识结构或改善行为性能,包括积累经验、发现规律、改进性能和适应环境。机器学习则是利用机器来完成这一过程,从而达到或部分达到学习的目的。

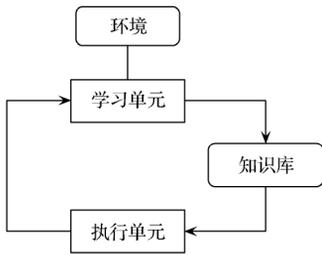


图 1-1 学习模型 I

人工智能的机器学习模型 I 如图 1-1 所示,其中包含学习系统的四个基本环节,环境提供外界信息;学习单元处理外部环境提供的信息;知识库以某种知识的表示形式来储存信息;执行单元则是利用知识库中的知识来完成某种任务,并将结果反馈至学习单元,以改善系统的性能。这种模型的特征是倾向于使用符号表示而不是数值表示,常使用启发式方法而不是数值计算方法。在实际的学习方法设计中,主要有归纳学习、类比学习和解释学习等方法。

#### 1. 归纳学习

归纳学习是从给定的关于某个概念的一系列已知的正例和反例中归纳出一个一般的概念描述,而归纳推理是从有限的、不完全的知识状态推出完全的知识状态。归纳学习可以分为有教师的学习和无教师的学习两类。前者,如实例学习,又叫概念获取,其目的是确定某一个概念的一般描述且该描述应能解释所有给定的正例并排除所有给定的反例;后者,如观察与发现学习,其目的是在无导师条件下产生解释所有或大多数观察到的规律的规则,包括概念聚类、构造分类等。与演绎推理相比归纳推理的开始前提是具体事实而不是一般公理,而推理目标是形式化解释事实的一般断言或预见新事实,这是一种非单调推理。

因此,归纳学习的一般模式是:对给定的观察语句集  $F$ , 寻求归纳断言  $H$  使得  $H$  重言或弱蕴涵观察语句集  $F$ , 并满足背景知识。所谓  $H$  弱蕴涵  $F$ , 是指事实  $F$  是  $H$  合理的或部分的结论。

归纳学习的一般操作是范化和特化,前者用于扩展某一假设的语义使其包含更多的正例;后者用于限制概念描述的应用范围,是泛化的反操作。用程序设计语言描述上述操作过程是归纳学习程序,而执行归纳学习程序完成特定归纳学习任务的系统是归纳学习系统。

## 2. 类比学习

类比学习是把两个或两类事物进行比较,找出它们在某一抽象层面上的相似性或相似关系,并以此为据把一事物的有关知识对应到另一事物,从而获得求解另一事物的知识,即是一种允许知识在具有相似性质的领域中进行转换的策略。类比的形式是多样性的、跨越式的,如概念类比、形式类比、方法类比等,其关键在于相似性的定义和度量,类比学习需要先验的领域知识,因而是一种深层次的非孤立的学习行为。同样,类比推理是根据已知域的知识,用类比来回答另一未知域的问题,即当类似的前提成立时类似的结论是否成立的推理过程。

同归纳学习相比,类比学习更缺少灵活性,在知识表示和知识库的快速检索上也存在更大的困难。

## 3. 解释学习

解释学习是在领域知识的指导下,通过对单个问题求解实例的分析,构造出求解过程的因果解释结构和泛化控制规则,以便指导类似问题的求解。

软计算的机器学习模型Ⅱ如图 1-2 所示,其中数据(随机数据)  $X$  输入学习系统和实际系统,使二者分别为输出  $Y'$  和  $Y$ 。同时实际系统的输出  $Y$  也反馈至学习系统,目的是使学习系统在某个函数集  $\Xi$  上选取某个函数  $f \in \Xi$ ,使得  $Y' = f(X)$  且  $Y'$  和  $Y$  的距离  $d(Y, Y')$  为最小,这是一个有导师学习模型,例如,分类;同样,可设计无导师学习模型,例如,聚类,这种模型的特征是由海量数据驱动,通过模型算法使学习过程成为一个算子作用下的动力系统,而学习目标则是该动力系统长时间作用下的形态,如吸引子等。在许多实际问题中,这类模型往往是随机模型,即该动力系统是随机动力系统。显然,软计算的机器学习模型完全不同于归纳、类比和解释等方法,而是基于数据的学习算法,属于数据挖掘的范畴。

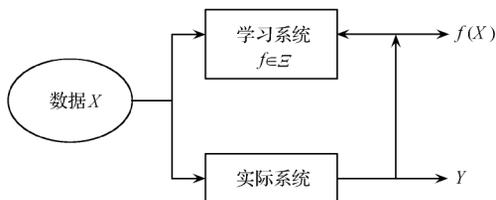


图 1-2 学习模型 II

数据挖掘是从数据中探知结构,从结构中获取知识的知识发现过程。由于计算机网络和信息技术迅速发展,产生大量分布异构的数据存储系统,为数据挖掘提供了丰富的数据资源,从而大大地推动了数据挖掘理论与技术的发展。

### 1.4.2 搜索

搜索是对问题的一种求解方法、技术和过程。因而,搜索是面向问题的,对于不同的问题,有不同的搜索方法、技术和过程。如果将问题域的解集构造成一个解空间,那么搜索就是在解空间中找到满足问题条件的目标集或解集。这里,关注的是“找到”,而不仅仅是“证明”其存在性,故搜索方法、技术和过程是构造性的。如果问题域的解集是“新知识”或“新概念”,那么搜索过程就是学习;如果问题域的解集是“新结论”,那么搜索过程就是推理;如果问题域的解集是“最优解”,那么搜索过程就是优化。

一般性的机器搜索模型如图 1-3 所示。对于问题域的一般性描述,应该表示成适合于机器搜索的模型形式,例如,数学模型、逻辑模型、知识模型。对于模型的生成和存储,往往要借助于模型库中的“模板”。其中参数的确定就是模式识别要解决的问题。生成模型需要一定的方法,求解模型也需要一定的方法,这就需要完成模型与方法匹配,形成一个新的适合于机器搜索的搜索模型。模型的识别、匹配,以及方法的生成,需要在知识库和学习机的支持下才能进行,而这一过程的进行,又需要通过控制器的管理,在机器搜索模型原理的指导下,产生人工智能中通常采用的三库结构,即模型库、方法库和知识库。

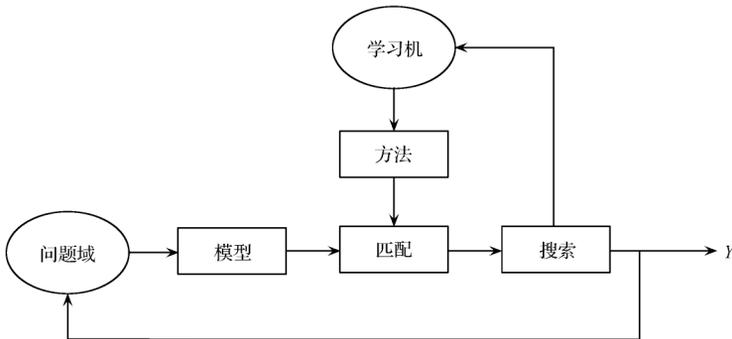


图 1-3 一般性的机器搜索模型

#### 1. 面向数学模型的搜索

数学模型的解空间从维数上可以分为两类,即有穷维解空间和无穷维解空间,在无穷维空间中的搜索,要么利用可求解的解析方法,要么利用有穷维空间的数值

逼近方法转化为有穷维空间中的搜索问题,有穷维空间中的有界解集如果只有有限个点,可采用有限遍历式的盲目搜索,即不运用知识信息的穷举法。盲目搜索包括宽度优先搜索、深度有限搜索和代价搜索等算法。但是,盲目搜索的效率太低,代价太大。于是,又有运用启发式信息的启发式搜索法,即运用某种准则、经验、问题特征使搜索沿着某个被认为最有希望达到目标的前沿区段扩展。这些准则、经验、特征可以来源于数学模型本身,也可以来源于数学模型所处问题域的知识背景。对于前者,例如,无约束优化模型中的负梯度方向为最速下降方向的极小化算法;对于后者,例如,人工智能中的估价函数法、最好优先搜索法、双向搜索法和问题归纳技术等。

## 2. 面向专家系统的搜索

专家系统是在生产式系统的基础上发展起来的一种模拟人类专家解决领域问题的计算机程序系统,即是内部含有大量的某个领域专家水平的知识和经验,能够利用人类专家的知识 and 解决问题的方法,借助人工智能技术和计算机技术,进行推理和判断,模拟专家的决策过程,以解决那些需要人类专家处理的复杂问题。专家系统的结构主要由知识库、推理机和控制器组成。专家系统关键是表达和运用专家知识,即来自人类专家的,并已经被证明对解决有关领域内的典型问题是有效的方法和过程,其特点是所要解决的问题一般没有算法解,并且经常要在不完全、不精确或不确定的信息基础上做出结论。因而,基于知识表示和非经典推理的高级求解技术,是建立专家系统的基础。

## 3. 面向机器人的自动规划

规划是优化搜索,意味着在行动之前决定行动的进程。面向机器人的自动规划是面向真实物理世界而不是抽象数学模型的一种重要的问题求解技术,而机器人规划与问题求解是自动规划的典型模型,因而,有时亦称自动规划为机器人规划,这种技术是从某个特定的问题状态出发,寻找一系列行动动作,并建立一个操作序列,直到求得目标状态。与一般问题求解相比,自动规划更侧重于问题的求解过程,而不是求解结构,即注重建立一个动作序列,并通过这个序列使机器人达到预期的工作目标,完成规定的任务。在机器人高层规划系统中,广泛采用了启发式搜索技术、基于一阶谓词演算公式集的消解反演推理技术、管理式学习技术、语义网络技术、分层规划技术与非线性规划等。

这些典型的人工智能搜索技术,在实际应用中很难处理因“组合爆炸”引起的搜索规模问题和复杂性很高的计算问题,例如,背包问题、旅行商问题。一个判定问题  $A$ , 即是其每一个实例只有“是”或“否”两个答案的问题,成为 NP 问题,如果对每一个 NP 问题,即非多项式确定的问题,均可在多项式时间内归约为该判定问

题  $A$ , 特别当  $A \in NP$  时, 判定问题  $A$  称为完成  $NP(NPC)$  问题。当前, 人们公认  $NP$  问题是极难在有效时间内完成计算问题的。

计算智能的一个目标, 正是针对  $NP$  问题的模型, 研究完全并行性的有效算法。这种算法具有启发式的特征。启发式算法是一种在可接受的计算时间内去搜索最好的解, 但是不一定能保证所得解的最优性, 甚至在许多情况下, 无法说明所得解同最优解的近似程度。不过, 以计算智能为基础的现代搜索技术, 仍以搜索全局最优解为目标。

### 1.4.3 推理

推理是人类基于逻辑的一种思维形式, 也是计算机基于知识表示的一种知识利用, 即根据一定的规则, 从已知的断言或知识得出另一个新的断言式知识的过程。推理所依据的断言称为前提或前置断言。由前提或前置断言推出的断言称为结论或后置断言。知识表示和机器推理构成了计算机“智能系统”或“智能行为”的基础。

以严格的经典逻辑为基础的推理是一种演绎推理。演绎推理建立在清晰而无二义性的概念之上, 使用抽象而严格的定义以及在一定理论框架下绝对正确的定理与公式。因而, 演绎推理是一种确定型的“保真”推理, 即如果前提是正确的, 那么由前提经演绎推出两个相反的命题; 同时, 还具有单调性, 即增加新的事例不会影响原有结论。这是一种从普遍到特殊的推理, 在其推理过程中不承认任何由已知前提推不出来的断言, 也不承认任何不经过演绎的假设和含有例外的结论。1965年, J. A. Robinson 根据 Herbrand 定理提出一阶谓词演算中的归结原理, 使演绎推理机械化达到了使用的程度。从而使后来许多基于逻辑的智能系统一般都使用一阶逻辑或者它的一些扩张形式, 例如, 引进各种认知概念的模态逻辑, 引进时间概念的时序逻辑等。

但是, 人类对客观世界的认识不仅是不确定的, 而且往往是不望权的, 因而人类知识的增长过程并不一定遵守演义推理的单调性, 而是非单调的发展过程。非单调逻辑是一种典型的非经典逻辑, 它的特征是推理系统的定理集合并不随推理过程的进行而单调增大, 新推出的定理很可能会否定、改变原来的一些结论, 使得原来能够解释的现象变得不能解释了。非单调逻辑始建立于 20 世纪 70 年代末。现在, 人工智能界在非单调推理的研究中, 有代表性的理论和系统是: 赖特的缺省理论、麦克德莫特的非单调逻辑、McCarthy 的界限理论和多伊尔的正确性维持系统。

非单调谓词逻辑存在推理效率低, 甚至不可计算的缺点。于是, 人工智能界在 20 世纪 80 年代又展开了约束推理、定性推理和范例推理等高级求解技术的研究。约束是包含若干变量的关系表达式, 而约束推理是在一个满足约束的域上缩小搜索空间的一种推理形式。定性推理是从物理系统、生命系统的结构描述出发, 导出

系统的行为描述,以便预测系统的行为特征并给出相应的解释。而范例推理则是从目标范例(要解决的问题)的提示中获得记忆中的源范例(记忆中的情况),并由源范例来知道目标范例求解的一种策略。

推理的不确定性,来源于以下两个方面:

① 结构和规则是明确的,只是事件的发生是随机的;

② 结构和规则本身就是不明确的。

对于前者,在处理不精确信息推理方面,20世纪80年代展开了“概率逻辑”的研究;对于后者,从关注人类语言的模糊性出发而导入 Fuzzy 推理。

人类语言的模糊性是人类模糊性思维的表现形式和模糊推理的工具。模糊推理,又称 Fuzzy 推理,是基于 Fuzzy 集理论的一种不确定性推理。Fuzzy 逻辑的研究,始于20世纪60年代,而至1973年 Zadeh 提出解决简单 Fuzzy 推理的 CRI 方法后,这种推理技术广泛应用于工业控制。Fuzzy 推理是以模糊判断为前提,使用模糊推理规则,推出模糊判断结论的推理方式,尤其注重模糊命题演算,是软计算在逻辑领域内研究的重要分支。Fuzzy 推理有不同的形式,例如,Fuzzy 假言推理、Fuzzy 关系推理、Fuzzy 混合关系推理等。Fuzzy 假言推理的小前提与大前提的前件由同一语言度量联系,结论与大前提的后件由同一语言变量联系,前提与结论间存在 Fuzzy 的逻辑联系。例如,

大前提(规则)	若 $x$ 大	则 $y$ 小
小前提(已知证据)	$x$ 很大	
结论	$y$ 几乎很小	

Fuzzy 关系推理是大前提、小前提及结论都是由 Fuzzy 关系判断的推理,这种推理的规则必然具有模糊性。例如,

大前提(规则)	$x$ 和 $y$ 近似相等
小前提(已知证据)	$y$ 比 $z$ 大得多
结论	$x$ 比 $z$ 大约大得多

Fuzzy 混合关系推理是前提中包含一个 Fuzzy 关系判断和一个 Fuzzy 性质判断的推理。例如,

大前提(规则)	$x$ 和 $y$ 近似相等
小前提(已知证据)	$x$ 很大
结论	$y$ 几乎很大

## 第 2 章 模拟退火算法

### 2.1 概 述

1982 年 Kirkpatrick 等将退火思想引入组合优化领域,提出了一种解大规模组合优化问题的有效近似算法——模拟退火算法(Simulated Annealing Algorithm, SA)。它源于对固体退火过程的模拟,采用 Metropolis 接受准则,并用一组称为冷却进度表的参数控制算法进程,使算法在多项式时间里给出一个近似最优解。

固体退火过程的物理现象和统计性质是模拟退火算法的物理背景;Metropolis 接受准则使算法跳离局部最优的“陷阱”;而冷却进度表的合理选择是算法应用的前提。

#### 2.1.1 物理退火过程

物理退火是先将固体加热熔化,再徐徐冷却使之凝固成规整晶体的热力学过程,属于热力学与统计物理研究的范畴。

热力学与统计物理所研究的对象通常称为热力学系统,是指在给定范围内,由大量微观粒子所组成的宏观物体,如气体、液体、固体、等离子体等。对同一研究对象,热力学从经验总结的定律出发,找出系统宏观量之间的联系以及宏观量变换的规律;统计物理学从物质的微观结构出发,把宏观量作为相应微观量的统计平均值来计算,可以从理论上计算某些宏观量及其涨落,因此更能反映热运动的本质。

固体加热时,固体粒子的热运动不断增强,随着温度的升高,粒子与其平衡位置的偏离越来越大。当温度升至熔解温度后,固体的规则性被彻底破坏,固体熔解为液体,粒子排列从较有序的结晶态变为无序的液态,这个过程称为熔解。熔解过程的目的是消除系统中原先可能存在的非均匀状态,使随后进行的冷却过程以某一平衡点为始点。熔解的过程与系统的熵增过程相联系,系统能量也随温度升高而增大。冷却时,液体粒子的热运动渐渐减弱,随着温度的徐徐降低,粒子运动渐趋有序。当温度降至结晶温度后,粒子运动变为围绕晶体格点的微小振动,液体凝固成固体的晶体,这个过程称为退火。退火过程之所以必须“徐徐”进行,是为了使系统在每一温度下都达到平衡态,最终达到固体的基态。退火过程中系统的熵值不断减小,系统能量也随着温度降低趋于最小值。冷却时若急剧降低温度,则将引

起淬火效应,即固体只能冷凝为非均匀的亚稳态,系统能量也不会达到最小值。退火过程中系统在每一温度下达到平衡态的过程,可以用封闭系统的等温过程来描述,根据 Boltzmann 有序性原理,退火过程遵循应用于热平衡封闭系统的热力学定律——自由能减少定律,即:

“对于与周围环境交换热量而温度保持不变的封闭系统,系统状态的自发变化总是朝着自由能减少的方向进行,当自由能达到最小值时,系统达到平衡态”。

系统的自由能  $F = E - TS$ , 其中,  $E$  是系统的内能,  $T$  是系统温度,  $S$  是系统的熵, 设  $i$  和  $j$  是恒温系统的两个状态, 即

$$F_i = E_i - TS_i, \quad F_j = E_j - TS_j \quad (2-1)$$

而

$$\Delta F = F_j - F_i = (E_j - E_i) - T(S_j - S_i) = \Delta E - T\Delta S \quad (2-2)$$

若系统状态由  $i$  自发变化到  $j$ , 则应有  $\Delta F < 0$ 。显然, 能量减少 ( $\Delta E < 0$ ) 与熵增加 ( $\Delta S > 0$ ) 有利于自发变换。因此, 任一恒定温度下, 系统状态从非平衡态自发变化到平衡态, 都是能量和熵竞争的结果, 温度决定着这两个因素的相对权重。在高温下, 熵占统治地位, 有利于变换的方向就是熵增加的方向, 因此显出粒子的无序状态; 而低温对应于低熵, 低温下能量占优势, 能量减少的方向有利于自发的变化, 因而得到有序和低能的晶体结构。在这种结构内部, 每个粒子都被它与相临粒子间相互作用的势能“囚禁”着。

### 2.1.2 Metropolis 算法

#### 1. 统计物理的基本假设

##### 1) 系统的宏观状态与微观状态

系统的宏观状态是指系统的热力学状态。处于平衡态的系统, 用体积、温度、压强、总能量等宏观量来描述。系统的微观状态是指系统的动力学状态。若系统由  $N$  个粒子组成, 每个粒子的自由度为  $s$ , 则系统的微观状态可用  $Ns$  个广义坐标与  $Ns$  个广义动量组成的  $2Ns$  维的相空间来描述, 相空间中的一点代表系统的一个微观状态。

平稳态是系统的一种宏观状态, 可以对应各种不同的微观状态, 系统平稳时, 宏观状态已经确定, 但是组成系统的粒子的动力学状态却在不断变化, 使得系统的微观状态不断变化, 而这一系列不同的微观状态属于同一宏观状态。例如, 系统第  $i$  个粒子的状态与第  $j$  个粒子的状态发生交换。对于系统来说, 这是两种不同的微观状态, 在相空间中以不同的代表点来代表, 但是它们又属于同一宏观状态。

##### 2) 系综

统计物理学首先要解决的是: 具有一定宏观状态的系统, 处于某一微观状态或

某一微观状态附近的概率是多少?

Gibbs 引进的系综概念可以形象地表示系统处于某一微观状态或某一微观状态附近的概率。系综是大量系统的集合,这些系统具有相同的粒子数,相同的化学性质和相同的宏观状态,其微观状态则按照各自的统计规律分布着。如果真实系统存在  $M$  个微观状态,则该系综便有  $M$  个系统组成,真实系统的每个微观状态都与系综中的一个系统相对应,对于系综中所有的系统取平均(称为系综平均)就能正确给出真实系统的时间平均值。因此,只要将系综中所有系统的代表点画到相空间中,则代表点的密度分布就形象地表示了真实系统处于某一微观状态附近的概率。

### 3) 微正则系综与等概率原理

微正则系综由这样一些系统组成,它们具有确定的粒子数目  $N$  和确定的体积  $V$ ,系统与外界的联系很微弱,使得系统的能量只能在  $E$  到  $E + \Delta E$  之间变动,因此,系统基本上是孤立的,微正则系综描述了一孤立系统的统计性质:系统的状态只能处于限定的能量  $E$  到  $E + \Delta E$  之内,处在这个能量内的任一微观状态出现的概率相等,而能量区间外的任一微观状态都是不可能出现的,这就是统计物理的基本假设——等概率原理,即对于处在平衡态的孤立系统,系统处于每一微观状态的概率是相等的。

## 2. 正则系综的分布函数

为了描述固体退火过程的统计性质,设想固体与一热源接触而处于平衡态,如图 2-1 所示。图中 I 为所研究的系统,II 为热源,系统 I 可以与热源交换能量,但是系统 I 的粒子数目和体积不变。

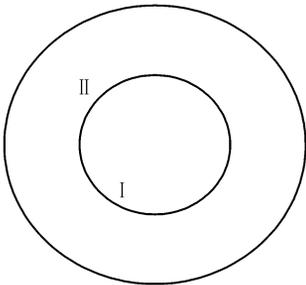


图 2-1 固体热平衡态

正则系综由这样一些系统组成,它们具有确定的粒子数  $N$  和确定的体积  $V$ ,系统与一个热源相接触而平衡。正则系综是能量可变,但粒子数守恒的系统的统计分布。利用微正则系综的等概率分布可以导出正则系综的概率分布。

设  $E_1$  代表系统 I 的能量,  $E_2$  代表热源 II 的能量,且  $E_2 \gg E_1$ 。将 I 和 II 合在一起成为大系统(I + II),该大系统是一个孤立的系统,其能量为

$$E \rightarrow E + \Delta E, \quad \Delta E \rightarrow 0$$

而

$$E = E_1 + E_2 \tag{2-3}$$

这里忽略了系统 I 与热源 II 之间的相互作用能量。大系统(I + I)既然是一个孤

立的系统,那么其统计性质就可以用微正则系综来描述,即处于任一微观状态的概率相等。

系统 I 处在能量为  $E_i$  的一个微观状态  $i$  上时,热源 II 可以处在能量区间  $[E - E_i, E - E_i + \Delta E]$  的所有微观态上,这些微观态以  $\Omega_2(E - E_i)$  来表示,则大系统 (I + II) 的微观态数是

$$1 \times \Omega_2(E - E_i) = \Omega_2(E - E_i) \quad (2-4)$$

若系统 I 的  $E_i$  能量的简并度(相同能量的状态数目,称为该能量的简并度)为  $g(E_i)$ , 则大系统 (I + II) 的总微观态数为

$$\Omega(E) = \sum_i g(E_i) \Omega_2(E - E_i) \quad (2-5)$$

由等概率原理可知,系统 I 处在  $E_i$  某一微观态  $i$  的概率为

$$P_i = \frac{\Omega_2(E - E_i)}{\Omega(E)} \quad (2-6)$$

其中,  $\Omega(E)$  为一常数,由于  $E_i \ll E$ , 可以将  $\ln \Omega_2$  展开为  $E_i$  的 Taylor 级数

$$\ln \Omega_2(E - E_i) = \ln \Omega_2(E) - E_i \left[ \frac{\partial \ln \Omega_2}{\partial E_2} \right]_{E_2=E} \quad (2-7)$$

其中,  $\Omega_2(E)$  为热源作为一孤立系统的状态数,按熵的定义,热源的熵为

$$S_2(E) = k \ln \Omega_2(E)$$

其中,  $k$  为 Boltzmann 常量。而

$$\left[ \frac{\partial S_2}{\partial E_2} \right]_v = \frac{1}{T} \quad (2-8)$$

于是

$$\left[ \frac{\partial \ln \Omega_2}{\partial E_2} \right] = \frac{1}{kT} \quad (2-9)$$

式(2-7)变为

$$\ln \Omega_2(E - E_i) = \frac{S_2(E)}{k} - \frac{E_i}{kT} \quad (2-10)$$

即有

$$\Omega_2(E - E_i) = \exp\left[\frac{S_2(E)}{k}\right] \exp\left[\frac{-E_i}{kT}\right] \quad (2-11)$$

代入式(2-5)得

$$P_i = A \exp\left[\frac{-E_i}{kT}\right] \quad (2-12)$$

其中,  $A$  是与  $E_i$  无关的常数。将系统 I 的所有可能状态的概率总和归一化,即

$\sum_i P_i = 1$ 。于是

$$A \sum_i \exp\left[\frac{-E_i}{kT}\right] = 1 \quad (2-13)$$

令  $Z = \sum_i \exp(-E_i/kT)$  为系统的配分函数,得  $A = 1/Z$ , 代入式(2-12)得

$$P_i = \frac{1}{Z} \exp\left\{\frac{-E_i}{kT}\right\} \quad (2-14)$$

其中,  $\exp(-E_i/kT)$  称为 Boltzmann 因子;  $T$  为绝对温度;  $k$  为 Boltzmann 常量。

式(2-14)的分布称为 Gibbs 正则分布。这种分布给出温度  $T$  时固体处于能量为  $E_i$  的微观态  $i$  的概率,显然,固体处于能量较低的微观态的概率较大。在温度较低时,那些能量相比最低的微观态最有可能出现。随着温度  $T$  的降低,概率集中在低能状态的一个更小的子集上,当温度趋于零时,固体只能处于能量为最小值的基态上。

### 3. Metropolis 准则

1953年, Metropolis 等提出重要性采样法。采用下述方法产生固体的状态序列:先给定以粒子相对位置表征的初始态  $i$ , 作为固体的当前状态,该状态的能量是  $E_i$ 。然后用扰动装置使随机选取的某个粒子的位移随机地产生一微小变化,得到一个新的状态  $j$ , 新状态的能量是  $E_j$ 。如果  $E_j < E_i$ , 则该新状态就作为“重要状态”;如果  $E_j > E_i$ , 则考虑到热运动的影响,该新状态是否是“重要状态”,要依据固体处于该状态的概率来判断。由式(2-14)可知,固体处于状态  $i$  和  $j$  的概率的比值等于相应 Boltzmann 因子的比值,即

$$r = \exp\left\{\frac{E_i - E_j}{kT}\right\} \quad (2-15)$$

$r$  是一个小于 1 的数,用随机数发生器产生一个  $[0, 1)$  的随机数  $\xi$ , 若  $r > \xi$ , 则新状态  $j$  作为重要状态,否则舍去。若新状态  $j$  是重要状态,就以  $j$  取代  $i$  成为当前状态,否则仍以  $i$  为当前状态,再重复以上新状态的产生过程。在大量迁移后,系统趋于能量较低的平衡态,固体状态的概率分布趋于式(2-14)的 Gibbs 正则分布。

由式(2-15)可知,高温下可接受与当前状态能差较大的新状态为重要状态。而在低温下只能接受与当前状态能差较小的新状态为重要状态。这与不同温度下热运动的影响完全一致。在温度趋于零时,就不能接受任何  $E_j > E_i$  的新状态  $j$  了。

上述接受新状态的准则称为 Metropolis 准则,相应的算法称为 Metropolis 算法。

#### 2.1.3 模拟退火算法

1983年, Kirkpatrick 等意识到组合优化与物理退火的相似性(表 2-1),并受到 Metropolis 准则的启迪,提出了模拟退火算法。模拟退火算法是基于 Monte Carlo 迭代求解策略的一种随机寻优算法,其出发点是基于物理退火过程与组合优化之

间的相似性,模拟退火算法由某一较高初温开始,利用具有概率突跳特性的 Metropolis 抽样策略在解空间中进行随机搜索,伴随温度的不断下降重复抽样过程,最终得到问题的全局最优解。

表 2-1 优化问题与物理退火问题的比较

优化问题	物理退火
解	分子状态
最优解	能量最低状态
目标函数	能量
设定初始温度	熔解过程
Metropolis 抽样	等温过程
温度的下降	冷却过程

设优化问题的一个解  $i$  及其目标函数  $f(i)$  分别与固体的一个微观状态  $i$  及其能量  $E_i$  等价,则随着算法进程递减其值的控制参数  $t$  担当固体退火过程中的温度  $T$  角色,则对于控制参数  $t$  的每一取值,算法持续进行“产生新解—判断—接受或舍弃”的迭代过程就对应着固体在某一恒定温度下趋于热平衡的过程,也就是执行了一次 Metropolis 算法。Metropolis 算法从某一初始状态出发,通过计算系统的时间演化过程,求出系统最终达到的状态相似,模拟退火算法从某一个初始解出发,经过大量解的变化后,可以求得给定控制参数值时优化问题的相对最优解。然后减少控制参数  $t$  的值,重复执行 Metropolis 算法,就可以在控制参数  $t$  趋于零时,最终求得优化问题的整体最优解。由于固体退火必须“徐徐”降温,才能使固体在某一温度下达到热平衡,最终趋于能量最少的基态,控制参数的值也必须缓慢衰减,才能确保模拟退火算法最终趋于优化问题的整体最优解集。

模拟退火算法用 Metropolis 算法产生优化问题解的序列,并由与 Metropolis 算法准则所对应的转移概率

$$P_i(i \Rightarrow j) = \begin{cases} 1, & f(j) \leq f(i) \\ \exp\left[-\frac{f(i)-f(j)}{t}\right], & f(j) > f(i) \end{cases} \quad (2-16)$$

确定是否接受从当前解  $i$  到新解  $j$  的转移。式(2-16)中的  $t \in R^+$  表示控制参数。开始让  $t$  取较大的值(与固体的熔解温度相对应),在进行足够多的转移后,缓慢减小  $t$  的值(与“徐徐”降温相对应),如此重复,直至满足某个停止准则时算法终止。因此,模拟退火算法可视为控制参数值递减时,Metropolis 算法的迭代。

假设存在邻域结构和产生器,  $t_k$  表示 Metropolis 算法第  $k$  次迭代时控制参数  $t$  的值,  $L_k$  表示 Metropolis 算法第  $k$  次迭代时产生的变换个数,则模拟退火算法的具体步骤如下:

- step 1 设定初始温度  $t = t_{\max}$ , 任选初始解  $i = i_0$ ;
- step 2 (内循环)
- step2.1 从  $i$  的邻域中随机变动, 选一个解  $j$ , 计算  $i$  和  $j$  对应的目标函数数值;
- 如果  $f(j) \leq f(i)$ , 则  $i = j$ ;
- 如果  $f(j) > f(i)$ , 则
- 若  $\exp((f(i) - f(j))/t) > \text{random}(0, 1)$ , 则  $i = j$ ;
- 否则重复 step2.1;
- step 3 (外循环)
- step3.1 降温  $t = \text{decrease}(t)$ ;
- step3.2 若不满足外循环停止条件, 则转 step2; 否则算法结束。

模拟退火算法依据 Metropolis 准则接受新解, 因此除接受优化解外, 还在一个限定范围内接受恶化解, 这正是模拟退火算法与局部搜索算法的本质区别所在。开始时  $t$  值大, 可能接受较差的恶化解; 随着  $t$  值的减小, 只能接受较好的恶化解; 最后在  $t$  值趋于零时, 就不再接受任何恶化解了。这就使得模拟退火算法既可以从局部最优的“陷阱”中跳出来, 更有可能求得整体最优解, 又不失简单性和通用性。因此, 对于大多数优化问题而言, 模拟退火算法要优于局部搜索算法。

## 2.2 模拟退火算法的收敛性分析

### 2.2.1 模拟退火算法的 Markov 链描述

为了便于阅读和理解, 这里先引用几个随机过程的概念和结论。

**定义 2.1** 设随机过程  $\{X(n), n \geq 0\}$ , 只能取可列个值  $i_0, i_1, \dots$ , 对于任意的  $n$  及  $i_0, i_1, \dots, i_n$ , 如果

$$P\{X(0) = i_0, X(1) = i_1, \dots, X(n) = i_n\} > 0$$

必有

$$\begin{aligned} P\{X(n+1) = i_{n+1} \mid X(0) = i_0, X(1) = i_1, \dots, X(n) = i_n\} \\ = P\{X(n+1) = i_{n+1} \mid X(n) = i_n\} \end{aligned} \quad (2-17)$$

则称  $\{X(n), n \geq 0\}$  为时间离散、状态离散的 Markov 链, 简称 Markov 链。

**定义 2.2** 称  $P\{X(m+n) = j \mid X(m) = i, n > 0\}$  为 Markov 链的  $n$  步转移概率, 记为  $P_{ij}^{(n)}$ 。

$P_{ij}^{(n)}$  具有下面两条性质:

- (1)  $P_{ij}^{(n)} \geq 0$ ;
- (2)  $\sum_{j \in I} P_{ij}^{(n)} = 1$ 。

**定义 2.3** 对于 Markov 链,如果

$$\begin{aligned} P_{ij}(m, m+1) &= P\{X(m+1) = j \mid X(m) = i\} \\ &= p_{ij}, \quad i, j \in I \end{aligned} \quad (2-18)$$

即从状态  $i$  出发转移到状态  $j$  的转移概率与时间起点  $m$  无关,则称这类 Markov 链为齐次 Markov 链。

**定义 2.4** 对于齐次 Markov 链,称  $P_{ij}$  为一步转移概率,全部  $P_{ij}(j, i \in I)$  所组成的一个矩阵  $P = (P_{ij})$  称为一步转移概率矩阵或随机矩阵。

**定义 2.5** 对转移矩阵为  $P$  的 Markov 链,若对任意  $i, j$ , 存在  $n \geq 1$ , 使得  $P_{ij}^{(n)} > 0$ , 则称该 Markov 链不可约。

**定义 2.6** 对转移矩阵为  $P$  的 Markov 链,若对任意  $i$ , 集合  $D_i = \{n \mid P_{ij}^{(n)} > 0\}$  的最大公约数  $d(i) = 1$ , 则称该 Markov 链为非周期的;若  $d(i) > 1$ , 则称 Markov 链为周期的。

**引理 2.1** 对于转移矩阵为  $P$  的不可约 Markov 链,若对任意  $i$ ,  $P_{ii} > 0$ , 则称该 Markov 链是非周期的。

**定义 2.7** 如果一步转移概率满足等式  $q_j = \sum_{i=1}^{\infty} q_i P_{ij}$ , 则称  $\{q_j \geq 0, j \in Z\}$  为 Markov 链的平稳分布。

在模拟退火算法中,算法从一个初始状态开始后,每一步状态转移均是在当前状态  $i$  的邻域  $N_i$  内随机产生新状态  $j$ , 然后以一定的概率进行接受。可见接受概率仅依赖于新状态和当前状态,并由温度加以控制,因此模拟退火算法对应了一个 Markov 链。

设  $P_{ij}(t)$  为在温度  $t$  下由状态  $i$  到状态  $j$  的转移概率,则对于任意  $i, j$ , 有

$$P_{ij}(k) = P_{ij}(t_k) = P_{ij}(t = t_k) = \begin{cases} G_{ij}(t_k) A_{ij}(t_k), & i \neq j \\ 1 - \sum_{l \in N_i, l \neq i} P_{il}(t_k), & i = j \end{cases} \quad (2-19)$$

其中,  $G_{ij}(t_k)$  表示由状态  $i$  产生状态  $j$  的概率,对任意  $i, j$ , 其定义为  $G_{ij}(t_k) = G_{ij} = \chi_{N_i}(j) / |N_i|$ , 如果新状态在当前状态的领域中以等概率产生,则  $G_{ij}(t_k) = 1 / |N_i|$ , 其中  $|N_i|$  为状态  $i$  的领域中状态总数,  $\chi_{N_i}(j)$  为示性函数。 $A_{ij}(t)$  表示当前状态  $i$  接受状态  $j$  的接受概率,一般定义为  $A_{ij}(t) = \min\{1, \exp((f(i) - f(j))/t)\}$ , 这里  $f(\cdot)$  为目标函数。

### 2.2.2 模拟退火算法的收敛性

我们对齐次情况讨论模拟退火算法的收敛性。

**定义 2.8** 设  $\Omega_{opt}$  是模拟退火算法的最优解集,若模拟退火算法经过充分地迭代后,有