

Artificial Intelligence

Principles and Applications

人工智能及其应用

杨天奇 编著



暨南大学出版社
JINAN UNIVERSITY PRESS

Artificial Intelligence
Principles and Applications

人工智能及其应用

杨天奇 编著



暨南大学出版社
JINAN UNIVERSITY PRESS

中国·广州

图书在版编目 (CIP) 数据

人工智能及其应用/杨天奇编著. —广州:暨南大学出版社, 2014. 6
ISBN 978 - 7 - 5668 - 1038 - 0

I. ①人… II. ①杨… III. ①人工智能—研究 IV. ①TP18

中国版本图书馆 CIP 数据核字 (2014) 第 111984 号

出版发行: 暨南大学出版社

地 址: 中国广州暨南大学

电 话: 总编室 (8620) 85221601

营销部 (8620) 85225284 85228291 85228292 (邮购)

传 真: (8620) 85221583 (办公室) 85223774 (营销部)

邮 编: 510630

网 址: <http://www.jnupress.com> <http://press.jnu.edu.cn>

排 版: 广州市科普电脑印务部

印 刷: 佛山市浩文彩色印刷有限公司

开 本: 787mm × 1092mm 1/16

印 张: 17.25

字 数: 368 千

版 次: 2014 年 6 月第 1 版

印 次: 2014 年 6 月第 1 次

定 价: 38.00 元

(暨大版图书如有印装质量问题, 请与出版社总编室联系调换)

内容简介

本书系统地阐述了人工智能的基本原理、方法和应用技术，反映了人工智能领域比较成熟的研究内容、最新进展和动态。全书共有11章，除第1章绪论外，其余可划分为两大部分。第一部分为基本人工智能，包括第2至第4章的知识表示、搜索技术和不确定性推理，论述了人工智能的基本理论与技术。第二部分为人工智能的高级理论与技术，包括第5至第11章，主要涉及智能优化算法、神经计算、支持向量机、粗糙集、数据挖掘算法、智能搜索引擎和分形理论及应用等人工智能的研究热点。

本书由浅入深、循序渐进、条理清楚，力求科学、实用、可读性强，能让学生在有限的时间内掌握人工智能的基本原理与应用技术，为将来的开发和研究工作打下基础。

本书可作为高等院校信息类各专业的研究生教材，高年级本科生可选取部分内容学习。也可供从事人工智能研究与应用的科技工作者学习参考。

前 言

人工智能自 1956 年问世以来，一直是引领未来发展的主导学科之一，是一门新思想、新观点、新理论、新技术不断涌现的前沿交叉学科。相关的研究成果已经广泛应用到各个领域，为推动科学技术的进步和发展发挥了巨大作用。

人工智能是一门研究机器智能的学科，作为一门前沿交叉学科，它的研究领域十分广泛。人工智能的远期目标是揭示人类智能的根本机理，用智能机器去模拟、延伸和扩展人类的智能。人工智能涉及脑科学、认知科学、计算机科学、系统科学、控制论等多种学科，并依赖于它们共同发展。近期目标是研究如何使现有的计算机更聪明，即使它能够运用知识去处理问题，能够模拟人类的智能行为。目前，人工智能仍处于发展时期，很多问题解决得还不够好，甚至不能求解，很多问题的求解还需要一定的条件。

本书基于“让现有的计算机更聪明”的理念，结合笔者多年来的硕士研究生教学经验和研究实践，吸取了国内外人工智能教材的优点，参考了国际上最新的研究成果编写而成。本书主要介绍人工智能中最基本、最经典的理论和方法，以及目前较为成熟和广泛应用的最新研究成果，为信息科学技术领域的研究生以及其他对人工智能感兴趣的科技工作者提供人工智能技术的知识。

全书共有 11 章，划分为两个部分。

第一部分：

第 1 章“绪论”，介绍了人工智能的基本概念、发展简史、研究的内容和领域；人工智能的观点和学派以及发展和未来。

第 2 章介绍了知识表示方法，包括谓词逻辑法、产生式表示方法、语义网络法等。

第 3 章介绍了搜索技术，包括搜索的基本概念、状态空间的盲目搜索、状态空间的启发式搜索、与/或树的盲目搜索、与/或树的启发式搜索和博弈搜索。

第 4 章介绍了不确定性推理原理，包括不确定性的概念、可信度方法、主观贝叶斯方法和证据理论。

第二部分：

第5章介绍了智能优化算法，包括组合优化问题、模拟退火算法、遗传算法、粒群优化和蚁群优化算法。

第6章介绍了神经计算的部分内容，包括感知器、BP神经网络和径向基神经网络原理；探讨了影响泛化能力的几个因素。

第7章介绍了支持向量机，包括支持向量机分类、核函数特征空间、支持向量机回归。

第8章介绍了粗糙集理论，包括信息系统、不可辨识关系、下近似和上近似、知识简约、决策表的简约算法等。

第9章介绍了数据挖掘算法，包括决策树学习算法、 K 均值聚类算法、 K 最近邻分类算法和 Apriori 算法。

第10章介绍了智能搜索引擎的概念，包括智能搜索引擎的特征、智能搜索引擎的主要策略、智能搜索引擎的技术和智能爬虫搜索算法等。

第11章介绍了分形理论及应用，包括分形与分形市场学说、多重分形理论、多重分形分析及应用等。

本书力求科学、实用、可读性强；内容由浅入深，循序渐进，条理清楚。教材采用逐层深入的策略和简洁、精练、提纲式的语言编写，以提高阅读效率。书中除第2章至第4章有连贯性之外，其他各章都可以任意选取，以适合不同学时、不同专业、不同层次的教学需求。

本书在多年人工智能教学经验和科研成果的基础上，吸取了国内外同类教材和网络上的有关文献的精华。他们的丰硕成果和贡献是本书学术思想的重要源泉，在此谨向这些教材和文献的作者致以崇高的敬意。他们的著作和资源为本书提供了丰富的营养，使我们受益匪浅。

本书的顺利编写得到了暨南大学研究生院、暨南大学计算机科学系等各级领导的支持和帮助。同时暨南大学计算机科学系的老师、同学在部分内容的编写、文字录入、图表制作、校对等方面做了大量的工作，特别是周文全、辛雅菲、张谏奇、张锐华、邓嘉兴、付琼芳、王开杰和陈玲等，在此一并表示感谢。

本书获得了“暨南大学研究生教材建设项目”的资助。

人工智能是多学科的交叉学科，涉及的内容广泛，而且一直在不断地发展，随时都在产生新的方法和理论，我们只能根据教学有针对性地选取。由于水平有限，书中可能会出现各种错误和疏漏，敬请读者给予批评指正。

杨天奇
2014年4月

目 录

前 言	1
第1章 绪 论	
1.1 人工智能的基本概念	1
1.2 人工智能的发展简史	2
1.3 人工智能的三种观点和三个学派	4
1.4 人工智能研究的内容和领域	5
1.5 人工智能的发展及未来	7
习 题	9
第2章 知识表示	
2.1 知识表示的概述	10
2.2 谓词逻辑法	12
2.2.1 一阶谓词的基本概念	12
2.2.2 谓词逻辑表示方法	14
2.2.3 谓词逻辑表示的特点	16
2.3 产生式表示方法	17
2.3.1 产生式基本形式	17
2.3.2 产生式系统的组成和求解过程	17
2.3.3 产生式系统的推理方式	19
2.3.4 产生式系统的特点	22
2.4 语义网络法	22
2.4.1 语义网络的基本概念	22
2.4.2 语义网络的表示方法	25

2.4.3 语义网络的推理过程 26

2.5 其他知识表示方法 27

2.5.1 框架表示法 27

2.5.2 剧本表示法 30

习 题 32

第3章 搜索技术

3.1 搜索的基本概念 34

3.1.1 状态空间法 34

3.1.2 问题归约法 37

3.1.3 搜索的分类与性能 39

3.2 状态空间的盲目搜索 40

3.2.1 一般图搜索过程 40

3.2.2 广度优先搜索 42

3.2.3 深度优先搜索 43

3.2.4 代价树搜索 44

3.3 状态空间的启发式搜索 46

3.3.1 启发性信息和估价函数 46

3.3.2 A 算法 47

3.3.3 A* 算法 48

3.3.4 A* 算法的性能分析 49

3.4 与/或树的盲目搜索 50

3.4.1 与/或树的一般搜索 50

3.4.2 与/或树的广度优先搜索 50

3.4.3 与/或树的深度优先搜索 52

3.5 与/或树的启发式搜索 53

3.5.1 解树的代价与希望树 53

3.5.2 与/或树的启发式搜索过程 54

3.6 博弈搜索 56

3.6.1 博弈问题 57

3.6.2 极小极大搜索过程 58

3.6.3 $\alpha - \beta$ 搜索过程 61

习 题	63
第 4 章 不确定性推理	
4.1 概述	67
4.1.1 不确定性的概念	67
4.1.2 不确定性推理要解决的问题	68
4.1.3 不确定性推理与确定性推理的差别	68
4.2 可信度方法	68
4.2.1 知识不确定性	69
4.2.2 证据不确定性	71
4.2.3 不确定性推理	72
4.3 主观贝叶斯方法	73
4.3.1 知识表示方式	73
4.3.2 证据不确定性及推理	75
4.4 证据理论	80
4.4.1 基本理论	80
4.4.2 证据理论的推理模型	84
习 题	90
第 5 章 智能优化算法	
5.1 基本概念	92
5.1.1 组合优化问题	92
5.1.2 局部搜索	93
5.2 模拟退火算法	96
5.2.1 固体退火	96
5.2.2 模拟退火算法	99
5.3 遗传算法	103
5.3.1 概述	104
5.3.2 遗传算法模型	104
5.3.3 基本遗传操作	107
5.3.4 遗传算法的应用	111
5.4 粒群优化	116

5.4.1	基本概念	117
5.4.2	粒群优化原理	117
5.4.3	局部和全局最优算法	119
5.4.4	参数分析	119
5.4.5	遗传算法和 PSO 的比较	120
5.5	蚁群优化算法	121
5.5.1	蚁群算法的基本概念	121
5.5.2	基于 TSP 的基本蚁群算法	123
习 题		126

第6章 神经计算

6.1	神经网络的基本概念	128
6.2	感知器	131
6.2.1	感知器模型	131
6.2.2	单层感知器学习过程	134
6.3	BP 神经网络	136
6.3.1	BP 算法原理	137
6.3.2	BP 网络学习过程	138
6.3.3	影响泛化能力的几个因素	139
6.3.4	初始样本探究	140
6.3.5	权值和阈值对网络泛化能力的影响	141
6.3.6	隐含层的影响	142
6.4	径向基神经网络	143
6.4.1	RBF 神经网络模型	144
6.4.2	RBF 神经网络的学习算法	146
6.4.3	RBF 网络的逼近性质	149
习 题		151

第7章 支持向量机

7.1	概述	152
7.1.1	机器学习的基本问题	152
7.1.2	统计学习理论的核心问题	154

7.2 支持向量机分类	156
7.2.1 线性 SVM	156
7.2.2 非线性 SVM	157
7.3 核函数特征空间	158
7.4 支持向量机回归	160
习 题	166
第8章 粗糙集	
8.1 信息系统	167
8.2 粗糙集相关理论	168
8.3 知识简约	171
8.4 决策表的简约算法	175
习 题	180
第9章 数据挖掘算法	
9.1 数据挖掘概述	182
9.2 决策树学习算法	184
9.2.1 信息熵理论	184
9.2.2 信息论在决策树学习中的应用	185
9.2.3 ID3 学习算法	186
9.3 K均值聚类算法	189
9.4 K最邻近分类算法(KNN)	190
9.5 Apriori 算法	192
9.5.1 关联规则的基本概念	192
9.5.2 关联规则的种类	193
9.5.3 Apriori 算法	194
习 题	197
第10章 智能搜索引擎	
10.1 智能搜索引擎的概念	199
10.1.1 搜索引擎	199

10.1.2	搜索引擎的分类	201
10.1.3	传统搜索引擎存在的问题	202
10.1.4	智能搜索引擎的提出	202
10.2	智能搜索引擎的特征	203
10.2.1	智能网络爬虫	203
10.2.2	智能人机接口	206
10.2.3	智能信息反馈	208
10.3	智能搜索引擎的主要策略	209
10.3.1	协同过滤策略	209
10.3.2	点击特征的智能化	212
10.3.3	用户自定义信息智能化	213
10.4	智能搜索引擎的技术	215
10.4.1	人机交互技术	215
10.4.2	文本分类技术	217
10.4.3	文档信息压缩技术	223
10.4.4	智能搜索代理 Agent 技术	224
10.4.5	自然语言理解技术	227
10.5	智能爬虫搜索算法	231
10.5.1	网页搜索策略	231
10.5.2	智能搜索策略(个性化搜索)	232
10.6	智能搜索引擎的发展对策	234
	习 题	235

第11章 分形理论及应用

11.1	分形与分形市场学说	236
11.1.1	分形理论的创立、发展和意义	236
11.1.2	什么是分形	237
11.1.3	各种分形实例	238
11.1.4	分形的特征	239
11.1.5	分形市场理论	240
11.2	多重分形理论	243
11.2.1	分形维的概念与意义	243

11.2.2 分形维的类别	244
11.2.3 多重分形	246
11.3 多重分形分析	252
11.3.1 多重分形谱的程序设计算法	252
11.3.2 多重分形谱分析和特征推导	254
11.4 预测应用举例	255
11.4.1 股票价格波动预测模型	256
11.4.2 聚类分析算法	257
11.4.3 股票价格波动的预测应用	259
参考文献	263

绪 论

人工智能(Artificial Intelligence, AI)主要研究如何用人工的方法和技术,模仿、延伸和扩展人的智能,实现机器智能。自从人工智能诞生以来,取得了许多令人兴奋的成果,在很多领域得到了广泛的应用。1956年麦卡锡正式提出“人工智能”这个术语,并把它作为一门新兴科学的名称。

1.1 人工智能的基本概念

随着相关学科的发展,人们对智能和人工智能的认识在不断变化,研究人工智能的方法和途径也在不断变化。关于“人工智能”的定义,还没有一个统一的认识,本书将介绍一些目前使用较多的概念和定义。

1. 智能的概念

自然界四大奥秘是:物质的本质、宇宙的起源、生命的本质和智能的发生。

当前学术界对智能还没有确切的定义,相关的主要流派有:

(1)思维理论:认为智能来源于思维活动,智能的核心是思维,人的一切知识都是思维的产物。可望通过对思维规律和思维方法的研究,来揭示智能的本质。

(2)知识阈值理论:认为智能取决于知识的数量及其可运用程度。一个系统所具有的可运用知识越多,其智能就会越高。

(3)进化理论:认为智能取决于感知和行为,取决于对外界复杂环境的适应,智能不需要知识、不需要表示、不需要推理,可由逐步进化来实现。

一般认为,智能是知识与智力的总和,知识是一切智能行为的基础,智力是获取知识并应用知识求解问题的能力。

2. 智能的特征表现

(1)感知能力:通过视觉、听觉、触觉、嗅觉等感觉器官感知外部世界的的能力。80%以上的信息可通过视觉得到,10%的信息可通过听觉得到。

(2)记忆能力:存储由感知器官感知到的外部信息以及由思维所产生的知识,并对记忆的信息进行处理。

(3)思维能力:

①逻辑思维(抽象思维):依靠逻辑进行思维,思维过程是串行的,容易形式化,思维过程具有严密性、可靠性。

②形象思维(直感思维):依据直觉,思维过程是并行协同式的,形式化困难,信息变形或缺少的情况下仍有可能得到比较满意的结果。

③顿悟思维(灵感思维):具有不定期的突发性、非线性的独创性及模糊性的特点,穿插于形象思维与逻辑思维之中。

(4)学习能力:学习既可能是自觉的、有意识的,也可能是不自觉的、无意识的;既可以是有教师指导的,也可以是通过自己实践的。

(5)行为能力(表达能力):信息的输出。

3. 人工智能

(1)人工智能:用人工的方法在机器(计算机)上实现的智能,或者说是人们使机器具有类似于人的智能。

(2)人工智能学科:一门研究如何构造智能机器(智能计算机)或智能系统,使它能模拟、延伸、扩展人类智能的学科。

(3)图灵测试:1950年图灵发表的《计算机与智能》中设计了一个测试,用以说明人工智能的概念。

1.2 人工智能的发展简史

自从“人工智能”在1956年达特茅斯学会上被提出,人工智能的概念也随之扩展,并影响到了其他技术的发展。

1. 孕育期(1956年之前)

关键事件有:

(1)亚里士多德(Aristotle)的三段论的提出。

(2)培根(F. Bacon)的归纳法的提出。

(3)莱布尼茨(G. W. Leibnitz)的符号表示和推理计算的提出。

(4)布尔(G. Boole)用符号语言描述思维活动的基本推理法则的提出。

(5)1936年,图灵的图灵机的发明。

(6)1943年,麦克洛奇(W. McCulloch)、匹兹(W. Pitts)的M-P模型的建立。

(7)美国爱荷华州立大学的阿塔纳索夫教授和他的研究生贝瑞在1937年至1941年间开发的世界上第一台电子计算机为人工智能的研究奠定了物质基础。

2. 形成期(1956年—1969年)

(1)1956年夏,由斯坦福大学教授麦卡锡,哈佛大学数学和神经学家、麻省理工学院教授明斯基,IBM公司信息研究中心负责人洛切斯特,贝尔实验室信息部数学研究

员香农共同发起,邀请普林斯顿大学的莫尔,IBM公司的塞缪尔,麻省理工学院的塞尔夫里奇和索罗莫夫以及兰德公司和卡内基-梅隆大学的纽厄尔、西蒙等10名年轻学者在达特茅斯大学召开了两个月的学术研讨会,讨论机器智能问题。会上经麦卡锡提议正式采用“人工智能”(Artificial Intelligence)这一术语,标志着人工智能学科正式诞生。

(2)1956年以后,人工智能的研究在机器学习、定理证明、模式识别、问题求解、专家系统及人工智能语言等方面都取得了许多令人瞩目的成就。

(3)1969年,召开了第一届国际人工智能联合会议(International Joint Conferences on Artificial Intelligence, IJCAI)。

(4)1970年,创办了国际性的人工智能杂志(*Artificial Intelligence*)。

3. 发展期(1970年—20世纪80年代末)

(1)20世纪60年代末,人工智能研究遇到困难,如机器翻译等。1966年美国顾问委员会的报告裁定,还不存在通用的科学文本机器翻译,也没有很好的实现前景。英国、美国中断了对大部分机器翻译项目的资助。

(2)1977年,费根鲍姆在第五届国际人工智能联合会议上提出了“知识工程”概念,推动了以知识为中心的研究。

(3)我国自1978年开始把“智能模拟”作为国家科学技术发展规划的主要研究课题。1981年成立了中国人工智能学会。

(4)人工智能已经成为计算机、航空航天、军事装备、工业等众多领域的关键技术。

4. 学派分立到综合期(20世纪80年代—21世纪初)

随着人工神经网络的再度兴起和布鲁克(R. A. Brooks)的机器虫的出现,人工智能研究形成了符号主义、连接主义和行为主义三大学派。

(1)符号主义学派。符号主义学派是指基于符号运算的人工智能学派,他们认为知识可以用符号来表示,认知可以通过符号运算来实现,例如专家系统等。

(2)连接主义学派。连接主义学派是指神经网络学派。在神经网络方面,继鲁梅尔哈特研制出BP网络之后,1987年,首届国际人工神经网络学术大会在美国的圣迭戈(San-Diego)举行,掀起了人工神经网络研究的第二次高潮。之后,随着模糊逻辑和进化计算的逐步成熟,又形成了“计算智能”这个统一的学科范畴。

(3)行为主义学派。行为主义学派是指进化主义学派。在行为模拟方面,麻省理工学院的布鲁克教授于1991年成功研制了能在未知的动态环境中漫游的有六条腿的机器虫。

(4)三大学派的综合集成。随着研究和应用的深入,人们又逐步认识到,三大学派各有所长、各有所短,应相互结合、取长补短、综合集成。

5. 智能科学技术的兴起期(21世纪初以来)

目前,一个以人工智能为核心,以自然智能、人工智能、集成智能为一体的新的智能科学技术学科正在逐步兴起,并引起了人们的极大关注。

该学科研究的主要特征包括以下几个方面：

- (1) 由人工智能的单一研究走向以自然智能、人工智能、集成智能为一体的协同研究。
- (2) 由人工智能学科的独立研究走向重视与脑科学、认知科学等学科的交叉研究。
- (3) 由多个不同学派的独立研究走向多学派的综合研究。
- (4) 由个体、集中智能的研究走向群体、分布智能的研究。

1.3 人工智能的三种观点和三个学派

1. 三种主要观点

1971 年美国电气电子工程师协会建立了一个工作组，关于“人工智能”现象，这个小组概括了三种主要观点。

(1) 乐观的观点。以纽厄尔和西蒙为代表的专家认为：机器模拟人类思维的前提是假设思维的基础是一套加工信息的基本规则，因此可以脱离大脑的有机基体的特性去认识思维，并能用现代电子计算机复现思维，计算机原则上最终能在智能活动的一切领域超过人。这两位启发式程序设计的创始人提出了令人倍受鼓舞的预测：不用十年，电子计算机便可成为世界象棋冠军了。但是十年过去了，他们的预言并没有成真。直到 1997 年 5 月，IBM 的“深蓝”才战胜世界象棋冠军卡斯帕罗夫。而“深蓝”的设计师还是认为，“深蓝”仅仅是处理并记牢了人所撰写的程序，它甚至连最笨的人也比不上。

(2) 悲观的观点。否定人的思维跟“人工智能”在实质上有任何相同之处。认为“机器的思维跟人的思维是根本不同的”。指出了人的思维不完全等于现代计算机所复现的信息过程，但是他们不能完全否定人跟电子计算机的活动之间有相似之处。

(3) 犹豫和怀疑态度。美国哲学家普特南提出，人的思维跟“人工智能”的差别是一个虚假的问题，这个问题的解决或者取决于社会是否承认机器人是人（如果人能做的事机器人都能做的事），或者取决于机器人本身（如果向它们提出这个问题的话）。他认为将图灵的计算机的逻辑状态等同于人的“思维状态”是臆想出来的。

2. 三个主要学派

(1) 符号主义学派，亦称功能模拟学派。这个学派的符号主义方法以物理符号系统假设和有限性原理为基础，主要观点是智能活动的基础是物理符号系统，思维过程是符号模式的处理过程。这个学派可谓实力雄厚，约维尔、西蒙两位 ACM 图灵奖得主的支持足以说明其底蕴之厚。

(2) 连接主义学派，亦称结构模拟学派。这个学派则以人工神经网络模型为核心，他们认为人工智能基于神经网络及网络间的连接机制和学习算法，希望通过研究搞清楚大脑的结构以及它进行信息处理的过程和机理，可望揭示人类智能的奥秘，从而真正实现人类智能在机器上的模拟。这个学派的代表成果是 M-P 模型和 PDP 理论。这