

人工智能

计算Agent基础

(加) David L. Poole Alan K. Mackworth 著

董红斌 董兴业 童向荣 汪廷华 译

黄厚宽 审校

Artificial Intelligence
Foundations of Computational Agents

ARTIFICIAL INTELLIGENCE

FOUNDATIONS OF COMPUTATIONAL AGENTS



DAVID L. POOLE
ALAN K. MACKWORTH



机械工业出版社
China Machine Press

计 算 机 科 学 丛

人工智能

计算Agent基础

(加) David L. Poole Alan K. Mackworth 著

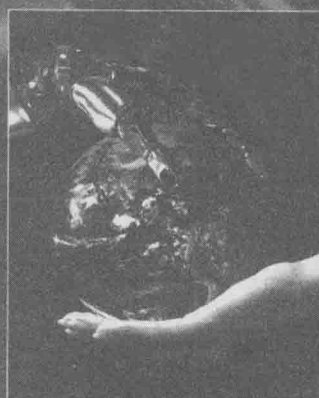
董红斌 董兴业 童向荣 汪廷华 译

黄厚宽 审校

Artificial Intelligence
Foundations of Computational Agents

ARTIFICIAL INTELLIGENCE

FOUNDATIONS OF COMPUTATIONAL AGENTS



DAVID L. POOLE
ALAN K. MACKWORTH



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

人工智能: 计算 Agent 基础 / (加) 普尔 (Poole, D. L.), (加) 麦克沃思 (Mackworth, A. K.) 著; 董红斌等译. —北京: 机械工业出版社, 2014.12

(计算机科学丛书)

书名原文: Artificial Intelligence: Foundations of Computational Agents

ISBN 978-7-111-48457-8

I. 人… II. ①普… ②麦… ③董… III. 人工智能 - 研究 IV. TP18

中国版本图书馆 CIP 数据核字 (2014) 第 258294 号

本书版权登记号: 图字: 01-2011-7728

This is a(n) simplified Chinese of the following title(s) published by Cambridge University Press:

David L. Poole and Alan K. Mackworth: Artificial Intelligence: Foundations of Computational Agents (ISBN 978-0-521-51900-7).

Original English language edition copyright © 2010 by David L. Poole and Alan K. Mackworth.

This simplified Chinese for the People's Republic of China (excluding Hong Kong, Macau and Taiwan) is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press and China Machine Press in 2015.

This simplified Chinese is authorized for sale in the People's Republic of China (excluding Hong Kong, Macau and Taiwan) only. Unauthorized export of this simplified Chinese is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of Cambridge University Press and China Machine Press.

本书原版由剑桥大学出版社出版。

本书简体字中文版由剑桥大学出版社与机械工业出版社合作出版。未经出版者预先书面许可, 不得以任何方式复制或抄袭本书的任何部分。

此版本仅限在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 销售。

本书非常全面地介绍了人工智能科学, 涵盖智能体、表达和推理、学习与规划以及个体和关系的推理等内容, 并充分考虑了 AI 科学中理论描述的形式化及实验的难以控制等特点, 对理论部分和实验部分做了很好的平衡, 展示了理论与实验之间的联系, 兼顾了科学与工程共同发展。本书提供的材料丰富、完整, 并很好地兼顾了各种层次的读者, 使之都能在书中发现自己感兴趣的东西。

本书适合作为计算机科学或者相关学科 (如计算机工程、哲学、认知科学和心理学) 的高年级本科生和研究生教材, 同时也适合研究人员阅读。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 朱秀英

责任校对: 董纪丽

印 刷: 三河市宏图印务有限公司

版 次: 2015 年 1 月第 1 版第 1 次印刷

开 本: 185mm×260mm 1/16

印 张: 29.25

书 号: ISBN 978-7-111-48457-8

定 价: 79.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光/邹晓东

文艺复兴以降,源远流长的科学精神和逐步形成的学术规范,使西方国家在自然科学的各个领域取得了垄断性的优势;也正是这样的传统,使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中,美国的产业界与教育界越来越紧密地结合,计算机学科中的许多泰山北斗同时身处科研和教学的最前线,由此而产生的经典科学著作,不仅擘划了研究的范畴,还揭示了学术的源变,既遵循学术规范,又自有学者个性,其价值并不会因年月的流逝而减退。

近年,在全球信息化大潮的推动下,我国的计算机产业发展迅猛,对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇,也是挑战;而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下,美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此,引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用,也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始,我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力,我们与Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage等世界著名出版公司建立了良好的合作关系,从他们现有的数百种教材中甄选出Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson等大师名家的一批经典作品,以“计算机科学丛书”为总称出版,供读者学习、研究及珍藏。大理石纹理的封面,也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力襄助,国内的专家不仅提供了中肯的选题指导,还不辞劳苦地担任了翻译和审校的工作;而原书的作者也相当关注其作品在中国的传播,有的还专程为其书的中译本作序。迄今,“计算机科学丛书”已经出版了近两百个品种,这些书籍在读者中树立了良好的口碑,并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑,这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化,教育界对国外计算机教材的需求和应用都将步入一个新的阶段,我们的目标是尽善尽美,而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正,我们的联系方式如下:

华章网站: www.hzbook.com

电子邮件: hzsj@hzbook.com

联系电话: (010)88379604

联系地址: 北京市西城区百万庄南街1号

邮政编码: 100037



华章教育

华章科技图书出版中心

人工智能自 20 世纪 50 年代诞生起就受到学者们的广泛关注,到目前为止国内外很多著名高校都成立了专门研究人工智能学科的研究机构。人工智能的研究涉及多门学科的技术,因此,人工智能是一门具有挑战性的学科。伴随计算机技术的发展和全球信息化大潮的推动,人工智能的研究课题和应用领域也在不断扩展。

本书非常详尽、系统地阐述了人工智能相关技术,内容丰富、连贯,其中涵盖智能体部分、表达和推理部分、学习和规划部分以及个体和关系的推理部分等。全书共 15 章,各章相对独立却也存在一定的逻辑关系。该书在各章节中都针对不同的概念给出了示例或图形,且难度由浅入深,方便读者更加透彻地理解复杂概念或方法。本书既适合人工智能的初学者,也适合熟悉人工智能的读者,但是建议读者对启发式算法、命题与推理、概率以及算法与数据结构有一定基础之后再进行阅读。

在翻译过程中,我们发现该书有以下特点:

1) 示例鲜活,书中所采用的示例都取材于实际生产、生活。该方式既阐明了技术问题,又说明了应用问题。

2) 伪代码简洁易懂,该书对算法的伪代码描述省去不必要的部分,增加适当的文字描述,算法脉络十分清晰。

3) 重点突出,该书特别强调对知识点的归纳总结,并对需要注意的问题进行提炼。

本书是一本有关 AI 的书籍,供计算机科学或者相关学科(例如计算机工程、电子工程、哲学、认知科学和心理学)的高年级本科生或研究生使用。

本书的翻译工作由董红斌、董兴业、童向荣和汪廷华共同完成,全书由黄厚宽审校。本书第 1、2、3 和 10 章由童向荣翻译,第 4、5、6、8 和 13 章由董红斌翻译,第 7、11、15 章和附录由汪廷华翻译,第 9、12 和 14 章由董兴业翻译。本书的翻译工作得到了国家自然科学基金的资助(No. 61472095),在此表示谢意。感谢王玲玲、孙雪姣、郭艳燕、翟一鸣、侯薇、崔晓晖、滕旭阳和杨雪等老师和同学对本书翻译过程的参与和支持。

在翻译过程中,我们力求忠实、准确地把握原著并保持原著的风格。但由于水平有限和时间仓促,书中表达不准确之处在所难免,在此恳请各位专家和广大读者批评指正。

译者

2014 年 7 月

本书是一本关于人工智能(AI)科学的图书。本书认为 AI 所要研究的是如何设计智能计算 Agent(智能体)。本书采用教科书的组织形式,适合广大读者阅读。

过去几十年,我们见证了 AI 这门综合学科的兴起。正如其他学科一样, AI 具有清晰、规范的理论 and 难操控的实验部分。本书平衡了理论和实验,并将两者密切地结合。俗话说“好的理论必须有其实用价值”,因此我们将工程应用融入到 AI 的科学研究中。本书所述方法都秉承了格言“凡事都应尽量从简,但不可过简”。我们认为科学必须有其坚实的基础,因此本书介绍了基础梗概,并且提供了构建有效智能系统的相关实例,这些实例也说明了智能系统所必需的复杂事物。尽管有可能生成复杂的系统,但是其基础和构建应该是简单的。

本书可作为人工智能的教科书,适用于计算机工程、哲学、认知科学以及心理学专业的高年级本科生和研究生。本书倾向于技术层面的思考,其中包含部分技术挑战;并且本书注重实践学习:设计、构造和实现系统。每个具有科学求知欲的读者都会从本书的学习中受益。书中是根据需要来提出概念的,所以要求读者具备关于计算系统的过往经验,但在我们构造的概念上不需要具备前期基础理论的研究,诸如逻辑、概率、微积分和控制理论。

认真的学生可从本书的专业知识中获得不同层次的宝贵技能,包括规范和设计智能 Agent,在具有挑战性的应用领域中进行软件系统的实现、测试和改进。不仅如此,学生应该很高兴能够参与到智能 Agent 这门学科的兴起。能够操控现实世界中普遍存在的、智能的、嵌入的 Agent,这类实用技能也是市场的巨大需求。

智能 Agent 的研究重点在于其在环境中起到的作用。开始时我们针对简易、静态的环境来研究简单 Agent 的行为,随后逐渐增加 Agent 的能力来应对更具有挑战的环境。我们从九个方面来研究复杂性,这样可以渐进地、模块化地向读者介绍构建智能 Agent 所面临的挑战。本书结构的安排尽量使得读者可以分别理解每个方面,并且我们在四种不同的 Agent 任务中反复论证使其更加具体,这四个 Agent 任务分别是:传送机器人、诊断助手、制导系统和交易 Agent。

学生可以把 Agent 理解为一个分层设计的 Agent,该 Agent 在随机环境中具有智能行为,但是它仅能得到部分观察。该部分观察是通过个体与个体间关系的推理所得到,具有复杂的参数、行动中学习、考虑环境中的其他 Agent 和给定自身计算限制能正确处理的特点。当然我们不能一开始就设计这样的 Agent,构建此类 Agent 仍然是一个研究课题。所以我们首先介绍最简单的 Agent,进而说明如何用模块化的方法将这些复杂功能添加其中。

我们给出了几个设计选项,这些可以说明本书与同类图书的区别,包括作者早期撰写

的书籍：

- 我们给出了用于理解 AI 的清晰框架。本书不介绍那些不适合在一起研究的分散主题。例如，我们不介绍 AI 中不相关的逻辑观点和概率观点，但是却提供了一个多维设计空间，在这个空间中学生可以得到一个整体情况，并且在该空间内概率和逻辑推理可以共存。
- 我们认为最好将构成复杂技术的基础理论清晰地介绍给读者，而不是直接介绍这些复杂技术。这意味着本书所覆盖的内容与当前科学前沿有着较大的差距，但同时意味着学生可以具备较好的基础去理解当前和未来的科学研究。
- 我们难以决定的问题之一是如何线性化设计空间。在我们上一本书中 (Poole, Mackworth 和 Goebel, 1998)，较早就提出了一种关系语言，并且根据该语言构建了基础理论。但对于不具备关联的工作，该方法对于学生来说比较难以理解，比如强化学习是根据状态展开而非关系。本书采用后期关系化的方法，此方法在基于特征表达的情况下较好地反映了过去几十年研究所取得的进步。该方法使得学生可以理解概率推理和逻辑推理的互补性。虽然是采用后期关系化的方法，但是该书结构使得老师可以较早地说明各章节的关系。

本书实用的例子来自于 AIspace.org (<http://www.aispace.org>)，我们在设计过程中一直采用这些教学程序。学生如果想更多地获得构建智能系统的经验，那么也应该尝试使用 LISP 或 Prolog 这种高级的符号操作语言进行实验。我们同样提供了本书许多问题在 AILog (与 Prolog 相关的简洁的逻辑编程语言) 上的实现。当然此部分对于理解和使用本书思想不是最重要的。

通过对 Agent 性能强度和表示语言这两方面的深入研究，本书方法在总结和分类各种 AI 应用上比传统方法更加简洁有效。但是例如计算视觉和计算语言学的应用在本书没有提及。

本书没有给出一个百科全书式的 AI 视图，也没去研究每一个被提出的重要思想。我们选择详尽地讲解一些基本思想，在这些基本思想上可以构造其他更复杂的技术，我们阐明了这些复杂技术是如何在此基础上被扩展的。

图 1 展示了本书所涉及的各个标题。实线给出了前提条件。前提结构通常不包含所有的子标题。为了撰写书籍，我们对各个主题进行章节编号。但是本书的讲解可以不按照章节编号，只要其满足前提条件的结构。

各章结尾处给出的参考文献并不全面：我们只是引用了我们直接使用的文献，以及那些我们认为通过引用经典工作和最近研究而提供了良好概述的文献。希望研究人员不要因为自己的研究工作被遗漏而不悦。如果您觉得书中思想有纰漏，欢迎您反馈给我们。再次重申本书不是 AI 研究的综述类图书。

诚邀您加入我们的智力探险活动：构建智能 Agent 科学。

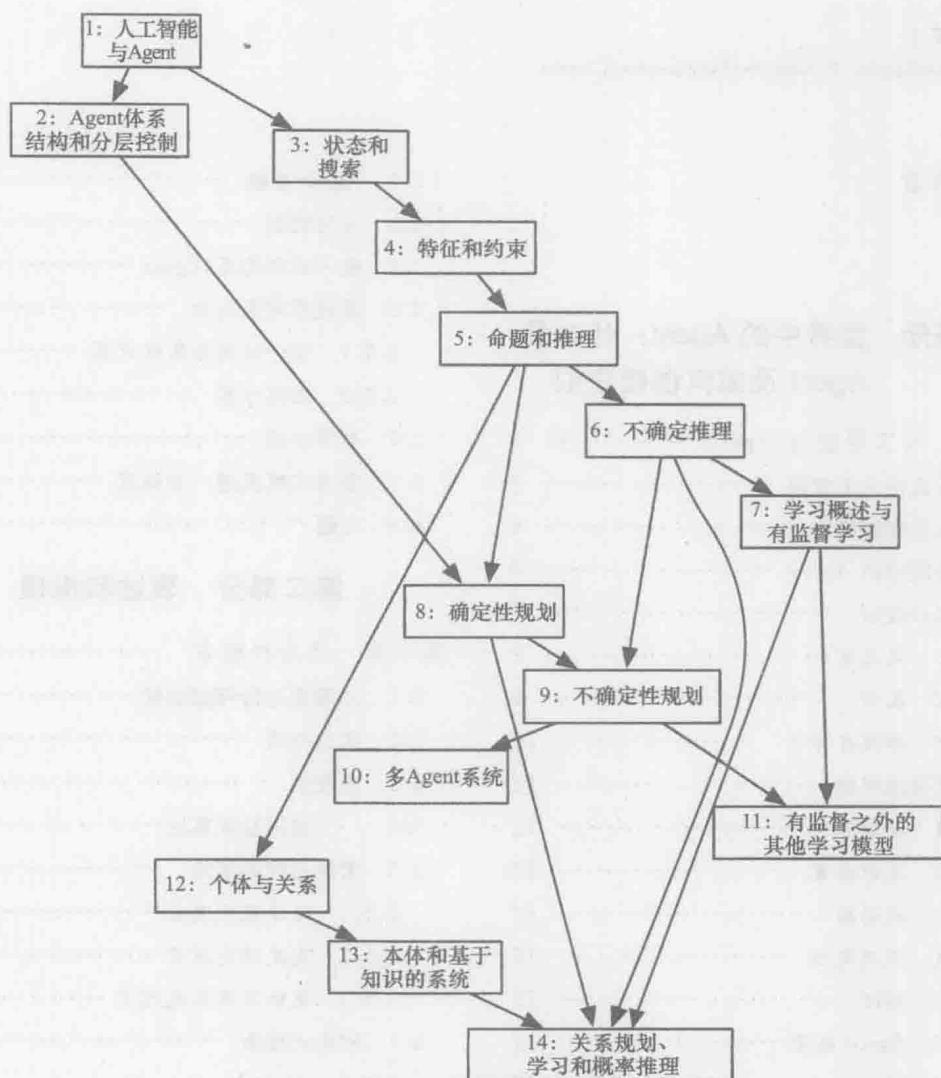


图1 章节和依赖关系的概述

致谢

感谢 Randy Goebel 对本书提出的宝贵意见。也非常感谢在早期起草本书时 Giuseppe Carenini、Cristina Conati、Mark Crowley、Pooyan Fazli、Holger Hoos、Manfred Jaeger、Mohammad Reza Khojasteh、Jacek Kisynski、Bob Kowalski、Kevin Leyton-Brown、Marian Mackworth、Gabriel Murray、Alessandro Provetti、Marco Valtorta 和匿名审稿人提出的有益意见。感谢在草稿中指出过许多错误的学生。

感谢 Jen Fernquist 设计的网站，以及 Tom Sgouros 对 hyperlatex 所做的维护。非常感谢 James Falen 允许我们引用他的诗。感谢编辑 Lauren Cowles 和剑桥大学出版社的工作人员的支持、鼓励和帮助。如发现错误请与我们联系。

David Poole

Alan Mackworth

推荐阅读

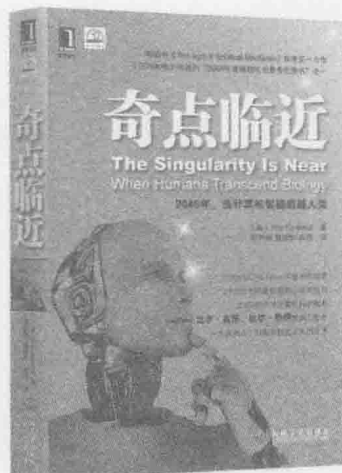


人工智能：复杂问题求解的结构和策略（原书第6版）

作者：George F. Luger ISBN: 978-7-111-28345-4 定价：79.00元

人工智能：智能系统指南（原书第3版）

作者：Michael Negnevitsky ISBN: 978-7-111-38455-7 定价：79.00元



奇点临近

作者：Ray Kurzweil ISBN: 978-7-111-35889-3 定价：69.00元

机器学习

作者：Tom Mitchell ISBN: 978-7-111-10993-7 定价：35.00元

推荐阅读



神经网络与机器学习 (原书第3版)

作者: Simon Haykin ISBN: 978-7-111-32413-3 定价: 79.00元



机器学习导论 (原书第2版)

作者: Ethem Alpaydin ISBN: 978-7-111-45377-2 定价: 59.00元



数据挖掘: 实用机器学习工具与技术 (原书第3版)

作者: Ian H. Witten 等 ISBN: 978-7-111-45381-9 定价: 79.00元



机器学习基础教程

作者: Simon Rogers 等 ISBN: 978-7-111-40702-7 定价: 45.00元

推荐阅读



中文版
第6版

作者: Abraham Silberschatz 著
中文翻译版: 978-7-111-37529-6, 99.00元
英文精编版: 978-7-111-40086-8, 69.00元
本科教学版: 978-7-111-40085-1, 59.00元



中文版
第3版

作者: Jiawei Han 等著
英文版: 978-7-111-37431-2, 118.00元
中文版: 978-7-111-39140-1, 79.00元



中文版
第3版

作者: Ian H. Witten 等著
英文版: 978-7-111-37417-6, 108.00元
中文版: 978-7-111-45381-9, 79.00元



英文版
第5版

作者: Andrew S. Tanenbaum 著
书号: 978-7-111-35925-8, 99.00元



中文版

作者: Behrouz A. Forouzan 著
英文版: 978-7-111-37430-5, 79.00元
中文版: 978-7-111-40088-2, 99.00元



中文版
第6版

作者: James F. Kurose 著
书号: 978-7-111-45378-9, 79.00元



中文版
第3版

作者: Thomas H. Cormen 等著
书号: 978-7-111-40701-0, 128.00元



英文版
第5版

作者: John L. Hennessy 著
书号: 978-7-111-36458-0, 138.00元



作者: Edward Ashford Lee 著
书号: 978-7-111-36021-6, 55.00元

目 录

Artificial Intelligence: Foundations of Computational Agents

出版者的话

译者序

前 言

第一部分 世界中的 Agent: 什么是 Agent 及如何创建它们

第 1 章 人工智能与 Agent	2
1.1 什么是人工智能	2
1.2 人工智能简史	4
1.3 环境中的 Agent	6
1.4 知识表示	7
1.4.1 定义解	8
1.4.2 表示	9
1.4.3 推理与行为	10
1.5 复杂性维度	12
1.5.1 模块性	12
1.5.2 表示方案	13
1.5.3 规划期	14
1.5.4 不确定性	15
1.5.5 偏好	15
1.5.6 Agent 数量	16
1.5.7 学习	16
1.5.8 计算限制	17
1.5.9 多维交互	18
1.6 原型应用	19
1.6.1 自主传送机器人	19
1.6.2 诊断助手	21
1.6.3 智能指导系统	23
1.6.4 交易 Agent	24
1.7 本书概述	25
1.8 本章小结	26
1.9 参考文献及进一步阅读	26
1.10 习题	27
第 2 章 Agent 体系结构和分层控制	28
2.1 Agent	28

2.2 Agent 系统	29
2.3 分层控制	32
2.4 嵌入式和仿真 Agent	37
2.5 通过推理来行动	38
2.5.1 设计时间与离线计算	39
2.5.2 在线计算	40
2.6 本章小结	41
2.7 参考文献及进一步阅读	42
2.8 习题	42

第二部分 表达和推理

第 3 章 状态和搜索	46
3.1 用搜索进行问题求解	46
3.2 状态空间	47
3.3 图搜索	48
3.4 一个通用搜索算法	50
3.5 无信息搜索策略	51
3.5.1 深度优先搜索	51
3.5.2 宽度优先搜索	55
3.5.3 最低花费优先搜索	56
3.6 启发式搜索	56
3.6.1 A* 搜索	58
3.6.2 搜索策略总结	59
3.7 更复杂的搜索方法	60
3.7.1 环检查	60
3.7.2 多路径剪枝	60
3.7.3 迭代深化	62
3.7.4 分支界限法	63
3.7.5 搜索方向	65
3.7.6 动态规划法	66
3.8 本章小结	68
3.9 参考文献及进一步阅读	68
3.10 习题	69
第 4 章 特征和约束	71
4.1 特征和状态	71

4.2 可能世界、变量和约束	72	5.5.2 完备知识的验证程序	130
4.2.1 约束	74	5.6 溯因推理	132
4.2.2 约束满足问题	75	5.7 因果模型	135
4.3 生成-测试算法	76	5.8 本章小结	137
4.4 使用搜索求解 CSP	76	5.9 参考文献及进一步阅读	137
4.5 一致性算法	78	5.10 习题	138
4.6 域分割	81	第6章 不确定推理	145
4.7 变量消除	83	6.1 概率	145
4.8 局部搜索	85	6.1.1 概率的语义	146
4.8.1 迭代最佳改进	86	6.1.2 概率公理	147
4.8.2 随机算法	87	6.1.3 条件概率	149
4.8.3 评估随机算法	90	6.1.4 期望值	152
4.8.4 局部搜索中利用命题结构	92	6.1.5 信息理论	153
4.9 基于种群的方法	92	6.2 独立性	153
4.10 最优化	94	6.3 信念网络	155
4.10.1 最优化的系统方法	96	6.4 概率推理	164
4.10.2 局部搜索最优化	98	6.4.1 信念网络中的变量消除	164
4.11 本章小结	99	6.4.2 通过随机模拟进行近似推理	169
4.12 参考文献及进一步阅读	100	6.5 概率和时间	176
4.13 习题	100	6.5.1 马尔可夫链	176
第5章 命题和推理	103	6.5.2 隐马尔可夫模型	176
5.1 命题	103	6.5.3 监听和平滑算法	179
5.1.1 命题演算的语法	103	6.5.4 动态信念网络	180
5.1.2 命题演算的语义	104	6.5.5 时间粒度	181
5.2 命题确定子句	107	6.6 本章小结	181
5.2.1 问题与解答	109	6.7 参考文献及进一步阅读	181
5.2.2 验证	110	6.8 习题	182
5.3 知识表示问题	115		
5.3.1 背景知识与观察	115	第三部分 学习与规划	
5.3.2 询问用户	116	第7章 学习概述与有监督学习	186
5.3.3 知识层的解释	117	7.1 学习问题	186
5.3.4 知识层的调试	119	7.2 有监督学习	189
5.4 反证法验证	122	7.2.1 评估预测	190
5.4.1 Horn 子句	122	7.2.2 无输入特征的点估计	193
5.4.2 假说与冲突	123	7.2.3 概率学习	195
5.4.3 基于一致性的诊断	124	7.3 有监督学习的基本模型	196
5.4.4 通过假设和 Horn 子句推理	126	7.3.1 决策树学习	196
5.5 完备知识假设	127	7.3.2 线性回归与分类	200
5.5.1 非单调推理	130	7.3.3 贝叶斯分类器	203

7.4 组合模型	206	9.5.3 值迭代	268
7.4.1 神经网络	207	9.5.4 策略迭代	270
7.4.2 集成学习	210	9.5.5 动态决策网络	271
7.5 避免过拟合	210	9.5.6 部分可观察决策过程	273
7.5.1 最大后验概率和最小描述 长度	211	9.6 本章小结	273
7.5.2 交叉验证	213	9.7 参考文献及进一步阅读	274
7.6 基于案例的推理	213	9.8 习题	274
7.7 改进假设空间的学习	215	第10章 多 Agent 系统	280
7.7.1 变型空间学习	216	10.1 多 Agent 框架	280
7.7.2 可能近似正确学习	218	10.2 博弈的表示	281
7.8 贝叶斯学习	220	10.2.1 博弈的标准形式	281
7.9 本章小结	224	10.2.2 博弈的扩展形式	282
7.10 参考文献及进一步阅读	225	10.2.3 多 Agent 决策网络	283
7.11 习题	225	10.3 完全信息的计算策略	284
第8章 确定性规划	229	10.4 部分可观察的多 Agent 推理	286
8.1 状态、动作以及目标的表示	229	10.4.1 纳什均衡计算	290
8.1.1 显式状态空间表示法	230	10.4.2 学习协调	292
8.1.2 基于特征的动作表示	231	10.5 群体决策	294
8.1.3 STRIPS 表示法	232	10.6 机制设计	294
8.1.4 初始状态和目标	233	10.7 本章小结	296
8.2 前向规划	233	10.8 参考文献及进一步阅读	297
8.3 回归规划	235	10.9 习题	297
8.4 CSP 规划	236	第11章 有监督之外的其他学习 模型	298
8.5 偏序规划	238	11.1 聚类	298
8.6 本章小结	241	11.1.1 期望最大化	298
8.7 参考文献及进一步阅读	241	11.1.2 k -均值	299
8.8 习题	241	11.1.3 用于软聚类的期望最大化	300
第9章 不确定性规划	244	11.2 信念网络学习	303
9.1 偏好和效用	245	11.2.1 概率学习	303
9.2 一次性的决策	252	11.2.2 未观察到的变量	304
9.3 序贯决策	255	11.2.3 缺失数据	304
9.3.1 决策网络	256	11.2.4 结构学习	305
9.3.2 策略	258	11.2.5 信念网络学习的一般情形	306
9.3.3 决策网络的变量消除	259	11.3 增强学习	306
9.4 信息与控制的价值	262	11.3.1 演化算法	308
9.5 决策过程	264	11.3.2 时间差	308
9.5.1 策略值	267	11.3.3 Q -学习	309
9.5.2 最优策略值	267	11.3.4 探索与利用	312

11.3.5 增强学习算法的评估	313	13.2 灵活的表示	363
11.3.6 在策略学习	314	13.2.1 选择个体和关系	364
11.3.7 为路径分配信用和责任	315	13.2.2 图形化表示	366
11.3.8 基于模型的方法	317	13.2.3 原始关系与导出关系	369
11.3.9 基于特征的增强学习	319	13.3 本体与知识共享	373
11.4 本章小结	320	13.3.1 描述逻辑	376
11.5 参考文献及进一步阅读	321	13.3.2 顶层本体	380
11.6 习题	321	13.4 查询用户和其他知识来源	382
第四部分 个体与关系的推理		13.4.1 函数化关系	383
第12章 个体与关系	324	13.4.2 更普遍的问题	383
12.1 在特征之外利用结构	324	13.5 实现基于知识的系统	384
12.2 符号与语义	325	13.5.1 基语言和元语言	385
12.3 Datalog: 一个关联规则语言	326	13.5.2 普通的元解释器	386
12.3.1 基 Datalog 的语义	328	13.5.3 扩展基语言	387
12.3.2 解释变量	329	13.5.4 深度有限搜索	388
12.3.3 带变量的查询	333	13.5.5 元解释器构建证明树	389
12.4 证明与替换	334	13.5.6 可询问用户的元解释器	390
12.4.1 带变量的自底向上过程	335	13.5.7 推迟目标	391
12.4.2 带变量的确定性归结	337	13.6 本章小结	391
12.5 函数符号	339	13.7 参考文献及进一步阅读	392
12.6 在自然语言处理中的应用	344	13.8 习题	392
12.6.1 在上下文无关文法中使用 限定子句	345	第14章 关系规划、学习和概率 推理	396
12.6.2 增强文法	347	14.1 规划个体与关系	396
12.6.3 为非终结符号建立结构	348	14.1.1 情景演算	396
12.6.4 封装的文本输出	348	14.1.2 事件演算	401
12.6.5 强制约束	349	14.2 个体与关系的学习	402
12.6.6 建立自然语言与数据库的 接口	350	14.3 概率关系模型	406
12.6.7 局限	351	14.4 本章小结	410
12.7 相等	352	14.5 参考文献及进一步阅读	411
12.7.1 允许相等断言	352	14.6 习题	411
12.7.2 唯一名字假设	353	第五部分 宏观图景	
12.8 完备知识假设	355	第15章 回顾与展望	416
12.9 本章小结	358	15.1 复杂性维度回顾	416
12.10 参考文献及进一步阅读	358	15.2 社会与道德后果	418
12.11 习题	359	15.3 参考文献及进一步阅读	420
第13章 本体和基于知识的系统	363	附录A 数学基础与记号	421
13.1 知识共享	363	参考文献	425
		索引	439

计算机人（Agent）是人工智能领域的一个重要概念，它是指能够在环境中自主地感知、推理和行动的智能体。在人工智能领域，Agent 的概念已经得到了广泛的应用，从简单的机器人到复杂的智能系统，Agent 的概念无处不在。在本章中，我们将介绍 Agent 的基本概念、分类、设计原则以及如何创建它们。

第一部分

Artificial Intelligence: Foundations of Computational Agents

世界中的 Agent: 什么是 Agent 及如何创建它们

世界中的 Agent 1.1

Agent 是人工智能领域的一个重要概念，它是指能够在环境中自主地感知、推理和行动的智能体。在人工智能领域，Agent 的概念已经得到了广泛的应用，从简单的机器人到复杂的智能系统，Agent 的概念无处不在。在本章中，我们将介绍 Agent 的基本概念、分类、设计原则以及如何创建它们。

Agent 是人工智能领域的一个重要概念，它是指能够在环境中自主地感知、推理和行动的智能体。在人工智能领域，Agent 的概念已经得到了广泛的应用，从简单的机器人到复杂的智能系统，Agent 的概念无处不在。在本章中，我们将介绍 Agent 的基本概念、分类、设计原则以及如何创建它们。

世界中的 Agent 1.1

世界中的 Agent 1.1

世界中的 Agent 1.1

世界中的 Agent 1.1

人工智能与 Agent

人工智能的历史充满幻想、可能、验证和希望。自从荷马描绘机器“鼎”服侍在众神的餐桌旁，那想象中的机器佣人便成为我们文化的一部分。然而，我们人工智能的研究者，直到 50 年前，才首次制造出实验性机器来验证那些假想，即有关具备思维和智能行为机器人的假想，使得之前仅在理论上具备可能性的机器人得到验证。

——Bruce Buchanan [2005]

历经几个世纪的思想构建，人工智能学科被公认为有超过 50 年的历史，本书正是与此有关。如同以上 Buchanan 所指出的，我们现已验证这种思想假设，并用于解决实际问题。许多科学和工程问题已经得到解决，但更多其他问题亟待解决。纵然已存在人工智能的实际应用，仍有很多潜在应用等待挖掘。本书中，我们将揭示具有智能行为的计算 Agent 的基本原理，这将有助于大家理解人工智能现在和未来的工作，并有助于对这些原理做出进一步的贡献。

1.1 什么是人工智能

人工智能(AI)是一个研究能够产生智能行为的计算 Agent 的综合与分析的领域。下面阐述此定义的部分。

Agent 是指在某个环境中进行某种行为的个体——它完成了某件事。Agent 可以是蠕虫、狗、恒温器、飞机、机器人、人类、公司、国家中的任何一个。

我们感兴趣的是一个 Agent 做什么？怎么做？我们通过 Agent 的行为来辨别不同的 Agent。

当 Agent 具有以下行为时，我们说此 Agent 是智能的：

- Agent 的行为能够与其环境及目标相适应。
- 能够灵活地适用于改变的环境及目标。
- 能够从经验中学习。
- 给定感知和计算限制时能够做出恰当的选择。Agent 通常不能直接观察到整个世界，它只有有限的记忆，并且没有无限时间进行某个行为。

计算 Agent 可以通过一系列的计算对其行为进行决策。也就是说，决策可以分解成能够在硬件设备上实现的一系列原始操作。这种计算有很多形式，人类是通过一些软件完成的，而计算机则是由一些硬件完成的。尽管有一些 Agent 可以认为是不可计算的，例如自然现象中侵蚀景观的风和雨。但是，是否所有的 Agent 都是可计算的，仍是一个开放性问题。

人工智能的核心科学目标是理解自然系统或人工系统中一些智能行为的原理，主要通过以下几个部分来进行：

- 对自然 Agent 及人工 Agent 进行分析。
- 对如何构造 Agent 进行形式化定义，并进行验证。
- 设计、构建和实验计算系统，通过执行某些智能性任务来进行验证。