



“十二五”普通高等教育本科国家级规划教材
高等学校计算机教育规划教材

人工智能

● (第3版) ●

贲可荣 张彦铎 编著
陈志刚 主审

清华大学出版社





“十二五”普通高等教育本科国家级规划教材

高等学校计算机教育规划教材

人工智能

(第3版)

贲可荣 张彦铎 编著

清华大学出版社
北京

内 容 简 介

人工智能是研究理解和模拟人类智能、智能行为及其规律的一门学科,其主要任务是建立智能信息处理理论,进而设计可以展现某些近似于人类智能行为的计算系统。本书介绍人工智能的理论、方法和技术及其应用,除了讨论那些仍然有用的和有效的基本原理和方法之外,着重阐述一些新的和正在研究的人工智能方法与技术,特别是近期发展起来的方法和技术。此外,用比较多的篇幅论述人工智能的应用,包括新的应用研究。

本书包括下列内容:①简述人工智能的起源与发展,讨论人工智能的定义、人工智能与计算机的关系以及人工智能的研究和应用领域;②论述知识表示、推理和不确定推理的主要方法,包括谓词逻辑、产生式系统、语义网络、框架、知识图谱、归结推理、非单调推理、主观 Bayes 方法、确定性理论、证据理论、模糊逻辑和模糊推理等;③讨论常用搜索原理,如盲目搜索、启发式搜索、min-max 搜索、 α - β 剪枝和约束满足等,并研究一些比较高级的搜索技术,如贪婪局部搜索、局部剪枝搜索、模拟退火算法、遗传算法等;④介绍分布式人工智能与 Agent、计算智能、反向传播神经网络、深度学习、竞争网络支持向量化等已成为当前研究热点的人工智能技术和方法;⑤比较详细地分析人工智能的主要应用领域,涉及自动规划系统、自然语言处理、信息检索、语言翻译、语音识别、计算机视觉、群体智能机器人等。

本书适合作为高等学校计算机及相关专业大学高年级和非计算机专业研究生人工智能的教材,也可作为希望深入学习人工智能的科技人员的参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

人工智能/贲可荣,张彦铎编著. —3 版. —北京:清华大学出版社,2018
(高等学校计算机教育规划教材)

ISBN 978-7-302-51198-4

I. ①人… II. ①贲… ②张… III. ①人工智能—高等学校—教材 IV. ①TP18

中国版本图书馆 CIP 数据核字(2018)第 209918 号

责任编辑:张瑞庆 常建丽

封面设计:常雪影

责任校对:梁毅

责任印制:李红英

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:39.5

字 数:937 千字

版 次:2006 年 2 月第 1 版 2018 年 12 月第 3 版

印 次:2018 年 12 月第 1 次印刷

定 价:79.90 元

产品编号:076525-01

在当今社会，尤其是融合了社交内容的移动互联网的普及，如何更好地连接人与信息，已成为人类社会的一个重要基础命题。个性化的信息流已经成为一种新的连接方式，人与信息、万物互联。那么，在信息流产品平台与服务这个领域，如何高效地处理、分析、挖掘、理解和组织海量文字、图片（视频），更好地连接人与信息呢？如何根据对用户的深度理解，进行信息的智能推送呢？无疑，人工智能扮演着重要角色。

从内容创作、过滤、分发、消费以及互动的每个环节，都可以使用大规模机器学习，包括文本分析、自然语言理解、计算机视觉和数据挖掘等技术，向用户进行智能推送。同时，还可以基于信息流丰富多样的应用场景和用户，持续累积大量的训练样本和数据，让机器学习系统形成闭环，不断地改善和进化，在机器人辅助内容创作、自动视频分析与理解、个性化推荐和问答等方面发展人工智能核心技术。

移动互联网时代，很多信息都藏在应用里面，虽然不能利用搜索引擎将每个App里的信息轻松找出，但是在这股潮流中涌现出一些新的应用，让我们能够重新定义信息的源头。我们可以与很多信息供应商、内容提供商、媒体创作者一起构建新的内容平台和生态系统。

当前，有更多的公司开始大规模使用人工智能做个性化推荐。因为人们使用智能手机有了很多碎片化时间，产生了学习、娱乐等需求。这些需求也产生了各式各样的应用场景。在推荐引擎领域有了一个将人与信息相连接的新机会。搜索引擎里所有的排序算法、内容分析等技术，都可用于进一步的个性化精准推荐，从而变成信息流。“信息流”是一种新的、更智能的方式，让人能够随时随地在需要时得到所需要的信息。人工智能可以做个性化推荐，使人们不受地域限制享受服务——用无所不在的超级机器智能帮助人类创作、发现、使用、分发信息，并进行社交场景的互动。

当你使用在线系统搜索网页、编辑文档、存储图片、听音乐、看视频、玩游戏，并享受着行云流水般的顺畅服务时，正有几十万到上百万台服务器坚守在大后方，为你提供着7×24小时的可靠服务。超大的规模和超高的复杂度给服务的可靠性、可用性和性能都带来了极大的挑战。可靠服务的背后是人们利用人工智能前沿技术解决大规模在线系统服务的运维

问题,即利用大规模数据挖掘、机器学习等人工智能技术对纷繁复杂的运维大数据进行实时分析,为系统维护提供有效的决策方案。

随着技术的进一步发展,人工智能与人可以互相帮助,从而让彼此变得更聪明。人工智能需要很多标注数据和训练样本,在信息流的场景,人们有更多机会拿到更多标注数据以及更细颗粒度的标注,帮助人们做自然语言理解、自然语言生成、图像视频理解和图像视频生成。人将与人工智能进一步共同进化。

软件正在改变全世界,而软件产业本身正在被人工智能的发展所颠覆。越来越多的软件开发不再只是依靠软件工程师的想法、逻辑和认知,而这些软件的核心已变成非常大的模型,有上千亿的参数,有各式各样的大数据。通过训练各种各样的模型,包含统计模型、符号、逻辑、知识表达,软件产业已被人工智能化。

今天,视频、图像、文字都已经被数字化,下一个阶段就是语义化,如图像理解。在数字原始表达空间,计算机很难做语义理解,我们需要用深度学习模型来学习非线性的转化。今天人工智能的本质其实是软件产业的革命,借由大数据、大计算和机器学习训练大模型,“编写”越来越智能的软件。

人工智能主要分三层。最底层是基础架构(infrastructure),包括云计算、芯片以及TensorFlow这样的框架。中间层主要是使能技术(enabling technology),如图像识别、语音识别、语义理解、机器翻译等。基础层和中间层是互联网巨头的必争之地,如芯片领域,英特尔、英伟达、高通公司都投入巨资,竞争极其激烈。同样,云计算、框架也都不是小公司能够涉足的领地。

创业公司的机会在最上层,就是拿着下两层的成果服务垂直行业,也就是所谓的“AI+”。这样的趋势在2017年展现了不少案例。数据+AI算法正在带来更多的想象。例如,英国正在用AI酿造啤酒,瑞典通过深度学习分析马匹行为,纽约通过AI定制千人千面的素食健康食品。随着AI算法和机器学习更加民主化,每一个行业都可能进一步得到“AI+”式的改变。

深入垂直行业的“AI+”又可细分为两类情况:“AI+行业”和“行业+AI”,它们之间有明显的区别。

“AI+行业”在AI技术成熟之前,这个行业、产品从未存在过。例如,自动驾驶、亚马逊的Echo智能音箱、苹果的Siri语音助手。在人工智能技术未突破前,不存在这样的产品。因为AI创造出了一条全新的产业链。

“行业+AI”就是行业本身,一直存在,产业链条成熟,只是以前完全靠人工,效率比较低,现在加入AI元素后,使得行业效率有了明显提高,如安防、医疗等领域。

“AI+行业”“行业+AI”通常因为大家起跑线一样,行业纵深会比较浅,而后者则有巨大的行业壁垒。未来行业壁垒才是人工智能创业最大的护城河,因为每个行业都有垂直纵深。以医疗+AI举例,什么最重要?大量准确的被医生标注过的数据最重要。没有数据,再天才的科学家也无用武之地。

根据行业和应用场景不同,人工智能的创新应用可分为“关键性应用”和“非关键性应用”。“关键性应用”如手术机器人、无人车,要追求99.9%后的多个9,做不到就没法商业化。这类项目研发周期都很长。人工智能领域的创业,95%都是“非关键性应用”,如人脸识别门禁系统。“非关键性应用”不追求高大上,简单、实用、性

价比高更重要，这样的项目通常是比拼综合实力，包括对行业的洞察理解、产品和工程化能力、成本控制、供应链能力、营销能力。

人工智能的“智能”可以扩充为以下4个含义：第一部分是“聚合智能”，即不同类型智能的聚合。例如，语音和图形属于两个不同领域。然而，人类的智能往往是将各种感官获得的信息（如语音和图形）聚合，共同形成认知。微软小冰和微软学术搜索都是聚合智能的典型示例。第二部分是“自适应智能”，人工智能必须能够不断学习、与时俱进。我们在图像识别和智商测试方面的研究都很好地体现了自适应智能的特性。第三部分是“隐形智能”，目前的人工智能还只是在人需要的时候才被动开启，未来，我们更期待人工智能可以实现主动学习、主动服务。微软小娜的跨平台应用和视频分析技术未来能提供的诸多服务，也都让我们直观地感受到自动、无缝实现的隐形智能。最后是“增强智能”，就是用人工智能这个了不起的“左脑”，与人类的右脑相配合，充分利用人类才有的创造力延伸出无限可能。

为抢抓人工智能发展的重大战略机遇，构筑我国人工智能发展的先发优势，加快建设创新型国家和世界科技强国，2017年7月，国务院颁发了《新一代人工智能发展规划》。

经过60多年的演进，特别是在移动互联网、大数据、超级计算、传感网、脑科学等新理论、新技术以及经济社会发展强烈需求的共同驱动下，人工智能加速发展，呈现出深度学习、跨界融合、人机协同、群智开放、自主操控等新特征。大数据驱动知识学习、跨媒体协同处理、人机协同增强智能、群体集成智能、自主智能系统成为人工智能的发展重点，受脑科学研究成果启发的类脑智能蓄势待发，芯片化、硬件化、平台化趋势更加明显，人工智能发展进入新阶段。当前，新一代人工智能相关学科发展、理论建模、技术创新、软硬件升级等整体推进，正在引发链式突破，推动经济社会各领域从数字化、网络化向智能化加速跃升。

《新一代人工智能发展规划》提出以下三步走策略。

到2020年，人工智能总体技术和应用与世界先进水平同步，新一代人工智能理论和技术取得重要进展。大数据智能、跨媒体智能、群体智能、混合增强智能、自主智能系统等基础理论和核心技术实现重要进展，人工智能模型方法、核心器件、高端设备和基础软件等方面取得标志性成果。

到2025年，人工智能基础理论实现重大突破，部分技术与应用达到世界领先水平，人工智能成为带动我国产业升级和经济转型的主要动力，智能社会建设取得积极进展。

到2030年，人工智能理论、技术与应用总体达到世界领先水平，智能经济、智能社会取得明显成效。在类脑智能、自主智能、混合智能和群体智能等领域取得重大突破。

本书力图体现如下3方面特点。

一是“基础性”。包括智能感知、智能推理、智能学习和智能行动所涉及的基本概念、基础理论、基本方法，并通过简单实例加深读者对概念的理解。内容涵盖中国计算机学会（CCF）术语百科平台人工智能分支的各个方面：人工智能逻辑、自动推理、多智能体系统、机器学习、神经网络、智能机器人、模式识别、计算语言学、知识工程，参见CCFPedia（<http://term.ccf.org.cn/>）和手机版（<http://term.ccf.org.cn/mobile>）。

二是“实践性”。案例具有实践性、情境性,知识、概念、理论在教材中详略适当地建构起体系后,就要通过“湿漉漉”的案例让读者有找到感觉、投身其中的可能。本书案例包括饭店脚本、在互联网金融行业中的应用、知识图谱应用、九宫图、传教士-野人问题、五子棋、洞穴探宝、八皇后、旅行商问题、帆船分类专家系统、抵押申请评估决策支持系统、火星探矿机器人、通过 EBG 学习概念 cup、基于反向传播网络拟合曲线、深度学习在计算机视觉中的应用、学习向量量化解决分类问题、XOR 问题、音节划分、深度学习在语音识别中的应用、规划问题的建模与规划系统的求解过程、Shakey 世界、仿真机器人运动控制算法、反恐作战数据挖掘、微博博主特征行为数据挖掘、智能网联汽车、城市计算等。每一章后的思考题都为学生课后实践提供了拓展空间。

三是“前瞻性”。近几年,人工智能蓬勃发展,基础性、应用性成果很多。在介绍智能感知、智能推理、智能学习和智能行动应用进展时,体现了最新研究成果。此外,1.5节“人工智能发展展望”中包括新一轮人工智能发展特征、未来40年的人工智能问题、人工智能鲁棒性和伦理、新一代人工智能发展规划。11.5节“集体智能”中包括社群智能、集体智能系统、全球脑、云脑、物联网等前瞻理论和应用场景。

第2版附录A“人工智能程序设计语言 Prolog”在本版调整为附录A“人工智能编程语言 Python”和附录B“手写体识别案例”。在教学中,如果需要继续使用 Prolog 相关资料,可取阅清华大学出版社提供的相应电子版资源。

第2版附录B“人工智能大作业”,在本书中做了如下处理:部分问题在正文中以案例的方式给出了解答,如九宫图(第2章)、传教士-野人问题(第2章)、五子棋(第3章)、洞穴探宝(第3章)、八皇后(第3章)、音节划分(第8章)。部分问题以思考题的形式放到各章习题中,如NIM问题(第1章)、人工智能军事应用跟踪(第1章)、计算机游戏如何产生娱乐效果(第1章);水壶问题(第2章)、魔方(第2章)、奥木(第2章)、合一算法(第2章)、归类测试算法(第2章);中国象棋(第3章)、围棋(第3章);小型动物分类专家系统(第5章);火星探测器 Agent(第6章)、用于电力管理的多 Agent 系统(第6章);用神经网络对大写字母分类(第7章);美国地理(第8章);Elsevier 的横向信息产品(第11章)、奥迪的数据整合(第11章)、人寿保险公司的技能寻获(第11章)、在线学习(第11章)、警察局的多媒体收藏索引(第11章)、康富的在线采购(第11章)、数码设备的可共用性(第11章)。

本书第3、4、9、10章由张彦铎撰写,毛新军参与第6章撰写,其余各章由贲可荣撰写。全书由贲可荣统稿。蔡敦波参与了第3、4、9、10章的修订,何智勇撰写了“附录A 人工智能编程语言 Python”和“附录B 手写体识别案例”,张毅伟参与了绘图工作,张献参与了审校,陈志刚教授对全书进行了审校,清华大学出版社张瑞庆编审推动和指导了本书撰写,在此一并致谢。

感谢所有参考文献的作者,感谢“中国计算机学会通讯”“中国人工智能学会通讯”、中国计算机学会“技术动态”相关论文作者。

贲可荣

2018年6月

智能与意识、思维和记忆以及问题求解、直觉、创造性、语言和学习有关，而且还与理解和感觉，如运动技能、预测环境能力、处理复杂世界的的能力、在学校和 IQ 测试的表现等有关。

智能与人工智能

1. 智能的定义

智能的定义有多种，如“智能是进行抽象思维的能力”“智能是学习或者从经验中获益的能力”“智能是学习到的或者学习以调整自己适应环境的能力”“智能是充分调整自己去适应生命中的相对较新的情况”“智能是接受知识的能力以及所拥有的知识”等，但要理解这些定义，最好的方法是通过问题体会。

通过寻找感兴趣的问题推进研究（例如，狗如何会跑并且抓住飞盘；为什么老鼠很快学会在迷宫中寻找方向；蚂蚁在寻找食物的长途旅行之后如何寻找到回到巢穴的路；人如何走路；人如何在一堆人中认出一张面孔），然后尝试了解这种特定的行为如何产生。是否愿意把这些行为称作智能，依赖于个人偏好。

我们在直觉上认为的智能，总是包括两个特殊性质：顺应性和多样性。简言之，主体（Agent）总是遵从其所处环境的自然和社会规则，利用这些规则去产生多样行为。例如，所有动物、人类以及机器人必须遵从存在重力和摩擦以及移动需要能量的事实，绝不可能完全摆脱这些。但是，适应并以一定方式利用这些限制条件使我们能够走路、奔跑、从杯子中喝水、将盘子放到桌子上、踢足球或骑自行车。多样性意味着主体能够进行多种行为，这样他/它能够对给定的情况做出适当反应。一个仅走路，或仅下棋，或仅奔跑的主体在直觉上比起能用积木搭出玩具汽车、把啤酒倒进杯子和在一群挑剔的听众面前讲课的主体缺少智能。在智能的许多定义中被提到的学习是一种随时间而增加行为多样性的有效手段。

直觉上，我们认为一些行为比另外一些需要更多的思考，一些动物比另外一些更聪明。在一般用法中，思考经常与有意识的思想相联系；认知

则更一般化,用于同感觉-运动过程非直接相关的行为中;智能则更加一般化,包含各种对主体有益的认知和思考等抽象行为。智能是先天的还是后天的,即智能在多大程度上来自遗传或是在一生中所获取的?可以用智能区分人类和其他物种吗?

2. 人工智能的研究内容

人工智能研究的是智能行为中的机制,它是通过构造和评估那些试图采用这些机制的人工制品进行研究的。在这个定义中,人工智能不像是关于智能机制的理论,更像是一种经验主义的方法学,它的主要任务是构造和测试支持这种理论的可能模型。它是一种对实验进行设计、运行和评估的科学方法,其目的是精练模型和进行更深入的实验。

人工智能的经验主义的方法学是一个重要的工具,也许它对于探索智能的本质来说是最好的工具。

人工智能属于交叉学科研究领域,本质上具有3个目标:①了解生物系统(也就是引起人类或动物智能行为的机制);②智能行为一般原则的抽象提取;③应用这些原则设计有用的人造物。这里的机制不仅是指神经机制或者大脑过程,也指主体的身体及其同真实世界的交互。如肌肉具有弹性,当一条腿抬起时,另一条腿承受的重量增加的事实,和与步行紧密相关的反射和大脑中枢一样是步行机制的有机组成。

如果人工智能想达到科学的水平并成为智能系统科学的关键组成部分,就必须在它制造的人工制品的设计、执行和分析中包含分析和经验式的方法。从这种观点看,每个人工智能程序都可以看作是一个实验:它向现实世界提出问题,而答案就是现实世界对此做出的响应。现实世界对我们的设计做出的响应和程序式的承诺构成了我们对于智能的形式方法、机理以及智能本质的理解。

3. 智能的计算特性

智能的计算特性开始于对计算设备的抽象规范说明。20世纪30年代到50年代的研究开启了这一探索,Turing(图灵)、Post(波斯特)、Markov(马尔可夫)、Church(丘奇)等人在对计算的形式化描述方面做出了极大贡献。这些研究的目的是不仅指出计算的含义,还指出关于可计算的界限。通用图灵机是大家熟知的规约,波斯特重写规则可作为产生式系统计算基础。基于部分递归函数的丘奇模型,支持诸如Scheme、Ocaml和Standard ML等现代高级函数式语言。

所有这些形式化方法都具有等价的计算能力。可以说,通用图灵机等价于现代的任意计算设备。没有哪个计算模型能够定义得比这些已知模型更强(丘奇-图灵命题)。一旦建立了计算规约的等价性,我们就从这些规约的机械化工具中解放出来:我们可以用电子管、硅芯片、细胞质或者普通玩具实现我们的算法。在一种媒介上的自动设计机制等价于在另外一种媒介上设计的机制。因此,当我们在一种媒介上测试在另一种媒介上实现的机制时,就使得经验式探索的方法变得更加重要。

有一种可能就是,图灵和波斯特的通用机器也许太泛化、太通用了。这里矛盾的是,智能可能并不需要很强的带集中控制的计算机制。Levesque等人建议人的智能可能需要更多的计算性的有效的表示(如用于推理的Horn子句)、对基文字的实际知识的约束以及可计算跟踪的真值维护系统的使用。智能的基于主体的模型和涌现模型似乎也支持这种观点。

由我们的机制模型的形式化等价性引出的另一个论点是，二元性问题和心身（mind-body）问题。笛卡儿时代以后，哲学家们就提出了智能、意识和身体之间的交互和整合问题。他们给出了每种可能的反映，从完全的唯物主义到对物质存在的否定，甚至到支持上帝的介入！人工智能和认知科学的研究否认了笛卡儿的二元论，而支持基于物理实现或者符号实例的物质的物质模型，支持管理这些符号的计算机制的形式化规约，支持表示范例的等价性，支持在具体模型中知识和技能的机械化。这种研究的成功表明了这种模型的有效性。

关于物理系统中智能的认识论基础

1. 表示问题

Allen Newell 和 Herbert Simon 假定物理符号系统和搜索对于智能的特性是充分必要的。神经模型或子符号模型的成功、智能的遗传和涌现方法的成功是否是对物理符号假设的一种驳斥，或者它们是这种假设的简单实例吗？

连这种假设的弱解释——物理符号系统是智能的一个充分模型——在现代认知科学领域中也产生了许多强大的、有用的结果。这种观点认为，我们可以实现那些能说明智能行为的物理符号系统。充分性使得我们能够为人所具有的许多方面的性能创建和测试基于符号的模型。但是，这个假设的强解释——物理符号系统和搜索对于智能活动是必要的——仍然是一个有待研究和解决的问题。

2. 认知中具体化的作用

物理符号系统假设的主要假定之一就是，物理符号系统的特定实例化是与其性能无关的；其主要内容是其形式化结构。许多研究者都对这一点提出了挑战，他们指出智能行为的需求要求一种允许主体整合到世界中的物理具体化。现代计算机的结构并不允许这种程度的情形，而是要求一个人工智能通过极端有限的窗口（同时代的输入输出设备）同世界进行交互。如果这种挑战是正确的，则尽管出现机器智能，它仍需要同时代的计算机提供一个非常相同的接口。

3. 文化与智能

传统上，人工智能侧重于把个体智能作为智能的来源；我们的行动好像在说，对于大脑编码和怎样管理知识的方法的解释可能是原始智能的一种完整解释。然而，我们也会认为知识最好被看作是基于一社会的，而不是一个个体所构造的。在基于记忆的智能理论中，社会本身也带有智能的本质组件。对于智能理论来说，对知识的社会环境和人类行为的理解，同对个体智能/大脑的理解是同等重要的。

4. 刻画解释的本质

在表示传统研究中，大多数模型一般工作在已经解释好的领域中，即对于解释的上下文，系统设计者通常都会给出一些隐含的、先验约定，在这种约定下，很难随着问题求解过程的进展而将上下文、目标或表示进行转换。目前还很少有成果能够阐明人类构造解释的过程。

Tarskian 的观点“将语义作为符号和对象之间的映射”还是太弱并不能解释一些事实，如一个领域在不同实践目标的指引下可能会有不同的解释。语言学家试图通过语

用理论弥补 Tarskian 语义的局限性。论述分析基本依赖于上下文中符号的使用,已经在近几年中广泛地讨论了这些问题,但是,这个问题涉及的内容事实上还要更加广泛,因为它通常还要处理参考工具的失败。

C. S. Peirce 最先倡导符号语言学的传统,后续的研究者还有 Eco、Sebeok 以及其他学者,他们对于语言采用了更激进的方法。这种符号语言学的传统把符号表达式放在广泛的记号和记号解释中,它表明,符号的含义只有在它用作解释的上下文中才能够被理解,即在解释的上下文中或在与环境的交互中才能被理解。

5. 表示的不确定性

Anderson 的表示不确定性猜想指出,在熟练性能的特定动作这种环境下,确定哪种表示模式最接近于人的问题求解器在理论上是不可能的。这种猜想是基于这样的事实,即每个表示模式不可避免地被连接到一个大型的计算结构,就像搜索策略一样。在对人类技能的详细分析中,我们不可能充分控制这个过程,使得我们能决定这个表示;也不可能为过程被唯一确定的那些点建立一个表示。由物理的不确定原理,现象可以通过检验这个过程加以改变,因此,构造智能模型是需要重点关注的,但没有必要限制它们的利用。

6. 设计可以反证的计算模型的必要性

Popper 等人指出科学理论必定是可以反证的,这就是说,必定存在一种环境,使得在此环境下的这个理论模型并不是对这种现象的成功近似。任何数目的确定性实验实例都不能充分地确定一个模型。许多研究是在已有理论的失败的基础上进行的。

物理符号系统假设的一般本质正如智能的情景模型和涌现模型一样,作为一种模型在使用上受到限制。同样,可以对关于现象学传统的假设进行批评。一些人工智能数据结构(如语义网)还是很普通的,使得它们可以建模几乎所有能够描述出的东西,或者正像通用图灵机一样,使得它们可以建模任意的可计算函数。因此,一个人工智能研究者或者认知科学家被问到在什么条件下他们的智能模型不能用时,给出答案经常是很难的。

7. 科学方法的局限性

许多研究者们宣称智能的最重要方面就是没有被模型化,并且原则上不可能被模型化,且特别是不能使用任意的符号表示来模型化。这些领域包括学习、自然语言理解、说话动作的产生等。这些问题已经深深地植根于我们的哲学理念中。

现代人工智能的大部分假设追其根源,可以回溯到 Carnap、Frege 和 Leibniz,再远回溯到 Hobbes、Locke 和 Hume,直至回溯到亚里士多德。这种传统观点认为,智能的处理过程符合通用法则,并且在原则上是可以被理解的。

Heidegger 和他的追随者们描述了一种可选择的方法理解智能。对于 Heidegger 来说,思考的意识源于具体经验的世界(一种生命世界)。Winograd、Flores 和 Dreyfus 等人认为一个人对事物的理解是扎根于在每天的世界中“使用”这些理解的实际活动中。这种世界在本质上是一种环境,其中包括按社会方式组织的各种作用和目的。而这种环境以及其中的人的功能不是通过命题解释的,也不是能够被定理所理解的。它更像是一种不断形成的流程。在基本意义上,人类专家并不知道“是什么”,而只是知道在进化的社会标准和隐含的目的不断发展的世界中,它是怎么样。我们不能自

然地就把我们的知识和大多数智能行为放入语言中，不管是形式的，还是自然的。

现在让我们来考虑上述这种观点。首先，作为对纯理性主义传统的批判，这种观点是正确的。理性主义者断言，所有的人类活动、智能和责任，至少原则上能够被表示、形式化和理解。大多数喜欢思考的人们并不相信这种情形，他们认为情感、自我主张和有责任的承诺等（至少）也是很重要的。在科学方法的领域之外，还有很多人人类活动在可靠的人类交互中起着本质的作用。这些不可能被机器再生或者取消。

然而，检查数据、构造模型、运行实验以及为了进一步实验而使用模型精练检查结果等这些科学传统已经进入理解、解释和预言人类社会能力这样一个重要的层次。科学方法是提高人类理解能力的一个有力工具。尽管如此，对于这种方法，这里仍然有许多的告诫是研究人员必须理解的。

首先，研究人员不要把这个模型与被建模的现象相混淆。模型能允许我们不断地逼近这种现象；通常，这里必然有一些不能使用经验解释的“残留物”。在这种意义上，表示不确定性并不是一个问题。一个模型是用来探索、解释和预言的；如果它允许研究人员完成这些任务，则它就是成功的。对于一种简单的现象，不同的模型可以用来解释这种现象的不同方面。

此外，当研究者们主张智能现象的各个方面已经在科学传统的范围和方法之外时，这种说法本身也只能用那些科学传统验证。科学方法只是一种工具，它可用来解释在什么意义上问题仍然是在我们当前的理解之外。每种观点，甚至是来自于现象学传统的观点，如果它是有一定含义的，那么它一定跟我们当前某些解释的概念相关，甚至它是与那些不能解释的现象相关联的。

人工智能研究中最让人振奋的方面是对我们必须解决的这些问题做出不懈的努力和贡献。为了解决问题求解、学习和语言，必须领会表示和知识的哲学层面含义。我们被要求用一种谦卑的方式解决亚里士多德的理论和实践之间的关系问题，以形成理解和实践的统一、理论和实践的统一，在科学与艺术中生活。

人工智能工作者是工具的制造者。我们的表示、算法和语言都是一些工具，用来设计和建立那些展现智能行为的机制。通过实验，我们同时检验了它们解决问题的计算适合性，也检验了我们自己对智能现象的理解。

智能系统

人工智能研究的一个最重要的动力是建立智能系统，以求解困难问题。20世纪80年代以来，知识工程成为人工智能应用最显著的特点，专家系统、知识库系统、智能决策系统等智能系统得到广泛应用。知识库系统是把知识以一定的结构存入计算机，进行知识的管理和问题求解，实现知识的共享。

建造智能系统可以模仿、延伸和扩展人的智能，实现某些“机器思维”，具有极大的理论意义和实用价值。根据智能系统具有的知识和处理范型的情况，可以分成4类：①单领域知识单处理范型智能系统；②多领域知识单处理范型智能系统；③单领域知识多处理范型智能系统；④多领域知识多处理范型智能系统。

1. 单领域知识单处理范型智能系统

系统具有单一领域的知识,并且只有一种处理范型。例如,第一代、第二代专家系统和智能控制系统都属于这种类型。

专家系统是运用特定领域的专门知识,通过推理模拟通常由人类专家才能解决的各种复杂的、具体的问题,达到与专家具有同等解决问题能力的计算机智能程序系统。它能对决策的过程作出解释,并有学习功能,即能自动增长解决所需的知识。第一代专家系统(如 DENDRAL、MACSYMA 等)以高度专业化、求解专门问题的能力为特点,但在体系结构的完整性、可移植性等方面存在缺陷,求解问题的能力弱。第二代专家系统(如 MYCIN、CASNET、PROSPECTOR、HEAR-SAY 等)属单学科专业型、应用型系统,其体系结构较完整,移植性方面也有所改善,而且在系统的人机接口、解释机制、知识获取技术、不确定推理技术、增强专家系统的知识表示和推理方法的启发性、通用性等方面都有所改进。

2. 多领域知识单处理范型智能系统

多领域知识单处理范型智能系统具有多种领域的知识,而处理范型只有一种。大多数分布式问题求解系统、多专家系统都属于这种类型。一般采用专家系统开发工具和环境研制这种大型综合智能系统。

由于智能系统在工程技术、社会经济、国防建设、生态环境等各个领域的广泛应用,对智能系统的功能提出多方面的要求。许多实际问题的求解,例如,医学诊治、经济计划、军事指挥、金融工程、作物栽培、环境保护等,往往需要应用多学科、多专业的专家知识和经验。现有的许多专家系统大多数是单学科、专门性的小型专家系统,不能满足用户的实际需求。建立多领域知识单处理范型智能系统在一定程度上可以达到用户的要求。

这类智能系统的特点是:面向用户实际的复杂问题求解;应用多学科、多专业、多专家的知识 and 经验,进行并行协同求解;基于分布式、开放性软硬件和网络环境;利用专家系统开发工具和环境;实现知识共享与知识重用。

3. 单领域知识多处理范型智能系统

单领域知识多处理范型智能系统具有单一领域的知识,而处理范型有多种。例如,混合智能系统属于这种类型。一般可以用神经网络通过训练获得知识,然后转换成产生式规则提供给推理机在求解问题时使用。

在进行问题求解时,也可以采用多种机制处理同一个问题。例如,疾病诊断系统既可采用符号推理的方法,也可通过人工神经网络让它们同时处理相同的问题,然后比较它们的结果,这样容易取得正确的结果,避免片面性。

4. 多领域知识多处理范型智能系统

这种系统具有多种领域的知识,而且处理范型也有多种。这种系统包含一集体智能模块,其含义是,在多种处理范型的环境下,各种处理机制各行其是,各司其职,协调工作,表现为集体的智能行为。

综合决策系统、综合知识系统属于多领域知识多处理范型智能系统。在这种系统中,基于推理的抽象思维采用符号处理的方法;而基于模式识别、图像处理之类的形象思维采用神经计算。

在总结和分析已有智能系统的设计方法和实现技术的基础上,采用智能主体(Agent)技术,实现具有多种知识表示、综合知识库、自组织协同工作、自动知识获取等功能的大型综合智能系统。这类系统是当前实现多领域知识多处理范型智能系统的主要途径。

现实世界的问题多数具有病态结构,研究的对象也在不断变化,很难找到一种精确的算法进行求解。构造人机统一,与环境进行交互、反馈的开放系统是解决这类智能问题的途径。所谓开放系统,是指系统在操作过程中永远有难以预料的后果,并能在任何时候从外部接收新的信息。

互联网已经成为各类信息资源的聚集地。在这些海量的、异构的Web信息资源中,蕴涵着具有巨大潜在价值的知识。通过Web内容发现、结构发现、使用发现等,能够从Web上快速、有效地发现资源和知识,提高在Web上检索信息、利用信息的效率。互联网上的维基百科反映了集体智能的特点和优势,受到社会人士的欢迎和好评。

展望智能科学与技术

智能科学与技术本身的发展正在向理论创新的深入和大规模实际应用发展。在2009年中国科学技术协会公布的“10项引领未来的科学技术”评选结果中,作为智能科学与技术核心的“人工智能技术”排在第4位,作为智能科学与技术重要应用的“未来家庭服务机器人”排在第2位,这充分显示了智能科学与技术的巨大潜力和极其广泛的社会影响。

在过去50多年的人工智能研究中,人们一直沿着“模拟脑”的方向做出努力,分别从智能系统的结构、功能、行为3个基本侧面展开对智能的研究,这样便先后形成了模拟大脑抽象思维功能的符号智能学说、模拟大脑结构的神经网络学说以及模拟智能系统行为的感知-动作系统学说。由于社会的迫切需要,呼唤着智能科学与技术理论上取得突破,在应用上广泛普及。展望智能科学与技术的发展,可以开展以下3方面的研究。

(1) 智能科学。智能科学是脑科学、认知科学、人工智能等的交叉学科,研究智能的理论和技術。智能科学不仅要进行智能的功能仿真,而且要研究智能的机理。脑科学从分子层次、细胞层次、行为层次研究自然智能机理,建立脑模型,揭示人脑的本质。认知科学是研究人类感知、学习、记忆、思维、意识等人脑心智活动过程的科学。人工智能研究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。

(2) 互联网智能。互联网为智能科学与技术提供了重要的研究、普及和应用平台。作为知识处理和智能行为交互的基本环境,今天的互网络最丰富的就是信息,最缺乏的就是智能。如何为在海量信息面前无所适从的用户提供有效的检索手段,如何剔除有害的、无用的垃圾邮件,如何使远方的机器人成为你放心的智能代理,都对网络信息的智能化提出迫切的需求,也是智能科学与技术发展的巨大动力。基于互联网的集体智能,通过大规模协作、综合集成,将为科学决策提供有效的途径。

(3) 智能机器人。智能机器人是将体力劳动和智力劳动高度结合的产物。智能

机器人是一种具有智能的、高度灵活性的、自动化的机器,具备感知、规划、动作、协同能力,是多种高新技术的集成体。

欧盟委员会2011年5月评出了对未来影响最大的6项前沿技术。这6大技术目前还处于研究阶段,前3项分别是:未来信息分析模拟技术、石墨烯技术和纳米级传感器技术。后3项与人工智能有关,分别是:

“人脑工程技术”。这一技术可用于对人脑的低能耗、高效率进行研究。人脑的学习功能、联想功能、创新功能都是目前计算机不具备的。另外,具有如此巨大功能的人脑又是节能减排的典范,它的功耗只有20~30W,相当于一盏白炽灯。人脑的这些神奇之处一旦被破解,将为信息技术研发提供借鉴。

“医学信息技术”。有关研究旨在推动信息技术在医药领域的大规模应用。此类技术还将对海量传输健康信息、利用人工智能技术处理这些信息并做出个性化治疗方案提出新要求。

“伴侣型机器人”开发。这一项目旨在研制具有一定感知、交流和情感表达能力的仿真机器人,为人类(特别是小孩和老人)提供无微不至的服务。这一项目将有两大亮点:一是依靠先进的人工智能技术,使机器人初步具有像人一样的感知、交流和情感表达能力;二是开发出制造机器人的新材料,可以让机器人看起来、摸起来像真人一样。

中国计算机学会《技术动态》评出“2011年度十大新闻”,其中3条与人工智能有关:“脑神经元连接同步定位首获成功”“IBM成功构建模拟人脑功能的认知计算机芯片”和“美研发光子神经元系统引计算机速度革命”。

本书是在《人工智能》(2006版)的基础上,吸取了国内外人工智能教材的优点,增补了国际上最新的研究成果修订而成。参考史忠植教授的教材,新增一章“互联网智能”,包括语义网与本体、Web技术、Web挖掘和集体智能等内容。结合教学实际,以附录形式增加了“人工智能大作业”,包括28个问题(选题),并明确了大作业组织形式及要求。

本书第3、4、9、10章由张彦铎撰写,其余各章由贲可荣撰写。全书由贲可荣统稿。蔡敦波对第3章进行了修订,郑笛参与了第4、10章的修订,陈志刚教授对全书进行了认真审校,特此致谢。

感谢所有参考文献的作者,感谢“中国计算机学会通讯”、中国计算机学会“技术动态”部分论文的作者。

贲可荣

2012年8月

智能 (intelligence) 是人类所特有的区别于一般生物的主要特征。智能解释为“感知、学习、理解、知道的能力, 思维的能力”。智能通常被理解为“人认识客观事物并运用知识解决实际问题的能力……往往通过观察、记忆、想象、思维、判断等表现出来”。

人工智能就是用计算机模拟人的智能, 因此又叫作机器智能。研究人工智能的目的, 一方面是要造出具有智能的机器, 另一方面是要弄清人类智能的本质。通过研究和开发人工智能, 可以辅助、部分代替, 甚至拓宽人的智能, 使计算机更好地造福于人类。

信息经抽象结晶为知识, 知识构成智能的基础。因此, 信息化到知识化再到智能化, 必将成为人类社会发展的趋势。人工智能已经并正在广泛而深入地结合到科学技术的各门学科和社会的各个领域中, 它的概念、方法和技术正在各行各业广泛渗透。智能已成为当今各种新产品、新装备的发展方向。

随着人工智能学科的发展, 课程的内容也要不断更新。在美国, 由 IEEE Computer Society 和 ACM 计算教程联合工作组共同制订了《计算教程 2001》(*Computing Curricula 2001*, CC2001), 它主要修订了 CC1991, 以反映计算机领域十余年来发展。从 CC2001 可以看出, 人工智能课程除包括人工智能概论、问题状态与搜索、知识表示、机器人学等传统部分外, 还增加了机器学习、智能体、自然语言处理、语音处理、知识库系统、神经网络、遗传算法等内容。这充分反映了 CC2001 对人工智能课程的重视。在我国, 从 20 世纪 70 年代末开始, 随着改革开放政策的实施, 人工智能的教学和科研逐步展开。

本书介绍人工智能的理论、方法和技术及其应用, 除了讨论那些仍然有用的和有效的基本原理和方法外, 着重阐述一些新的和正在研究的人工智能方法与技术, 特别是近期发展起来的方法和技术。此外, 用比较多的篇幅论述人工智能的应用, 包括人工智能新的应用研究。具体包括下列内容。

(1) 简述人工智能的起源与发展, 讨论人工智能的定义、人工智能与计算机的关系以及人工智能的研究和应用领域。

(2) 论述知识表示、推理和不确定推理的主要方法,包括谓词逻辑、产生式系统、语义网络、框架、面向对象、归结推理、非单调推理、主观 Bayes 方法、确定性理论、证据理论、模糊逻辑和模糊推理等。

(3) 讨论常用搜索原理,如盲目搜索、启发式搜索、minimax 搜索、 α - β 剪枝和约束满足等,并研究一些比较高级的搜索技术,如贪婪局部搜索、局部剪枝搜索、模拟退火算法、遗传算法等。

(4) 介绍近期发展起来的已成为当前研究热点的人工智能技术和方法,即分布式人工智能与 Agent、计算智能、机器学习、反向传播神经网络、Hopfield 神经网络、知识发现等。

(5) 比较详细地分析人工智能的主要应用领域,涉及自动规划系统、自然语言处理、信息检索、语言翻译、语音识别、机器人等。

本书第 3、4、9、10 章由张彦铎撰写,其余各章由贲可荣撰写。全书由贲可荣统稿。吴荣华撰写了附录初稿。陈志刚教授对全书进行了认真审校,特此致谢。

在本书编写过程中,参考和引用了许多专家、学者的著作和论文,正文中未一一注明。在此,作者谨向相关参考文献的作者表示衷心的感谢。

不当之处,恳请读者批评指正。

作者

2006 年 1 月