



智能科学技术著作丛书

面向智能体的知识工程

程显毅 刘一松 晏立 编著



科学出版社

www.sciencep.com

TP182
664
1-

智能科学技术著作丛书

面向智能体的知识工程

程显毅 刘一松 晏立 编著

科学出版社

北京

内 容 简 介

本书是从代表未来计算模式——智能体的视角论述知识工程的基本概念、方法和技术。全书共分 11 章,第 1 章论述知识工程问题的提出;第 2 章给出智能体的评价标准——理性;第 3、4、7、8 章分别讨论了智能体工作的四种策略:目标、逻辑、效用、学习;第 5 章讨论了智能体工作的基础知识表示——本体;第 6 章讨论了常识及其推理;第 9 章和第 10 章详细介绍了知识系统开发环境 Prolog 及知识系统应用案例;第 11 章讨论了多智能体及其通信问题。

本书既是一本专著,也可作为高等学校电子信息、自动化、机电工程、计算机及其他相关专业研究生和本科高年级知识工程或人工智能的课程教材,还可供从事知识系统教学、研究、开发和应用的科技工作者参考。

图书在版编目(CIP)数据

面向智能体的知识工程/程显毅,刘一松,晏立编著.—北京:科学出版社,2008

(智能科学技术著作丛书)

ISBN 978-7-03-022737-9

I. 面… II. ①程…②刘…③晏… III. 知识工程 IV. TP182

中国版本图书馆 CIP 数据核字(2008)第 120979 号

责任编辑:张海娜/责任校对:张 琪

责任印制:刘士平/封面设计:陈 敬

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>

骏杰印刷厂印刷

科学出版社发行 各地新华书店经销

*

2008 年 9 月第 一 版 开本: B5(720×1000)

2008 年 9 月第一次印刷 印张: 17 3/4

印数: 1—3 000 字数: 353 000

定价: 40.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

《智能科学技术著作丛书》编委会

名誉主编：吴文俊

主 编：涂序彦

副 主 编：钟义信 史忠植 何华灿 蔡自兴 孙增圻 谭 民

秘 书 长：韩力群

编 委：（按姓氏汉语拼音排序）

蔡庆生（中国科学技术大学）

杜军平（北京邮电大学）

何华灿（西北工业大学）

黄河燕（中国科学院计算语言研究所）

焦李成（西安电子科技大学）

刘 宏（北京大学）

秦世引（北京航空航天大学）

阮秋琦（北京交通大学）

孙增圻（清华大学）

涂序彦（北京科技大学）

王家钦（清华大学）

吴文俊（中国科学院系统科学研究所）

尹怡欣（北京科技大学）

张琴珠（华东师范大学）

庄越挺（浙江大学）

蔡自兴（中南大学）

韩力群（北京工商大学）

何 清（中国科学院计算技术研究所）

黄心汉（华中科技大学）

李祖枢（重庆大学）

刘 清（南昌大学）

邱玉辉（西南师范大学）

史忠植（中国科学院计算技术研究所）

谭 民（中国科学院自动化研究所）

王国胤（重庆邮电学院）

王万森（首都师范大学）

杨义先（北京邮电大学）

于洪珍（中国矿业大学）

钟义信（北京邮电大学）

《智能科学技术著作丛书》序

“智能”是“信息”的精彩结晶，“智能科学技术”是“信息科学技术”的辉煌篇章，“智能化”是“信息化”发展的新动向、新阶段。

“智能科学技术”(intelligence science & technology, IST)是关于“广义智能”的理论方法和应用技术的综合性科学技术领域,其研究对象包括:

- “自然智能”(natural intelligence, NI),包括:“人的智能”(human intelligence, HI)及其他“生物智能”(biological intelligence, BI)。

- “人工智能”(artificial intelligence, AI),包括:“机器智能”(machine intelligence, MI)与“智能机器”(intelligent machine, IM)。

- “集成智能”(integrated intelligence, II),即:“人的智能”与“机器智能”人机互补的集成智能。

- “协同智能”(cooperative intelligence, CI),指:“个体智能”相互协调共生的群体协同智能。

- “分布智能”(distributed intelligence, DI),如:广域信息网,分散大系统的分布式智能。

1956年,“人工智能”学科诞生,50年来,在起伏、曲折的科学征途上不断前进、发展,从狭义人工智能走向广义人工智能,从个体人工智能到群体人工智能,从集中式人工智能到分布式人工智能,在理论方法研究和应用技术开发方面都取得了重大进展。如果说,当年“人工智能”学科的诞生是生物科学技术与信息科学技术、系统科学技术的一次成功的结合,那么,可以认为,现在“智能科学技术”领域的兴起是在信息化、网络化时代又一次新的多学科交融。

1981年,“中国人工智能学会”(Chinese Association for Artificial Intelligence, CAAI)正式成立,25年来,从艰苦创业到成长壮大,从学习跟踪到自主研发,团结我国广大学者,在“人工智能”的研究开发及应用方面取得了显著的进展,促进了“智能科学技术”的发展。在华夏文化与东方哲学影响下,我国智能科学技术的研究、开发及应用,在学术思想与科学方法上,具有综合性、整体性、协调性的特色,在理论方法研究与应用技术开发方面,取得了具有创新性、开拓性的成果。“智能化”已成为当前新技术、新产品的发展方向和显著标志。

为了适时总结、交流、宣传我国学者在“智能科学技术”领域的研究开发及应用成果,中国人工智能学会与科学出版社合作编辑出版《智能科学技术著作丛书》。需要强调的是,这套丛书将优先出版那些有助于将科学技术转化为生产力以及对社会和国民经济建设有重大作用和应用前景的著作。

我们相信,有广大智能科学技术工作者的积极参与和大力支持,以及编委们的共同努力,《智能科学技术著作丛书》将为繁荣我国智能科学技术事业、增强自主创

新能力、建设创新型国家做出应有的贡献。

祝《智能科学技术著作丛书》出版,特赋贺诗一首:

**智能科技领域广
人机集成智能强
群体智能协同好
智能创新更辉煌**



中国人工智能学会荣誉理事长

2005年12月18日

前 言

人类对于知识的研究与探索自始至终都未停止过。在人类进入信息化社会并向知识化社会迈进的过程中，人类通过计算机的应用把知识从概念真正跃升到知识科学。知识工程便是一门新兴的关于知识获取、表示和推理以及用一种特定形式把知识表示为计算机可操作对象的科学，其研究的目标是挖掘和抽取人类知识，这也使得计算机具有人类的一定智能。

知识工程是一个庞大的交叉领域，本书将知识处理定义为对环境中接收感知信息并执行行动的智能体，每个这样的智能体都实现把感知序列映射到行动函数；讨论了表达这些函数的各种方法，诸如一个从当前状态的条件到行动的直接映射（反应式智能体），一种从感知序列利用关于世界发展方式的信息，以及关于智能体可以采用的可能行动的结果信息推断世界的相关属性（逻辑智能体）；指示了对世界状态的愿望度的效用信息及对行动的愿望度的行动值信息（效用智能体）；表明了学习智能体的感知不仅应该对进行中的行动有用，而且应该能够改进智能体未来行动的能力（泛化能力）。

全书共分为 11 章，第 1 章讨论了在图灵测试意义上的人工智能学派、应用领域，进而引出知识工程的概念；第 2 章围绕智能体概念（能够决定要做什么，然后执行行动的系统），给出了理性的判别标准和问题环境；第 3 章详细设计了目标智能体；第 4 章讨论了实现智能体推理的基础——逻辑；第 5 章讨论了智能体推理所需要的知识的表示方法——本体；第 6 章讨论常识及其表示；第 7 章集中讨论了当世界存在不确定因素时的推理与决策——效用；第 8 章讨论了智能体推理所需要的知识的获取方法——机器学习；第 9 章和第 10 章详细介绍知识处理环境 Prolog 及知识系统应用案例；第 11 章讨论多智能体规划、协作和通信问题。

本书第 2~4、9、11 章由刘一松编写，第 6~8 章由晏立编写，第 1、5、10 章由程显毅编写，全书由程显毅统稿。

感谢陈小波、梁军、朱嫣、杨长隅、谢玉宇等在资料整理过程中所做的工作。感谢江苏大学研究生部和计算机学院给予的支持。

感谢人工智能学会荣誉理事长涂序彦教授、中国科学院计算技术研究所史忠植研究员、清华大学石纯一教授和原西南师范大学邱玉辉教授对本书提出的宝贵意见。

本书得到江苏省高校自然科学研究计划指导性项目基金（03kj520175）和江苏大学研究生教材建设规划项目（1293000404）的资助。

知识工程是一门比较年轻的学科，正处于蓬勃发展时期，对许多问题作者并未做深入研究，一些有价值的新内容也来不及收入本书，加上作者知识水平和实践经验有限，书中难免存在不足之处，敬请读者批评指正。

目 录

《智能科学技术著作丛书》序

前言

第 1 章 绪论	1
1.1 图灵测试与人工智能	1
1.1.1 图灵测试	1
1.1.2 人工智能的不同学派	2
1.2 知识工程	5
1.2.1 问题的提出	5
1.2.2 知识	6
1.2.3 知识工程的基础、原理	7
1.2.4 知识模型中的构件	8
1.2.5 人的因素：知识工程中人的角色	9
1.2.6 知识工程的发展	11
1.3 习题	12
第 2 章 智能体	13
2.1 智能体和环境	13
2.2 理性	14
2.3 任务环境 PEAS 属性	16
2.4 智能体结构	17
2.4.1 反应式智能体	17
2.4.2 慎思式智能体	18
2.4.3 混合式智能体	19
2.5 小结	20
2.6 习题	21
第 3 章 目标智能体	22
3.1 目标智能体结构	22
3.2 问题形式化	23
3.3 不完全可观察环境的搜索	25
3.4 利用启发式知识的目标搜索	26
3.4.1 贪婪最佳优先搜索算法	26

3.4.2	A* 搜索算法	28
3.4.3	爬山搜索算法	29
3.4.4	模拟退火搜索算法	31
3.4.5	遗传算法	32
3.5	目标测试标准化	33
3.5.1	约束满足问题	33
3.5.2	CSP 问题的回溯搜索	35
3.5.3	通过约束传播信息	35
3.5.4	CSP 问题的最小冲突搜索	37
3.6	小结	37
3.7	习题	37
第 4 章	逻辑智能体	39
4.1	逻辑智能体结构	39
4.2	wumpus 世界	40
4.3	逻辑与知识库	42
4.3.1	命题逻辑	43
4.3.2	一个简单的知识库	44
4.3.3	真值表推理	45
4.4	命题逻辑智能体	45
4.5	一阶逻辑智能体	47
4.5.1	一阶逻辑	47
4.5.2	一阶逻辑知识表示	48
4.6	逻辑智能体的推理	51
4.7	小结	53
4.8	习题	53
第 5 章	本体论	55
5.1	本体概念	55
5.1.1	本体特征	55
5.1.2	本体与一般术语的区别	56
5.1.3	本体分类	58
5.2	本体的形式化定义	61
5.2.1	OWA 形式化定义	61
5.2.2	Guarino 形式化定义	61
5.2.3	KAON 形式化定义	63
5.3	本体建模	65

5.3.1	本体建模基元	65
5.3.2	建立本体的一般方法	67
5.4	领域本体知识库	69
5.4.1	领域本体知识库概念	69
5.4.2	领域本体知识库的构建	71
5.4.3	领域本体的建模	74
5.4.4	领域本体的复用	75
5.4.5	领域本体的应用	77
5.5	本体编辑工具	78
5.5.1	本体编辑工具概述	78
5.5.2	Protégé 开发过程	81
5.5.3	推理机 Racer	90
5.6	本体描述语言	96
5.6.1	本体语言概述	96
5.6.2	HTML 扩展	96
5.6.3	本体标记语言 OML	97
5.6.4	基于 XML 的文本交换语言 XOL	97
5.6.5	XML	97
5.6.6	RDF	99
5.6.7	本体交互语言 OIL	102
5.6.8	DAML+OIL 语言	103
5.6.9	子语言	103
5.6.10	描述逻辑	105
5.7	小结	107
5.8	习题	108
第 6 章	常识	109
6.1	常识的概念	109
6.1.1	常识的例子	109
6.1.2	常识推理的特点	109
6.1.3	常识的重要性	110
6.1.4	常识的表示	111
6.2	常识推理	115
6.2.1	缺省推理	115
6.2.2	真值维护系统	118
6.3	小结	119

6.4	习题	120
第7章	效用智能体	121
7.1	效用智能体结构	121
7.2	不确定性	122
7.2.1	不确定环境下智能体的行动	122
7.2.2	不确定知识的处理	123
7.2.3	不确定性与理性决策	125
7.3	信念网推理	126
7.3.1	全联合分布推理	126
7.3.2	信念网推理	130
7.3.3	信念网的数值语义	132
7.3.4	信念网拓扑语义	135
7.3.5	信念网的精确推理	135
7.4	定性推理	140
7.4.1	定性推理概述	140
7.4.2	定性推理的基本方法	142
7.4.3	ENVISION 方法	143
7.4.4	空间定性推理概述	147
7.5	集对分析	149
7.5.1	集对分析基本原理	149
7.5.2	联系数	152
7.5.3	集对势	152
7.5.4	基于集对分析的不确定性理论	153
7.5.5	集对分析在人工智能中的应用	154
7.6	小结	162
7.7	习题	162
第8章	学习智能体	164
8.1	学习智能体结构	164
8.2	机器学习概述	165
8.2.1	基本概念	166
8.2.2	机器学习的发展史	167
8.3	从观察中学习	168
8.3.1	假设及假设空间	168
8.3.2	决策树 CLS 算法	169
8.4	统计机器学习	171

8.4.1	贝叶斯决策	171
8.4.2	最大似然参数估计	172
8.4.3	无监督聚类	173
8.5	PAC 学习	176
8.5.1	泛化问题	176
8.5.2	SVM	177
8.5.3	表示问题	179
8.6	集成机器学习	180
8.6.1	弱可学习定理	180
8.6.2	集成机器学习方法	181
8.6.3	经验性研究问题	182
8.7	强化学习	183
8.7.1	完全可观察的 MDP 问题	183
8.7.2	部分可观察的 MDP 问题	186
8.7.3	强化学习的组成部分	186
8.7.4	强化学习原理	187
8.8	小结	189
8.9	习题	189
第 9 章	知识处理环境 Prolog	191
9.1	Prolog 的特点	191
9.1.1	事实和规则的描述	191
9.1.2	Prolog 程序结构	192
9.2	一个简单的 Prolog 程序	193
9.2.1	表达事实	193
9.2.2	事实查询	195
9.2.3	规则表达	197
9.2.4	规则查询	197
9.2.5	规则使用	198
9.3	运算	200
9.3.1	算术运算	200
9.3.2	逻辑运算	201
9.3.3	数据管理	201
9.3.4	递归	203
9.3.5	表	205
9.3.6	截断	208

9.3.7 循环	210
9.4 Visual Prolog 编程入门	211
9.4.1 创建项目	211
9.4.2 编写简单的应用程序	216
9.4.3 创建窗口	217
9.4.4 给窗口添加控件	222
9.4.5 画鼠标掠影	224
9.5 小结	227
9.6 习题	227
第 10 章 知识系统案例	229
10.1 案例 1——简单的医疗诊断系统	229
10.1.1 规则库的构造	229
10.1.2 源程序	231
10.1.3 功能扩展	232
10.1.4 程序运行	235
10.2 案例 2——汽车故障检修咨询系统	237
10.2.1 知识库	237
10.2.2 编程实现	238
10.2.3 运行结果	243
10.3 习题	244
第 11 章 多智能体	245
11.1 多智能体	245
11.1.1 效用和偏好	245
11.1.2 多智能体交互	246
11.1.3 优势策略和 Nash 平衡	248
11.1.4 竞争与零和交互	250
11.2 智能体之间的通信	253
11.2.1 交谈	253
11.2.2 理解语言字符串	255
11.2.3 有效通信	262
参考文献	267

第 1 章 绪 论

理解智能包括理解：知识如何获取、表达和存储；智能行为如何产生和学习；动机、情感和偏好如何发展和运用；传感器信号如何转换各种符号；怎样利用各种符号执行逻辑运算、对过去进行推理及对未来进行规划；智能机制如何产生愿望、信念和意图等现象。

从 1956 年正式提出人工智能五十多年来，人工智能已取得长足的发展，成为一门广泛的交叉和前沿学科。现在计算机似乎已经变得十分“聪明”了。例如，1997 年 5 月，IBM 公司研制的深蓝（Deep Blue）计算机战胜了国际象棋大师卡斯帕洛夫（Kasparov）。计算机编程语言和其他计算机软件都因为有了人工智能的进展而得以存在。

进入 21 世纪，人工智能理论正酝酿着新的突破——人工生命的提出，这意味着人类不仅试图从传统的工程技术途径，而且将从新开辟的生物工程技术途径，去发展人工智能；同时人工智能的发展，还将作为人工生命科学的重要支柱和推动力量。可以预言，人工智能的研究成果将能够创造出更多更高级的智能“制品”，并使之在越来越多的领域超越人类智能；人工智能将为发展国民经济和改善人类生活作出更大贡献。

人工智能的近期研究目标是建造智能计算机，用以代替人类从事脑力劳动，即使现有的计算机更“聪明”更有用。正是根据这一近期研究目标，才把人工智能理解为计算机科学的一个分支。人工智能的远期研究目标是探究人类智能和机器智能的基本原理，研究用自动机模拟人类的思维过程和智能行为。这个长期目标远远超出计算机科学的范畴，几乎涉及自然科学和社会科学的所有学科。

1.1 图灵测试与人工智能

1.1.1 图灵测试

一些哲学家试图证明人工智能是不可能的，即机器不可能智能地行动。显然，人工智能是否可能，取决于它是如何被定义的。本质上，人工智能寻求的是在给定的体系结构之上最好的智能体（在第 2 章给出定义）程序。根据此公式，人工智能在理论上是可能的：对于任何由 k 位存储所组成的数字体系结构，刚好有 2^k 个智能体程序，为了找到最好的那个，需要进行枚举并逐一测试。如果 k

很大，这是不实际的。

“在给定的体系结构上建立最好的智能体”，这是一个工程问题。1950年，著名的英国科学家图灵发表了题为《计算机能思考吗?》的论文，给人工智能下了一个定义，他建议，我们应该问“能否通过关于行为的智能测试”(后被称为图灵测试)，而不是问“机器能否思考”。图灵测试的本质就是让人在不看外形的情况下不能区别机器的行为和人的行为，那么这个机器就是智慧的。

该测试是让两个程序 A、B(通过联机打字输入消息)与一个询问人 C 进行 5 分钟的对话，C 与 A、B 被隔离。询问者只知道二人的称呼是 X、Y，通过提问以及回答来判断，最终作出“X 是 A，Y 是 B”或者“X 是 B，Y 是 A”的结论。测试中，A 必须尽量使 C 判断错误，而 B 的任务是帮助 C。

用一个机器代替测试中的 A，并且机器将试图使 C 相信它是一个人。如果机器通过了图灵测试，就认为它是“智慧”的(见图 1.1)。

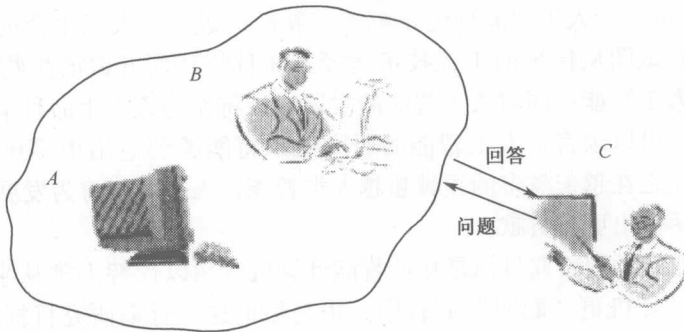


图 1.1 图灵测试

问与答从表面上看是语言的对话，其核心是语义的理解和与理解相关联的背景知识以及理解后作出的相应的处理，这就涉及语义如何定义(显语义与隐语义、浅语义与深语义)，语义之间如何关联，语义如何形式化表示，计算机如何能够“理解”语义或做出近似于“理解”语义的操作等。

图灵还列出了“一台机器永远做不了……”的清单：如善良和蔼、足智多谋、美丽大方、友好地、积极主动、有幽默感、明辨是非、犯错误、坠入爱河、从经验中学习、用词恰如其分、反思自我、作出真正创新之事等。

1.1.2 人工智能的不同学派

不同科学或学科背景的学者对人工智能有不同的理解，提出了不同的观点，主要观点有符号主义、联结主义和行为主义等。

1. 符号主义

符号主义 (symbolism, 1956 年) 又称为逻辑主义 (logicism)、心理学派 (psychologism) 或计算机学派 (computerism), 其原理主要是物理符号系统 (即符号操作系统) 假设和有限合理性原理。

符号主义认为人的认知基元是符号, 而且认知过程即符号操作过程。它认为人是一个物理符号系统, 计算机也是一个物理符号系统, 因此, 就能够用计算机来模拟人的智能行为, 即用计算机的符号操作来模拟人的认知过程。也就是说, 人的思维是可操作的。它还认为, 知识是信息的一种形式, 是构成智能的基础。人工智能的核心问题就是知识表示、知识推理和知识运用。知识可用符号表示, 也可用符号进行推理, 因而有可能建立起基于知识的人类智能和机器智能的统一理论体系。

符号主义认为人工智能源于数理逻辑。数理逻辑从 19 世纪末起获得迅速发展, 到 20 世纪 30 年代开始用于描述智能行为。计算机出现后, 其又在计算机上实现了逻辑演绎系统。正是这些符号主义者, 在 1956 年首先采用“人工智能”这个术语。后来又发展了启发式算法、专家系统和知识工程理论与技术, 并在 80 年代取得很大发展。符号主义曾长期一枝独秀, 为人工智能的发展作出了重要贡献, 尤其是专家系统的成功开发与应用, 对人工智能走向工程应用和实现理论联系实际具有特别重要的意义。在人工智能的其他学派出现之后, 符号主义仍然是人工智能的主流派。这个学派的代表有 Newell、Shaw、Simon Nilsson 等。

一个完善的符号系统应具有下列六种基本功能:

- (1) 输入符号 (input);
- (2) 输出符号 (output);
- (3) 存储符号 (store);
- (4) 复制符号 (copy);
- (5) 建立符号结构 (built symbol structure): 通过找出各符号间的关系, 在符号系统中形成符号结构;
- (6) 条件性迁移 (conditional transfer): 根据已有符号, 继续完成活动过程。

人具有上述六种功能, 现代计算机也具备物理符号系统的这六种功能。

符号主义附带三个推论:

推论一: 既然人具有智能, 那么他就一定是个物理符号系统。

推论二: 既然计算机是一个物理符号系统, 那么它就一定能够表现出智能。

推论三: 既然人是一个物理符号系统, 计算机也是一个物理符号系统, 那么