

人工智能 原理与应用

● 王文杰 叶世伟 编著



人民邮电出版社
POSTS & TELECOM PRESS

人工智能原理与应用

王文杰 叶世伟 编著

人民邮电出版社

图书在版编目 (CIP) 数据

人工智能原理与应用/王文杰, 叶世伟编著. —北京: 人民邮电出版社, 2004.3

ISBN 7-115-11903-1

I. 人... II. ①王... ②叶... III. 人工智能 IV. TN911.6

中国版本图书馆 CIP 数据核字 (2004) 第 004736 号

内 容 提 要

本书系统地阐述了人工智能的基本原理、基本技术、研究方法和应用领域等内容, 比较全面地反映了国内外人工智能研究领域的最新进展和发展方向。全书共分为 16 章, 内容涉及到人工智能基本概念、结构化和非结构化知识表示技术、搜索技术和问题求解、确定性推理和不确定性推理、非标准逻辑、专家系统、机器学习、神经网络、统计学习、遗传算法和智能 Agent、自然语言处理等几个方面。每章后面均附有习题, 以供读者练习。本书充分考虑到人工智能学科的整体结构, 注重系统性、先进性、新颖性、实用性。内容由浅入深、循序渐进, 条理清晰。

本书是为人工智能原理课程而编写的, 同时也涵盖了国家学位委员会对同等学历人员申请计算机科学技术专业硕士学位“人工智能考试大纲”的内容, 可作为相关学科的本科生、研究生以及在职研究生的教材, 也可以供从事相关专业的教师和广大科技人员作为参考书。

人工智能原理与应用

◆ 编 著 王文杰 叶世伟

责任编辑 邹文波

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号

邮编 100061 电子函件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

读者热线 010-67194042

北京汉魂图文设计有限公司制作

北京隆昌伟业印刷有限公司印刷

新华书店总店北京发行所经销

◆ 开本: 787×1092 1/16

印张: 30.75

字数: 748 千字

2004 年 3 月第 1 版

印数: 1-4 000 册

2004 年 3 月北京第 1 次印刷

ISBN 7-115-11903-1/TP·3735

定价: 39.00 元

本书如有印装质量问题, 请与本社联系 电话: (010) 67129223

前 言

人工智能是一门前沿性的学科,它研究计算机实现智能的基本原理,研究使计算机达到高层次的有效应用;同时人工智能也是一门多学科交叉的综合学科,它涉及计算机科学、数学、心理学、认知科学等众多领域。从人工智能正式提出至今已经走过了 50 多个年头,但是人工智能的发展和前进过程充满了坎坷,有高潮也有低谷。随着信息化、计算机网络和 Internet 技术的发展,人类已步入信息社会和网络经济的时代,它们为人工智能提出了更多具体的研究目标和研究课题。人工智能作为计算机学科的重要分支,已成为人类在信息社会和网络经济时代所必须具备的一项核心技术,并将在新的世纪发挥更大的作用。

本教材是在我们多年为中国科学院研究生院计算机学科和相关专业的研究生以及在职研究生讲授人工智能原理这门课程的基础上编写的。由于人工智能研究涉及众多学科,同时它也是一门不断发展和成熟的学科,新的理论和方法不断涌现,新的研究成果不断丰富着它的研究内容,新的应用不断为它提出新的研究课题,目前人工智能所涉及的内容更加广泛,技术更加复杂,应用更加深入和具体。这使得人工智能尚未形成完整和成熟的理论体系,为人工智能的教学和科研带来了一定的难度。因此,在教材的选材方面,我们充分考虑到人工智能学科的整体结构和最新研究进展,同时也参考了国家学位委员会颁布的《同等学力人员申请硕士学位计算机科学与技术学科综合水平考试大纲及指南》中的“人工智能考试大纲”的要求,既结合了全日制学生的特点,介绍了非标准逻辑、统计学习、智能 Agent 等概念,又考虑到了在职研究生的具体情况,在紧扣全国考试大纲的基础上,对数理逻辑、归结原理、不确定性推理、知识表示等内容做了详细的介绍。本教材内容力求做到由浅入深、循序渐进、条理清晰、前后一致,既强调基本原理和工程应用,又要全面反映国内外研究和应用的新的进展,具有先进性、新颖性、实用性。

全书共分为 16 章。第 1 章对人工智能的基本概念、发展历史、各学派研究特点和应用领域等进行了必要的讨论。第 2 章介绍了问题求解的基本技术,讨论了各种搜索技术,包括盲目搜索、启发式搜索和博弈问题以及约束满足问题。第 3 章是知识表示和数理逻辑的基础性介绍,对基于逻辑的知识表示方法进行了介绍。第 4 章介绍了归结推理规则和归结原理,并对逻辑程序设计语言 Prolog 做了简单的介绍。第 5 章对非标准逻辑进行了介绍,包括封闭世界假设、缺省推理、限制理论、真值维护系统等,并对人工智能的难题之一的框架问题进行了介绍。第 6 章介绍不确定性推理,讨论了主观 Bayes 方法、确定性理论、证据理论以及模糊逻辑和模糊推理等内容。第 7 章和第 8 章介绍了其他几种主要的知识表示方法,包括产生式系统、语义网络、框架理论,并介绍了 Petri 网、概念依赖、脚本等表示方法,还简单介绍了产生式系统 CLIPS,包括产生式系统的特点、RETE 匹配算法等。第 9 章和第 10 章介绍了知识工程研究的重要领域专家系统以及机器学习基本技术,包括实例学习、ID3 算法、加强学习等。第 11 章到第 13 章介绍了近年来研究比较热的领域,包括神经网络、统计学习和 SVM,以及遗传算法

等内容。第 14 章和第 15 章介绍 Agent 和智能 Agent 的概念,以及多 Agent 系统相关的一些概念。第 16 章介绍了与自然语言处理相关的一些议题。我们在本书的每章后面都附有一定数量的习题,并在最后列出了大量的参考文献,供读者学习参考。

本教材是为研究生教学而编写的,既可以作为全日制研究生相关课程的学习,也可以作为在职研究生和高年级本科生学习人工智能的教材和参考书。由于本书涉及面广,内容繁多,在教学过程中可以根据实际需要进行取舍。本教材包含了作者多年的教学实践和体会,也吸取了国内外同类教材和有关文献的精华,在此谨向这些教材和文献的作者表示感谢,也向为我们提供帮助的多位老师和学生表示感谢。

由于作者水平有限,在书中难免出现错误,再加上人工智能研究领域广泛,在内容的取舍和安排上,以及前后连贯上难免有不足之处,殷切希望广大读者批评指正。

编 者

2003 年 12 月

于中科院研究生院

目 录

第 1 章 绪论	1
1.1 人工智能的界定	1
1.1.1 什么是人工智能	2
1.1.2 人工智能的研究目标	4
1.1.3 人工智能中的通用问题求解方法	5
1.2 人工智能的学科范畴	8
1.2.1 人工智能的研究范畴	8
1.2.2 人工智能技术的应用	12
1.3 人工智能的研究方法	15
1.3.1 人工智能研究的特点	15
1.3.2 人工智能的研究途径	15
1.4 人工智能的发展简史和趋势	18
1.4.1 人工智能的发展简史	18
1.4.2 人工智能发展趋势	22
1.5 本书的目标和范围	23
1.6 小结	24
习题	24
第 2 章 搜索技术	25
2.1 引言	25
2.2 通用问题求解方法	26
2.2.1 宽度优先搜索	27
2.2.2 深度优先搜索	29
2.2.3 迭代加深搜索	30
2.3 启发式搜索	32
2.3.1 启发性信息和评估函数	33
2.3.2 启发式 OR 图搜索算法	33
2.4 问题归约和 AND-OR 图启发式搜索	45
2.4.1 问题归约的描述	45
2.4.2 AND-OR 图表示	45

2.4.3 AO* 算法	47
2.5 博弈	52
2.5.1 极大极小过程	53
2.5.2 α - β 过程	55
2.6 约束满足搜索	57
2.7 小结	60
习题	61
第3章 知识表示及逻辑推理	63
3.1 概述	63
3.1.1 知识和知识表示	63
3.1.2 人工智能对知识表示方法的要求	65
3.2 知识表示语言问题	65
3.3 命题逻辑	67
3.3.1 语法	67
3.3.2 语义	68
3.4 谓词逻辑	70
3.4.1 语法	70
3.4.2 语义	76
3.5 逻辑推理	78
3.5.1 谓词公式的等价性和永真蕴涵	78
3.5.2 谓词公式的有效性和可满足性	80
3.6 符号逻辑中的范式	83
3.7 一阶谓词逻辑的应用	84
3.8 小结	86
习题	87
第4章 归结原理及其应用	89
4.1 引言	89
4.2 命题演算中的归结	90
4.2.1 归结推理规则	90
4.2.2 归结反演	91
4.2.3 命题逻辑归结反演的合理性和完备性	93
4.2.4 归结反演的搜索策略	93
4.3 谓词演算中的归结	95
4.3.1 子句型	95
4.3.2 置换和合一	98
4.3.3 合一算法	100
4.3.4 归结式	102
4.3.5 归结反演	103
4.3.6 答案的提取	105

4.4 谓词演算归结反演的完备性和合理性	108
4.4.1 Herbrand 域	108
4.4.2 Herbrand 解释	109
4.4.3 语义树	110
4.4.4 Herbrand 定理	112
4.4.5 完备性和合理性	114
4.5 逻辑程序设计原理	117
4.5.1 逻辑程序定义	117
4.5.2 Prolog 数据结构和递归	118
4.5.3 SLD 归结	119
4.5.4 非逻辑成分: CUT	121
4.5.5 其他逻辑程序设计语言	123
4.6 小结	124
习题	124
第 5 章 缺省和非单调推理	128
5.1 引言	128
5.2 非单调推理	129
5.3 封闭世界假设	131
5.4 缺省推理	133
5.5 限制理论	137
5.6 非单调逻辑	139
5.7 真值维护系统	141
5.8 框架问题	144
5.8.1 什么是框架问题	145
5.8.2 框架问题的非单调解决方案	149
5.9 小结	154
习题	155
第 6 章 不确定性推理	157
6.1 概述	157
6.1.1 什么是不确定推理	157
6.1.2 不确定推理要解决的基本问题	158
6.1.3 不确定性推理方法分类	160
6.2 主观 Bayes 方法	161
6.2.1 全概率公式和 Bayes 公式	161
6.2.2 主观 Bayes 方法	163
6.3 确定性理论	171
6.3.1 建造医学专家系统时的问题	171
6.3.2 C-F 模型	172
6.3.3 确定性方法的说明	177

6.4 证据理论	178
6.4.1 假设的不确定性	178
6.4.2 证据的不确定性	182
6.4.3 证据的组合函数	182
6.4.4 规则的不确定性	183
6.4.5 不确定性的传递	183
6.4.6 不确定性的组合	184
6.4.7 证据理论的特点	186
6.5 模糊逻辑和模糊推理	186
6.5.1 模糊集合及其运算	186
6.5.2 模糊关系	188
6.5.3 语言变量	189
6.5.4 模糊逻辑	189
6.5.5 模糊推理	190
6.6 小结	198
习题	199
第7章 产生式系统	201
7.1 概述	201
7.2 产生式专家系统 CLIPS	203
7.2.1 产生式系统语言 CLIPS	203
7.2.2 CLIPS 中知识的表示模式	203
7.2.3 CLIPS 运行	207
7.3 产生式系统的匹配算法	209
7.3.1 索引记数法	210
7.3.2 Rete 算法	210
7.3.3 书写规则时效率的考虑	214
7.4 产生式系统的冲突消解策略	216
7.5 元知识	217
7.5.1 什么是元知识	217
7.5.2 元知识的作用	218
7.5.3 元知识的使用模式	219
7.6 产生式系统的推理方向	220
7.7 产生式系统的类型	222
7.8 产生式系统的特点	223
7.9 小结	224
习题	224
第8章 知识的结构化表示	226
8.1 引言	226
8.2 语义网络	227

8.2.1	语义网络的基本概念	227
8.2.2	基本事实的表示	228
8.2.3	表示情况和动作	230
8.2.4	多元语义网络的表示	231
8.2.5	连接词与量词的表示	231
8.2.6	语义网络的推理过程	234
8.2.7	语义网络表示方法的特点	238
8.3	框架	239
8.3.1	框架理论	239
8.3.2	框架	240
8.3.3	框架系统	243
8.3.4	框架中预先定义的槽	245
8.3.5	框架中的推理	247
8.3.6	框架表示方法的特点	249
8.3.7	框架系统、语义网络、面向对象知识表示的比较	250
8.4	Petri 网	251
8.5	概念依赖	253
8.6	脚本	255
8.7	小结	256
	习题	257
第 9 章	专家系统	259
9.1	专家系统概述	259
9.1.1	什么是专家系统	259
9.1.2	专家系统的结构	260
9.2	问题求解组织结构	263
9.2.1	议程表	263
9.2.2	问题求解的黑板模型	264
9.2.3	面向对象的问题组织结构	266
9.3	知识获取	266
9.3.1	知识获取的手工方法	266
9.3.2	知识获取的机器学习方法	269
9.4	有效性、验证和维护问题	269
9.4.1	专家系统的有效性	270
9.4.2	知识库的验证	272
9.4.3	知识库的维护	275
9.5	专家系统开发工具	276
9.6	小结	278
	习题	279

第 10 章 机器学习	280
10.1 机器学习的概念和方法	280
10.1.1 什么是机器学习	280
10.1.2 机器学习研究概况	281
10.1.3 机器学习研究方法	282
10.2 归纳学习	285
10.2.1 归纳学习的基本概念	285
10.2.2 版本空间学习	287
10.2.3 基于决策树的归纳学习方法	289
10.2.4 Induce 算法	292
10.3 其他学习方法	294
10.3.1 类比学习	294
10.3.2 解释学习	295
10.4 加强学习	297
10.5 归纳逻辑程序设计学习	300
10.6 基于范例的学习	301
10.6.1 CBR 的过程模型	302
10.6.2 范例的表示和索引	303
10.6.3 范例的推理	304
10.6.4 范例的学习	306
10.7 小结	306
习题	307
第 11 章 人工神经网络理论	309
11.1 概述	309
11.1.1 基本的神经网络模型	309
11.1.2 神经网络的学习方法	310
11.2 人工神经元及感知机模型	311
11.2.1 基本神经元	311
11.2.2 感知机模型	312
11.3 前馈神经网络	313
11.3.1 前馈神经网络模型	313
11.3.2 多层前馈神经网络的误差反向传播(BP)算法	314
11.3.3 BP 算法的若干改进	316
11.4 反馈神经网络	317
11.4.1 离散 Hopfield 网络	317
11.4.2 连续 Hopfield 网络	322
11.4.3 Hopfield 网络应用	323
11.4.4 双向联想记忆(BAM)模型	324
11.5 随机神经网络	325

11.5.1 模拟退火算法	325
11.5.2 玻尔兹曼机	327
11.6 自组织特征映射神经网络	330
11.6.1 网络的拓扑结构	330
11.6.2 网络自组织算法	331
11.6.3 有教师学习	332
11.7 小结	332
习题	332
第 12 章 统计学习理论与支持向量机	334
12.1 统计学习理论	334
12.1.1 机器学习的表示	334
12.1.2 学习过程一致性的条件	335
12.1.3 VC 维理论	337
12.1.4 推广性的界	337
12.1.5 结构风险最小化	338
12.2 支持向量机	339
12.2.1 最优分类超平面	340
12.2.2 广义最优分类超平面	341
12.2.3 支持向量机	342
12.2.4 核函数	343
12.2.5 SVM 的算法及多类 SVM	344
12.2.6 ϵ -不敏感损失函数	345
12.2.7 用于非线性回归的支持向量机	346
12.2.8 SVM 的应用研究	348
12.3 小结	348
习题	349
第 13 章 遗传算法	350
13.1 简介	350
13.2 遗传算法的历史	351
13.3 遗传算法的研究内容和取得的进展	353
13.4 遗传算法的描述	353
13.5 遗传算法的特点	357
13.6 遗传算法的马尔柯夫链分析	359
13.6.1 概率意义下的收敛	359
13.6.2 有限 Markov 链的预备知识	359
13.6.3 Markov 链的分析方法回顾	360
13.6.4 GA 的收敛性分析	361
13.7 遗传算法的研究展望	362
13.8 小结	364

习题	364
第 14 章 智能 Agent	365
14.1 引言	365
14.2 分布式人工智能	366
14.2.1 分布式问题求解	367
14.2.2 多 Agent 系统	369
14.3 Agent 与智能 Agent	369
14.3.1 什么是 Agent	370
14.3.2 什么是智能 Agent	371
14.3.3 Agent、专家系统与对象	373
14.3.4 智能 Agent 的抽象结构	374
14.4 模态逻辑	375
14.4.1 可能世界模型	376
14.4.2 模态逻辑	377
14.5 知道和信念逻辑	381
14.5.1 知道和信念逻辑	381
14.5.2 公共和分布式知识	383
14.5.3 自认知逻辑	384
14.6 逻辑万能问题	388
14.6.1 演绎规则的不完全性方法	389
14.6.2 隐式和显式信念	391
14.7 理性 Agent 模型	393
14.7.1 Cohen 和 Levesque 的理性逻辑	396
14.7.2 Rao 和 Georgeff 的 BDI 逻辑	399
14.7.3 KARO 框架	402
14.8 Agent 结构	403
14.8.1 基于逻辑的 Agent	404
14.8.2 反应 Agent	406
14.8.3 BDI Agent	408
14.8.4 层次 Agent	410
14.9 小结	412
习题	412
第 15 章 多 Agent 系统	414
15.1 引言	414
15.2 Agent 通信	415
15.2.1 通信与 DAI	415
15.2.2 通信的层次	417
15.2.3 言语动作	417
15.2.4 知识查询处理语言 KQML	421

15.2.5 Agent 通信语言 ACL	425
15.2.6 Agent 通信与对话	427
15.2.7 Agent 通信与 WWW	429
15.3 协调与合作	429
15.3.1 协调	430
15.3.2 合作	434
15.3.3 协商	435
15.3.4 市场机制	438
15.4 Agent 的社会性	439
15.4.1 识别 (Recognition)	439
15.4.2 团队形成	441
15.4.3 规划形成	442
15.4.4 团队行动	443
15.5 小结	443
习题	444
第 16 章 自然语言理解	445
16.1 自然语言理解的一般问题	445
16.1.1 自然语言理解的概念及意义	445
16.1.2 自然语言理解研究的发展	446
16.1.3 自然语言理解的层次	448
16.2 词法分析	449
16.3 句法分析	450
16.3.1 短语结构语法和 Chomsky 语法体系	450
16.3.2 句法分析树	452
16.3.3 转移网络	453
16.3.4 转移网络的神经网络实现	455
16.3.5 扩充转移网络	456
16.4 语义分析	459
16.4.1 语义文法	459
16.4.2 格文法	460
16.5 大规模真实文本的处理	461
16.5.1 语料库语言学及其特点	461
16.5.2 统计学方法的应用及所面临的问题	464
16.5.3 汉语语料库加工的基本方法	465
16.6 对话和语用分析	468
16.7 小结	469
习题	469
参考文献	470

第 1 章 绪论

人工智能 (Artificial Intelligence, AI) 是一门正在发展中的综合性的前沿学科, 是计算机科学与技术学科领域的一个重要的研究方向, 主要研究如何使用机器 (计算机) 来模拟和实现人类的智能行为, 是在计算机、控制论、信息论、数学、心理学、哲学、语言学等多种学科相互综合、相互渗透的基础上发展起来的一门新兴边缘学科。自 1956 年人工智能诞生至今已经走过了曲折的 40 多年, 其间风风雨雨, 高潮低潮几经交替, 遇到过不少的争论、困难和挑战, 同时也获得了迅速的发展, 在很多领域得到了广泛的应用。本章将对人工智能学科做一简要的回顾, 介绍智能系统所包括的一些研究问题, 以及应用范围等。

1.1 人工智能的界定

人类的自然智能伴随着人类的活动无处不在, 如解题、下棋、猜谜、讨论问题、编制计划和编制程序、驾车等都需要智能。因此, 智能是和人类的各项活动紧密联系在一起。在古代, 人类就开始用各种方式来记录信息和描绘自己的形象, 不仅如此, 人们更希望建立一种“智能”的模型, 以便指导人们在各种活动中的行为并做出决策。但是, 对那时的人类来说, 由于缺乏有效的手段和工具, 这是一件非常复杂和困难的工作。经过成千上万年的不断努力, 人们构造出了可以机械地进行一定算术运算的“分析机”。Babbage 的分析机是现代意义上计算的第一个重要的成果。基于这一成果, 经过人们多年的努力以及各种相关条件的不断成熟, 人类实现了第一代计算机。在晶体管发明后, 人们很快就研制出了体积上更小的第二代计算机。这些计算机主要用于商业数据处理等领域。经过 10 年左右的时间, 当集成电路出现以后, 第三代计算机进入了商业应用。这些计算机具有极强的功能, 可以实时地进行大量的计算。很多的电子机械机器人就是用这些计算机设计的。又经过了 10 年左右的时间, 随着高速 VLSI 的出现, 第四代计算机产生了。人们研究出电子机器人, 它通过摄像机来对物体定位, 并把物体放在需要和合适的位置上。在 1981~1990 年期间, 人们开始了第五代计算机计划或类似的高技术计划, 第五代计算机除了要具有第四代计算机所有的能力外, 还需要能够处理“智能”。当前的计算机可以处理自然语言、玩游戏、识别物体的图像和证明数学定理等, 所有这些都是人工智能的研究领域。那么到底什么是人工智能呢? 下面就来讨论这个问题。

1.1.1 什么是人工智能

John McCarthy 给出“Artificial Intelligence”这一术语已经有 40 多年了，到目前一直没有一个简明的和形式的定义。随着人工智能研究和应用的不断深入，人们对人工智能的理解和认识有了深刻的变化，对人工智能的定义也出现了许多不同的说法，如 Rich 和 Knight 认为“人工智能是研究如何让计算机做现阶段只有人才能做得好的事情”；Nilsson 认为“人工智能是关于知识的科学，即怎样表示知识，怎样获取知识和怎样使用知识的科学”；Buchanan 和 Shortliffe 认为“人工智能是计算机科学的分支，它用符号、非算法的方法进行问题求解”。这些定义反映了人工智能学科的基本思想和基本范畴。

Stuart Russell 和 Peter Norvig 在专著^[11]中对当时已有的主要的“AI”定义进行了分析。他们把这些定义分成下面 4 个部分：

类似人一样进行思考和推理的系统	理性思考的系统
类似人一样进行动作（act）的系统	理性动作的系统

“类似人一样动作”认为人工智能是一门艺术，它能够创造出完成一定任务的机器，而当人类对这些任务进行处理的时候，需要一定的智能。如证明定理、下棋、做外科手术等等。采用的方法是选择人类做得比较好的智力型的任务，交给计算机去做。在类似人一样动作的系统中，最著名的就是 Turing 测试。Turing 测试的目的是对智能给出一个满意的操作定义。Turing 把智能行为定义为一种能力：在所有的认知任务中能够表现出和人类行为同样的程度。简单地说，Turing 测试就是计算机可以由人通过电传机来进行询问，如果询问者不能说出在另一端是人还是计算机，则该测试通过。一个计算机要通过该测试，它应该具有下面的能力。

- 自然语言处理能力：可以某种语言进行通信。
- 知识表示：在询问之前或询问过程中存储信息。
- 自动推理：根据存储的信息回答问题并得出新的结论。
- 机器学习：适应新的环境，并发现或增加新的模式。

在 AI 领域，人们并没有为通过 Turing 测试而做很大的努力。当编制需要和人进行交互的 AI 程序时，“类似人一样动作”的问题就会自然出现：专家系统需要解释如何得到其结论；自然语言处理系统和用户之间需要有一个对话系统。这些程序的行为表现必须基于人类交互的正常模式，以便使他们能够相互理解。这类系统的表示和推理可以基于人的模型，也可以不基于人的模型。

“类似人一样进行思考和推理”考虑的是怎样能够让计算机更好地完成任务。类似人一样进行思考和推理的系统中，主要采用的是认知模型的方法——关于人类思维工作原理的可检测的理论。认知科学是研究人类感知和思维信息处理过程的一门学科，它把来自 AI 的计算机模型和来自心理学的实验技术结合在一起，目的是要对人类大脑的工作原理给出准确和可测试的模型。如果说某个程序能够像人一样思考，那么就必须以某种方式确定人是如何思考的。为确定人类思维的内部是怎样工作的，可以有两种方法：通过内省（Introspec-

tion)——在人的思考过程中,掌握人自己的想法;或者通过心理学实验。一旦有了关于人类思维足够精确的理论,就可能把这种理论表示为计算机程序。如果该程序的输入/输出和实时行为与人的行为相匹配,则说明该程序的某些机制可能也是人类自身所采用的机制。例如,Newell 和 Simon 开发了“通用问题求解器”GPS,Newell 和 Simon 并不满足他们的程序能够正确地求解问题,他们更关心对某个问题 GPS 的求解过程和步骤与人对同一个问题的求解过程和步骤的比较。这是与当时的其他研究人员不同的研究方法,其他的人员关心的是如何得到正确的答案,而不考虑人是如何做的。

“理性思考”主要指的是思维方法的规范化。古希腊哲学家 Aristotle 可能是第一个试图把“正确地思考”(Right Thinking)规范化的人,即不能驳倒的(Irrefutable)推理过程,其著名的三段论方法给出了推理的模式,当前提为真的时候,就可以得到正确的结论。例如“Socrates 是人,所有的人都会死,所以 Socrates 也会死的”。人们认为这种模式就体现出了人类的思维模式,因此也开创了“逻辑学”研究领域。

19 世纪后期和 20 世纪早期发展起来的形式逻辑给出了描述事物的语句以及事物之间关系的精确的记号。到 1965 年,已经有了用逻辑符号描述问题的程序:只要给出足够的时间和空间,如果有解,就可以找到解(如果没有解,这些程序可能会永远不停止地寻找解答)。在 AI 中被称为“逻辑理论家”(Logician)的系统就是希望创建智能系统的程序。

该方法有两个主要的问题。第一,把非形式的知识用形式的逻辑记号表示是不容易做到的,特别是当这些知识不是 100%确定的时候。第二,在“理论上”可以解决一个问题和在实际中这样做是有很大的不同的,甚至只要很少的知识就可能耗尽任何计算机的资源,除非在推理时给出一定的指导。

“理性动作”的系统主要指的是理性的主体 Agent。行为上的理性指的是已知某些信念,执行某些动作以达到某个目标。Agent 可以看作是可以进行感知和执行动作的某个系统。在这种方法中,AI 可认为就是研究和建造理性 Agent。

在“理性的思考”方法中,它所强调的是正确的推理。做出正确的推理有时被作为理性 Agent 的一部分,因为理性行动的一种方法是逻辑地推出结论。另一方面,正确的推理并不是理性的全部,因为在有些情景下,往往没有某个行为是一定正确的,而其他的是错误的,也即是说没有可以证明是正确的应该做的事情,但是还必须要做某件事情。

把人工智能作为理性 Agent 进行研究有两个好处。第一,它比“理性思维”方法更加通用,因为正确的推理仅仅是达到理性的一种有用的机制,但不是必须的^[111]。第二,更应该重视的是那些基于人类行为和人类思想的科学的发展,而不是某种方法的出现。

当知识是完全的,并且资源是无限的时候,就是所谓的逻辑推理。当知识是不完全的,或者资源有限时,就是理性的行为。理性的思考和行动常常能够根据已知的信息(知识、时间、资源等)做出最合适的决策。

通过上面的讨论可以看到,虽然有很多关于 AI 的定义,但是没有一个定义能够被广泛地接受,这可能是因为这些定义都涉及到了“智能”这一概念。目前,“智能”还是一个抽象的、不可测量的量。因此一个好的 AI 定义,可能需要首先对“智能”给出形式的定义。心理学家和认知科学家认为“智能”能够帮助人们在做决策时选择合适的知识。因此可以认为:AI 是在机器上对人类智能的仿真或模拟,以便机器可以在求解问题时有效地选择和使用正确的知识。如果我们把理性定义为,系统能够在适当的时间合理地规划和执行任务,则