

计算机科学丛书

人工智能

(美) Nils J. Nilsson 著

郑扣根 庄越挺 译

潘云鹤 校



机械工业出版社
China Machine Press

本书从一个新颖的角度对人工智能各方面的问题进行了探讨。由浅入深地介绍了整个人工智能系统和agent的发展历程。首先,描述了仅能对周围环境中可感知特征做出反应的原始agent,以及它们所涉及的机器视觉、机器学习和机器进化等问题;然后,逐步介绍了agent可以从无法立即感知的任务环境中获取信息的技术。本书不仅是对人工智能技术的介绍,而且能为人工智能的研究提供参考和建议。

本书作为人工智能的入门教材,适合所有对人工智能这门学科感兴趣的读者参考,尤其适合大专院校的计算机专业及相关专业的学生用做教材或教学参考书。

Nils J. Nilsson: Artificial Intelligence: A New Synthesis.

Copyright © 1998 by Morgan Kaufmann Publishers, Inc.

Chinese edition published by arrangement with Morgan Kaufmann.

All rights reserved.

本书中文简体字版由美国Morgan Kaufmann 出版公司授权机械工业出版社独家出版。未经出版者书面许可,不得以任何方式复制或抄袭本书内容。

版权所有,侵权必究。

本书版登记号:图字:01-2000-1856

图书在版编目(CIP)数据

人工智能/(美)尼尔森(Nilsson, N. J.)著;郑扣根等译. -北京:机械工业出版社, 2000.9

(计算机科学丛书)

书名原文:Artificial Intelligence: A New Synthesis

ISBN 7-111-07885-3

I. 人… II. ①尼… ②郑… III. 人工智能 IV. TP18

中国版本图书馆CIP数据核字(2000)第30597号

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:陈贤舜

北京第二外国语学院印刷厂印刷·新华书店北京发行所发行

2000年9月第1版第1次印刷

787mm × 1092mm 1/16 · 20.5印张

印数:0 001-5 000册

定价:30.00元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换

译者序

人类总是在不断深入地研究自然界，最复杂的研究对象便是人类本身。人工智能在20世纪50年代诞生并随之兴起，此后掀起了用机器来研究与模拟人类思维的阵阵热浪。人工智能的发展历史反复经历了高峰和低谷的转换。起初，研究者们对之极为乐观，期望在不远的将来，人类所创造的机器能够成为作曲家、艺术家、工程师和象棋大师等。在这种目标的激励下，人工智能的研究热潮一浪高过一浪，很快在问题求解、博弈、演绎逻辑及机器自动证明理论和技术等方面有了突飞猛进的发展。但是人类智能之复杂及计算机软硬件之局限，决定了人工智能的发展道路崎岖不平。过高的期望也给许多人带来了失望。跳棋程序到了一定程度水平难以再提高、通用解题程序遇到了难以逾越的困难、归结方法难以克服“组合爆炸”等等，人工智能的研究曾一度陷入低谷。但近几年来，通过人工智能研究者的不断争论、探索和创新，并随着相关领域、特别是计算机技术的飞速发展，人工智能又进入一个新的发展期，一些新技术、新观念被集成到这个领域，本书即是作者在此背景下撰写而成的。

本书的作者Nilsson教授是人工智能研究中逻辑学派的奠基人之一，他早期的一些人工智能著作堪称为该领域的经典，曾经在全世界许多国家被用做大学的教材。本书是Nilsson教授在综合了当前人工智能领域的最新研究、计算机及相关领域的最新发展，增添了大量新的研究内容之后撰写完成的。本书取材新颖，概念清楚，通俗易懂，在广度、深度及先进性方面都作了综合的考虑。与作者的其他著作风格不同的是，在每章的最后一节均附有“补充阅读和讨论”，系统地总结了与所讨论的章节相关的参考文献和评论，为读者进一步深入了解书中的基本原理提供了很大的帮助。

本书的翻译力求忠于作者原意。考虑到本书可作为一本人工智能的基础教材，我们在许多人工智能的专业术语后面的括号内注上英文原文。这一方面是为了方便读者能对照理解，为其以后的学习打下基础；另一方面也为了避免以往就存在的不同中文译法而带来的歧义，以节省读者的宝贵时间。另外，本书频繁地使用了“agent”这一术语，书中可译为“智能体”或“智能主体”，但由于“agent”这一单词近几年在人工智能和计算机界使用非常广泛，且大都沿用原文，故我们在本书中也按此习惯，未将其译为中文。译者希望本书的翻译既能吸引更多的读者进入这个领域，同时也能给人工智能研究者们提供知识的更新和补充。

本书的翻译由郑扣根和庄越挺合作完成，全书由潘云鹤教授校阅。

在本书的翻译过程中，得到了马社亮、田鹏、任海波、田稷等同志的许多帮助，在此表示深深的谢意。同时也非常感谢机械工业出版社的编辑们给予我们耐心的等待和支持。

由于种种原因，书中错误和不妥之处在所难免，恳请读者批评指正。

译者

2000年4月

译、校者介绍



潘云鹤，男，1946年11月生，浙江省杭州市人。计算机专家，中国工程院院士，现任浙江大学教授、校长兼研究生院院长。

1970年毕业于上海同济大学建筑学专业。1970年8月至1978年10月在湖北省襄樊市工作，曾任襄樊市自动化研究所所长和市科委副主任。1978年至1981年为浙江大学计算机应用专业研究生，1981年毕业并获硕士学位，后留校任教，1985年9月晋升为副教授。1986年至1988年在美国卡内基—梅隆大学计算机系作访问学者。1990年任浙江大学人工智能研究所所长，1990年8月晋升为教授，1993年被评为博士生导师。1991年9月至1994年7月任浙大计算机系主任，1994年7月至1995年5月任浙江大学副校长，1995年5月起任浙江大学校长。现兼任中国计算机学会理事、中国人工智能学会理事、中国智能CAD/CAM专委理事长、《中国科学》等5种杂志编委等。



郑扣根，男，1964年11月出生于江苏镇江。1986年东北重型机械学院自动控制系本科毕业，同年考入浙江大学科仪系攻读硕士学位，1987年公派至英国Warwick大学继续攻读工程硕士，并于1990年获Warwick大学博士学位。同年至英国Leicester大学做博士后，1993年12月回国至浙江大学计算机系做博士后，1994年5月晋升为副研究员，1996年1月出站并留在浙江大学计算机系任教至今。主要研究方向为操作系统、人工智能、地理信息系统、并行算法等。



庄越挺，1965年6月出生于浙江慈溪。获浙江大学计算机应用博士学位。现为浙江大学计算机系教授、系副主任、博士生导师，浙江省“151人才工程”第一层次培养人员，曾多次获浙江大学优秀青年教师称号。主要的研究领域为多媒体数据库及信息检索、智能动画、人工智能、CAD等，主持过国家基础研究“攀登计划”、国家“八五”攻关、国家自然科学基金等10多个项目，并取得丰硕的科研成果，92年获中科院科技进步一等奖，93年获国家科技进步二等奖。共发表50多篇学术论文。1997年2月至1998年8月间，获浙江大学包氏奖学金资助，赴美国伊利诺斯大学Urbana-Champaign(UIUC)的计算机系和Beckman研究中心作访问学者。

前 言

本书从一个新颖的角度对人工智能(Artificial Intelligence, 简称AI)各方面的问题进行了探讨, 由浅入深地介绍整个人工智能系统或agent的发展历程。首先, 将介绍仅能对周围环境中可感知特征做出反应的原始agent, 以及这些简单的机器所涉及的机器视觉、机器学习和机器进化等问题; 然后, 将逐步介绍使agent可以在无法立即感知的任务环境中获取信息的技术。这些信息可以采用环境状态、环境图标模型、状态空间图和逻辑表示等描述性信息的形式。因为AI的发展历程类似于动物的进化过程, 因此我称其为演化人工智能(*evolutionary artificial intelligence*)。希望本书不仅是对人工智能技术的介绍, 而且能为研究人工智能提供参考(建议)。为此, 书中的例子为人工智能的学习提供激励和基础。

尽管我运用agent来说明人工智能技术, 但人工智能技术本身拥有更广泛的应用。许多人工智能研究者的思想已经渗透到计算机科学中, AI已广泛应用于专家系统、自然语言处理、人机交互、信息检索、图形图象处理、数据挖掘和机器人技术(对此将会举例说明)。这里, agent旨在把一系列看似不相关的主题组织到一起。

本书涉及的范围, 将力图控制在理论和实践之间的中间地带。这一地带拥有丰富而重要的人工智能的思想, 并且, 在本书中我将尽力选取并说明那些在人工智能领域中具有经久不衰的价值观点(当然, 在选取论题并做出结论时不可避免地会出现遗漏和错误)。同时, 在书中将对某些论题进行深入探讨——不仅因为这些论题更加重要, 而且我想要在书中提供一些深层剖析的例子。虽然书中出现了伪代码算法, 但本书并非人工智能编程的教材(“人工智能技术”的书包括: [Shoham 1994, Norvig 1992, Tracy & Bouthoorn 1997])。我并不对所有重要的理论结果给出证明, 但会对那些形式证明尽力提供直观的论据和引用。我的目的是为一个学期的大学初级课程提供一本厚度适中的入门教材, 激发学生和读者的兴趣, 为进一步学习更高级的人工智能课程做好准备, 同时使大量关于人工智能的文献易于查阅。

本书的一个打破常规之处是机器学习(*machine learning*)并未作为单独的论题讨论, 而是将其贯穿本书始终。首先讨论神经网络(*neural net*)和受监督的学习(*supervised learning*)的基本思想; 接着在“搜索(*search*)”章节中将讨论学习启发式搜索和动作策略的技术; 然后, 在有关“逻辑(*logic*)”章节之后将讨论规则学习(*rule learning*), 归纳逻辑编程(*inductive logic programming*)和基于解释的学习(*explanation-based learning*); 最后, 在讨论了基于逻辑的计划(*logic-based planning*)之后, 将讨论有关学习规划(*learning plan*)。

我以前的书中每章末尾均提供“参考书目和历史评价”(有的读者或许觉得仍有用), 但是在这本书中我并没有这样做。因为随着人工智能的发展, 它所包含的内容已经愈加广泛, 而且另一本更加详尽的教材已经作了这项工作(Russell & Norvig 1995)。但我在本书中引入了适当的评价和引用, 并且在多数章节末尾的讨论小节中给出了另外一些讨论。那些有志于以人工智能为研究方向的学生可以查阅这些参考书, 希望这些大量的引用不会给一般读者带来困扰。

每章末尾均附有习题, 有些只是书中概念的简单应用, 有些则稍具挑战性, 难度不一。我希望教师能根据自己的需要扩充习题, 包括上机练习和项目编程(为了与以思想而不是以程序

为中心的初衷保持一致，我并没有在书中涉及任何计算机上机练习和项目编程。在专门的人工智能编程技术教材中可找到相关内容。）

本书使用如下的排版约定：黑体大写字母如W和X用来表示矢量、矩阵和模式操作符。小写希腊字母表示谓词演算表达式和子表达式所涉及的元变量，有时也表示替换。大写希腊字母用来表示谓词演算公式的集合。小写字母p表示概率。

通过万维网（WWW），学生和研究者可以找到大量关于人工智能的资料，这里并没有列出它们的网址，因为现在列出的清单数月后会变得不完整和不确切。另外，通过Web搜索引擎，读者可迅速查找到应用实例、常见问题、参考书目、研究论文、程序、交互演示、研究所和会议的公告及研究者的主页等等。

在出版者的Web站点www.mkp.com/nils的网页上可找到本书的相关资料，如发现错误，请通过以下地址给出版商发电子邮件：aibugs@mkp.com。错误及更正可以在以下网址中找到：<http://www.mkp.com/nils/clarified>。

我的前一本人工智能教材《人工智能原理》（Principles of Artificial Intelligence）现在已经过时，但书中某些内容仍有价值，因此本书的编写直接采用了这些内容。同时，与其他人工智能教材（特别是[Russell & Norvig 1995, Rich & Knight 1991, Stefik 1995]）相互对照学习，亦十分有益。

14.4 归结反驳搜索策略	142	17.5.3 基于解释的一般化	183
14.4.1 排序策略	143	17.6 补充读物和讨论	184
14.4.2 精确策略	143	第18章 表示常识知识	187
14.5 Horn子句	144	18.1 常识世界	187
第15章 谓词演算	146	18.1.1 什么是常识知识	187
15.1 动机	146	18.1.2 表示常识知识的困难	188
15.2 谓词演算语言和它的句法	146	18.1.3 常识知识的重要性	189
15.3 语义	147	18.1.4 研究领域	189
15.3.1 世界	147	18.2 时间	190
15.3.2 解释	147	18.3 用网络表示知识	191
15.3.3 模型及其相关的概念	148	18.3.1 分类的知识	191
15.3.4 知识	149	18.3.2 语义网络	192
15.4 量化	150	18.3.3 语义网络的非单调推理	193
15.5 量词语义学	150	18.3.4 框架	194
15.5.1 全称量词	150	18.4 补充读物和讨论	194
15.5.2 存在量词	151	第19章 用不确定信息进行推理	197
15.5.3 有用的等价式	151	19.1 概率论简介	197
15.5.4 推理规则	151	19.1.1 基本思想	197
15.6 谓词演算作为一种表示知识的语言	151	19.1.2 条件概率	199
15.6.1 概念化	151	19.2 概率推理	201
15.6.2 举例	152	19.2.1 一个一般的方法	201
15.7 补充读物和讨论	153	19.2.2 条件独立	202
第16章 谓词演算中的归结	155	19.3 贝叶斯网	203
16.1 合一	155	19.4 贝叶斯网的推理模式	204
16.2 谓词演算归结	157	19.5 不确定证据	205
16.3 完备性和合理性	158	19.6 D分离	205
16.4 把任意的合式公式转化为子句形式	158	19.7 在polytree中的概率推理	206
16.5 用归结证明定理	160	19.7.1 证据在上方	207
16.6 回答提取	161	19.7.2 证据在下方	208
16.7 等式谓词	161	19.7.3 证据在上下两方	209
16.8 补充读物和讨论	163	19.7.4 一个数值例子	210
第17章 基于知识的系统	166	19.8 补充读物和讨论	211
17.1 面对现实世界	166	第20章 用贝叶斯网学习和动作	214
17.2 用Horn子句进行推理	166	20.1 学习贝叶斯网	214
17.3 动态知识库的维持	170	20.1.1 已知网络结构	214
17.4 基于规则的专家系统	173	20.1.2 学习网络结构	216
17.5 规则学习	176	20.2 概率推理与动作	219
17.5.1 学习命题演算规则	177	20.2.1 一般设置	219
17.5.2 学习一阶逻辑规则	180	20.2.2 一个扩展的例子	220

20.2.3 一般化举例	222	23.2.1 模型种类	255
20.3 补充读物和讨论	223	23.2.2 模拟策略	256
第四部分 基于逻辑的规划方法			
第21章 状态演算	227	23.2.3 模拟数据库	257
21.1 状态和动作推理	227	23.2.4 有意识思维方式	257
21.2 存在的一些困难	229	23.3 知识模式逻辑	258
21.2.1 框架公理	229	23.3.1 模式算子	258
21.2.2 条件	230	23.3.2 知识公理	259
21.2.3 分枝	230	23.3.3 关于其他agent知识的推理	260
21.3 生成计划	231	23.3.4 预测其他agent的动作	261
21.4 补充读物和讨论	231	23.4 补充读物和讨论	261
第22章 规划	234	第24章 agent之间的通信	263
22.1 STRIPS规划系统	234	24.1 交谈	263
22.1.1 描述状态和目标	234	24.1.1 计划交谈	264
22.1.2 向前搜索方法	235	24.1.2 实现交谈	264
22.1.3 递归STRIPS	236	24.2 理解语言字符串	265
22.1.4 带有运行时条件的计划	238	24.2.1 短语结构语法	265
22.1.5 Sussman异常	238	24.2.2 语义分析	267
22.1.6 向后搜索方法	239	24.2.3 扩展语法	271
22.2 计划空间和部分有序规划	242	24.3 有效通信	272
22.3 层次规划	246	24.3.1 上下文的使用	272
22.3.1 ABSTRIPS	246	24.3.2 使用知识解决歧义性	273
22.3.2 层次规划和部分有序规划的组合	248	24.4 自然语言处理	274
22.4 学习计划	248	24.5 补充读物和讨论	275
22.5 补充读物和讨论	250	第25章 agent体系结构	277
第五部分 通信与集成			
第23章 多agent	255	25.1 三级体系结构	277
23.1 交互agent	255	25.2 目标仲裁	278
23.2 其他agent模型	255	25.3 三层塔式结构	279
		25.4 自举	280
		25.5 补充读物和讨论	280
		参考文献	282

第1章 绪 论

我认为，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、情感和优先权如何发展和运用；传感器信号如何转换成各种符号；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；智能机制如何产生幻觉、信念、希望、畏惧、梦幻甚至善良和爱情等现象。我相信，对上述内容有一个根本的理解将会成为与拥有原子物理、相对论和分子遗传学等级相当的科学成就。

— James Albus “答复 Henry Hexmoor”，摘自URL：

<http://tommy.jsc.nasa.gov/er6/mrl/papers/symposium/albus.txt>

1995年2月13日

1.1 什么是人工智能

广义地讲，人工智能是关于人造物的智能行为，而智能行为包括知觉、推理、学习、交流和在复杂环境中的行为。人工智能的一个长期目标是发明出可以像人类一样或能更好地完成以上行为的机器；另一个目标是理解这种智能行为是否存在于机器、人类或其他动物中。因此，人工智能包含了科学和工程的双重目标。本书主要从工程角度讨论AI，集中说明构成智能机器设计基础的重要概念和思想。

长期以来，围绕着人工智能有很多争议。“机器是否能思考？”这一问题吸引了许多哲学家、科学家和工程师。在一篇著名的文章中，计算机科学的创始人之一，艾伦·图灵（Alan Turing），重述了这一问题，使其更经得起一种实验的测试，这种测试后来被称为图灵测试[Turing 1950]。下面将描述这一测试，但图灵同时指出对“机器是否能思考”这一问题的答案取决于人们如何定义“机器”和“思考”。他也许还可指出，这一问题还依赖于人们如何定义“能”。

让我们先来考虑“能”这个词。我们认为机器现在或将来能思考吗？我们认为原则上机器应该可以思考吗（即使我们不可能制造出这样的机器）？或者，我们真的要求实际的演示吗？由于人造物尚未具有广泛的思考能力，这些问题就变得非常重要。

一些人认为，能够思考的机器必定十分复杂且拥有复杂的经验（如与其所处的环境和其他能够思考的机器交流）。以致于我们永远也无法设计并制造出它们。产生全球气象的过程是一个很好的例子。尽管我们知道有关天气的一切重要现象，这些知识也无法让我们完整、详尽地复制天气现象。因为再没有比地球表层、大气层和海洋这些存在于宇宙之中、汲取太阳的光和热并受潮汐影响的更简单的系统能够完整详尽地复制天气现象了。同样，完全与人类相当的智能会十分复杂，或者至少会十分依赖于人类严密的生理机能，从而使其不能脱离处于特定环境的人的主体（*embodiment*）而单独存在（关于“主体”这一概念的重要性的讨论，可参见[Lakoff 1987, Winograd & Flores 1986, Harnad 1990, Mataric 1997]）。至于我们是否能造出与人类水平相当的能思考的机器仍无定论。但人工智能朝着这一目标的发展是坚定不移的，虽然这一进展比早期开创者们的预计要慢。我对我们最终的胜利持乐观态度。

接着，我们考虑“机器”这一词。许多人认为，机器是一种相当愚钝的东西，它总让人联

想起齿轮转动、蒸汽嘶嘶、钢铁铿锵的景象。这样的机器能思考吗？但是，如今计算机已大大延伸了“机器”这一概念。同时，我们对生物机制的理解也有了前所未有的进展。譬如：一种名为E6抗菌素的简单过滤性病毒（如图1-1所示），它头部含过滤性病毒DNA。它用尾部须根与一个细菌的细胞壁相连，先刺入细胞壁，再将其DNA注入此细菌中，然后这些DNA使此细菌产生成千上万这一过滤性病毒DNA的复制品。这些复制品自动集合而形成新的过滤性病毒后，离开这一细菌，再重复以上过程。这一完整的集合看起来、运作起来均像一台机器，我们还不如其称之为由蛋白质构成的机器。

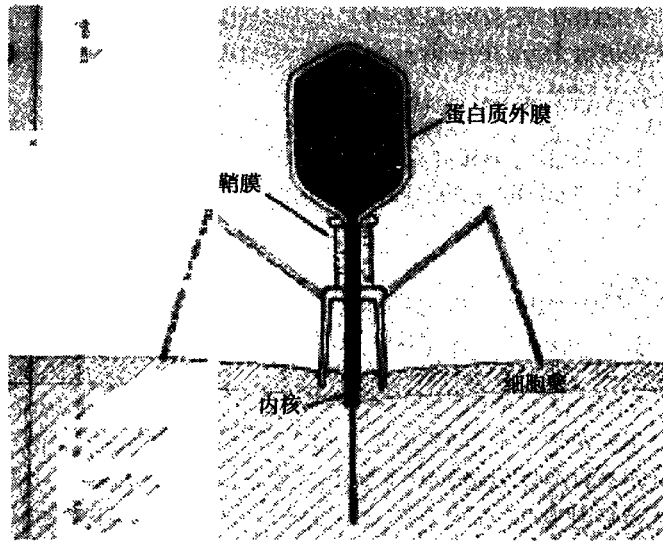


图1-1 E6抗菌素的示意图

其他生物过程和有机物又如何呢？最近，细菌*Haemophilus influenzae* Rd的基因排列已经完全被破解[Fleischmann, et al. 1995]。这一基因有1 830 137个基对（由字母A、G、C和T组成），大概占 3.6×10^6 比特；约半兆字节。尽管科学家们还不清楚其中1743个基因的功能，但他们已经开始像解释机器（当然是非常复杂的机器）的发展和功能那样解释这些机制的发展和功能。运用逻辑电路时序图这一计算机科学家熟知的技术，对理解可感染细菌的过滤性病毒的复杂生化基因的规则是十分有益的[McAdams & Shapiro 1995]。对其他有机物包括人类的基因组的排序工作正在进行中。一旦知道这些结果，我们会把细菌、寄生虫、果蝇、老鼠、海豚与人类一起视为机器吗？如果人类是机器，那么机器便能思考！因为我们拥有活生生的证据。只是我们还不知道人类这种机器是如何运作的。

即使我们就什么是“机器”达成一致观点，这种观点仍存在其他许多争议。尽管由蛋白质构成的机器能思考，由硅片构成的机器则未必能。一位知名哲学家John Searle认为，我们由什么构成直接影响着我们的智能[Searle 1980, Searle 1992]。他认为思考仅发生在那些十分特殊的机器上——有生命且由蛋白质构成的机器。

与Searle的观念（和前面所提到的“主体”概念）截然相反，Newell 和 Simon提出了物理符号系统假说[Newell & Simon 1976]。这一假说指出，物理符号系统具备必要且足够的方法来进行普通智能行为。Newell 和 Simon指出物理符号系统是类似数字计算机的机器，具备灵活处理符号数据的能力——加数、重排符号序列（如按字母顺序排列一组姓名）及符号替换等等。

这一假说的重要之一是它指出这种物理符号系统由什么构成并不重要！这一假说是完全中性的。一个智能实体只要能处理符号，它可以由蛋白质、机械传动、半导体或其他什么构成[⊖]。

还有一些人也认为，机器是由蛋白质还是由硅片构成无关紧要，但他们认为智能行为是他们所谓的“亚符号处理”，即“信号处理”，而不是“符号处理”的结果。如识别熟悉面孔对人类来说易如反掌，而我们却不知道机器该如何运作。他们认为这一过程最好的解释便是人类把图象或图象各部分作为多维信号而不是符号来处理。

能列出很多其他关于什么样的机器才具有人类的思维能力的看法，我们经常听到的有：

- 人脑对信息进行并行处理，而传统的计算机则是串行处理。我们需要建造各种新型的并行计算机来加快人工智能的发展。
- 传统的计算机以非真即假（双态）逻辑为基础，而真正的智能系统应运用某种模糊逻辑。
- 动物神经元远比开关——计算机的基本模块——要复杂，我们需要在智能机器中运用更现实的人造神经元。

尽管许多人工智能研究者接受物理符号系统这一假说，但在人工智能领域关于究竟需要哪种机器达成共识还为时过早。

最后，我们来看看“思考”这一最难的词。图灵没有企图对这个词下定义，只是提出了“图灵测试”。通过这一测试即可判断某一特定机器是不是智能机器。这一测试最初被描绘成一种游戏。从图灵的文章中摘录如下[Turing 1950]：

游戏由一男（A）、一女（B）和一名询问者（C）（性别不限）进行。询问者单独在一间房间里与其他两人分别通过电传打字机联系。在游戏中，询问者的目的是分辨两人的性别。开始，他只知道两人的称呼X、Y，最终，他需要在“X是A，Y是B”或者“X是B，Y是A”中选择答案。询问者允许问A和B以下问题：

C：X能告诉我你的头发的长度吗？

如果X是A，那么他必须回答。游戏中，A必须尽力使C判断错误。

...

而B的任务则是帮助询问者。

...

现在我们提出这样一个问题：一个机器代替游戏中的A会如何？询问者会依然像当游戏由一男一女进行时一样经常判断错误吗？这些问题代替了最初的问题：机器能思考吗？

图灵测试常被简化为让一个机器试图使询问者相信它是一个人。许多更简单的测试层出不穷，然而由于就连一些陈旧的机器也可以愚弄询问者一段时间，这些简单的测试已经不再被视为测试机器智能的良方了。譬如：Joseph Weizenbaum的ELIZA程序运用一些相当简单、但对一个宽容的使用者却是虽显空洞却十分现实的对话技巧。Mauldin的JULIA程序是更新和更复杂的对话程序[Mauldin 1994][⊖]。

除了运用图灵测试，我们有必要在标榜一台机器是智能机器之前，了解这样的机器应具备

⊖ 当然，如果我们考虑到速度、永久性、可靠性、并行处理的适合性和温度敏感度等实际因素，模块材料必定有好有坏。

⊖ 1991年，Hugh Loebner开始举行一个有奖竞赛，他向第一个能通过无限制图灵测试的计算机程序的开发者提供10万美元的奖金。另外，每年这一竞赛都为能通过有限图灵测试的最佳程序的开发者提供数额略少的奖金。

怎样的能力。许多计算机程序已经完成了大量不可思议的事——设计高效省油的最佳航空路线、模拟全球气象状况、统筹安排工厂的机器使用等等。这些是智能程序吗？它们能体现人工智能的主旨吗？本书一开始我便描绘那些难以被人们称为智能机器的机器，随着其复杂性的增强，它们会变得越来越智能吗？毫无疑问，别人会有不同的观点，但至少我这样认为。

1.2 人工智能的研究方法

尽管人工智能已经创造了一些实用系统，但人们不得不承认这些远未达到人类的智能水平。正因为如此，就选择人工智能研究的最佳方法——既为人工智能的最终研究目标打好基础，又能创造出短期效益——存在大量的讨论和争辩。这样，在过去的四十年里涌现出大量方法，每一种方法均有其拥护者，有些甚至有趣得令人爱不释手。也许所有这些方法应该综合起来运用。总之，所有这些拥护者都认为自己的研究方法具有突破性进展，值得特别关注。其中的一些方法可分为两大类。

第一类包括符号处理的方法。它们基于Newell和Simon的物理符号系统的假说。尽管不是所有人都赞同这一假说，但几乎大多数被称为“经典的人工智能”（即哲学家John Haugeland所谓的“出色的老式人工智能”或GOFAI）均在其指导之下。这类方法中，突出的方法是将逻辑操作应用于说明性知识库。最早由John McCarthy的“采纳意见者”备忘录提出[McCarthy 1958]，这种风格的人工智能运用说明语句来表达问题域的“知识”，这些语句基于或实质上等同于一阶逻辑中的语句。采用逻辑推理可推导这种知识的结果。这种方法有许多变形，包括那些强调对逻辑语言中定义域的形式公理化的角色的变形。当遇到“真正的问题”，这一方法需要掌握问题域的足够知识，通常就称作基于知识的方法。许多系统的构建都运用了这些方法，在本书后面将会提到一些。

在大多数符号处理方法中，对需求行为的分析和为完成这一行为所做的机器合成要经过几个阶段。最高阶段是知识阶段，机器所需知识在这里说明。接下来是符号阶段，知识在这里以符号组织表示（例如列表可用列表处理语言LISP来描述），同时在这里说明这些组织的操作。接着，在更低级的阶段里实施符号处理。多数符号处理采用自上而下的设计方法，从知识阶段向下到符号和实施阶段。

第二类包括所谓的“子符号”方法。它们通常采用自下而上的方式，从最低阶段向上进行。在最低层阶段，符号的概念就不如信号这一概念确切了。在子符号方法中突出的方法是“*Animat approach*”。偏爱这种方式[Wilson 1991, Brooks 1990]的人们指出，人的智能经过了在地球上十亿年或更长时间的进化过程。他们认为，为了制造出真正的智能机器，我们必须沿着这些进化的步骤走。因此，我们必须集中研究复制信号处理的能力和简单动物如昆虫的支配系统，沿着进化的阶梯向上进行。这一方案不仅能在短期内创造实用的人造物，又能为更高级智能的建立打好坚实的基础。

第二类方法也强调符号基础。[Brooks 1990]将物理符号系统和他的物理基础假说相对照。在物理基础假说中，一个agent不采用集中式的模式而运用其不同的行为模块与环境相互作用来进行复杂的行为（然而，他也承认，要达到人类智能水平的人工智能也许需要将两种途径相结合）。

机器与环境的相互作用产生了所谓的“自然行为（*emergent behavior*）”。一名研究人员这样说[Maes 1990b, p.1]：

一个agent的功能可视为该系统与动态环境密切相互作用的自然属性。agent本身对其行为的说明并不能解释它运行时所表现的功能；相反，其功能很大程度上取决于环境的特性。不仅要动态地考虑环境，而且环境的具体特征也要运用于整个系统之中。

由于符号派制造的著名样品机器包括所谓的“神经网络 (Neural network)”。受到生物学方法的启发，这些系统主要因其学习的能力而十分有趣。根据模拟生物进化方面的进程，一些有趣的机器应运而生，包括：Sexual crossover、Mutation和Fitness-proportional reproduction。其他自下而上、含animat风格的方法是基于控制理论和动态系统的分析（参见[Beer 1995, Port & van Gelder 1995]）。

介于自上而下和自下而上之间的方法是一种动机“环境自动机 (situated automata)” [Kaelbling & Rosenschein 1990, Rosenschein & Kaelbling 1995]的方法。Kaelbling 和 Rosenschein 建议编写一种程序设计语言来说明agent在高水平上所要求的行为，并编写一编译程序，以从这种语言编写的程序中产生引发行为的线路。

1.3 人工智能简史

当20世纪40~50年代数字计算机研制成功时，几位研究者就编写了能够完成原始推理工作的程序。其中突出的是第一个可以下国际象棋[Shannon 1950, Newell, Shaw & Simon 1958]、担当实验员[Samuel 1959, Samuel 1967]和证明平面几何定理[Gelernter 1959]的计算机程序。1956年，John McCarthy和Claude Shannon合作编著了一本名为《Automata Studies》(自动机研究)的书[Shannon & McCarthy 1956]。由于对书中主要针对automata的数学理论感到遗憾，所以McCarthy决定把1956年的Dartmouth会议用人工智能来命名。在该次会议上发表了許多重要论文，包括由Allen Newell、Cliff Shaw和Herbert Simon 编写的名为《Logic Theorist》(逻辑理论家) [Newell, Shaw & Simon 1957]的程序，它可以证明命题逻辑中的定理。尽管人们试着用许多其他名称来为该领域命名，包括复杂信息处理、机器智能、启发式编程和认知技术，但人工智能这一名称最终保留下来。毫无疑问，这主要归因于一系列的教科书、大学课程、会议和期刊均用这一命名。

很久以前，亚里士多德（公元前384~322年）在着手解释和编纂他称之为三段论的演绎推理时就迈出了向人工智能发展的早期步伐。一些使智能自动化的努力对于今天来说显得太不实际。一位加泰罗尼亚的诗人兼神秘主义者，Ramon Llull（大约1235~1316年），构建了一套称为Ars Mgna的转轮，据说是一部可以回答任何问题的机器。同时许多科学家和数学家开始探讨推理自动化。Martin Gardner[Gardner 1982, p. 3]把“有一天所有的知识，包括精神和无形的真理，能够通过通用的代数演算放入一个单一的演绎系统”的梦想归功于莱布尼兹（1646~ 1716年）。莱布尼兹称这个系统为微积分原理机，或推理机。当然，这个梦想运用当时的技术设备是无法实现的。直到布尔[Boole 1854]建立并发展了命题逻辑，这方面才有了实质性的进展。布尔的意图是要“把有关人类意识的本质和构成的某些可能的暗示收集起来”。到了19世纪末期，Gottlieb Frege提出了用于机械推理的符号表示系统，从而发明了我们现在熟知的谓词演算[Frege 1879]，他称之为Begriffsschrift，可以译为“概念书写 (concept - writing)”。

1958年，John McCarthy 建议在他称之为“意见采纳者”的系统中采用谓词演算这种语言来表示和运用知识。这一系统被告知它所需要知道的而不是事先程序设计好的知识。Covdell

Green在他所谓的QA3系统中[Green 1969a]适度地、颇有影响地实现了这些思想。尽管在研究者中还存在大量争议，谓词演算和一些它的变形构成了人工智能知识表示的基础。

20世纪的逻辑学家，包括Kurt Gödel、Stephen Kleene、Emil Post、Alonzo Church和Alan Turing，对哪些能和哪些不能由逻辑和计算机系统完成的任务做了形式化分类。最近，计算机科学家，包括Stephen Cook和Richard Karp证明有些计算在原则上可能需要根本不切实际的时间和存储空间。

许多从逻辑学和计算机科学中所得到的结果是：“真理不可能被演绎”、“计算不可能被执行”。也许这些负面的发现令许多哲学家和其他人振奋，他们将之理解为再一次否定了人类的智能可以机械化[Lucas 1961, Penrose 1989, Penrose 1994]，他们猜想人类不存在机械所固有的计算局限。然而多数逻辑学家和计算机科学家却认为这些负面结果并不暗示机器具有人类所不具有的任何局限。

在现代，第一篇讨论把人类智能机械化的可能性的文章是由Alan Turing所著的（前面已经引用）[Turing 1950]。同一时期，Warren McCulloch和Walter Pitts总结出简单计算元素和生物神经元之间关系的理论[McCulloch & Pitts 1943]。他们证明了运用逻辑网络系统计算可计算功能的可能性（参见 [Minsky 1967]有关McCulloch Pitts神经元计算方面有价值的论述）。另外，由Frank Rosenblatt[Rosenblatt 1962]所著的书中探讨了称作perceptrons的网络由类似于神经元的部件组成运用于学习和模式识别的可行性。一些其他学派的工作，如控制论[Wiener 1948]、认知心理学、计算语言学[Chomsky 1965]和自适应控制理论[Widrow & Hoff 1960]，均对人工智能的发展作出了贡献。

许多人工智能的早期工作（从20世纪60年代至70年代初）探讨了问题表示、搜索技术和通用启发等一系列问题——并把它们运用于计算机程序中来解迷、博弈和检索信息，其中有影响的程序是由Allen Newell、Cliff Shaw和Herbert Simon[Newell, Shaw & Simon 1959, Newell & Simon 1963]共同编写的通用问题求解程序（General Problem Solver (GPS)）。由这些早期系统解决的实例问题包括符号集成[Slagle 1963]、代数词汇问题[Bobrow 1968]、类比难题[Evans 1968]及机器人的控制[Nilsson 1984b]。在这些系统中，许多都是《Computers and Thought》这卷书中的主题[Feigenbaum & Feldman 1963]。

为了应用于更重要的现实问题而对这些程序和技术进行升级的尝试表明这些系统只能解决“玩具问题”。更有效的系统要求对应用领域具有更多内在的知识。20世纪70年代末80年代初发展了一些更高级的程序，包括在完成一定任务时模拟专业人员的知识，如分析、设计和诊断等。一些表达具体问题的知识得到了探讨和发展。第一个能演示具体领域知识的重要程序DENDRAL是一个根据所提供的化学分子式和质谱分析图来预测有机物分子结构的系统[Feigenbaum, Buchanan & Lederberg 1971, Lindsay, et al. 1980]。接着，其他“专家系统”，包括医疗诊断[Shortliffe 1976, Miller, Pople & Myers 1982]、计算机系统的配置[McDermott 1982]和矿藏评估（evaluated potential ore deposits）[Campbell, et al. 1982, Duda, Gaschnig & Hart 1979]。[McCorduck 1979]撰写了这一阶段的人工智能简史。

通过升级游戏问题，博弈这一领域有了实质性的进展。1997年5月11日，一个名为“深蓝”的IBM程序在六局比赛中以3.5比2.5的总比分战胜了世界象棋冠军Garry Kasparov(盖利·卡斯帕洛夫)。这次成功是运用复杂的搜索算法、高速计算机和国际象棋专用硬件才得以实现的。

人类的智能包括洞察和分析可视场景、理解并运用语言等许多方面的能力。关于这些能力的专题均得到了高度重视。Larry Roberts开发了早期场景分析程序之一[Roberts 1963]。这一工作之后对机器视觉作了大量研究([Nalwa 1993]是一本很好的通用教材)，对动物视觉系统[Letvinn, et al. 1959, Hubel 1988, Marr 1982]也作了研究。一个早期自然语言理解系统也由Terry Winograd开发成功[Winograd 1972]。20世纪70年代，在一个多方项目中，连续语言理解系统原型被开发出来；由William Woods开发的LUNAR系统能回答用口语提出的关于由美国航空航天局(NASA)从月球收集的岩石样品的问题。尽管现在存在一些自然语言理解系统，但它们的能力仅局限于特定的话题和词汇。打破这些局限有待于开发出更大量的常识表示。CYC项目[Guha & Lenat 1990, Lenat & Guha 1990, Lenat 1995]的一个目标就是尽可能多地收集、表达这些所需的知识。

20世纪50年代末在Frank Rosenblatt所领导的开创性工作之后，对神经网络的研究虽然一度萎靡，但是到20世纪80年代又恢复了活力。具有强度可调互连系统的非线性元素网络如今已被视为一类重要的非线性建模工具。现在已存在神经网络方面的一些重要应用。动态方法与神经网络相结合，促使人工智能的研究集中到把符号处理过程与处于物理环境中的机器人的传感器与受动器联系起来的问题上来。

立足当前，展望未来，我认为人们将重视集成的、自治的系统——机器人和Softbots。Softbots[Etzioni & Weld 1994]是在互联网中查找他们认为用户会感兴趣的信息的软件agent。今后，不断提高和完善机器人和软件agent的能力将促进并引导人工智能研究。

1.4 本书规划

许多人工智能研究者已提出了一些有关智能机械化的观点和技术，我会在介绍一系列逐步弹大逐步复杂的agent时陈述这些内容。我们本可以考虑各类agent及其环境，如：在太空失重的情况下、在海底深水域中、在办公楼或工厂中及在互联网的“符号数据世界”中的机器人。然而，在这样的真实世界中，真正实用的agent将会十分复杂，这样会难以清晰地展示赋予agent智能的人工智能概念。因此，我将在“网格空间世界”这一假想空间中采用一系列“玩具”agent。虽然简单世界很容易描述，但各方面的发展使之变得太复杂而迫使其中的agent需要有智能才行。

网格空间世界是一个三维空间，它以二维的地面为界限，而地面是由一系列单元格组成。单元格集合可以容纳具有各种特性的物体。单元格集合之间可能会存在像墙一样的边界，agent不能离开地面，但可以在单元格之间移动。物体必须在地面上或必须由在地面上的其他物体支撑。有时我会采用仅包含地面的二维子空间。一个典型的网格空间世界如1-2图所示。图中有两个机器人，一个是原始的二维机器人，它用感知邻近单元格是否为空的传感器来判断是否向其移动；另一个略微复杂，它有一个可操纵物体的手臂。

熟悉人工智能历史的读者会发现网格空间世界能够被定制为其他许多用于人工智能研究

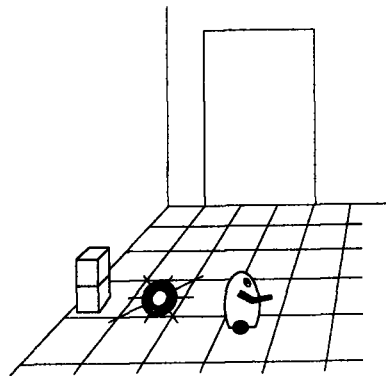


图1-2 网格空间世界