



高等学校教材

# 人工智能原理 及其应用

王万森 编著



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

URL: <http://www.phei.com.cn>

高等学校教材

# 人工智能原理及其应用

王万森 编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

本书全面系统地介绍了人工智能的基本原理、方法及研究应用领域。全书共 10 章,分为三大部分,第一部分第 1 章至第 5 章,介绍人工智能的基本原理和方法,包括人工智能概述和人工智能的三大技术(知识表示、推理及搜索);第二部分第 6 章至第 8 章,介绍人工智能的三个重要研究领域(机器学习、神经网络学习和自然语言理解);第三部分第 9 章和第 10 章,介绍人工智能的两个重要应用领域(专家系统和智能决策支持系统)。全书内容既符合国家学位委员会 1998 年 11 月颁发的《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全国统一考试大纲及指南》中的“人工智能考试大纲”的要求,也充分考虑到了人工智能学科的整体结构和最新研究进展。

本书可作为计算机学科申请硕士学位人员的学习用书,也可作为计算机、信息处理、自动化等学科的研究生和本科高年级学生教材,还可供相关专业技术人员使用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。  
版权所有,翻版必究。

### 图书在版编目(CIP)数据

人工智能原理及其应用/王万森编著. - 北京:电子工业出版社,2000.9  
高等学校教材  
ISBN 7-5053-5305-5

I.人... II.王... III.人工智能-高等学校-教材 IV.TP18

中国版本图书馆 CIP 数据核字(2000)第 40859 号

丛 书 名: 高等学校教材

书 名: 人工智能原理及其应用

编 著 者: 王万森

策划编辑: 应月燕

责任编辑: 张孟玮

特约编辑: 朱强国

排版制作: 电子工业出版社计算机排版室

印 刷 者: 北京东郊印刷厂

装 订 者: 三河市新伟装订厂

出版发行: 电子工业出版社 URL: <http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销: 各地新华书店

开 本: 787×1092 1/16 印张: 19 字数: 486 千字

版 次: 2000 年 9 月第 1 版 2000 年 9 月第 1 次印刷

书 号: ISBN 7-5053-5305-5  
G·460

印 数: 4000 册 定价: 26.00 元

凡购买电子工业出版社的图书,如有缺页、倒页、脱页、所附磁盘或光盘有问题者,请向购买书店调换;  
若书店售缺,请与本社发行部联系调换。电话 68279077

## 前 言

随着信息社会和知识经济时代的来临,信息正在以前所未有的速度膨胀。面对浩如烟海的庞大信息资源,人类的自然智能越来越显得难于驾驭。如何用人造的智能去模仿和扩展人类的自然智能,实现信息的智能化处理,是信息社会所面临的一个重大课题。

人工智能作为一门研究机器智能的学科,其目的是要用人工的方法和技术,研制智能机器或智能系统,来模仿、延伸和扩展人的智能。因此,它是人类迈向信息社会、迎接知识经济挑战所必须具备的一项核心技术。

信息社会对智能的巨大需求是人工智能发展的强大动力。人工智能自 1956 年问世以来,已经取得了引人瞩目的成就,形成了专家系统、机器学习、自然语言理解、人工神经网络、智能决策支持系统等诸多研究和应用领域。尤其是近几年来,随着计算机网络、Internet、多媒体、分布式人工智能和开放分布式环境下的多智能体协同工作等计算机主流技术的兴起,人工智能又迎来了一个蓬勃发展的新时期。

由于人类对大脑的认识尚处于探索阶段,人类智能的奥秘还远未彻底揭示,因此,研究人类智能的机理、用机器代替人脑,还任重而道远。

全书共分 10 章。第 1 章介绍了人工智能的基本概念、研究领域、学派之争及发展趋势;第 2 章讨论了知识表示的基本概念和各种确定性知识表示方法;第 3 章讨论了推理的基本概念及归结、演绎等确定性推理方法;第 4 章讨论了不确定性推理的有关概念及各种不确定性的表示和推理方法;第 5 章讨论了搜索的基本概念和状态空间、与或树的各种搜索算法;第 6 章讨论了机器学习的基本概念和各种符号学习方法;第 7 章讨论了人工神经网络的概念和各种连接学习方法;第 8 章讨论了自然语言理解的基本概念和分析方法;第 9 章介绍了人工智能的一个重要应用领域,即专家系统,它目前正在从集中、封闭模式向分布、开放模式发展;第 10 章介绍了人工智能的另一个重要应用领域,即智能决策支持系统,它是目前迅速兴起的网络商务中的一项重要技术,有着广阔的应用前景。

本书的内容安排既符合国家学位委员会 1998 年 11 月颁发的《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全国统一考试大纲及指南》中的“人工智能考试大纲”的要求,也充分考虑了人工智能学科的整体结构和最新研究进展。

本书是作者在多年人工智能教学基础上形成的。作为一本教科书,它既包含了作者的研究与教学实践,也吸取了国内外同类教材和有关文献中的精华。在此,谨向这些教材和文献的作者表示感谢。

在本书的编写、出版过程中,得到了电子工业出版社的大力支持,在此谨表示诚挚的谢意。

由于作者水平所限,加之时间仓促,书中难免存在一些缺点和错误,恳请各位专家和读者不吝指教。

王万森

2000 年 3 月于北京

# 目 录

|                           |        |
|---------------------------|--------|
| <b>第 1 章 人工智能概述</b> ..... | ( 1 )  |
| 1.1 人工智能及其研究目标 .....      | ( 1 )  |
| 1.1.1 人工智能的定义 .....       | ( 1 )  |
| 1.1.2 人工智能的研究目标 .....     | ( 4 )  |
| 1.2 人工智能的产生与发展 .....      | ( 4 )  |
| 1.2.1 孕育期 .....           | ( 4 )  |
| 1.2.2 形成期 .....           | ( 6 )  |
| 1.2.3 知识应用期 .....         | ( 7 )  |
| 1.2.4 综合集成期 .....         | ( 9 )  |
| 1.3 人工智能研究的基本内容及其特点 ..... | ( 9 )  |
| 1.3.1 人工智能研究的基本内容 .....   | ( 9 )  |
| 1.3.2 人工智能研究的特点 .....     | ( 11 ) |
| 1.4 人工智能的研究和应用领域 .....    | ( 12 ) |
| 1.4.1 机器学习 .....          | ( 12 ) |
| 1.4.2 自然语言理解 .....        | ( 13 ) |
| 1.4.3 专家系统 .....          | ( 13 ) |
| 1.4.4 模式识别 .....          | ( 13 ) |
| 1.4.5 计算机视觉 .....         | ( 14 ) |
| 1.4.6 机器人学 .....          | ( 15 ) |
| 1.4.7 博弈 .....            | ( 16 ) |
| 1.4.8 自动定理证明 .....        | ( 16 ) |
| 1.4.9 自动程序设计 .....        | ( 17 ) |
| 1.4.10 智能控制 .....         | ( 17 ) |
| 1.4.11 智能决策支持系统 .....     | ( 17 ) |
| 1.4.12 人工神经网络 .....       | ( 17 ) |
| 1.4.13 知识发现和数据挖掘 .....    | ( 18 ) |
| 1.4.14 分布式人工智能 .....      | ( 18 ) |
| 1.5 人工智能研究的不同学派及其争论 ..... | ( 18 ) |
| 1.5.1 人工智能的三大学派 .....     | ( 18 ) |
| 1.5.2 人工智能理论的争论 .....     | ( 20 ) |
| 1.5.3 人工智能研究方法的争论 .....   | ( 20 ) |
| 1.6 人工智能的近期发展分析 .....     | ( 21 ) |
| 习题 .....                  | ( 22 ) |
| <b>第 2 章 知识表示</b> .....   | ( 23 ) |
| 2.1 知识与知识表示的概念 .....      | ( 23 ) |

|            |               |             |
|------------|---------------|-------------|
| 2.1.1      | 知识            | (23)        |
| 2.1.2      | 知识表示          | (26)        |
| 2.2        | 一阶谓词逻辑表示法     | (27)        |
| 2.2.1      | 一阶谓词逻辑表示的逻辑基础 | (28)        |
| 2.2.2      | 谓词逻辑表示方法      | (31)        |
| 2.2.3      | 谓词逻辑表示的应用     | (32)        |
| 2.2.4      | 谓词逻辑表示的特性     | (36)        |
| 2.3        | 产生式表示法        | (37)        |
| 2.3.1      | 产生式表示的基本方法及特性 | (37)        |
| 2.3.2      | 产生式系统的基本结构    | (38)        |
| 2.3.3      | 产生式系统的基本过程    | (40)        |
| 2.3.4      | 产生式系统的控制策略    | (41)        |
| 2.3.5      | 产生式系统的类型      | (42)        |
| 2.3.6      | 产生式系统的特点      | (43)        |
| 2.4        | 语义网络表示法       | (44)        |
| 2.4.1      | 语义网络的基本概念     | (45)        |
| 2.4.2      | 事物和概念的表示      | (47)        |
| 2.4.3      | 情况和动作的表示      | (49)        |
| 2.4.4      | 逻辑关系的表示       | (51)        |
| 2.4.5      | 语义网络的推理过程     | (52)        |
| 2.4.6      | 语义网络表示法的特征    | (53)        |
| 2.5        | 框架表示法         | (54)        |
| 2.5.1      | 框架理论          | (54)        |
| 2.5.2      | 框架和实例框架       | (54)        |
| 2.5.3      | 框架系统          | (56)        |
| 2.5.4      | 框架系统的问题求解过程   | (63)        |
| 2.5.5      | 框架表示法的特性      | (64)        |
| 2.6        | 脚本表示法         | (65)        |
| 2.6.1      | 脚本的结构         | (65)        |
| 2.6.2      | 脚本的推理         | (67)        |
| 2.7        | 过程表示法         | (67)        |
| 2.7.1      | 表示知识的方法       | (68)        |
| 2.7.2      | 过程表示的问题求解过程   | (68)        |
| 2.7.3      | 过程表示的特性       | (69)        |
| 2.8        | 面向对象表示法       | (70)        |
| 2.8.1      | 面向对象的基本概念和特征  | (70)        |
| 2.8.2      | 知识的面向对象表示     | (71)        |
|            | 习题            | (72)        |
| <b>第3章</b> | <b>确定性推理</b>  | <b>(74)</b> |
| 3.1        | 推理的基本概念       | (74)        |

|              |                        |              |
|--------------|------------------------|--------------|
| 3.1.1        | 什么是推理 .....            | (74)         |
| 3.1.2        | 推理方法及其分类 .....         | (74)         |
| 3.1.3        | 推理的控制策略及其分类 .....      | (77)         |
| 3.1.4        | 正向推理 .....             | (77)         |
| 3.1.5        | 逆向推理 .....             | (78)         |
| 3.1.6        | 混合推理 .....             | (80)         |
| 3.1.7        | 推理的冲突消解策略 .....        | (82)         |
| 3.2          | 推理的逻辑基础 .....          | (83)         |
| 3.2.1        | 谓词公式的解释 .....          | (83)         |
| 3.2.2        | 谓词公式的永真性与可满足性 .....    | (85)         |
| 3.2.3        | 谓词公式的等价性与永真蕴含性 .....   | (85)         |
| 3.2.4        | 谓词公式的范式 .....          | (87)         |
| 3.2.5        | 置换与合一 .....            | (87)         |
| 3.3          | 自然演绎推理 .....           | (90)         |
| 3.4          | 归结演绎推理 .....           | (91)         |
| 3.4.1        | 子句集及其化简 .....          | (91)         |
| 3.4.2        | 海伯伦理论 .....            | (95)         |
| 3.4.3        | 鲁宾逊归结原理 .....          | (98)         |
| 3.4.4        | 归结演绎推理的归结策略 .....      | (106)        |
| 3.4.5        | 用归结反演求取问题的答案 .....     | (109)        |
| 3.5          | 基于规则的演绎推理 .....        | (111)        |
| 3.5.1        | 规则正向演绎推理 .....         | (111)        |
| 3.5.2        | 规则逆向演绎推理 .....         | (115)        |
| 3.5.3        | 规则双向演绎推理 .....         | (118)        |
| 3.6          | 规则演绎推理的剪枝策略 .....      | (119)        |
|              | 习题 .....               | (119)        |
| <b>第 4 章</b> | <b>不确定与非单调推理 .....</b> | <b>(122)</b> |
| 4.1          | 不确定性推理的基本概念 .....      | (122)        |
| 4.1.1        | 不确定性推理的含义 .....        | (122)        |
| 4.1.2        | 不确定性推理的基本问题 .....      | (123)        |
| 4.1.3        | 不确定性推理的类型 .....        | (124)        |
| 4.2          | 不确定性推理的概率论基础 .....     | (125)        |
| 4.2.1        | 样本空间与随机事件 .....        | (125)        |
| 4.2.2        | 事件的概率 .....            | (126)        |
| 4.2.3        | 全概率公式与 Bayes 公式 .....  | (127)        |
| 4.3          | 确定性理论 .....            | (128)        |
| 4.3.1        | 可信度的概念 .....           | (128)        |
| 4.3.2        | C-F 模型 .....           | (128)        |
| 4.3.3        | 带加权因子的可信度推理 .....      | (133)        |
| 4.4          | 主观 Bayes 方法 .....      | (135)        |

|            |              |              |
|------------|--------------|--------------|
| 4.4.1      | 知识不确定性的表示    | (135)        |
| 4.4.2      | 证据不确定性的表示    | (137)        |
| 4.4.3      | 组合证据不确定性的计算  | (138)        |
| 4.4.4      | 不确定性的更新      | (138)        |
| 4.4.5      | 结论不确定性的合成    | (140)        |
| 4.5        | 证据理论         | (142)        |
| 4.5.1      | D-S理论的形式描述   | (142)        |
| 4.5.2      | 证据理论的推理模型    | (147)        |
| 4.5.3      | 推理实例         | (151)        |
| 4.6        | 可能性理论和模糊推理   | (153)        |
| 4.6.1      | 模糊逻辑基础       | (153)        |
| 4.6.2      | 模糊知识表示       | (158)        |
| 4.6.3      | 模糊概念的匹配      | (160)        |
| 4.6.4      | 模糊推理         | (161)        |
| 4.7        | 非单调推理        | (165)        |
|            | 习题           | (166)        |
| <b>第5章</b> | <b>搜索策略</b>  | <b>(169)</b> |
| 5.1        | 搜索的基本概念      | (169)        |
| 5.1.1      | 搜索的含义        | (169)        |
| 5.1.2      | 状态空间法        | (169)        |
| 5.1.3      | 问题归约         | (172)        |
| 5.2        | 状态空间的盲目搜索    | (175)        |
| 5.2.1      | 一般图搜索过程      | (175)        |
| 5.2.2      | 广度优先搜索       | (177)        |
| 5.2.3      | 深度优先搜索       | (178)        |
| 5.2.4      | 有界深度优先搜索     | (180)        |
| 5.2.5      | 代价树搜索        | (181)        |
| 5.3        | 状态空间的启发式搜索   | (183)        |
| 5.3.1      | 启发性信息和估价函数   | (183)        |
| 5.3.2      | A算法          | (184)        |
| 5.3.3      | A*算法         | (185)        |
| 5.3.4      | A*算法应用举例     | (190)        |
| 5.4        | 与/或树的盲目搜索    | (191)        |
| 5.4.1      | 与/或树的一般搜索    | (191)        |
| 5.4.2      | 与/或树的广度优先搜索  | (192)        |
| 5.4.3      | 与/或树的深度优先搜索  | (193)        |
| 5.5        | 与/或树的启发式搜索   | (193)        |
| 5.5.1      | 解树的代价与希望树    | (194)        |
| 5.5.2      | 与/或树的启发式搜索过程 | (195)        |
| 5.6        | 博弈树的启发式搜索    | (196)        |



|              |                       |       |
|--------------|-----------------------|-------|
| 5.6.1        | 概述                    | (196) |
| 5.6.2        | 极大极小过程                | (197) |
| 5.6.3        | $\alpha$ - $\beta$ 剪枝 | (198) |
| 习题           |                       | (199) |
| <b>第 6 章</b> | <b>机器学习</b>           | (202) |
| 6.1          | 机器学习的基本概念             | (202) |
| 6.1.1        | 学习和机器学习               | (202) |
| 6.1.2        | 机器学习的发展过程             | (203) |
| 6.1.3        | 学习系统                  | (204) |
| 6.1.4        | 机器学习的分类               | (207) |
| 6.2          | 机械式学习                 | (208) |
| 6.3          | 指导式学习                 | (209) |
| 6.4          | 归纳学习                  | (210) |
| 6.4.1        | 归纳学习的类型               | (210) |
| 6.4.2        | 示例学习                  | (211) |
| 6.4.3        | 观察与发现学习               | (214) |
| 6.5          | 基于类比的学习               | (219) |
| 6.5.1        | 类比学习的概念               | (219) |
| 6.5.2        | 属性类比学习                | (220) |
| 6.5.3        | 转换类比学习                | (221) |
| 6.6          | 基于解释的学习               | (222) |
| 6.6.1        | 解释学习概述                | (222) |
| 6.6.2        | 解释学习的空间描述及学习模型        | (223) |
| 6.6.3        | 解释学习的基本原理             | (224) |
| 6.6.4        | 解释学习的基本过程             | (224) |
| 6.6.5        | 领域知识的完善性              | (226) |
| 习题           |                       | (226) |
| <b>第 7 章</b> | <b>神经网络及连接学习</b>      | (227) |
| 7.1          | 人工神经网络概述              | (227) |
| 7.1.1        | 生物神经元及脑神经系统的结构与特征     | (227) |
| 7.1.2        | 人工神经元及人工神经网络          | (229) |
| 7.1.3        | 人工神经网络的发展过程           | (231) |
| 7.1.4        | 人工神经网络的局限性            | (232) |
| 7.2          | 人工神经网络的互连结构及其学习机理     | (232) |
| 7.2.1        | 人工神经网络的互连结构           | (232) |
| 7.2.2        | 人工神经网络学习和记忆的心理学基础     | (234) |
| 7.2.3        | 人工神经网络的学习算法           | (234) |
| 7.3          | 感知器模型及其学习             | (237) |
| 7.3.1        | 感知器模型                 | (237) |
| 7.3.2        | 感知器的学习                | (237) |

|                               |              |
|-------------------------------|--------------|
| 7.3.3 有关感知器 XOR 问题求解的讨论 ..... | (238)        |
| 7.4 误差反向传播网络及其学习 .....        | (240)        |
| 7.4.1 B-P 网络结构 .....          | (240)        |
| 7.4.2 B-P 网络学习的传播公式 .....     | (241)        |
| 7.4.3 B-P 网络的学习算法 .....       | (242)        |
| 7.4.4 B-P 网络学习的讨论 .....       | (243)        |
| 7.5 Hopfield 网络及其学习 .....     | (243)        |
| 7.5.1 Hopfield 网络的结构 .....    | (243)        |
| 7.5.2 Hopfield 模型的稳定性 .....   | (244)        |
| 7.5.2 Hopfield 网络学习算法 .....   | (244)        |
| 习题 .....                      | (245)        |
| <b>第 8 章 自然语言理解</b> .....     | <b>(246)</b> |
| 8.1 语言及其理解的基本概念 .....         | (246)        |
| 8.1.1 自然语言与自然语言理解 .....       | (246)        |
| 8.1.2 自然语言理解的研究任务 .....       | (247)        |
| 8.1.3 自然语言理解的发展 .....         | (247)        |
| 8.1.4 自然语言理解的层次 .....         | (248)        |
| 8.2 语法规则的表示方法 .....           | (249)        |
| 8.2.1 句子结构的表示 .....           | (249)        |
| 8.2.2 上下文无关文法 .....           | (250)        |
| 8.2.3 变换文法 .....              | (251)        |
| 8.3 语法分析 .....                | (251)        |
| 8.3.1 自顶向下与自底向上分析 .....       | (251)        |
| 8.3.2 扩充转移网络分析 .....          | (253)        |
| 8.4 语义的分析 .....               | (256)        |
| 8.4.1 语义文法 .....              | (256)        |
| 8.4.2 格文法 .....               | (256)        |
| 8.5 自然语言的生成 .....             | (258)        |
| 8.6 自然语言理解系统的层次模型 .....       | (259)        |
| 习题 .....                      | (260)        |
| <b>第 9 章 专家系统</b> .....       | <b>(261)</b> |
| 9.1 专家系统的基本概念 .....           | (261)        |
| 9.1.1 什么是专家系统 .....           | (261)        |
| 9.1.2 专家系统的分类 .....           | (261)        |
| 9.1.3 专家系统的特点 .....           | (265)        |
| 9.2 专家系统的基本结构 .....           | (265)        |
| 9.2.1 知识库 .....               | (266)        |
| 9.2.2 数据库 .....               | (266)        |
| 9.2.3 推理机 .....               | (266)        |
| 9.2.4 解释机构 .....              | (267)        |

|               |                       |              |
|---------------|-----------------------|--------------|
| 9.2.5         | 知识获取机构 .....          | (267)        |
| 9.2.6         | 用户界面 .....            | (267)        |
| 9.3           | 知识获取 .....            | (268)        |
| 9.3.1         | 知识获取的任务 .....         | (268)        |
| 9.3.2         | 知识获取方法的分类 .....       | (269)        |
| 9.3.3         | 非自动知识获取 .....         | (270)        |
| 9.3.4         | 自动知识获取 .....          | (271)        |
| 9.4           | 专家系统的开发与评价 .....      | (271)        |
| 9.4.1         | 专家系统的开发条件 .....       | (272)        |
| 9.4.2         | 专家系统开发过程的生命期概念 .....  | (273)        |
| 9.4.3         | 专家系统开发过程的各个阶段 .....   | (273)        |
| 9.4.4         | 专家系统的评价 .....         | (275)        |
| 9.5           | 专家系统开发工具与环境 .....     | (276)        |
| 9.5.1         | 程序设计语言 .....          | (276)        |
| 9.5.2         | 知识工程语言 .....          | (277)        |
| 9.5.3         | 辅助型工具 .....           | (279)        |
| 9.5.4         | 支持工具 .....            | (279)        |
| 9.5.5         | 专家系统开发环境 .....        | (280)        |
| 9.6           | 专家系统的进一步发展 .....      | (281)        |
| 9.6.1         | 新一代专家系统的特征 .....      | (281)        |
| 9.6.2         | 分布式专家系统 .....         | (282)        |
| 9.6.3         | 协同式专家系统 .....         | (282)        |
|               | 习题 .....              | (283)        |
| <b>第 10 章</b> | <b>智能决策支持系统</b> ..... | <b>(284)</b> |
| 10.1          | 智能决策支持系统的基本概念 .....   | (284)        |
| 10.1.1        | 决策与决策过程 .....         | (284)        |
| 10.1.2        | 决策支持系统 .....          | (284)        |
| 10.1.3        | 智能决策支持系统 .....        | (285)        |
| 10.2          | 决策支持新技术 .....         | (285)        |
| 10.2.1        | 数据仓库 .....            | (286)        |
| 10.2.2        | 数据开采 .....            | (287)        |
| 10.2.3        | 数据仓库和数据开采的结合 .....    | (288)        |
| 10.3          | 智能决策支持系统的结构 .....     | (289)        |
| 10.3.1        | 智能决策支持系统的基本结构 .....   | (289)        |
| 10.3.2        | 智能决策支持系统的新结构体系 .....  | (289)        |
|               | 习题 .....              | (290)        |
|               | <b>参考文献</b> .....     | <b>(291)</b> |

# 第 1 章 人工智能概述

人工智能(Artificial Intelligence,简称 AI)是在计算机科学、控制论、信息论、神经生理学、心理学、哲学、语言学等多种学科相互渗透的基础上发展起来的一门新兴边缘学科。它主要研究如何用机器(计算机)来模仿和实现人类的智能行为。有人把人工智能同原子能技术、空间技术一起称为 20 世纪的三大尖端科技成就。

随着信息社会和知识经济时代的来临,信息和知识已成为人们的一个热门话题。然而,在这个话题的背后还蕴含着另外一个更深层的问题——智能。一般来说,信息是由数据所表达的客观事实,知识是信息经过智能性加工后的产物,智能是用来对信息和知识进行加工的加工器。在信息社会,人类面对的信息将非常庞大,仅靠人脑表现出来的自然智能是远远不够的,必须开发那种由机器实现的人工智能。就像在工业社会人类需要用机器去放大和延伸自己的体能一样,在信息社会人类又需要用机器去放大和延伸自己的智能,实现脑力劳动的自动化。

人工智能前景诱人,同时也任重而道远。本章作为概述,主要讨论人工智能的定义、形成过程、研究内容、研究方法、技术特点、应用领域及发展趋势等。目的在于展示一个处于不断发展中的人工智能的概貌,使读者对它能有一个初步了解。

## 1.1 人工智能及其研究目标

### 1.1.1 人工智能的定义

谈到人工智能的定义,首先需要指出以下两点:第一,人工智能和其他许多新兴学科一样,至今尚无一个统一的定义,所谓人工智能的定义,只能是人工智能学者根据对它的已有认识所作的一些不同解释;第二,人工智能的定义依赖于智能的定义。因此,要定义人工智能,首先应该定义智能。

#### 1. 什么是智能

通俗地说,智能是一种认识客观事物和运用知识解决问题的综合能力。至于其确切定义,还有待于对人脑奥秘的彻底揭示。事实上,为揭示这一奥秘,人类一直在进行着不懈的探索。但遗憾的是这一问题至今尚未完全解决,人类目前对人脑的认识还比较肤浅。在这种情况下,要从本质上确切定义智能还为时过早。尽管如此,人类通过对人脑奥秘的探索,毕竟对智能有了一个初步的认识。

#### (1) 认识智能的不同观点

人类在认识智能的过程中提出了许多不同的观点,其中最具有代表性的观点有以下三种。

##### ① 智能来源于思维活动

这种观点被称为思维理论,它强调思维的重要性,认为智能的核心是思维,人的一切智慧或智能都来自于大脑的思维活动,人的一切知识都是思维的产物,因而通过对思维规律与思维方法的研究可望揭示智能的本质。

## ② 智能取决于可运用的知识

这种观点被称为知识阈值理论,它把智能定义为:智能就是在巨大的搜索空间中迅速找到一个满意解的能力。知识阈值理论着重强调知识对智能的重要意义和作用,认为智能行为取决于知识的数量及其可运用的程度,一个系统所具有的可运用知识越多,其智能性就会越高。

## ③ 智能可由逐步进化来实现

这种观点被称为进化理论,它是由美国麻省理工学院(MIT)的布鲁克(R. A. Brooks)教授在对人造机器虫研究的基础上提出来的。他认为智能取决于感知和行为,取决于对外界复杂环境的适应,智能不需要知识,不需要表示,不需要推理,智能可以由逐步进化来实现。这一观点目前还没有形成完整的理论体系,反对者大有人在。但由于其与众不同且来源于实践,故引起了人们的关注。

由于上述三种观点对智能的认识角度不同,有些看起来好像是相互对立的,但如果把它们放到智能的层次结构中去考虑,又是统一的。

### (2) 智能的层次结构

人类的智能总体上可分为高、中、低三个层次,不同层次智能的活动由不同的神经系统来完成。其中,高层智能以大脑皮层为主,大脑皮层也称抑制中枢,主要完成记忆和思维等活动;中层智能以丘脑为主,丘脑也称感觉中枢,主要完成感知活动;低层智能以小脑、脊髓为主,主要完成动作反应。并且,智能的每个层次都还可以再进行细分。例如,对思维活动可按思维的功能分为记忆、联想、推理、学习、识别、理解等;或按思维的特性分为形象思维、抽象思维、灵感思维等。对感知活动可按感知功能分为视觉、听觉、嗅觉、触觉等。对行为活动可按行为的功能分为运动控制、生理调节、语言生成等。

可见,前述不同观点中的思维理论和知识阈值理论对应于高层智能,而进化理论则对应于中层智能和低层智能。

### (3) 智能所包含的能力

智能是一种综合能力。具体地说,它包含的各种能力如下。

#### ① 智能具有感知能力

感知能力是指人们通过感觉器官感知外部世界的能力。它是人类最基本的生理、心理现象,也是人类获取外界信息的基本途径。人类对感知到的外界信息,通常有两种不同的处理方式:一种是对简单或紧急情况,可不经大脑思索,直接由低层智能做出反应。例如,人们骑自行车遇到紧急情况时的急刹车动作;另一种是对复杂情况,一定要经过大脑的思维,然后才能作出反应,例如,人们平常所说的“三思而后行”。

#### ② 智能具有记忆与思维能力

记忆与思维是人脑最重要的功能,也是人类智能最主要的表现形式。记忆是对感知到的外界信息或由思维产生的内部知识的存储过程。思维是对所存储的信息或知识的本质属性、内部规律等的认识过程。人类基本的思维方式有抽象思维、形象思维和灵感思维。

抽象思维也称为逻辑思维,是一种基于抽象概念,根据逻辑规则对信息或知识进行处理的理性思维形式。例如,推理、证明、思考等活动。神经生理学认为,抽象思维是由左半脑实现的。

形象思维也称为直感思维,是一种基于形象概念,根据感性形象认识材料对客观现象进行处理的一种思维方式。例如,视觉信息加工、图像或景物识别等。神经生理学认为,形象思维是由右半脑实现的。

灵感思维也称为顿悟思维,是一种显意识与潜意识相互作用的思维方式。平常,人们在考

1-5  
虑问题时往往会因获得灵感而顿时开窍。这说明人脑在思维时除了那种能够感觉到的显意识在起作用外,还有一种潜意识也在起作用,只不过人们意识不到而已。灵感思维在创造性思维中起着十分重要的作用,它比形象思维更为复杂,对其产生机理和实现方法人们至今还不能确切描述。

在人类的思维机制中,形象思维和抽象思维通常是被结合起来使用的。即先用形象思维形成假设,然后再用抽象思维进行论证。至于形象思维到抽象思维的过渡,则是一个有待研究的问题。

### ③ 智能具有学习和自适应能力

学习是一个具有特定目的的知识获取过程。学习和自适应是人类的一种本能,一个人只有通过学习,才能增加知识、提高能力、适应环境。尽管不同人在学习方法、学习效果等方面有较大差异,但学习却是每个人都具有的一种基本能力。

### ④ 智能具有行为能力

行为能力是指人们对感知到的外界信息做出动作反应的能力。引起动作反应的信息可以由感知直接获得的外部信息,也可以是经思维加工后的内部信息。完成动作反应的过程,一般通过脊髓来控制,并由语言、表情、体姿等来实现。

## 2. 什么是人工智能

人工智能是一个含义很广的词语,在其发展过程中,具有不同学科背景的人工智能学者对它有着不同的理解,提出了一些不同的观点,如符号主义观点、连接主义观点和行为主义观点等。这些不同观点将在后面专门讨论,这里主要考虑人工智能的定义。

综合各种不同的人工智能观点,可以从“能力”和“学科”两个方面对人工智能进行定义。从能力的角度来看,人工智能是相对于人的自然智能而言的,所谓人工智能是指用人工的方法在机器(计算机)上实现的智能;从学科的角度来看,人工智能是作为一个学科名称来使用的,所谓人工智能是一门研究如何构造智能机器或智能系统,使它能模拟、延伸和扩展人类智能的学科。

如何衡量机器是否具有智能?早在1950年人工智能还没有作为一门学科正式出现之前,英国数学家图灵(A. M. Turing)就在他发表的一篇文章《Computing Machinery and Intelligence (计算机与智能)》中提出了“机器能思维”的观点,并设计了一个很著名的测试机器智能的实验,称为“图灵测试”或“图灵实验”。该测试的参加者由一位测试主持人和两个被测试者组成。要求两个被测试者中的一个是人,另一个是机器。测试规则是:让测试主持人和每个被测试者分别位于彼此不能看见的房间中,相互之间只能通过计算机终端进行会话。测试开始后,由测试主持人向被测试者提出各种具有智能性的问题,但不能询问测试者的物理特征。被测试者在回答问题时,都应尽量使测试者相信自己是“人”,而另一位是“机器”。在这个前提下,要求测试主持人区分被测试者哪个是人,哪个是机器。如果无论如何更换测试主持人和被测试者中的人,测试主持人能分辨出人和机器的概率都小于50%,则认为该机器具有了智能。作为人的一方不能判定对方是人还是机器,那么就认为对方的那台机器达到了人的智能。

对图灵的这个测试标准,也有人提出了疑义:认为该测试仅反映了结果的比较,既没有涉及思维的过程,也没有明确参加实验的人是小孩还是具有良好素质的成年人。尽管如此,它对人工智能这门学科的发展所产生的影响则是十分深远的。

### 1.1.2 人工智能的研究目标

关于人工智能的研究目标,目前还没有一个统一的说法。1978年,索罗门(A. Sloman)对人工智能给出了以下三个主要目标:

- (1) 对智能行为有效解释的理论分析;
- (2) 解释人类智能;
- (3) 构造智能的人工制品。

要实现索罗门的这些目标,需要同时开展对智能机理和智能构造技术的研究。即使对图灵所期望的那种智能机器,尽管它没有提到思维过程,但要真正实现这种智能机器,却同样离不开对智能机理的研究。因此,揭示人类智能的根本机理,用智能机器去模拟、延伸和扩展人类智能应该是人工智能研究的根本目标,或者叫远期目标。

人工智能的远期目标涉及到脑科学、认知科学、计算机科学、系统科学、控制论及微电子等多种学科,并有赖于这些学科的共同发展。但从目前这些学科的现状来看,实现人工智能的远期目标还需要有一个较长的时期。

在这种情况下,人工智能研究的近期目标是研究如何使现有的计算机更聪明,即使它能够运用知识去处理问题,能够模拟人类的智能行为,如推理、思考、分析、决策、预测、理解、规划、设计和学习等。为了实现这一目标,人们需要根据现有计算机的特点,研究实现智能的有关理论、方法和技术,建立相应的智能系统。

实际上,人工智能的远期目标与近期目标是相互依存的。远期目标为近期目标指明了方向,而近期目标则为远期目标奠定了理论和技术基础。同时,近期目标和远期目标之间并无严格界限,近期目标会随人工智能研究的发展而变化,并最终达到远期目标。

## 1.2 人工智能的产生与发展

人工智能这个术语自1956年正式提出,并作为一个新兴学科的名称被使用以来,已经有四十多年的历史了。回顾其产生与发展过程,可大致分为孕育、形成、知识应用、综合集成这四个阶段。

### 1.2.1 孕育期(1956年以前)

自远古以来,人类就有着用机器代替人们的脑力劳动的幻想。早在公元前900多年,我国就有歌舞机器人流传的记载。到公元前850年,古希腊也有了制造机器人帮助人们劳动的神话传说。此后,在世界上的许多国家和地区也都出现了类似的民间传说或神话故事。为追求和实现人类的这一美好愿望,很多代科学家为之付出了艰辛的劳动和不懈的努力。人工智能可以在顷刻间诞生,而孕育这个学科却需要经历一个相当漫长的历史过程。

(1) 古希腊伟大的哲学家和思想家亚里斯多德(Aristotle)创立了演绎法。他在其名著《工具论》中提出了形式逻辑中的一些基本规律,为形式逻辑奠定了基础。他提出的三段论至今仍然是演绎推理的最基本出发点。

(2) 英国哲学家和自然科学家培根(F. Bacon)创立了归纳法。其归纳法和亚里斯多德的演绎法一起,构成了思维的基本法则。此外,培根还提出了“知识就是力量”的名言。这一名言对研究人类的思维过程,对几百年后人工智能研究从一般思维探讨转向专门知识运用的这一重

大突破都起到了积极的促进作用。

(3) 德国数学家和哲学家莱布尼茨(G. W. Leibnitz)把形式逻辑符号化,奠定了数理逻辑的基础,从而使得人们可以对人的思维进行运算和推理。他认为可以建立一种通用的符号语言以及在此符号语言上进行推理的演算,提出了“万能符号”和“推理计算”的思想。这一思想不仅是数理逻辑的基础,同时也是现代机器思维设计思想的萌芽。

(4) 法国物理学家和数学家帕斯卡(B. Pascal)制造成功了世界上第一台加法器,对后来计算机领域的许多发明都产生了积极的影响。

(5) 英国数学家和发明家巴贝奇(C. Babbage)发明了差分机和分析机,为研制“思维机器”做出了巨大贡献。在他所发明的分析机中,已经包括了电子计算机的大部分特点,其设计思想与现代电子计算机十分相似。除此之外,巴贝奇还阐明了一台通用计算机系统应该包括的重要组成部分:输入(把数字输入机器)、存储器(保存数字和程序指令)、运算器(执行运算)、控制器(控制执行各种命令)和输出(把运算结果告诉用户)。只是由于当时的条件限制而未能实现,但这一思想为计算机的发展奠定了基础。

(6) 英国数学家布尔(G. Boole)创立了布尔代数。他在其著作《思维法则》中,首次用符号语言描述了思维活动中推理的基本法则,实现了莱布尼茨的理想。并在1854年发表的论文“An Investigation on the Law of Thoughts”(对思维规律的探讨)中,试图找出思维模拟的机械化规律,并明确提出符号逻辑代数是基于“机器是否放大智力”的探讨。可见,布尔所关注的是研制“智能机器”。

(7) 英国数学家、超时代的天才、图灵机的发明者图灵(A. M. Turing)1936年创立了自动机理论,并为人工智能做了大量的开拓性工作。自动机理论亦称图灵机,是图灵在他26岁那年提出的一个理论计算机模型。这一理论推进了思维机器的研究,并为电子计算机的诞生奠定了理论基础。

(8) 匈牙利数学家、博弈论的创立者冯·诺依曼(John. Von. Neumann)1945年提出了存储程序的概念,在计算机领域建立了不朽的功勋。冯·诺依曼的这一思想被誉为电子计算机时代的开始。至今,计算机的体系结构还基本上是冯·诺依曼型。

(9) 美国数学家、电子数字计算机的先驱莫克利(J. W. Mauchly)与他的研究生埃克特(J. P. Eckert)合作,1946年研制成功了世界上第一台通用电子计算机ENIAC(Electronic Numerical Integrator And Computer),为机器智能的研究与实现提供了物质基础。

(10) 美国著名数学家、控制论创始人维纳(N. Wiener)1948年创立了控制论。控制论是一门研究和模拟自控制的生物和人工系统的学科。这一学科的出现,标志着人们根据动物心理和行为学进行计算机模拟研究与分析的基础已经形成。控制论向人工智能的渗透,形成了现在的人工智能行为主义学派。

(11) 美国应用数学家、信息论的创始人香农(C. E. Shannon)创立了信息论。他认为人的心理活动可用信息的形式来进行研究,并提出了描述心理活动的数学模型。他1956年与麦卡锡(J. McCarthy)一起主编的《自动机研究》一书,又汇编了有关思维机器研究的多篇论文。

(12) 美国神经生理学家麦克洛奇(W. McCulloch)和皮兹(W. Pitts)一起于1943年建成了第一个神经网络模型(MP模型)。这一研究开创了用微观人工智能方法,从结构上模拟人脑的研究途径,并且为后来人工神经网络的研究奠定了理论基础。

通过上面的讨论可以看出,在人工智能诞生之前世界上的一些著名科学家就已经创立了数理逻辑、自动机理论、控制论和信息论,并发明了通用电子数字计算机。这些成就已经为人工



智能的产生准备了必要的思想、理论和物质技术条件。

## 1.2.2 形成期(1956年——1970年)

人工智能诞生于一次历史性的聚会。为使计算机变得更“聪明”，或者说使计算机具有智能，1956年夏季，当时在达特莫斯(Dartmouth)大学的年轻数学家、计算机专家麦卡锡，后为MIT教授；和他的三位朋友，哈佛大学数学家、神经学家明斯基(M. L. Minsky, 后为MIT教授)，IBM公司信息中心负责人洛切斯特(N. Lochester)，贝尔实验室信息部数学研究员香农共同发起，并邀请IBM公司的莫尔(T. More)和塞缪尔(A. L. Samuel)，MIT的塞尔夫里奇(O. Selfridge)和索罗蒙夫(R. Solomonoff)以及兰德(RAND)公司和卡内基(Carnegie)工科大学的纽厄尔(A. Newell)和西蒙(H. A. Simon)共10个人，在美国达特莫斯大学举行了一次为期两个月的夏季学术研讨会。这10位来自美国数学、神经学、心理学、信息科学和计算机科学方面的杰出年轻科学家，在一起共同学习和探讨了用机器模拟人类智能的有关问题，并由麦卡锡提议正式采用了“人工智能 AI(Artificial Intelligence)”这一术语。从而，一个以研究如何用机器来模拟人类智能的新兴学科——人工智能诞生了。

这次会议之后，在美国很快就形成了三个以人工智能为研究目标的研究小组。一个是纽厄尔和西蒙的卡内基—兰德小组(也称心理学小组)，另一个是塞缪尔的IBM公司工程课题研究小组，还有一个是明斯基和麦卡锡的MIT研究小组。人工智能在其诞生以后的10多年中，很快就在定理证明、问题求解、博弈等领域取得了重大突破。这一时期的主要研究大致有以下一些方面。

### (1) 心理学小组

1957年，纽厄尔、肖(J. Shaw)和西蒙等人的心理学小组研制了一个称为逻辑理论机(Logic Theory Machine, 简称LT)的数学定理证明程序。该程序模拟了人类用数理逻辑证明定理时的思维规律，纽厄尔用它证明了怀特海(A. N. Whitehead)和罗素(B. A. W. Russell)的历史名著《数学原理》中的38条定理，开创了用计算机研究人类思维活动规律的工作。

LT程序是针对具体领域的，为打破这种局限性，他们随后又开始研究一种不依赖具体领域的通用问题求解程序。通过心理学实验，他们总结出了人们在解决问题时思维过程的普遍规律，并把它归结为三个阶段：

- ① 先想出大致的解题计划；
- ② 根据记忆中的公理、定理和解题规划组织解题过程；
- ③ 在解题过程中不断进行方法和目标分析，修正解题计划。

基于这一规律，他们又于1960年研制了通用问题求解(General Problem Solving, 简称GPS)程序。该程序当时可以解决11种不同类型的问题，如不定积分、三角函数、代数方程、猴子摘香蕉、河内梵塔、人一羊过河等。

### (2) IBM工程课题研究小组

1956年，塞缪尔在IBM 704计算机上研制成功了具有自学习、自组织和自适应能力的西洋跳棋程序。这个程序可以从棋谱中学习，也可以在下棋过程中积累经验、提高棋艺。通过不断学习，该程序1959年击败了塞缪尔本人，1962年又击败了一个州的冠军。塞缪尔的下棋程序是用机器模拟人类学习过程的一次成功探索，其主要贡献在于发现了启发式搜索是表现智能行为的最基本机制。

### (3) MIT小组