



新一代计算机与人工智能系列新形态教材

人工智能 技术与应用

主 编 黄风华 黄家乾

 南京大学出版社

南京大学出版社
版权所有



新一代计算机与人工智能系列新形态教材

人工智能 技术与应用

主 编 黄风华 黄家乾

副主编 曾祥飞 胡神权

陆林花 林 韧

胡津阁 蔡子堃

课程资源



◎ 视频学习

◎ 拓展阅读

◎ 实践案例

 南京大学出版社

南京大学出版社
版权所有

图书在版编目(CIP)数据

人工智能技术与应用 / 黄风华, 黄家乾主编.

南京: 南京大学出版社, 2026. 8. -- ISBN 978-7-305-30281-7

I. TP18

中国国家版本馆 CIP 数据核字第 2026ZX6494 号

出版发行 南京大学出版社
社 址 南京市汉口路 22 号 邮 编 210093

书 名 人工智能技术与应用
RENGONG ZHINENG JISHU YU YINGYONG
主 编 黄风华 黄家乾
责任编辑 刘 飞

照 排 南京布克文化发展有限公司
印 刷 南京凯德印刷有限公司
开 本 787 毫米×1092 毫米 1/16 开 印张 22.75 字数 553 千
版 次 2026 年 8 月第 1 版 2026 年 8 月第 1 次印刷
I S B N 978-7-305-30281-7
定 价 59.00 元

网 址 <http://www.njupco.com>
官方微博 <http://weibo.com/njupco>
官方微信 NJUYUNSHU
销售咨询热线 025—83594756

* 版权所有, 侵权必究
* 凡购买南大版图书, 如有印装质量问题, 请与所购
图书销售部门联系调换

南京大学出版社
版权所有

前言

PREFACE

人工智能正在深刻地改变人们的生活方式。从手机上的语音助手到医院里的影像诊断系统,从短视频平台的智能推荐到自动驾驶汽车的路测车队, AI 技术已经从实验室走入了每个人的日常生活。2022 年以来, ChatGPT 和 DeepSeek 的发布更是引发了全球范围的 AI 热潮, 让“人人都在谈论 AI”成为现实。在这样的时代背景下, 人工智能素养正在成为当代大学生不可或缺的基本素养之一, 正如计算机基础和外语能力在过去几十年中逐渐成为大学教育的“标配”一样。

本书正是为回应这一时代需求而编写的。编写本书的初衷, 不是要把每一位读者培养成 AI 工程师或算法研究者, 而是希望帮助来自各个专业背景的大学生建立起对人工智能的系统认知——理解它的核心概念与基本原理, 了解它的能力边界与现实局限, 掌握与 AI 工具协作的基本方法, 并具备对 AI 相关社会议题进行独立思考和理性判断的能力。简言之, 编者希望读者在学完本书后, 能够“看懂 AI、用好 AI、思考 AI”。

全书共十三个模块, 按照“技术基石→经典方法→数据与感知→当代前沿→应用与伦理”的逻辑脉络展开。模块一至三奠定技术基石: 从人工智能的基本概念与发展历程出发, 依次讲解机器学习的核心思想和神经网络的工作原理, 帮助读者建立“人工智能→机器学习→深度学习”的知识框架。模块四至五拓展经典 AI 视野: 专家系统展示了“知识+推理”的符号主义路径, 智能机器人则将 AI 从“纯软件”引向了与物理世界交互的“软硬结合”领域。模块六至七引入数据与群体的视角: 数据挖掘揭示了从海量数据中发现知识的方法论, 群体智能则展现了“简单个体涌现复杂智能”的迷人现象。模块八至九聚焦感知与大数据: 计算机视觉让机器学会“看”, 模糊逻辑与大数据技术则为 AI 提供了处理不确定性和海量信息的工具。模块十至十一代表当代最前沿: 大语言模型和智能体正在重新定义人与 AI 的交互方式。模块十二展示 AI 在视听内容创作领域的革命性应用。模块十三以 AI 伦理、法律法规和人机关系收束全书, 引导读者从技术认知走向价值思考。

在编写理念上, 本书坚持“问题驱动、案例穿插、概念着陆、思辨收束”的教学设计原则。每个模块以若干核心问题串联知识脉络, 每个新概念都配有贴近生活的类比或具体案例来帮助读者“着陆”理解, 每节末尾的思考题从识记、分析到应用逐层递进, 引导读者从“记住知识”走向“运用知识”和“批判知识”。全书不出现数学公式和程序代码, 所有技

术原理均以文字、图表和直觉化描述呈现,始终以通识课的标准来控制技术深度。

案例选取方面,本书兼顾国际经典与中国实践,覆盖互联网、医疗、金融、制造、教育、交通、农业等多个行业领域,力求让不同专业背景的读者都能看到 AI 与自身领域的关联。在时效性管理上,本书有意识地将正文主体建立在“稳定层”内容之上(基本概念、经典算法、已成定论的历史事件),而将易变的产品名称、最新数据和前沿动态集中安排在案例框和思考题中,以延长教材的有效期,降低再版维护成本。

每个模块采用统一的结构设计:以“模块导读”引出核心问题,以“学习目标”明确认知层次,以“思维导图”呈现知识全景,然后通过若干“任务”(即各节)展开具体内容。每节内设置引导思考框、案例框、知识拓展框、对比表、结构图、本节小结和思考题等标准化教学元素。模块末尾的“思想引领”板块超越技术层面,引导读者思考 AI 背后的哲学、伦理和社会问题;“任务实践”板块则提供可在课堂上开展的动手活动和小组研讨方案,将纸面知识转化为切身体验。

编者深知, AI 领域的发展速度之快,常常让教材“未写完即过时”。本书力图在“追踪前沿”和“经得起时间检验”之间取得平衡:前沿内容让读者感受到 AI 技术的脉搏和温度,经典内容则提供不因技术迭代而失效的认知框架和思维方法。编者希望读者从本书中获得的不只是对当下 AI 产品和技术的了解,更是一种能够持续自主学习、理性判断的“AI 素养”——无论未来的技术如何演变,这种素养都将是宝贵的。

本书可作为普通高等院校计算机通识课程的教材,也可供对 AI 感兴趣的各类读者自学参考。建议授课学时为 32—48 学时(2—3 学分)。教师可根据专业特点和学时安排灵活选取模块:模块一至三为核心基础,建议全部讲授;模块四至十二可根据教学重点选取;模块十三建议作为课程收束。各模块在保持逻辑衔接的同时也具备基本的独立可读性,教师可根据需要调整讲授顺序。

本书由校企联合构思,编写期间得到了众多同仁的支持与帮助,包括来自阳光学院相关课程的各种任课老师,以及来自知鱼智联科技股份有限公司的吴聿建工程师、杭州半云科技有限公司的叶舜宇工程师。在此,编者向所有参与内容设计、审稿、提出宝贵意见的专家和教师表示衷心感谢,向为本书提供案例素材和教学反馈的一线教师和企业工程师致以诚挚谢意。同时,编者也要感谢在编写过程中提供技术支持的出版团队。

限于编者水平,书中难免存在疏漏和不足之处,恳请广大读者和同仁批评指正。编者也将持续关注 AI 领域的最新发展,在后续修订中不断完善本书的内容。

今天的大学生,是与 AI 共同成长的一代人。他们既是 AI 技术的使用者和受益者,也将是 AI 发展方向的参与者和决策者。希望本书能为读者走进 AI 世界提供一扇清晰的窗口,也为读者思考 AI 与人类的未来关系提供有价值的参照。

编者

2026 年 5 月

目录

CONTENTS

模块一 人工智能概论

模块导读 学习目标	003
思维导图	004
任务一 AI 的基本知识	005
任务二 AI 的发展历程	012
任务三 AI 发展的三大引擎	022
任务四 AI 的应用现状	027
思想引领	033
任务实践	034
实践一 “我身边的 AI”调研报告	034
实践二 AI 发展史的关键时刻	034
实践三 “AI 能做什么,不能做什么”辩论赛	035
实践四 绘制一张行业 AI 应用地图	035

模块二 机器学习

模块导读 学习目标	039
思维导图	040
任务一 机器学习概述	041
任务二 基于学习方式的分类	047
任务三 机器学习的基本结构	053
任务四 机器学习算法	059
任务五 机器学习的应用	066

思想引领	072
任务实践	073
实践一 “拆解一个 ML 应用”分析报告	073
实践二 “监督学习 vs 无监督学习”体验实验	074
实践三 “算法选型”情境挑战	074
实践四 “推荐算法的信息茧房”调研与辩论	075

模块三 神经网络

模块导读 学习目标	079
思维导图	080
任务一 神经网络概述	081
任务二 神经信息处理的基本原理	083
任务三 人工神经网络	087
任务四 前馈神经网络	092
任务五 Hopfield 神经网络	096
任务六 随机神经网络	101
思想引领	105
任务实践	106
实践一 “人体神经网络”课堂模拟实验	106
实践二 “联想记忆”纸笔实验	107
实践三 “神经网络家族”知识图谱	107
实践四 “AI 中的跨学科”小组研讨	108

模块四 专家系统

模块导读 学习目标	111
思维导图	112
任务一 专家系统概述	113
任务二 专家系统基本结构	118
任务三 专家系统工具	122
任务四 专家系统建立	128
任务五 新型专家系统	132

思想引领	136
任务实践	137
实践一 “迷你专家系统”规则设计	137
实践二 技术选型辩论	138
实践三 “知识图谱”小实验	138

模块五 智能机器人

模块导读 学习目标	141
思维导图	142
任务一 机器感知.....	143
任务二 智能机器人概述	148
任务三 智能机器人的体系结构	152
任务四 机器人视觉系统	157
思想引领	162
任务实践	163
实践一 “拆解一台机器人”分析报告	163
实践二 “机器人感知挑战”场景实验	163
实践三 “自动驾驶的伦理困境”角色扮演研讨	164

模块六 数据挖掘

模块导读 学习目标	167
思维导图	168
任务一 从数据到知识	169
任务二 数据挖掘方法	173
任务三 数据挖掘的经典算法	178
任务四 机器学习与数据挖掘	181
思想引领	186
任务实践	187
实践一 “发现隐藏的关联”纸笔实验	187
实践二 “CRISP-DM 全流程”方案设计	187
实践三 “相关不等于因果”辩论与反思	188

模块七 群体智能

模块导读 学习目标	191
思维导图	192
任务一 什么是群体智能	193
任务二 典型算法模型	197
任务三 群体智能背后的故事	202
任务四 群体智能的应用与发展	205
思想引领	208
任务实践	209
实践一 “模拟蚁群觅食”桌面实验	209
实践二 “群体智能解决校园问题”方案设计	209
实践三 “去中心化组织”案例研讨	210

模块八 计算机视觉

模块导读 学习目标	213
思维导图	214
任务一 模式识别.....	215
任务二 图像识别.....	218
任务三 计算机视觉技术	221
任务四 智能图像处理技术	224
任务五 计算机视觉技术的应用	228
思想引领	232
任务实践	233
实践一 “拆解手机相机 AI”体验报告	233
实践二 “CV 应用行业调研”报告	233
实践三 图像鉴别挑战与伦理讨论	234

模块九 大数据与人工智能

模块导读 学习目标	237
思维导图	238

任务一 模糊逻辑的概念	239
任务二 模糊逻辑系统	243
任务三 大数据思维与变革	246
任务四 大数据与人工智能	250
思想引领	253
任务实践	255
实践一 “模糊控制器”纸笔设计	255
实践二 “我的数字画像”大数据体验与反思	255
实践三 “精确 vs 模糊”辩论	256

模块十 大模型

模块导读 学习目标	259
思维导图	260
任务一 大模型概述	261
任务二 自然语言	267
任务三 AI 大语言	272
任务四 大模型的局限和挑战	276
思想引领	280
任务实践	281
实践一 “大模型的能力边界”测试实验	282
实践二 “提示工程”技巧体验	282
实践三 “大模型与学术诚信”圆桌讨论	283

模块十一 智能体

模块导读 学习目标	287
思维导图	288
任务一 智能体概述	289
任务二 基于大模型的智能体	295
任务三 主流智能体产品与应用	300
任务四 局限和挑战	303
思想引领	307

任务实践	308
实践一 “设计一个智能体”方案展示	308
实践二 “智能体可靠性”红队测试	309
实践三 “AI 的自主性边界”哲学讨论	309

模块十二 AI 的视听应用

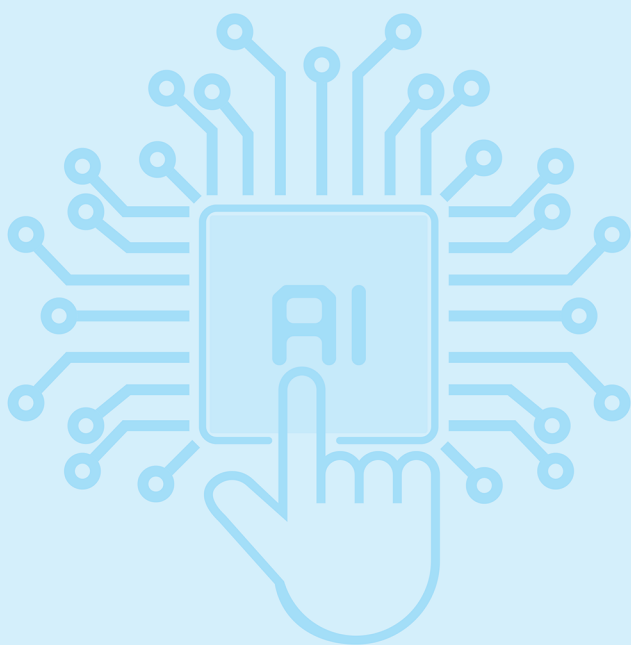
模块导读 学习目标	313
思维导图	314
任务一 AI 与视听语言	315
任务二 AI 视频生成软件	319
任务三 AI 语言写作软件	324
任务四 AI 视听应用案例	328
思想引领	332
任务实践	333
实践一 “AI 辅助短视频”创作实战	333
实践二 “同题创作对比”实验	333
实践三 “AIGC 伦理”辩论与公约制定	334

模块十三 AI 的挑战与未来

模块导读 学习目标	337
思维导图	338
任务一 AI 面临的挑战	338
任务二 AI 对未来的新展望	342
任务三 构建友好人机关系	345
思想引领	348
任务实践	348
实践一 “AI 影响力评估”综合报告	348
实践二 “给 10 年后的自己写一封信”	349
实践三 《AI 公民宣言》	349
参考文献	351

» 模块一

人工智能概论



南京大学出版社
版权所有

南京大学出版社
版权所有

◆ 模块导读

从清晨被手机闹钟唤醒,到刷短视频时收到精准的个性化推荐;从对智能汽车、智能家居发出语音指令,到享受各类便捷的智能化生活服务……AI 已深度融入日常生活的方方面面,而多数人对此并未清晰察觉。AI 究竟是什么?它与科幻电影中能够自主思考的全能机器人是否等同?

关于“如何让机器具备智能”这一核心问题,科学界形成了符号主义、连接主义、行为主义三大理论流派。在工程实践中,AI 的实现方式也在持续演进:从早期依靠人工编写规则的专家系统,到让计算机从数据中自动学习的统计方法,再到如今能够实现对话、写作、编程等复杂任务的大语言模型,每一次技术突破都显著拓展了 AI 的能力边界。然而,当前的 AI 是否已经“无所不能”?它距离科幻作品中设想的通用 AI 还有多远?

本模块将围绕上述核心问题展开论述,带领读者系统认识 AI。

◆ 学习目标

【了解】

- AI 的基本概念与定义,理解其与传统计算机程序的本质区别;
- AI 发展过程中的关键里程碑事件与代表性人物;
- AI 三大学派(符号主义、连接主义、行为主义)的基本观点。

【理解】

- AI 从萌芽到深度学习时代各阶段的特征与演进逻辑;
- 数据、算力、算法三大引擎对 AI 发展的驱动机制及其协同关系;
- 弱 AI、强 AI 与超 AI 的分类标准与内涵差异。

【掌握】

- 运用 AI 发展的时间线框架,分析关键技术突破的因果关系;
- 结合产业链视角,识别 AI 在不同行业中的典型应用场景。

【能够】

- 综合运用本模块知识,对身边的 AI 产品与服务进行识别和初步分析;
- 针对 AI 的发展现状与趋势,形成自己的判断与思考。

◆ 思维导图

下图展示了本模块的知识结构与内容框架,帮助读者从整体上把握各任务之间的逻辑关系:

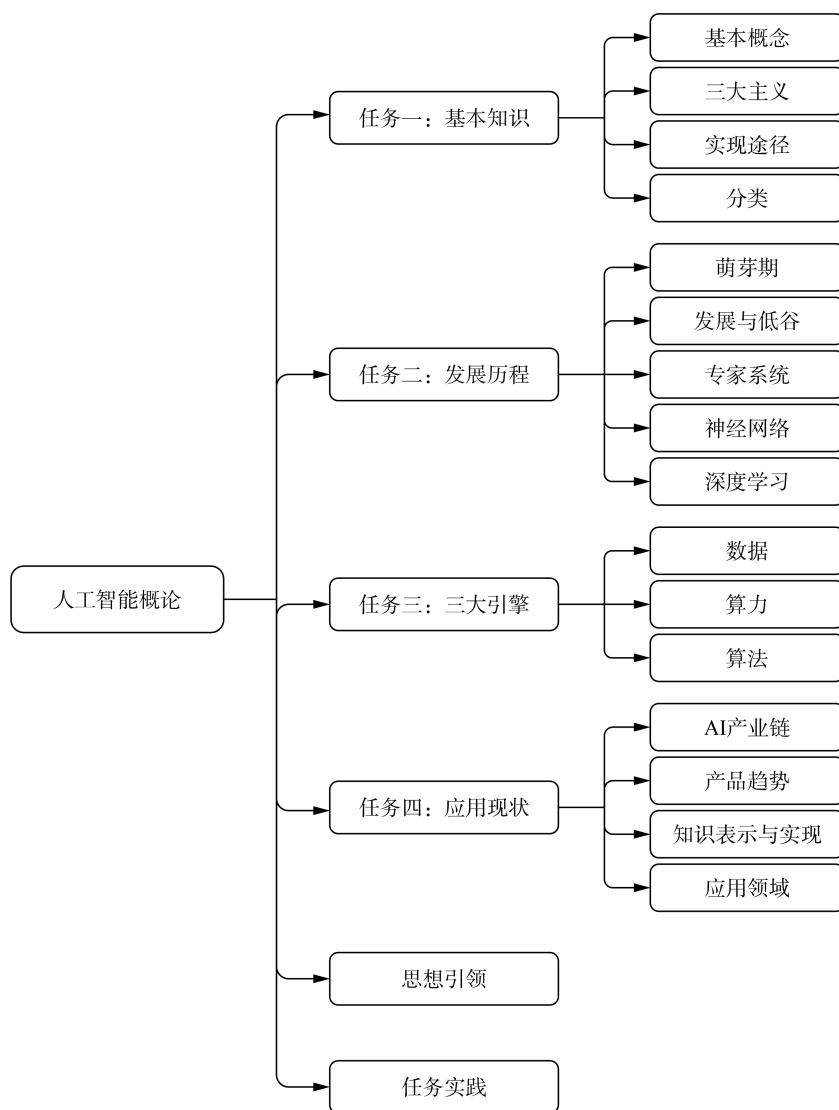


图 1-1 模块一知识结构思维导图

任务一 AI 的基本知识

◆ 引导思考

对着手机说“帮我定一个明天早上8点的闹钟”，手机便自动完成了设置；打开短视频平台，系统似乎总能猜到用户想看什么——这些日常场景的背后，都是AI在发挥作用。

那么，AI到底是什么？它与在科幻电影中看到的全能机器人是一回事吗？科学家们对于“如何让机器变得聪明”这件事，有哪些截然不同的思路？AI发展到今天，真的已经无所不能了吗？它的能力边界在哪里？

一、AI 的基本概念

1. 什么是 AI

人工智能(Artificial Intelligence, AI)是计算机科学中一个极富魅力和挑战性的研究领域。尽管这一概念自1956年正式提出以来已有近七十年的历史，但学术界至今仍未形成一个被所有人接受的统一定义。

从不同的理论视角出发，AI的定义大致可以归纳为四种取向。第一种试图构建一个能够模拟人类认知过程的AI系统。第二种关注的是AI系统在外部表现上是否能够达到与人类相当的水平。第三种追求合理地思考，让AI系统按照逻辑推理的规则正确地进行推理和判断。第四种则着眼于AI系统能否在给定环境中作出最优的决策和行为。

表 1-1 AI 的四种定义取向

定义取向	核心关注	典型代表
像人一样思考	模拟人类认知过程	认知建模、脑科学启发的 AI
像人一样行动	外部行为达到人类水平	图灵测试、聊天机器人
合理地思考	遵循逻辑规则正确推理	专家系统、定理证明
合理地行动	在环境中作出最优决策	自动驾驶、智能推荐

2. 图灵测试

在上面介绍的四种定义取向中，“图灵测试”是最广为人知的一个。它是由英国数学家艾伦·图灵(Alan Turing)在1950年发表的论文《计算机器与智能》中提出的。图灵没有直接回答“机器能否思考”这个哲学问题，而将其转化为一个可操作的实验：让一位人类评委通过文字对话的方式分别与一个人和一台机器交流，如果评委无法区分哪个是机器、

哪个是人,那么就有理由认为这台机器具备了某种意义上的智能。

图灵测试的伟大之处在于,它首次为一个看似无法回答的哲学问题提供了一个具体的、可操作的判断标准。但有的批评者也指出,图灵测试只关注了语言交流这一种智能表现形式,而忽略了感知、运动、情感等其他重要的智能维度;此外,系统可能并不真正理解对话的内容。

知识拓展:“中文房间”思想实验

哲学家约翰·塞尔(John Searle)在1980年提出了这样一个思想实验:假设有一个完全不懂中文的人被关在一个房间里,房间里有一本详细的规则手册。外面的人通过门缝递进用中文写的问题,房间里的人按照手册查找对应的回答、写好后递出去。外面的人看到的回答完全正确、流利,会以为房间里有一个懂中文的人。但事实上,房间里的人对中文一无所知,他只是在执行规则。

这个实验的核心问题是:如果一个系统只是按照规则处理符号,而并不理解这些符号的含义,它真的具有智能吗?这一问题至今仍然是AI哲学讨论中的核心议题之一。

围绕图灵测试的争论至今没有定论,但这些讨论本身对理解AI有着重要的启发价值。目前的AI技术,无论多么先进,都只能够使AI的输出结果足以以假乱真,而AI还不能够理解自己在做什么。

二、AI的三大学派

AI领域有一个非常根本的问题:机器的智能应该从哪里来?对于这个问题,学者们形成了三种截然不同的回答,被称为AI的“三大学派”:符号主义、连接主义、行为主义。

1. 符号主义

符号主义(Symbolism),又称逻辑主义,是AI历史上最早占据主导地位的学派。它的核心观点可以用一个简单的等式来概括:智能=知识+推理。

符号主义者认为,人类之所以聪明,是因为人掌握了大量的知识,并且能够运用逻辑推理将这些知识组合起来解决问题。因此,要让机器变得聪明,只需要做两件事:第一,把人类的知识用计算机能够理解的符号和规则表达出来;第二,让计算机按照逻辑规则操作这些符号,从已有的知识中推导出新的结论。

符号主义最辉煌的成果之一是20世纪70年代,斯坦福大学开发的MYCIN系统。MYCIN系统内置了大约600条形如“如果患者有发热症状,并且血液培养发现革兰氏阴性菌,那么感染的可能原因是……”的医学规则,能够辅助医生诊断出血液感染疾病。在当时的测试中,MYCIN给出的诊断建议在某些指标上甚至优于部分非专业的医生。

然而,符号主义首先需要相关领域的专家将自己的专业知识一条一条地翻译成计算机能处理的语言,这个过程耗时且困难。并且符号主义要求知识清晰明了,便于推理,但现实世界的信息往往是模糊、残缺的,很难用非黑即白的逻辑规则来表达。

2. 连接主义

连接主义(Connectionism),又称仿生学派,它的目标是让 AI 模仿人类大脑的工作方式。

人类大脑包含大约 860 亿个神经元,每个神经元都是一个相对简单的细胞,但它们通过约 100 万亿个突触相互连接,形成了一个极其复杂的网络。人类的思考、记忆、感知、决策等所有智能活动,其实都是涌现于这些神经元之间电信号和化学信号的传递过程中。

连接主义者受到这一事实的启发,提出了一个大胆的想法:既然智能涌现于大量简单单元的连接与协作,那么能不能在计算机中也构建一个类似的网络呢?这就是人工神经网络的基本思路的由来。

目前,随着大数据的积累和 GPU 算力的飙升,以人工神经网络为基础的深度学习方法迎来了全面爆发阶段,在图像识别、语音识别、自然语言处理等领域取得了划时代的突破。日常使用的 AI 工具,底层都是人工神经网络,通过深度学习方法构建的。

案例: AlphaGo——连接主义的里程碑式胜利

2016 年 3 月,谷歌 DeepMind 团队开发的围棋程序 AlphaGo 以 4:1 击败世界围棋冠军李世石,震惊全球。围棋的可能局面数量极高(约 10^{170}),长期以来被认为是 AI 最难攻克的棋类游戏之一。

AlphaGo 使用了两个人工神经网络:“策略网络”负责预测下一步棋最可能下在哪里,“价值网络”负责评估当前局面的胜率。这两个网络通过数百万盘棋局的训练,学会了连人类棋手都难以言说的“棋感”。

AlphaGo 成功是连接主义思想的一次有力证明:智能不一定需要明确的规则和逻辑推理,也可以从大量数据和经验中自发涌现出来。

3. 行为主义

行为主义(Behaviorism),又称进化主义或控制论学派。与符号主义和连接主义不同,行为主义把注意力放在机器与外部环境的互动上。

行为主义者认为,智能不需要人类设计,而是在与外部环境的反复互动中学习出来的。行为主义的 AI 会在环境中不断尝试,做对了就给奖励,做错了就给惩罚,智能行为最终会自然出现。

行为主义的思想在当代 AI 中也有重要的传承。近年来广受关注的强化学习技术,其核心思想就是让 AI 在环境中通过试错来优化自己的行为策略。强化学习已经在游戏对战、机器人控制、资源调度等领域取得了令人瞩目的成果。

4. AI 三大主义的融合

在 AI 发展的早期阶段,三大主义之间存在激烈的学术竞争,甚至互相排斥。但随着研究的深入,人们越来越认识到,单一学派的方法都有其局限性,真正强大的 AI 需要融合多种思想。

日常生活中常用的大语言模型就是三大主义融合的例子。它的底层架构是基于连接主义的深度神经网络,训练过程中使用了行为主义的强化学习技术,而在应用层面又可以通过“提示词”来引导其进行类似符号推理的操作。

表 1-2 AI 三大主义对比

学派	核心观点	代表技术	局限性
符号主义	智能=知识+推理	专家系统、定理证明	知识获取困难,难应对模糊信息
连接主义	涌现于神经元连接	神经网络、深度学习	需要大量数据,“黑箱”难解释
行为主义	在与环境交互中涌现	强化学习、自主机器人	学习效率低,安全性难保证

三、AI 的实现方法

当前,AI 的工程实现主要有四种途径。这四种途径之间大致呈现出历史演进的关系。

1. 基于规则的方法

这是 AI 最早期、最贴合大众直观认知的实现思路。原理并不复杂:由行业资深专家梳理总结实操经验与专业知识,整理成一条条明确准则,再交由计算机程序依照准则自动运行。这类规则大多采用“如果……,就……”的逻辑形式编写。

在现代,基于规则的技术手段,依旧能在范围固定、界限明确的场景中发挥作用。例如银行反洗钱风控体系,当单笔交易金额达到五十万元,系统便自动推送至人工复核,系统依靠既定规则完成基础判定。但一旦进入场景灵活、情况多变的复杂环境中,例如体会语句深层含义、识别图片内各类物体,这套方法就很难发挥作用。

2. 基于统计学习的方法

基于规则模式暴露的局限,推动 AI 行业研究思路发生了转变:让计算机直接从海量真实数据里自主摸索内在规律,这便是统计学习的核心思路。

统计学习无需人为设定标准,只需收集大量真实数据,日常应用中常见的统计学习算法有决策树、支持向量机、朴素贝叶斯、随机森林等。这类学习方式如今早已融入生活,广泛用于垃圾邮件拦截、个人信用评估、疾病风险预判、电商商品推荐等诸多场景。邮箱自动筛选屏蔽垃圾邮件的功能,背后大多就是统计学习模型在发挥作用。

3. 基于深度学习的方法

深度学习是统计学习的升级版。传统统计学习虽具备从数据中摸索规律的能力,但在处理图片像素、音频波形这类原始数据时,都需要研发人员提前手动设计数据特征来完成信息梳理。这套人工设计特征的流程十分考验专业经验水平,也长久制约着整套技术体系的发展效率。

深度学习最核心的突破,便是可以直接对接原始数据,自主学习适配各类任务的数据

特征,并直接输出结果。这种省去中间繁琐步骤、直接完成输入到输出的模式,称为端到端学习。

案例: ImageNet 竞赛——深度学习的成名之战

ImageNet 拥有超 1 400 万张带标注的海量图像数据,业内常年在此基础上举办图像识别赛事。在 2012 年之前,业内顶尖算法的识别错误率始终居高不下,一直维持在 25% 以上。

2012 年,多伦多大学辛顿团队携 AlexNet 深度卷积神经网络参赛,直接将识别错误率压低至 15.3%,和第二名拉开整整 10 个百分点的巨大差距。

以往行业算法每年的性能提升幅度仅有 1 至 2 个百分点,如此大幅度的突破瞬间震撼了整个学术领域,也正式宣告深度学习时代全面到来。在此之后,深度学习模型的识别精度一路攀升,到 2020 年,模型识别错误率已经远远低于人类肉眼识别约 5% 的误差水平。

4. “一个模型,万种用途”的大模型

早期 AI 模型大多偏向专项能力,各司其职,擅长图像识别的无法完成语言翻译,专攻翻译的也不能编写程序,用途十分单一。而 2020 年前后兴起的大模型,逐步具备了全能通用的发展特质。

大模型遵循“先广泛学习,再针对性深耕”的研发逻辑。第一步开展大规模预训练,依托海量数据夯实基础,积累充足的通用知识。以大语言模型为例,大语言模型的预训练阶段会学习网络中海量文本内容,涵盖书籍文献、资讯内容、各类文章以及程序代码等,在学习语言规律的过程中,建立起对人类语言完整且透彻的理解。完成基础学习后,再借助微调或者指令提示的方式,把通用能力适配到各类具体工作场景中。

如今各类主流大语言模型走进大众视野,普通人只需用日常语言下达需求,就能让它协助撰写文稿、完成翻译、整理分析数据、讲解专业知识,或是编写程序代码。这种简单易上手的使用方式,打破了 AI 的技术壁垒,渐渐成为和搜索引擎一样,融入人类日常使用的基础通用工具。

四、AI 的分类

在新闻和社交媒体上经常有这样的话题:AI 到底发展到了什么程度? 它离像人一样聪明还有多远? 要回答这些问题,就需要了解 AI 的分类体系。

1. 按智能水平分类

依据智能层级由低到高划分,AI 主要分为三类:弱 AI、强 AI 以及超 AI,这套划分标准,也是认清 AI 能力界限最核心的参考依据。

弱 AI 也叫专用 AI,核心特点是只擅长完成单一固定任务。这类智能系统能在专属领域做到极致,部分表现甚至超越人类,但一旦跳出自身擅长的范围,就无法开展任何相

关工作。

生活中的弱 AI

拿起手边的手机稍加留意,就能发现随处都有弱 AI 在发挥作用:

- 依托用户的浏览习惯与互动偏好推送内容的短视频推荐算法;
- 借助人脸识别自动完成人像分组整理的相册分类功能;
- 依靠语言规律预判输入内容的输入法智能联想;
- 整合实时路况规划最优出行路径的地图导航;
- 结合海量历史数据与实时场景,预估送餐时长的外卖配送时效测算功能。

这些应用各自出色,但推荐算法不会帮用户导航,人脸识别也无法预测配送时间。这就是弱 AI 只精通自身对应的功能,各项智能能力相互独立,不能通用互换的典型特征。

强 AI 也叫通用 AI,目标是打造出拥有和人类同级别的综合智能体系。这类智能不再局限于单一事务,能够像人一样自主学习、逻辑推理、统筹规划、创意创作、与人沟通交流,还能感知情绪变化,灵活适配各类不同场景,自由切换完成多种任务。

如果一款 AI 模型既能开车,还会日常烹饪,还能与人闲谈谈心、读懂诗词背后的深意,又能快速适应陌生全新环境,并且完成这些事无需单独搭建不同模型,这才算真正达到强 AI 的标准。就目前而言,市面上还没有任何智能产品能够实现这一水准。

像 DeepSeek 这类全能对话大模型,还不能够代表强 AI 已经到来。它本质上仍然属于一种文字处理工具,能力依旧局限在语言交互范畴。在没有外接视觉、语音配套功能的前提下,它无法直观视物、聆听外界声音,也不能在现实世界开展实际行动,不存在独立思想与主观想法,更谈不上真正理解事物内涵。这类模型只能算作性能顶尖的专用 AI,距离真正的通用智能相差甚远。

对于强 AI 的落地时间,行业内始终没有统一答案,各方看法差异明显。持乐观态度的从业者认为,未来十到二十年便有望实现;多数深耕 AI 领域的科研人员态度更为谨慎,认为至少还需要五十年甚至更久;还有一部分学者提出疑问,依靠当下现有的技术路径,或许根本无法研发出真正的通用 AI。这一问题,也成为 AI 领域最具探讨价值的未解难题。

至于超 AI,则是一个更为遥远和思辨性的概念,指的是在几乎所有认知领域都远远超越人类最优秀个体的智能系统。超 AI 目前仅存在于理论探讨和科幻想象中,但围绕它引发的关于 AI 安全性、可控性、伦理规范的讨论,对于 AI 技术的长期健康发展具有重要的预警意义。

表 1-3 AI 按智能水平的分类

类别	核心特征	发展现状	典型举例
弱 AI	专注单一任务,仅限固定领域使用	技术成熟,已全面普及	语音助手、人脸识别、AlphaGo
强 AI	具备综合能力,可灵活适配各类场景	暂未研发落地	—
超 AI	几乎所有认知领域都远远超越人类	纯理论阶段	—

2. 按应用领域分类

除了按照智能高低划分, AI 还能依据实际应用场景细分出多个主流分支领域。

● 计算机视觉

让机器看懂图片与视频。生活和行业中应用十分普遍, 常见的有手机人脸解锁、公共场所安防人脸识别, 医学影像筛查、工业产品质量检测, 以及自动驾驶车辆实时感知周边路况等。

● 自然语言处理

让机器读懂人类语言, 同时也能自主生成通顺语句。日常使用的在线翻译工具、线上智能客服、文章内容精简提炼、用户评论情绪判断, 还有各类大模型对话、文案创作等功能, 都依托这项技术实现。

● 语音处理

实现人与机器的语音交互, 把人声转成文字的语音识别, 以及将文字转为真人音色的语音合成; 除此之外还能通过声纹特征识别谁在说话。

● 机器人技术

让实体机械设备拥有环境感知能力, 并且能在现实世界自主完成各类动作。涵盖汽车制造、芯片生产所用的工业机器人, 餐厅送餐、医院引导的服务机器人, 以及航拍、物流运输无人机等。

● 智能决策

依靠数据运算, 让机器在复杂场景中辅助甚至独立作出合理判断。多用于金融信贷审核、交易风险排查, 城市交通调度、航班行程安排, 还有医疗辅助诊断、用药方案建议等场景。

在实际的 AI 应用中大多会融合多种技术。例如自动驾驶需要视觉技术识别路况、语音功能接收指令、智能决策规划行驶方案, 再通过机器人相关技术完成车辆操控; 智能客服则整合了语言理解、语音交互与智能应答决策等多项能力。

案例：智能手机里的多元 AI 技术融合

日常随身携带的智能手机, 正是各类 AI 技术融合运用的典型缩影。

唤醒语音助手发送消息需要先通过语音识别收录话语内容, 再依靠自然语言处理读懂用户需求, 最后由智能决策模块匹配对应的操作指令。

拍照 APP 的美颜优化功能需要依托计算机视觉完成场景分辨与人脸捕捉, 再借助智能决策自动调配合适的滤镜与修图参数。

使用实景导航时, 设备需要依靠视觉技术识别现实道路环境, 再通过智能决策将路线信息精准叠加在实景画面上。

3. 正确看待 AI 的能力边界

学完本部分内容, 读者需要建立一个核心认知: 当前生活中能够接触到的所有 AI 技术, 都还处于弱 AI 阶段, 它们都不具备真正的理解能力, 也没有常识推理能力, 更没有自我意识。即使它们有时候输出的内容, 看起来仿佛像拥有这些能力。所以, 不是所有问题

都适合用 AI 工具解决, AI 工具生成的内容也不能不经过谨慎验证就直接使用。

如今的大模型热潮声势浩大,但是每一轮技术热潮都会伴随着过度期待和炒作,而真正能长久的进步,从来都源于对技术的理性认知和扎实应用。对于未来的技术人才来说,对 AI 技术保持不盲从、不恐惧的理性态度,培养对 AI 技术的批判性思维,比掌握某一种具体的 AI 工具更加重要。

本节小结

1. AI 旨在让机器胜任以往只有人类才能完成的智能工作,从思维与行为、人类标准与理性标准两个角度,可梳理出四类主流定义方向。

2. AI 三大主流学派,划定了不同智能实现路径:符号主义依托知识与逻辑推理,连接主义仿生神经网络结构,行为主义侧重在环境互动中习得能力,如今三类技术理念正逐步融合发展。

3. AI 落地应用历经四大发展阶段,依次为规则设定、统计学习、深度学习与大模型,它们顺着时代顺序迭代升级,现阶段各类方式依旧相辅相成、共同使用。

4. 依照智能能力高低, AI 划分为三类:已经成熟落地的弱 AI、暂未实现的通用强 AI,以及仅存在理论设想的超 AI,目前市面上所有 AI 产品,全都属于弱 AI 范畴。

5. AI 主流应用涵盖视觉识别、自然语言交互、语音交互、智能机器人、智能决策等方向,现实场景中大多都需要多项技术协同完成一项任务。

思考题

1. [识记] 请简述 AI 的四种定义取向,并用自己的话解释“合理地行动”这一取向为什么被认为最具实用价值。

2. [理解] 有人说“ChatGPT 已经是强 AI 了”,请结合本任务所学的弱 AI/强 AI 分类标准,分析这一说法是否准确,并说明理由。

3. [应用] 请在日常生活中找出 3 个 AI 应用的实例,分别判断它们属于哪个应用领域,并分析它们使用了哪种实现途径。

任务二 AI 的发展历程

◆ 引导思考

1956 年,美国达特茅斯学院举办暑期研讨会,正式确立了“AI”这一名称。自此 AI 踏上发展之路,一路走来几经沉浮,为何业内都说它的发展史,是一部热潮与寒冬轮番上演的演变史?

历经近八十年岁月沉淀,哪些标志性事件与核心人物,深刻影响了 AI 的发展走向? 又是什么核心力量,推动它走出实验室,从最初仅供科研探索的研究方向,转变为如今全民都能便捷使用的日常工具?

一、AI 研究与图灵奖

正式梳理 AI 发展历程前,先来了解计算机领域一个重要的奖项——图灵奖。
图灵奖素有计算机界诺贝尔奖之称,是整个计算机科学领域的最高荣誉,由美国计算机协会在 1966 年设立,每年评选颁发,用以表彰在计算机领域作出深远突出贡献的研究者。这项奖项以图灵命名,就是为了纪念提出图灵测试的英国数学家艾伦·图灵。
历年图灵奖获奖者的研究成果,串联起了一部精简的计算机科学发展史。其中,诸多在 AI 领域开创先河的学者都曾获此殊荣。

表 1-4 部分 AI 领域的图灵奖获得者

年份	获奖者	主要贡献	对 AI 发展的意义
1969	马文·明斯基	人工神经网络与 AI 基础理论	AI 领域的奠基人之一,推动了早期 AI 研究
1971	约翰·麦卡锡	提出“AI”术语,发明 Lisp 语言	AI 学科的命名者和创始人之一
1975	艾伦·纽厄尔、赫伯特·西蒙	符号推理与问题求解	开创了符号主义 AI 的研究范式
1994	拉吉·雷迪	大规模 AI 系统与语音识别	推动了 AI 从实验室走向实际应用
2011	朱迪亚·珀尔	贝叶斯网络与因果推理	为 AI 提供了处理不确定性的数学框架
2018	约书亚·本吉奥、杰弗里·辛顿、杨立昆	深度学习的基础工作	开创了当今 AI 技术的核心方法论

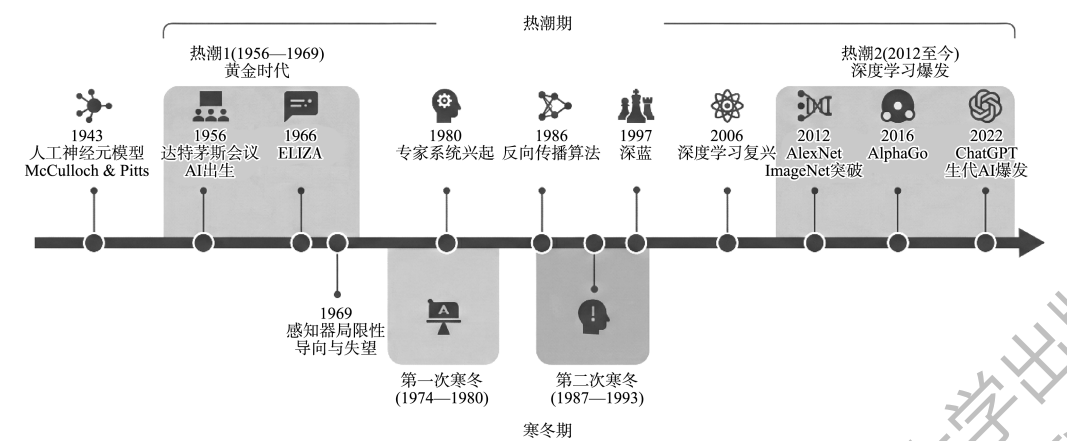


图 1-2 AI 发展历程时间线 (1943 年至今)

二、启蒙时期(从人工神经元开始,1943—1956 年)

AI 的起源,始于一个朴素的问题:人类大脑究竟如何运转?能否借助数学手段,描摹出它的工作逻辑?

1943 年,神经生理学家沃伦·麦卡洛克联合数学家沃尔特·皮茨,推出史上首个人工神经元数学模型——MCP 模型。MCP 模型做了大量简化设计,高度模拟了生物神经元的运作模式:接收多路输入信号并加权汇总,数值达到设定阈值就启动输出,未达标则保持静默。该模型证实了生物神经元具备的逻辑作用,完全能够依靠数学公式完成描述与模拟。

1956 年夏季,约翰·麦卡锡、马文·明斯基等一众顶尖学者,在达特茅斯学院筹办了为期两个月的学术研讨会。会议筹备期间,麦卡锡首次正式提出 AI 这一专业名称。这场会议标志着 AI 正式成为一门独立的研究学科,也确立了此后数十年行业研究的方向。

知识拓展: 达特茅斯会议

这场会议全称为达特茅斯 AI 暑期研讨项目。会议提案里写下过一段广为流传的愿景:团队计划探索实现多项智能能力,让机器熟练运用语言、建立抽象思维、攻克各类人类专属难题,同时具备自我迭代优化的能力。时至今日,这段愿景依旧是 AI 长久以来的核心发展目标。

当年参与会议的学者,日后大多成为 AI 领域的奠基人。麦卡锡在斯坦福大学搭建起 AI 实验室,明斯基于麻省理工学院创立同类研究平台,香农更是奠定了信息论的核心理论基础。现代 AI 几乎所有主流研究脉络,都能从这场会议的思想交流中找到源头。

三、第一次黄金时期(早期期望无限,1956—1969 年)

达特茅斯会议结束后的十几年,AI 正式步入首个黄金发展期。科研人员满怀热情与乐观心态,在多个研究领域接连取得亮眼的早期成果。

在博弈搜索领域,1952 年阿瑟·塞缪尔研发出可自主学习的跳棋程序。这款程序能够通过自我对弈积累对战经验,不断精进棋艺,最终达到业余高手水准。它也是机器学习最早的实践成果之一,而“机器学习”这个专业名词,正是由塞缪尔率先提出。

在数学定理研究领域,艾伦·纽厄尔与赫伯特·西蒙打造出逻辑理论家程序,实现了数学定理自动推导。1956 年,该程序成功证明了经典著作中的三十余条定理,部分推导思路甚至比原著写法更为简洁。此后二人又研发通用问题求解器,尝试打造能够解决各类难题的通用智能程序。

自然语言研究方面,1966 年麻省理工学院的约瑟夫·维森鲍姆开发出 ELIZA 对话程序,用来模拟心理咨询师与人交流。它仅依靠简单句式匹配、关键词替换完成回复,很

容易让使用者产生被理解的错觉。虽然这款程序原理简单,不具备语义理解能力,却在社会上掀起不小热度,也让学界开始深入探讨人机交互的核心本质。

在机器感知领域,1958年弗兰克·罗森布拉特发明感知器,这是一款能够从数据中自主学习的简易神经网络,可完成基础图形分类工作。这项成果在当时轰动业界,甚至被媒体称作可以独立思考的智能机器。罗森布拉特也对此满怀期待,直言未来感知器能够实现行走交流、图文创作、自我复刻,甚至拥有自我意识。

这一阶段的研究还获得了政府大力扶持,美国国防高级研究计划局等机构投入巨额科研资金,全力推进机器翻译、语音识别、图像解析等实用技术落地。充足的经费加上不断涌现的技术成果,让整个行业信心高涨。1957年赫伯特·西蒙预言,十年之内计算机就能战胜人类顶尖棋手拿下国际象棋冠军;1967年马文·明斯基更是认为,短短一代人的时间,就能彻底攻克AI研发难题。事实却证明,人类远远低估了智能研究的复杂难度,近七十年过去,当初定下的诸多目标,至今还有不少尚未实现。

案例: ELIZA——初代聊天机器人

ELIZA的工作原理并不复杂,核心依托预设的模式匹配规则运行。比如它的应答逻辑:一旦捕捉到用户语句里“我的XX”这类句式,就会对应回复“你的XX怎么了,不妨再多说说”。

这款程序本身不具备语义理解能力,全程只是机械套用句式完成文字应答。但让人意外的是,即便提前知晓它只是一款简易程序,很多使用者依旧会觉得它读懂了自己的情绪与心事。开发者维森鲍姆也为此深感忧虑,还专门出书提醒大众,不要将这种简单的文字匹配,误认为是真正的智能思考。

以如今的眼光来看,ELIZA算是现代大语言模型对话产品的早期雏形。二者技术水平差距悬殊:ELIZA只靠几十条人工编写规则运行,而如今的大模型依托千亿级参数深度神经网络搭建而成。

四、第一次寒冬时期(现实瓶颈未获突破,1974—1980年)

黄金时代高涨的乐观氛围没能一直延续。从二十世纪六十年代末期开始,科研人员渐渐意识到,AI研发路上的实际难题,远比最初预想的更加棘手。接连出现的各类现实难题,彻底打破了行业内的美好憧憬。

最先迎来冲击的是机器翻译领域。1966年,美国相关学术委员会发布权威评估报告,直接否定了当时机器翻译的实用价值。报告指出,彼时的翻译成果质量达不到落地使用的要求,短时间内也找不到彻底改善的办法。这份报告出台后,美国官方直接缩减相关科研拨款,自然语言处理整个领域的发展信心也随之受挫。

第二重打击来自感知机自身的先天不足。1969年,明斯基和西摩·佩珀特出版了《感知机》一书,书中指出感知机无法完成异或逻辑的处理。异或逻辑是一种基础逻辑问题,两个条件仅满足其一即可成立,同时满足或同时不满足则不成立。感知机无法学习掌握基础的逻辑运算,这一结论否定了整个神经网络研究方向。受此影响,神经网络相关研

究沉寂十余年,对应的科研经费也基本全面中断。

第三个难以突破的困境便是组合爆炸问题。早期 AI 程序能在小型简易问题中发挥作用,依托的是有限的问题范围。但问题规模稍有扩大,需要遍历推算的可能性就会成倍激增,即使用上当时运算速度最快的计算机,也无法在合理时间内完成运算。

1973 年,英国学者莱特希尔受政府委托撰写行业调研报告,直言 AI 研究并未兑现最初许下的发展愿景。这份报告出台后,英国大幅削减 AI 科研资金,其他国家也纷纷效仿, AI 就此步入发展史上第一个行业寒冬。

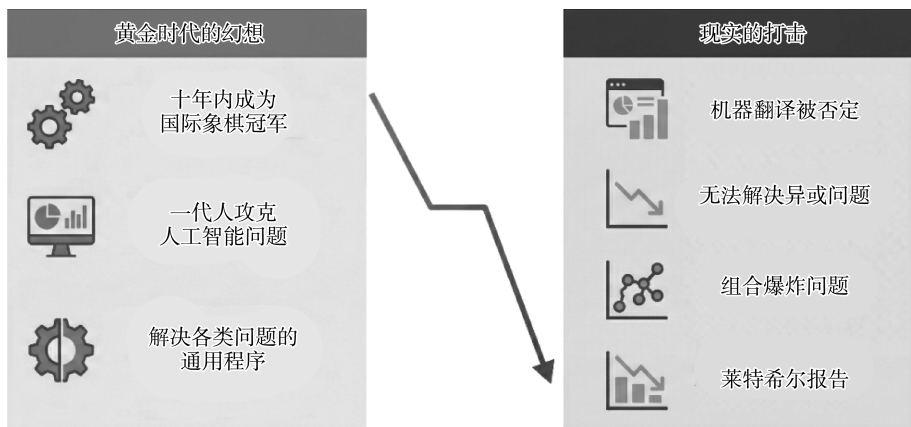


图 1-3 AI 的“期望与现实”——第一次热潮到第一次寒冬

五、第二次寒冬时期(专家系统弊端,1987—1993 年)

第一次行业寒冬并未让 AI 研究彻底停滞,反而倒逼研究者调整研发方向。既然通用 AI 的目标遥不可及,学界便选择稳中求进,深耕垂直细分领域,打造具备实际使用价值的智能程序,专家系统就此顺势兴起。

1969 至 1979 年间,多款经典专家系统陆续问世。例如斯坦福大学研发的两款系统: DENDRAL 系统主打化学领域应用; MYCIN 系统则专注医疗领域。这两个系统在对应领域表现出色,充分证实了 AI 在细分行业里的实用价值。

进入八十年代,专家系统走出实验室正式开启商业化落地, AI 迎来第二波发展热潮。1980 年, 卡内基梅隆大学研发的订单配置系统投入企业使用, 专门负责计算机订单搭配规划, 成为专家系统商业化成功的典范。受此影响, 各大企业纷纷入局研发专属专家系统, 到八十年代中期, 各行各业已有数千套同类系统投入运行。

看似蓬勃发展的专家系统, 背后却藏着难以规避的缺陷: 把专家积累的实战经验转化为规则耗时费力, 许多只可意会的隐性经验更是无法用规则表述; 构建的规则运行灵活性不足, 规则之间还易出现矛盾冲突, 一旦遇到突发特殊情况, 容易输出错误结论。

八十年代末期, 专家系统的种种弊端彻底暴露在人们面前, AI 行业再度陷入信任危机。各大企业纷纷缩减相关投资, 各国政府也下调科研扶持力度, 1987 年至 1993 年, AI 迎来第二次发展寒冬。

案例：AI 两次寒冬的启示

复盘 AI 发展史上的两次行业寒冬，不难总结出规律：行业热度褪去、陷入低谷，往往都是过高的发展期待，与当下实际技术实力严重脱节所导致。

第一次寒冬出现在 1974 至 1980 年，当时科研人员急于朝着通用智能的方向推进，早早定下宏大发展目标，但受限于落后的算力与尚未成熟的理论体系，根本无力实现。各大资助方看清现实差距后，纷纷失去信心，科研经费也随之大幅缩减。

第二次寒冬介于 1987 年到 1993 年，专家系统商业化落地掀起热潮，市场对其寄予厚望，可这类系统应对复杂开放场景时漏洞百出，实际性能远达不到大众使用预期。

这种过度夸大前景、最终愿景落空、行业整体降温的发展轨迹，就是典型的技术炒作周期。这个轨迹同样适用于如今火热的大模型浪潮：既要正视 AI 如今取得的巨大技术突破，也不能盲目跟风，轻信 AI 能够包揽所有事务的片面言论。

六、复苏时期(神经网络的回归,1986 年至今)

就在专家系统逐渐走向没落之际，曾经一度被学界放弃的神经网络研究，开始悄然复苏。

早在 1969 年，受单层感知器理论缺陷的影响，神经网络相关研究近乎陷入停滞，但仍有部分科研人员始终坚守这一方向。行业迎来关键转机是在 1986 年 10 月，大卫·鲁梅尔哈特、杰弗里·辛顿和罗纳德·威廉姆斯联合发布论文，提出将反向传播算法运用在多层神经网络中的可行方案。它的原理非常简单：当模型输出结果出现偏差时，算法会从结果端向前逐层追溯，算出每一层节点对误差造成的影响，再据此调整网络的权重，让后续输出不断贴近标准答案。

反向传播算法的问世，让训练大规模的人工神经网络成为现实，不仅能够轻松解决感知器难以解决的异或逻辑问题，理论上还可以拟合各类复杂函数。

不过这一阶段的人工神经网络依旧存在明显局限，受限于当时硬件算力不足，训练大型网络耗时极久，同时可供使用的优质训练数据也十分短缺。即便理论层面成果颇丰，实际落地效果依旧比不上成熟的传统统计学习方法。而神经网络真正迎来全面腾飞，还要等待大数据时代来临与高性能算力全面普及。

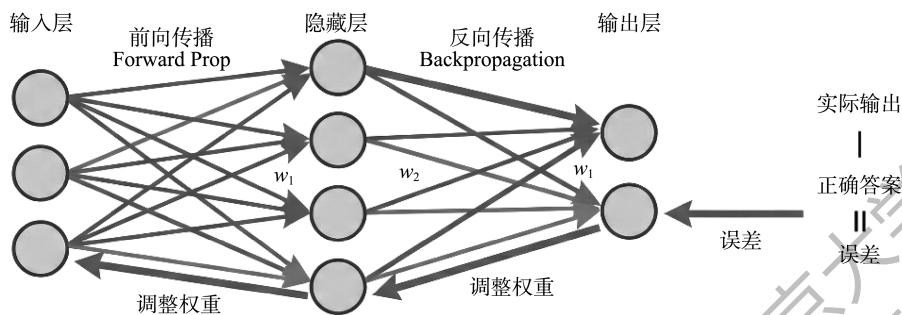


图 1-4 反向传播算法的直观理解

七、持续复苏时期(概率推理和机器学习的兴起,1987 年至今)

在神经网络复苏的同一时期, AI 领域还发生了另一场深刻的变革: 研究者们开始系统地引入概率论和统计学的工具来处理不确定性问题。这一转变从根本上改变了 AI 研究的方法论基础。

早期的符号主义 AI 有一个重要的假设: 世界是可以用确定的逻辑规则来描述的。但现实世界充满了不确定性——明天会不会下雨? 这封邮件是不是垃圾邮件? 这张 X 光片上的阴影是不是肿瘤? 对于这类问题, “非黑即白”的逻辑规则往往力不从心, 人们更需要的是“大概率是”或“有 80% 的可能性”这样的概率性判断。

1988 年, 朱迪亚·珀尔出版了一本影响深远的著作, 系统地提出了贝叶斯网络的理论框架。贝叶斯网络用一种直观的图形结构来表示变量之间的概率依赖关系, 使得 AI 系统能够在不确定的情况下进行合理的推理。用一个生活化的例子来说: 如果早上看到地面是湿的, 人们会推测“昨晚可能下过雨”; 但如果人们同时注意到邻居的洒水器开着, 对“下过雨”的信心就会降低。贝叶斯网络能够精确地处理这种“多个线索相互影响”的复杂推理场景。

与此同时, 机器学习作为一种通用的方法论开始获得越来越多的关注。20 世纪 90 年代, 一系列重要的机器学习算法相继成熟, 包括决策树、支持向量机、随机森林等(这些算法的原理将在模块二中详细介绍)。这些算法的共同特点是: 它们不依赖于手工编写的规则, 而是从数据中自动学习模式和规律。

机器学习方法的实用价值在 1997 年得到了一次戏剧性的展示。IBM 开发的国际象棋程序深蓝(Deep Blue)在六局比赛中以 3.5 : 2.5 击败了当时的世界国际象棋冠军加里·卡斯帕罗夫。虽然深蓝主要依靠的是强大的搜索能力(每秒评估 2 亿个棋局位置)而非“学习”能力, 但这场胜利在公众中产生了巨大的影响, 让全世界第一次直观地看到: 在某些特定领域, 计算机已经可以超越人类的最高水平。

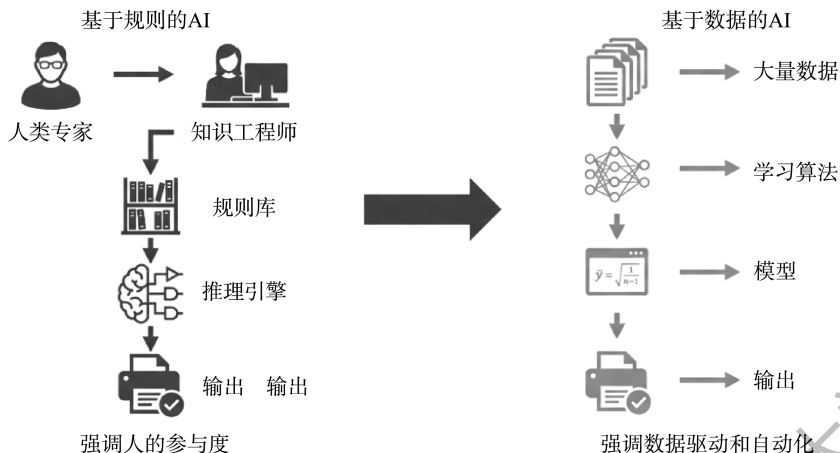


图 1-5 AI 方法论的范式转变

这一时期的另一个重要变化是:AI 研究变得更加“科学化”。研究者们不再满足于在特定例子上展示系统的能力,而是开始重视严格的实验方法论——使用标准数据集进行测试、用统计指标衡量性能、与其他方法进行公平对比。这种方法论的成熟为后续的技术突破奠定了坚实的基础。

表 1-5 AI 方法论的演变:从规则到数据

维度	早期符号主义 AI	概率/统计学习 AI
知识来源	专家手工编写规则	从数据中自动学习
处理不确定性	困难,倾向于非黑即白	擅长,用概率表示信心程度
适用场景	规则明确的封闭领域	数据丰富的开放领域
典型产物	专家系统	垃圾邮件过滤、推荐系统
主要瓶颈	知识获取困难	需要大量高质量数据

八、第二次黄金时期(大数据时代来临,2010 年至今)

步入 21 世纪,大数据强势崛起,成为彻底扭转 AI 发展走向的核心力量。互联网飞速普及,催生了规模空前的数据浪潮。搜索引擎留存海量用户搜索记录,社交平台每日产出海量图文动态,电商平台沉淀着完整的用户消费行为数据,手机各类传感器也持续采集定位、影像、动态等多元信息。2010 年前后,人类每两天产生的数据总量,就已经追平了从文明起源至 2003 年积攒的全部数据。

数据之所以能推动 AI 前行,核心原因契合机器学习的本质:机器学习依靠数据总结规律,数据便是模型学习成长的根基。缺少充足优质的数据支撑,再先进的算法也难以发挥作用,而大数据时代的到来,恰好为 AI 发展储备了充足的学习资源。

在数据持续暴涨的同时,硬件算力也迎来跨越式升级。依照摩尔定律,传统 CPU 性能大致每两年实现翻倍。而针对神经网络所需的大规模并行运算,GPU 的跨界运用更是带来了巨大突破。

GPU 最初专为游戏画面渲染打造,其强大的并行运算能力,刚好适配神经网络训练需求,单块高性能 GPU 的运算效率,远超大量普通 CPU 组合。好比原本用于基建搬运的专用设备,恰好完美适配粮食转运工作,实现了用途价值的全新挖掘。

充足的数据资源搭配强劲的硬件算力,为深度学习迎来全面爆发筑牢了全部基础。

知识拓展：数据的“质”与“量”

大数据对于 AI 的价值,不仅体现在数量,更在于内容。例如训练一个人脸识别模型时,倘若仅用一千张人脸正面照片训练,模型很难识别出侧面、远景照片中的人脸;而如果投入数百万张涵盖不同角度、光影、肤色、年龄的人脸照片,模型的识别水平便会大幅提升。

九、第三次黄金时期(深度学习爆发,2012 年至今)

2010 年起,海量数据、高性能算力与成熟的神经网络理论正式汇聚交融,一场颠覆性的 AI 技术革命顺势爆发,深度学习就此强势崛起,成为 AI 发展进程中声势最为浩大的发展浪潮。

2012 年的 ImageNet 图像识别赛事中,使用深度学习技术的 AlexNet 模型,直接将图像分类的错误率从 26% 降至 15.3%,在业内引发巨大震动。这项成果证明了深度学习可直接依托原始数据自主提取特征,大幅度超越了依靠人工提炼特征的传统方法。

在此之后,深度学习在各大领域接连实现突破。语音识别精度大幅提升,让各类智能语音助手走进大众生活;自然语言翻译效果实现质的飞跃,全面超越传统规则式翻译技术;2016 年 AlphaGo 战胜人类围棋顶尖棋手,更是直观展现出深度学习强大的实战能力。

深度学习飞速发展之下,全新的技术研发模式应运而生,预训练大模型登上行业舞台。2017 年谷歌专为文本数据打造了 Transformer 架构。依托这一架构,各类大语言模型接连诞生,模型中的参数规模也从数亿一路增长至万亿级别。

不止语言交互领域,AI 还广泛应用于气象预报、新型材料研发、数学定理推导等前沿科研工作。如今深度学习早已不再是单一技术,而是融入各行各业、遍及日常与科研的新型数字基础设施。

生活中的深度学习

如今深度学习早已深度融入日常生活,每天都在使用的功能,背后都依靠这项技术支撑,只是大多数人未曾留意。

手机拍照时,按下快门的瞬间,深度学习模型会自动识别风景、人像、美食等不同拍摄场景,智能调校色彩与曝光参数。即便在暗光环境下,也能通过算法优化画面亮度,拍出清晰好看的照片。

日常使用语音输入,说话转文字的全过程,都由深度学习语音识别模型实时完成。这项技术早年识别准确率不足八成,如今在安静环境中,识别准确率已经超过 97%。

刷短视频时,平台推送的内容都经过深度学习推荐模型筛选。模型结合用户浏览习惯、视频内容、兴趣偏好等诸多信息,精准预判喜好,完成个性化内容推送。

在外卖服务场景里,平台预估配送时长、规划骑手出行路线、智能分配订单等核心环节,都大量运用深度学习技术,有效提升整体配送效率与调度精准度。

案例：从深蓝到 AlphaGo 再到 ChatGPT

1997 年,深蓝战胜国际象棋世界冠军,它依靠的是暴力搜索策略,每秒可推演两亿个棋局,凭借超强算力穷举最优落子思路。虽说实战表现十分出色,但深蓝本质只是一台高速运算机器,不具备自主学习的能力。

2016 年,AlphaGo 拿下围棋赛事胜利。围棋棋局复杂度远高于国际象棋,传统暴力搜索已然失效。这款模型结合深度学习与强化学习,研读海量对弈棋局,慢慢练就出连顶尖棋手都难以言说的行棋思路。

2022 年面世的 ChatGPT 风靡全球,它不再局限于下棋这类单一任务,在语言场景中展现出极强综合能力,可完成写作、翻译、编程、逻辑分析等各类工作,实现了 AI 从专用领域工具,向通用交互助手的重要跨越。

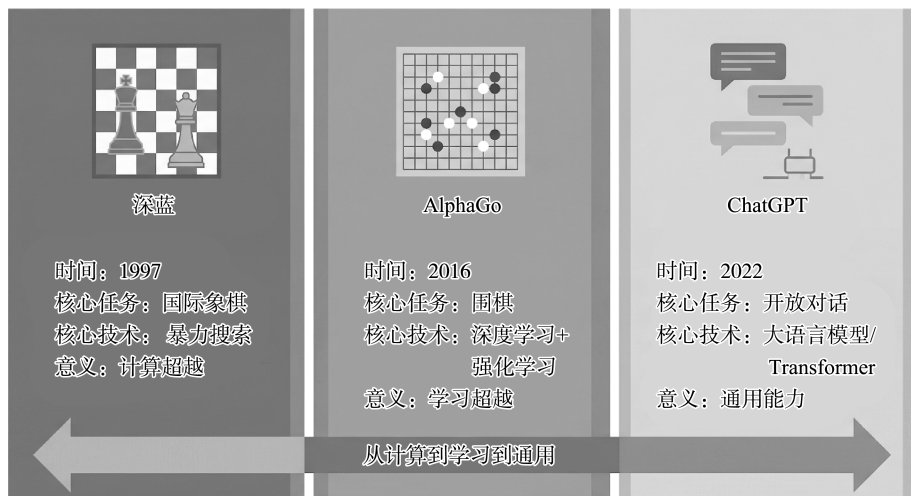


图 1-6 AI 发展的三次标志性飞跃:深蓝、AlphaGo 与 ChatGPT

本节小结

1. AI 研究的主线由早期符号逻辑推理,过渡到概率统计学习,最终迈入深度学习时代。
2. AI 概念萌芽于 1943 至 1956 年,MCP 神经元模型、图灵测试奠定了 AI 理论根基,达特茅斯会议正式确立 AI 学科名称,标志 AI 领域的诞生。
3. AI 的发展整体呈热潮与寒冬交替前行的态势:1952 至 1969 年迎来首个发展黄金期,因预期过高陷入第一次行业寒冬;1969 至 1986 年专家系统兴起并走向商业化,后期弊端凸显,催生第二次 AI 寒冬。
4. 1986 年反向传播算法落地应用,让沉寂已久的神经网络重回研究视野;九十年代概率推理体系成型,各类经典机器学习算法也逐步完善。
5. 进入 21 世纪,海量数据资源与 GPU 高性能算力持续赋能,为深度学习崛起筑牢根基。2012 年 ImageNet 赛事拉开深度学习发展大幕,2020 年开始,各类语言大模型的问世,让 AI 走入大众生活,成为全民通用工具。

思考题

1. [识记] 请按时间顺序列出 AI 发展史上的五个关键事件,并用自己的话说明每个事件为什么重要。
2. [理解] 有人认为“AI 的两次寒冬说明 AI 注定是一场泡沫”,也有人认为“寒冬恰恰是为下一次突破积蓄力量”。你更倾向于哪种看法?请结合本任务所学的历史事实,阐述选择的理由。
3. [应用] 参考 AI 发展史上“技术炒作周期”的规律,分析当前大语言模型热潮可能面临的风险和挑战。你认为这一轮热潮会重蹈“寒冬”的覆辙吗?为什么?

任务三 AI 发展的三大引擎

◆ 引导思考

AI 经历了多次“热潮”与“寒冬”的交替。那么,是什么力量最终推动 AI 走出寒冬,迎来今天的全面爆发?

人们常说,数据、算力和算法是驱动 AI 发展的“三大引擎”。这三者各自扮演着怎样的角色?它们之间又是如何相互配合的?缺少其中任何一个,AI 还能运转吗?

理解这三大引擎的协同关系,对认识当前 AI 技术的能力与瓶颈有什么帮助?

从 1943 年至今,纵观 AI 的整个发展历程,不难发现 AI 每一次跨越式突破,都并非依靠单一技术,而是多重条件齐备、多方力量共同推动的成果。

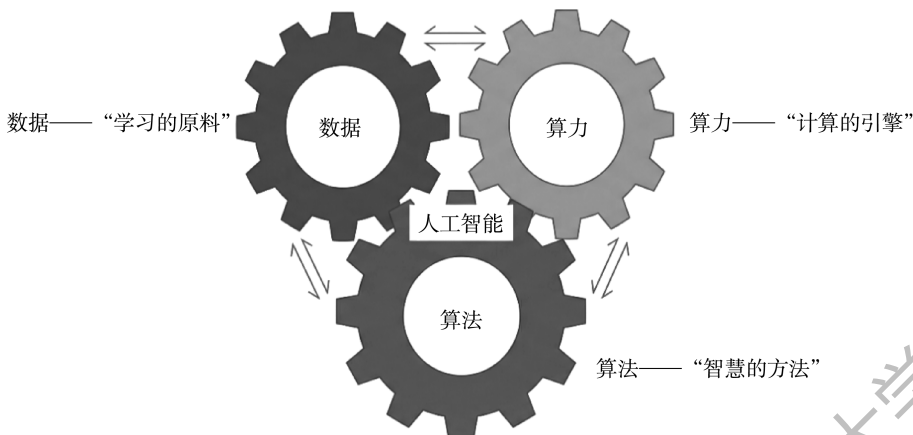


图 1-7 AI 的三大引擎

2012 年深度学习迎来全面爆发就是最好的印证,这场技术变革的到来,离不开三大核心条件同步成熟:互联网沉淀下的海量训练数据、GPU 带来的强劲并行算力,再加上深度神经网络、反向传播等成熟高效的训练算法。

这三大核心要素,称作驱动 AI 前行的三大引擎。如同汽车行驶离不开燃油、发动机与传动系统协同配合,AI 想要持续发展进步,数据、算力、算法三者相辅相成,缺一不可。接下来,分别深入了解每一个引擎的作用。

一、数据

机器学习的核心逻辑,就是让机器依托数据自主总结规律。其中数据便是模型学习必不可少的原材料,AI 想要完成学习训练,也必须依靠足量数据作为支撑。

互联网问世之前,数据大多依托纸质资料、实体照片、录音磁带等形式存储,不仅整体存量有限,还很难被计算机直接读取与运用。互联网全面普及后,这一情况彻底被改写。人类社会的数据总量,正以惊人的速度迅猛增长。搜索引擎留存着海量用户搜索记录,社交平台每日产出海量图文动态,电商平台积累了庞大的交易数据与用户行为信息,手机摄像头、各类智能传感器更是源源不断生成图片、视频、地理位置等多种类型数据。

数据的爆发式增长,为 AI 提供了无比充足的学习资源。但数据数量庞大,并不代表模型效果就一定出色。倘若训练数据存在大量错误标注、重复内容或是固有偏差,模型学习总结出的规律也必然存在漏洞。

为此,AI 项目研发首先要面对的是数据治理问题,涵盖了数据的采集、清洗、标注、存储,其重要性不亚于算法研发。AI 行业中流传着一句话:普通项目开发中,八成精力都耗费在数据准备工作上,仅有两成时间用于模型训练。这句话凸显了数据工作在 AI 项目开发过程中的重要性。

案例:数据标注——AI 背后的“隐形劳动”

人们时常惊叹 AI 工具能够精准识别图片里的猫狗、车辆,却很少知晓这份能力,全都依托海量人工标注工作搭建而成。数据标注,就是工作人员逐一核对图片等原始素材,手动添加对应类别标签,而 AI 模型正是依靠这些标注信息完成知识学习。

据行业统计,全球有数百万从业者专职从事数据标注工作,其中多数集中在发展中国家。他们日复一日完成勾选、框选、分类标记等重复性工作,默默为 AI 的表现筑牢根基。AI 所展现出的各类智能能力并非凭空产生,它们都扎根在庞大的数据标注等人力劳作之上。

二、算力

数据是 AI 的学习原料,算力便是加工原料的核心动力。算力就是计算机完成各类运算任务的综合能力,运算速度越快、并行处理数据的规模越大,代表算力越强。

AI 研发离不开强劲算力支撑,当下主流大语言模型的一次完整训练,需要上千块高

端 GPU 持续运转数周甚至数月,训练消耗的电能,不亚于一个大型社区。

传统 CPU 处理器等同于一位擅长处理各类复杂事务的全能型人才,但是它同一时段只能完成少量工作。GPU 则截然不同,它如同由海量基础运算单元组成的作业团队,每个单元只能进行简单的运算,却能同步协作、批量处理数据。而神经网络训练中最依赖的是大规模并行的简单乘法计算,非常适合使用 GPU 中的众多运算单元同步完成。

凭借 GPU 算力的不断突破,训练超大型神经网络渐渐成为现实。从 2012 年 AlexNet 的六千万参数,到如今大语言模型数千亿级别的参数规模,模型每一次体量升级,都建立在算力稳步提升的基础之上。

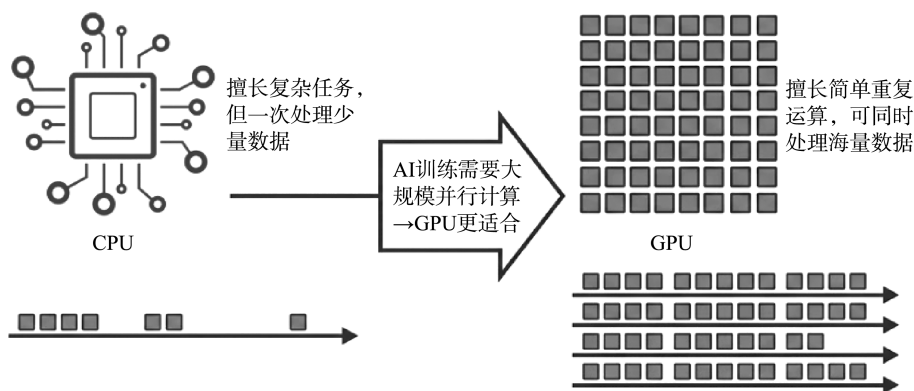


图 1-8 CPU 与 GPU 的工作方式对比

三、算法

有了作为原料的数据、充当动力的算力,还需要对应的加工方式完成转化,这就是算法承担的作用。解决问题的流程与思路、指引计算机依照既定规则与步骤处理各类数据的过程即是算法。

结合 AI 的发展历程不难发现,行业每一次重大跨越,都离不开算法层面的革新。1986 年反向传播算法问世,打通了多层神经网络的训练壁垒;2012 年 AlexNet 依托卷积神经网络结构优化与工程调优大放异彩;2017 年 Transformer 架构诞生,更是为大语言模型的飞速发展筑牢根基。足以见得,算法创新是持续拓宽 AI 能力上限的核心动力。

除此之外,算法升级不止追求识别、预测准确率提升,更注重运行效率优化。优质的算法优化方案,能够在不额外增加算力投入的前提下,显著提升模型整体表现,相当于低成本扩充运算资源。

表 1-6 AI 领域的关键算法里程碑

时间	算法/架构	核心贡献	影响
1986 年	反向传播算法	实现多层人工神经网络高效训练	正式推动连接主义重回主流研究视野

续表

时间	算法/架构	核心贡献	影响
1989 年	卷积神经网络 (CNN)	专门处理图像的网络结构	计算机视觉的核心技术
1997 年	长短期记忆网络 (LSTM)	擅长解析处理序列和时间信息	推动语音识别和翻译的发展
2014 年	生成对抗网络 (GAN)	依靠双模型博弈式学习	开辟 AI 图文内容生成新赛道
2017 年	Transformer 架构	依托注意力机制优化序列数据处理	大语言模型的基石

知识拓展：算法创新与“站在巨人肩膀上”

AI 算法的发展是持续积累、迭代优化的过程。每一项重大算法突破，都建立在前人研究成果之上。比如 Transformer 架构里的注意力机制，最早可溯源至上世纪 90 年代的相关研究；反向传播算法的数学原理，也曾在不同领域被多次独立推导发现。

创新大多并非从零起步，而是对现有知识进行重组整合，再加以创造性运用。牛顿曾说“我之所以看得远，是因为站在巨人的肩膀上”，AI 算法的演进之路，正是这句话最鲜活的现实印证。

四、数据、算力、算法的协同关系

数据、算力与算法彼此联系紧密、相辅相成。缺少足量优质数据，再精妙的算法也无从施展；没有充足算力支撑，再先进的算法构想也只能停留在理论层面；算法设计落后，再

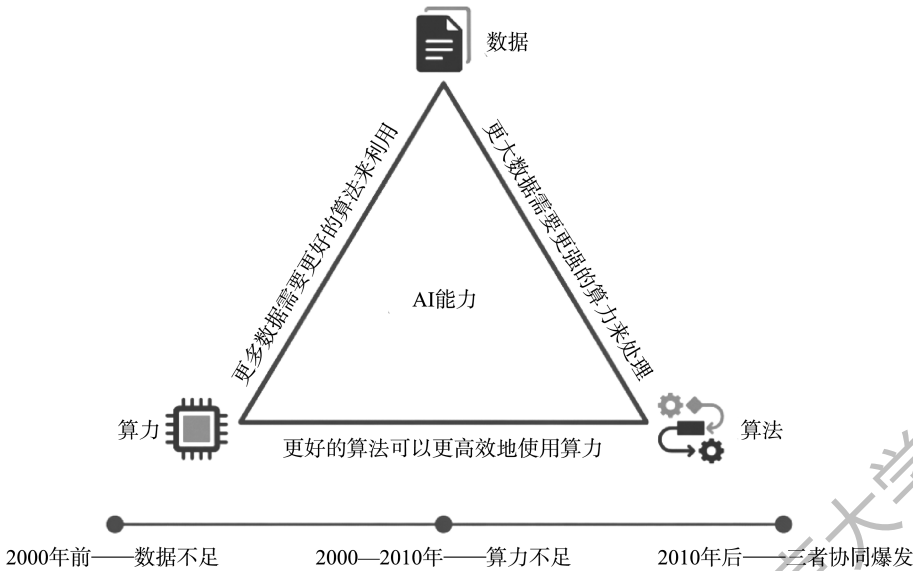


图 1-9 三大引擎的协同驱动关系

多数据与强悍算力也只会造成资源浪费。

在 AI 行业的不同发展阶段,总会受限于最薄弱的一环:2000 年以前,发展瓶颈主要是数据资源匮乏;深度学习起步阶段,算力不足成为最大阻碍;如今数据与算力条件日趋成熟,算法创新再度成为实现技术突破的核心关键。三者彼此助推,海量数据推动算法向更大规模、更复杂结构迭代升级,大型模型又进一步催生更高算力需求;算力升级能够轻松处理海量数据,高效算法也可在同等算力下充分挖掘数据价值,这套循环持续驱动 AI 技术不断向前发展。

表 1-7 三大引擎的角色与关系

引擎	核心作用	当前挑战
数据	为模型供给学习样本,支撑规律挖掘	质量高低不一,人工标注成本高,存在隐私泄露、数据偏见等问题
算力	保障模型训练与推理的运算能力	使用成本昂贵,能耗碳排放高,高端芯片供应受限
算法	制定数据处理逻辑,高效提炼知识规律	前沿理论突破存在不确定性,核心创新高度依赖专业顶尖人才

从更宏观的视角来看,数据、算力、算法的协同发展不仅是一个技术问题,也深刻地影响着产业格局和社会发展。谁掌握了高质量的数据、强大的算力和领先的算法,谁就有可能在 AI 时代占据竞争优势。正因如此,全球各大经济体纷纷将数据中心建设、高端芯片研发、前沿算法攻关等 AI 基础设施建设,纳入重点国家发展战略。

本节小结

1. 数据、算力、算法是驱动 AI 发展的三大核心引擎,三者相辅相成、缺一不可。
2. 数据是 AI 学习的基础原料,其体量、丰富度与质量直接决定模型效果,数据采集、清洗、标注等治理工作是 AI 开发的重中之重。
3. 算力为模型训练与推理提供运算支撑,GPU 的普及助力深度学习快速兴起,目前算力仍存在成本高、能耗大、芯片供应受限等现实难题。
4. 算法是 AI 的核心方法,技术迭代与重大突破均依托算法创新,从反向传播到 Transformer 架构,持续拓宽 AI 应用能力边界。
5. 数据、算力、算法相互依存、彼此赋能,同时存在明显短板制约效应。

思考题

1. [识记] 请用自己的话,分别解释数据、算力和算法在 AI 发展中各自扮演的角色,并为每个角色找一个生活中的类比。
2. [理解] 有人说“只要有足够多的数据,AI 就能解决一切问题”。请结合三大引擎的协同关系,分析这一说法的合理性和局限性。
3. [应用] 假设你所在的学校计划开发一个“智能课程推荐系统”,根据学生的兴趣和

成绩推荐选修课程。请从数据、算力和算法三个维度分析：这个系统需要什么样的数据？对算力的要求高吗？可能采用什么类型的算法？开发过程中最大的挑战可能在哪个“引擎”上？

任务四 AI 的应用现状

◆ 引导思考

日常使用手机、浏览短视频、下单外卖，诸多生活场景背后，都有 AI 在默默赋能运转。那么放眼产业层面，AI 行业已然形成了怎样完整的分工协作体系？从最底层的智能芯片，到日常使用的各类手机应用，其间又串联着哪些产业链环节？

当下 AI 产品正在医疗、教育、工业制造、现代农业、金融服务等众多行业中实现落地应用。AI 在哪些行业已经大量应用，哪些仍处于探索起步阶段？认清 AI 在各行各业的真实应用现状，又能为人们自身的学习规划与职业发展方向，带来哪些实用启示？

一、AI 产业链

想要全面认清 AI 的应用全貌，首先要掌握产业链这一核心概念。产业链，指的是一项前沿技术从实验室构想，逐步打磨成型，最终转化为大众可用产品所历经的全部流程环节。目前 AI 产业链主流划分为三大层级，分别是基础层、技术层与应用层。

1. 基础层

基础层为 AI 运行筑牢底层根基，主要包含数据资源、计算硬件、云服务平台三大板块。

数据是 AI 学习的核心原料。如今数据采集、清洗、标注及流通，已经形成了一个规模庞大的成熟市场。计算硬件以支撑 AI 模型、AI 服务运行的专用芯片为主，为模型训练与推理全程提供算力支持。云服务平台整合算力与数据资源，以按需租用的形式面向开发者开放，大大降低了 AI 模型的研发门槛。依托云平台，中小型研发团队无需采购高价硬件设备，也能顺利完成模型训练工作。

2. 技术层

技术层是衔接基础层与应用层的重要桥梁，负责将底层资源整合转化为可直接落地使用的 AI 核心能力。主要分为三类：TensorFlow、PyTorch 等通用 AI 开发框架；提供通用语言理解、文本生成等基础能力的各类预训练大模型；以及计算机视觉、语音识别、自然语言处理等专项智能技术。

当今技术层最显著的发展趋势便是平台化与 API 接口化。以往企业想要运用 AI 技

术,必须从零开始搭建团队、收集数据、自主训练模型;如今多数模型都以 API 接口形式对外开放,企业直接调用接口,就能快速获取对应功能。这就如同无需自建发电厂,接入电网插座便能用电一般,大幅简化了 AI 落地应用的流程。

3. 应用层

应用层是最容易直观感受到的层级,它把技术层的 AI 能力和各行各业实际需求相结合,打造出面向普通用户的各类智能产品与服务。目前日常使用的语音助手、聊天 APP、人脸识别系统等,全都隶属于应用层。

行业深度定制是应用层最突出的特点。深度学习等底层 AI 技术具备通用属性,但想要真正在行业内落地见效,就必须理解对应行业的业务流程、数据特征与实际使用需求。AI 想要顺利落地应用,离不开技术人才与行业专业人才的携手协作。

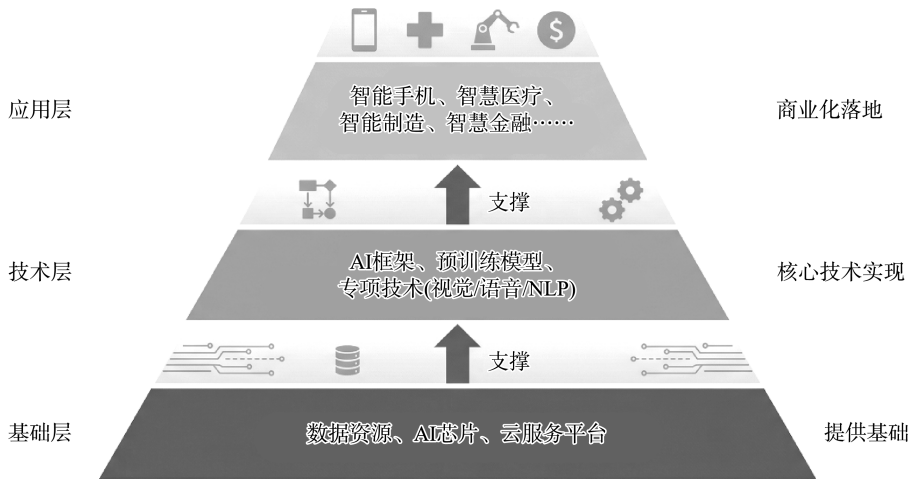


图 1-10 AI 产业链三层结构

表 1-8 AI 产业链三层结构概览

层次	核心内容	代表性企业/产品
基础层	数据资源、AI 专用芯片、云计算服务平台	英伟达、华为昇腾芯片、阿里云
技术层	AI 开发框架、预训练大模型、计算机视觉、语音识别等专项技术	谷歌 TensorFlow、百度文心大模型
应用层	贴合各行各业需求的 AI 产品与落地服务	智能语音助手、智能安防系统

二、AI 的三大产品趋势

1. 从专用智能走向通用智能

传统 AI 大多是专项智能,单个模型只能完成单一任务。而以大语言模型为代表的通

用大模型,彻底打破了这一局限。一款通用大模型可同时实现文案写作、语言翻译、代码编写、数据分析等多项功能,用户只需用日常自然语言下达指令即可操作,无需掌握任何编程知识。

这场变革深刻影响着整个 AI 产业。过去开发一款全新 AI 应用,需要从零采集数据、训练模型,研发周期长、成本高昂;如今开发者依托成熟大模型就能快速搭建各类智能应用,就像在手机系统上开发应用软件一样便捷。这种发展生态,大幅降低了 AI 落地门槛,催生出大量全新应用场景。

2. 从云端算力走向边缘终端

早期 AI 应用高度依赖云端服务器,所有数据上传至远程云端运算处理,再将结果传回设备。这种模式存在响应延迟高、用户隐私易泄露、必须依托网络使用等诸多弊端。

随着小型化 AI 芯片量产与轻量化算法持续优化,越来越多的 AI 能力直接部署在手机、智能汽车、监控摄像头等终端设备上,实现本地离线运算。

日常手机拍照智能识别场景、自动优化画质,新一代手机离线运行小型语言模型实现无网语音交互,都是典型应用。端侧 AI 不仅让智能服务响应更快、隐私安全性更高,还能让 AI 工具在无网络、弱网络的特殊场景中稳定运行,进一步拓宽了应用边界。

3. 从单模态走向多模态

计算机视觉、自然语言处理、语音识别等技术原本各自独立,早期 AI 仅能单独处理文字、图像、音频其中一种数据,属于单模态智能。但人类依靠视觉、听觉等多种感官协同认知世界,多模态 AI 正是朝着这一方向不断进化。

目前主流大模型已能够同步处理文字、图片、语音、视频等多种信息,实现看图释义、语音创作、图文互转等跨模态功能。例如拍摄家中食材照片,便可自动匹配菜品并给出详细制作教程。多模态技术的成熟,让人机交互愈发自然流畅,更加贴合人类的认知与思维习惯。

三、AI 的应用

1. 医疗健康

医疗是 AI 社会价值极高的核心应用领域。主流应用包含医学影像诊断、智能药物研发、大众健康监测三大方向。

AI 可快速分析 X 光、CT、病理影像,精准定位病变位置,大幅提升诊断效率。在实际诊疗中,AI 仅作为辅助筛查工具,帮助医生快速标注可疑病灶,减轻阅片压力,并不会取代医护人员。

传统新药研发耗时 10 至 15 年、投入成本极高,依托 AI 分析海量生物数据与分子结构,能够快速筛选优质候选药物,极大压缩研发周期,AlphaFold 蛋白质结构预测技术,更是为新药研发开辟了全新路径。除此之外,结合智能穿戴设备,AI 还能实时监测人体健

康数据,及时推送健康风险预警。

2. 教育

AI 正持续革新传统教学模式。依托智能学习系统,可依据学生学习情况精准定位知识短板,定制专属学习方案,实现一对一个性化辅导;智能批改系统能够高效完成作文、客观习题批阅,减轻教师工作负担;智能备课工具也可协助老师快速制作课件、设计课堂内容。

与此同时,AI 教育也存在现实争议:过度依赖智能辅导容易弱化学生自主思考能力,AI 难以精准评判写作中的创新思维,大模型的普及也滋生了学术诚信等问题,技术落地需要与教育理念协同发展。

3. 制造业

制造业是 AI 落地最成熟、普及度最高的行业。依托机器视觉实现产品智能质检,可全天候高精度排查产品瑕疵,效率与准确率远超人工;通过采集设备运行温度、震动等数据,实现设备故障提前预判与预防性检修,规避生产停工损失;智能算法还可优化生产排班计划,合理调配产能,提升整体生产效益。

国内制造业智能化转型势头强劲,3C 电子生产线已大规模普及 AI 视觉质检,众多企业打造全流程智能化灯塔工厂,实现从原料入库到成品出库的智能管控,全面提质增效。

4. 交通出行

交通领域的 AI 应用贴近日常生活,辨识度极高。融合视觉感知、传感技术与智能决策的自动驾驶技术,正从辅助驾驶稳步向高阶自动驾驶迭代升级;智慧交通系统依托车流大数据,动态调控红绿灯、规划通行路线,有效缓解城市交通拥堵;AI 也深度融入智慧物流,优化仓储调度、配送路线,助力无人配送落地。

日常出行导航软件,更是数据、算力、算法三大引擎协同运作的典型实例,依靠海量路况数据、强大云端算力与智能路径算法,实时规划最优出行路线。

5. 金融

金融行业数据体量庞大,天然适配 AI 技术落地。智能风控系统依托机器学习分析海量交易数据,秒级识别异常消费、金融诈骗行为,守护用户资金安全;智能投顾结合用户风险承受能力与市场行情,推送合理化理财投资建议;基于自然语言处理的智能客服,可高效处理常规业务咨询,大幅压缩服务成本、提升服务效率。

6. 农业

AI 农业应用虽然关注度较低,但发展潜力十足。借助无人机航拍、卫星遥感采集农田影像,AI 精准判别农田病虫害、缺水缺肥区域,指导农户精准灌溉、科学施药,兼顾增产与生态环保;智能养殖设备实时监测畜禽生长状态与健康数据,及时排查疫病隐患;利用

视觉识别技术完成农产品自动分级筛选,实现农副产品标准化、自动化分拣包装,助力传统农业向智慧农业转型。

生活里的 AI: 细数日常一天邂逅的智能应用

跟随一名普通大学生的日常作息,便能发现 AI 早已渗透全天方方面面:
7:00 智能闹钟依据睡眠节律智能调整唤醒时刻,实现人性化起床提醒;
7:30 浏览资讯时,智能推荐算法精准推送契合喜好的内容;
8:00 出行骑行共享单车,平台依托大数据预判车辆分布,就近推荐取车点位;
10:00 课间拍摄课堂板书,借助 OCR 视觉识别快速转化为可编辑文字;
12:00 点购外卖,平台智能预估配送时长,同时为骑手规划最优配送路线;
15:00 借助大语言模型梳理论文框架、理清写作思路;
18:00 浏览短视频,智能推荐持续匹配个人兴趣偏好;
22:00 语音下达设置闹钟指令,手机语音助手快速识别语义并完成操作。
短短一日间,与 AI 的相遇频次多达二十余次,多数智能服务悄然运行,早已融入日常却不易被察觉。

表 1-9 AI 主要应用领域一览

应用领域	典型场景	涉及 AI 技术	应用成熟度
医疗健康	医学影像分析、新药研发、日常健康监测	计算机视觉、深度学习	部分场景已成熟,仍在持续拓展落地
教育行业	个性化辅导、作业批改、智能备课	自然语言处理、通用大模型	处于快速普及、稳步发展阶段
制造业	产品质检、设备预测性维护、生产排产	计算机视觉、机器学习	AI 落地最成熟、普及度高
交通出行	自动驾驶、智慧导航、智能仓储	多传感器融合、强化学习	分级落地稳步推进,逐步迭代升级
金融行业	风险防控、理财顾问、智能客服	机器学习、自然语言处理	商业化落地较为成熟
现代农业	种植管护、病虫害排查、农产品自动分选	计算机视觉、遥感数据分析	处于起步探索阶段,发展潜力巨大

纵观各行业 AI 应用现状,其落地普及程度存在明显差异。整体而言,数据资源充足、业务流程标准化程度高的金融、制造等行业,AI 应用发展更为成熟;而自动驾驶这类对安全等级要求严苛,以及教育、医疗等高度依赖人际沟通与专业经验的领域,AI 目前仍以辅助赋能为主,想要实现全流程自主运行,还需要较长的发展周期。

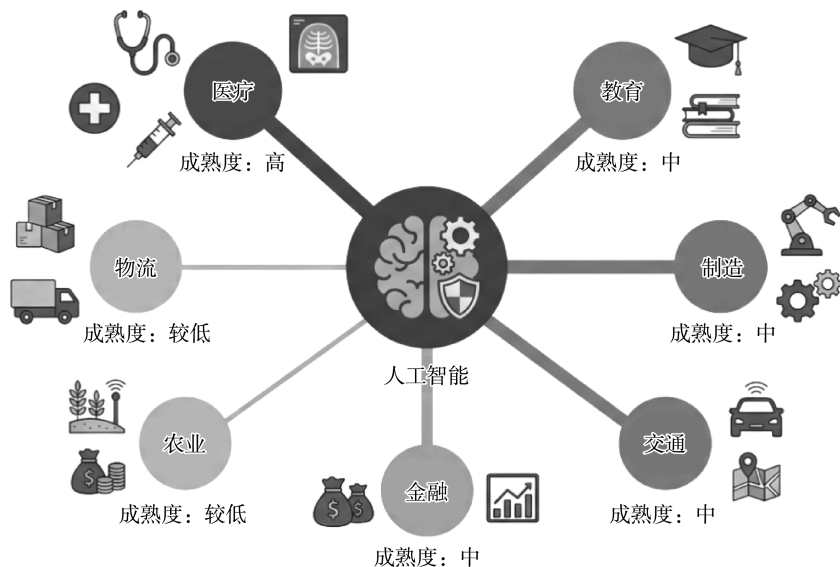


图 1-11 AI 在各行业的渗透

知识拓展：中国 AI 产业的发展

中国是全球 AI 发展最为活跃的国家之一。在基础层,华为的昇腾 AI 芯片、寒武纪的智能计算芯片等产品正在逐步构建自主可控的 AI 算力生态;阿里云、腾讯云、百度智能云等平台提供了丰富的 AI 云服务。

在技术层,百度的文心大模型、阿里巴巴的通义千问、科大讯飞的星火大模型等中国自主研发的大语言模型相继推出,在中文理解和生成方面表现出色。

在应用层,中国在智慧城市、智能制造、金融科技、智慧医疗等领域的 AI 应用处于全球领先水平。大疆的无人机在农业、测绘、物流等领域被广泛使用;旷视科技和海康威视在计算机视觉领域积累了丰富的行业落地经验。

中国 AI 产业的发展得益于几个独特的优势:庞大的人口基数带来了丰富的数据资源,完善的互联网基础设施提供了数据采集的便利条件,活跃的移动互联网生态为 AI 应用提供了广阔的落地场景。

AI 行业迭代更新速度极快,如今尚处于起步阶段的应用领域,短短数年后便有望发展成熟。因此重要的是掌握分析 AI 应用的通用思维框架:从所需数据、配套算力、核心算法、解决的实际问题以及现存局限五大维度,去剖析各类智能应用场景。这套思考方式,远比零散的案例知识更具备长期实用价值。

这也预示着未来行业人才需求的新方向:优质 AI 从业者不仅要精通专业技术,更要深耕所属行业,善于发掘实际工作中的痛点难点,并搭配合适的 AI 技术落地解决问题。对于正在规划学业与职业方向的大学生而言,这一发展趋势值得深思与借鉴。

本节小结

1. AI 产业链分为基础层、技术层、应用层三大层级,基础层提供数据、芯片、云平台等底层资源,技术层搭建开发工具与智能能力,应用层面向各行各业落地产品服务,三层相互依托、协同运转。
2. 当下 AI 发展呈现三大主流趋势:由专用智能迈向通用智能、运算由云端下沉至终端边缘端、信息处理由单模态走向多模态融合。
3. AI 广泛落地医疗、教育、制造、交通、金融、农业等领域,数据充足、流程标准的行业应用更为成熟;高安全要求、强人际交互领域中,AI 仅作辅助使用,全面自主应用尚不成熟。

思考题

1. [识记] 请用自己的话解释 AI 产业链的三层结构,并为每一层举出一个具体的例子。
2. [理解] 有人说“AI 将很快取代医生和教师”,请结合本任务中关于 AI 在医疗和教育领域的应用现状,分析这一说法是否合理,并说明理由。
3. [应用] 请选择一个你熟悉或感兴趣的行业,分析该行业中有哪些环节可能受益于 AI 技术,你认为最有价值的 AI 应用场景是什么? 实施过程中可能遇到的最大挑战是什么?

思想引领

一、中国 AI 的发展

AI 的发展,核心依托数据、算力、算法三大核心要素的协同支撑。当下全球科技竞争日趋白热化,AI 早已不再局限于学术研究范畴,其核心技术的自主掌控,直接关系到国家发展与国家安全的整体格局。

近些年,我国 AI 产业实现了快速发展,创新成果层出不穷。华为昇腾 AI 芯片、百度文心大模型、大疆智能无人机、科大讯飞智能语音技术等一众成果,覆盖了 AI 产业链多个核心环节,充分彰显了国内企业与科研机构扎实的创新实力。

但也要客观正视现存短板,在高端 AI 芯片研发制造、原创基础算法等核心关键领域,我国仍存在技术瓶颈,面临诸多外部制约。

也正因如此,科技自立自强成为我国创新发展的核心导向。如今 AI 技术快速普及、跨界渗透,无论未来深耕哪个专业,当代大学生都大概率会与 AI 产生交集。主动了解 AI 的基础原理与发展逻辑,既能丰富自身知识体系、提升综合素养,更能为日后参与国家科技创新建设筑牢基础。

二、辩证地看待 AI 发展

纵观 AI 的发展历程,行业始终呈现出热潮与寒冬交替往复的波动特征。每一轮技术风口来临,都会掀起“AI 将颠覆一切”的舆论狂热;而当技术落地遇阻、热度褪去,又会滋生“AI 价值有限”的消极论调。如今,大语言模型带动新一轮 AI 浪潮席卷全球,各类观点充斥网络与社交媒体,有人担忧 AI 会快速替代大部分人类工作,也有人夸大风险,渲染超级智能、人类存续危机等极端说法。

面对纷繁杂乱的网络言论和两极化观点,在 AI 全面渗透生活、学习与工作的当下,跳出极端舆论、摒弃跟风心态,以理性、辩证的视角看待 AI 发展,是新时代大学生必备的数字素养与基本能力。

◆ 任务实践

实践一 “我身边的 AI”调研报告

任务说明

目标:识别和分析日常生活中的 AI 应用,建立对 AI 应用现状的感性认知。

步骤:

(1) 每位小组成员记录自己一天中接触到的所有可能涉及 AI 技术的产品或服务(至少 10 项)。

(2) 小组汇总后,选择其中 5 个最具代表性的应用进行深入分析。对每个应用,讨论以下问题:它属于哪个应用领域?它可能使用了哪种实现途径?它属于弱 AI 还是强 AI?它的数据来源是什么?

(3) 撰写一份不超过 2 000 字的调研报告,附上分析表格。

评价维度:应用识别的全面性与代表性、AI 技术分析的准确性、分类与判断的合理性、调研报告的逻辑性与清晰度。

实践二 AI 发展史的关键时刻

任务说明

目标:深入理解 AI 发展历程中的关键事件,锻炼信息整合和口头表达能力。

步骤:

(1) 小组从以下事件中选择一个(或自选一个本模块提及的事件)进行深入研究:
a. 1956 年达特茅斯会议;b. 1969 年《感知器》出版与第一次 AI 寒冬;c. 1997 年深蓝击败卡斯帕罗夫;d. 2012 年 AlexNet 在 ImageNet 竞赛中的突破;e. 2016 年 AlphaGo 击败李世石;f. 2022 年 ChatGPT 发布。

(2) 收集该事件的背景资料,理解其技术内涵和历史意义。

(3) 准备一次 5—8 分钟的课堂演讲,回答以下问题:这个事件发生的背景是什么?它为什么重要?它对后续的 AI 发展产生了什么影响?它给到读者什么启示?

(4) 演讲结束后接受其他同学 2—3 分钟的提问。

评价维度:内容准确性、分析深度、表达清晰度、回答提问的应变能力。

实践三 “AI 能做什么,不能做什么”辩论赛

任务说明

目标:培养对 AI 能力边界的理性认知和批判性思维能力。

步骤:

(1) 全班分为正方和反方。辩题:“以当前技术的发展速度,强 AI(AGI)将在 20 年内实现。”

(2) 各方利用本模块学到的知识组织论据。鼓励课外查阅相关资料,但所有论据必须有事实或逻辑支撑,不得使用未经验证的说法。

(3) 辩论流程建议:正方一辩陈述(3 分钟)→反方一辩陈述(3 分钟)→自由辩论(10 分钟)→正方总结(2 分钟)→反方总结(2 分钟)。

(4) 辩论结束后,教师引导全班讨论:双方的论据中,哪些最有说服力?这场辩论是否改变了你原有的看法?

评价维度:论点与论据的科学性、逻辑论证的严密性、对 AI 能力边界的理解深度、批判性思维与反思能力。

实践四 绘制一张行业 AI 应用地图

任务说明

目标:将 AI 应用分析框架运用到具体行业场景中,培养“AI+行业”的跨学科思维。

步骤:

(1) 每个小组选择一个与组员专业背景相关的行业。优先选择本模块任务四中未详细讨论的行业。

(2) 通过网络调研和小组讨论,梳理该行业的主要业务环节,识别其中 3—5 个可能受益于 AI 技术的环节。

(3) 对每个环节,分析:a. 该环节目前的痛点是什么? b. AI 可以如何帮助解决? c. 需要什么类型的数据? d. 对算力的要求如何? e. 可能采用什么类型的算法? f. 实施过程中最大的挑战是什么?

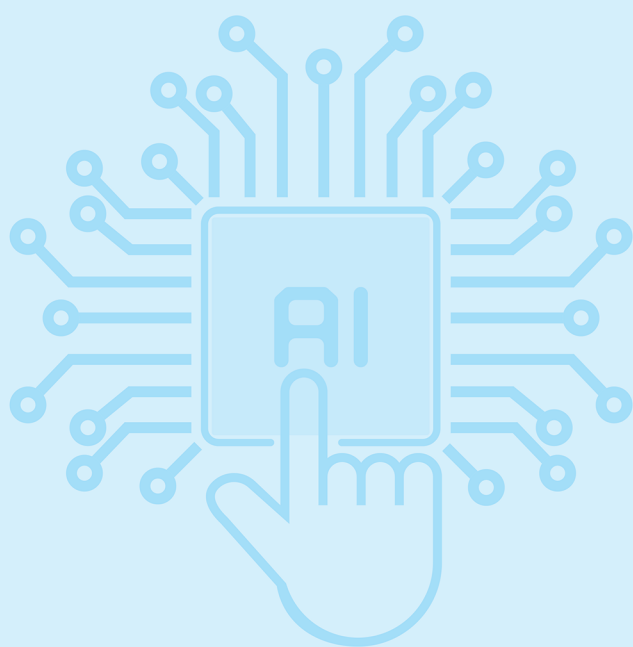
(4) 将分析成果整理为一张可视化的流程图、思维导图或信息图,配以不超过 1 500 字的说明文字。

评价维度:行业理解深度、AI 技术匹配合理性、分析框架的运用能力、可视化表达的清晰度。

南京大学出版社
版权所有

» 模块二

机器学习



南京大学出版社
版权所有

南京大学出版社
版权所有

◆ 模块导读

从模块一中可以了解到“让计算机从数据中自动学习规律”是实现人工智能的核心途径之一,并且知道了机器学习是驱动当今 AI 爆发式发展的关键方法论。那么,机器学习究竟是如何“学习”的?它和人类的学习有什么异同?

当提及“训练一个模型”时,这个“训练”过程具体包含哪些步骤?为什么同样的数据,用不同的算法会得到截然不同的结果?监督学习、无监督学习、强化学习这些听起来颇为专业的术语,分别对应着怎样的学习方式?它们各自适合解决什么类型的问题?

本模块将围绕以上核心问题展开,通过五个任务带领读者系统理解机器学习。任务一从宏观视角介绍机器学习的基本概念、典型应用及其与人工智能的关系;任务二深入解析基于学习方式的四种分类;任务三梳理从数据准备到模型评估的完整工作流程;任务四以直觉化的方式介绍几种经典的机器学习算法;任务五展示机器学习在图像识别、自然语言处理、语音识别、推荐系统等领域的实际应用。通过本模块的学习,读者将建立起对机器学习技术的整体认知,理解“数据驱动”这一核心范式,并能够对日常生活中的机器学习应用进行初步分析与判断。

◆ 学习目标

【了解】

- 机器学习的基本概念与定义,理解机器学习与传统编程方式的本质区别;
- 监督学习、无监督学习、半监督学习和强化学习四种学习方式的基本含义;
- 线性回归、逻辑回归、决策树、支持向量机、聚类算法、集成学习等经典算法的名称和基本用途。

【理解】

- 机器学习的工作流程,即从数据收集、预处理到模型训练、评估、优化的完整过程;
- 四种学习方式各自的适用场景、优势与局限,以及它们之间的区别和联系;
- 经典机器学习算法的核心思想(直觉层面),以及不同算法适用于不同类型问题的原因。

【掌握】

- 针对给定的实际问题场景,判断应当采用哪种学习方式和哪类算法;
- 运用机器学习的基本工作流程框架,分析一个 AI 应用案例中数据、模型、评估各环节的作用。

【能够】

- 结合日常生活中的具体场景(如推荐系统、语音助手、图像搜索等),识别其中运用

的机器学习技术,并对其工作原理进行初步分析;

——对机器学习技术的能力与局限性形成独立判断,能够批判性地评价“AI 预测”“AI 推荐”等应用的可靠性。

◆ 思维导图

下图展示了本模块的知识结构与内容框架,帮助读者从整体上把握各任务之间的逻辑关系:

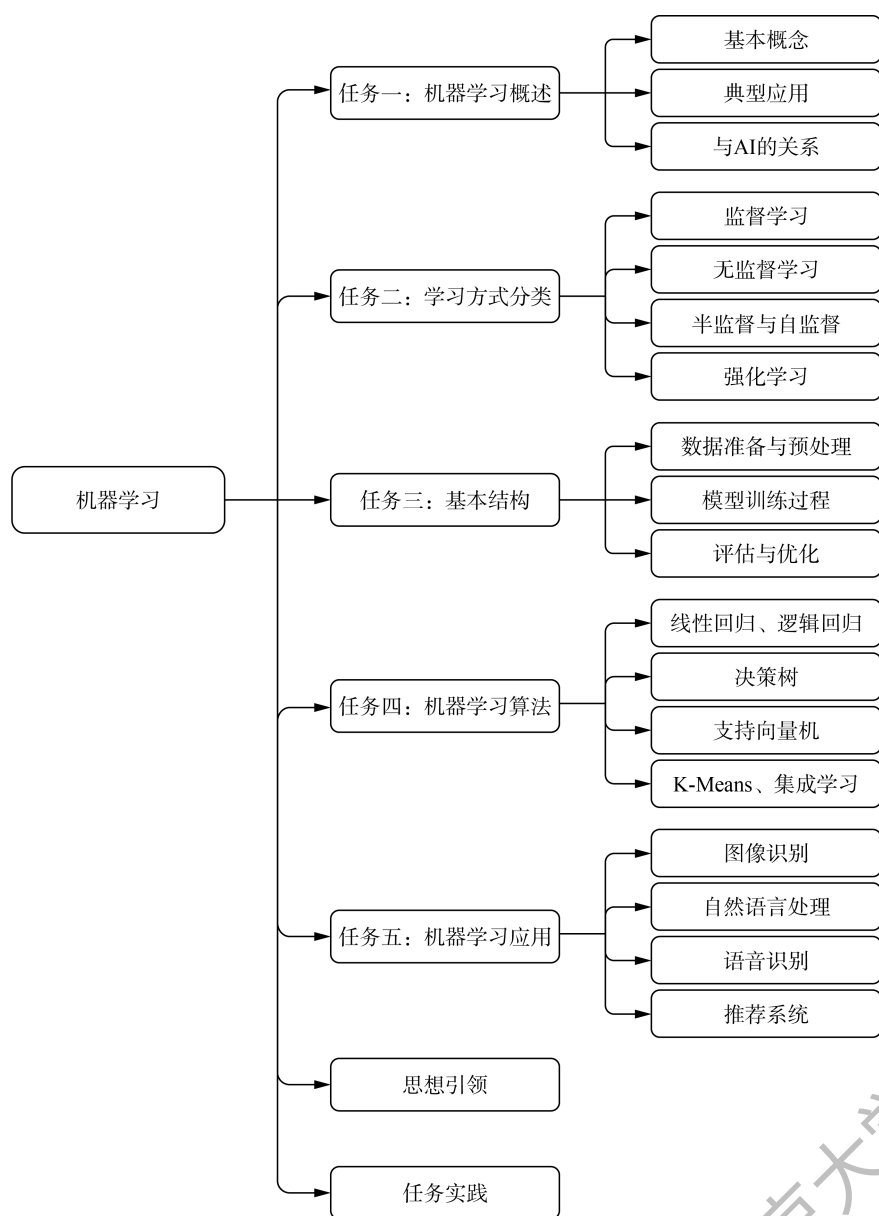


图 2-1 模块二知识结构思维导图

任务一 机器学习概述

◆ 引导思考

你有没有注意到,购物网站似乎越来越“懂你”——它推荐的商品经常正中你的心意。邮箱里的垃圾邮件也像被一只“看不见的手”自动过滤掉了。这些功能的背后,到底是怎样的技术在起作用?

在模块一中,了解到机器学习是实现人工智能最重要的途径之一。那么,机器学习和人类的“学习”有什么相同和不同?计算机是怎样从数据中“学”到东西的?

机器学习与人工智能是什么关系?它们是同一个概念的不同说法,还是有着本质的区别?

一、什么是机器学习

1. 从“编写规则”到“从数据学习”

在模块一的任务一中,已经初步了解了机器学习的基本思想:让计算机从数据中自动学习规律,而不是由程序员逐条编写规则。

传统的计算机程序是“规则驱动”的。程序员需要预先想好解决问题的每一个步骤,然后用代码把这些步骤写出来。计算机严格按照这些预设的步骤执行,不会多走一步,也不会少走一步。这种方式对于规则明确的问题非常有效——比如计算工资、排序数据、查找文件。但当面对规则难以穷举的复杂问题时,传统方式就力不从心了。

举一个最直观的例子:如何让计算机判断一封邮件是不是垃圾邮件?如果用传统方式,可能会写一些规则:“如果邮件标题包含‘中奖’或‘免费领取’,就判定为垃圾邮件。”但垃圾邮件的制造者很快就会变换花样——把“中奖”写成“zhong jiang”,把“免费”写成“免·费”。这样每天都要更新规则,疲于奔命,永远追不上对方的变化。

机器学习的思路完全不同。它不要求告诉计算机“什么是垃圾邮件”的具体规则,而是给计算机“看”大量已经标注好的邮件样本(“这封是垃圾邮件”,“这封不是”),让计算机自己从这些样本中“摸索”出区分垃圾邮件的规律。经过充分的“学习”之后,当一封新邮件到来时,计算机就能根据它学到的规律自动作出判断——即使遇到了全新的伪装手法,它也往往能正确识别,因为它学到的是更深层次的“模式”,而不是死板的关键词规则。

用一个生活化的比喻来总结:传统编程就像“给你一本菜谱,严格按步骤做菜”;机器学习则像“带你去餐厅吃一百道菜,让你自己悟出什么是好菜”。前者依赖“明确的指令”,后者依赖“足够的经验”。

2. 机器学习的定义

关于机器学习,学术界有一个被广泛引用的经典定义,由美国计算机科学家汤姆·米切尔提出:“如果一个计算机程序在某类任务 T 上的性能 P ,随着经验 E 的积累而不断提高,那么就说这个程序对任务 T 进行了学习。”

这个定义中包含三个关键要素:任务(T)——希望计算机完成什么?性能度量(P)——怎么衡量它做得好不好?经验(E)——它从什么数据中学习?用垃圾邮件过滤的例子来对应这三个要素:任务 T 是“判断邮件是否为垃圾邮件”;性能度量 P 是“判断的准确率”;经验 E 是“大量已标注的邮件样本”。当系统处理的邮件样本越多(经验积累),它的判断准确率(性能)就越高——这就是“学习”。

案例：从“死记硬背”到“举一反三”——机器学习的学习方式

为了更好地理解机器学习,可以拿它和人类学生的学习做一个类比。

假设老师让学生学会区分猫和狗的图片。一种方式是“死记硬背”:老师给学生看 100 张图片,学生记住每一张图片的答案。考试时,如果出现的是这 100 张中的某一张,学生能答对;但如果出现一张全新的图片,学生就没辙了。这种方式不能称为真正的“学习”,因为它没有概括出规律。

另一种方式是“举一反三”:学生在看了 100 张图片后,自己归纳出了一些规律——比如“猫的耳朵通常是尖的,狗的耳朵通常是垂的”“猫的脸型偏圆,狗的脸型偏长”。有了这些规律,即使遇到一张从未见过的新图片,学生也能大概率判断正确。

机器学习追求的正是第二种“学习”——不是记住具体的样本,而是从样本中提取出可以泛化到新数据的规律。在机器学习术语中,这种“在新数据上也能表现良好”的能力叫作泛化能力,它是衡量一个机器学习模型好坏的最重要标准之一。

3. 机器学习与人类学习的异同

机器学习和人类学习有一些有趣的相似之处,但也存在本质的差异,理解这些差异有助于正确地认识机器学习的能力边界。

表 2-1 机器学习与人类学习的对比

维度	人类学习	机器学习
学习素材	多感官、多渠道(看、听、触、嗅)	数字化数据(数字、文字、图像像素等)
学习数量	少量样本即可概括规律	通常需要大量数据
学习速度	学习单个概念快,但整体知识积累慢	处理海量数据快,但“理解”深度有限
迁移能力	强:学会骑自行车有助于学摩托车	弱:图像识别模型无法自动做语音识别
创造力	能产生全新的想法和解决方案	只能在训练数据的“范围”内生成输出
遗忘	会遗忘,但遗忘也有筛选功能	不遗忘已学内容,但也不会自动更新

从表 2-1 可以看出,机器学习在处理海量数据方面远超人类,但在迁移能力和创造力方面与人类相比仍有很大差距。这也正是为什么当前的机器学习属于“弱人工智能”的范畴(详见模块一任务一):它在特定任务上可以做到极为出色,但无法像人类那样灵活地将一个领域的知识迁移到完全不同的领域。

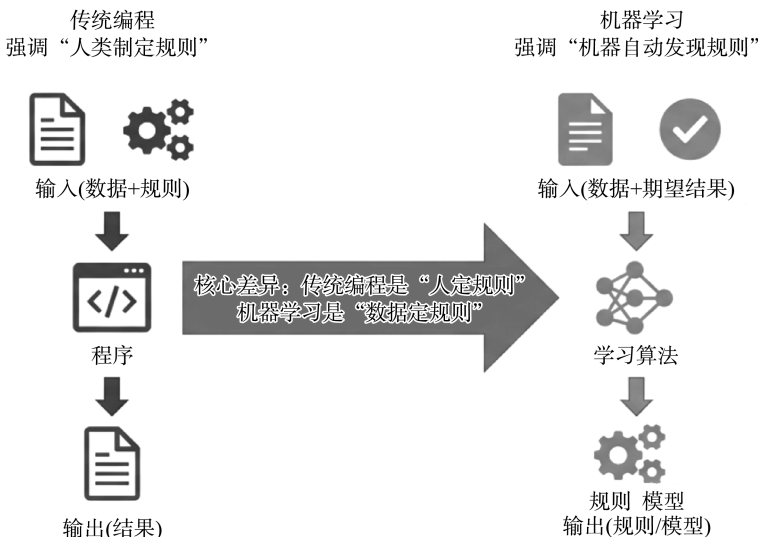


图 2-2 传统编程与机器学习的对比

二、机器学习的典型应用

机器学习并不是一个遥远的学术概念,它已经深入渗透到日常生活的方方面面。了解这些典型应用,有助于建立对机器学习“能做什么”的具体认知。

1. 推荐系统

打开购物 App,首页展示的商品仿佛“量身定制”;刷短视频时,平台总能契合浏览喜好——这些都是推荐系统在发挥作用。推荐系统的核心思路是:通过分析用户过去的行为数据(浏览了什么、购买了什么、点赞了什么),机器学习模型“学”出用户的偏好模式,然后据此预测用户接下来可能感兴趣的内容。

推荐系统是当今互联网产品最核心的技术之一。据行业估算,电商平台上相当大比例的交易额是由推荐算法驱动的;短视频平台的用户停留时长也在很大程度上取决于推荐算法的精准度。可以说,用户在互联网上看到的大部分内容,都已经经过了机器学习模型的“筛选”和“排序”。

2. 语音识别与智能助手

对手机下达“帮我设一个明天早上 7 点的闹钟”指令时,语音助手需要完成两个关键步骤:先将语音转化为文字(语音识别);然后,理解这段文字的含义并执行相应的操作(自然语言理解)。这两个步骤都大量依赖机器学习技术。现代的语音识别系统经过数万小

时的语音数据训练,在安静环境下的准确率已经相当高,使得语音交互成为了一种日常化的人机沟通方式。

3. 图像识别

机器学习在图像识别领域的应用同样无处不在。手机的人脸解锁、相册的自动分类、社交媒体上的滤镜效果,以及更专业的医学影像分析(如从CT片中检测肿瘤)、工业视觉质检(如检测产品表面瑕疵),都建立在机器学习驱动的图像识别技术之上。在模块一的任务二中提到的2012年ImageNet竞赛,正是图像识别领域机器学习(具体而言是深度学习)全面爆发的起点。

4. 自然语言处理

自然语言处理是让计算机理解和生成人类语言的技术。机器翻译(如将中文翻译成英文)、文本情感分析(如判断一条产品评论是好评还是差评)、智能写作(如辅助撰写邮件和报告),以及模块一中详细讨论过的大语言模型(如ChatGPT),都属于自然语言处理的范畴。这些应用的背后,是机器学习模型通过海量文本数据“学”到了语言的结构和规律。

5. 其他应用

除了上述几个领域,机器学习还广泛应用于金融风险评估(预测贷款违约概率)、精准医疗(根据患者数据推荐个性化治疗方案)、智能制造(预测设备故障)、自动驾驶(感知周围环境并作出驾驶决策)、天气预报(提高中短期天气预测的准确率)等众多领域。可以说,在任何“有大量数据可供学习且需要从中发现规律”的场景中,机器学习都有用武之地。

生活中的机器学习: 用户可能没注意到的“隐形应用”

以下这些日常功能的背后,都有机器学习在默默工作:

手机输入法的联想词:根据用户的打字习惯和当前上下文,预测用户接下来要输入的词。

邮箱的垃圾邮件过滤:自动识别并拦截垃圾邮件,准确率通常高于99%。

社交媒体的信息流排序:决定用户看到哪些朋友的动态排在前面。

打车软件的动态定价:根据供需关系和交通状况实时调整价格。

银行卡的异常交易提醒:监测消费模式,发现异常时立即预警。

这些应用的共同特点是:它们都在“悄无声息”地工作,用户享受着它们带来的便利,却几乎不会意识到它们的存在。

三、机器学习与人工智能的关系

在日常交流中,“人工智能”和“机器学习”这两个词经常被混用,很多人以为它们是同一个东西的不同叫法。实际上,这两个概念之间存在着明确的层次关系。模块一任务一中,曾以“套娃”比喻初步阐释该关系,下文将对此进行系统化梳理。

1. 包含关系: AI>ML>DL

人工智能(AI)是最大的概念范畴,它涵盖了所有让机器表现出智能行为的技术和方法。机器学习(Machine Learning, ML)是实现人工智能的一种途径——目前也是最成功、最主流的途径。深度学习(Deep Learning, DL)则是机器学习中的一个特定分支,它使用深层神经网络来处理数据。三者的关系可以表示为:深度学习属于机器学习,机器学习属于人工智能。如图 2-3 所示。

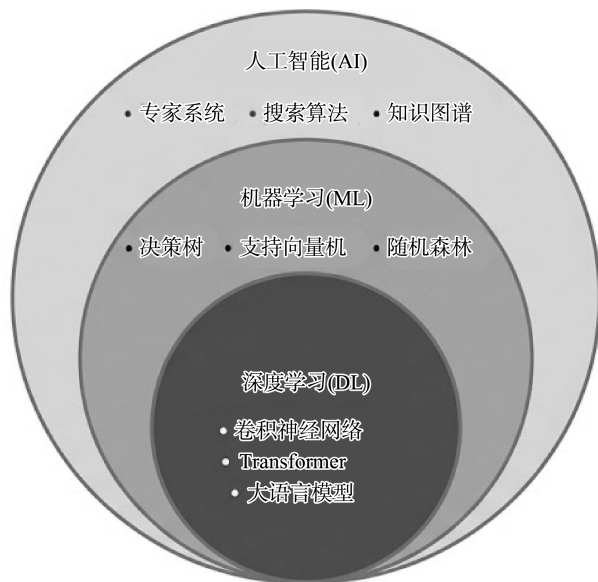


图 2-3 人工智能、机器学习与深度学习的包含关系

2. 为什么机器学习成为了 AI 的“主流路线”

在模块一中可了解到,人工智能有多种实现途径:基于规则的方法、基于统计学习的方法、基于深度学习的方法、基于大模型的方法。那么,为什么机器学习(涵盖后三种途径)逐渐成为了 AI 的主流?

核心原因在于:现实世界中大多数有价值的问题(如理解语言、识别图像、预测行为),其内在规律极其复杂,人类专家很难用明确的规则完整地表达。而机器学习恰恰不需要人类预先定义规则——它直接从数据中“发现”规律。再加上互联网时代海量数据的积累和 GPU 算力的飞速提升(模块一任务三的“三大引擎”),机器学习的优势被进一步放大,成为当今 AI 应用背后最核心的技术驱动力。

3. 并非所有 AI 都是机器学习

理解了包含关系后,一个重要的推论是:并非所有的人工智能应用都基于机器学习。模块一中介绍的专家系统(基于手工编写的规则进行推理)就是一种不使用机器学习的 AI 技术。在某些规则明确、变化不大的场景中(如简单的流程自动化、基于固定逻辑的决策树),传统的基于规则的方法仍然有效,不需要“大动干戈”地引入机器学习。

选择使用机器学习还是传统方法,关键取决于问题的性质:如果规则可以被人类专家清晰地定义,且规则数量有限,传统方法可能更简单高效;如果规则难以穷举,或者规律隐藏在大量数据中需要被“发现”,那么机器学习就是更好的选择。这种“根据问题选方法”的思维方式,是理解 AI 技术应用的一个重要框架。

知识拓展:“没有免费的午餐”定理

机器学习领域有一个著名的理论结论,叫作“没有免费的午餐”定理。它的核心含义是:不存在一种“万能”的机器学习算法,能在所有类型的问题上都表现最好。每种算法都有自己擅长的问题类型和不擅长的问题类型。

这个定理的名字非常形象——它告诉人们:在机器学习中没有“免费的午餐”,用户不可能什么都不付出(不了解问题的特点)就获得最好的结果。要想让机器学习发挥最大价值,必须深入理解用户要解决的问题,然后选择最合适的算法和方法。

这个道理对读者的学习也有启发:不要迷信任何一种“万能方法”,关键是理解问题、选对工具。编者将在本模块的任务四中介绍几种经典的机器学习算法,届时读者可以体会到,不同的算法确实适合不同类型的问题。

本节小结

1. 机器学习是一种让计算机从数据中自动学习规律的技术,其核心区别于传统编程:传统编程由人类制定规则,机器学习由数据“发现”规则。
2. 机器学习的经典定义包含三个关键要素:任务(T)、性能度量(P)和经验(E)。当程序在任务上的性能随经验积累而提高时,就说它在“学习”。
3. 学习的核心能力是泛化——即在从未见过的新数据上也能作出正确判断,而不仅仅是“记住”训练数据。
4. 机器学习已广泛应用于推荐系统、语音识别、图像识别、自然语言处理等众多领域。
5. 人工智能、机器学习和深度学习是三个层次递进的概念($AI > ML > DL$)。机器学习是当今 AI 最主流的实现途径,但并非所有 AI 都基于机器学习。

思考题

1. [识记] 请用一个自己的生活场景为例,向一位没有技术背景的朋友解释“什么是机器学习”,以及它与传统的计算机程序有什么不同。
2. [理解] 有人说“机器学习就是让计算机记住大量数据”,请结合本任务关于“泛化能力”的讨论,分析这种说法是否准确,并说明理由。
3. [应用] 假设某学校的图书馆想开发一个“智能图书推荐系统”,根据学生的借阅历史推荐可能感兴趣的书籍。请用米切尔的机器学习定义框架(任务 T 、性能度量 P 、经验 E)来描述这个系统,并思考:这个系统可能面临哪些挑战?(提示:可以从数据质量、冷启动问题等角度思考)

任务二 基于学习方式的分类

◆ 引导思考

在任务一中可了解到,机器学习的核心是“从数据中学习规律”。但“学习”的方式并不只有一种——就像人类既可以在老师的指导下学习,也可以通过自己的探索和尝试来学习。

机器学习领域根据“数据是否带有标准答案”这一关键区分,划分出了几种截然不同的学习方式。它们各自有什么特点?分别适合解决什么类型的问题?

如果遇到一个实际问题,如何判断应该用哪种学习方式?掌握这一判断能力,是理解机器学习的关键一步。

任务一中以“从数据中自动学习规律”概括机器学习的核心思想。但这只回答了“是什么”的问题,还没有回答“怎么学”的问题。就像人类的学习有“课堂听讲”“自习探索”“实践试错”等不同形式一样,机器学习也有多种不同的学习方式。根据训练数据中是否包含“标准答案”(即标签),以及学习的目标和反馈机制的不同,机器学习主要分为四种学习方式:监督学习、无监督学习、半监督学习和强化学习。

一、监督学习

1. 什么是监督学习

监督学习是机器学习中最直观、应用最广泛的一种范式。它的核心特点是:训练数据中每一个样本都附带了“标准答案”(在机器学习术语中叫作标签)。模型的任务就是从数据和标签的对应关系中学习规律,以便对新的、没有标签的数据作出预测。

用一个最贴近学生生活的比喻:监督学习就像“有老师批改作业”的学习方式。教师提供习题数据,并配套标准答案。学生做完一道题,就可以对照答案检查自己做得对不对,从错误中学习。经过大量的练习之后,学生就掌握了解题的规律,面对新的题目(新数据)也能给出正确的答案——即使这道新题从没见过。

2. 两大核心任务:分类与回归

监督学习主要解决两类问题:分类和回归。

分类是将数据分配到预定义的类别中。例如:判断一封邮件是“垃圾邮件”还是“正常邮件”(两个类别);识别一张照片中的动物是“猫”“狗”还是“鸟”(多个类别);根据病理切片判断组织是“良性”还是“恶性”。分类问题的标签是离散的类别,模型的输出是“属于哪个类”的判断。

回归是预测一个连续的数值。例如：根据房屋面积、地段、楼层等特征预测房价（输出是一个具体的价格数字）；根据历史销售数据预测下个月的销量；根据天气数据预测明天的最高气温。回归问题的标签是连续的数值，模型的输出是“具体是多少”的预测。

区分分类和回归的关键在于：如果所要预测的是一个“类别”（是 A 还是 B），就是分类问题；如果所要预测的是一个“数值”（具体是多少），就是回归问题。

3. 监督学习的典型应用

监督学习的应用几乎无处不在。任务一中提到的垃圾邮件过滤（分类）、图像识别（分类）、语音识别（分类）都属于监督学习的范畴；房价预测（回归）、股票价格预测（回归）、天气预报（回归）也是典型的监督学习任务。可以说，当知道“要预测什么”并且“有历史数据和对应的答案”时，监督学习通常是首选方案。

在中国的互联网产业中，监督学习有着极为广泛的应用。电商平台利用监督学习模型预测用户的购买意向（分类：会买还是不会买？），从而精准推送广告；外卖平台利用回归模型预测每一单的配送时长，为用户提供更准确的等待时间估计；金融机构利用分类模型评估贷款申请人的信用风险（分类：违约还是不违约？），辅助信贷审批决策。

案例：监督学习如何帮助医生诊断疾病

假设训练一个 AI 模型，辅助医生判断一张胸部 X 光片是否存在肺炎。这是一个典型的监督学习任务：

数据：数千张胸部 X 光片。

标签：每张 X 光片都由经验丰富的医生标注为“正常”或“肺炎”。

学习过程：模型反复“看”这些标注好的 X 光片，逐渐学会从图像的纹理、阴影、结构等特征中识别肺炎的迹象。

预测：当一张新的、未标注的 X 光片输入模型时，模型根据学到的规律给出判断：“正常”或“肺炎”，并附上一个置信度（如“肺炎的概率为 87%”）。

需要强调的是，这个模型的质量高度依赖于标签的质量——如果医生的标注有大量错误，模型也会“学错”。这再次印证了在模块一任务三中强调的：数据质量是 AI 系统的“生命线”。

二、无监督学习

1. 什么是无监督学习

与监督学习相对，无监督学习的训练数据中没有标签——也就是说，数据只有“题目”，没有“标准答案”。模型的任务是自己从数据中发现隐藏的结构或模式。

继续用学习的比喻：无监督学习就像“没有老师指导的自学”。没人告诉学生哪些知识点是重点，也没有标准答案可以参照。学生只能自己阅读大量材料，尝试发现其中的规律和结构——比如“这几个概念似乎有关联”“这些内容可以归为一类”。这种学习方式更

自由、更开放,但也更具挑战性。

2. 核心任务:聚类与降维

无监督学习最常见的任务是聚类和降维。

聚类是将相似的数据自动分成若干组。例如:电商平台将消费者按购买行为自动分成不同的“客户群”(如“价格敏感型”、“品质追求型”、“冲动消费型”),以便实施差异化营销;社交媒体分析中将大量帖子按话题自动归类;生物学中将基因表达相似的样本自动分组。聚类与分类的关键区别是:分类有预定义的类别(事先知道要分成哪几类),聚类没有(让算法自己决定分几组以及怎么分)。

降维是将高维的数据压缩到低维空间,同时尽量保留有用的信息。打个比方:一份简历上有姓名、年龄、学历、工作经验、技能等数十个字段(高维),但招聘经理在初筛时可能只关注“学历+经验”这两个维度(低维)就能作出基本判断。降维的本质是从海量原始特征里提取少数代表性特征,以此更高效地理解与处理数据。

3. 无监督学习的典型应用

无监督学习在数据探索和用户洞察方面有着独特的价值。客户分群(帮助企业理解不同类型的客户)、异常检测(在没有“异常标签”的情况下发现数据中的异常模式,如检测信用卡盗刷)、数据可视化(将高维数据降维后进行可视化展示)都是无监督学习的典型应用场景。

举一个具体的例子:一家音乐流媒体平台拥有数千万首歌曲,但不可能让人工为每首歌都标注风格类型。无监督学习的聚类算法可以根据歌曲的音频特征(节奏、音调、配器等)自动将歌曲分成若干组——算法可能“发现”了“舒缓钢琴曲”、“重金属摇滚”、“电子舞曲”等聚类,尽管没有人预先定义这些类别。这些聚类结果可以被推荐系统利用:当用户喜欢某一首歌时,系统推荐同一聚类中的其他歌曲。

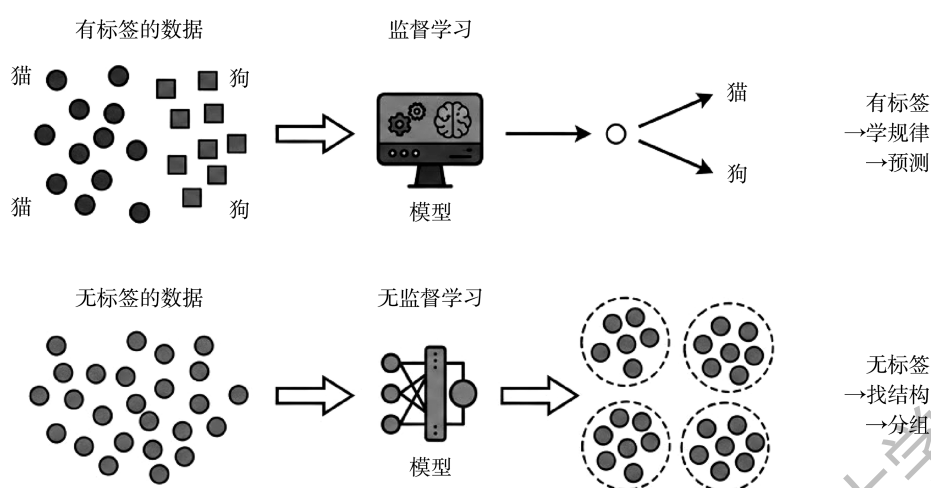


图 2-4 监督学习与无监督学习的对比

三、半监督学习

在实际应用中会经常面临这样的困境:收集大量的原始数据并不难(比如从互联网上下载百万张图片),但给这些数据逐一标注标签却非常昂贵和耗时(让人类专家逐张标注“这是猫”“这是狗”)。有没有一种折中的方案,既能利用少量标注数据的“指导性”,又能利用大量未标注数据的“信息量”?

半监督学习正是为了解决这个问题而生的。它的训练数据中同时包含少量有标签的数据和大量没有标签的数据。模型先从少量有标签数据中学到初步的规律,然后利用大量无标签数据来完善和修正这些规律。

用一个生活化的比喻:半监督学习就像一个学生先跟着老师学了几道“例题”(有标签数据),掌握了基本方法;然后自己做了大量的“练习题”(无标签数据),虽然没有标准答案,但通过反复尝试和自我验证,进一步巩固和深化了对知识的理解。

半监督学习在医学影像分析(大量影像数据但只有少量经专家标注)、自然语言处理(海量文本但标注成本高)等场景中具有重要的实用价值。近年来大语言模型的预训练过程(先在海量无标签文本上学习语言模式,再用少量标注数据进行微调)也体现了半监督学习的思想。

四、强化学习

1. 什么是强化学习

前面三种学习方式(监督、无监督、半监督)都是从一批“静态的”数据集中学习规律。但现实世界中有一大类问题不是这样的——智能体需要在一个动态的环境中不断作出决策,并通过环境给出的反馈(奖励或惩罚)来学习最优的行为策略。这就是强化学习的领域。

强化学习的核心概念包括三个要素:智能体(做决策的主体)、环境(智能体所处的外部世界)和奖励信号(环境对智能体行为的评价)。智能体在每一步都要选择一个动作,环境会根据这个动作给出一个奖励(正面的或负面的),智能体的目标是学会一套行为策略,使得长期累积的总奖励最大化。

用一个最直观的比喻:强化学习就像“训练一只小狗”。主人(环境)不会告诉小狗(智能体)每一步该怎么做,但当它做对了(比如听到“坐下”指令后真的坐下了),主人会给它一块零食(奖励);当它做错了(比如随地乱跑),主人会表示不满(惩罚)。小狗通过反复尝试,逐渐“摸索”出哪些行为能获得奖励,哪些会招致惩罚,从而学会了正确的行为方式。

2. 强化学习的独特之处

强化学习与监督学习有一个根本性的区别:在监督学习中,模型在每一步都能得到“标准答案”(这道题的正确答案是什么);但在强化学习中,智能体得到的只是一个“好坏

信号”(这一步做得好不好),而且这个信号往往是延迟的——某个动作的好坏可能要等到很久之后才能看出来。例如,下棋时一步“看似随意”的走法可能在 30 步之后带来决定性的优势。强化学习必须学会在“即时收益”和“长远利益”之间进行平衡,这使得它在解决序列决策问题方面具有独特的优势。

3. 强化学习的典型应用

强化学习最著名的成就是在游戏和棋类领域。模块一任务二中有详细讨论过的 AlphaGo 就是强化学习的典范之作——它通过与自己对弈数百万局(试错和积累经验),学会了连人类棋手都难以想象的围棋策略。此后,DeepMind 的 AlphaZero 更进一步,仅凭游戏规则(不依赖任何人类棋谱),就通过纯强化学习达到了超越人类的水平。

除了游戏和棋类,强化学习在自动驾驶(在复杂交通环境中作出驾驶决策)、机器人控制(让机器人学会行走、抓取等动作)、资源调度(如数据中心的能耗优化)、个性化推荐(将推荐过程建模为与用户的持续互动)等领域也展现出了巨大的潜力。

案例：AlphaGo 的“自我对弈”——强化学习的经典示范

2016 年 AlphaGo 击败围棋世界冠军李世石的故事(详见模块一任务二),堪称强化学习最戏剧性的展示。从强化学习的视角重新审视这个案例:

智能体:AlphaGo 程序。

环境:围棋棋盘和对弈规则。

动作:在棋盘上落子(每一步有数百个可能的位置)。

奖励信号:赢棋获得正奖励,输棋获得负奖励。

AlphaGo 的训练过程分为两个阶段:首先,通过监督学习从大量人类棋谱中学习基本的下棋模式;然后,通过强化学习让两个 AlphaGo 副本相互对弈数百万局,在不断“试错”中超越人类水平。特别值得注意的是,AlphaGo 在与李世石的对局中下出了若干“人类棋手从未想到”的走法,这说明强化学习能够发现人类经验之外的全新策略。

表 2-2 四种学习方式的对比

学习方式	数据特点	通俗比喻	核心任务	典型应用
监督学习	有标签(有标准答案)	有老师批改作业	分类、回归	垃圾邮件过滤、房价预测、医学影像
无监督学习	无标签(无标准答案)	自学找规律	聚类、降维	客户分群、异常检测、数据可视化
半监督学习	少量有标签 + 大量无标签	例题 + 大量练习	结合两者优势	医学影像(少量标注)、NLP 预训练
强化学习	环境反馈(奖惩信号)	训练小狗(奖罚)	序列决策	游戏 AI、自动驾驶、机器人控制

从表 2-2 可以清晰地看到,四种学习方式各有所长,适用于不同类型的问题。在实际的 AI 项目中,选择哪种学习方式取决于两个关键因素:一是“有什么样的数据”(有标签还

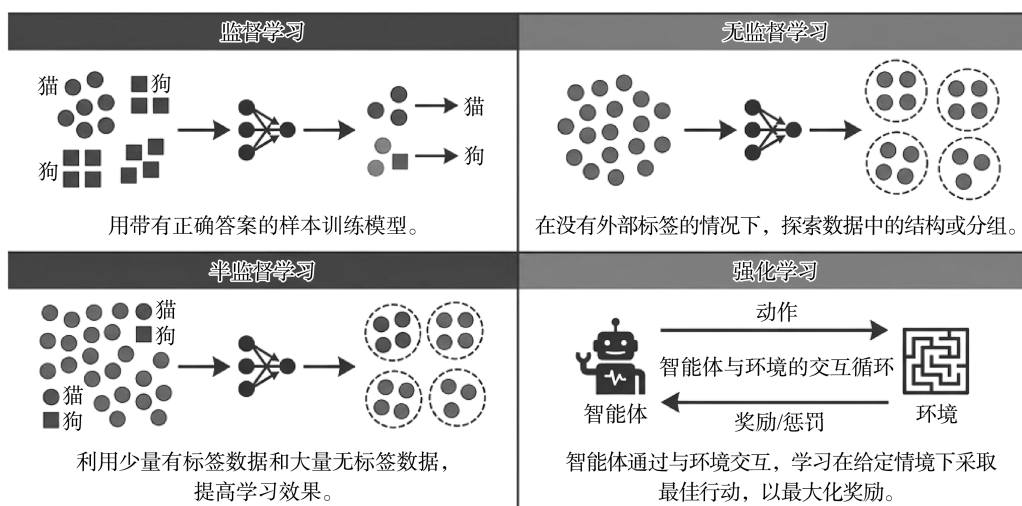


图 2-5 四种学习方式的概念示意图

是没有？数据量大不大？标注成本高不高？），二是“要解决什么问题”（预测类别？预测数值？发现分组？作出决策？）。理解这一选择逻辑，是从“理解机器学习”到“运用机器学习思维”的关键一步。

值得一提的是，在实际的 AI 系统中，不同的学习方式常常被组合使用。例如，一个智能推荐系统可能同时使用无监督学习进行用户聚类（发现不同的用户群体），使用监督学习进行点击率预测（预测用户是否会点击某个推荐），再使用强化学习优化整体的推荐策略（在用户长期满意度和即时点击率之间寻找平衡）。这种“混合使用”在工业界非常普遍，在后续学习中会看到更多的例子。

知识拓展：自监督学习——大模型背后的“秘密武器”

近年来，随着大语言模型的兴起，一种被称为“自监督学习”的方法引起了广泛关注。自监督学习可以看作无监督学习的一种巧妙变体：它从无标签数据中自动“构造”出标签。

以大语言模型的预训练为例：模型的训练任务是“根据前面的文字预测下一个词”。在这个任务中，“下一个词”就是自动从文本中获取的“标签”——不需要任何人工标注。通过在海量文本上反复执行这个“预测下一个词”的任务，模型逐渐“学会”了语言的语法规则、词汇含义、常识知识甚至一定的推理能力。

自监督学习的出现极大地缓解了标注数据匮乏的问题，是推动大模型技术爆发的关键方法论创新之一。从学习方式的分类角度来看，它模糊了监督学习和无监督学习的传统边界，代表了机器学习方法不断融合与创新的趋势。

本节小结

1. 机器学习根据数据标签情况和学习机制,主要分为监督学习、无监督学习、半监督学习和强化学习四种方式。
2. 监督学习使用有标签的数据,主要解决分类(预测类别)和回归(预测数值)两类问题,是应用最广泛的学习方式。
3. 无监督学习使用无标签的数据,主要任务是聚类(自动分组)和降维(特征压缩),擅长从数据中发现隐藏的结构和模式。
4. 半监督学习结合少量有标签数据和大量无标签数据,在标注成本高昂时具有重要的实用价值。
5. 强化学习通过智能体与环境的交互来学习最优策略,适用于需要序列决策的场景。

思考题

1. [识记] 请分别用一句话概括监督学习、无监督学习和强化学习的核心特点,并为每一种学习方式举一个任务一中提到的应用实例。
2. [理解] 有人说“监督学习比无监督学习更好,因为有标准答案就能学得更准”。请分析这种说法的合理性和局限性。(提示:考虑标注成本、适用问题类型、各自的独特价值)
3. [应用] 假设你是一家连锁奶茶店的数据分析师,手上有以下数据:过去一年的销售记录(每天的销量、温度、是否周末、促销活动等)、会员的消费行为数据(购买频次、偏好口味、消费金额)。请思考:a. 如果你想预测明天的销量,应该用哪种学习方式?这是分类问题还是回归问题? b. 如果你想将会员分成不同的客户群以便实施差异化营销,应该用哪种学习方式?为什么?

任务三 机器学习的基本结构

◆ 引导思考

在任务一中了解了机器学习“是什么”,在任务二中了解了“有哪些学习方式”。那么,一个完整的机器学习项目从头到尾是怎样运作的?从拿到原始数据到最终交付一个可用的模型,中间要经历哪些步骤?

人们常说“数据决定了 AI 的上限,算法只是逼近这个上限”。这句话背后的含义是什么?数据准备工作在整个机器学习流程中到底有多重要?

当一个模型训练完成后,怎么知道它“够好”还是“不够好”?有哪些方法可以进一步提升模型的表现?

前两个任务分别回答了“什么是机器学习”和“有哪些学习方式”的问题。本任务将回答第三个关键问题：一个完整的机器学习项目是怎样从头到尾运作的？可以将机器学习的工作流程拆解为三个主要阶段：数据收集与预处理、模型选择与训练、模型评估与优化。理解这个“全流程框架”，就像拿到了一张机器学习项目的“施工蓝图”——即使不亲自编程，也能看懂一个 AI 项目的关键环节和决策点。

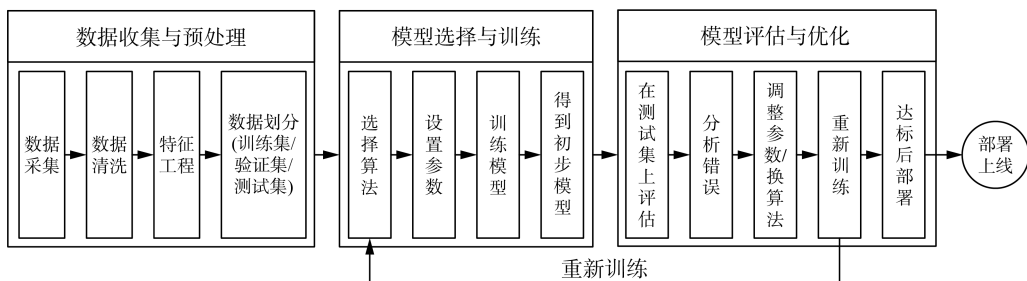


图 2-6 机器学习的完整工作流程

一、数据收集与预处理

数据是机器学习的“原料”，该内容已在模块一任务三中展开深入探讨。在一个实际的机器学习项目中，数据相关的工作往往占据了整个项目周期的大部分时间。业界有一个广为引用的说法：“80%的时间花在数据上，20%的时间花在模型上。”虽然这个比例不是精确的统计结论，但它生动地反映了数据工作在机器学习中的核心地位。

1. 数据采集

数据采集是整个流程的起点。数据的来源多种多样：可以是企业内部的业务数据库（如电商平台的交易记录、医院的病历档案），可以通过传感器、摄像头等设备实时采集的物联网数据，也可以是从互联网上爬取的公开数据，还可以使用已有的公开数据集（如在前面多次提到的 ImageNet）。

数据采集阶段需要关注两个关键问题：一是数据量是否足够——在任务一中了解到，机器学习（尤其是深度学习）通常需要大量的训练数据；二是数据是否具有代表性——如果采集的数据只覆盖了部分场景（例如只收集了晴天的自动驾驶数据），模型在其他场景（如雨天、雪天）上的表现就会大打折扣。

2. 数据清洗

现实中的原始数据几乎不可能直接用于模型训练，它往往充满了各种“瑕疵”：缺失值（某些记录的部分字段为空）、异常值（明显不合理的数据，如年龄为一5岁）、重复数据（同一条记录被录入多次）、格式不一致（同一个城市名有“北京”“Beijing”“BJ”三种写法）。如果不加处理就直接送入模型，这些“脏数据”会严重影响学习效果。

数据清洗就是系统地识别和修复这些问题的过程。用一个生活化的比方：做菜之前，需要先洗菜、择菜、去掉烂叶子——这个步骤看似不起眼，但如果跳过它，做出来的菜质量

肯定好不了。数据清洗在机器学习中扮演着完全相同的角色。

3. 特征工程

清洗完的数据还不能直接交给模型,还需要经过一个叫作特征工程的步骤。所谓特征,就是用来描述数据的“属性”或“维度”。比如,要预测一套房子的价格,可能用到的特征包括:面积、楼层、地段、建造年份、周边学校数量等。

特征工程的核心任务是:从原始数据中提取出对预测目标最有帮助的特征。有时候,原始数据中的信息需要经过“加工”才能成为有效的特征。例如,电商平台可能记录了每位用户每次登录的时间戳,但直接使用时间戳并没有太大意义;将其加工成“过去7天的登录次数”“最近一次登录距今的天数”等特征,对预测用户是否会流失就有价值得多。

在传统的机器学习方法中,特征工程的质量往往决定了模型性能的上限。不过,如模块一中所述,深度学习的一个重大优势正是“自动特征学习”——深层神经网络可以从原始数据中自动提取有用的特征,大大减少了人工特征工程的工作量。

4. 数据划分

在正式开始训练之前,还有一个至关重要的步骤:将数据划分为训练集、验证集和测试集三个部分。

为什么要这样划分?回顾任务一中探讨的泛化能力——一个好的模型应该在从未见过的新数据上也能表现良好,而不仅仅是“记住”了训练数据。训练集用于模型的学习;验证集用于在训练过程中监控模型的表现,帮助调整参数;测试集则“密封”到最后,专门用来评估模型的最终泛化能力。这就像考试制度:训练集是课堂练习(用来学习),验证集是模拟考试(用来检测学习效果并调整策略),测试集是正式考试(用来评判最终水平)——如果“正式考试”的题目提前泄露(测试集在训练中被使用),考试成绩就失去了参考价值。

二、模型选择与训练

1. 选择合适的算法

数据准备好之后,下一步是选择一个合适的机器学习算法(也叫“模型”)来进行训练。在任务二中了解了四种学习方式(监督、无监督、半监督、强化),首先需要根据问题的性质确定使用哪种方式;然后,在该方式下选择具体的算法。例如,如果确定是一个监督学习的分类问题,可以选择决策树、支持向量机、神经网络等不同的算法(这些算法将在任务四中详细介绍)。

选择算法时需要综合考虑多个因素:数据的规模(数据量大时深度学习更有优势,数据量小时传统方法可能更合适)、问题的复杂度(简单问题用简单模型,复杂问题用复杂模型)、可解释性要求(某些应用场景如医疗和金融要求模型的决策过程“可解释”,这时候决策树可能比深度神经网络更合适)。任务一中提到的“没有免费的午餐”定理表明:没有哪个算法在所有问题上都是最优的,关键是找到最匹配当前问题的那一个。

2. 训练过程

训练是整个机器学习流程中最核心的环节。通俗地说,训练就是让模型“看”大量的训练数据,通过不断调整内部参数,使其输出越来越接近正确答案的过程。

在任务一中以“做菜”来比喻训练:第一次做出来味道不对,调整配料比例;第二次颜色不好,调整火候;反复尝试直到满意。机器学习的训练遵循类似的循环:模型给出一个预测→与正确答案对比,计算“差多远”(损失)→根据差距调整参数→再预测→再对比……如此循环数千甚至数百万次,模型的预测就越来越准确了。

案例:训练一个图像分类模型——“看图学习”的过程

假设要训练一个模型来区分猫和狗的照片。训练过程大致如下:

第1轮:模型看到一张猫的照片,猜测“这是狗”(错了!)。系统计算损失(猜错的程度),并据此调整模型内部参数,让它“下次更倾向于把这类图案判断为猫”。

第2轮:模型看到一张狗的照片,猜测“这是狗”(对了!)。损失很小,参数只做微调。

第3轮:模型看到另一张猫的照片,这次猜对了。损失更小,参数继续微调。

.....

第10 000轮:经过大量数据的反复“学习”,模型已经学会了从图像中提取猫和狗的区分特征(如耳朵形状、面部轮廓等),对新图片的判断准确率达到了95%以上。

整个过程的关键在于:模型不是一次性“学会”的,而是通过成千上万次的“预测—纠错—调整”循环,逐渐积累起识别能力。

选读:什么是“损失”?

(本框内容为选读,不影响后续学习)

在训练过程中,“损失”(Loss)是衡量模型预测值与真实值之间差距的数值指标。损失越小,说明模型的预测越接近正确答案。

用考试来类比:如果满分是100分,某考生得了85分,那么其“损失”就是15分。训练的目标就是让这个“损失”尽可能地小。

定义损失的数学公式叫作“损失函数”,不同类型的问题使用不同的损失函数。例如,回归问题常用“均方误差”(预测值与真实值之差的平方的平均值),分类问题常用“交叉熵”(衡量预测的概率分布与真实分布之间的差异)。

读者不需要记住这些公式的具体形式,只需要理解核心思想:损失函数是训练过程的“指南针”——它告诉模型“你离目标还有多远”,指引模型朝着“更准确”的方向前进。

三、模型评估与优化

1. 为什么需要评估

模型训练完成后,一个自然的问题是:这个模型“够好”吗? 能不能直接投入使用? 直

觉上,似乎可以直接采用训练数据来测试模型的表现——但这就像让学生用复习过的原题来考试,即使考了满分也说明不了什么。在前面讨论数据划分时提到,真正有意义的评估必须使用模型“从未见过”的测试集数据。

2. 常见的评估指标

对于不同类型的机器学习任务,有不同的评估指标。

表 2-3 常见的模型评估指标

问题类型	常用指标	通俗含义
分类问题	准确率	所有预测中,猜对了多少?
	精确率	模型说“是”的那些里面,真正“是”的有多少?
	召回率	实际为“是”的那些里面,模型找出了多少?
回归问题	均方误差	预测值与真实值的平均偏差有多大?
	决定系数(R-squared)	模型解释了数据中多少比例的变化?

在实际应用中,不同的场景对不同指标的侧重是不一样的。例如,在癌症筛查中,召回率比准确率更重要——宁可“多查出一些疑似病例”(即使部分是误报),也不希望“漏掉真正的患者”;而在垃圾邮件过滤中,精确率可能更重要——不希望正常邮件被误判为垃圾邮件(否则可能错过重要邮件)。选择合适的评估指标,是机器学习实践中的一个重要判断。

3. 两大常见问题:过拟合与欠拟合

在模型评估中,有两个最常见的问题需要特别注意:过拟合和欠拟合。

过拟合是指模型在训练数据上表现非常好,但在新数据(测试集)上表现明显变差。这就像一个学生把考试的原题和答案全部背了下来(训练集上满分),但换一套新题就完全不会做(测试集上很差)。过拟合的本质是模型“记住”了训练数据中的细节和噪声,而没有真正学到可以泛化的规律。

欠拟合是指模型在训练数据上就表现不佳。这说明模型太“简单”,无法捕捉数据中的规律。用考试的比喻来说,就是这个学生连课堂练习都做不好——不是题太难,而是基础太薄弱。

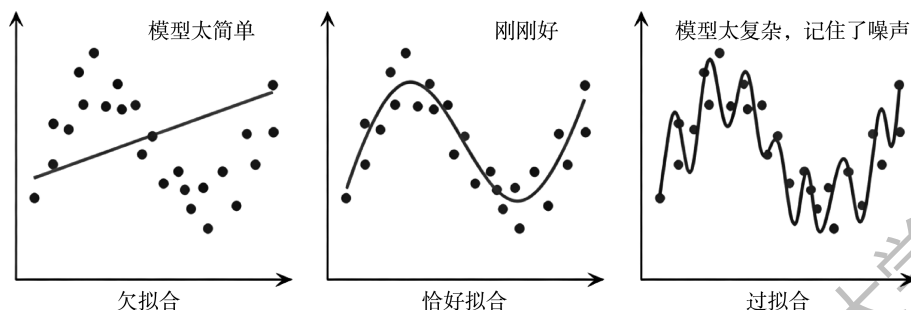


图 2-7 过拟合与欠拟合的直观理解

4. 优化策略

当模型评估不达标时,需要进行优化。常见的优化策略包括:增加或减少数据量(过拟合时可增加数据,让模型有更多的“练习题”)、调整模型的复杂度(欠拟合时用更复杂的模型,过拟合时用更简单的模型)、调整训练参数(如学习率等)、进行更细致的特征工程等。

模型优化通常是一个反复迭代的过程:训练→评估→发现问题→调整→重新训练→再评估……如此循环,直到模型的表现达到可接受的水平。如图 2-6 所示的流程中,第三阶段到第二阶段之间的循环箭头正体现了这种迭代特征。实际项目中,这个循环可能要进行数十甚至数百次。

机器学习项目的全过程包括数据采集、数据清洗、特征提取、算法选择、模型训练、模型评估、模型优化、模型部署等多个阶段,项目“全景视图”如图 2-8 所示。

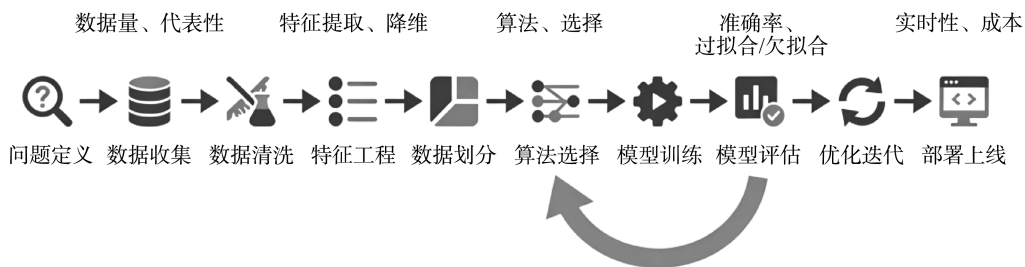


图 2-8 机器学习项目的“全景视图”

知识拓展：“奥卡姆剃刀”与模型选择

14 世纪的英国哲学家奥卡姆的威廉提出了一个著名的思维原则:“如无必要,勿增实体”——在能够同样好地解释事物的多种理论中,应该选择最简单的那个。这个原则后来被称为“奥卡姆剃刀”。

奥卡姆剃刀在机器学习中有着重要的指导意义。当两个模型在测试集上的表现差不多时,通常应该选择那个更简单的模型。原因在于,简单的模型更不容易过拟合(因为它没有足够的“容量”去记忆噪声),而且更容易被人类理解和维护。

这个原则提醒着读者:在机器学习中,“更复杂”不等于“更好”。选择恰当复杂度的模型,在拟合数据和保持泛化之间找到平衡,是机器学习实践中最核心的“艺术”之一。

本节小结

1. 一个完整的机器学习项目分为三个主要阶段:数据收集与预处理、模型选择与训练、模型评估与优化。
2. 数据阶段包括数据采集、数据清洗、特征工程和数据划分(训练集/验证集/测试

集),占据了项目的大部分时间。

3. 模型训练的核心是“预测—计算损失—调整参数”的反复循环,选择算法时需综合考虑数据规模、问题复杂度和可解释性等因素。

4. 模型评估需要使用独立的测试集,常见指标包括准确率、精确率、召回率等;过拟合和欠拟合是最常见的两大问题。

5. 模型优化是一个反复迭代的过程。“奥卡姆剃刀”原则:在同等效果下,优先选择更简单的模型。

思考题

1. [识记] 请按顺序列出机器学习项目的主要工作步骤,并用自己的话说明“数据划分”(训练集/验证集/测试集)的目的。

2. [理解] 在癌症筛查场景中,为什么“召回率”比“准确率”更重要? 如果反过来,在什么场景下“精确率”会更重要? 请各举一个例子。

3. [应用] 假设某教师训练了一个模型来预测大学生的期末考试成绩。模型在训练数据上的预测误差非常小(几乎完美匹配),但在新学期的学生数据上预测误差很大。根据本任务所学的知识:a. 这最可能是什么问题(过拟合还是欠拟合)? b. 可尝试采用哪些优化策略来改善模型?

任务四 机器学习算法

◆ 引导思考

在任务三中了解了机器学习项目的完整工作流程,其中“选择算法”是关键步骤之一。那么,机器学习领域有哪些常用的算法? 每种算法的核心思想是什么?

不同的算法就像工具箱里不同的工具——锤子适合钉钉子,螺丝刀适合拧螺丝。面对不同类型的问题,应该如何选择合适的“工具”?

需要提醒的是,本任务的目标是帮助读者建立对各类算法的“直觉理解”,而不是掌握算法的数学细节。读者关注的是“每种算法像什么”“能做什么”“适合什么场景”,而不是“它内部怎么算”。

在任务二中了解了监督学习、无监督学习等学习方式;在任务三学习了从数据准备到模型评估的完整流程。本任务将走进机器学习的“工具箱”,逐一介绍六种经典且常用的算法。每种算法均围绕三点展开阐述:它的核心思想是什么(直觉层面)? 它适合解决什么类型的问题? 它有什么优势和局限?

一、线性回归

线性回归是机器学习中最简单、最基础的算法,也是理解其他更复杂算法的起点。它属于监督学习,用于解决回归问题(即预测一个连续的数值)。

线性回归的核心思想可以用一句话概括:找到一条“最佳拟合直线”,使得这条线尽可能地贴近所有的数据点。用一个生活化的例子来说明:收集班级同学的“每周学习时长”和“期末考试成绩”的数据,画在一张图上,横轴是学习时长,纵轴是成绩。可以发现,数据点大致呈现出一个趋势:学习时间越长,成绩往往越高。线性回归的任务就是找到一条直线,最好地描述这个趋势。有了这条直线,当想知道一个新同学的学习时长时,就可以沿着直线“查”到他可能的成绩。

线性回归的优势在于简单直观、计算效率高、结果易于解释(可以直接看到“每多学习1小时,成绩平均提高多少分”)。它的局限也很明显:现实世界中的关系往往不是简单的直线(例如,学习时间超过一定量后,成绩的提升会放缓甚至停滞),线性回归在这些情况下就不够用了。

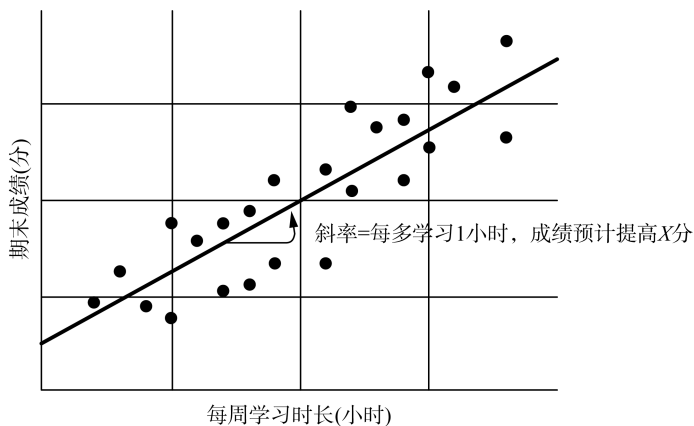


图 2-9 线性回归的直觉理解

二、逻辑回归

尽管名字里有“回归”两个字,逻辑回归实际上是一种分类算法。它属于监督学习,主要用于解决“是/否”类的二分类问题(如判断邮件是否为垃圾邮件、患者是否患有某种疾病、用户是否会购买某个商品)。

逻辑回归的核心思想是:在线性回归的基础上,增加一个“压缩”步骤,把原本可以取任意值的预测结果“压缩”到0和1之间,使其可以被解释为一个概率。比如,模型输出0.85,可以解释为“有85%的概率是垃圾邮件”;输出0.12,则意味着“只有12%的概率是垃圾邮件”。然后设定一个阈值(通常是0.5),高于阈值就判定为“是”,低于则判定为“否”。

打个比方:如果线性回归是一把“直尺”(用来量长度),那么逻辑回归就是在直尺上加

了一个“开关”——把测量值转化为“开/关”(是/否)的判断。这个小小的改变,使得逻辑回归成为了分类问题的有力工具。

逻辑回归在金融风控(判断贷款是否会违约)、医学诊断(判断检测结果是否阳性)、互联网广告(预测用户是否会点击广告)等领域应用极为广泛。它的优势是简单高效、结果可解释(可以看到每个因素对结果的影响方向和大小);局限是只能处理线性可分的问题,对于复杂的非线性关系需要更强大的模型。

三、决策树

决策树可能是所有机器学习算法中最直观、最容易理解的一种。它既可以用于分类,也可以用于回归,属于监督学习。

决策树的核心思想是:通过一系列的“提问”来逐步缩小答案的范围,最终作出判断。每一个“问题”对应树上的一个分叉节点,每一个可能的回答对应一条分支,而最终的“答案”(预测结果)对应树的叶子节点。

用一个极其贴近生活的例子:决定今天中午吃什么。决策过程可能是这样的——“想吃辣的吗?”→如果是,“想花超过 30 元吗?”→如果是,“去吃火锅”;如果不是,“去吃麻辣烫”→如果不想吃辣,“想吃面食吗?”……这种“一步步提问、层层筛选”的过程,就是一棵决策树。

机器学习中的决策树也是如此,只不过“问题”是关于数据特征的判断(如“年龄是否大于 30?”“收入是否高于 10 万?”),而“答案”是要预测的结果(如“会购买”或“不会购买”)。算法会自动从数据中找出“最有效的提问顺序”——即那些最能帮助区分不同结果的特征。

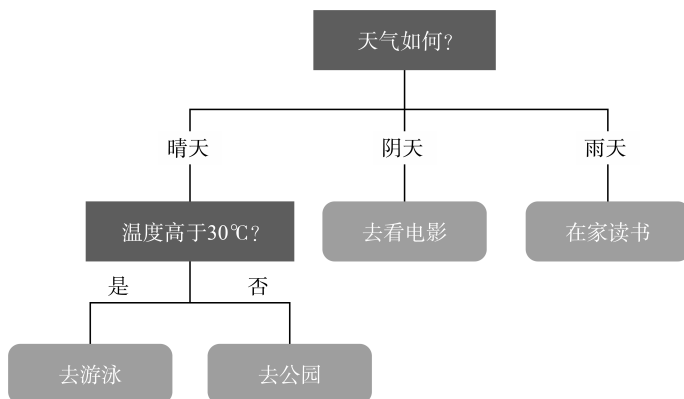


图 2-10 决策树的直觉理解——以“周末活动推荐”为例

决策树的最大优势是可解释性强:可以清楚地看到模型是基于哪些条件作出判断的,这在医疗、法律、金融等要求“决策透明”的领域尤为重要。它的主要局限是容易过拟合(任务三中讨论过的问题)——如果树长得太“深”太“茂盛”,就会把训练数据中的噪声也当成规律来学习。

四、支持向量机

支持向量机(Support Vector Machine, SVM)是一种功能强大的分类算法,属于监督学习。它的核心思想可以用一个形象的比喻来理解:在两类数据之间找到一条“最宽的分界线”。

想象一个操场上有两组学生——篮球队(红色)和足球队(蓝色)——随意站在各处。任务是在地上画一条线,把两组学生尽可能好地分开。直觉上,会希望这条线不仅能分开两组人,而且离两边最近的人都尽可能远——这样,即使有人稍微移动位置(新数据出现),这条线仍然有效。支持向量机做的就是这件事:找到那条“间隔最大”的分界线(在高维空间中,这条“线”变成了一个“超平面”)。

“支持向量”这个名字中的“支持”指的是那些离分界线最近的数据点——它们“支撑”着分界线的位置。只有这些“边界上的”关键点决定了分界线的位置,其他更远处的点对结果没有影响。这也是 SVM 的一个有趣特性:它的判断只依赖于少量的关键样本。

SVM 的一个重要扩展是使用核技巧:当两类数据在原始空间中无法用一条直线分开时(比如红蓝两色的点混在一起,呈环形分布),SVM 可以通过“核技巧”将数据映射到更高维的空间,在高维空间中找到一个可以分开它们的超平面。用一个比喻来说:就像在一张纸上画了两个嵌套的圆环(红色和蓝色),在纸面上(二维)无论怎么画直线都无法分开它们;但如果把纸“捏”起来变成三维的形状,就有可能用一个平面把它们分开了。

SVM 在文本分类、手写数字识别、生物信息学等领域有着优秀的表现,尤其适合中小规模、高维度的数据集。它的局限在于当数据量非常大时计算效率较低,而且结果的可解释性不如决策树。

生活中的机器学习算法：它们其实就在身边

读者可能没想到,本任务介绍的这些“听起来很学术”的算法,其实就潜伏在读者每天使用的产品背后:

线性回归:外卖平台可预测订单送达时间,背后的基础模型很可能包含线性回归的成分。

逻辑回归:消费者每次刷信用卡时,银行的反欺诈系统都在用类似逻辑回归的模型实时判断这笔交易是否异常。

决策树:保险公司评估客户的保费等级时,使用的风险评估模型可能就是一棵(或一组)决策树。

支持向量机:手写输入法将用户在触摸屏上的笔迹识别为文字,SVM 曾是这一领域的主力算法。

K-Means 聚类:音乐 App 的“每日推荐歌单”可能先用聚类算法将歌曲按风格分组,再从用户喜欢的组里挑选歌曲。

集成学习:电商平台判断一条商品评论是真实评价还是“刷单”水军,背后往往是一个集成学习模型。

五、聚类算法

前面介绍的四种算法都属于监督学习。接下来介绍一种无监督学习的代表性算法——聚类算法。在任务二中了解到,聚类的目标是将相似的数据自动分成若干组,而不需要事先定义“应该分成哪几类”。

最经典的聚类算法叫作 K 均值聚类(K-Means)。它的核心思想非常直观:首先随机选择 K 个“中心点”,然后将每个数据点分配到离它最近的中心点所在的组;接着,重新计算每组数据的“中心”(取平均值);然后再重新分配……如此反复迭代,直到分组结果不再变化。

用一个生活化的比喻:假设一个城市要新建 3 个快递分拣中心($K=3$),需要决定它们的位置,使得每个分拣中心到它负责区域内的快递点的总距离最短。K-Means 的做法就是:先随便选 3 个位置(初始中心),把每个快递点分配给最近的分拣中心;然后把每个分拣中心移到它所负责的快递点的“重心”位置;再重新分配……不断调整,直到找到最优的布局方案。

K-Means 的优势是简单高效,容易理解;局限在于需要事先指定 K 值(即分几组),而且对“球形”分布的数据效果最好,对形状不规则的数据聚类效果可能不理想。

案例:用聚类分析发现“隐藏的客户画像”

一家连锁超市拥有数十万会员的消费数据(购买金额、购买频次、最近一次消费时间等),但从未对会员进行分类。运营团队决定用 K-Means 聚类来发现不同类型的客户群体。

经过分析,算法将客户自动分为了四组:

第一组:“高频高消费”——经常来,每次花费较多(VIP 客户)。

第二组:“低频高消费”——偶尔来,但每次花费很多(节日/囤货型客户)。

第三组:“高频低消费”——经常来,但每次只买少量商品(日常便利型客户)。

第四组:“沉睡客户”——很久没来,消费也很少(流失风险客户)。

这些分组在数据分析之前并不存在,是算法从数据中“发现”的。运营团队据此制定了差异化的营销策略:对 VIP 客户推送高端新品,对沉睡客户发送唤醒优惠券,对日常便利型客户推荐捆绑套餐。这就是无监督学习“从数据中发现隐藏结构”的典型价值。

六、集成学习

在日常生活中有一个朴素的经验:遇到重要决定时,多听几个人的意见往往比只听一个人的更靠谱。集成学习正是这一智慧在机器学习中的体现——它不是一种单独的算法,而是一种“组合多个模型来提升整体效果”的策略。

集成学习的核心思想是：三个臭皮匠，顶个诸葛亮。单个模型可能在某些数据上判断失误，但如果同时训练多个不同的模型，让它们各自独立作出预测，然后通过“投票”（分类问题）或“取平均”（回归问题）来综合它们的结果，整体的准确率往往会显著提高。

1. 随机森林

随机森林是集成学习最经典的代表。它的思路很简单：同时训练很多棵决策树（通常是几十到几百棵），每棵树在训练时随机选择一部分数据和一部分特征，因此每棵树都略有不同。预测时，让所有的树分别“投票”，最终采用“多数票”作为结果。

随机森林为什么有效？因为“随机性”带来了“多样性”。每棵树因为看到的数据和特征不同，会犯不同类型的错误；当这些“犯不同错的树”一起投票时，个别树的错误往往会被其他树纠正，整体的准确率和稳定性远高于单棵决策树。这就像班级选课代表：如果只问一个同学的意见，可能带有个人偏见；但如果全班投票，结果通常更能反映多数人的真实意愿。

2. 梯度提升

另一种重要的集成学习方法是梯度提升（如 XGBoost、LightGBM 等）。与随机森林“同时训练很多树”不同，梯度提升采用“逐步改进”的策略：先训练一棵简单的树，找出它犯错的地方；再训练第二棵树，专门“补救”第一棵树犯的 error；然后训练第三棵树，“补救”前两棵树仍然存在的 error……依次叠加，每一步都在前一步的基础上进行改进。

梯度提升在各种数据科学竞赛和工业应用中都表现极为出色，被誉为“表格数据的王者”（即当数据以表格形式组织时，梯度提升往往是最强的传统机器学习方法）。在互联网公司的推荐系统、广告点击率预测、金融风控等核心业务中，梯度提升模型的应用极为普遍。

用修改论文来比喻梯度提升的过程：第一个审稿人（第一棵树）指出了论文的几个主要问题；根据意见修改后（减少了误差），请第二个审稿人（第二棵树）看，他指出了第一个审稿人没注意到的细节问题；再次修改，请第三个审稿人看……每一轮修改都针对前一轮

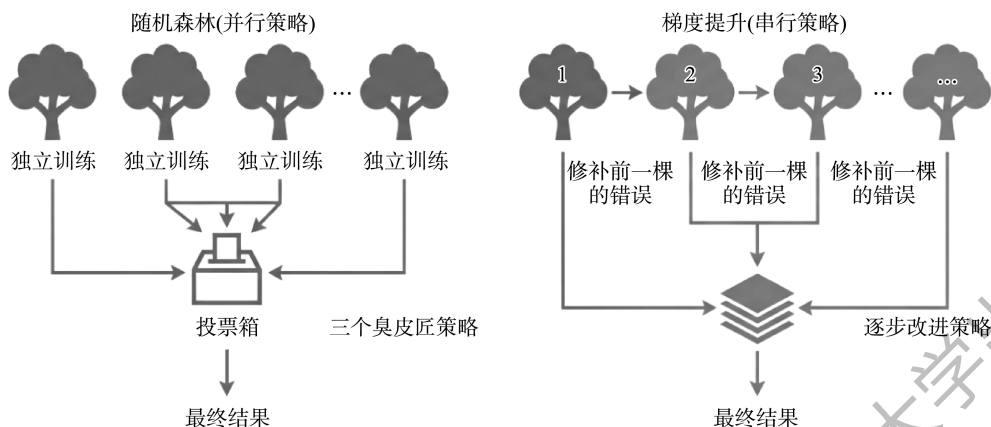


图 2-11 随机森林 vs 梯度提升两种策略对比

遗留的问题,论文(模型)的质量因此越来越高。这种“逐步修补”的策略,使得梯度提升在实践中往往能达到非常高的预测精度。

表 2-4 六种经典机器学习算法对比

算法	类别	核心比喻	适用问题	主要优势	主要局限
线性回归	监督/回归	画一条最合适的直线	预测连续数值	简单、高效、可解释	只能处理线性关系
逻辑回归	监督/分类	直尺+开关	二分类问题	简单、概率输出、可解释	只能处理线性可分
决策树	监督/分类+回归	逐步提问,层层筛选	分类和回归	最强可解释性	容易过拟合
支持向量机	监督/分类	找最宽的分界线	分类(尤其高维)	中小数据上表现优秀	大数据时计算慢
K-Means 聚类	无监督/聚类	找最优分簇中心	数据自动分组	简单直观	需指定 K,对形状敏感
集成学习	监督/分类+回归	三个臭皮匠	提升任何单模型	准确率高、稳定	复杂度高、可解释性弱

知识拓展：从传统算法到深度学习——何时需要“升级”？

本任务介绍的六种算法都属于“传统机器学习”的范畴。那么,在什么情况下需要“升级”到深度学习？

一般来说,当数据是结构化的表格数据(如用户行为记录、金融交易数据)时,传统机器学习算法(尤其是集成学习)往往已经足够好,甚至可能优于深度学习。但当数据是非结构化的(如图像、音频、自然语言文本)时,深度学习的自动特征提取能力就展现出巨大的优势。

此外,数据量也是一个关键因素。当数据量较小(几百到几千条)时,传统算法通常更稳定;当数据量很大(几十万到几百万条以上)时,深度学习的性能优势才会充分体现。

记住任务一中“没有免费的午餐”定理的启示:没有哪种算法在所有场景下都是最好的。选对工具,比追逐最新最炫的技术更重要。关于深度学习和神经网络的详细内容,将在模块三中系统介绍。

本节小结

- 1. 线性回归是最基础的回归算法,通过拟合直线预测连续数值;逻辑回归是在线性回归基础上增加了概率转换的分类算法。
- 2. 决策树通过“逐步提问”的方式作出判断,可解释性最强;支持向量机通过寻找“最宽分界线”来实现分类,在高维数据上表现优秀。
- 3. K-Means 是最经典的聚类算法,属于无监督学习,通过迭代寻找最优聚类中心将

数据自动分组。

4. 集成学习(随机森林、梯度提升)通过组合多个模型来提升整体效果,是工业界应用最广泛的传统机器学习方法之一。

5. 不同算法适用于不同的问题场景,选择合适的算法比追求最复杂的算法更重要。结构化表格数据适合传统机器学习算法,非结构化数据(图像/文本/音频)更适合深度学习。

思考题

1. [识记] 请从本任务介绍的六种算法中选择三种,分别用一句话概括它们的核心思想,并为每种算法举一个实际应用场景。

2. [理解] 有人说“集成学习既然总是比单个模型好,那为什么不永远使用集成学习?”请结合本任务关于各算法优劣势的讨论,分析这种说法可能忽略了哪些因素。(提示:考虑可解释性、计算成本、部署复杂度等)

3. [应用] 假设你在一家视频网站工作,面临以下三个任务:a. 预测一部新上线电影的播放量(数值预测);b. 判断一条用户评论是好评还是差评(分类);c. 将网站的数百万用户按观影偏好自动分成若干群体(无标签分组)。请分别判断每个任务应该使用哪种学习方式和哪种(或哪类)算法,并简要说明理由。

任务五 机器学习的应用

◆ 引导思考

在前面四个任务中,我们系统地学习了机器学习的概念、学习方式、工作流程和经典算法。那么,这些知识在现实世界中是如何被运用的?

当我们用手机拍照、和语音助手对话、刷短视频时,背后的机器学习系统是怎样工作的?它用的是哪种学习方式?可能用了哪些算法?经历了怎样的数据处理和模型训练流程?

带着本模块所学的“分析工具”——学习方式(任务二)、工作流程(任务三)、算法选择(任务四)——来“拆解”几个最重要的机器学习应用领域。

在模块一的任务四中,已从宏观产业视角扫描了 AI 在各行业的应用。本任务将换一个角度:从机器学习的技术视角出发,聚焦四个最具代表性的应用领域——图像识别、自然语言处理、语音识别和推荐系统,分析每个领域中机器学习是如何发挥作用的。任务目标不是面面俱到地介绍每个领域的技术细节,而是帮助读者学会用本模块构建的“分析框架”来理解任何一个机器学习应用。

一、图像识别

图像识别是让计算机“看懂”图片内容的技术。在模块一中多次提到过 2012 年 ImageNet 竞赛的标志性事件——深度学习在这次竞赛中的突破,正式开启了机器学习在图像领域的黄金时代。

1. 机器学习是如何“看懂”图片的

对于人类来说,看一张照片并说出“这是一只猫”几乎是瞬间完成的,毫不费力。但对于计算机来说,一张照片只是一堆数字——每个像素由一组数值(代表红、绿、蓝三种颜色的强度)表示。计算机需要从这些“毫无意义”的数字中,学会识别出有意义的物体。

这个过程恰恰是监督学习(任务二)的经典应用:准备大量已标注的图片(“这张是猫”“这张是狗”),使用深度学习模型(如卷积神经网络,将在模块三中详细介绍)进行训练,经过任务三所述的数据准备、模型训练、评估优化的完整流程后,模型就能对新的图片作出准确的分类判断。

2. 典型应用场景

图像识别技术已经深入到日常生活的方方面面。手机的人脸解锁利用图像识别技术判断摄像头前的人是否是机主;手机相册的“按人物分类”功能利用人脸识别将同一个人的照片自动归组;医学影像分析利用图像分类技术辅助医生从 X 光片、CT 片中检测病变区域;工业质检利用目标检测技术在生产线上自动识别产品表面的划痕、裂纹等瑕疵。

中国企业在图像识别领域有着深厚的技术积累。旷视科技、商汤科技、海康威视等企业在人脸识别、视频分析、智能安防等场景中积累了大量落地经验。例如,在校园门禁、支付验证、交通违章识别等场景中,图像识别技术已经实现了大规模的商业化部署。

二、自然语言处理

自然语言处理(Natural Language Processing, NLP)是让计算机理解和生成人类语言的技术。如果说图像识别让机器学会了“看”,那么自然语言处理就是让机器学会了“读”和“写”。

1. 机器学习是如何“理解”语言的

人类语言的复杂性远超想象。同一句话在不同的语境中可能有完全不同的含义(“这个杀手不太冷”既是电影名,也可以是对天气的感叹);语言中充满了隐喻、反讽、省略和歧义。让机器“理解”语言,长期以来被认为是 AI 最困难的挑战之一。

早期的自然语言处理主要依赖基于规则的方法(模块一中介绍过的符号主义路线),但效果有限。机器学习,尤其是深度学习的引入,带来了根本性的改变。现代 NLP 系统通过在海量文本数据上进行训练,让模型自己“学会”语言的语法规则、词汇含义、语义关

系和上下文推理能力。任务二中介绍的“自监督学习”(如“预测下一个词”)正是当前大语言模型训练的核心方法。

2. 典型应用场景

自然语言处理的应用涵盖了从日常工具到前沿研究的广阔范围:机器翻译(如将中文翻译为英文,早期的规则方法已被基于深度学习的神经机器翻译全面替代)、情感分析(分析大量用户评论中的正面和负面情绪,帮助企业了解产品口碑)、文本摘要(自动从长文章中提取核心信息)、智能客服(理解用户的自然语言提问并给出回答),以及大语言模型的各种生成任务(写作、编程、分析等)。

案例:机器翻译的进化——从“伏特加是好的”到“信达雅”

在模块一的任务一中,提到过早期基于规则的机器翻译闹出的笑话(“The spirit is willing, but the flesh is weak”被翻译成“伏特加是好的,但肉烂了”)。这个笑话生动地展示了规则方法在面对语言复杂性时的无力。

现代的神经机器翻译系统采用了完全不同的思路:它不再试图编写语法规则,而是通过监督学习,在数百万对“源语言—目标语言”的平行语料(即同一段内容的不同语言版本)上进行训练,让模型自己“学会”如何翻译。

这种方法的效果远超传统方法。截至本书编写时,主流的机器翻译系统在常见语种之间的翻译质量已经相当流畅,虽然在文学翻译、专业术语和文化隐喻等方面仍有不足,但对于日常交流和信息获取来说已经足够实用。这是机器学习“从数据中学习规律”这一核心思想的又一次有力验证。

三、语音识别

语音识别是将人类的语音信号转换为文字的技术。它让机器学会了“听”——当对手机说“帮我查一下明天的天气”时,语音识别系统在毫秒级的时间内就将语音转写成了文字。

1. 机器学习是如何“听懂”语音的

语音信号本质上是一段随时间变化的声波。语音识别系统需要完成的任务是:从这段连续的声波中,“切分”出一个个音节或音素,再将它们“组合”成词语和句子。这个过程面临着巨大的挑战:不同的人说同一个词,发音可能差异很大(口音、语速、音调);环境中的噪声也会干扰信号。

早期的语音识别主要使用统计学习方法(如隐马尔可夫模型),准确率有限。深度学习的引入带来了质的飞跃——深层神经网络能够从大量的语音数据中自动学习声学特征和语言模式,使得语音识别的准确率在安静环境下已经非常高。这是一个典型的监督学习应用:训练数据是大量的“语音+对应文字”配对,模型从中学习“什么声音对应什么文字”的映射关系。

2. 典型应用场景

语音识别技术支撑着人们日常使用的大量产品：手机语音助手（Siri、小爱同学等）、语音输入法（说话代替打字）、会议转录（将会议录音自动转写为文字记录）、智能家居控制（“小爱同学，关灯”）、车载语音交互（“导航到最近的加油站”）等。中国的科大讯飞在中文语音识别领域处于全球领先水平，其语音转写技术已广泛应用于教育、司法、医疗等行业。

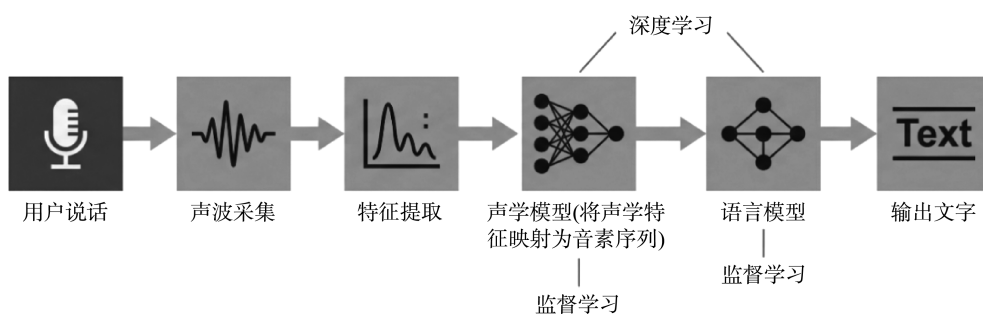


图 2-12 语音识别系统的工作流程

四、推荐系统

推荐系统可能是日常“接触”最多，却最少“意识到”的机器学习应用。当打开购物 App 看到“猜你喜欢”，打开短视频看到“为你推荐”，打开音乐 App 听到“每日歌单”时，背后都是推荐系统在工作。

1. 推荐系统是如何“猜中”消费者的喜好的

推荐系统的核心问题是：在海量的商品（或视频、歌曲、文章）中，预测哪些最可能引起某个特定用户的兴趣。解决这个问题主要有两种经典思路：

协同过滤：“和你品味相似的人喜欢什么，你也可能喜欢。”这种方法不需要理解商品本身的内容，只需要分析用户的行为数据（谁买了什么、谁看了什么）。它通过分析用户行为数据的相似性，找到“品味相似的用户群体”，然后在群体内进行交叉推荐。例如，如果用户 A 和用户 B 都购买了同样的 5 本书，而 B 还购买了第 6 本书，那么系统就可能把第 6 本书推荐给 A。

基于内容的推荐：“你喜欢什么类型的内容，就推荐同类型的。”这种方法分析商品本身的特征（如书籍的类别、电影的导演和演员、歌曲的风格），并结合用户的历史偏好，推荐特征匹配的新内容。它更多地运用监督学习的思路，将“用户是否会喜欢某个商品”建模为一个分类（或回归）问题。

在实际的工业级推荐系统中，这两种方法通常会被结合使用，再叠加深度学习模型来捕捉更复杂的用户行为模式，形成“混合推荐”方案。任务四中介绍的集成学习思想（组合多个模型提升效果）在推荐系统中也有广泛应用。

2. 推荐系统的“双刃剑”效应

推荐系统在带来便利的同时,也引发了值得关注的社会问题。最突出的是信息茧房现象:当推荐系统总是推荐用户“喜欢”的内容时,用户接触到的信息范围会越来越窄,逐渐被“困”在一个由算法为你构建的“信息茧房”中,难以接触到不同的观点和多元的信息。

此外,推荐算法对用户注意力的“精准捕获”也可能导致过度使用问题——很多人刷短视频“停不下来”,正是因为推荐算法经过大量数据训练后,精确地掌握了“什么内容能让你继续看下去”的规律。这些问题表明:机器学习作为一种强大的工具,其应用需要受到伦理和社会责任的约束。后续将在模块十三中进一步讨论这些问题。

知识拓展: 用本模块的“分析框架”拆解任何 ML 应用

通过本任务的学习,读者可以尝试用以下框架来分析任何一个机器学习应用:

1. 问题定义:这个应用要解决什么问题? 是分类、回归、聚类,还是序列决策?
2. 学习方式:它用的是监督学习、无监督学习、半监督学习还是强化学习? 为什么?
3. 数据:它需要什么样的数据? 数据从哪里来? 标注成本高不高?
4. 算法:它可能使用了什么类型的算法(传统 ML 还是深度学习)? 为什么选这个?
5. 评估:怎么衡量它做得好不好? 最关键的评估指标是什么?
6. 局限:它有什么做不到的? 可能存在什么偏差或风险?

掌握了这个框架,读者就拥有了一副“机器学习眼镜”——透过它,读者能看到日常产品背后的技术逻辑,也能更理性地评估 AI 技术的能力与边界。

表 2-5 四大应用领域的机器学习“拆解”

应用领域	核心问题	主要学习方式	关键算法/技术	典型应用
图像识别	从像素中识别物体	监督学习	卷积神经网络(CNN)	人脸解锁、医学影像、工业质检
自然语言处理	理解和生成人类语言	监督/自监督学习	Transformer/大模型	机器翻译、智能客服、文本生成
语音识别	将语音转换为文字	监督学习	深度神经网络+语言模型	语音助手、语音输入、会议转录
推荐系统	预测用户兴趣偏好	监督+无监督学习	协同过滤+深度学习	短视频推荐、电商“猜你喜欢”

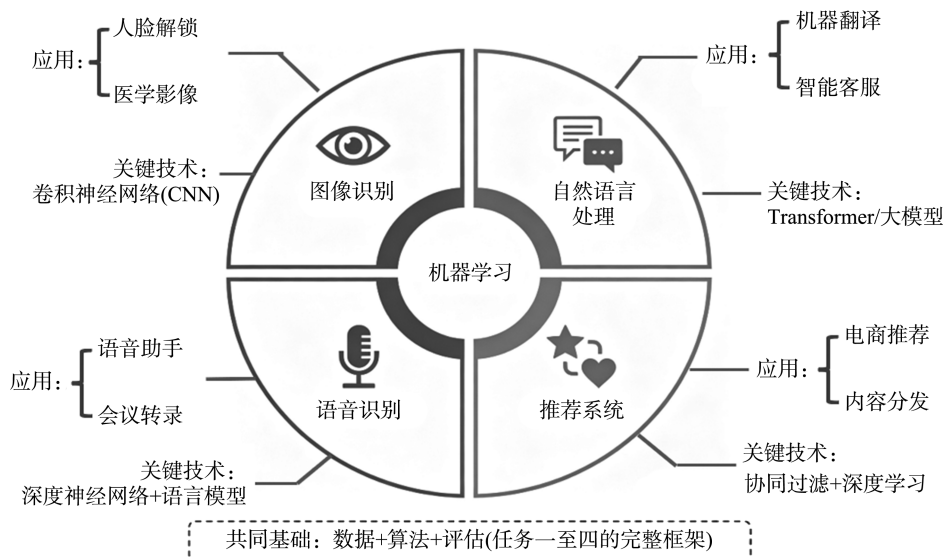


图 2-13 机器学习四大应用领域全景图

本节小结

1. 图像识别是让机器“看懂”图片的技术，主要采用监督学习+卷积神经网络，在人脸识别、医学影像、工业质检等领域已大规模应用。
2. 自然语言处理是让机器“读懂”和“写出”人类语言的技术，深度学习（尤其是Transformer架构和自监督学习）的引入带来了革命性突破。
3. 语音识别是将语音转换为文字的技术，深度学习大幅提升了准确率，支撑了语音助手、语音输入等日常应用。
4. 推荐系统通过分析用户行为数据预测兴趣偏好，协同过滤和基于内容的推荐是两种经典方法，现代系统普遍采用深度学习的混合方案。
5. 用“问题定义→学习方式→数据→算法→评估→局限”的分析框架，可以系统地拆解和理解任何一个机器学习应用。

思考题

1. [识记] 请从图像识别、自然语言处理、语音识别、推荐系统中选择两个领域，分别说明它们主要使用哪种学习方式，以及为什么这种学习方式适合该领域。
2. [理解] 推荐系统的“信息茧房”问题是一个被广泛讨论的社会议题。请从技术角度分析：为什么推荐算法会导致信息茧房？如果你是推荐系统的设计者，可以从哪些方面入手来缓解这个问题？（提示：可以考虑在推荐策略中引入“多样性”指标）

3. [应用] 请选择你最常使用的一款 App(如短视频、外卖、音乐、购物等),运用本任务知识拓展框中的“六步分析框架”,对该 App 中的核心机器学习功能进行一次完整的“拆解分析”。请分析该功能最大的技术挑战可能在哪个环节?

◆ 思想引领

一、数字时代的基本公民能力

本模块学习表明,机器学习的一切能力都建立在数据之上。任务一阐明“机器从数据中学习规律”,任务三揭示了数据准备工作在整个项目中占据 80% 的时间,任务五展示了图像识别、推荐系统等应用如何依赖海量的训练数据。数据,已经成为这个时代最重要的“生产资料”之一。

在这样的背景下,数据素养——即理解数据如何被采集、处理、分析和应用的基本能力——正在成为每一个现代公民都需要具备的核心素养。它不仅包括技术层面的理解(数据是怎么变成“智能”的),更包括伦理层面的判断(数据的采集和使用是否合理、公正)。

任务三中关于数据质量的讨论带来重要警示:“脏数据”会导致“错误的智能”。如果训练数据本身存在偏差,模型学到的“规律”就会带有偏见。例如,一个用主要来自城市用户数据训练的推荐系统,可能无法准确理解农村用户的需求;一个用单一性别数据训练的医学诊断模型,可能对另一性别的诊断准确率显著降低。未来社会建设进程中,需重视数据公平性,保障技术发展造福全体人群,避免加剧社会失衡。

二、“工具理性”与“价值理性”

在任务四中,介绍了多种机器学习算法,见识了它们在各种问题上的出色表现。面对如此强大的技术工具,一种常见的心态是过度信赖——“既然算法能从数据中发现规律,那就让算法来做所有决策吧。”这种将机器学习视为“万能答案”的倾向,值得人们警惕。

任务一中的“没有免费的午餐”定理表明:没有哪种算法能在所有问题上都表现最好。任务三中的过拟合和欠拟合现象印证了模型的预测能力是有边界的。任务五中推荐系统的“信息茧房”问题更是说明即使算法在技术指标上表现优秀,其社会影响也可能是负面的。

德国社会学家马克斯·韦伯曾经区分了两种理性:工具理性(关注“如何高效地实现目标”)和价值理性(关注“这个目标本身是否值得追求”)。机器学习是一种极其强大的工具理性——它能以前所未有的效率从数据中发现规律、作出预测。但“该不该用这个预测来做决策”“这个决策是否公平”“它会不会伤害某些人”这些问题,不是算法能回答的,需要人类的价值判断。

对于非技术专业的学生来说,理解机器学习的能力边界,可能比掌握算法的技术细节更加重要。在未来的职业生涯中,他们很可能不会亲自编写算法,但几乎一定会在工作中

使用基于机器学习的工具,或者参与涉及 AI 技术应用的决策。此时,需要的不是“盲目信任算法的输出”,而是“有能力判断算法的适用条件和潜在风险”。这种判断力,正是通识教育的核心价值所在。

三、终身学习

AI 技术正在以惊人的速度迭代。本模块介绍的很多具体算法和应用,在几年后可能就会被更新的方法所超越。那么,学习这些内容的意义何在?

答案在于:真正需要学习的并非单一具体算法,而是分析问题的思维框架。任务二教我们根据数据特点选择学习方式(监督、无监督、强化),任务三教我们如何系统地思考一个机器学习项目的全流程,任务四教我们结合问题性质匹配算法工具,任务五教我们借助统一的分析框架“拆解”任何一个 ML 应用。这些思维方式是“不变”的,它们不会因为某个算法的过时而失效。

正如“学会钓鱼比得到一条鱼更重要”——在技术快速变化的时代,培养“学习如何学习”的元能力,远比记住某一种具体工具的用法更有长期价值。本模块的学习,正是这种元能力培养的一次实践。

◆ 任务实践

以下实践任务旨在帮助读者将本模块所学的机器学习理论知识与实际问题相联系,通过调研、分析和动手体验,加深对机器学习概念、流程和应用的理解。建议以小组(3—5 人)为单位完成,每组提交一份书面报告并进行课堂汇报。

实践一 “拆解一个 ML 应用”分析报告

任务说明

目标:运用本模块学习的“六步分析框架”(任务五知识拓展框),对一个真实的机器学习应用进行系统分析。

步骤:

(1) 小组选择一个日常生活中的机器学习应用(如短视频推荐、外卖配送时间预测、手机人脸解锁、输入法联想、语音转文字等)。

(2) 按照以下框架进行分析:a. 问题定义——这个应用解决的是什么问题? 属于分类、回归还是聚类? b. 学习方式——它最可能使用哪种学习方式(监督/无监督/强化)? 为什么? c. 数据——它需要什么样的训练数据? 这些数据从哪里来? 标注成本高不高? d. 算法——它可能使用了什么类型的算法? e. 评估——怎么衡量它的效果? 最关键的评估指标是什么? f. 局限——它有什么做不到的? 可能存在什么偏差或风险?

(3) 撰写一份不超过 2 000 字的分析报告,要求论据清晰,引用本模块所学的具体知识点。

评价维度:分析框架的完整运用、技术判断的合理性、批判性思考的深度。

实践二 “监督学习 vs 无监督学习”体验实验

任务说明

目标:通过动手体验,直观感受监督学习和无监督学习的核心差异。

步骤:

(1) 准备材料:收集 30—50 张动物图片(混合猫、狗、鸟三种),打印或在电脑上展示。

(2) 模拟“监督学习”:将图片分为两组。第一组(约 30 张)每张图片附上标签(“猫”“狗”“鸟”),交给一位同学“学习”(观察有标签的图片,总结区分规律)。然后用第二组(约 20 张,无标签)测试这位同学,记录正确率。

(3) 模拟“无监督学习”:将全部 50 张图片(无标签)交给另一位同学,只告诉他“请把这些图片分成若干组”。观察他如何分组,以及分出的组是否与实际的猫/狗/鸟类别一致。

(4) 小组讨论:a. 两种“学习”方式的效果有何差异? b. “监督学习”的同学在测试中犯了什么类型的错误? 这些错误对应任务三中的过拟合还是欠拟合? c. “无监督学习”的同学分出的组和预期的类别一致吗? 如果不一致,他的分组逻辑是什么? d. 该实验如何帮助我们理解“标签”在机器学习中的作用?

提示:这个实验的目的不是测试同学的图片识别能力,而是通过人类的“学习”过程来类比机器学习的工作方式。鼓励大家在讨论中联系本模块的具体知识点。

评价维度:数据标注的准确性、实验辨析的合理性、分析总结的有效性。

实践三 “算法选型”情境挑战

任务说明

目标:锻炼在具体问题场景中选择合适的学习方式和算法的判断能力。

步骤:

(1) 每个小组从以下情境中抽取 2 个(或自行设计类似情境):

情境 A:一家连锁餐厅想预测每天中午的客流量,以便提前准备食材。

情境 B:一家电商平台想把数百万用户自动分成不同的客户群体,以便实施差异化营销策略。

情境 C:一家银行想判断每笔信用卡交易是否存在欺诈风险。

情境 D:一家游戏公司想让游戏中的 NPC(非玩家角色)自动学会与玩家对战的策略。

情境 E:一家医院拥有大量的胸部 X 光片,但只有少部分经过了专家标注,想开发一个辅助诊断系统。

(2) 对每个情境,讨论并回答:a. 这是什么类型的问题(分类/回归/聚类/序列决策)? b. 应该采用哪种学习方式? c. 推荐使用什么算法(可以参考任务四的对比表)? d. 需要什么数据? 可能面临什么数据挑战? e. 如何评估模型效果? 最重要的评估指标是什么?

(3) 各小组课堂汇报方案,其他小组提出改进建议。

评价维度:问题分析的准确性、算法选择的合理性、团队协作与表达能力。

实践四 “推荐算法的信息茧房”调研与辩论

任务说明

目标:结合技术知识与社会思考,培养对机器学习社会影响的批判性认知。

步骤:

(1) 个人调研(课前完成):每位同学选择一个自己常用的内容平台(短视频、新闻、音乐等),连续3天记录平台推荐给自己的内容类型。观察:推荐内容是否越来越“同质化”?是否感觉自己被“困”在了某种类型的内容中?

(2) 小组汇总:整理全组成员的调研结果,分析推荐系统的内容分布是否存在“信息茧房”的迹象。

(3) 课堂辩论:全班分为正反两方。辩题:“内容平台的推荐算法应当被强制要求增加推荐内容的多样性。”正方论证:算法导致信息茧房,平台有社会责任;反方论证:用户有选择权,强制干预可能降低用户体验。

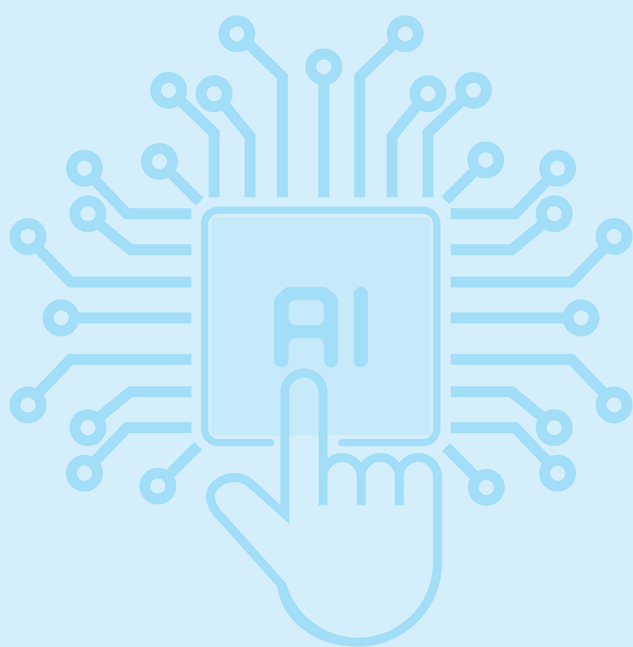
(4) 辩论后反思:请每位同学写一段不超过300字的个人反思,回答:这次调研和辩论是否改变了自己对推荐算法的看法?作为用户,可以采取什么措施来“打破”信息茧房?

评价维度:调研数据的真实性和分析深度、辩论中论据的技术含量和逻辑性、个人反思的独立思考深度。

南京大学出版社
版权所有

» 模块三

神经网络



南京大学出版社
版权所有

南京大学出版社
版权所有

◆ 模块导读

在模块一中,了解到连接主义的核心思想是“模拟大脑中神经元的连接方式来实现智能”;在模块二中,知道了深度学习是机器学习中最强大的分支,它使用“深层神经网络”在图像识别、语音识别、自然语言处理等领域取得了划时代的突破。那么,神经网络究竟是什么?它为什么要模仿大脑?一个人工神经元是如何工作的?多个神经元又是怎样组合成一个能够“学习”的网络的?

从生物神经系统到人工神经元模型,从简单感知器到复杂多层网络,神经网络经历了怎样的演进?前馈网络、反馈网络、随机网络等不同结构的神经网络,各自具备何种特点,适合解决哪些类型的问题?

本模块围绕以上核心问题展开,通过六个任务,带领读者系统理解神经网络。任务一从宏观视角介绍神经网络的发展脉络与基本思想;任务二深入解析生物神经网络的工作原理及其对人工神经网络的启发;任务三系统讲解人工神经网络的基本结构、学习规则和工作方式;任务四介绍应用最广泛的前馈神经网络,含感知器与BP网络;任务五介绍具备“联想记忆”能力的Hopfield网络;任务六介绍引入随机机制的玻尔兹曼机。通过本模块学习,读者可理解神经网络从生物灵感得到工程实现的完整逻辑链条,为后续深度学习相关模块的学习奠定坚实基础。

◆ 学习目标

【了解】

- 神经网络的发展历程,包括关键人物如麦卡洛克、皮茨、罗森布拉特、鲁梅尔哈特、辛顿等,以及里程碑事件;
- 生物神经系统,包括神经元、突触、信号传递的基本工作原理,以及它对人工神经网络设计的启发;
- 前馈神经网络、Hopfield网络和玻尔兹曼机等不同网络结构的名称、特点和典型应用场景。

【理解】

- 人工神经元模型的基本工作方式:输入、加权求和、激活函数、输出;
- 人工神经网络的学习过程,即通过调整连接权重使网络输出逼近目标的“训练”机制;
- 前馈网络与反馈网络的结构区别,以及反向传播算法在多层网络训练中的核心作用。

【掌握】

- 根据模式分类、函数逼近、联想记忆、优化求解等问题特点,判断适配的神经网络类型;
- 运用层、节点、权重、激活函数、训练等神经网络基本概念,分析、描述简单神经网络系统。

【能够】

- 结合前两个模块的知识,将神经网络置于“人工智能→机器学习→深度学习”的知识体系中,理解其在整体技术版图中的位置;

—对人脸识别、语音助手、智能翻译等日常生活中的神经网络 AI 应用,形成深入技术理解,并解释其基本工作原理。

◆ 思维导图

下图展示了本模块的知识结构与内容框架,帮助读者从整体上把握各任务之间的逻辑关系:

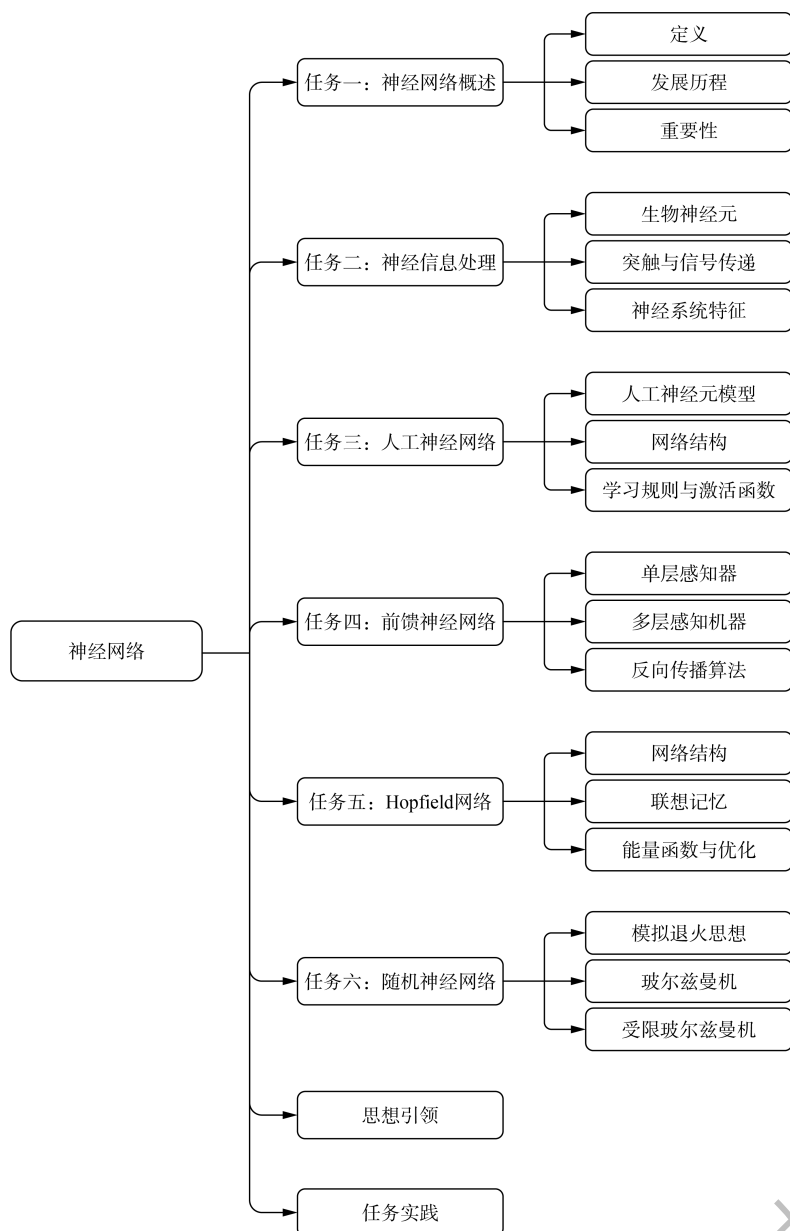


图 3-1 模块三知识结构思维导图

任务一 神经网络概述

◆ 引导思考

在模块一中,连接主义认为“智能涌现于大量简单单元的连接与协作”。在模块二中,深度学习是当今机器学习最强大的分支。作为连接主义与深度学习共同的技术基石,神经网络经历了怎样的发展历程?

科学家为何想到用数学模型“模拟大脑”?这看似大胆的想法,如何从理论设想变为改变世界的实用技术?

一、什么是神经网络

人工神经网络(Artificial Neural Network,简称 ANN 或“神经网络”),是受生物大脑启发的计算模型。该模型由大量相互连接的简单计算单元组成,此类单元称为人工神经元或节点。神经网络通过调整节点间连接强度,从数据中学习规律,该连接强度称为权重。

神经网络得名“神经”网络,源于其设计灵感直接来自生物大脑神经元的工作方式。人工神经网络对生物神经系统的模拟高度简化,如同飞机模拟鸟的飞行原理,却并非鸟类本身。但这种“从自然获取灵感”的思路,催生了一系列极为强大的技术。

二、神经网络的发展脉络

模块一任务二中,已沿时间线梳理神经网络的关键节点。现从“神经网络”主题视角,串联这些节点,形成清晰发展脉络。

表 3-1 神经网络发展的关键里程碑

时间	事件	意义
1943 年	麦卡洛克和皮茨提出 MCP 人工神经元模型	证明神经元的逻辑功能可以用数学描述
1958 年	罗森布拉特发明感知器(Perceptron)	第一个能从数据中“学习”的神经网络模型
1969 年	明斯基和帕珀特出版《感知器》	指出单层感知器的严重局限,引发第一次“寒冬”
1986 年	鲁梅尔哈特等推广反向传播算法	使多层网络可训练,神经网络“复活”
1989 年	杨立昆提出卷积神经网络(CNN)	为图像识别开辟了专用的网络结构
2006 年	辛顿提出深度信念网络	标志着“深度学习”概念的正式提出
2012 年	AlexNet 在 ImageNet 竞赛中大获全胜	深度学习时代全面开启
2017 年	谷歌提出 Transformer 架构	大语言模型的基石,开启大模型时代

由表 3-1 可见,神经网络发展并非一帆风顺。1943 年至 2012 年近七十年间,经历至少两次“被判死刑又复活”的戏剧性转折。“热潮—寒冬—复兴”的波动模式,是模块一总结的 AI 发展规律缩影。2012 年后十余年,神经网络以深度学习形式,进入“大放异彩”的黄金时代。

三、神经网络为什么重要

深度学习的核心技术载体,正是深层神经网络。理解神经网络,即抓住当今 AI 技术最核心的“发动机”。

神经网络的强大,源于通用逼近能力。理论上,足够大的多层前馈神经网络,可学会逼近任意复杂函数关系。通俗而言,只要提供足够数据与算力,神经网络几乎能“学会”任意输入到输出的映射规律,无论规律的复杂程度如何。这是神经网络能在图像识别、语音识别、自然语言处理、游戏对弈等截然不同领域取得突破的原因。从数学角度看,此类问题本质均为“从输入数据到输出结果的复杂映射”。

当然,“理论上能学会”与“实际上学得好”存在较大距离。模块二任务三讨论的数据质量、模型训练、过拟合等问题,在神经网络中同样存在,且更为突出。后续任务从生物原理出发,逐步深入理解神经网络的结构、原理与工作方式。

知识拓展:“仿生”的智慧——从鸟到飞机,从大脑到神经网络

神经网络“模仿大脑”的思路,在科学史上并不罕见。飞机的发明是经典仿生案例。莱特兄弟从鸟的飞行中获得灵感,却并未制造“机械鸟”,这是因为历史上扇翅膀飞行器屡试屡败。莱特兄弟提取飞行核心物理原理,即气流与翼面关系,以工程方法,用固定翼加发动机,实现了比鸟更快更远的飞行。

神经网络与大脑的关系同理。借鉴大脑“大量简单单元通过连接协作”的核心设计理念,而非在细胞层面完全复制大脑,后者为神经科学任务。人工神经元远比生物神经元简单,人工网络规模远小于真实大脑。如同固定翼飞机无需扇翅膀即可飞行,人工神经网络无需完全模拟大脑,即可展现强大学习能力。

这种“提取核心原理而非照搬细节”的仿生策略,是工程创新的重要思维方式。

本节小结

1. 人工神经网络是受生物大脑启发、由大量简单计算单元(神经元)相互连接组成的计算模型,通过调整连接权重从数据中学习。
2. 神经网络发展经历“诞生(1943)→热潮(1958)→寒冬(1969)→复活(1986)→爆发(2012)”的曲折历程。
3. 神经网络是深度学习核心技术载体,强大的通用逼近能力使其在图像、语音、语言等领域取得突破性成果。

思考题

1. [识记] 列出神经网络发展史上三个最关键的里程碑事件,用文字说明各事件重要性。
2. [理解] 有观点认为“人工神经网络就是在计算机里模拟了一个大脑”。结合本任务“仿生策略”讨论,分析该观点的合理之处与不准确之处。

任务二 神经信息处理的基本原理

◆ 引导思考

人类大脑由约 860 亿个神经元组成,通过约 100 万亿个突触相互连接,构成已知宇宙中最复杂的信息处理系统。人类的感知、思考、记忆、决策,均涌现于这巨大网络活动之中。

生物神经元如何接收、处理、传递信息? 单个神经元是否具备智能? 大量简单神经元如何协作产生智能?

理解生物学基础,对人工神经网络设计原理的理解,存在何种帮助?

一、生物神经元的结构

理解人工神经网络,需先了解灵感来源——生物神经元。人工神经元是对生物神经元的高度简化,理解生物原型,有助于把握人工模型设计逻辑。

典型生物神经元(神经细胞)由细胞体(soma)、树突(dendrite)、轴突(axon)三部分组成。

树突为神经元“输入端”,如树枝般从细胞体向外延伸,接收其他神经元信号。单个神经元可拥有成千上万根树突,同时接收大量神经元输入。

细胞体为神经元“处理中心”,汇总所有树突接收的信号并综合处理。汇总信号强度超过阈值时,细胞体激活,产生电脉冲。

轴突为神经元“输出端”,是长形“电线”,传递细胞体产生的电脉冲至其他神经元。轴突末端分叉为突触小体,每个小体与下一神经元树突形成连接。

单个神经元类似微型投票站:树突为投票箱,接收输入;细胞体为计票员,汇总判断;轴突为广播喇叭,传递结果。兴奋性信号(赞成票)超过抑制性信号(反对票)达到阈值时,神经元激活,通过轴突传递信号。

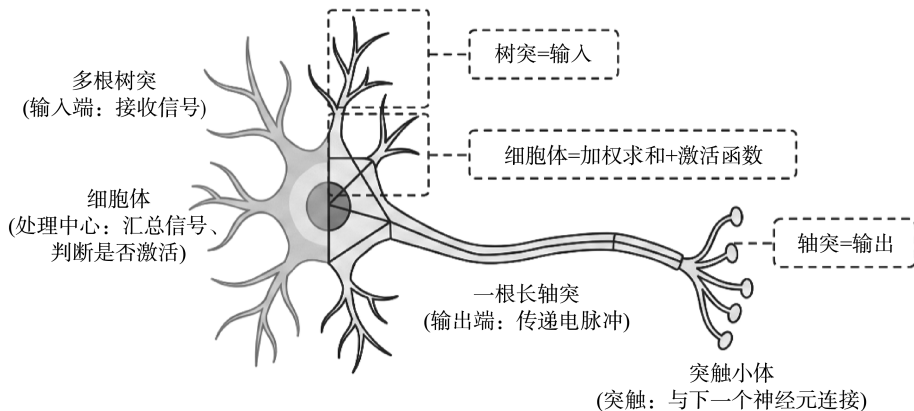


图 3-2 生物神经元的基本结构

二、突触与信号传递

神经元间不直接接触,通过突触传递信号。突触为两神经元间微小间隙,宽度约 20 至 30 纳米。神经元轴突末端电脉冲到达突触时,触发释放神经递质,化学物质跨越间隙,与下一神经元树突受体结合,完成信号传递。

突触具备关键特性——可塑性。突触传递信号效率,即对下一神经元影响强度,并非固定不变,可根据使用情况增强或减弱。两神经元频繁同时活跃,突触连接增强;反之减弱。此为赫布规则核心思想,表述为:“一起放电的神经元会连在一起。”

突触可塑性被认为是学习与记忆的生物学基础。反复练习弹钢琴时,控制手指运动的神经元间突触连接逐渐增强,手指配合愈发默契,即“熟能生巧”的神经层面解释。

人工神经网络中,节点间连接权重,是突触强度的数学模拟。人工网络的“学习”过程,即调整权重使输出逼近目标,与突触可塑性生物机制存在直观对应关系。赫布规则被视为人工神经网络学习理论的思想源头。

案例:“伦敦出租车司机的大脑”——突触可塑性的生动证据

科学家对伦敦出租车司机进行脑部扫描,发现其负责空间记忆的海马体,明显大于普通人。伦敦街道复杂,出租车司机需数年记忆全城路线与地标,高强度空间记忆训练,使相关脑区神经连接更密集高效。

司机从业时间越长,海马体越大。说明大脑结构并非固定不变,可根据“训练数据”,即反复路线记忆,持续调整优化。此过程与人工神经网络通过数据训练提升性能逻辑相似。

三、神经系统的信息处理特征

生物神经系统经数亿年进化,形成独特信息处理特征,为人工神经网络设计提供重要启发。

1. 大规模并行处理

单个生物神经元处理速度缓慢,神经脉冲传递速度约每秒 100 米,远低于计算机芯片电信号每秒约 3 亿米的传播速度。大脑拥有约 860 亿个神经元同步工作,通过大规模并行方式,使大脑整体完成复杂任务。

该特征直接启发人工神经网络设计:网络中每个节点独立完成简单计算,大量节点同步工作。此为 GPU 比 CPU 更适合训练神经网络的原因,GPU 擅长大规模并行计算。

2. 分布式存储

传统计算机中,信息存储于特定内存地址,地址损坏则会造成信息丢失。大脑记忆不存储于特定位置,分散于大量神经元连接模式中,即分布式存储。信息编码于整个网络权重分布,而非单个节点。

分布式存储赋予大脑惊人容错能力,部分神经元受损,整体网络信息仍可大部分恢复。模糊或缺照片仍可识别,人工神经网络具备类似容错特性,部分节点失效,网络整体性能不会急剧下降。

3. 自组织与自适应

大脑可根据外部环境变化,自动调整内部结构。新生儿大脑出生后数年,经历大规模突触生成与修剪,形成适应生长环境的神经回路。自组织与自适应能力,使大脑灵活应对变化环境。

人工神经网络依靠训练数据自动调整权重完成学习,该过程是对生物大脑自适能力的工程模拟。不过,人工神经网络的灵活性与自适应能力仍远不及生物大脑。

表 3-2 生物神经系统与人工神经网络的对应关系

生物神经系统	功能	人工神经网络对应物
神经元(细胞体)	信号处理与判断	节点(人工神经元)
树突	接收输入信号	输入连接
轴突	传递输出信号	输出连接
突触	连接两个神经元,传递信号	连接权重(weight)
突触强度/可塑性	影响信号传递效率,可调节	权重大小(可通过训练调整)
兴奋/抑制	促进或抑制下一个神经元激活	正权重(兴奋)/负权重(抑制)
激活阈值	信号超过阈值才触发输出	激活函数(activation function)

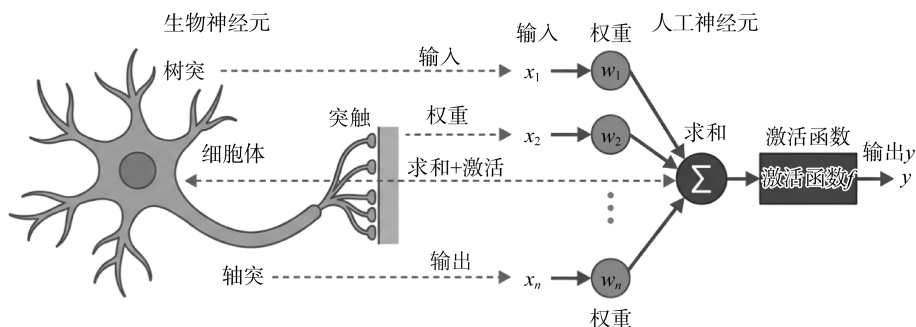


图 3-3 从生物神经元到人工神经元的映射

知识拓展：人工神经网络“不是”大脑模拟器

人工神经网络设计灵感来自生物大脑,但是两者存在巨大差异:

规模差距:人类大脑约 860 亿个神经元、100 万亿个突触连接;截至本书编写时,最大人工神经网络参数规模达万亿级别,复杂度仍远低于大脑。

工作机制:生物神经元通过电化学信号传递信息,涉及复杂分子生物学过程;人工神经元仅执行简单数学运算(加权求和+激活函数)。

能耗差距:人类大脑功耗约 20 瓦,相当于小灯泡;训练大型人工神经网络则需消耗数百万度电能。

学习效率:人类仅需数张照片即可认识新动物;人工神经网络则需成千上万张标注图片来完成同类任务。

因此,将人工神经网络理解为“受大脑启发的数学模型”,比“大脑的计算机模拟”更准确。

本节小结

1. 生物神经元由树突(输入)、细胞体(处理)和轴突(输出)三部分组成,通过突触与其他神经元连接和通信。
2. 突触可塑性,即连接强度可调节,是学习与记忆的生物学基础,为人工神经网络“权重调整”的灵感来源。赫布规则,即“一起放电的神经元会连在一起”,是神经网络学习理论思想源头。
3. 生物神经系统具备大规模并行处理、分布式存储、自组织自适应三大信息处理特征,均在人工神经网络设计中得到借鉴。
4. 人工神经网络是受大脑启发的数学模型,非大脑精确模拟,两者在规模、机制、能耗、学习效率上存在巨大差异。

思考题

1. [识记] 说明生物神经元树突、细胞体、轴突、突触的功能,及其在人工神经网络中的对应概念。
2. [理解] 赫布规则表述为“一起激活的神经元会连接在一起”。以外语学习或乐器练习为例,解释该规则,并说明其与人工神经网络“训练调整权重”的关系。
3. [应用] 人类仅需数张照片即可认识新动物,人工神经网络需成千上万张。结合本任务生物神经系统特征讨论,分析原因。(提示:考虑大脑的已有知识积累、多感官信息融合、进化带来的先天“预装”能力等)

任务三 人工神经网络

◆ 引导思考

任务二已了解生物神经元结构与工作原理。科学家如何将生物学知识“翻译”为数学模型? 人工神经元的数学版本结构如何?

单个人工神经元仅能完成简单任务。大量神经元组装成网络时,如何排列? 信息在网络中如何流动?

网络如何“学习”? “调整权重”过程如何实现?

一、人工神经元模型

1. MCP 模型:最早的人工神经元

1943年,麦卡洛克与皮茨提出 MCP 模型,为首个人工神经元模型。工作原理简单:接收多个二值输入(0 或 1),加权求和,结果超过阈值输出 1(激活),否则输出 0(不激活)。

以投票站比喻:每个输入对应选民,投赞成票(1)或反对票(0);选民影响力对应权重;计票员汇总加权总分,超过通过线(阈值),宣布通过,输出 1。

MCP 模型虽然简单,但首次证明了数学与逻辑可以描述神经元的基本功能,是生物学到工程学的关键翻译步骤。

2. 现代人工神经元模型

在 MCP 模型基础上,现代人工神经元模型完成两项重要扩展:输入不再局限于 0 和 1,可取任意实数值,如 0.3、-1.7、42.5 等,可处理更丰富信息;输出不再局限于 0 和 1,通过激活函数产生连续输出值。

现代人工神经元工作过程分三步：接收、加权求和、激活。该神经元首先从多输入端接收信号；其次将各输入与对应权重相乘并累加，再叠加偏置项完成加权求和；最后把求和结果送入激活函数运算，生成最终输出。

以做菜比喻：接收为从冰箱取食材；加权求和为按菜谱比例混合调料；激活为下锅烹饪，把控合适火候方可出锅。

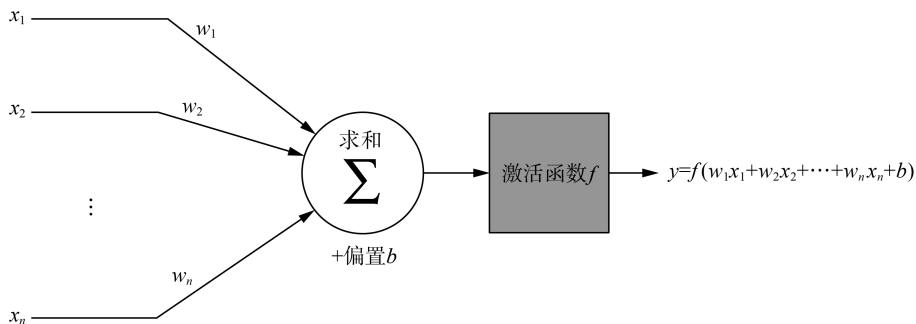


图 3-4 人工神经元的工作流程

3. 激活函数：给神经元注入“非线性”

激活函数是人工神经元的关键组件，其作用是将加权求和结果转化为最终输出。若无激活函数，多层网络整体等价于简单线性运算，仅能绘制直线，无法捕捉现实中的复杂非线性关系。

激活函数引入非线性，使多层神经网络弯曲决策边界，拟合任意复杂曲线。无激活函数时，神经网络如同仅能画直线的直尺；加入激活函数后，变为可弯曲的曲线尺，能绘制任意形状。

常见激活函数包括：阶跃函数，MCP 模型所用，输出非 0 即 1；Sigmoid 函数，输出压缩至 0 到 1，常用于概率输出；ReLU 函数，简单高效，避免梯度消失，加速深度网络训练。读者无需记忆各类激活函数的数学表达式，只需把握其共同作用：引入非线性，使网络学习复杂规律。

二、网络结构

单个人工神经元能力有限，仅能完成简单加权投票。神经网络真正的威力，源于大量神经元的组合协作。根据神经元排列方式、信息流动方向，神经网络分为多种基本结构类型。

1. 网络的层次结构

多数人工神经网络采用分层结构，神经元组织为若干层，每层含多个神经元。基本分层结构包括三类：输入层，接收外部原始数据，如图像像素值、用户行为数据；隐藏层，输入层与输出层间的中间加工层，可设置一层或多层；输出层，输出最终结果，如“类别为猫”“销量预计 500”。

以工厂流水线理解:输入层为原材料仓库;隐藏层为加工车间,多层串联形成流水线,即“深度”含义;输出层为成品出口。

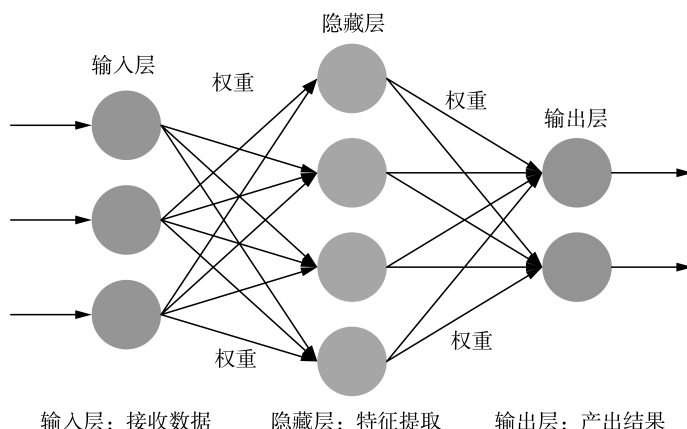


图 3-5 神经网络的层次结构示意图

2. 前馈网络与反馈网络

根据信息在网络中的流动方向,神经网络分为前馈网络、反馈网络(递归网络)两大类。

前馈网络中,信息严格从输入层向前流向输出层,无回头路,类似水流从高处向低处流动。前馈网络是最常见的网络结构,适用于多数分类、回归问题,任务四会详细介绍前馈网络代表——感知器、BP 网络。

反馈网络中,信息可回流,部分层输出反馈至前层作为输入,形成环路。该结构赋予网络记忆能力,可处理时间、序列相关任务,如语音、文本。本模块任务五介绍的 Hopfield 网络为典型反馈网络,主要用于联想记忆;处理序列任务的 RNN、LSTM 将在后续模块介绍。

表 3-3 前馈网络与反馈网络的对比

维度	前馈网络	反馈网络
信息流向	单向:输入→输出	存在回路:输出可反馈回输入
记忆能力	无:每次输入独立处理	有:可记忆之前的状态
典型结构	感知器、BP 网络、CNN	Hopfield 网络、RNN、LSTM
适用问题	分类、回归、图像识别	序列处理、联想记忆、时间序列预测
通俗比喻	流水线:原料从头走到尾	反馈环:做完一步看看效果再调整

三、网络是怎样“学会”的

神经网络的核心能力是自主学习,即通过训练数据自动调整权重,使网络输出逼近目标。模块二任务三已从宏观角度了解“预测、计算损失、调整参数”训练循环,现从神经网络角度具体理解该过程。

1. 赫布学习规则

最早提出的神经网络学习规则,直接源自任务二介绍的赫布规则:“一起放电的神经元,连接会增强。”数学层面可表述为:相连神经元同步激活时,权重增大;激活模式不一致,权重减小。赫布规则为无监督学习规则,无需标准答案,主要用于自组织、联想记忆类网络。

2. 误差修正学习规则

监督学习中,常用误差修正规则:对比网络实际输出与人工标注标准答案(标签),依据两者差值(误差)反向调整权重,使下次输出逼近正确答案。此为模块二熟悉的监督学习模式,类似教师批改作业。

误差修正学习代表为反向传播算法(Backpropagation,简称BP算法)。模块一任务二以足球队赛后复盘比喻反向传播:从最终结果(误差)出发,逐层向前追溯,计算每个节点对误差的贡献,按比例调整权重。过程在每批训练数据上重复执行,网络预测愈发准确。任务四会进一步讨论反向传播算法在多层前馈网络中的应用。

3. 竞争学习规则

第三种重要学习规则为竞争学习。网络中多个神经元竞争对输入的响应权值,采用“胜者通吃”的原则:只有获胜的神经元被激活并更新其权重,其余神经元不变。该规则常用于聚类、特征发现任务,与模块二无监督学习聚类思想呼应。

表 3-4 三种主要学习规则对比

学习规则	核心思想	学习方式	典型应用
赫布规则	一起激活则增强连接	无监督	联想记忆(Hopfield 网络)
误差修正(BP 算法)	根据误差调整权重	监督	分类、回归(前馈网络)
竞争学习	胜者通吃,只更新赢家	无监督	聚类、特征发现

案例：用“学骑自行车”理解神经网络的学习过程

学骑自行车是理解神经网络学习的绝佳类比：

首次尝试：摇摇晃晃易摔倒，类似未训练神经网络，权重随机，输出（平衡控制）与目标（稳定骑行）差距大。

感知误差：摔倒后，大脑计算偏差——身体倾斜过度，类似计算损失（实际输出与目标差距）。

调整权重：下次尝试，反向用力纠正偏差，类似依据误差调整权重，减小错误连接、增强正确连接。

反复迭代：经数十次尝试、摔倒、调整循环，权重（肌肉协调模式）优化，稳定骑行。

学会后权重固定：多年不骑车，重新骑行可快速恢复，因神经网络权重（神经连接模式）已训练至良好状态。

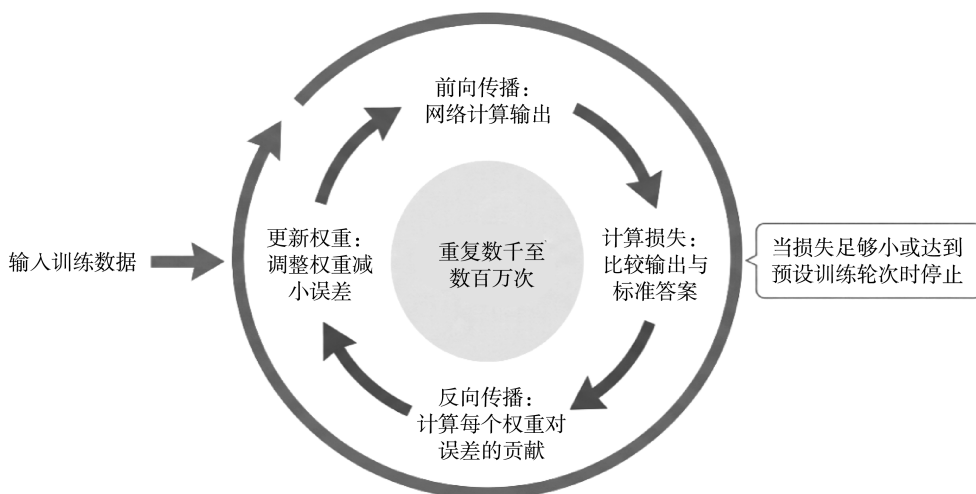


图 3-6 神经网络学习过程的直觉理解

本节小结

1. 人工神经元模型将生物神经元简化为“接收输入→加权求和→激活函数→输出”的数学过程。MCP 模型(1943)是最早的人工神经元,现代模型扩展支持连续值输入输出。
2. 激活函数给网络注入“非线性”,使得多层网络能够学习任意复杂的规律。常见的激活函数包括 Sigmoid 和 ReLU 等。
3. 神经网络采用分层结构,包括输入层、隐藏层、输出层。根据信息流向分为前馈网络(单向流动)和反馈网络(存在回路)。
4. 网络学习本质为调整权重,三种主要学习规则:赫布规则(一起激活则增强)、误差修正/BP 算法(依据误差调整)、竞争学习(胜者通吃)。
5. 反向传播算法是训练多层前馈网络最重要的方法,将在任务四中进一步展开。

思考题

1. [识记] 描述人工神经元工作过程(接收→加权求和→激活→输出),并解释激活函数必要性。
2. [理解] 简述前馈网络与反馈网络核心区别,各举一个适配场景并说明原因。
3. [应用] 训练神经网络判断文字评论情感(正面/负面),运用本任务知识描述:
 - a. 输入层接收数据;
 - b. 输出层输出内容;
 - c. 训练所用学习规则;
 - d. 训练过程损失衡量方式。

任务四 前馈神经网络

◆ 引导思考

在任务三中了解了人工神经网络的基本结构和学习规则,知道了前馈网络是信息“单向流动”的网络类型。那么,最简单的前馈网络,也就是感知器,能做什么?又不能做什么?它的局限性引发了怎样的“学术风暴”?

当把多个层叠加起来,形成“多层感知器”后,网络的能力发生了怎样的质变?“反向传播算法”又是如何让这种多层网络真正“学会”东西的?

从简单的感知器到深度卷积神经网络,前馈网络是如何一步步成为当今 AI 最核心的“工作马”的?

一、单层感知器

1958年,美国心理学家罗森布拉特基于MCP模型,提出感知器(Perceptron),它是首个能从数据中自动学习的神经网络模型。感知器结构简单,仅含输入层、输出层,无隐藏层。每个输入通过权重连接至输出节点,输出节点对加权求和结果应用激活函数(阶跃函数,大于阈值输出1,否则输出0),产生最终输出。

感知器学习过程直观:将带有已知标准答案的训练数据逐一输入网络,输出与正确答案一致则不调整,不一致则按误差修正规则调整权重,减小错误方向权重、增大正确方向权重。多轮训练后,权重调整到位,网络学会正确分类规则。

1. 感知器能做什么

感知器可解决线性可分分类问题。线性可分指二维空间可用直线、三维空间可用平面,完全分隔两类数据。如依据身高、体重判断偏瘦、偏胖,可用斜线分隔两类人群,感知器可解决。

逻辑运算中“与”(AND)、“或”(OR)线性可分。以“与”运算为例,两输入均为1时输出1,其余情况输出0;将四种输入组合绘于平面,可用直线分隔输出1与输出0的点,感知器可学习这条分界线。

2. 感知器不能做什么:“异或”难题

感知器存在致命局限,无法解决非线性分类问题,典型案例为异或(XOR)问题,两输入相同时输出0,不同时输出1。将四种输入组合绘于平面,代表输出1和输出0的点呈交叉排列,无法用一条直线将其分隔。

1969年,MIT明斯基、帕珀特出版《感知器》,以严格数学证明指出局限,影响深远,导

致研究者、资助机构认为神经网络不可行,引发第一次寒冬。

明斯基的批评仅针对单层感知器,书中承认多层网络理论可解决异或问题,当时缺乏有效训练多层网络方法,训练难题解决后,神经网络重获生机。

二、多层感知器

突破感知器局限的思路十分直观:输入层与输出层间加入一层或多层隐藏层,这类网络称为多层感知器(Multi-Layer Perceptron,MLP),即多层前馈网络。

增加一层解决异或问题,单层感知器仅能绘制一条分界线,多层网络可绘制多条分界线并组合。异或问题中,直线无法分隔四点,两条直线可划分平面为三个区域,输出1的点位于两线之间。隐藏层作用为绘制多条分界线,输出层作用为组合分界线,实现复杂分类。

数学上已证明:包含足够多隐藏节点的两层前馈网络,理论上可以逼近任意复杂函数,这被称为万能逼近定理,从数学层面保证多层网络强大的表达能力。

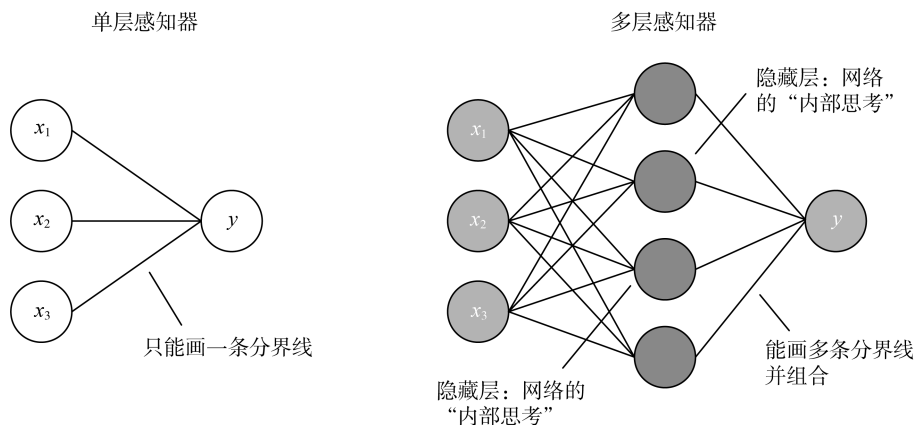


图 3-7 单层感知器与多层感知器的结构对比

三、反向传播算法

多层网络包含多层权重,输出误差是所有层权重共同作用的结果,难以确定每一层、每一个权重对最终误差的贡献,这便是信用分配问题(Credit Assignment Problem)。

1986年,鲁梅尔哈特、辛顿、威廉姆斯发表里程碑式论文,系统化提出了反向传播算法(Backpropagation,简称BP算法),解决了多层网络训练难题。反向传播算法主要包含两个核心阶段:前向传播与反向传播。在前向传播阶段,训练数据由输入层逐层前馈,经各节点加权求和与激活函数处理后,于输出层生成预测值,并据此计算预测结果与真实标签间的误差。在反向传播阶段,误差信号从输出层逐层逆向传导,利用链式法则计算各权重对误差的梯度贡献,进而完成权重的更新与优化。这一闭环过程宛如足球赛后的战术复盘,通过精准追溯误差根源,指导模型不断调整与改进。

两个阶段循环执行:输入数据→前向传播预测→计算误差→反向传播调权→输入下一

批数据。如此循环成千上万次,网络的预测结果会越来越准确,逐步掌握数据中的规律。

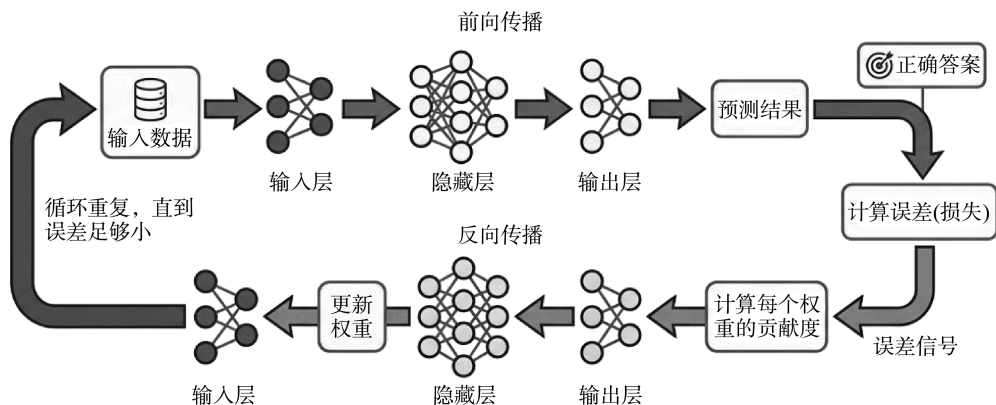


图 3-8 反向传播算法的工作流程

案例：从“不会认字”到“认字高手”——BP 网络的学习过程

以训练网络识别手写数字(0—9)为例：

初始状态：网络权重随机，类似不识字者，输入手写 7 图片，输出 3，完全错误。

第 1 轮：输出 3、正确答案 7，误差大，反向传播调整权重，下次倾向输出 7。

第 100 轮：识别 7，其他数字仍出错，继续训练。

第 1 000 轮：多数手写数字识别正确，准确率超 95%。

过程类似人类认字，反复输入、纠错、掌握规律；人类数十个例子学会新字，神经网络需数万例子，即前文学习效率差距。

四、从 BP 网络到卷积神经网络

标准多层感知器存在效率问题：每层都是全连接，上层每个节点连接到下层所有节点；输入数据量大时，如高分辨率图片上百万像素，全连接导致权重参数庞大，训练缓慢、易过拟合。

1989 年，杨立昆(Yann LeCun)提出了卷积神经网络(CNN, Convolutional Neural Network)，解决上述效率问题。CNN 的核心创新在于局部连接与权值共享。

局部连接：节点不再连接上层所有节点，仅连接上层小区域节点，这借鉴了生物视觉系统，视网膜神经元只负责视野中的小块区域，CNN 节点仅观察图像小块区域(如 3×3 像素)，提取局部特征。

权值共享：同一组权重(卷积核、滤波器)扫描整张图像不同位置，类似同一放大镜逐块检查图像；卷积核学习检测特定局部特征(竖线、角点)，滑动扫描定位特征位置。

这两项设计大幅减少了网络参数的数量，同时保留了图像特征的捕捉能力。CNN 典型结构为：多层卷积层(逐层提取简单到复杂特征)→池化层(压缩数据、降计算量)→全连接层(综合特征、最终判断)。

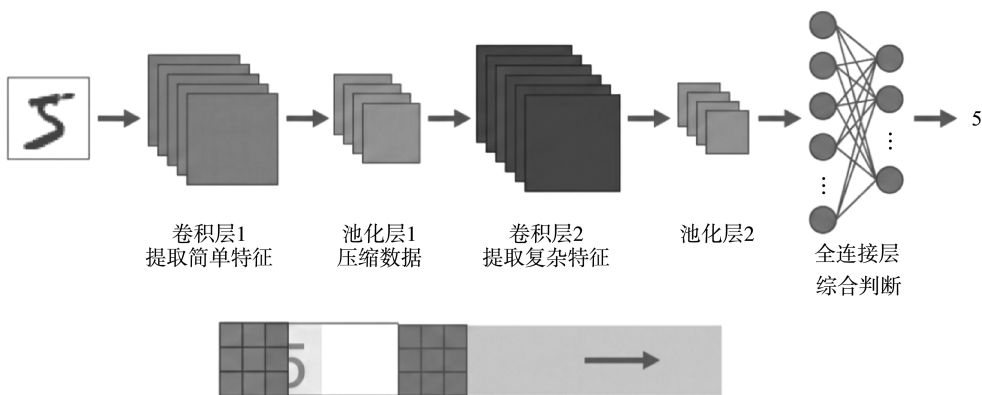


图 3-9 卷积神经网络(CNN)的基本结构

2012 年,基于 CNN 架构的 AlexNet 在 ImageNet 图像识别大赛中大幅领先传统方法,标志深度学习时代的全面开启。此后,CNN 成为计算机视觉领域的核心技术工具,广泛应用于人脸识别、医学影像分析、自动驾驶物体检测等场景。

表 3-5 前馈神经网络的三种代表形态

网络类型	结构特点	核心能力	关键局限	通俗比喻
单层感知器	无隐藏层,只能画一条分界线	线性分类	无法处理非线性问题	“只会一刀切”的门卫
多层感知器(BP 网络)	有隐藏层,可画多条分界线并组合	任意复杂的分类和回归	需要大量数据和计算资源	“会综合判断”的评审团
卷积神经网络(CNN)	局部连接+权值共享,专为图像设计	高效的图像特征提取	主要适用于网格状数据(图像)	“拿着放大镜逐块检查”的质检员

知识拓展：“深度”到底意味着什么？

深度学习(Deep Learning)中的深度,指神经网络层数多。浅层网络含一至两层隐藏层,深度网络含数十至上百层隐藏层。

深度带来的核心价值是层次化特征提取。图像识别中,第一层检测边缘、颜色;第二层组合边缘为角点、纹理;第三层组合纹理为局部形状(眼睛、鼻子);第四层组合局部形状为完整物体(人脸)。每一层都会提取比上一层更抽象、高级的特征。

本节小结

- 1. 单层感知器(1958)是首个具备自主学习能力的神经网络,仅能解决线性可分问题;异或问题暴露其固有局限,引发第一次寒冬。
- 2. 多层感知器通过新增隐藏层突破线性分类限制,具备万能逼近理论能力。隐藏层的作用是绘制、组合多条分界线,实现复杂非线性分类。

3. 反向传播算法(1986)解决多层网络训练的难题,通过前向传播计算误差、反向传播计算梯度并更新权重,循环往复,使多层网络自动学习数据规律。

4. 卷积神经网络(CNN)通过局部连接、权值共享,提升图像数据的处理效率,是计算机视觉领域的核心技术工具。

5. 深度指网络层数多,每层提取比上一层更高级的特征,实现简单到复杂的层次化特征提取。

思考题

1. [识记] 解释单层感知器无法解决异或问题的原因,说明多层感知器如何通过隐藏层克服该局限的。

2. [理解] 有观点认为明斯基《感知器》杀死神经网络。结合本任务讨论,分析批评合理之处、误导之处,及对学术批评的启示。

3. [应用] CNN 的局部连接、权值共享设计借鉴了生物视觉系统。请借鉴人类听觉系统的生理特点,给出设计音频识别神经网络的核心思路。(提示:人的耳蜗对不同频率的声音有不同的“感受区域”)

任务五 Hopfield 神经网络

◆ 引导思考

我们在上一任务中学习的前馈网络,信息只能单向向前传递。生活里你是否常有这样的体验:只凭一张静态的照片,大脑就能自动补齐完整的场景、声音与气味,拼凑出一段完整的回忆。那么,仅依靠前馈网络能不能实现从部分恢复整体的效果?

倘若让网络的输出结果回流、反复迭代作用于输入,会催生出哪些全新的能力?

物理学中的能量最小化原理与神经网络之间存在何种关系? 网络的记忆模式与物理系统的稳定状态之间,又有哪些深层相通之处?

一、从前馈到反馈

在前馈网络中,信息只能从输入层单向流向输出层,不存在反向传递路径,适合给定输入、直接计算对应输出的分类、预测类任务。反观人类大脑的信息处理机制,脑中神经元存在海量反馈连接,信息能够来回交互传递,最终收敛至稳定状态。

1982年,美国物理学家约翰·霍普菲尔德(John Hopfield)提出 Hopfield 网络,这是一种全连接反馈网络:网络内所有神经元两两互连,但神经元不与自身建立连接;网络没有严格划分的独立输入层与输出层,信息支持双向传递,每一个神经元同时承担输入、输

出两种功能。

前馈网络好比单向通行的单行道,数据流只能一路向前;Hopfield 网络则类似一场圆桌讨论会,每个神经元都会输出信号、互相影响,经过多轮意见交换后,整体状态达成一致,网络趋于稳定。

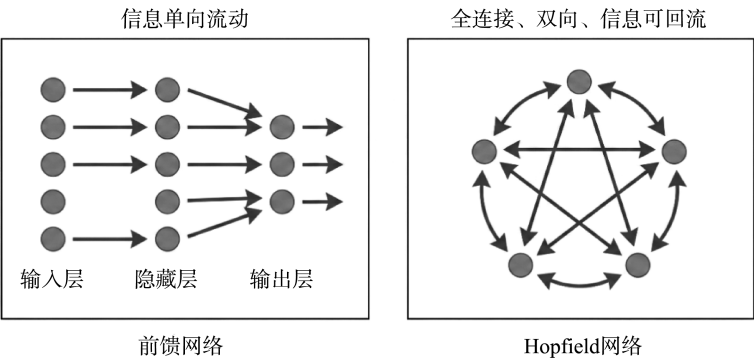


图 3-10 前馈网络与 Hopfield 网络的结构对比

表 3-6 前馈网络与 Hopfield 网络的对比

对比维度	前馈网络(BP 网络)	Hopfield 网络
信息流向	单向(输入→输出)	双向(全连接,可回流)
网络结构	分层(输入层/隐藏层/输出层)	单层全连接(不分层)
核心能力	分类、回归、函数逼近	联想记忆、优化求解
学习方式	监督学习(需要标注数据)	无监督(按赫布规则设定权重)
工作方式	一次前向计算即出结果	反复迭代直到状态稳定
理论工具	微积分(梯度下降)	物理学(能量函数、统计力学)
通俗比喻	“从入口到出口的流水线”	“反复讨论直到达成共识的圆桌会”

二、联想记忆

Hopfield 网络最核心的能力是联想记忆(Associative Memory):向网络输入残缺、带噪声的局部样本,网络能够自动迭代还原出与之匹配、完整清晰的记忆模式。

生活中这类场景随处可见:打电话时只听到半句话,大脑就能补齐完整对话;看到一张被咖啡渍遮挡局部的旧照片,我们依然能辨认出照片里的人物形象。Hopfield 网络正是用数学模型复刻了这种“由局部推全貌”的记忆逻辑。

联想记忆的工作过程分为两个完整阶段:首先是训练存储阶段,该阶段将多组完整记忆模式(例如字母 A、B、C 的标准图案)存入网络,依据样本特征设定神经元间的连接权重;计算规则为将记忆中同步激活的神经元数值相乘,叠加至二者对应的连接权重。接着是回忆复原阶段,该阶段向网络输入残缺、受损的样本(例如缺角的字母 A),神经元之间

反复传递信号、迭代交互,网络状态会持续调整,最终收敛至能量最低的稳定状态,也就是与残缺样本最贴合的完整记忆(标准字母 A)。

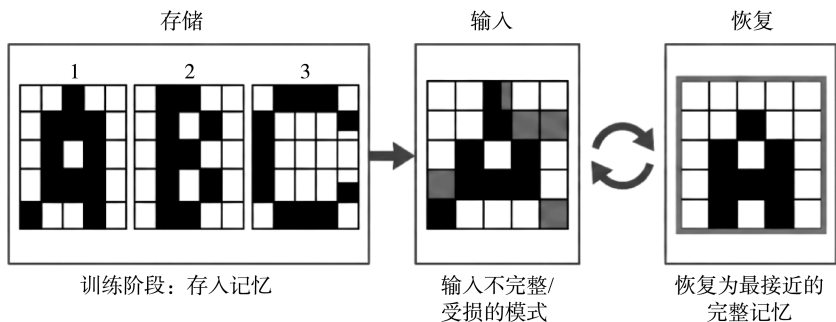


图 3-11 Hopfield 网络的联想记忆过程

三、能量函数

Hopfield 网络最巧妙的设计,是借用物理学的能量概念描述网络整体运行规律。霍普菲尔德本身是物理学家,他为网络定义了专属能量函数:网络每一种运行状态,都会对应一个固定能量数值;该函数具备核心特性——无论神经元如何更新自身状态,网络的总能量只会保持不变或持续降低,永远不会升高。

我们可以用能量地形图直观理解:横轴代表网络的全部可能状态,纵轴代表对应能量大小,整张地形图分布着高低不一的山峰与谷底。网络从初始输入状态启动,就像一颗小球从山坡释放,只会持续向下滚动,最终停留在谷底;每一处谷底,都对应网络一个稳定记忆状态。

联想记忆的全过程可以对应这套地形模型:

存储记忆,相当于在能量地形中人工挖出谷底,通过调整权重,在目标记忆样本的位置创造能量最低点;

回忆复原,相当于把小球放在任意初始坡位,小球会依靠重力自然滚落至距离最近的谷底,对应网络收敛还原出完整记忆。

能量函数的引入,让 Hopfield 网络的运行规律能够借助成熟的物理学工具定量分析。这项工作的深远价值,在于打通了神经网络与物理学的研究边界,后续玻尔兹曼机等经典模型,均诞生于这条跨学科研究思路。

知识拓展: 霍普菲尔德 —— 以物理学家视角重构神经网络

约翰·霍普菲尔德的本职是物理学家,并非计算机或神经科学研究者。他的突破性发现,恰恰源于这种跨学科的旁观者视角:他发现神经科学家早已熟知的神经元交互规律,与物理学中的自旋玻璃(磁性材料)系统拥有完全一致的数学表达。

自旋玻璃系统里,每一个原子磁矩(自旋)都会受周边原子影响,整个系统会自发演化至能量最低的稳定状态。霍普菲尔德建立了一一对应的类比关系:原子对应神经元、自旋朝向对应神经元激活状态、原子间磁力相互作用对应神经元突触权重,由此借用自旋玻璃的成熟数学体系解析神经网络。这套跨学科洞察,正是 Hopfield 网络诞生的核心思想来源。

2024 年,霍普菲尔德因基于人工神经网络开创机器学习基础理论而斩获诺贝尔物理学奖,这也是物理学与人工智能交叉领域的最高学术认可。

四、Hopfield 网络的应用与局限

1. 核心应用:联想记忆、组合优化求解

第一类典型应用是联想记忆场景,例如带噪图像修复、残缺文字补全、破损数据复原等。虽然如今深度学习模型的处理效果远优于传统 Hopfield 网络,但它首次搭建起联想记忆完整理论框架,对后续相关算法发展影响深远。

第二类重要用途是复杂优化问题求解:旅行商最短路线规划这类现实组合优化任务,均可等价转化为“寻找能量最低点”的求解问题。我们只需把优化任务的目标函数映射为 Hopfield 网络的能量函数,网络自主演化至能量谷底时,就能得到问题的近似最优解。1985 年,霍普菲尔德与塔克利用该思路完成 30 座城市的旅行商问题求解,直观证明了神经网络处理组合优化难题的巨大潜力。

2. 固有局限:存储容量上限、局部最优陷阱

Hopfield 网络存在两处无法回避的核心短板:

一是存储容量存在严格上限:若网络包含 N 个神经元,仅能稳定可靠存储约 $0.15 N$ 组独立记忆样本。举例来说,拥有 100 个神经元的网络,最多只能完整记住 15 种模式;一旦强行存入过多样本,不同记忆会相互干扰,回忆阶段极易出现混淆、生成不存在的虚假记忆,这和我们记忆大量电话号码后互相混淆、记混号码的体验十分相似。

二是容易陷入局部最优。结合能量地形图的比喻:小球只会滚落至距离起点最近的谷底,未必是整张地形图里深度最大、能量最低的全局谷底;小球被困在浅层小坑时,就无法找到全局最优解。针对该缺陷的改进方案,将在任务六随机神经网络中详细介绍。

案例:人类大脑每天都在做“联想记忆”

Hopfield 网络的联想逻辑,在日常生活中随处可见:听到歌曲前奏,就能完整回想起整首旋律与歌词;闻到某种熟悉的气味,瞬间回忆起多年前对应的场景;仅凭远处模糊的身影,就能判断出对方身份。

人脑接收局部线索后,会自动匹配脑海中已存储的完整记忆,这套逻辑和 Hopfield 网络依靠能量最小化还原样本的机制高度相似。但人类的联想记忆能力远比 Hopfield 网络复杂、强大:人脑拥有海量神经元,记忆存储容量远远超出 Hopfield 网络的理论上限。即便如此,Hopfield 网络仍是首个用数学公式完整建模联想记忆底层机制的模型,具备不可替代的理论研究价值。

五、Hopfield 网络的历史意义

Hopfield 网络的价值不局限于工程落地应用,它在人工智能发展史中有三大里程碑式贡献:

第一,Hopfield 网络的提出为寒冬中的神经网络研究注入了信心。1969 至 1982 年,受《感知器》一书影响,神经网络的研究陷入低谷,大量研究者失去研究信心;霍普菲尔德借助严谨的物理学理论,严格证明了神经网络具备不可忽视的理论潜力,重新点燃全球学者的研究热情。

第二,Hopfield 网络打通了物理学与神经网络的研究桥梁。首次将能量函数、统计力学引入神经网络分析体系,为玻尔兹曼机、生成式模型等后续重要理论奠定基础。

第三,Hopfield 网络开创了反馈式神经网络研究范式,与前馈网络形成互补。如今广泛用于文本、语音等时序任务的循环神经网络(RNN)、长短期记忆网络(LSTM),核心理念思路均来自“信息回流、自带记忆”的反馈结构;只是网络具体设计存在差异。

本节小结

1. Hopfield 网络属于全连接反馈网络,神经元之间信号双向传递、迭代交互,最终收敛至稳定状态,和前馈网络单向数据流形成鲜明区别。
2. Hopfield 网络的核心功能为联想记忆:输入残缺、带噪声的局部样本,可还原出匹配度最高的完整记忆,运行逻辑与人脑记忆恢复过程相近。
3. 能量函数是分析 Hopfield 网络的核心工具:存储记忆等价于在能量地形创建谷底,回忆样本等价于小球滚落至就近谷底;神经元状态异步更新时,网络总能量只会不变或降低。
4. Hopfield 网络存在两大局限:存储容量上限为 $0.15 N$ (N 为神经元总数)、易陷入局部最优;任务六将介绍引入随机扰动的随机神经网络,以此缓解该问题。
5. Hopfield 网络拥有深远的历史意义:走出神经网络的第一次寒冬、搭建物理与 AI 跨学科研究的桥梁、开创反馈式网络的研究范式;2024 年霍普菲尔德凭借相关成果获得诺贝尔物理学奖。

思考题

1. [识记] 请用“球在能量地形上滚动”的比喻,分别解释 Hopfield 网络中“存储记忆”和“回忆”的过程。
2. [理解] Hopfield 网络的存储容量只有约 $0.15 N$,其中 N 为神经元数量。请思考:如果存在一个 100 个神经元的网络被强制存储了 30 个模式(远超容量),回忆时可能会出现什么情况? 请用“太多电话号码混在一起”的比喻来解释。

3. [应用] 在人们的专业领域或日常生活中,有哪些问题可以被理解为“从不完整的信息恢复完整信息”的联想记忆问题?请举出至少两个例子,并分析 Hopfield 网络的思路是否可以为此类问题提供启发。(提示:考虑文物修复、缺失数据填补、模糊照片清晰化等场景)

任务六 随机神经网络

◆ 引导思考

使用 Hopfield 网络进行寻优计算时,模型很容易卡在局部最优的浅层能量谷底,无法搜寻到全局最优解;通过引入随机扰动能够让模型跳出浅层极小值,完成全局最优的求解,这也是模拟退火策略的核心逻辑。

生活中人们做决策也常常趋于保守,固守现有选择而错失最优方案,适度冒险、主动试错反而更容易得到更优结果,这套随机探索的思路也被应用到神经网络优化当中。

那么,随机扰动为什么能解决复杂优化难题?在处理复杂问题时,带有不确定性的寻优方案,为何反而比固定不变的确定性策略更有优势?

一、为什么需要“随机性”

在任务五中,用“球在能量地形上滚动”来比喻 Hopfield 网络的工作过程:球从高处出发,沿着地形滚向最近的谷底。这个“最近的谷底”不一定是“最深的谷底”。它可能只是局部最优,而真正的全局最优,在更远的地方。由于球只会“往下滚”,一旦落入浅层洼地便无法自行脱出。

怎样才能让球跳出浅坑,继续寻找更深的谷底?一个巧妙的思路是:给球一些随机的“扰动”,比如偶尔把球“弹”起来一点。如果扰动足够大,球就有可能翻过浅坑的边缘,滚向更远处的深谷;如果扰动太小,球仍然会留在原来的浅坑里。关键在于,这种“扰动”应该是逐渐减小的。一开始大幅“弹跳”以便探索更广阔的地形,后来逐渐“安定”以便精确地落入最深的谷底。

这个思路在物理学中有一个著名的对应:模拟退火。在金属冶炼中,“退火”是一种热处理工艺:先将金属加热到很高的温度,此时原子剧烈运动,可以“跳”出不稳定的晶格位置,然后缓慢降温,原子逐渐安定,最终排列成能量最低的稳定晶体结构。如果降温太快,原子来不及找到最优排列,金属就会产生内应力和缺陷;只有足够缓慢的降温,才能获得最优的晶体结构。

将“模拟退火”的思想引入神经网络,就产生了随机神经网络。其中最著名的代表就是玻尔兹曼机。

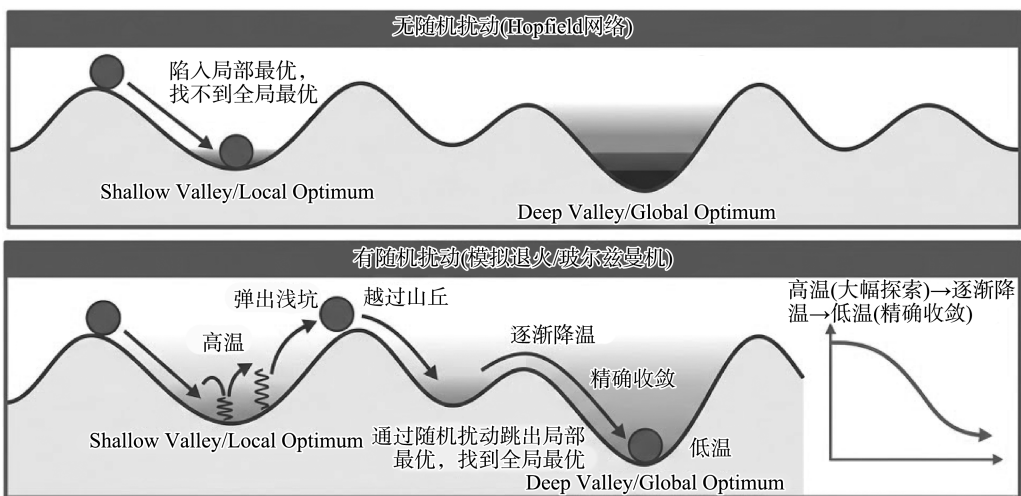


图 3-12 “局部最优”陷阱与“模拟退火”策略

二、玻尔兹曼机

1. 基本思想

玻尔兹曼机(Boltzmann Machine)由辛顿和塞杰诺夫斯基于 1985 年提出,以物理学家路德维希·玻尔兹曼的名字命名。它可以被看作 Hopfield 网络的“升级版”——在 Hopfield 网络的基础上引入了随机性,使网络在寻找最优状态时不再是“确定性地往下滚”,而是“概率性地探索”。

具体来说,玻尔兹曼机中的每个神经元在更新状态时,不是简单地根据输入信号确定性地选择“激活”或“不激活”,而是以一定的概率作出选择。这个概率受到一个“温度”参数的控制:温度高时,神经元的行为更加随机,更容易作出“出人意料”的选择;温度低时,神经元的行为趋于确定性,倾向于选择能量最低的状态。

训练过程就是一个“模拟退火”过程:一开始设置较高的“温度”,让网络大范围地随机探索。然后逐渐降低“温度”,让网络慢慢“收敛”到一个好的状态。这种“先粗后细”的策略使得玻尔兹曼机有更大的概率找到全局最优解,而不是被困在局部最优中。

2. 可见单元与隐含单元

玻尔兹曼机中的神经元分为两类:可见单元和隐含单元。可见单元直接与外部世界交互,它们接收输入数据并给出输出结果,相当于网络的“窗口”。隐含单元则不直接暴露在外,它们在网络内部工作,负责捕捉数据中隐藏的规律和结构,相当于网络的“内部思考过程”。

可见单元就像一家公司的“前台”,直接面对客户。隐含单元就像公司的“后台团队”,客户看不到,但实际上在做大量的分析和决策工作。隐含单元的存在使得玻尔兹曼机能

够学习到数据中更深层次的特征和规律。

三、受限玻尔兹曼机

完全版的玻尔兹曼机虽然理论上很优美,但在实际应用中有一个严重的问题:训练速度极慢。因为所有神经元之间都相互连接,网络中的状态变化极其复杂,训练所需的计算量随着网络规模呈指数级增长。

为了解决这个问题,辛顿等人提出了一种简化版本——受限玻尔兹曼机(Restricted Boltzmann Machine,RBM)。“受限”体现在两个方面。一是将神经元严格分为可见层和隐含层两层;二是限制连接方式:可见层和隐含层之间有连接,但同一层内的神经元之间没有连接。这种“层间有连接、层内无连接”的结构大大简化了训练过程,使得高效的学习算法成为可能。

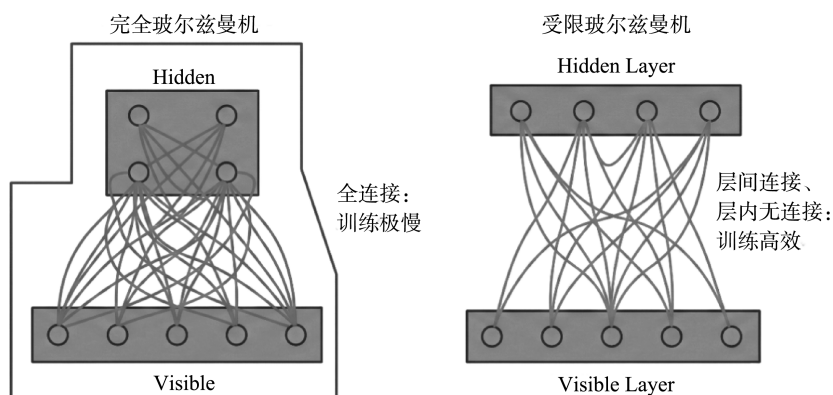


图 3-13 完全玻尔兹曼机与受限玻尔兹曼机的结构对比

受限玻尔兹曼机在神经网络的发展历史中扮演了重要的角色。2006 年,辛顿提出了将多个 RBM 逐层堆叠来构建深度信念网络(DBN)的方法。先用无监督的方式逐层预训练每一层 RBM,然后再用反向传播算法进行整体微调。这一方法有效地解决了当时深层神经网络难以训练的问题,标志着“深度学习”概念的正式提出。虽然后来更先进的训练技巧(如 ReLU 激活函数、批量归一化等)使得直接训练深层网络成为可能,逐步取代了 RBM 预训练的方法,但 RBM 在深度学习发展史上的“破冰”贡献不可磨灭。

案例：“模拟退火”在日常决策中的影子

模拟退火的“先大胆探索后逐渐精细化”策略,其实在日常决策中随处可见:

找工作:刚开始找工作时(高温阶段),求职者可能会广泛投递简历,尝试不同行业和岗位;随着面试经验的积累,求职者逐渐缩小范围,聚焦到最适合自己的方向;最后做出最终选择。如果一开始就只投一家公司,求职者可能找到了“还不错”的工作,但错过了“最好”的机会。

旅行规划:规划一次旅行时,先粗略地浏览大量目的地,确定几个候选后再深入比较,最后敲定具体行程。

学术研究:一个好的研究者在选题阶段会广泛阅读不同领域的文献,逐渐聚焦到一个有前景的方向,最终深入钻研。

这种“先发散后收敛”的策略之所以有效,核心原因是:在问题空间很大的时候,局部搜索很容易陷入“局部最优”;适度的“随机性”可以跳出固有思维定式,发现更优的全局方案。

表 3-7 模块三主要神经网络类型总览

网络类型	结构特点	核心能力	学习方式	典型应用	关键局限
单层感知器	单层前馈,无隐藏层	线性分类	监督(误差修正)	简单模式分类	无法处理非线性问题
多层前馈/BP网络	多层前馈,有隐藏层	任意复杂的分类/回归	监督(反向传播)	图像识别、语音识别	需大量标注数据
卷积神经网络(CNN)	局部连接、权值共享	图像特征提取	监督(反向传播)	人脸识别、医学影像	主要适用于网格数据
Hopfield 网络	全连接反馈网络	联想记忆、优化求解	无监督(赫布规则)	模式恢复、旅行商问题	容量有限、局部最优
玻尔兹曼机/RBM	随机反馈/分层网络	概率建模、特征学习	无监督(概率学习)	深度网络预训练	训练复杂(完全版)

知识拓展:从随机网络到生成模型——“不确定性”的力量

玻尔兹曼机引入的“概率性”思想,在后来的 AI 发展中结出了丰硕的果实。它启发了一系列基于概率和随机性的生成模型,这些模型不仅能“识别”数据中的模式,还能“生成”全新的、高逼真度的数据。

例如,近年来大热的图像生成技术的核心思想就与模拟退火有着深刻的联系:先给一张“纯噪声”的图片,然后逐步“去噪”,最终生成一张清晰的图像。这种“从混沌中涌现秩序”的过程,正是随机性和确定性巧妙结合的典范。

本节小结

1. 确定性网络(如 Hopfield 网络)容易陷入“局部最优”。引入随机扰动可以帮助网络跳出浅谷底,寻找全局最优。这就是“模拟退火”策略的核心思想。
2. 玻尔兹曼机是将模拟退火思想引入神经网络的代表,其神经元以概率方式更新状态,“温度”参数控制随机性的强弱:高温时大胆探索,低温时精确收敛。
3. 受限玻尔兹曼机(RBM)通过限制连接方式大幅提高了训练效率,并在 2006 年成为构建深度信念网络、推动深度学习“破冰”的关键技术。
4. 模块三介绍的五种主要网络类型(感知器、BP 网络、CNN、Hopfield 网络、玻尔兹

曼机/RBM)各有特长,适用于不同类型的问题。

思考题

1. [识记] 请用“球在能量地形上滚动”的比喻,解释 Hopfield 网络为什么会陷入“局部最优”,以及玻尔兹曼机是如何通过“随机扰动”来克服这一问题的。

2. [理解] 模拟退火策略的核心是“先高温探索,后低温收敛”。如果反过来,先低温再高温,会发生什么? 为什么这种顺序不合理? 请结合“找工作”的比喻来说明。

3. [应用] 回顾本模块学习的所有神经网络类型,包括感知器、BP 网络、CNN、Hopfield 网络、玻尔兹曼机,请判断以下每个任务最适合使用哪种网络,并说明理由:
a. 识别手写数字图片;b. 从一张模糊的旧照片中恢复出清晰的人脸;c. 为一个有 50 个城市的物流公司规划最短送货路线。

思想引领

一、从物理学到神经网络的启示

本模块的学习呈现了一个令人印象深刻的跨学科图景:神经网络的灵感来自生物学的神经元和突触的结构,Hopfield 网络的理论工具来自物理学的能量函数和统计力学,玻尔兹曼机的训练策略来自材料科学的退火工艺,而这一切最终以数学和计算机科学为载体得以实现。

这种跨学科交叉正是人工智能领域最迷人的特质之一。霍普菲尔德是一位物理学家,辛顿的早期研究受到了认知心理学的深刻影响,而“深度学习三巨头”之一的杨立昆最初的学术训练是电气工程。AI 的每一次重大突破,几乎都发生在不同学科的“交汇处”。

对于非计算机专业的学生来说,这是一个特别值得关注的信号:不管专业背景是物理学、生物学、心理学、经济学还是艺术设计,都可能为 AI 技术的发展提供独特的灵感和视角。未来 AI 领域最具创新性的突破,很可能来自那些能够将多个学科的知识创造性地结合起来的人。

二、“黑箱”问题

在任务四中了解到,多层神经网络尤其是深度网络虽然性能强大,但有一个广受关注的问题:可解释性差。一个深层神经网络可能包含数百万甚至数十亿个参数,它的决策过程隐藏在这些参数的复杂交互之中,即使是设计者也很难准确解释“模型为什么作出了这个判断”。这被称为“黑箱”问题。

在某些应用场景中,黑箱问题并不严重。比如推荐系统推荐了一首用户不喜欢的歌,后果很小。但在另一些场景中,它可能带来严重的伦理和法律风险:如果一个基于神经网络

络的医疗诊断系统给出了“建议手术”的判断,但无法解释为什么,医生应该信任它吗? 如果一个 AI 系统拒绝了某人的贷款申请,申请人有权要求系统解释原因吗?

“可解释 AI”已经成为当前 AI 研究中最活跃的方向之一。不同类型的神经网络在可解释性上存在显著差异:决策树天然可解释,感知器的决策边界也比较直观,但深层网络的内部机制至今仍难以完全解读。在选择 AI 技术时,需要根据应用场景对“可解释性”和“性能”进行权衡。

三、“涌现”的哲学

本模块最深刻的思想启示或许在于:单个人工神经元极其简单,不过是“加权求和+激活函数”,但大量简单神经元的连接与协作,却能涌现出令人惊叹的复杂智能。识别人脸、理解语言、下围棋——这些看似“高级”的智能行为,竟然可以从“低级”的数学运算中“长”出来。

这种“从简单到复杂”的涌现现象在自然界中无处不在:单个蚂蚁的行为极其简单,但蚁群能够建造精巧的巢穴、开辟高效的觅食路线。单个水分子没有“湿”的属性,但大量水分子聚在一起就有了“湿”的特性。单个神经元不具备“思考”的能力,但 860 亿个神经元的协作产生了人类的意识和智慧。

理解“涌现”有助于以更谦逊的态度面对复杂系统。不要试图用简单的因果链去解释一切,也不要因为组成部件很简单就低估整体的能力。这种思维方式,不仅适用于理解神经网络,也适用于理解生态系统、社会组织、经济市场等一切复杂系统。

◆ 任务实践

以下实践任务旨在帮助读者将本模块所学的神经网络理论知识与直觉理解相结合,通过动手体验、分析和讨论,加深对神经网络结构、学习过程和应用的认识。建议以小组(3—5 人)为单位完成。

实践一 “人体神经网络”课堂模拟实验

任务说明

目标:通过全班同学的协作,用“人体”模拟一个简单神经网络的前向传播和反向传播过程,直观感受网络的工作方式。

准备:全班分为 3 组,分别担任“输入层”(3 人)、“隐藏层”(4 人)和“输出层”(2 人)。每个人代表一个神经元。准备一些数字卡片和计算器。

步骤:

(1) 前向传播:教师给“输入层”每人发一个数字(输入值)。每个“隐藏层”成员从自己连接的“输入层”成员那里收集数字,用预设的权重(写在纸条上)进行加权求和,然后通过一个简单的激活规则(如“大于 0 就输出 1,否则输出 0”)得到输出值,传递给“输出层”。

“输出层”同理计算,得到网络的最终输出。

(2) 计算误差:教师宣布“正确答案”,全班计算输出误差。

(3) 反向传播:从“输出层”开始,根据误差调整每条连接的权重(用简化的规则,如“如果输出太大就减小权重,太小就增大权重”)。然后向前传递调整信号,让“隐藏层”也相应调整。

(4) 重复:用新的输入重复以上过程 2—3 轮,观察网络的输出是否越来越接近正确答案。

讨论:a. 上述“人体网络”的前向传播和反向传播过程,与任务四中描述的 BP 网络有什么对应关系? b. 在模拟过程中你有什么感受? 对理解神经网络有什么新的认识?

提示:这个实验不追求数学上的精确,目的是让每位同学“身临其境”地体验信息在网络中的流动和权重的调整过程。

评价维度:计算与传递的规范性、团队协作与参与度、原理理解与反思深度。

实践二 “联想记忆”纸笔实验

任务说明

目标:通过纸笔实验,直观体验 Hopfield 网络“联想记忆”的核心原理。

准备:每人准备一张 5×5 的方格纸(共 25 个格子)。

步骤:

(1) 存储“记忆”:每人在方格纸上用黑色和白色(或打勾和留空)画出一个简单的图案(如字母 T、L、十字形等),这就是自己的“存储记忆”。

(2) 制造“损坏”:将方格纸交给旁边的同学,由他随机翻转其中 5 个格子(黑变白或白变黑),模拟“信息损坏”。

(3) 尝试“恢复”:拿到损坏图案的同学,尝试猜测原始图案是什么,并修复被翻转的格子。

(4) 对比:与原图案对比,看恢复的准确率。

讨论:a. 你是根据什么“规律”来修复图案的? 该“规律”与 Hopfield 网络寻找“能量最低点”有什么相似之处? b. 如果同时记忆了多个图案,恢复会变得更加容易还是更困难? 这与任务五中讨论的“存储容量限制”有什么关系? c. 如果翻转的格子数量更多(比如翻转 15 个),恢复会变得怎样? 这说明了什么?

评价维度:修复策略的合理性、干扰对比实验的完成度、与 Hopfield 网络原理的类比能力。

实践三 “神经网络家族”知识图谱

任务说明

目标:整理和可视化本模块所学的所有神经网络类型,建立系统性的知识结构。

步骤:

(1) 每个小组绘制一张“神经网络家族”的知识图谱(可以是树形图、思维导图或关系图),包含本模块介绍的所有网络类型:单层感知器、多层感知器(BP 网络)、卷积神经网络(CNN)、Hopfield 网络、玻尔兹曼机、受限玻尔兹曼机(RBM)。

(2) 对每种网络标注以下信息:a. 结构特点(前馈/反馈、层数、连接方式);b. 核心能力(分类/回归/联想记忆/特征学习);c. 学习方式(监督/无监督);d. 一个典型应用;e. 一个核心局限。

(3) 在图谱中用箭头标注各网络之间的“演化关系”(如:感知器的局限→催生多层感知器;Hopfield 网络的局部最优问题→催生玻尔兹曼机;RBM 的逐层预训练→推动深度学习的诞生)。

(4) 课堂展示并互评:每组 5 分钟展示,其他小组提出补充或疑问。

评价维度:知识的完整性和准确性、网络之间关系的梳理深度、可视化表达的清晰度。

实践四 “AI 中的跨学科”小组研讨

任务说明

目标:围绕本模块思想引领中提出的“跨学科思维”主题,展开深入的调研和讨论。

步骤:

(1) 每个小组选择以下一个主题进行调研:

主题 A:神经网络中的“生物学启发”——除了神经元和突触,还有哪些 AI 技术借鉴了生物学的思想?(如遗传算法模拟进化、蚁群算法模拟蚂蚁觅食等,将在模块七群体智能中进一步讨论)

主题 B:神经网络中的“物理学启发”——除了 Hopfield 网络的能量函数和玻尔兹曼机的模拟退火,物理学还为 AI 提供了哪些重要思想?(如扩散模型与热力学的关系)

主题 C:“黑箱”问题的伦理思考——在医疗诊断、司法判决、信贷审批等高风险场景中,使用不可解释的深度神经网络是否合理?应该如何平衡“性能”和“可解释性”?

(2) 小组进行课外调研(可参考网络资源和本教材后续模块的内容预览),撰写不超过 1 500 字的研讨报告。

(3) 课堂汇报(每组 8 分钟)+全班讨论(10 分钟)。

评价维度:调研的广度和深度、跨学科联系的建立能力、批判性思考的质量。