第 1 章
人工智能概述

机器人是人类的古老梦想。古希腊神话中已经出现了机器人，至今机器人仍然是众多科幻小说的重要元素。实现这个梦想的第一步是了解如何将人类的思考过程形式化和机械化。科学家们被这一梦想深深吸引，开始研究记忆、学习和推理。在 20 世纪 30 年代末到 50 年代初，神经学研究发现大脑是由神经元组成的电子网络，克劳德·香农提出的信息论则描述了数字信号，图灵的计算理论证明了一台仅能处理 0 和 1 这样简单二元符号的机械设备能够模拟任意数学推理。这些密切相关的成果暗示了构建电子大脑的可能性。在 1956 年的达特茅斯会议上，“人工智能”（Artificial Intelligence，AI）一词被首次提出，其目标是“制造机器模仿学习的各个方面或智能的各个特性，使机器能够读懂语言，形成抽象思维，解决人们目前的各种问题，并能自我完善”。这也是我们今天所说的“强人工智能”的概念，其可以理解为，人工智能就是在思考能力上可以和人做得一样好。今天所说的“弱人工智能”是指只处理特定问题的人工智能，如计算机视觉、语音识别、自然语言处理，不需要具有人类完整的认知能力，只要看起来像有智慧就可以了。一个弱人工智能的经典例子就是那个会下围棋并且仅仅会下围棋的 AlphaGo。

虽然强人工智能仍然是人工智能研究的一个目标，但是强人工智能算法还没有真正的突破。大多数的主流研究者希望将解决局部问题的弱人工智能的方法组合起来实现强人工智能。业界的共识是，大部分的应用都是弱人工智能（如监督式学习），实现近似人类的强人工智能还需要数十年，乃至上百年。在可见的未来，强人工智能既非人工智能讨论的主流，也看不到其成为现实的技术路径。弱人工智能才是这次人工智能浪潮中真正有影响力的主角，本书将聚焦于更具有现实应用意义的弱人工智能技术。

从政府到资本、业界都热情拥抱人工智能，以人工智能驱动的智能化变革正在引发第 4 次工业革命。虽然人工智能在某些方面还处于炒作周期的顶峰，但我们可以预测，人工智能正变得更加实用和有用。在此大背景下，我们有必要知道人工智能是什么、火在哪里、是否已经成熟。人工智能技术的壁垒在哪里？了解商业化的边界在哪里，才能更好地理解人工智能。

1.1 AI 是什么

人工智能是一门利用计算机模拟人类智能行为科学的统称，它涵盖了训练计算机使其能够完成自主学习、判断、决策等人类行为的范畴。AI 是人工智能的英文 Artificial Intelligence 的首字母的组合，它是当前人类所面对的最为重要的技术变革。AI 技术给予了机器（这里的机器不仅仅指机器人，还包括消费产品，如音箱、汽车等范围更广的物体）一定的视听感知和思考能力。例如，苹果 Siri 和亚马逊 Echo 智能音箱可以帮助我们通过语音控制的方式设置闹钟、播放音乐、回复信息、询问天气，还可以聊天；滴滴出行和 Uber 应用也是在人工智能技术的驱动下帮助司机选择最佳路线。

除了日常生活外，人工智能在工业、金融、安防、医疗、司法等领域也发挥了巨大的作用。工业机器人代替人类完成焊接、铸造、装配、包装、搬运、分发货物等单调、重复、繁重的工作；在金融领域，人工智能技术可以帮助金融机构提供投资组合建议，创建高精度的风险控制模型，实现精准营销等金融活动；对于安防行业，以图像识别、人脸识别为代表的人工智能技术对摄像头获取的海量视频信息进行解析，已被广泛应用于门禁系统、车辆检测、追踪嫌犯等场景中，对增强安防水平、维护社会稳定、提高刑侦效率等都有重大意义；在医疗领域，IBM 的人工智能系统 Watson（沃森）已被多家医疗机构采用，它可以帮助医生更快、更准确地诊断疾病，还能提出对医疗方案的疗效及风险的评估，这将有效地弥补有些地区医疗资源不足的缺陷；美国人工智能律师 Rose Intelligence 可以理解律师向它提出的问题，收集已有的法律条文、参考文献和法律案件等数据，进行推论，给出基于证据的高度相关性答案，这样的系统可以减少法律服务成本，使更多的人能够获得法律帮助。

1.1.1 火热的 AI

人工智能发展到今天已经有 60 多年了。它实际上经历了三个阶段：第一个阶段，1956 年到 1976 年，注重逻辑推理。第二个阶段，从 1976 到 2006 年，以专家系统为主。2006 年起进入重视数据、自主学习的认知智能时代，这是第三个阶段，它会持续多长时间，没有人知道。

最近几年，在算法、大数据、计算力等技术的推动下，人工智能开始真正解决问题，在各行业的应用场景逐渐明朗，并带来实际的商业价值。目前，无论在学术界、投资界，还是在职场，AI 异常火热。AI 论文发表数量激增：自从 1996 年以来，每年发表的 AI 论文数量增加了 9 倍以上，如图 1-1 所示。斯坦福大学入学选修人工智能和机器学习入门课程的学生人数从 1996 年以来增长了 11 倍以上。在美国，有资本投资的 AI 创业公司数量从 2000 年以来增加了 14 倍，如图 1-2 所示。在美国，投资 AI 创业的基金数量也在增长，从 2000 年以来，每年投入 AI 创业的资本额增加了 6 倍。美国最近几年中，每年都有几十亿美元的风险资本（VC）进入 AI 领域，人工智能相关岗位的需求也在急剧增长。图 1-3 展示了 Indeed.com 平台上，从 2013 年 1 月份起，AI 技术相关工作岗位的份额的增长。在开源软件使用和生态上，AI 软件也是异

常火热的。图 1-4 展示了 AI 各个软件包在 GitHub 上加星标的次数。排在第一的 TensorFlow 是排在第二的 scikit-learn 的 4 倍左右。本书的例子都是在 TensorFlow 和 scikit-learn 软件包上实现的。

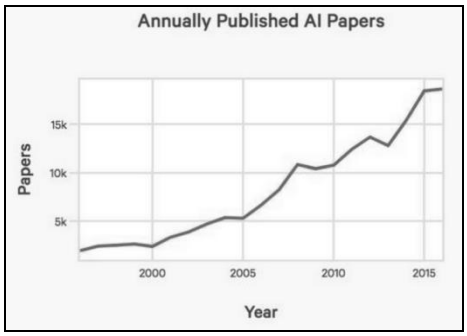


图 1-1　AI 学术论文每年发表情况

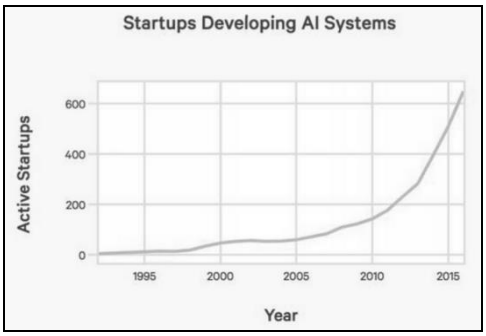


图 1-2　美国 AI 创业公司数量

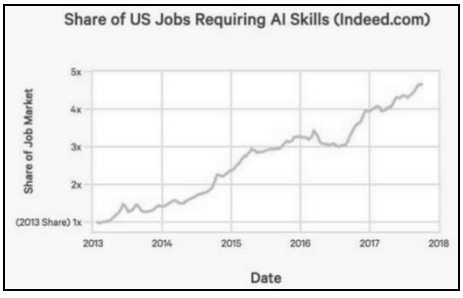


图 1-3　需要 AI 技能的工作岗位

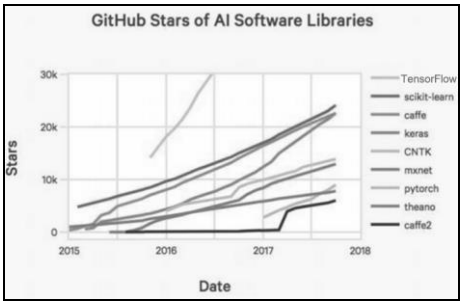


图 1-4　AI 开源软件

1.1.2　AI 的驱动因素

某著名咨询公司预计到 2025 年，全球 AI 市场规模将达到 3 万亿美元。AI 持续火热的驱动力主要来自于技术本身的提高，包括数据、算法、计算力、大数据和物联网等技术，而这些正是人工智能技术发展的基础。

- 高质量和大规模的海量数据使得 AI 成为可能

海量数据为 AI 技术的发展提供了充足的原材料。表 1-1 展示了数据量与医疗图像准确性的关系，表明了训练数据量越大，准确性越高。

表 1-1　训练数据量与医疗图像模型准确性的关系（%）

训练数据集大小	5GB	10GB	50GB	200GB
大脑识别	0.3%	3.39%	59.7%	98.4%
脖颈识别	21.3%	30.63%	99.34%	99.74%
肩部识别	2.98%	21.39%	86.57%	92.94%
胸腔识别	23.39%	34.45%	96.18%	99.61%
腹部识别	0.1%	3.23%	65.38%	95.18%
胯部识别	0%	1.15%	55.9%	88.45%
平均准确性	8.01%	17.37%	77.15%	95.67%

- 计算力提升突破瓶颈

以 GPU 为代表的新一代计算芯片提供了更强大的计算力，使得运算更快。同时，在集群上实现的分布式计算帮助 AI 模型可以在更大的数据集上快速运行。

- 机器学习算法取得重大突破

以多层神经网络模型为基础的算法，使得机器学习算法在图像识别等领域的准确性取得了飞跃性的提高。

- 物联网和大数据技术为 AI 技术的发展提供了关键要素

物联网为 AI 的感知层提供了基础设施环境，同时带来了全面的海量训练数据。大数据技术为海量数据在存储、清洗、整合方面提供了技术保障，帮助提升了深度学习算法的性能。

1.2 AI 技术的成熟度

顾名思义，AI 就是能够让机器做一些之前只有“人”才做得好的事情。主要集中在这几个领域：视觉识别（看）、自然语言理解（听）、机器人（动）、机器学习（自我学习能力）等。在技术层面，AI 分为感知、认知、执行三个层次。感知技术包括机器视觉、语音识别等各类应用人工智能技术获取外部信息的技术，认知技术包括机器学习技术，执行技术包括人工智能与机器人结合的硬件技术以及智能芯片的计算技术。这些领域目前还比较散，它们正在交叉发展，走向统一的过程中。

很自然地，我们会在同一个任务上将 AI 系统和人类的表现进行比较。在某些任务中，计算机比人类要优秀得多，例如，70 年代的小计算器就可以比人类更好地完成算术运算。但是，AI 系统在处理诸如回答问题、医学诊断等更通用的任务时更加困难。AI 系统的任务往往是在非常窄的背景下进行的，这样能在特定的问题或应用上取得进展。虽然机器在特定的任务上表现出卓越的性能，但是有时任务稍微有所改动，系统性能就会大大降低。

1.2.1 视觉识别

以图像识别和人脸识别为代表的感知技术已经走向了应用市场，特别是在交通、医疗、工业、农业、金融、商业等领域，带动了一批新业态、新模式、新产品的突破式发展，带来了深刻的产业变革。苹果公司的 iPhone 手机包含 Face ID、A13 芯片等 AI 技术。苹果的 Face ID 技术有人脸验证功能。iPhone 的顶部集成了实现 Face ID 功能的器件，包括红外镜头、泛光感应元件、点阵投影器和普通摄像头。从原理上讲，当红外摄像头发现一张面孔时，点阵投影器会闪射出 3 万个光点，接着红外摄像头会捕捉这些光点的反馈，从而采集一张人脸的 3D 数据模型，并与 A13 芯片中存储的模型进行比对。如果互相匹配，就可以解锁了，iPhone 随即被唤醒。为了更加精确地进行人脸识别，苹果芯片中包含了一个神经引擎，用神经网络处理图像和

点阵模式，并邀请好莱坞特效面具公司制作面具来训练神经网络，以保证安全性。The Verge（美国科技媒体网站）曾借用了一台具有夜视功能的摄像机，成功拍摄到这些肉眼不可见的红外光点，可以看到这 3 万个光点非常密集，不只是投射至人脸，连衣服上也有，视觉效果极其震撼。

如图 1-5 所示，在大规模视觉识别挑战赛（LSVRC）比赛中，图像标签的错误率从 2010 年的 28.5%下降到了 2.5%，AI 系统对物体识别的性能已经超越了人类。在国内，视觉与图像领域的融资排在第一，在整个 AI 投资中占比 23%（数据来源：腾讯的《中美两国人工智能产业发展报告》），说明国内投资者非常看好这一领域。

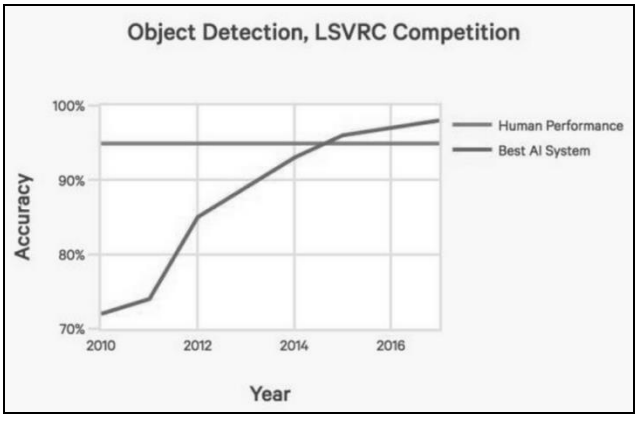


图 1-5　物体识别能力比较（直线为人类，曲线为 AI）

1.2.2　自然语言理解

自然语言理解是指机器接受人类提问的语音输入，先通过语音识别将人类语音转化为文字，再运用自然语义分析理解人类提问的含义（即理解人类的行为），最后反馈给人类以所提问相关的精准搜索结果，其核心技术在于用自然语义分析来理解人类日常说话中的提问。在词语解析方面，AI 系统在确定句子语法结构上的能力已经接近人类能力的 94%。在从文档中找到既定问题的答案的能力已经越来越接近人类（见图 1-6 左图）。AI 系统识别语音录音的表现早在 2016 年就已经达到了人类水平（见图 1-6 右图）。

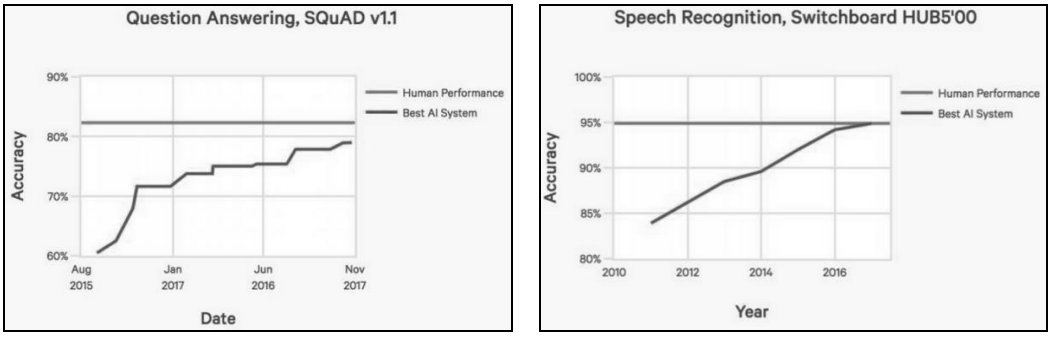


图 1-6　问答准确性比较（左图）和语音识别能力比较（右图：直线为人类，曲线为 AI）

从 PC 互联网到移动互联网再到 AI 时代，每个时代都伴随着一次交互式的变革。利用语音识别、自然语言处理和自然语言理解等技术研发的对话机器人，正在改变着传统的人机交互方式。它们或内嵌到应用程序中，或与硬件相结合，致力于成为用户的个性化“助理”。目前，这些“助理”已经具备了基本的问答、对话以及上下文理解功能。它们正在打造全新的人机交互方式，为用户提供多场景的便捷服务。例如，智能音箱是最近几年美国消费中的热门产品。虽然语音交互的老大依然是苹果公司的 Siri，但是 Amazon Alexa 正在快速崛起（见图 1-7 左边的产品），它不仅可以对话应答，还可以和多种智能家居设备进行交互，比如：语音关灯等。谷歌的智能音箱产品（见图 1-7 中间的产品），功能类似 Alexa。苹果也于 2018 年 2 月 9 日正式上市 HomePad 智能音箱（见图 1-7 右边的产品）。

图 1-7　智能音箱产品

语音交互可以说是人与机器“交流”的重要环节，这对于未来的人工智能而言是非常关键的入口。在国内，自然语音处理领域的融资排在第二，在整个 AI 投资中占比 19%。国内企业中，京东与科大讯飞公司合作布局了智能音箱，致力于成为家庭控制中心。阿里推出了名叫“天猫精灵”的智能音箱，小米推出了小米 AI 音箱。激烈的音箱之争背后其实是下一代服务入口之争。

搭载百度 DuerOS 的智能硬件产品也在陆续面世。DuerOS 是百度基于 AI 技术打造的对话式人工智能系统。搭载 DuerOS 的设备可让用户以自然语言对话的交互方式（比如“小度小度，我想听陈百强的歌”）实现影音娱乐、信息查询、生活服务、出行路况等多项功能。目前，腾讯的所有语音端都采用自己研发的 AI 技术，而阿里的淘宝、支付宝电话客服、天猫精灵、优酷、虾米音乐等都应用了自己的语音技术。除了使用自家语音技术外，BAT 也在加速对外开放平台，滚动扩张。阿里云、腾讯云小微、百度 DuerOS 平台都开放了语音识别、视觉识别等 AI 技术。百度还宣布语音技术全系列接口永久免费开放。

在谷歌 I/O 大会上，语音助手 Google Assistant 更像人。作为谷歌 AI 用户感观最直接的语音助手，谷歌试图将其打造得更近似人：其一是声音拟人化，其二是对话日常化。I/O 大会现场展示了指令 Google Assistant 预定餐厅座位，然后发出指令的人即可忙自己的事，而 AI 将自行打电话给餐厅，通过多轮对话与餐厅工作人员敲定好时间。在这个展示上，突显的亮点是，对话能力加强，近似日常交流习惯，极大地提高了与机器对话的用户体验。

语音是下一代人机交互的入口，未来语音技术会向各场景渗透。它们不但可以响应用户命令并执行任务，如回答问题、设置闹钟、检查航班行程等，而且与搜索、手机、智能家居等紧密结合。除了产品市场本身之外，争夺未来以语音交互为核心的智能家居生态的入口，是科技巨头纷纷推出智能音箱的重要原因。智能语音这块蛋糕有多大，目前尚未可知。有一点越来越清晰，未来肯定是通过人工智能核心技术+应用数据+领域支持的方式构建垂直入口或行业刚需。

1.2.3　机器人

大部分智能机器人目前还处于产业发展初期，但随着全球人工智能步入第三次高潮期，智能化成为当前机器人重要的发展方向，人工智能与机器人融合创新，进一步提升机器人的智能化程度。智能机器人有自主的感知、认知、决策、学习、执行和社会协作能力。美国波士顿动力公司（Boston Dynamics）的研究重点是像狗一样的细长机器人，它可以爬楼梯，在与人类的拔河中保持住姿势，并可以开门，让其他机器人通过。这些功能不禁让人联想到快速、强大，有时甚至令人生畏的未来机器人。如图 1-8 所示是波士顿动力公司的两款机器人。

图 1-8　波士顿动力公司的两款机器人

从全球范围来看，日本 ASMO Actroid-F 仿人机器人、Pepper 智能机器人、美国 BigDog 仿生机器人等一大批智能机器人快速涌现，巨头企业也纷纷通过收购机器人企业，将智能机器人作为人工智能重要的载体，推动人工智能发展，例如，谷歌相继收购 Schaft、Redwood Robotics 等 9 家机器人公司，在类人型机器人制造、机器人协同等方面积极布局。值得指出的是，机器人进展有时不尽人意。以前日本人常常炫耀他们的机器人能跳舞，结果一个福岛核辐射事故一下子把所有问题都暴露了，发现他们的机器人一点招都没有。美国也派了机器人过去，同样出了很多问题。比如一个简单的技术问题，机器人进到灾难现场，背后拖一根长长的电缆，要供电和传数据，结果电缆就被缠住了，动弹不得。所以，智能服务机器人仍处于产业化起步阶段。

1.2.4　自动驾驶

AI 的智能程度决定了无人驾驶的可靠性，苹果、谷歌、特斯拉、百度等公司持续研发无人驾驶技术。虽然出行环境变化多样，当前的技术水平还无法直接应用于日常上路。但在出行

过程中，人工智能技术已经开始发挥作用，包含行车记录仪、测距仪、雷达、传感器、GPS等设备的ADAS系统，已经可以帮助汽车实时感知周围情况并发出警报，实现高级辅助驾驶，保证用户出行安全。自动驾驶的技术核心包括高精度地图、定位、感知、智能决策与控制四大模块。自动驾驶汽车依托交通场景物体识别技术和环境感知技术，实现高精度车辆探测识别、跟踪、距离和速度估计、路面分割、车道线检测，为自动驾驶的智能决策提供依据。

汽车行业正经历大规模的颠覆，汽车厂商越来越意识到，半自动和全自动驾驶车辆将需要基于AI的计算机视觉解决方案，以确保安全驾驶。特斯拉推出了多款电动车，包括Model S、Model 3（这两个为小轿车）、Model X（SUV）、Semi电动卡车等车型。这些车型配备了半自动化驾驶技术，包括自动制动、车道保持以及车道偏离警告等功能。在国内，自动驾驶/辅助驾驶的融资在整个国内AI投资中占比18%。中国的自动驾驶/辅助驾驶企业虽然只有31家，但融资额却排在第三。

与人类水平相当的无人驾驶可能需要更长时间的测试才能成熟起来，但是，我们预估，在未来几年中，越来越多的汽车厂商和IT公司会进入自动驾驶领域。目前，自动驾驶研究领域基本分为两大阵营：

（1）传统汽车厂商和Mobileye公司合作的“递进式”应用型阵营——“在任何区域里发挥局部功能”，强调“万无一失”的复杂传感器组合（Redundancy in System）识别周围环境。通过低精度导航地图在任何区域实现无人驾驶。

（2）以谷歌、百度以及初创科技公司为主的“越级式”研究型阵营——“在特定区域里发挥全效功能”，强调通过采集某一区域的高精度3D地图信息配合激光雷达在某一区域实现无人驾驶。

但是殊途同归，两大阵营的终极愿景都是：“在任何区域里发挥全效功能。”

1.2.5 机器学习

人的大脑一直是一个未解之谜。人类如何思考，人类的大脑如何工作，智能的本质是什么，是古今中外的哲学家和科学家一直在努力探索和研究的问题。早期的研究者将逻辑视为人类智慧比较重要的特征。让计算机中的人工智能程序遵循逻辑学的基本规律进行运算、归纳或推理，是许多早期人工智能研究者的最大追求。但人们很快发现，人类思考实际上仅涉及少量逻辑，大多是直觉的和下意识的“经验”。基于知识库和逻辑学规则构建的人工智能系统（例如专家系统）只能解决特定的狭小领域问题，很难被扩展到宽广的领域和日常生活中。于是，一些研究者提出了一种全新的实现人工智能的方案，那就是机器学习。

人类的聪明之处就在于可以通过既有的认知触类旁通地推理出未知的问题。如图1-9所示，人类看书（书就是数据）时，依靠自身的思考与学习从书中提炼出智慧；机器学习是让计算机利用已知数据得出适当的模型，并利用此模型对新的情境给出判断的过程。机器学习本质上是一种计算机算法，计算机通过大量样本数据的训练，能够对以后输入的内容做出正确的反馈。训练的过程就是通过合理的试错来调整参数，使得出错率降低，当出错率低到满足预期的时候，就可以拿出来应用了。机器学习分为监督式学习和非监督式学习。

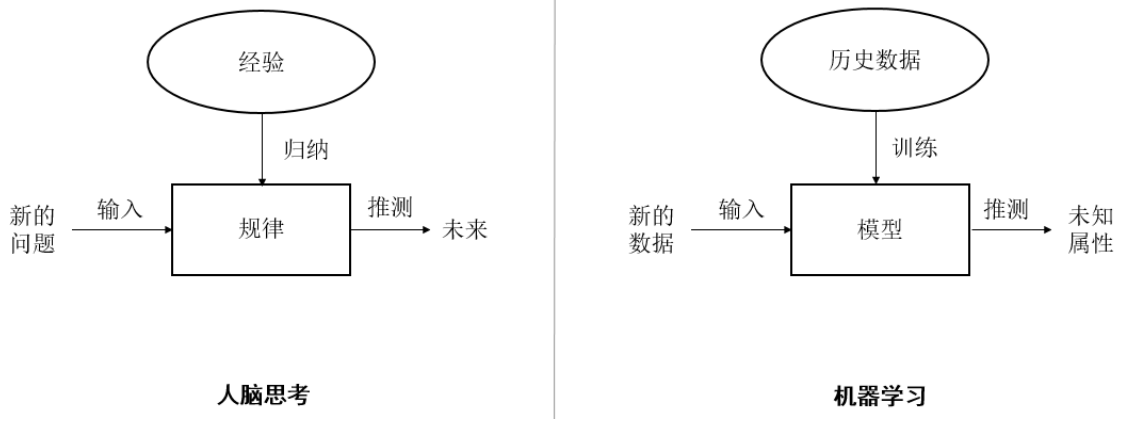


图 1-9　机器学习与人脑思考

机器学习的应用非常广泛，应用在文本方面就是自然语言处理，应用在图像方面就是图像（模式）识别，应用在视频上就是实体识别，应用在汽车上就是自动驾驶，等等。

机器学习重要的成果是 2006 年关于深度学习（Deep Learning）的突破。深度学习起源于 20 世纪八九十年代的神经网络研究。深度学习模型的灵感来自于人类大脑视觉皮层以及人类学习的方式，以工程化方法对功能进行简化。深度学习模型是否精确反映了人类大脑的工作方式还存在争议，但重要的是这一技术的突破让机器第一次在语音识别、图像识别等领域实现了与人类同等甚至超过人类的感知水平，从实验室走向产业，从而发挥价值。

美国大笔投资在机器学习应用上，占美国整个 AI 投资的 21%。这一领域是仅次于芯片的吸金领域（芯片投资的占比为 31%）。机器学习热潮是由三个基本因素的融合推动的：①深度学习算法的持续突破；②大数据的快速增长；③机器学习的计算加速，如 GPU 芯片这样的机器学习硬件，将训练时间从几个月缩短到几天、甚至几个小时。这些硬件芯片正在迅速发展，谷歌、英伟达、英特尔等公司都宣布推出下一代 GPU 芯片硬件，这将进一步加快训练速度的 10~100 倍。

1.2.6　游戏

游戏是一个相对简单和可控的实验环境，因此经常用于 AI 研究。在游戏领域，AI 已超过人类。

1. 国际象棋

20 世纪 50 年代，一些计算机科学家预测，到 1967 年，计算机将击败人类象棋冠军。但直到 1997 年，IBM 的“深蓝”系统才击败当时的国际象棋冠军盖瑞 • 卡斯帕罗夫（Gary Kasparov）。如今，在智能手机上运行的国际象棋程序可以表现出大师级的水平。

2. 围棋

2016 年 3 月，谷歌 DeepMind 团队开发的 AlphaGo 系统击败了围棋冠军。DeepMind 后来发布了 AlphaGo Master，并在 2017 年 3 月击败了排名世界第一的柯洁。2017 年 10 月，DeepMind

发表在 *Nature* 上的论文详细介绍了 AlphaGo 的另一个新版本——AlphaGo Zero，它以 100∶0 击败了最初的 AlphaGo 系统。

AlphaGo 成功的背后是结合了深度学习、强化学习（Reinforcement Learning）与搜索树算法（Tree Search）三大技术。简单来说，当时的 AlphaGo 有两个核心：策略网络（Policy Network）和评价网络（Value Network），这两个核心都是由卷积神经网络（Convolutional Neural Networks, CNN）所构成的。具体而言，首先在“策略网络”中输入大量棋谱，机器会进行监督式学习，然后使用部分样本训练出一个基础版的策略网络，并使用完整样本训练出“进阶版”的策略网络，让这两个网络对弈，机器通过不断新增的环境数据调整策略，也就是所谓的强化学习。而“策略网络”的作用是选择落子的位置，再由“评价网络”来判断盘面，分析每个步数的权重，预测游戏的输赢结果。当这两个网络把落子的可能性缩小到一个范围内时，机器计算需要庞大运算资源的负担减少了，再利用蒙特卡洛搜索树于有限的组合中算出最佳解。而 AlphaGo Zero 与 AlphaGo 不同，它没有被输入任何棋谱，而是从一个不知道围棋游戏规则的神经网络开始，仅通过全新的强化学习算法，让程序自我对弈，自己成为自己的老师，在这个过程中，神经网络不断被更新和调整。

中国工程院院士“高文”总结了什么样的 AI 系统不需要外部数据就可以战胜人，实际上需要满足以下三个条件：

（1）集合是封闭的。无论是状态集还是其他集，集合都是封闭的，我们知道围棋集合是封闭的。

（2）规则是完备的。也就是说，下棋时什么地方能下，什么地方不能下，这个规则是完全完备的，不能随便更改。

（3）约束是有限的。也就是说，在约束条件下，不可以继续递归，因为允许继续递归之后，往下推演就停不下来，而约束为有限的就能停下来。

满足这三个条件，不需要外部数据，系统自己产生数据就够了。所以可以预见，今后有很多情况，我们可以判断这个人和机器最后谁能赢，满足这三个条件机器一定能赢，无论是德州扑克还是围棋，类似的情况很多。

1.3 AI 与大数据的关系

人工智能如今处在发展的早期阶段，非常像十几年前互联网的成长。推动 AI 发展的三个动力是算法、算力和数据（见图 1-10）。第一个是算法，尤其是机器学习的算法在过去几年迅速发展，不断有各种各样的创新，深度学习、DNN、RNN、CNN 到 GAN，不停地有新的发明创造出来；第二个是计算能力，随着云计算的普及，计算的成本在不断下降，服务器也变得越来越强大，我们将在第 2 章中详细介绍人工智能芯片产业；第三个是数据，数据的产生仍然在以一个非常高的速度发展，数据越多，训练越全面，就会进一步推动算法的不断创新，以

及对计算能力提出更新的要求。数据是 AI 的根本和基础，AI 和大数据密不可分。没有海量数据支撑的人工智能就是人工智障。

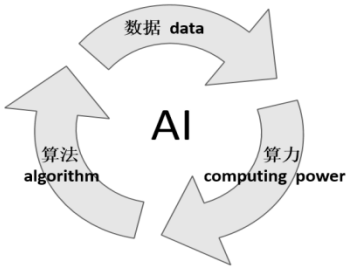


图 1-10　推动 AI 的动力

AI 的火热是与近几年大数据获得重大的突破紧密相关的。本轮 AI 浪潮是大数据驱动的，算法就是“炼数术”。因此，AI 面临的核心挑战之一依然是数据，尤其是进行监督式学习时所需要的高质量训练数据源。大数据与人工智能相辅相成，在人工智能的加持下，海量的大数据对算法模型不断训练，又在结果输出上进行优化，从而使人工智能向更为智能化的方向进步，大数据与人工智能的结合将在更多领域中击败人类所能够做到的极限。

1. 什么是大数据?

云计算、物联网、移动互联、社交媒体等新兴信息技术和应用模式的快速发展，促使全球数据量急剧增加，推动人类社会迈入大数据时代。一般意义上，大数据是指利用现有理论、方法、技术和工具难以在可接受的时间内完成分析计算、整体呈现高价值的海量复杂数据集合。

大数据呈现出多种鲜明的特征。

- 在数据量方面，当前全球所拥有的数据总量已经远远超过历史上的任何时期，更为重要的是，数据量的增加速度呈现出倍增趋势，并且每个应用所计算的数据量也大幅增加。
- 在数据速率方面，数据的产生、传播的速度更快，在不同时空中流转，呈现出鲜明的流式特征，更为重要的是，数据价值的有效时间急剧缩短，也要求越来越高的数据计算和使用能力。
- 在数据复杂性方面，数据种类繁多，数据在编码方式、存储格式、应用特征等多个方面也存在多层次、多方面的差异性，结构化、半结构化、非结构化数据并存，并且半结构化、非结构化数据所占的比例不断增加。
- 在数据价值方面，数据规模增大到一定程度之后，隐含于数据中的知识的价值也随之增大，并将更多地推动社会的发展和科技的进步。此外，大数据往往还呈现出个性化、不完备化、价值稀疏、交叉复用等特征。

大数据蕴含大信息，大信息提炼大知识，大知识将在更高的层面、更广的视角、更大的范围帮助用户提高洞察力，提升决策力，将为人类社会创造前所未有的重大价值。但与此同时，这些总量极大的价值往往隐藏在大数据中，表现出价值密度极低、分布极其不规律、信息隐藏程度极深、发现有用的价值极其困难的鲜明特征。

2. 大数据产业链

如图 1-11 所示，大数据生产全链条覆盖数据采集、计算引擎、数据加工、数据可视化、机器学习、AI 应用等。

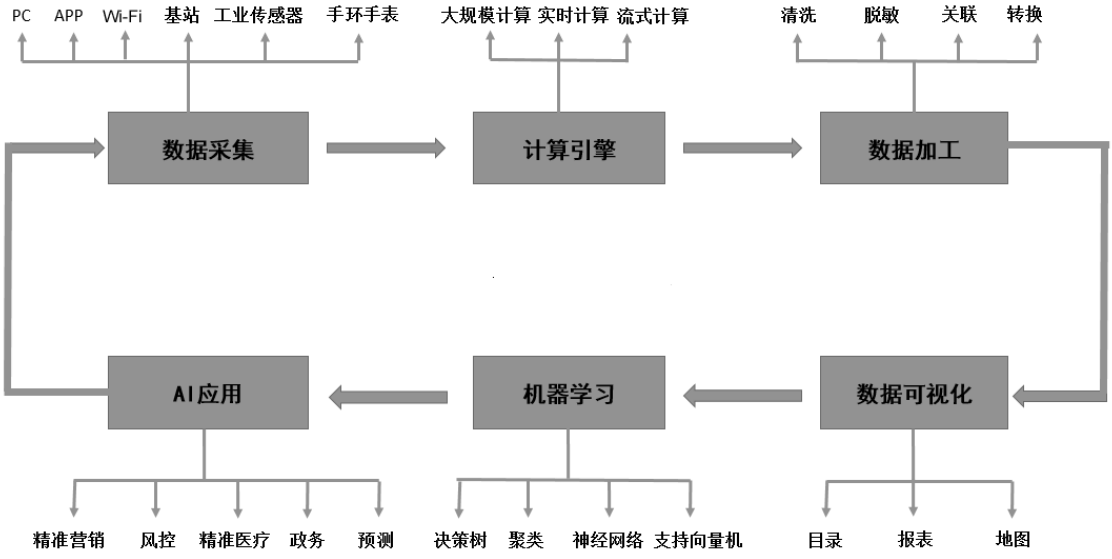


图 1-11 大数据产业链

3. 基于大数据的 AI 应用

如何把数据资源转化为 AI 应用，是我们特别关注的问题。现在看来，大数据和 AI 的结合主要有以下几种较为常用的场景。

- 追踪：互联网和物联网无时无刻不在记录数据，大数据可以追踪、追溯任何记录，形成真实的历史轨迹。历史数据是许多 AI 应用的起点，包括消费者购买行为分析、购买偏好分析等。在电商平台上，从前是人找货，现在是货找人。
- 识别：在对各种因素全面追踪的基础上，通过定位、比对、筛选可以实现精准识别，尤其是对语音、图像、视频进行识别，使 AI 可分析的内容更加丰富，得到的结果更为精准。
- 画像：通过对同一主体不同数据源的追踪、识别、匹配，形成更立体的刻画和更全面的认识。只要积累足够的用户数据，就能分析出用户的喜好与购买习惯，甚至做到“比用户更了解用户自己”。这样的画像就可以精准地推送广告和产品；对企业画像，可以准确地判断其信用及面临的风险。
- 预测：在历史轨迹、识别和画像基础上，对未来趋势及重复出现的可能性进行预测，当某些指标出现预期变化或超预期变化时给予提示、预警。以前也有基于统计的预测，大数据和 AI 技术大大丰富了预测手段，对建立风险控制模型有深刻意义。
- 匹配：在海量信息中精准追踪和识别，利用相关性、接近性等进行筛选比对，更有效率地实现产品搭售和供需匹配。

- 优化：按距离最短、成本最低等给定的原则，通过各种算法对路径、资源等进行优化配置。对企业而言，提高服务水平，提升内部效率；对公共部门而言，节约公共资源，提升公共服务能力。

总之，把用户、数据和算法巧妙地连接起来的是 AI 应用（或 AI 产品）。最终，大数据的成功最关键的一步往往是一个极富想象力的 AI 创新应用。比如金融行业的“秒贷”，就是基于 AI 算法的数据智能实时发挥作用，最终实现秒级放贷，这个是传统的金融服务没法想象的。这样的智能商业才是对传统商业的颠覆。再比如，美国的 UPS 快递公司建立了基于大数据的预测性分析 AI 系统来检测全美 60000 辆车辆的实时车况，以便及时地进行防御性修理。

4. 深度神经网络等新兴技术开辟大数据分析技术的新时代

传统的数据分析方法，无论是传统的 OLAP 技术还是数据挖掘技术，都难以应付大数据的挑战。首先是执行效率低，在处理太字节（TB）级以上的数据时效率更低。其次是数据分析精度难以随着数据量的提升而得到改进，特别是难以应对非结构化数据。目前来看，以深度神经网络等新兴技术为代表的大数据分析技术已经得到一定发展。神经网络是一种先进的人工智能技术，具有自行处理、分布存储和高度容错等特性，非常适合处理非线性的以及模糊、不完整、不严密的知识或数据，十分适合解决大数据挖掘的问题。深度学习是近年来机器学习领域最令人瞩目的方向。自 2006 年深度学习界泰斗 Geoffrey Hinton 在 *Science* 杂志上发表 *Deep Belief Networks* 的论文后，激活了神经网络的研究，开启了深度神经网络的新时代。学术界和工业界对深度学习热情高涨，并逐渐在语音识别、图像识别、自然语言处理等领域获得突破性进展，深度学习在语音识别领域的准确率获得了 20%~30%的提升，突破了近十年的瓶颈。图像识别领域早在 2013 年就通过深度学习将准确率提高到了 89%。神经网络算法的结构和流程特性非常适合大数据分布式处理平台进行计算，通过神经网络能够实现各领域的分析算法和应用。

1.4 AI 与云计算的关系

正如前文所提到的，推动 AI 发展的三个动力是算法、算力和数据（见图 1-10）。随着云计算的普及，计算的成本在不断下降，计算能力也变得越来越强大，这是推动 AI 发展的必备条件。AI 是今后产业发展的巨大引擎。无论是国内的 BATJ，还是美国的谷歌、亚马逊、微软、Facebook、苹果等公司，都已经拥有了海量的云计算基础设施。它们各自推出的 AI 功能都是为了给予云端客户更强的数据分析能力，从而构建基于人工智能的云服务，这符合未来云服务的“云+AI”发展趋势。例如，亚马逊利用 AWS 云正尝试为云端客户提供高效的 AI 解决方案，如图 1-12 所示。谷歌寄希望于借 AI 在云计算领域赶超 AWS。基于微软云平台 Azure 的智能 API 涵盖了五大方向的人工智能技术，包括计算机视觉、语音、语言、知识、搜索五大类 API。

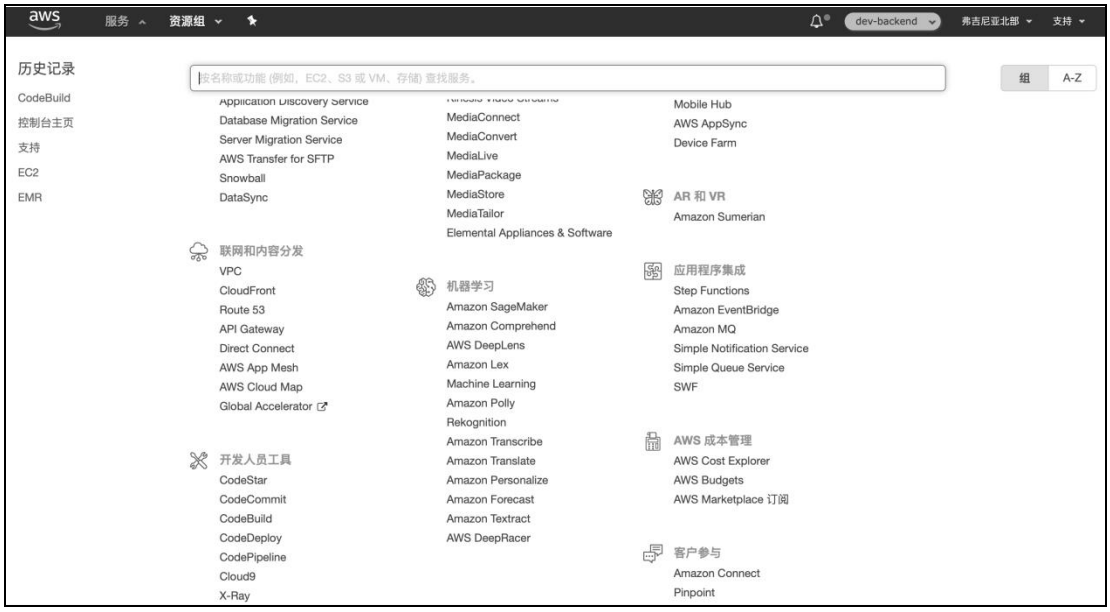


图 1-12　AWS 提供的云端机器学习组件

人工智能技术正在逐渐发展，距离真正的成熟期还有很长的路要走，单单依靠有限的企业去推动整个技术的发展，力量相对有限，而通过开源人工智能平台，能够群策群力，将更多的优秀人才调动到人工智能系统的开发中。开源人工智能平台可以增强云计算业务的吸引力和竞争力，比如用户使用谷歌开源的 TensorFlow 平台训练和导出自己所需要的人工智能模型，然后把模型导入 TensorFlow Serving 对外提供预测类云服务，实质上是将使用开源深度学习工具的用户直接变为其云计算服务的用户，现阶段包括阿里、亚马逊在内的云计算服务商都将机器学习平台嵌入其中，作为增强其竞争实力和吸引更多用户的方式。同时，开放的开发平台将带来下游应用的蓬勃发展。开源平台的建立在推动技术成熟的同时，对科技巨头来说，既整合了人才，又可以第一时间将开发成果接入自己的产品中，实现研发到商业化的快速过渡，从而在人工智能市场中占据先发优势。

数据是资产，云计算为大数据资产提供存储、访问和计算。通过 AI 挖掘价值性信息和预测性分析，是大数据的核心议题，也是云计算的最终方向。

1.5 AI 技术路线

AI 的常用开发框架包括 scikit-learn、谷歌的 TensorFlow、Facebook 的 Torch、微软的 CNTK 等，这些框架都是开源软件。

scikit-learn 简称为 sklearn，是针对 Python 编程语言的开源软件机器学习库。它具有各种分类、回归和聚类算法，包括支持向量机、随机森林、梯度提升、k 均值和 DBSCAN。sklearn

是 GitHub 上最受欢迎的机器学习库之一。sklearn 与许多其他 Python 库很好地集成在一起，例如 matplotlib 和 plotly 用于绘图，NumPy 用于数组矢量化，Pandas 数据帧等。作为专门面向机器学习的 Python 开源框架，sklearn 实现了各种各样成熟的算法，容易安装和使用，样例丰富，而且教程和文档也非常详细。sklearn 的性能表现也是非常不错的。当然，sklearn 也有缺点，例如它不支持深度学习和强化学习，而它们在今天已经是应用非常广泛的技术。此外，它不支持 Python 之外的语言，对 GPU 的使用并不高效。本书的前面 6 章的例子都是基于 sklearn 实现的。

2015 年，谷歌发布第二代人工智能系统 TensorFlow，并宣布将其开源。TensorFlow 包括很多常用的深度学习技术、功能和例子的框架，本书用 3 章内容详细介绍 TensorFlow。2013 年，卷积神经网络发明者 Yann LeCun 加入 Facebook，带领公司在图像识别技术和自然语言处理技术方面得到大幅提升。Facebook 的深度学习框架是在之前的 Torch 基础上实现的，于 2015 年 12 月开源。表 1-2 列出了各个公司所提供的 AI 开源平台。

表 1-2　AI 开源平台列表

公司	开源时间	平台名称	简介
谷歌	2015.11	TensorFlow	谷歌的第二代深度学习系统
微软	2015.11	DMTK	一个将机器学习算法应用在大数据上的工具包
Facebook	2015.12	Torchnet	深度学习 Torch 框架，鼓励模块化编程
微软	2016.01	CNTK	通过一个有向图将神经网络描述为一系列计算步骤

除了上述的 AI 开源平台和框架之外，AWS 推出了 SageMaker，Apache 有 Spark MLlib。Spark MLlib 是一个具有高度拓展性的机器学习库，在 Java、Scala、Python 甚至 R 语言中都非常有用，因为它使用 Python 和 R 中类似 NumPy 这样的程序包，能够进行高效的交互。MLlib 可以很容易地插入 Hadoop 工作流程中。它提供了机器学习算法，如分类、回归、聚类等。这个强大的库在处理大规模的数据时，速度非常快。Spark MLlib 的官网地址是 https://spark.apache.org/mllib/。Spark MLlib 的优点是，对于大规模数据处理来说，非常快，可用于多种语言。缺点是，陡峭的学习曲线，仅 Hadoop 支持即插即用。

第 2 章
AI产业

人工智能是一门新兴的技术科学，该领域的研究包括机器人、语音识别、图像识别、自然语言处理等。人工智能从诞生以来，理论和技术日益成熟，应用领域也不断扩大，AI 赋予了机器一定的视听感知和思考能力，不仅会促进生产力的发展，而且会对经济与社会的运行方式产生积极作用。目前，随着数据资源和运算能力的大幅进步，深度学习算法、语音识别、图像识别等技术加速突破。数据资源、运算能力、核心算法在客观上构成人工智能的三大基本要素，在当前皆重新站上一个新台阶，共同推动当下人工智能从计算智能向更高层的感知、认知智能发展，并通过衍生出通用技术、解决方案输出以及具体人工智能大规模应用产品的落地，掀起人工智能第三次新浪潮。

人工智能作为全球科技革命和产业变革的制高点，已经成为推动经济社会发展的新引擎。人工智能产业是指一个以人工智能关键技术为核心的、由基础支撑和应用场景组成的、覆盖领域非常广阔的产业。与人工智能的学术定义不同，人工智能产业更多的是经济和产业上的一种概括。如图 2-1 所示，人工智能产业分为三层：基础层、技术层和应用层。其中，基础层包括芯片、大数据、网络等多项基础设施，为人工智能产业奠定硬件和数据基础。技术层包括计算机视觉、语音语义识别、机器学习等，多数人工智能技术公司以一项或多项技术细分领域为切入点。而最终人工智能技术能否落地且产生巨大的商业价值，还需要应用层中多场景的应用。目前，人工智能技术应用到多个行业中，包括金融、安防、智能家居、医疗、机器人、自动驾驶等。应用层市场空间大，参与的企业多，这些企业发展垂直应用，解决行业痛点，实现场景落地。

美国的 AI 产业布局非常完善，基础层、技术层和应用层都有涉及，尤其是在算法、芯片和数据等产业核心领域，积累了强大的技术创新优势，各层级企业数量全面领先中国。相比较而言，中国在基础元器件、基础工艺等方面差距较大。AI 的目标客户分为大众消费市场和政府企业。面向政府企业的 AI 商业模式类似于传统 IT 厂商的角色。

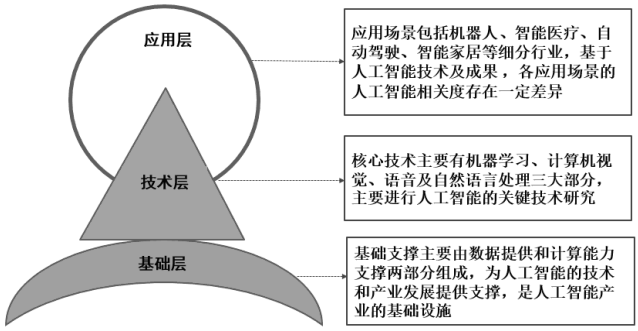


图 2-1　AI 产业层

2.1 基础层

人工智能产业链分为基础层、技术层和应用层。图 2-2 所示为基础层，包括芯片、传感器、大数据、云计算等领域，为 AI 提供数据或计算能力支撑。除了上述列出的领域外，其他领域，如大带宽也是人工智能基础层的内容。通过大带宽提供良好的基础设施，以便在更大范围内进行数据的收集，以更快的速度进行数据的传输，为大数据的分析、计算等环节提供时间和数据量方面的基本保障。

图 2-2　基础层

海量数据是人工智能发展的基础，各类信息系统和传感器的数据是未来大数据的核心。伴随着物联网的发展，数据开始以指数级规模增长，大量数据应用到人工智能算法模型的训练中，AI 得以快速发展。人工智能的技术也快速应用到大数据分析中，通过 AI 挖掘丰富数据背后的价值，从而可以极大地提高生产力。随着一些核心基础设施问题的解决，AI 应用层正在快速构建。一方面，专门的 AI 应用几乎在任何一个垂直行业都有出现。另一方面，在企业内部，已经出现了各种 AI 工具。例如，智能客服应用为用户提供个性化企业服务。

2.1.1 芯片产业

随着中兴事件和华为芯片断供事件的发生，大家都高度重视芯片。的确，AI 的“大脑”在于芯片和算法。AI 芯片也被称为 AI 加速器或计算卡，即专门用于处理人工智能应用中的大量计算任务的模块。当前，AI 芯片主要分为 GPU、FPGA、ASIC 和类脑芯片。在人工智能时代，它们各自发挥优势，呈现出百花齐放的状态。在美国人工智能企业中，融资占比排名第一的领域为芯片/处理器，占比 31%。AI 芯片由于投资周期长、专业技术壁垒厚，导致竞争非常激烈且难以进入。

AI 芯片的计算场景可分为云端 AI 和终端 AI。英伟达首席科学家 William Dally 将深度学习的计算场景分为三类，分别是数据中心的训练、数据中心的推理和嵌入式设备的推理。前两者可以总结为云端的应用，后者可以概括为终端的应用。终端智能芯片的一个经典案例是苹果的 A13 神经引擎，每秒运算次数最高可达万亿次。它实现了基于深度学习的高准确性人脸识别解锁方式（Face ID），并解决了云接口（Cloud-Based API）带来的延时和隐私问题，以及庞大的训练数据和计算量与终端硬件限制的矛盾。

AI 芯片目前有三个技术路径，通用的 GPU（既能作为图形处理器引爆游戏业务，又能渗透数据中心横扫训练端）、可编程的 FPGA（适用于迭代升级，各类场景化应用前景超大）以及专业的 ASIC（叩开终端 AI 的大门）。其中，英伟达、英特尔两大传统芯片巨头在三大路径，特别是通用芯片和半定制芯片都有布局，掌握强大的先发优势，在数据中心、汽车等重要蓝海布局扎实；在 ASIC 方面，谷歌从 TPU 出发开源生态进行布局，且 TPU 展露了训练端芯片市场的野心。ASIC 定制化的特点有效规避了传统巨头的垄断局面，有着可靠健康的发展路线。表 2-1 总结了目前几个主流的 AI 芯片厂商。

表 2-1 AI 芯片厂商列表

公司	芯片	说明
高通	骁龙	发布高通神经处理软件开发工具包，挖掘骁龙芯片的 AI 计算能力，致力于移动 AI、智能驾驶 AI 等领域
谷歌	TPU（TensorFlow Processing Unit）	专为其深度学习算法 TensorFlow 设计，也用在 Google 搜索、图片、Gmail、翻译、Google 助手等系统中，Cloud TPU 能够对机器学习模型的训练和运行带来显著的加速效果
英伟达	GPU	适合并行算法，占目前 AI 芯片市场最大份额，应用领域涵盖视频游戏、电影制作、产品设计、医疗诊断等各个门类
AMD	GPU	GPU 第二大市场占有率
英特尔	FPGA	来自 167 亿美元收购的 Altera，峰值性能逊色于 GPU，指令可编程，且功耗也要小得多，适用于工业制造、汽车电子系统等，可与至强处理器整合
	Xeon Phi Knights Mill	适用于包括深度学习在内的高性能计算，能充当主处理器，可以在不配备其他加速器或协处理器的情况下高效处理深度学习应用
微软	FPGA	自主研发，已被用于 Bing 搜索，能支持微软的云服务 Azure，速度比传统芯片快得多
Xilinx	FPGA	世界上最大的 FPGA 制造厂商，推出了支持深度学习的 reVision 堆栈

（续表）

公司	芯片	说明
IBM	TrueNorth 类脑芯片	是一种基于神经形态的芯片
苹果	专用芯片 Apple Neural Engine	该芯片定位于本地设备 AI 任务处理，把人脸识别、语音识别等 AI 相关任务集中到 AI 模块上，提升 AI 算法效率
Mobileye	EyeQ5	用于汽车辅助驾驶系统

英伟达是 GPU 的行业领袖。GPU 是目前深度学习领域的主流芯片，拥有强大的并行计算力。而另一个老牌芯片巨头英特尔则是通过大举收购进入 FPGA 人工智能芯片领域的。谷歌的 TPU 是专门为其深度学习算法 TensorFlow 设计的，TPU 也用在了 AlphaGo 系统中。第三代 Cloud TPU 理论算力达到了 420T Flops，能够对机器学习模型的训练和运行带来显著的加速效果。类脑芯片是一种基于神经形态工程，借鉴人脑信息处理方式，具有学习能力的超低功耗芯片。IBM 从 2008 年开始模拟人类大脑的芯片项目。苹果公司的“苹果神经引擎（Apple Neural Engine，简称 ANE）”是一款专用芯片。该芯片定位于本地设备的 AI 任务处理，把人脸识别、语音识别等任务集中到 AI 模块上，提升 AI 算法效率。苹果神经引擎芯片属于嵌入式神经网络处理器（NPU）。

自动驾驶系统与 AI 芯片紧密相关，比如，特斯拉的电动车使用的是英伟达的芯片。在美国市场上，正在逐渐形成英伟达与英特尔-Mobileye 联盟两大竞争者。Mobileye 已经被英特尔收购。Mobileye 的机器视觉算法与英特尔的芯片、数据中心、AI 和传感器融合，加上地图服务，正协同打造一个全新的自动驾驶供应商。在 Mobileye 的官网上，有一段 40 分钟的视频，它记录了在特别复杂的街道上自动驾驶的整个过程。对于自动驾驶感兴趣的读者，可以观看一下，我相信你一定会深有体会的。图 2-3 是视频中截取的两幅图，左边是在自动驾驶软件上呈现的分析图，右边是实况，其中上面有一个圆圈的汽车就是自动驾驶的汽车。

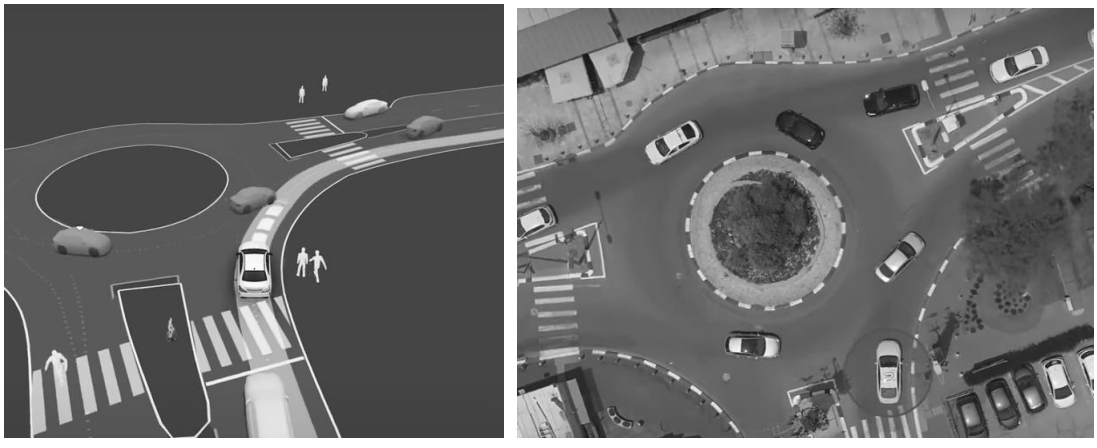


图 2-3　自动驾驶案例

人工智能芯片领域，这是一个包含数十家创业公司，以及英特尔、AMD、高通和英伟达

这样的传统硬件厂商的重要市场。随着时代的发展，谷歌和亚马逊已不再是纯粹的互联网企业，苹果和微软已不再是纯粹的终端设备和软件公司，它们都已或多或少地开始扮演起芯片制造者的角色。

2.1.2 GPU

随着 CPU 摩尔定律的终止，传统处理器的计算力已远远不能满足海量并行计算与浮点运算的深度学习训练需求，而在人工智能领域反映出强大适应性的 GPU 成为标配。GPU 比 CPU 拥有更多的运算器（Arithmetic Unit），只需要进行高速运算而不需要逻辑判断，其海量数据并行运算的能力与深度学习的需求不谋而合。因此，在深度学习上游训练端（主要用于云计算数据中心），GPU 是第一选择。目前，GPU 的市场格局以英伟达为主（超过 70%），AMD 为辅，预计 3~5 年内 GPU 仍然是深度学习市场的第一选择。截至目前（2020 年 8 月），英伟达毫无疑问是这波人工智能浪潮最大的受益者，它的市值超过了英特尔的体量。英伟达的崛起完全得益于这场突如其来的人工智能大革新。

有些芯片商除了做芯片之外，还会在整个 AI 生态上进行布局。例如，英伟达拥有一个较为成熟的开发生态环境（CUDA，见图 2-4），包括开发套件和丰富的库以及对英伟达 GPU 的原生支持。

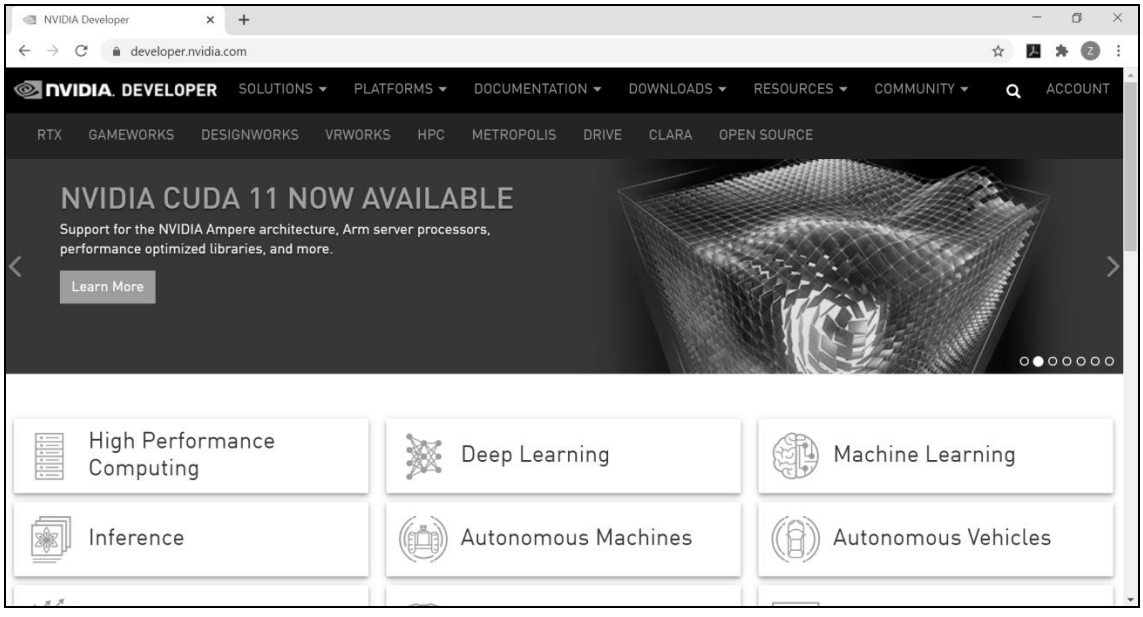


图 2-4　英伟达 GPU 开发环境 CUDA、开发库和工具

2.1.3 FPGA

FPGA 是现场可编程门阵列。下游推理端更接近终端应用，更关注响应时间而不是吞吐率，需求更加细分。目前来看，下游推理端虽可容纳 GPU、FPGA、ASIC 等芯片，但随着 AI 的发展，FPGA 的低延迟、低功耗、可编程性（适用于传感器数据预处理工作以及小型开发试错

升级迭代阶段）的优势将突显出来。

在 FPGA 的市场份额中，Xilinx 的份额约为 49%，主要应用于工业和通信领域，近年亦致力于云计算数据中心的服务器以及无人驾驶的应用。Altera（已被英特尔收购）的市场份额约为 40%，定位同 Xilinx 类似。莱迪斯半导体（Lattice Semiconductor）的市场份额约为 6%，主要市场为消费电子产品和移动传输，以降低耗电量、缩小体积及缩减成本为主。Microsemi（Actel）的市场份额约为 4%，瞄准通信、国防与安全、航天与工业等市场。目前，Altera 的 FPGA 产品被用于微软 Azure 云服务中，包括必应搜索、机器翻译等应用中。

2.1.4　ASIC

ASIC 是 Application Specific Integrated Circuit 的英文缩写，中文名为专用集成电路或特殊应用集成电路。终端设备的模型推理方面，由于低功耗、便携等要求，FPGA 和 ASIC 的机会优于 GPU。

谷歌的 TPU 就是 ASIC 类型的芯片。如图 2-5 所示，通过谷歌云平台（GCP）提供的 Cloud TPU 能帮助机器学习专家更快地训练和运行机器学习（ML）模型。Cloud TPU 是谷歌设计的一种硬件加速器，使用多个定制化 ASIC 构建，单个 Cloud TPU v3 的计算能力达到 420 万亿次浮点运算，具备 128GB 的高带宽内存。这些板卡可单独使用，也可以通过超快的专门网络联合使用，以构建数千万亿次级别的机器学习超级计算机（TPU Pod）。

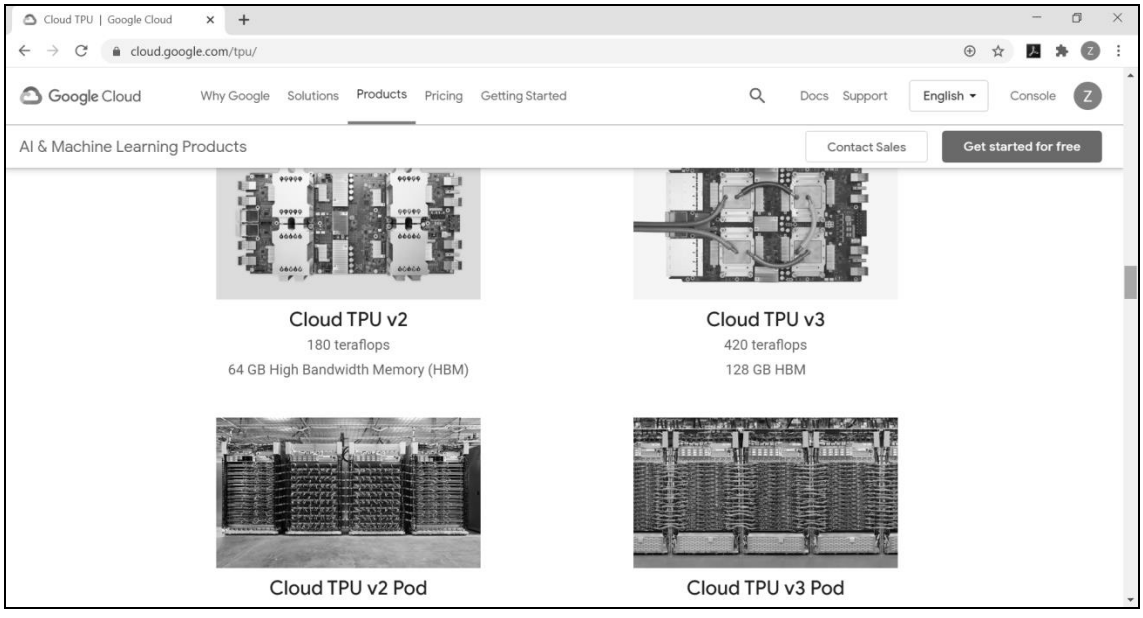


图 2-5　谷歌的 TPU

2.1.5　TPU

随着人工智能革新浪潮与技术进程的推进，AI 芯片成了该领域下一阶段的竞争核心。2016 年 5 月，谷歌发布了一款特别为 TensorFlow 设计的机器学习专用芯片：张量处理器（Tensor

Processing Unit，TPU），它旨在优化机器学习工作负载。2017 年又推出了它的第二代产品（Cloud TPU）。目前的最新产品是 Cloud TPU v3。这是一种被认为比 CPU，甚至 GPU 更加高效的机器学习专用芯片。价格大约为每个云 TPU 每小时几美元。这个曾支持了著名 AI 围棋程序 AlphaGo 的强大芯片将很快成为各家科技公司开展人工智能业务的强大资源。

据谷歌称，现在的 TPU 可用于机器学习模型的训练，这个机器学习过程中重要的部分完全可以在单块、强大的芯片上进行。TPU 帮助了谷歌的各类机器学习应用进行快速预测，并使产品迅速对用户需求做出回应。谷歌称，TPU 已运行在每一次谷歌搜索中，TPU 支持谷歌图像搜索（Google Image Search）、谷歌照片（Google Photo）和街景（Street View）等产品的基础精确视觉模型，TPU 也帮助了谷歌翻译质量的提升，而其强大的计算能力也在 DeepMind AlphaGo 的胜利中发挥了重要作用。

Cloud TPU 的目的是为 TensorFlow 工作负载提供差异化的性能，使机器学习工程师实现更快迭代，他们可以在一系列 Cloud TPU 上训练同样模型的不同变体，而且第二天就可以将准确率最高的训练模型部署到生产过程。使用单个 Cloud TPU 并遵循教程（https://cloud.google.com/tpu/docs/tutorials/resnet），就可以在不到一天的时间内训练 ResNet-50，使其在 ImageNet 基准上达到期望的准确率。传统上，编写自定义 ASIC 和超级计算机的程序需要极高的专业水平，而对于 Cloud TPU 而言，可以使用高级 TensorFlow API 进行编程。谷歌开源了一系列高性能 Cloud TPU 模型实现，比如 ResNet-50 和图像分类模型（https://github.com/tensorflow/tpu/tree/master/models/official），用于机器翻译和语言建模的 Transformer（https://cloud.google.com/tpu/docs/tutorials/transformer），用于目标检测的 RetinaNet（https://github.com/tensorflow/tpu/blob/master/models/official/retinanet）。

云 TPU 同样简化了对机器学习计算资源的规划和管理，随着需求的变化动态调整自己的容量。相比于花费资金、时间和专业人才来设计、安装、维护一个机器学习计算群（还需要专门的供能系统、冷却系统、网络设施和存储系统），我们可以从谷歌多年以来优化过的大规模、高集成的机器学习基础设施受益。另外，谷歌云服务提供了复杂的安全机制和实践的保护。谷歌云 TPU 还提供大量的高性能 CPU（英特尔）和 GPU（英伟达）。

有意思的是，谷歌 TPU 的全面开放让英伟达警觉的神经再次紧绷。我们可以认为，谷歌是英伟达在人工智能算力市场最大的竞争对手。随着第三代 TPU 的发布及其在人工智能专有领域，特别是在搭载了谷歌 TensorFlow 框架的深度神经网络训练效率方面的表现，外界越来越认识到二者间的差距逐渐缩小。谷歌在人工智能领域的雄心十分明显，从一开始对 TPU 的只字不提到后来开放上云，谷歌已逐渐认识到算力市场的巨大潜力并渴求牢牢抓住这一契机。TPU+TensorFlow+云训练的模式让谷歌获得了迄今为止其他科技巨头尚不具备的人工智能核心竞争实力。这一点已经引起其他科技公司的注意，他们认为，各行各业的公司都有自己的数据驱动业务，算力不应该被掌控在一家巨头手上。AI 芯片崛起的背后是算力的战争。

2.1.6 亚马逊的芯片

亚马逊最近推出了自己的芯片，名字叫 Graviton2，目前用于亚马逊自己的云计算数据中

心。此举让亚马逊成为继谷歌和苹果之后，又一家自主研发芯片的大型科技公司。这些科技公司之所以这样做，是为了实现自家产品的个性化。但对于英特尔和英伟达等传统芯片公司而言，他们的客户就要变成竞争对手了。当前，英特尔控制着服务器主芯片市场 98%的份额，而英伟达则为这些服务器开发与英特尔主芯片协同工作的人工智能芯片。FPGA 芯片授权初创公司 Flex Logix Technologies CEO（Geoff Tate）称：“如果这种趋势持续下去，将来，数据中心所有者将自主研发芯片，与当前的芯片供应商相竞争。”

2.1.7　芯片产业小结

摩尔定律的终止已成为业界共识，那么 AI 芯片的革命又从何说起？众所周知，当前的人工智能技术进程是奠定在神经网络与深度学习之上的。从人工智能发展史来看，经历了早期的控制论和简单神经网络、逻辑过程与编程革命、运筹学与博弈论、专家系统的兴起，人工智能技术进程在算法与算力的不断迭代中演化至今。而当前神经网络算法趋于稳固，在算法框架没有深刻变化的前提下，算力就成了唯一的更新焦点。

深度学习工程的两大关键环节 Training（训练）和 Inference（推理或推测）需要大量的算力支撑，而 GPU 在训练环节扮演着不可或缺的角色。但随着人工智能应用场景的延伸，GPU 并非所有深度学习计算任务的充分条件，FPGA（现场可编程门阵列）和 ASIC（专用集成电路）同样有着相当大的表现空间。前者通过内置可灵活组合的逻辑、IO、连线模块为专用计算服务，后者是不可配置的高度定制化芯片。谷歌 TPU 就是 ASIC 的一种方案。

凭借 GPU，英伟达公司一直是 AI 趋势的最大受益者。因为其图形处理器（GPU）是训练 AI 系统的早期选择。GPU 能够同时执行大量复杂的数学运算，这使它在早期成为 AI 应用的最佳选择。后来，科技巨头纷纷研发自己的 AI 芯片，包括谷歌的 TPU、苹果的神经引擎、微软的 FPGA，以及亚马逊正在为 Alexa 研发的定制 AI 芯片。

2016 年初，谷歌开始研发被称为 TPU 的定制 AI 芯片，它基于 ASIC，旨在为谷歌公司的深度学习 AI 应用程序提供更高效的性能。该芯片为 TensorFlow 奠定了基础，TensorFlow 是用于训练该公司的 AI 系统的框架。TPU 可以处理 AI 的训练和推理阶段。谷歌的优势在于凭借自身“TPU+TensorFlow+云”的资源吸引开发者和拓展企业级市场、专用领域，但该模式的前提必须是谷歌极力维系 TensorFlow 作为深度学习主流框架而长期存在，一旦神经网络算法主流架构有变，TPU 作为高度制定化的芯片产物，其单位成本之高恐酿成不可回避的风险。相反，倘若谷歌的计划顺利实施，其垄断的生态优势同样对英伟达形成巨大威胁。

苹果公司一直是用户隐私的支持者，并且走了一条与它的技术同行不同的道路。该公司的移动设备为传输到云端的任何数据添加电子噪声，同时剥离任何可识别个人身份的信息，从而更大程度地保证用户的隐私和安全。苹果公司开发了一种神经引擎，作为其新的 A13 仿生芯片的一部分，该芯片是一款可在本地处理多种 AI 功能的先进处理器。这大大减少了传输到云端的用户信息量，有助于保护用户数据。

微软公司早前投注于可定制处理器——现场可编程门阵列（FPGA），这是一种专用芯片，可为客户的特定用途进行配置。这些已经成为微软 Azure 云计算系统的基础，并且提供比 GPU

等传统产品更灵活的架构和更低的功耗。

虽然这些公司都采用了不同的处理器策略，但是仍在大量使用英伟达的 GPU，英伟达 GPU 的使用增长仍在继续。竞争是不可避免的，但到目前为止还没有解决方案能够完全取代 GPU。相比 GPU 集群，FPGA 因其定制化、低功耗和忽略延迟的特点，在终端推测环节有着广泛应用，所以它被微软、亚马逊等云商以及苹果、三星等手机制造商所接受。而 GPU 与 TPU 作为训练环节的主力，则开启了两种不同产品形态争锋对立的局面，也就是说，在深度学习训练领域，完全成了英伟达和谷歌两者之间的战争。AI 芯片战争已经全面打响，由人工智能进程引发的第二次芯片革命已经让业界嗅到了熟悉的工业革命的气息。正如 19 世纪蒸汽机、内燃机的迭代结束了大洋之上纵横数个世纪的风帆时代，人工智能算力的突破亦将成为摩尔定律的变革者，将延续了近一个世纪的计算机科学文明引入下一阶段。

2.1.8 传感器

如今的机器人已具有类似人一样的肢体及感官功能，有一定程度的智能，动作灵活，在工作时可以不依赖人的操纵，而这一切都少不了传感器的功劳。传感器是机器人感知外界的重要帮手，它们犹如人类的感知器官，机器人的视觉、力觉、触觉、嗅觉、味觉等对外部环境的感知能力都是由传感器提供的，同时，传感器还可用来检测机器人自身的工作状态，以及机器人智能探测外部工作环境和对象的状态，并能够按照一定的规律转换成可输出信号。为了让机器人实现尽可能高的灵敏度，在它的身体构造里会装上各式各样的传感器，那么机器人究竟要具备多少种传感器才能尽可能地做到如人类一样灵敏呢？

根据检测对象的不同可将机器人用的传感器分为内部传感器和外部传感器。内部传感器主要用来检测机器人内部系统的状况，如各关节的位置、速度、加速度、温度、电机速度、电机载荷、电池电压等，并将所测得的信息作为反馈信息送至控制器，形成闭环控制。而外部传感器用来获取有关机器人的作业对象及外界环境等方面的信息，是机器人与周围交互工作的信息通道；用来执行视觉、接近觉、触觉、力觉等传感器，比如距离测量、声音、光线等。

1. 视觉传感器

机器视觉是使机器人具有感知功能的系统，其通过视觉传感器获取图像进行分析，让机器人能够代替人眼辨识物体，测量和判断，实现定位等功能。目前在国内使用的简便的智能视觉传感器占了机器视觉系统市场 60％左右的份额。视觉传感器的优点是探测范围广、获取信息丰富，实际应用中常使用多个视觉传感器或者与其他传感器配合使用，通过一定的算法可以得到物体的形状、距离、速度等诸多信息。

以深度摄像头为基础的计算视觉领域已经成为整个高科技行业的投资和创业热点之一。有意思的是，这一领域的许多尖端成果都是由初创公司先推出的，再被巨头收购后发扬光大，例如英特尔收购 RealSense 实感摄像头，苹果收购 Kinect 的技术供应商 PrimeSense，Oculus 收购了一家主攻高精度手势识别技术的以色列技术公司 Pebbles Interfaces。

深度摄像头早在 20 世纪 80 年代就由 IBM 提出了相关概念，2005 年创建于以色列的 PrimeSense 公司是该技术民用化的先驱。当时，在消费市场推广深度摄像头还处在概念阶段，

此前深度摄像头仅使用在工业领域，为机械臂、工业机器人等提供图形视觉服务。由它提供技术方案的微软 Kinect 成为深度摄像头在消费领域的开山之作，并带动整个业界对该技术的民用开发。

2. 声音传感器

声音传感器的作用相当于一个话筒（麦克风），用来接收声波，显示声音的振动图像，但不能对噪声的强度进行测量。声觉传感器主要用于感受和解释在气体（非接触感受）、液体或固体（接触感受）中的声波。声波传感器的复杂程度可以从简单的声波存在检测到复杂的声波频率分析，直到对连续自然语言中单独语音和词汇的辨别。

从 20 世纪 50 年代开始，贝尔实验室开发了世界上第一个语音识别 Audry 系统，可以识别 10 个英文数字。到 20 世纪 70 年代，声音识别技术得到快速发展，动态时间规整（DTW）算法、向量量化（VQ）以及隐马尔可夫模型（HMM）理论等相继被提出，实现了基于 DTW 技术的语音识别系统。近年来，声音识别技术已经从实验室走向实用，国内很多公司都利用声音识别技术开发出了相应产品，比如科大讯飞、腾讯、百度等，共闯语音技术领域。

3. 距离传感器

用于智能移动机器人的距离传感器有激光测距仪（兼可测角）、声呐传感器等，近年来发展起来的激光雷达传感器是目前比较主流的一种，可用于机器人导航和回避障碍物。

4. 触觉传感器

触觉传感器主要是用于机器人模仿触觉功能的传感器。触觉是人与外界环境直接接触时的重要感觉功能，研制满足要求的触觉传感器是机器人发展中的关键技术之一。随着微电子技术的发展和各种有机材料的出现，已经提出了多种多样的触觉传感器的研制方案，但目前大部分属于实验阶段，达到产品化的不多。

5. 接近觉传感器

接近觉传感器介于触觉传感器和视觉传感器之间，可以测量距离和方位，而且可以融合视觉和触觉传感器的信息。接近觉传感器可以辅助视觉系统的功能，来判断对象物体的方位、外形，同时识别其表面形状。因此，为准确抓取部件，对机器人接近觉传感器的精度要求是非常高的。这种传感器主要有以下作用：

（1）发现前方障碍物，限制机器人的运动范围，以避免和障碍物碰撞。

（2）在接触对象物体前得到必要信息，比如与物体的相对距离、相对倾角，以便为后续动作做准备。获取物体表面各点间的距离，从而得到有关对象物表面形状的信息。

6. 滑觉传感器

滑觉传感器主要是用于检测机器人与抓握对象间滑移程度的传感器。为了在抓握物体时确定一个适当的握力值，需要实时检测接触表面的相对滑动，然后判断握力，在不损伤物体的情

况下逐渐增加力量，滑觉检测功能是实现机器人柔性抓握的必备条件。通过滑觉传感器可实现识别功能，对被抓物体进行表面粗糙度和硬度的判断。滑觉传感器按被测物体滑动的方向可分为三类：无方向性传感器、单方向性传感器和全方向性传感器。其中，无方向性传感器只能检测是否产生滑动，无法判别方向；单方向性传感器只能检测单一方向的滑移；全方向性传感器可检测多个方向的滑动情况，这种传感器一般制成球形以满足需要。

7. 力觉传感器

力觉传感器是用来检测机器人自身力与外部环境力之间相互作用力的传感器。力觉传感器经常装于机器人关节处，通过检测弹性体变形来间接测量所受力。装于机器人关节处的力觉传感器常以固定的三坐标形式出现，有利于满足控制系统的要求。目前出现的六维力觉传感器可实现全力信息的测量，因其主要安装于腕关节处而被称为腕力觉传感器。腕力觉传感器大部分采用应变电测原理，按其弹性体结构形式可分为两种：筒式和十字形腕力觉传感器。其中，筒式腕力觉传感器具有结构简单、弹性梁利用率高、灵敏度高的特点；而十字形腕力觉传感器结构简单、坐标建立容易，但加工精度要求高。

8. 速度和加速度传感器

速度传感器有测量平移和旋转运动速度两种，但大多数情况下，只限于测量旋转速度。利用位移的导数，特别是光电方法让光照射旋转圆盘，检测出旋转频率和脉冲数目，以求出旋转角度，并利用圆盘制成有缝隙，通过两个光电二极管辨别出角速度（转速），这就是光电脉冲式转速传感器。

加速度传感器是一种能够测量加速度的传感器。通常由质量块、阻尼器、弹性元件、敏感元件和适调电路等组成。传感器在加速过程中，通过对质量块所受惯性力的测量，利用牛顿第二定律获得加速度值。根据传感器敏感元件的不同，常见的加速度传感器包括电容式、电感式、应变式、压阻式、压电式等。

2.1.9 传感器小结

机器人要想做到如人类般灵敏，视觉传感器、声音传感器、距离传感器、触觉传感器、接近觉传感器、力觉传感器、滑觉传感器、速度和加速度传感器这 8 种传感器对机器人极为重要，尤其是机器人的五大感官传感器是必不可少的，从拟人功能出发，视觉、力觉、触觉最为重要，目前已进入实用阶段，但其他的感官，如听觉、嗅觉、味觉、滑觉等对应的传感器还等待一一攻克。

人工智能目前正在为社会的方方面面带来革新。比如，通过结合数据挖掘和深度学习的优势，我们可以利用人工智能来分析各种来源的大量数据，识别各种模式，提供交互式理解和进行智能预测。这种创新发展的一个例子就是将人工智能应用于由传感器生成的数据，尤其是通过智能手机和其他消费者设备所收集的数据。运动传感器数据及其他信息（比如 GPS 信息）可提供大量不同的数据集。本节最后以常见的运动传感器为例来说明 AI 和传感器的综合应用。

一个常见的应用是通过分析数据来确定用户在每个时间段的活动，无论是坐姿、走路、跑

步还是睡眠的情况下。在活动跟踪方面，原始数据通过轴向运动传感器得以收集，例如智能手机、可穿戴设备和其他便携式设备中的加速度计和陀螺仪。这些设备获取三个坐标轴（x、y、z）上的运动数据，以便于连续跟踪和评估活动。对于人工智能的监督式学习，需要用标记数据来训练“模型”，以便分类引擎可以使用此模型对实际用户行为进行分类。只获取原始传感器数据是不够的。我们观察到，要实现高度准确的分类，需要仔细确定一些特征。为了进行活动识别，指示性特征可以包括“滤波信号”，例如身体加速（来自传感器的原始加速度数据），或“导出信号”，例如高速傅里叶变换（FFT）值或标准差计算。举例来说，加州大学欧文分校（UCI）的机器学习数据库创建了一个定义了 561 个特征的数据集，这个数据集以 30 名志愿者的 6 项基本活动（即站立、坐姿、卧姿、行走、下台阶和上台阶）为基础。使用默认的 LibSVM 内核训练的模型进行活动分类的测试，准确度高达 91.84%。在完成培训和特征排名后，选择最重要的 19 项功能足以达到 85.38%的活动分类测试准确度。我们还发现最相关的特征是频域变换以及滑动窗口加速度原始数据的平均值、最大值和最小值。

通过传感器为用户提供真正的个性化体验已成为现实，通过人工智能，系统可以利用由智能手机、可穿戴设备和其他便携设备的传感器所收集的数据为人们提供更多深度功能。未来几年，一系列现在还难以想象的设备和解决方案将会进一步发展。人工智能和传感器为设计师和用户打开了一个激动人心的新世界。

2.2 技术层

技术层是在基础层之上，结合软硬件能力所实现的、针对不同细分应用开发的技术。如图 2-6 所示，技术层主要包括机器学习、计算机视觉、语音及自然语言处理三个方面。主要技术领域包括图像识别、语音识别、自然语言处理和其他深度学习应用等。涉及的领域包括机器视觉、指纹识别、人脸识别、视网膜识别、虹膜识别、掌纹识别、专家系统、自动规划、智能搜索、定理证明、博弈、自动程序设计、智能控制、机器人学习、语言和图像理解等。

图 2-6　AI 技术层

目前，技术层企业在计算机视觉、语音识别等领域竞争激烈。技术层涵盖的厂商以科技巨头、传统科研机构及新兴技术创业公司为主。除了综合性科技巨头外，创业企业也依赖自身技术的积累和细分领域的积累快速崛起。在发展路径上，以 2B、2C 或 2B2C 为主。一方面，面向企业级用户，为应用层厂商提供技术支持；另一方面，研发相应的软件及硬件产品，直接面对消费者，或者提供车载、家居等产品的人机交互技术，从而满足用户需求。

科技巨头仍然掌握技术、数据、资金优势，生态链相对完整。而传统技术厂商（如语音识别领域的科大讯飞）具有强大的科研背景，掌握一定的研发能力，同时获得政府的支持，与相关政府机构合作获取大量的数据来源，强化了人工智能技术。创业公司深耕垂直领域，创始团队多是技术专家，掌握研发技术，通过融资等方式弥补资本不足，逐渐积累资金、人才、技术实力，专攻细分领域，可以快速实现技术的落地，而其技术上的创新也弥补了传统技术提供商及科技巨头的不足，能够在竞争中实现技术的成熟。

2.2.1 机器学习

人工智能、机器学习、深度学习是我们经常听到的三个热词。关于三者的关系，简单来说，机器学习是实现人工智能的一种方法，深度学习是实现机器学习的一种技术（见图 2-7）。机器学习使计算机能够自动解析数据、从中学习，然后对真实世界中的事件做出决策和预测；深度学习是利用一系列“深层次”的神经网络模型来解决更复杂问题的技术。

图 2-7　人工智能、机器学习和深度学习的包含关系

人工智能的核心是通过不断地进行机器学习，而让自己变得更加智能。自 2015 年以来，人工智能开始大爆发。一方面是由于巨头整合了 AI 开源平台和芯片，技术快速发展，GPU 的广泛应用，使得并行计算变得更快、更便宜、更有效；另一方面在于云计算、云存储的发展和当下海量数据的爆发，各类图像数据、文本数据、交易数据等为机器学习奠定了基础。机器学习利用大量的数据来“训练”，通过各种算法从数据中学习如何完成任务，使用算法来解析数据、从中学习，然后对真实世界中的事件做出决策和预测。我们将在第 3~6 章中讲解机器学习。

深度学习是机器学习的重要分支，作为新一代的计算模式，深度学习力图通过分层组合多个非线性函数来模拟人类神经系统的工作过程，其技术的突破掀起了人工智能的新一轮发展浪潮。深度学习的人工神经网络算法与传统计算模式不同，本质上是多层次的人工神经网络算法，

即模仿人脑的神经网络，从最基本的单元上模拟了人类大脑的运行机制，它能够从输入的大量数据中自发地总结出规律，再举一反三，应用到其他的场景中。因此，它不需要人为地提取所需解决问题的特征。IT 巨头争相开源人工智能平台，各种开源深度学习框架层出不穷。2015 年以来，全球人工智能顶尖巨头陆续开源自身核心的人工智能平台，其中包括 Caffe、CNTK、MXNet、Neon、TensorFlow、Theano 和 Torch 等。我们将在第 7~10 章讲解深度学习。

深度学习的典型代表是谷歌 AlphaGo，而 AlphaGo Zero 采用纯强化学习的方法进一步扩展了人工智能技术，不需要人类的样例或指导，不需要提供基本规则以外的任何领域知识，在它自我对弈的过程中，神经网络被调整、更新，以预测下一个落子位置，并以 100∶0 的战绩击败 AlphaGo。深度学习使得机器学习能够实现众多的应用，使所有的机器辅助功能成为可能，拓展了人工智能的领域范围。谷歌作为人工智能领域的科技巨头，在软硬件领域都有布局，通过结合开源平台、智能芯片和相关硬件，谷歌建立了完整的人工智能生态。其中，谷歌自主研发的深度学习开源平台 TensorFlow，可用于编写并编译执行机器学习算法的代码，并将机器学习算法变成符号表达的各类图表。TensorFlow 目前已应用于谷歌搜索、谷歌翻译等服务。同时，大量开发者也接入到平台中，成为主流的深度学习框架。谷歌还进一步推出了 TensorFlow Lite，支持移动和其他终端设备，谷歌已成为人工智能领域不可或缺的巨头。本书的多个章节都是以 TensorFlow 为基础阐述机器学习技术的。

2.2.2　语音识别与自然语言处理

交互模式的变革贯穿了整个 IT 产业的发展史，语音交互成为下一代人机交互的主要模式（见表 2-2）。

表 2-2　交互模式的变革

交互时代	DOS	图形用户界面	触屏	自然语言
交互方式	键盘命令行	鼠标+键盘	触摸	声音
输入/输出	文字	文字、图片	文字、图片	声音、图像

语音识别与自然语音处理是机器能够“听懂”用户语言的主要技术基础，其中语音识别注重对用户语言的感知，目前在中文语音识别上，国内已经达到 97%的语音识别准确率，这要归功于深度神经网络的应用、算力的提高以及大数据的积累。语音识别是机器感知用户的基础，在听到用户的指令之后，更为重要的是如何让机器懂得指令的意义，这就需要自然语言处理，将用户的语音转化为机器能够反应过来的机器指令，包括自然语言理解、多轮对话理解、机器翻译技术等。对于自然语言处理方面，虽然深度学习能起到的作用还有待观察，但在语义理解和语言生成等领域都有了重要突破。如图 2-8 所示，很多提供语音技术服务的公司突破了原有的单纯语音识别或者语义理解的业务框架，开始提供整体的智能语音交互产品。

图 2-8　语音交互过程

1. 语音识别技术

语音识别技术已趋于成熟。语音识别的目标是将人类语音表达的内容转换为机器可读的输入，用于构建机器的“听觉系统”。语音识别技术经历了长达 60 年的发展。近年来，机器学习和深度神经网络的引入，使得语音识别的准确率提升到足以在实际场景中应用。早在 2016 年年初，美国麻省理工学院（MIT）主办的知名科技期刊《麻省理工科技评论》评选出了“2016 年十大突破技术”，语音识别位列第三，与其他技术一起“到达一个里程碑式的阶段或即将到达这一阶段”。

深度神经网络声学模型的几个重大发展阶段如下：

- 2006 年，Geoffrey Hinton 提出深度置信网络（DBN），促进了深度神经网络的研究。
- 2009 年，Geoffrey Hinton 将深度神经网络应用于声音的声学建模，当时在 TIMIT 上获得了很好的结果。
- 2011 年年底，微软研究院又把深度神经网络技术应用在了大词汇连续识别任务上，大大降低了语音识别的错误率。从此以后，基于深度神经网络声学模型技术的研究变得异常火热。

微软于 2016 年 10 月发布的 Switchboard 语音识别测试中，更是取得了 5.9%的词错误率，第一次实现了和人类一样的识别水平，这是一个历史性突破。语音识别整个过程（见图 2-9）包含语音信号预处理、声学特征提取、声学和语言模型建模、解码等多个环节。简单来说，声学模型用来模拟发音的概率分布，语言模型用来模拟词语之间的关联关系，而解码阶段就是利用上述两个模型将声音转化为文本。

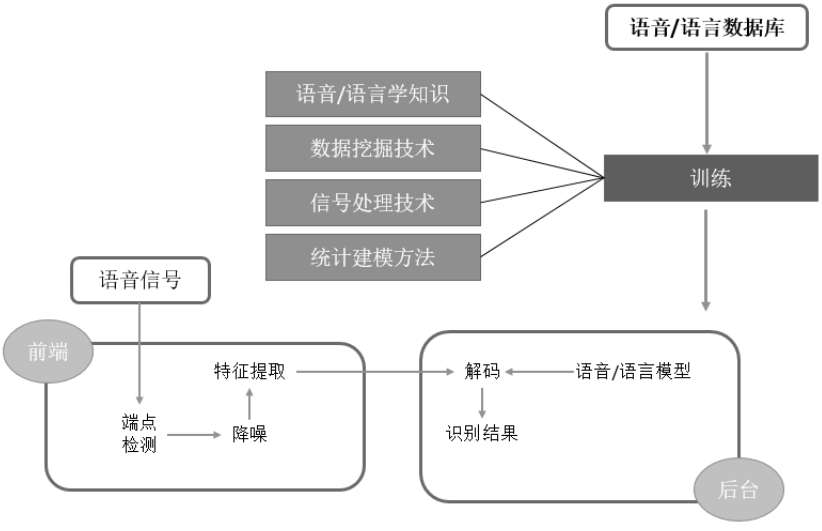


图 2-9　语音识别技术的运作流程

深度神经网络声学模型主要应用于声学、语言模型建模、解码等各个主要环节，模型主要包括深度神经网络、长短时记忆网络（LSTM）、双向长短时记忆网络（BLSTM）、深度卷

积神经网络（Deep CNN）、Residual/Highway 网络等模型，具体特点见表 2-3。

表 2-3　深度神经网络各部分及其特点

名称	特点
深度神经网络	包含至少 3 层以上的隐藏层，通过增加隐藏层数量来进行多层的非线性变换，大大地提升了模型的建模能力
长短时记忆网络	一种特殊的循环神经网络（RNN）。通过输入门、输出门和遗忘门可以更好地控制信息的流动和传递，具有长短时记忆能力，并在一定程度上缓解 RNN 的梯度消失和梯度爆炸问题
双向长短时记忆网络	相比 LSTM 还考虑了反向时序信息的影响，即“未来”对“现在”的影响，这在语音识别中也是非常重要的

总之，语音识别作为一类重要的基础技术，应用十分广泛，并且已有不少产品为人们所熟知，语音识别产业的增长主要靠渗透率的提升和应用的突破，主要的应用包括语音助手、语音输入、语音搜索等，可应用在各类移动 APP 应用和终端应用等对人机交互有较高要求的领域。对于语音识别技术而言，率先发展起来的服务机器人和语音助手已占据数据积累的领先地位，在家居、出行、运动等多个场景中，语音交互正在爆发，智能音箱、智能车载、智能手表等产品中，通过接入语音交互技术，实现随身陪伴、语音助理的功能。国内现已涌现出一批发展较好的智能语音相关企业，其中技术领先和产品成熟的企业主要有科大讯飞、百度、小米等。语音识别经过几年的技术积累已相对成熟，厂商仍在发展方言识别等更为精准的识别方式。

2. 自然语言处理

简单地说，自然语言处理（Natural Language Processing，NLP）就是用计算机来处理、理解以及运用人类语言，属于人工智能的一个分支，是计算机科学与语言学的交叉学科。实现人机间自然语言通信意味着要使机器既能理解自然语言文本的意义，也能以自然语言文本来表达给定的意图、思想等。前者称为自然语言理解，后者称为自然语言生成。

无论是实现自然语言理解，还是自然语言生成，都十分困难。从现有的理论和技术现状来看，通用的、高质量的自然语言处理系统仍然是较长期的努力目标，但是针对一定应用，具有相当自然语言处理能力的实用系统已经出现，有些已商品化，甚至开始产业化。

深度学习、算力和大数据的爆发极大地促进了自然语言处理技术的发展。表 2-4 中是几种常用的深度神经网络 NLP 模型。

表 2-4　几种常用的深度神经网络 NLP 模型

Word2Vec	Word2Vec 可以在百万数量级的词典和上亿的数据集上进行高效地训练。该工具得到的训练结果为词向量（Word Embedding，也称为词嵌入），可以很好地度量词与词之间的相似性
循环神经网络（Recurrent Neural Networks）	RNN 是 NLP 任务最常用的方法之一。RNN 模型的优势之一就是可以有效利用之前传入网络的信息
门控循环单元（Gated Recurrent Units）	目的是为 RNN 模型在计算隐藏层状态时提供一种更复杂的方法。这种方法将使模型能够保持更久远的信息

自然语言处理（NLP）领域还有很多其他种类的深度学习模型，有时候卷积神经网络（CNN）也会用在 NLP 任务中，但没有循环神经网络（RNN）这么广泛。总之，在自然语言处理领域，多轮对话理解日益完善，但语义理解仍然具有一定的缺陷，距离机器理解人类，实现自然的人机交互还有一些路要走。

2.2.3 计算机视觉

视觉是人脑最主要的信息来源，计算机视觉是指通过计算机或图像处理器及相关设备来模拟人类视觉，以让机器获得相关的视觉信息并加以理解，是机器能够“看懂”周围环境的计算基础，最终解决机器代替人眼的问题。

从技术流程来看，计算机视觉是将识别对象（如图像）转换成数字信号进行分析处理的技术。根据识别的种类不同，又分为图像识别、人脸识别、文字识别等。通过计算机视觉技术可以对图片、实物或视频中的物体进行特征提取和分析，从而为后续动作提供关键的感知信息。从技术流程来看，视觉识别通常需要几个过程：图像采集、目标提取、目标识别、目标分析，如图 2-10 所示。

图 2-10　视觉识别的几个过程

对于特征识别，有生物特征识别技术，用于识别人类的指纹、虹膜、人脸等；有 OCR 识别技术，用于识别图片和文字；有物体识别技术，用于识别图片或视频中的物体。

1. 视频分析

在进行视频识别与分析时，需要使用前端摄像头设备收集和传输数据，同时需要通过大数据训练，具备云计算能力的深度学习图像分析系统来实时进行视频检测和数据分析（见图 2-11）。由于机器不疲劳，而且可以全面识别整帧图像信息，通过使用该技术处理海量监控视频，可大大降低交管、公安部门的监控负担，具体的应用场景包括车辆识别、非法停车检测、嫌犯追踪等。

图 2-11　视频图像分析

在深度学习出现后，机器视觉的主要识别方式发生了重大转变，自学习状态成为视觉识别的主流，即机器从海量数据中自行归纳特征，然后按照该特征规律进行识别，图像识别的精准度也得到极大的提升（目前到了 95%以上）。机器不再只是通过特定的编程完成任务，而是通过不断学习来掌握本领，这主要依赖高效的模型算法进行大量数据训练。

近年来，与计算机视觉相关的视频监控和身份识别等行业的市场规模均逐渐扩大，伴随着技术的发展，计算机视觉技术和应用逐渐趋于成熟，被广泛应用到金融、安防、电商等场景中，技术进一步实现场景化落地，计算机视觉也成为目前人工智能领域最为火热和应用最为广泛的领域之一。计算机视觉厂商主要走技术和解决方案提供商的路径，通过研究通用型的技术，深耕图像处理和图像分析，提供软硬件全套服务，开放程序接口供其他厂商使用。另外，一部分厂商走技术应用的路径，将技术接入不同的领域和场景中，以技术为基础实现场景落地，为用户提供服务。

2. 人脸识别

人脸识别是基于人的脸部特征信息进行身份识别的一种识别技术。人脸识别技术被广泛应用于金融、安防、交通、教育等相关领域，主要应用场景包括企业、住宅的安全管理，公安、司法和刑侦的安全系统、自助服务等，刷脸支付、刷脸进站等项目逐渐实现。人脸识别包括 1∶1 的人脸对比和 1∶N 的人脸对比。1∶1 主要指用户真实的脸部信息与用户提交的身份证信息进行比对，常见于银行等金融机构和公安系统。1∶N 更常见于刑侦和国家安防领域，能够通过与 faceID 库的对比，快速找到犯罪分子或失踪人员，1∶N 识别精度的难度要远远高于 1∶1 人脸识别。厂商也针对 1∶N 的精确度做了技术深耕，百度曾宣布百度大脑的 1∶N 人脸识别监测准确率已达 99.7%。目前，人脸关键点检测技术可以精确定位面部的关键区域，还可以做到支持一定程度的遮挡以及多角度人脸，活体检测及红外光识别技术有效解决了照片、手机视频等二维人像的作弊行为，使三维人脸识别的准确率大幅度提升。但双胞胎识别、整容和易容前后的识别依然是人脸识别的难点，因此需要虹膜识别等其他识别技术进行补充。

人脸识别技术另一个关键层面在于 faceID 库的建立，三维人脸识别数据采集相对困难，需采集的数据量十分巨大，对计算机的计算存储能力要求较高，faceID 库的数据量是人脸识别技术算法训练的基础，数据越多，相应的准确度才会越高。各厂商仍需继续扩充自身的 faceID 库规模。在美国，亚马逊推出了人脸识别系统 Rekognition（注：亚马逊故意取名为 Rekognition，有别于“识别”对应的正确英文单词 Recognition），识别一个人脸只需要几分钟。亚马逊公司已经开始通过云计算模式推出计算机视觉识别功能，提供了基于机器学习的人脸识别服务。人脸识别技术不再是一个高价的服务了。

总之，随着计算机技术的发展，人类开始能够通过计算机实现不同模式（文本、声音、人物、物体等）的自动识别。但当前不存在一种单一模型和单一技术能够解决所有的模式识别问题，而是需要在具体场景中使用多种算法和模型。还有，计算机视觉可以与其他技术结合进行综合应用，比如与医疗系统结合形成疾病辅助监测，与汽车驾驶系统结合形成自动驾驶。

2.3 应用层

人工智能给各行各业带来了变革与重构，一方面将 AI 技术应用到现有的产品中，创新产品，发展新的应用场景；另一方面 AI 技术的发展也正在颠覆传统行业，人工智能对人工的替代成为不可逆转的发展趋势，尤其在工业、农业等简单、重复、可程序化强的环节中，而在国防、医疗、驾驶等行业中，人工智能可以提供能够适应复杂环境、更为精准、高效的专业化服务，从而取代或者强化传统的人工服务，服务形式在未来将趋于个性化和系统化。

人工智能与行业的深度结合，可以实现传统行业的智能化，包括 AI+金融、AI+医疗、AI+安防、AI+家居、AI+教育等，如图 2-12 所示。在各个垂直领域中，传统厂商具备产业链、渠道、用户数据优势，正通过接入互联网和 AI 搭载人工智能的浪潮进行转型。创业公司深耕垂直领域，快速崛起，致力于推动技术进步、场景落地。应用层厂商更直接地面对用户，或者遵循 2B、2C 的发展路径，相较于技术层和基础层，具有更多的用户数据，也需要进一步打磨产品，满足用户需求。

图 2-12　AI 应用场景

对于人工智能的应用来说，技术平台、产业应用环境、市场、用户等因素都对人工智能的产业化应用市场有很大的影响。如何实现人工智能产业自身的创新并应用到具体场景中将会是各行业发展的关键点。目前，人工智能技术的主要应用场景包括但不限于：安防、制造业、服务业、金融、教育、传媒、法律、医疗、家居、农业、汽车等。人工智能技术日益成熟，商业化场景逐渐落地，智能家居、金融、医疗、驾驶、安防等多个行业成为目前主要的应用场景。

2.3.1　安防

安防的应用场景较多，小到身份识别、家居安防，大到反恐国防。现代社会人口流动大，安防成为刚需。身份识别手段的多样性对于安防意义重大，因此安防领域对于图像识别的要求更高，也要求更多的手段通过多维度来进行识别，如图 2-13 所示，AI 技术的进步可以大大提高身份识别手段的多样性与准确率，对于安防的意义重大，尤其是安防在国防安全领域的应用，具有国家战略意义。

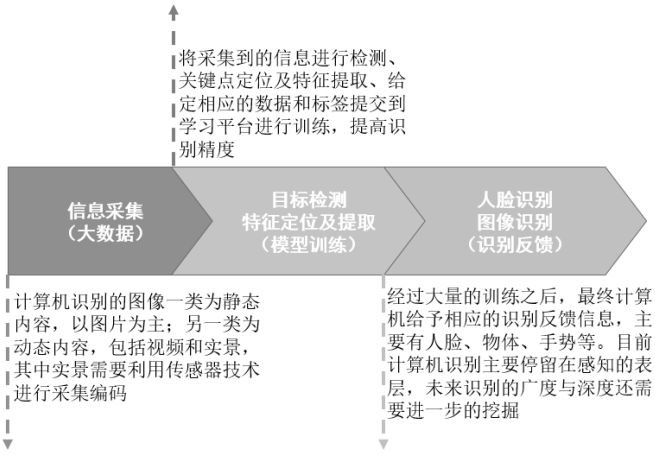


图 2-13　安防中的图像识别技术

在视频监控技术飞速发展的今天，视频监控画面的信息已成海量，远远超过了人力所能进行的有效处理范围。传统采用人工回放录像取证的方式具有效率低下、容易出错的缺点。而大数据技术恰好具有处理海量信息的能力，也能在人工智能技术的基础上实现实时监控、基准判断。智能视频分析（Intelligent Video Analysis，IVA）技术是解决海量视频数据处理的有效途径。IVA 采用计算机视觉方式，主要应用于两个方面，一是基于特征的识别，主要用于车牌识别、人脸识别。特征识别与视频智能分析应用于安防体系中，提高了安防的时效性、安全性和精准度。二是行为分析技术，包括人数管控、个体追踪、禁区管控、异常行为分析等，可以应用到监测交通规则的遵守、周界防范、物品遗留丢失检测、人员密度检测等。通过对视频内的图像序列进行定位、识别和追踪，智能视频分析能够做出有效分析和判断，从而实现实时监控并上报异常，使得安防由被动防范向提前预警方向发展，将实现对危险分子的主动识别，安防行为由被动向主动转变。

从应用领域来看，目前平安城市、智能交通仍然是安防行业最大的应用领域，与政府公安相关的交通、道路视频监控仍然是安防行业最重要的应用环节。计算机视觉广泛应用于机场、火车站等公共场合，在大规模视频监控系统中可实现实时抓拍人脸、布控报警、属性识别、统计分析、重点人员轨迹还原等功能，并做出及时有效的智能预警。这对于抓获有作案前科的惯犯帮助很大，目前多应用于公安事前、事中、事后敏感人员布控，失踪人员查找等。安全布防需要消耗大量的警力资源，尤其是运动会、国家会议、演唱会等重点区域和重点活动的安防，其中已经开始出现人工智能产品的身影，包括实时监测系统、巡逻机器人、排爆机器人等，未来这些机器人也将会更多地替代传统安防体系中重复且低效的工作，以节省警力资源。

有必要指出的是，安防体系中存储的信息将呈指数级增长，需要大数据平台及其配套的硬件设备进行整合。

2.3.2　金融

AI 在金融领域的应用主要集中在投资决策辅助、风控与智能支付三个方面。在投资决策

辅助方面，人工智能技术将协助金融工作者从数以万计的信息中迅速抓取有效信息，并进一步对数据进行分析，利用大数据引擎技术、自然语义分析技术等自动准确地分析与预测各市场的行情走向，从而实现信息的智能筛选与处理，辅助工作人员进行决策。在风控方面，人工智能也能帮助金融机构建立金融风控平台，进行风控管理，实现对投资项目的风险分析和决策、个人征信评级、信用卡管理等。在智能支付领域中，利用人工智能的人脸识别、指纹识别、声纹识别技术可实现“刷脸支付”“指纹支付”或者“语音支付”。

金融行业与整个社会存在巨大的交织网络，在长期的发展过程中沉淀了海量数据，如客户身份数据、资产负债情况数据、交易信息数据等，金融业对数据的强依赖性为人工智能技术应用到金融领域做好了准备。按金融业务执行前端、中端、后端的模块来看，人工智能在金融领域的应用场景主要有智能客服、智能身份识别、智能营销、智能风控、智能投顾、智能量化交易等。

身份认证主要通过人脸识别、指纹识别、声纹识别、虹膜识别等生物识别技术快速提取客户的特征。近年来，金融机构对远程身份识别、远程获客需求日益增加，而人脸信息凭借易于采集、较难复制和盗取、自然直观等优势，在金融行业中的应用不断增加。人脸识别可实现客户“刷脸”即可开户、登录账户、发放贷款等，让金融机构远程获客和营销成为可能。在互联网金融领域，“刷脸”也可以应用到刷脸登录、刷脸验证、刷脸支付等诸多领域。同时，人脸识别可以成为银行安全防控手段的有效选择。银行安防的难点之一是在动态场景下完成多个移动目标的实时监控，人脸识别技术在银行营业厅等人员密集的区域可有效实现多目标实时在线检索、比对，在 ATM 自助设备、银行库区等多个场景下都可以应用。中国人民银行发布《中国人民银行关于优化企业开户服务的指导意见》（银发〔2017〕288 号），对新设企业开立人民币银行结算账户服务提出意见。央行鼓励银行积极运用技术手段提升账户审核水平，包括鼓励银行将人脸识别、光学字符识别（OCR）、二维码等技术手段嵌入开户业务流程，作为读取、收集以及核验客户身份信息和开户业务处理的辅助手段。

人工智能技术可以助力金融行业形成标准化、模型化、智能化、精准化的风险控制系统，帮助金融机构、金融平台及相关监管层对存在的金融风险进行及时有效的识别和防范。如图 2-14 所示，人工智能应用于金融风险控制的流程主要包括：数据收集、行为建模、用户画像及风险定价。智能风控可以协助金融监管机构防范系统性金融风险。在信贷领域，智能风控可以应用到贷前、贷中、贷后全流程。贷前，助力信贷机构进行信息核验、信用评估、实现反欺诈；贷中，可以实现实时交易监控、资金路径关联分析、动态风险预警等；贷后，可以助力信贷机构进行催收、不良资产等价等。智能风控系统包含一组模型，会根据身份认证、还款意愿和还款能力三大维度，给申请贷款的用户进行信用评分，依据分值来决定是否应放款。有效提升了贷款审批速度和贷款获批率，并降低了贷款的逾期率。

图 2-14　智能风控分析流程

金融行业目前正在打造闭合的全产业链，提供的服务不仅针对客户成长中的某一阶段，而是全生命周期的服务。如图 2-15 所示，每个客户都要经历获取、提升、保持、流失和衰退几个阶段。在不同的发展阶段，风险特点及对金融服务需求的特点不尽相同。基于 AI 技术，我们可以对不同阶段的客户开展个性化金融业务。

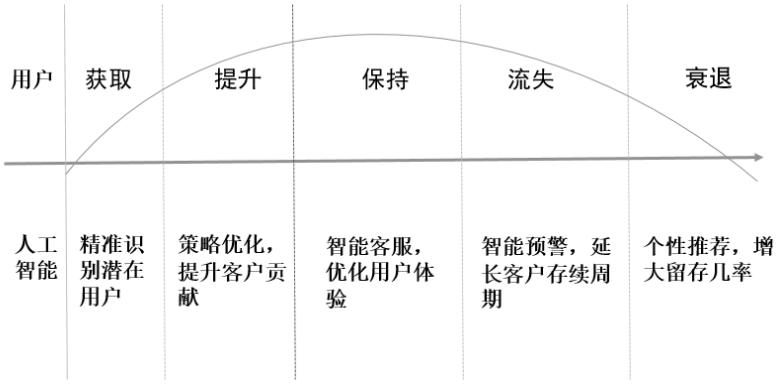


图 2-15　全生命周期客户服务

2.3.3　制造业

人工智能的应用有望实现制造业从半自动化生产到全自动化生产的转变，工业以太网的建立、传感器的使用及算法的革新将实现工业制造过程中所有生产环节的数据打通，人与机器、机器与机器实现互联互通，一方面人机交互更为便利，另一方面机器之间将协作办公，既能够精细化操作，又能及时地预测产品需求并调整产能。人工智能将推动机器在制造业中进一步取代人工，提高生产效率、降低生产成本，并通过低成本的个性化生产实现智能定制化服务。

2.3.4　智能家居

如图 2-16 所示，AI 在智能家居场景中，一方面将进一步推动家居生活产品的智能化，包括照明系统、音箱系统、能源管理系统、安防系统等，实现家居产品从感知到认知再到决策的发展；另一方面在于智能家居系统的建立，搭载人工智能的多款产品都有望成为智能家居的核

心，包括机器人、智能音箱、智能电视等产品，智能家居系统将逐步实现家居自我学习与控制，从而提供针对不同用户的个性化服务。

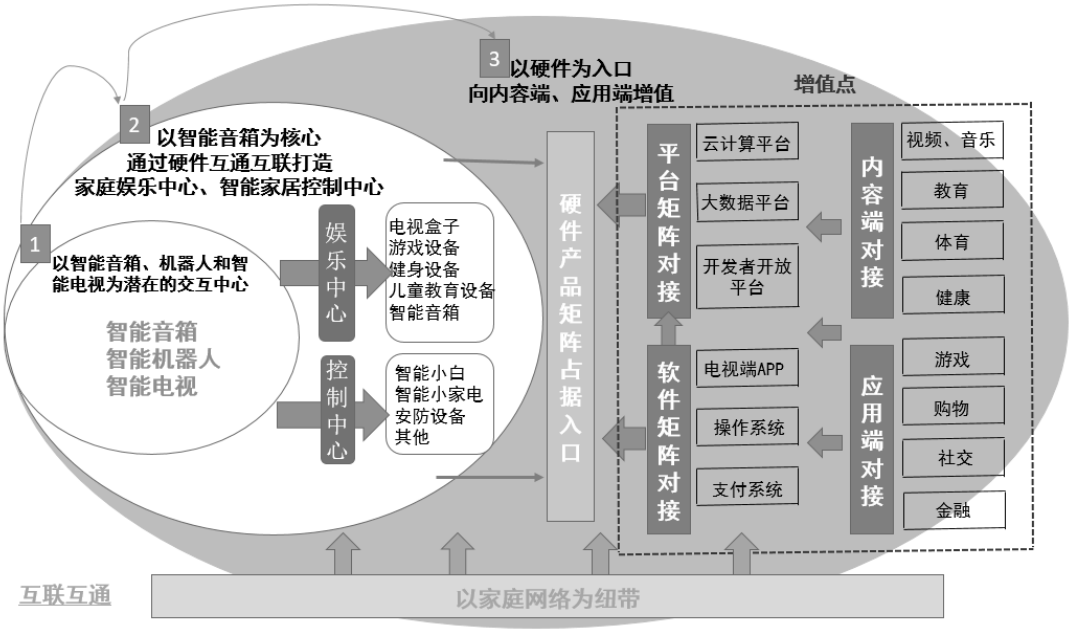


图 2-16　智能家居生态布局

目前，智能家居仍处于从手机控制向多控制结合的过渡阶段，手机 App 仍是智能家居的主要控制方式，但基于人工智能技术开发出来的语音助手、搭载语音交互的硬件等软硬件产品已经进入市场。通过语音控制，多产品联动的使用场景逐步变为现实。而在未来，人工智能将推动智能家居从多控制结合向感应式控制再到机器自我学习自主决策阶段发展。

传统的鼠标操作、触屏操作逐渐向语音交互这种更为自然的交互方式演进，语音交互的未来价值在于用户数据挖掘，以及背后内容、服务的打通，以语音作为入口的物联网时代将会产生新的商业模式。智能音箱、服务机器人、智能电视等智能化产品成为现阶段搭载语音识别技术和自然语言处理技术的载体，作为潜在的智能家居入口，智能音箱、服务机器人和智能电视等产品在提供原有服务的同时，接入更多的移动互联网服务，并实现对其他智能家居产品的控制。这些产品为付费内容、第三方服务、电商等资源开拓了新的流量入口，用户多方数据被记录分析，厂商将服务嫁接到生活中不同的场景中，数据成为基础，服务更为人性化。

2.3.5　医疗

目前，医疗行业存在医疗资源不足、医疗资源区域分布不均、医生培养周期长、医疗成本高、医疗误诊率高、疾病变化快等诸多痛点。同时，随着人口老龄化逐渐加剧、慢性疾病增长，对医疗服务的需求也逐渐增加。待解决的医疗痛点及逐渐增加的医疗服务需求，成为了人工智能技术应用于医疗行业的现实需求。医疗行业基于人工智能技术，将形成辅助诊断系统，通过图像识别、知识图谱等技术，将辅助医生决策，而医学大数据的发展将患者信息数字化，提高

发现潜在疾病的概率，并提供针对性解决方案。人工智能技术将为医疗领域中的医生与患者带来新的疾病治疗方式。

另一方面，政府在积极推动“人工智能医疗”的应用进程。2017 年 7 月 8 日，国务院发布《新一代人工智能发展规划》，提出发展便捷高效的智能服务，围绕教育、医疗、养老等需求，加快人工智能创新应用；提出推广人工智能治疗这种新模式、新手段，建立智能医疗体系，开发人机协同的手术机器人、智能诊疗助手等，实现智能影像识别、病理分型和智能多学科会诊；智能健康和养老方面，提出加强群体智能健康管理，突破健康大数据分析、物联网等技术，构建安全便捷的智能化养老基础设施体系，加强老年人产品智能化和智能产品适老化等。

在医疗领域，人工智能技术应用前景广泛。从全球企业实践来看，“人工智能+医疗”具体应用场景主要有医学影像、辅助诊疗、虚拟助理、新药研发、健康管理、可穿戴设备、急救室和医院管理、洞察与风险管理、营养管理及病理学、生活方式管理与监督等。

“人工智能+医学影像”是将人工智能技术应用在医学影像的诊断上，实际上是模仿人类医生的阅片模式。人工智能技术应用于医学影像主要包括数据预处理、图像分割、特征提取和匹配判断 4 个流程。人工智能强大的图像识别和深度学习能力，有助于解决传统医学影像中存在的准确度低、工作量大的问题，弥补影像科医生不足，提升读片准确度，提高医生工作效率，缓解放射科医生压力。同时，技术手段助力疾病早筛，及早为患者发现病灶，提高患者存活率。虽然影像识别在单病种的市场空间不大，但在政策推动背景下，影像科、检验科等科室市场化运营，成立病理中心，高端诊断服务将成为影像识别技术的巨大机会。

“人工智能+辅助诊疗”就是将人工智能技术应用于辅助诊疗中，让机器学习专家医生的医疗知识，通过模拟医生的思维和诊断推理来解释病症原因，最后给出可靠的诊断和治疗方案。在诊断中，人工智能需要获取患者病症，解释病症，通过推理判断疾病原因及发展走向，形成有效治疗方案。如图 2-17 所示，辅助诊疗的一般模式为：获取病症信息→做出假设→制定治疗方案。IBM Watson 融合了认知技术、推理技术、自然语言处理技术、机器学习及信息检索等技术，是目前“人工智能+辅助诊疗”应用中最为成熟的案例。IBM Watson 已经通过了美国职业医师资格考试，并在美国多家医院提供辅助诊疗服务。IBM Watson 可以在 17 秒内阅读 3 469 本医学专著、248 000 篇论文、69 种治疗方案、61 540 次试验数据、106 000 份临床报告。“人工智能+辅助诊疗”服务基于电子处方、医学文献、医学影像等数据，寻找疾病与解决方案之间的对应关系，构建医学知识图谱，在诊断决策层面有效优化医生的诊断效率。未来，“人工智能+辅助诊疗”的市场空间巨大，尤其在基层常见病诊疗方面能够发挥较大效能，有效提高基层医疗效率，降低医疗成本。

总之，人工智能广泛应用于医疗领域，有助于解决现阶段医疗资源不足的核心痛点。伴随着医疗成本高、人才培养周期较长等问题，人工智能高效分析能力有效提高医疗行业的产能。人工智能广泛应用于医疗领域有助于带动基层医疗服务。“人工智能+医疗”有望成为一种可复制的医疗资源，增加基层医生的诊断精准度。

图 2-17　人工智能+辅助诊疗

2.3.6　自动驾驶

自动驾驶也称为无人驾驶，指依靠人工智能、视觉计算、雷达、监控装置和全球定位系统协同合作，让计算机可以在没有任何人类主动的操作下，自动安全地操作机动车辆。先进驾驶辅助系统（Advanced Driver Assistant System，ADAS）利用安装于车上的各式各样的传感器，在第一时间收集车内外的环境数据，从而能够让驾驶者以最快的时间察觉可能发生的危险。ADAS 采用的传感器主要有摄像头、雷达、激光和超声波。ADAS 与自动驾驶的区别在于：ADAS 可以视为自动驾驶实现的一个路径，ADAS 可以最终演化为自动驾驶。

自动驾驶系统分为 4 个层级：感知层、识别层、决策层、执行层。自动驾驶各层级及其相互关系如图 2-18 所示。

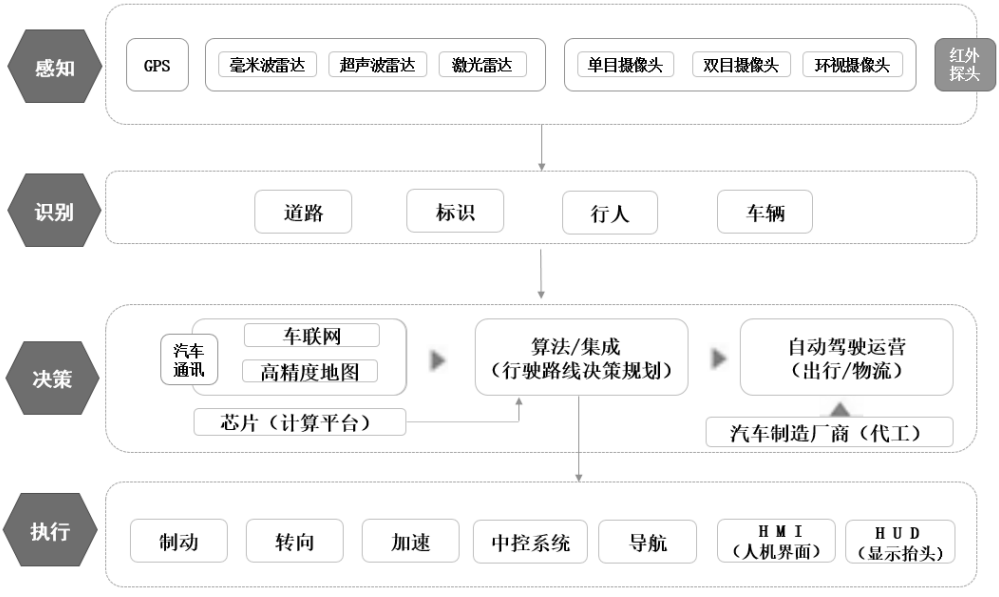


图 2-18　自动驾驶层次结构图

1. 感知（传感）

（1）车载摄像头

以摄像头为代表的机器视觉传感器是自动驾驶的核心感知技术。视觉系统不仅能够识别目标距汽车的距离，还能够识别目标的纹理和色彩，这是车载雷达所不能做到的。相比于其他传感器，摄像头的优势在于：技术成熟，成本较低；可以通过较小的数据量获得最全面的信息。但是，摄像头识别也存在一定局限性：受光线、天气影响大；物体识别基于机器学习数据库，需要的训练样本大，训练周期长，难以识别非标准障碍物；由于广角摄像头的边缘畸变，得到的距离准度较低。

目前摄像头的应用主要有以下几种：

- 单目摄像头：一般安装在前挡风玻璃上部，用于探测车辆前方的环境。
- 后视摄像头：一般安装在车尾，用于探测车辆后方的环境，应用于倒车可视系统。
- 立体摄像头，或称双目摄像头：利用两个经过精确标定的摄像头同时探测车辆前方的环境，实现更高的识别精度和更远的探测范围。
- 环视摄像头：一般至少包括 4 个摄像头，分别安装在车辆前、后、左、右侧，实现 360° 环境感知，应用于自动泊车和全景泊车系统。

（2）超声波雷达

超声波雷达主要是利用超声波原理，由探头发送超声波，撞击障碍物后反射回此超声波，而后计算出车体与障碍物间的实际距离。超声波雷达现在主要应用于倒车雷达。

（3）激光雷达

激光雷达的原理与超声波雷达相似，根据激光遇到障碍后的折返时间，计算与目标的相对距离。激光雷达的激光光束与超声波雷达的声波和毫米波雷达的电磁波相比更加聚拢，声波和电磁波在传播路径上遇到尺寸比波长小的物体时，将会发生衍射现象，因此，无法探测大量存在的小型目标，而激光雷达可以准确测量视场中物体轮廓边沿与设备间的相对距离。用于雷达系统的激光波长一般有微米的量级，因而它能够探测非常微小的目标，测量精度也远远高于毫米波雷达及其他车载标准雷达。激光雷达的劣势在于价格昂贵。激光雷达的测量精度与其雷达线束的多少有关，线束越多，测量精度越高，ADAS 自动驾驶系统的安全性也越高。同时线束越多，其价格也越贵。

激光雷达按有无机械旋转部件分为机械激光雷达和固态激光雷达。固态激光雷达无须旋转部件，尺寸较小，性价比较高，测量精度相对低一些。低成本化是激光雷达的一大趋势，目前行业有三种方式来降低整个激光雷达的成本与价格：①降维，即使用低线束、低成本激光雷达配合其他传感器；②采全固态激光雷达代替机械激光雷达；③通过规模效益降低激光雷达的单个成本。

（4）毫米波雷达

毫米波雷达指工作在毫米波波段的雷达。采用雷达向周围发射无线电，波长在 1mm~10mm，频率在 30GHz~300GHz，比较常见的汽车毫米波雷达工作频率在 24GHz、77GHz、

79GHz 三个频率附近。毫米波雷达通过测定和分析反射波以计算障碍物的距离、方向、角度、相对速度和大小。毫米波雷达可以做到让车辆自适应巡航及跟随前车。当汽车与周围的物体可能发生碰撞时，通过警告提醒装置告知驾驶员或车辆采取自动紧急制动避免碰撞。当碰撞不可避免时，通过对刹车、头靠、安全带等进行控制，减轻因碰撞而带来的危害。

2. 识别与决策

（1）识别芯片

芯片在自动驾驶系统中的行业集中度高，主要有 Mobileye、ADI 等公司。Mobileye 作为 ADAS 界的大佬，占领了全球汽车安全驾驶系统 70%以上的市场份额。在这个领域深耕细作十几年，有相当深厚的历史背景，这些经验并不是其他公司短时间可以超越的。

（2）决策算法

决策部分的算法和芯片主要由一些大公司，以及由大公司出来的科学家成立的创业公司研发。比如，谷歌公司的开放式无人驾驶算法，该项目由谷歌街景的共同发明人 Sebastian Thrun 领导。谷歌的工程人员使用 7 辆试验车，目前已经行驶上百万千米。再比如，Driver.ai 公司利用深度学习来开发无人驾驶技术，百度公司开发“阿波罗”无人驾驶系统，Comma.ai 公司的基于卷积神经网络的无人驾驶算法。

（3）决策芯片

自动驾驶决策芯片提供商有高通、英伟达、Intel+Mobileye 等公司。

（4）高精度地图

汽车需配备足够准确显示周围环境的高精度地图，误差不能大于 10 厘米（cm）。传感器和地图的结合使自动驾驶汽车能够及时修正数据上的误差，辨识车辆的准确位置并导航。并且，高精度地图能够核对传感器所接收的数据并帮助汽车精确监测周边环境。目前高精度地图已经被苹果、谷歌、国内的 BAT 等大公司垄断，表 2-5 是这些公司并购地图厂商的事件。

表 2-5　国内外巨头收购高精度地图公司一览表

公司	时间（年）	事件
谷歌	2013	13 亿美元收购众包地图公司 Waze
苹果	2013	收购在线交通导航应用开发商 HopStop
	2013	收购综合性地图公司 BoardMap
	2015	收购开发高精度全球定位系统的公司 Coherent Navigation
	2015	3000 万美元收购地图分析公司 Mapsense
阿里巴巴	2014	全资收购高德地图
腾讯	2014	以 11.73 亿元人民币收购四维图新 11.28%股权
德国三大汽车厂商戴姆勒-奔驰、宝马、奥迪组成的财团	2015	以 32 亿美元收购诺基亚地图业务

（5）车联网

车联网 V2X 是自动驾驶和未来智能交通运输系统的关键技术。V2X 是指联网无线通信技

术，实现车对外界的信息交换，V2X 包括 V2V（车-车）、V2I（车-基础设施）、V2R（车-道路信息）、V2P（车-行人）等方式的车联网通信技术。它可以弥补单车智能的软肋，当车辆环境感知系统无法做到全天候、全路况的准确感知时，V2X 可以利用通信技术、卫星导航对感知系统进行协调互补。

按照 IHS Automotive 保守估计，全球 L4/L5 自动驾驶汽车产量在 2025 年将达到接近 60 万辆，并在 2025—2035 年间获得高速发展。在这个“无人驾驶黄金十年”内复合增长率将达到 43%，并在 2035 年 L4/L5 自动驾驶汽车产量将达到 2100 万辆，另有接近 7600 万辆汽车具备部分自动驾驶功能，同时将带动产业链衍生市场的大规模催化扩张。

3. 自动驾驶趋势分析

（1）趋势 1：低成本激光雷达方案

激光雷达作为自动驾驶最昂贵的配件，精度高，性能好，是最被看好的车载传感器。激光雷达未来趋于固态化、小型化、低成本，目前特斯拉尚未采用激光雷达方案，主要在于成本太高，因此作为将来自动驾驶的核心配件来说，如果能够提供低成本的激光雷达方案，将会快速推动自动驾驶市场。

（2）趋势 2：多传感器融合方案

①融合感知是大势所趋。

毫米波雷达能解决所有情况下 30%左右的问题，激光雷达能解决 60%~70%的问题，单目配合雷达能够实现测距和预测碰撞时间，双目配合单目的识别技术也能够丰富双目在测距之外的感知能力，因此未来融合会是趋势。

②各类车厂的选择方案有所不同。

技术实力弱的车厂更多依靠 Tier1 来集成，实力强的车厂会自己来做整合。

（3）趋势 3：深度学习算法应用于 ADAS

传统算法仍然适用于 ADAS 阶段，深度学习满足最后关键 5%的识别精度。深度学习出现以后，视觉识别任务的精度都进行了大幅度的提升。因此，大量公司会将算法模型开放，其背后的动机在于收集更多数据训练自身的算法模型，同时改进算法，最终将改进的算法与车厂合作，将算法的商业价值变现。对于开放算法，将深度学习直接用于 ADAS 领域的公司，将迎来一次机会，如 Comma.ai、Driv.ai 这样的公司。

（4）趋势 4：自动驾驶深度学习专用集成电路（ASIC）处理器

专用集成电路（ASIC）是根据特定客户要求和特定电子系统的需要而设计、制造的集成电路，即芯片。在批量生产时，与通用集成电路相比具有体积更小、功耗更低、可靠性更高、保密性更强、成本更低等优点。将深度学习算法应用在自动驾驶并且利用专用芯片技术来实现深度学习功能的 ASIC 处理器，相比于 FPGA 而言，ASIC 处理器牺牲了灵活性以换取尺寸和功耗下降，ASIC 处理器去除了通用芯片中与算法实现无关的组件，在牺牲灵活性的同时，极大地提升了自动驾驶深度学习的效率。

（5）趋势 5：物流行业的无人驾驶应用

使用物流无人驾驶能为物流行业解决以下 3 个问题：

①路线较为固定，降低了环境的复杂性，有利于提升无人驾驶的安全性。

②该细分领域司机疲劳驾驶情况比较明显，无人驾驶可以提高安全性。

③有效降低运营的人力成本，提升行业效率。

2.4 AI 产业发展趋势分析

如图 2-19 所示，人工智能产业链可以分为基础设施层、技术层和应用产品层，各层的发展趋势如下：

- 基础设施层，主要有基础数据提供商、半导体芯片供应商、传感器供应商和云计算服务商。在过去的 5~10 年，人工智能技术得以商业化，主要得益于传感器等硬件价格快速下降、云服务的普及以及 GPU 等芯片使大规模并行计算能力得以提升。人工智能产业在基础设施层面的搭建已经基本形成。
- 技术层，主要有语音识别、自然语言处理、计算机视觉、深度学习技术提供商。与其他技术相比，语音识别在技术和应用方面都已经较为成熟，谷歌、亚马逊、苹果、百度、阿里等巨头的布局很深，科大讯飞等企业也显示了良好的增长势头。另外，计算机视觉尤其是人脸识别、自然语言处理等方向也将是技术和应用发展较快的领域。
- 处于应用产品层的企业，主要是把人工智能相关技术集成到自己的产品和服务中，然后切入特定场景（金融、家居、医疗、安防、车载等）。未来数据完整（信息化程度原本就比较高的行业或者数据洼地行业）、反馈机制清晰、追求效率动力比较强的场景或将率先实现 AI 技术的大规模商业化。目前来看，自动驾驶、医疗、安防、金融、营销等领域是业内人士普遍比较看好的方向。

图 2-19　AI 的结构

AI 产业发展还呈现了以下趋势。

（1）平台崛起，技术、硬件、内容多方面资源进一步整合

人工智能覆盖的行业及场景巨大，单一企业无法涉及人工智能产业的方方面面，厂商基于

自身的优势切入产业链条，并与其他厂商进行合作，技术、硬件、内容多方面资源进行整合，共同推动人工智能技术落地。在技术、内容及硬件的发展下，平台进一步崛起，生态化布局日益重要。

（2）人工智能技术继续向垂直行业下沉

通用型人工智能技术已不能满足各行业的需求，不同行业在应用侧重点上有所不同，数据资源也同样不同，需要市场从业者针对行业特点，设计不同的行业解决方案。人工智能技术将继续从场景出发实现技术落地，在垂直行业中，医疗、金融、安防、环境、教育、家居等行业已初具规模，未来发展前景巨大。

（3）产学研相结合，人才仍是抢夺的重点

AI、物联网成为主流的发展趋势，人才在其中发挥的价值越来越大，而产业发展速度与人才培养速度之间的矛盾在产学研发展路径下将逐渐缩小，专业型人才开始增多，具有核心知识的专家仍然成为厂商抢夺的重点。在人工智能领域中，国内人才集中在技术层及应用层，基础层人才薄弱，国内高校在人工智能人才培养方面也持续缺失，专业布局较晚，专家有限，国内外在教育系统之间的差距较大，这也导致国内在人工智能领域基础层研究的薄弱。在意识到人才方面的缺失之后，国家及企业采取各类措施进行追赶，比如采取“千人计划”“新一代人工智能发展规划”等政策吸引优秀专业人才回国，企业围绕其核心业务抢夺人工智能人才。未来需要继续建立核心技术人才培养体系，加强人工智能一级学科建设，实现产学研的有效融合，为人工智能产业持续不断输送优质人才。

（4）厂商进入卡位战，不断发掘新的商业模式

人工智能将通过 AI+的形式影响各行各业，技术厂商迅速崛起，但应用才是技术落地的关键。技术被集成到各类产品中，技术厂商本身议价能力不强，所获得的利益有限，因此技术厂商积极搭建平台，或发展硬件、布局生态，以集成商的角色获取更多的行业红利。

软件以及互联网对传统商业的冲击已呈颠覆之势，而 AI 所覆盖的领域更为庞大，冲击也更甚。随着人工智能的发展，由软件和互联网打造的流量价值被打破，数据为王成为新趋势，云端服务、后端收费等依托智能硬件而发展起来的新兴服务模式逐渐兴起。人工智能产业中的入局者需要在推动技术落地的同时不断发掘新的商业机会。

（5）中国仍需加大在算力、算法、大数据领域的发展，弥补技术弱势

人工智能底层基础层技术仍旧掌握在欧美国家手中，尤其是芯片、先进半导体等核心零部件，以及算法、开源框架等核心技术，这些技术将直接影响人工智能技术的发展进程。虽然国家通过“中国制造 2025”等战略推动先进技术的研发，但是国内研发基础相对薄弱，在基础算法研究领域仍处于劣势。教育不完善、人才短缺、研究领域集中、数据开放不足等问题成为限制中国人工智能发展的重要因素。因此，中国仍需加大在算法算力、大数据领域的布局，掌握核心技术能力。

（6）伦理之争不止，AI 终将取代部分人工

由人工智能引发的伦理问题一直无法达成共识。目前，业内普遍认为人工智能将经历三个

时间节点：第一个时间节点是这一波人工浪潮，其产业红利在 3~5 年之内会尘埃落定；第二个时间节点是 10 年之内，一半以上的现有工作会被人工智能替代；第三个时间节点是 30 年之内，人工智能将具备自我觉醒的能力。在硅谷备受推崇的观点也是在未来 30 年内，90%的工作会因人工智能技术的进步而被淘汰。

伴随着人工智能的兴起，技术威胁论引发的一系列谈论从未停止过，技术裹挟着变革力量推动时代向前发展，这也意味着与时代脱离的观念和行为将会被抛弃：工业革命瓦解小农经济，互联网时代颠覆线下经济实体，人工智能技术将会取代传统耗时、重复性、机械化的运动，机器成为生产主力，同时与之相对应的新兴职业增多，专业技术人才的竞争力加大。

在人工智能取代人类或人工智能增强人类能力的讨论之余，用户所能做的只有强化自身的能力，发挥主体的不可替代性。而在人工智能领域中的基因重组、机器人学等超人类主义项目，仍需要政府加大监管力度。

第 3 章
机器学习概述

从 1956 年达特茅斯会议“人工智能”这一概念被提出，到现在已经有六十多年了，这期间经过了多个阶段。2010 年以后，随着深度学习使得语音识别、图像识别和自然语言处理等技术取得了惊人发展，前所未有的人工智能商业化和全球化的浪潮席卷而来。总的来说，人工智能是最早出现的概念，其次是机器学习，最后出现的是深度学习，是当今人工智能大爆炸的核心驱动（见图 3-1）。

图 3-1　人工智能发展阶段

3.1　走进机器学习

大数据是怎么与图像识别、语音识别等具体的 AI 技术联系起来的呢？靠的就是机器学习。

3.1.1　什么是机器学习

机器学习（Machine Learning，简称 ML）是让机器从大量样本数据中自动学习其规则，并根据学习到的规则预测未知数据的过程。以上是机器学习的一个定义。如果在网络上搜索“机器学习”，你会看到很多版本的定义。毕竟，越是火热的词汇，人们越难给它下一个精准而权威的定义。在机器学习的众多定义中，我们认为这个定义是相对让初学者容易接受的一种说法。这个定义中的关键字是“学习”二字。机器学习的目标是发现数据中暗藏的规律，并由此来对

未知进行预测。这个过程要通过“学习”来实现，而学习用到的材料则是数据。

3.1.2 机器学习的感性认识

如果你是第一次接触机器学习，其实机器学习离你绝不遥远。机器学习可以类比小孩认知事物的过程，只不过这里的认知过程是交给机器来完成的，让机器发现事物的规律。在我们小的时候，我们对周边的事物还并不了解，不知道什么是苹果，什么是猫，什么是汽车。但在经过某种“训练”之后，我们逐渐能够自信而准确地判断出我们看到的、听到的、感知到的东西是什么。这个“训练”可能来自外部的教导，也可能源自于我们自身的探索和尝试。

举个例子来说，在我们很小的时候，家长带我们参观动物园。刚刚学会说话和识字的我们会看到很多令我们“惊奇”的未知生物。家长告诉我们那个长鼻子的动物是大象，我们似懂非懂地记住了。走了几步之后，我们又在另一个地方发现了另一只长得很像之前看到过的动物，家长又告诉我们“这也是一只大象”。回到家中，我们拿着洗出来的照片，指着照片上的“长鼻子”问妈妈，“这个是什么来着？”妈妈回答我们说是大象。这时我们已经逐渐发现了这个叫作大象的东西长相的特点（见图 3-2）。过了几天后，我们在电视机上的动物园短片中看到了一个熟悉的轮廓，凭借自己对“长鼻子”和其他特征的一些判断，我们向妈妈喊出：“看，是大象”。

图 3-2 大象识别样图

机器学习的原理和上面的过程极其相似。假如你想让机器完成从图像中识别大象的任务，就像人的知识不是与生俱来的一样，机器不是一上来就什么都会的，想让机器能够完成任务，要先提供大量的数据让它学习，告诉它大象长什么样子，大象之外的其他动物长什么样子。在经过大量样本的学习后，机器可以从一张新的图像中判定其中是否有大象，如图 3-3 所示。

图 3-3 大象识别样图

3.1.3 机器学习的本质

如图 3-4 所示，类似人脑思考，机器经过大量样本的训练（Training），获得了一定的经验（模型），从而产生了能够推测（Inference，推断或推理）新的事物的能力，就是机器学习。这种预测能力，本质上是输入到输出的映射。

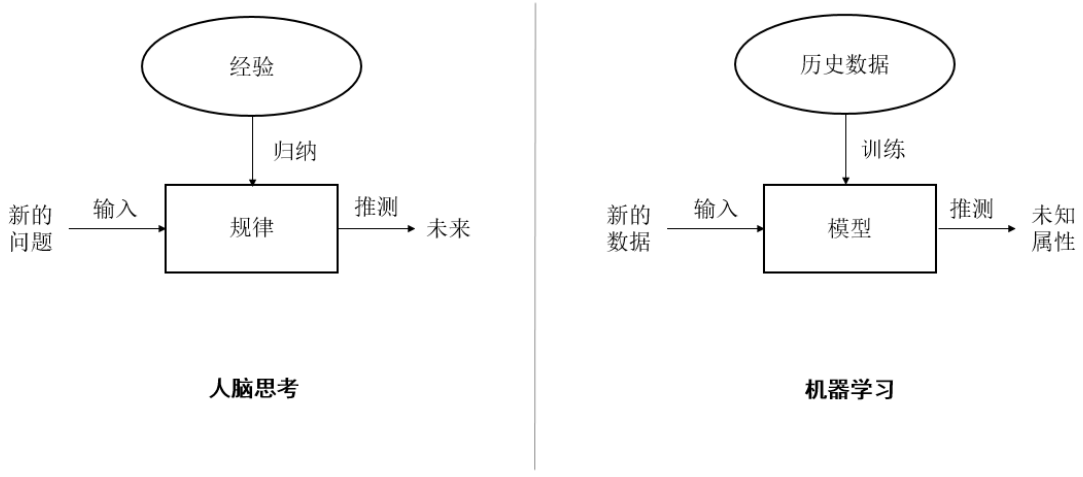


图 3-4　人脑思考与机器学习

给定一个输入，比如一段语音、一张图片或者一些数据型的信息，计算机能够建立一个函数（可以理解为一种对应关系），生成输出结果（见图 3-5）。机器学习的任务就是找到这个函数，找出从输入到输出的规则。

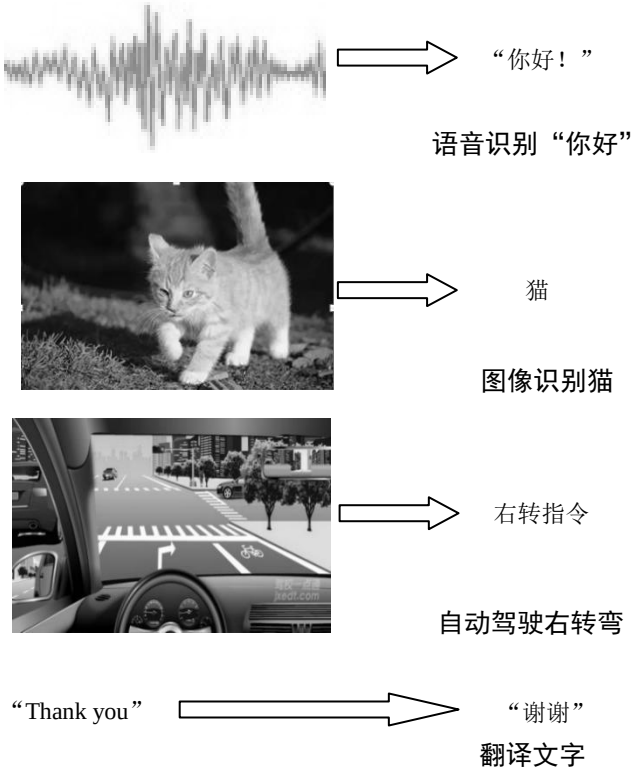


图 3-5　机器学习的例子

3.1.4 对机器学习的全面认识

机器学习是一门学科，它基于概率、统计、优化等数学理论的研究，理论严谨，也是已被广泛认可的成熟的知识体系。机器学习在高等教育院校中作为一门独立的课程存在，近年来受到包括数据科学、统计学、计算机、应用数学、运筹学、工程学等众多专业的学生的青睐。在学术研究方面，每年机器学习相关论文发表不计其数，是数理学科重要的学术研究方向之一。

机器学习也是一门技术，它被数据科学家（Data Scientist）广泛应用，是在数据分析中最常用的技术之一。而数据科学家是 21 世纪最火热的职业之一，其中机器学习是这个职位最重要的技能和工作内容之一。随着大数据时代信息量和数据量的爆炸式提升，人们对未知事物更加好奇，机器学习越来越能够“落地”而发挥使用价值。同时，随着 Python 等编程语言的普及以及 TensorFlow 等机器学习框架的完善，这个曾经似乎是高端学术的东西也越来越偏向应用，能够被更多人接受。

机器学习包含一系列算法，虽说它是这一系列算法的总称，但仅仅把机器学习视为算法或者模型是不准确的。机器学习是解决问题的一种方法，算法只是其中的一部分。这里所谓的算法或者模型，只是从输入到输出之间的一步，而机器学习是实现从输入到输出的全部过程，其中还包括对输入数据的清洗和转换、对特征的提取和整合、对数据的探索分析等。这些必要的步骤不做，拿到数据就盲目地直接套用模型，是不能解决问题的。

3.1.5 机器学习、深度学习与人工智能

机器学习、深度学习和人工智能都是现今人们热议的词汇，这三个概念通常被人们联系在一起讨论，但很多人理不清三者之间的关系。其实这三个概念虽然定义上的动机和角度有所差异，但事实上它们之间存在很清晰的包含关系。普遍认为，人工智能、机器学习、深度学习三者的关系如图 3-6 所示。

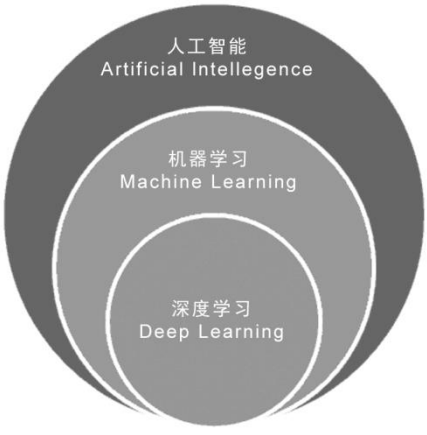


图 3-6　人工智能、机器学习和深度学习的包含关系

从概念上来说，机器学习是人工智能的一个分支。机器学习被看作解决人工智能问题的途径或者方案。而深度学习是机器学习的子类，或者可以理解为机器学习众多算法中的一种。只

不过因为它最近过于突出的表现，人们越来越习惯把这个概念单独提出来讨论。

从时间轴来看，三者也从前到后有着推进关系。人工智能这个词出现得最早。早在 1950 年，人工智能的概念就被提出，当时这个全新的领域令人兴奋不已。到了 1980 年左右，机器学习开始兴起，人们开始用机器学习的方法解决人工智能问题。深度学习在 2010 年开始流行，作为机器学习中最前沿的部分，推动人工智能发展取得了巨大突破。下面介绍深度学习，并重点说明深度学习和机器学习的关系。

最近几年，“深度学习”的概念越来越火热，并且这个词更多地与人工智能产生了直接联系。深度学习被视为进入人工智能领域的敲门砖，也是很多人工智能项目开展的基石。关于深度学习，我们会在后面的章节中进行详细介绍。这里我们先用几句话说明深度学习与机器学习之间的联系和区别。

前文提到，深度学习和机器学习是包含关系，深度学习是机器学习的一个子类。在传统的介绍机器学习算法的课程中，绝大多数会提到神经网络这个模型，而深度学习其实就是有多个隐藏层的神经网络。所以说，深度学习可以算作机器学习的一个分支领域，从算法来说是机器学习的一种，只不过这个分支由于具有其他算法不具备的显著优势，特别在 AI 领域的应用中，这些优势使得深度学习解决问题的效果尤为突出，所以在某种程度上，深度学习几乎成了人工智能模型算法的代名词。

在 AI 很多特定的领域中，我们看到的现象也确实如此，深度学习作为新兴的模型算法，表现出压倒性的统治力，而传统的机器学习算法应用只占极少数。那么，深度学习究竟为何能够在短时间内异军突起，击败它的“老前辈”们，成为新时代的宠儿？深度学习的优势究竟在哪里？在本书介绍完机器学习和深度学习后，相信读者会对它们的本质看得更透彻，也会对二者产生更直观的感性认识，对这个问题我们会在第 7 章给出解释。

在这个深度学习大有“一统江湖”趋势的时代里，我们还有必要学习传统的机器学习模型理论吗？答案是肯定的。首先，传统的机器学习模型并不落后，至今还有相当大的使用价值，在一些特定的问题中使用起来更为快速、灵活；其次，传统机器学习模型的思想对学习深度学习有重要的帮助。深度学习的许多理论和思想是基于传统机器学习模型的理念而来的。深度学习所用的神经网络可以由简单的机器学习模型（如感知机、逻辑回归）演变而来。

对于初学者来说，可以将深度学习理解为“多层神经网络”。严格来说，深度学习是一种学习的模式，是指采用具有“深度”的模型进行学习，其本身并不是一个模型。多层神经网络是具有“深度”特点的一个学习模型，它实际上是深度学习的一种形式。

3.1.6　机器学习、数据挖掘与数据分析

除了深度学习之外，另一个和机器学习类似而被人们广泛谈论的词语是数据挖掘（Data Mining）。另外，数据分析（Data Analysis）和大数据分析（Big Data Analysis）也似乎和机器学习有着密切的联系。下面我们把这几个概念放到一起谈谈。

无论是机器学习、数据挖掘还是数据分析，都没有一个学术上的权威定义。这几个词本身从定义上的出发点不同，并且提出概念的动机和背景上的差异比之前提到的人工智能、机器学

习和深度学习之间的差异更大，因此它们直接的界定更为模糊。因为它们在定义上有相通和重叠之处，我们也没有必要刻意去区分这几个概念之间的关系。这几个概念不像之前的人工智能、机器学习、深度学习那样容易通过维恩图来阐述。不过我们还是简单比较一下它们之间的异同。我们先来看一个数据分析的例子，如表 3-1 所示。

表 3-1　数据分析的例子

对照组（无药物）	药物 A	药物 B
14.4	14.3	18.1
13.9	15.9	17.9
12.4	16.1	17.2
15.8	12.8	16.6
15.0	13.0	19.8
14.6	14.6	18.5
13.2	17.4	17.6

对比上述三组数据，我们会发现 B 组的得分显著高于 A 组和对照组。通过建立统计学模型，我们会得出 B 组的得分显著高于 A 组的结论，于是证明药物 B 更加有效。根据已知信息进行分析总结，得出有意义的结论，就是一般数据分析要做的工作。

人们通常所说的数据分析是对小规模数据而言的，而当我们需要处理大规模数据时，往往会用大数据分析这个说法。大数据分析和数据分析相比，主要是数据量的区别以及因此而带来的运算模式和方法上的差异。大数据分析通常需要依赖多台计算机和分布式系统架构进行计算。除去这一点，大数据分析在目标上和数据分析是大致相同的。大数据分析与数据挖掘、机器学习的概念也更接近一些，因为数据挖掘和机器学习几乎都是建立在大数据基础上的。

相比数据分析，数据挖掘要做的事情更深入，跟机器学习的意义也更为贴近。数据挖掘不仅是对数据进行分析总结，还要“挖掘”表层所看不到的信息。从概念上来说，二者既有交叉，又有区别。数据挖掘的范围更大，是指从数据中获得有价值的信息。数据挖掘经常会通过机器学习来完成。事实上，两者的区别我们只需从字面上理解即可。数据挖掘侧重于“挖掘”二字，是从海量数据中发现和提取有用信息的过程。这里所谓有用的信息，可以是任何具有指导意义、在商业环境中能够帮助人们进行决策的信息。

数据挖掘的一个经典案例是“啤酒和尿布”的故事。在 20 世纪 90 年代，美国沃尔玛超市的管理人员从销售数据中发现了一个有趣的现象：在某些特定的情况下，“啤酒”与“尿布”两件看上去毫无关系的商品经常会出现在同一个购物篮中。经过后续分析发现，同时购买这两种商品的顾客通常是年轻的父亲。这样问题得到了解释：在美国有婴儿的家庭中，母亲一般在家中照看婴儿，而去超市买尿布的任务通常会落在父亲身上。父亲在购买尿布的同时，往往会顺便为自己购买啤酒。因此，啤酒和尿布竟然成为“会经常同时购买”的商品。这样一来，沃尔玛想出了一个点子，将啤酒和尿布尝试摆放在同一个货架区域，从而让年轻父亲能够在找到尿布的同时发现啤酒，于是大大提升了两种商品的购买率，最终为超市带来了更高的营业收入。通过数据挖掘，沃尔玛员工给公司挖掘出了商业价值。这就是数据挖掘的魅力所在。

反观机器学习，从机器学习的定义来说，它最终落在“预测”两个字上，由此可见，通常

机器学习是基于预测未知信息给人们带来决策上的收益的。但数据挖掘则不限于如此，发现数据的规律后不一定要跟着做预测，一条有价值的总结性信息可以直接帮助人们进行决断。

机器学习、数据挖掘、数据分析、大数据分析的相同点总结如下：

- 都是从数据中提取信息的过程。
- 都是数学和计算机结合的产物。
- 都可以帮助人们进行判断和决策。

3.2 机器学习的基本概念

在本节中，我们首先了解一个机器学习任务是如何进行的，分为哪些关键步骤；其次阐述机器学习中的几个重要的术语（样本、特征、目标）；最后了解如何根据目标形式对机器学习任务进行分类。

3.2.1 数据集、特征和标签

我们从一个实际问题出发。表 3-2 是纽约市某餐厅一个月内顾客消费和给予小费的数据，我们希望利用此数据研究顾客用餐给小费的规律。以这个数据集为例，我们先向读者介绍机器学习中的一些基本概念。

表 3-2　某餐厅小费支付表

ID	餐费	小费	性别	人数	星期	时间
1	17.8	2.34	男	4	周六	晚餐
2	21.7	4.3	男	2	周六	晚餐
3	10.1	1.83	女	1	周四	午餐
4	32.9	3.11	男	2	周日	晚餐
5	16.5	3.23	女	3	周四	午餐
6	13.4	1.58	男	2	周五	午餐

*数据节选自 Python Seaborn 数据包。原数据包中含 244 个样本，7 个变量。

我们通常把表 3-2 这样的样本数据叫作数据集（Dataset），该数据集以结构化的列表形式呈现。数据集由若干样本（Instance 或 Example）组成，每一个样本是一个观测数据的记录（Record），或者叫观测值（Observance），在表格中以行（Row）的形式体现。在机器学习中，一行、一条记录和一个样本的概念可以视为是等价的。在这个情景中，我们关注的是顾客给予小费的情况，小费这一列是我们关注的结果（Outcome），我们可以把这个变量称为因变量（Dependent Variable，也叫函数值），在机器学习领域中通常叫作目标（Target）或标签（Label），也有人把它称为响应值（Response）。以上几个概念可以视为一个意思，在本书中一般用目标

来指代这个变量，对应的数据称为标签数据。不同于“小费”，表中其他列表示的变量在这个问题中是用来解释和预测“小费”的，我们把这些变量叫作自变量（Independent Variable），在机器学习领域通常用特征（Feature）这个术语来表示。特征和标签在表中通常以列（Column）的形式呈现。整个关系如图 3-7 所示。

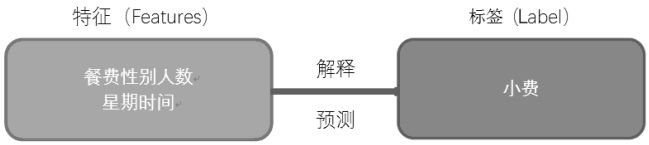


图 3-7　特征和标签

3.2.2　监督式学习和非监督式学习

并不是所有机器学习任务的数据集都带有标签数据，我们把具有标签数据的学习任务叫作监督式学习（Supervised Learning）。当目标变量是连续型（比如温度、价格）的时候，我们把这类问题叫作回归任务（Regression Task）；当目标变量是离散型（例如某种植物是否具有毒性、贷款人是否会违约、员工所属部门类别）的时候，我们遇到的问题则是分类任务（Classification Task）。回归问题和分类问题是监督式学习的两大类型。

监督式学习是目前最常见的机器学习。我们来看几个实际例子。比如：某银行具有贷款违约的客户详细信息，通过机器学习来预测哪些现有客户具有高的违约率。在这里的标签就是“是否违约”。再比如，某电商平台记录了所有访问产品页面的潜在客户信息，有些潜在客户购买了产品，成为了真正的客户，有些还没有。在这里的标签就是是否转化为真正的客户。通过机器学习来预测那些尚未转化的潜在客户中，哪些客户是最有可能被转化的，电商平台可针对这些客户进行针对性的促销和跟进，从而提高转化率。

有时我们遇到的样本数据并没有标签数据，我们把这个问题叫作非监督式学习（Unsupervised Learning）。非监督式学习虽然没有标签数据，但我们仍然可以挖掘特征数据的信息进行分析，聚类（Clustering）就是其中最常见的一种，它根据样本数据分布的特点将数据分成几个类，虽然我们不知道类别是什么。因此，我们可以把机器学习任务按图 3-8 进行分类。

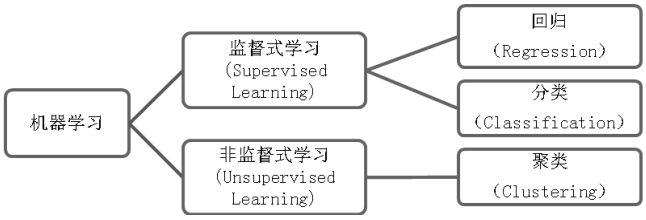


图 3-8　机器学习分类

3.2.3　强化学习和迁移学习

强化学习（Reinforcement Learning）是不同于监督式学习和非监督式学习的另一种机器学

习方法。在传统机器学习分类中不包括强化学习，而随着强化学习的飞速发展，越来越多的人倾向于把强化学习看作机器学习的第三类方法，如图 3-9 所示。

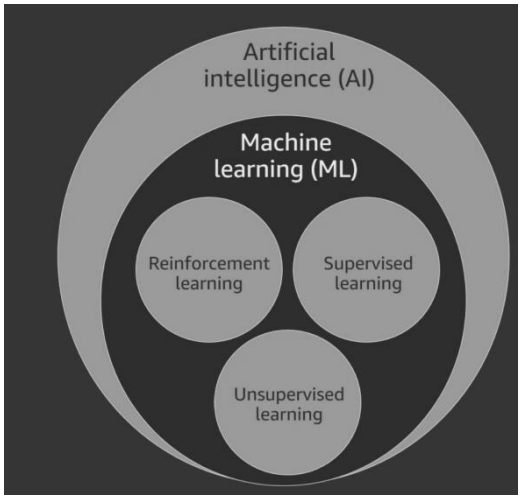


图 3-9　机器学习的 3 种方法

强化学习的一个经典例子是训练狗。让狗蹲下，如果狗听话蹲下，那么就给它一个奖励（小食品），通过这么不断的强化训练，狗最后就越来越听话了。所以，强化学习是基于“行动-反馈”的自我学习机制。所谓反馈，是一种基于行动对学习机的奖励。学习机以最大化奖励为目标，不断改进“行动”，从而适应环境。强化学习与监督式学习的主要区别是，前者是完全靠自己的经历去学习，没有人告知学习机正确的答案，“强化”的信号是对学习机行动的反馈；而后者则是有人在监督学习机。

强化学习就像人类刚出生时探索未知的大自然一样，是自我摸索寻找行为道路的过程。强化学习目前一个火热的应用是在游戏 AI 中。一个射击游戏的机器人要学会如何躲避敌人子弹，找到最合理的开枪和换子弹时机，这些用传统的机器学习来完成是相当困难的，因为游戏对局是动态的、瞬息万变的，有无数种可能。要用监督式学习“教会”电脑如何进行这些操作，需要训练的过程是漫长而烦琐的。强化学习很好地适应了这一问题。我们需要给电脑一个反馈机制，将“未能躲避子弹”作为惩罚，杀死敌人给予奖赏，剩下的就完全交给电脑去完成。这样电脑就能通过一遍又一遍的行为探索得到一套成熟有效的行动方案。

迁移学习指的是将已经训练好的参数提供给新的模型用作训练。现实中很多机器学习问题是存在相关性的。比如在图像识别中，识别狗和识别哈士奇，虽然具体任务不同，但它们具有相似性，用于识别狗的模型学习到的参数可以分享给识别哈士奇的任务，使得后者可以“从半路开始”，而不是从零开始学习参数，大大减少了学习时间。

迁移学习并不是一种新的机器学习分类，而是一种加快学习的模式。迁移学习在深度学习模型中的应用尤为明显。深度学习的模型庞大复杂，具有极多的参数需要训练。

3.2.4 特征数据类型

- 数值型（Numerical）：如年龄、访问某个网页的次数、花费金额、放在购物篮中的天数、价格等。
- 分类型（Categorical）：如性别、职业、年龄层（青年/中年/老年）、所在城市等。
- 文本（Text）：如姓名、地址等。
- 日期（Datetime）：如 2021-08-26。
- 数组型（Array）：如某客户的一系列感兴趣的关键词，曾经玩过的游戏等。

3.2.5 训练集、验证集和测试集

在机器学习任务中，我们通常将数据集分成三部分：训练集（Training Set）、验证集（Validation Set）和测试集（Test Set）。下面介绍这三个概念。

- 训练集：用于训练模型，确定模型中的参数。
- 验证集：用于模型的选择和优化。
- 测试集：用于对已经训练好的模型进行评估，评价其表现。

训练集和测试集的概念相对好理解。训练集顾名思义是用来训练的，机器使用训练集来学习样本。而测试集用来检验模型的效果。就像我们在学校学习功课，训练集如同教科书中的题库，测试集相当于考试试卷。我们通过“刷题库”获得知识，从而在考试中取得优异的成绩，如图 3-10 所示。

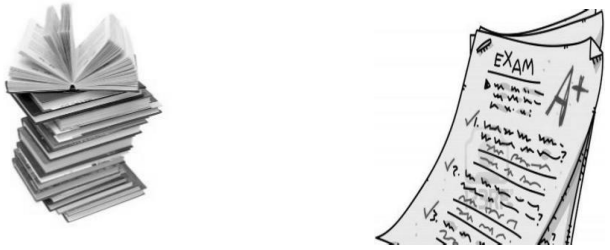


图 3-10　在考试中取得优异的成绩

为什么要建立测试集呢？不直接用训练集进行测试的原因是，模型是用训练集进行学习的，倾向于尽可能拟合训练集数据的特性，因此在训练集上的测试效果通常会很好，但在没有见过的数据集上表现效果可能会明显下降，这个现象叫作过度拟合（Overfitting，简称过拟合）。有关过度拟合的概念，后面会详细介绍。模型在没有见过的数据集上取得高准确率比在原训练集上获得好的效果更有说服力。因此，总是要设立测试集。就像只有考试才能最公平地衡量学生对功课的掌握程度一样。在常规的机器学习中，一般把 80%的数据放到训练集，而把 20%的数据放到测试集。

有了训练集和测试集，很多机器学习入门者可能不知道还有验证集这样一个概念。事实上，验证集是用来调参的。为了叙述的流畅性，这里读者可以先将调参理解为调整模型，相关概念

会在后续章节介绍并通过具体例子说明。验证集的作用是比较我们所尝试的多个模型，从中选择表现最好的一个。这个任务仅通过测试集其实也能实现，很多人会直接把测试集当作验证集来选择和优化模型，从而将测试集和验证集的概念混为一谈。但严格来说，验证集的单独存在是必要的。测试集用来衡量一个完整建好的模型，意味着这个模型在之前就被认定为已经调整到最优，而这个优化的过程就是通过验证集来实现的。如果我们延续上文中对训练集和测试集的比喻，验证集就相当于考前的模拟测试。

3.2.6　机器学习的任务流程

一个完整的任务流程大致可分为如图 3-11 所示的 6 个步骤。注意这个流程只是一般的思路，具体问题会有各自的差异和侧重。

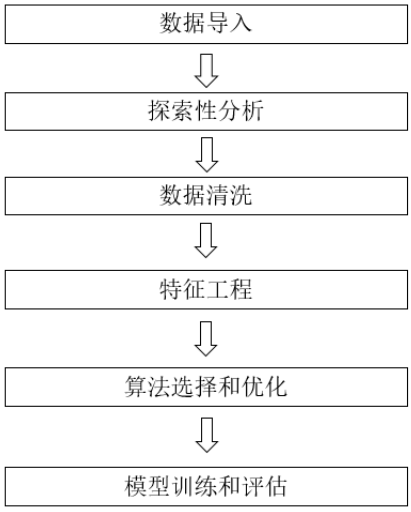


图 3-11　机器学习任务流程图

一般来说，在“数据导入”后，经过预处理，机器学习算法读入的是像表 3-2 一样的结构化数据（Structured Data）。在结构化数据中，特征都是以列的形式一条一条展开的。但是在图像识别、语音识别等任务中，原始数据以图片或音频的形式出现，所谓的特征我们是“看不见的”。这个时候，我们需要将这些原始信息转化为结构化的形式。

3.3　数据预处理

在拿到任务和数据后，应该先做数据预处理的基本流程，包括探索性分析、数据清洗和特征工程。本节阐述探索性分析的目的和途径、数据清洗的常见类型和手段、特征工程的重要性和常见方法。

3.3.1 探索性分析

探索性分析（Exploratory Analysis），也叫探索性数据分析（Exploratory Data Analysis）是通过图表等可视化工具对原始数据（Raw Data）进行大致了解和初步分析的过程。探索性分析的目的是让我们对陌生的数据集有个直观和感性的认识，从而在庞大的数据集中发现有价值的、值得挖掘的信息，找出数据集中的“亮点”。具体而言，通过探索性分析，我们可以：

- 了解数据集的基本信息。
- 给数据清洗提供方向。
- 为特征工程提供方向。

探索性分析是我们在拿到数据集还没有头绪的时候可以尝试的手段。探索性分析要避免时间过长，毕竟我们的目的是对数据进行初步探索，分析工作的大头在后面。

3.3.2 数据清洗

好的数据总是比好的算法要强得多（Better data beats fancier algorithms）。任何想从事机器学习的数据分析师，首先要记住这句话。如果数据质量差，杂乱无章，即使再好的算法也没有用，就好比加工垃圾一样，用再先进的技术加工出来的成品也是垃圾（Garbage in，garbage out）。之所以要进行数据清洗（Data Cleaning），是因为在现实生活中，我们遇到的绝大多数数据集都是“不干净的”。比如会出现以下情形（见表 3-3）：

- 存在重复记录的数据。比如人口数据中同一个人有两条完全相同的记录。
- 存在不相关记录。比如我们只关注中国人口数据，但数据集中有外国的人口信息。
- 无用的特征信息。例如身份 ID 等一些显然不会对结果有影响的编号类数据。
- 文字拼写错误。一些比较明显的信息输入错误。
- 信息格式不统一。例如大小写不一致，比如“beijing”和“Beijing”应该属同一类。表述形式不统一，比如“北京市”和“北京”也应该统一成一种。
- 明显错误的离群值（Outlier）。比如某个人的年龄数据显示为 175。
- 缺失数据。表格中有一些信息空缺，没有记录，如表 3-3 所示。

表 3-3 人口信息：杂乱的数据

姓名	ID	性别	年龄	学历	城市
张三	1400001	男	26	本科	北京市
李雷	1400002	男	21	研究生	上海
李雷	1400002	男	21	研究生	上海
王娜	1400003	女	NaN	高中	杭州
刘磊	1400026	男	175	NaN	深圳
林佳	1400027	Female	18	硕士	新加坡
孙瑶	1400031	女	24	本科	北京

*NaN 表示缺失数据。

设想一下，假如我们遇到表 3-3 这样的原始数据，想必一定会很头疼吧。如果不做数据清洗，后面的模型分析等操作根本就是寸步难行。然而数据乱不等于它没有分析价值，只要经过专业的数据清洗和特征工程处理，我们仍然能得到出色的分析效果。现实中数据来源纷杂多样，绝大多数做机器学习的人都需要花费大量时间进行数据预处理和清洗。

那么，如何清洗数据呢？对于离群值（见图 3-12），很多人会把离群值所在的记录去掉或者把它认定为缺失值，但其实很多时候这样做并不是最好的选择。离群值通常指样本中偏离均值较大的数据，在图像中通常处于“孤立”的位置。离群值所表示的数据很可能是有问题的。然而，离群值在被证明“有罪”之前都是清白的，仅当有确切、合适的理由的时候才可以去掉它，并且这样做能够提高模型的预测效果。因为“数值太大”，草率地将其去掉是不可取的，因为这个“大数值”本身可能包含了一定的信息。最后无论怎样，离群值所在行的其他特征数据依然是清白的，所以这一条记录不能因为一个特征出现离群值就去掉。

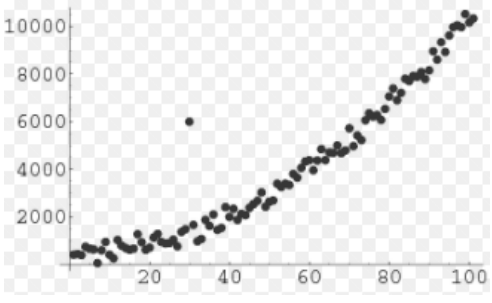


图 3-12　离群值实例

对于缺失值，有几种常见的处理方式。第一种处理方式是用均值或者中间值等进行填充。这样做的好处是比较快捷、方便，但可能不是最合理的方式。当缺失值比例较大时，这样做等于人为地向数据集中添加了噪声，因为这些数据并不是真实、准确的信息，从而可能会影响我们对结论的判断。类似的填充方式还有根据前后数值填充、插值填充、模型拟合填充等。第二种处理方式是去掉该特征。当一个特征大部分值都缺失时，如果我们认为这个特征对分析没有帮助，去掉该特征也并非不是一个可取的方法。第三种处理方式是保留缺失值的信息。对于分类型变量，我们可以将缺失值作为新的一类。

3.3.3　特征工程

特征工程（Feature Engineering）又称特征提取，是机器学习建模之前的一个重要步骤。前文提到，机器学习的本质是要找从特征（通常用 x 表示）到目标（通常用 y 表示）的映射，我们最终的目标是输出 y。如果 x“不给力”的话，后面的努力往往会成为徒劳。特征工程就是从原始数据中找到合适的特征集 x 的过程。比如，通过对客户流失数据集的特征提取，可以得到导致客户流失的一系列主要特征，利用这些特征可以有效地预测客户的流失（目标/标签）。

在很多机器学习任务中，特征工程是最重要，也是最耗时的环节，其重要性远远超过建模和训练。然而这个过程最容易被初学者忽视。特征工程是漫长而艰苦的过程。在 Kaggle 数据科学竞赛中，选手们平均花在构建特征集的时间在 70%以上。一个好的特征集通常能战胜一

个好的模型或者算法。特征工程往往还包含一些同特征选取和降维相关的预处理。我们将在第4章详细讲解。

3.4 算法

机器学习常用的算法有分类、聚类、回归、神经网络等。本节简要概述机器学习算法，让读者有一个整体的印象。第5章将详细讲解算法。

1. 分类算法

分类的含义就是找出数据集中的一组数据对象的共同特点，并按照分类模式将其划分为不同的类，比如：是否为垃圾邮件。分类的目的是通过分类模型将数据集中的数据项映射到某个给定的类别中。分类的应用包括客户的分类、客户的购买趋势预测等。分类在机器学习中属于有监督式学习的范畴。主要的分类算法包括决策树、KNN法（K-Nearest Neighbor）、SVM法、Bayes法、神经网络等。分类算法是有局限性的。分类作为一种监督式学习方法，要求必须事先明确知道各个类别的信息（如：是垃圾邮件、不是垃圾邮件两个类别），并且断言所有待分类项都有一个类别与之对应。但是很多时候上述条件得不到满足，尤其是在处理海量数据的时候，如果要通过预处理使得数据满足分类算法的要求，那么代价非常大，这时候可以考虑使用聚类算法。

2. 聚类算法

聚类的含义是指事先并不知道样本集的类别标签，按照数据对象的相似性和差异性，把一组对象划分成若干类，并且每个类里面对象之间的相似度较高，不同类里面对象之间的相似度较低或差异明显。我们并不关心某一类具体是什么，需要实现的目标只是把相似的东西聚到一起，聚类是一种非监督式学习方法。

聚类与分类的区别是，聚类类似于分类，但是与分类不同的是，聚类不依靠给定的类别（标签）对对象进行划分，而是根据数据的相似性和差异性将一组数据分为几个类别。聚类与分类的目的不同。聚类要按照对象的相似性和差异性将对象进行分类，属于同一类别的数据间的相似性很大，但不同类别之间数据的相似性很小，跨类的数据关联性很低。组内的相似性越大，组间差别越大，聚类就越好。

主要的聚类算法可以划分为5类，即划分方法、层次方法、基于密度的方法、基于网格的方法和基于模型的方法。每一类中都存在得到广泛应用的算法，划分方法中有K-Means、最近邻聚类算法，层次方法中有凝聚型层次聚类算法，基于模型的方法中有神经网络聚类算法。聚类可以应用到客户群体的分类、客户背景分析、客户购买趋势预测、市场的细分等。

3. 回归算法

回归算法有线性回归算法、逻辑回归算法等。它其实是一个统计预测模型，用以描述和评

估因变量（标签或目标）与一个或多个自变量（特征）之间的关系。回归算法被广泛地用于解释市场占有率、销售额、品牌偏好及市场营销效果。它可以应用到市场营销的各个方面，如客户寻求、保持和预防客户流失活动、产品生命周期分析、销售趋势预测及有针对性的促销活动等。

4. 神经网络

神经网络作为一种先进的人工智能技术，非常适合处理非线性的问题，以及那些以模糊、不完整、不严密的知识或数据为特征的问题。典型的神经网络有卷积神经网络（CNN）和循环神经网络（RNN）两种。

5. 预测分析

从已知的数据推测未知的数据或对象集中某些属性的值分布。建立预测模型的常用方法包括回归分析、线性模型、支持向量机、决策树预测、遗传算法、随机森林算法等。

3.5 初探机器学习的开源框架

大数据、算法和并行计算能力构成了人工智能高速发展的三要素，海量的数据积累是基础。开源的机器学习平台能够让开发者将复杂的数据传输给已有的框架进行分析和处理，缩短了开发时间，提升了训练效果，极大地推动了 AI 技术的商业化进程。在本节中，我们以 Scikit-learn 为例看看机器学习框架和库是什么样子的。

scikit-learn 项目最早由数据科学家 David Cournapeau 在 2007 年发起，是 Python 语言中专门针对机器学习应用而发展起来的一款开源框架。Python 是数据科学家最常用的编程语言之一，也是机器学习的首选工具。这是因为 Python 内置了很多实用的模块（Module），也称之为库（Library），它们就是预先写好的代码，可以通过导入（Import）直接用于应用程序中，省去重复编写代码的过程。多个相关 Python 模块组合在一起就组成一个包（Package），scikit-learn 就是其中之一。它可以很方便地直接拿来用于解决机器学习的实际问题。

在 Python 中，数据科学家常用的程序包如表 3-4 所示。

表 3-4　Python 中常用的程序包

包名	主要功能	例子
NumPy	最常见的包，科学计算的利器。用于向量和矩阵的存储和计算	np.dot(x,y)：进行 x 与 y 的矩阵乘法 np.max(x, axis=1)：返回矩阵每行的最大值
Pandas	基于 NumPy 构建的具有更高级数据结构的数据分析包，实现了结构化表数据的基本操作	pd.read_csv('file name')：读取 CSV 格式的数据源
Matplotlib	用于制作图表的基础包	plt.scatter(x,y)：制作变量 x 和 y 的散点图
Seaborn	专业用于统计制图，适合描绘数据集特征的分布和关系，基于 Matplotlib	sns.boxplot：显示一组数据分散情况的统计图

对于初次接触 Python 语言的读者，我们推荐下载 Anaconda。Anaconda 集成了 Python 基本环境和常见程序包，安装方便，非常适合初学者使用。读者只需从 Anaconda 官网上免费下载其最新版本，就可以编写 Python 程序。本书中所有实例代码都可以在 Anaconda Python 中实现。在 Anaconda 上，我们可以选择 Spyder 进行代码的编写。Spyder 是 Python 的一款 IDE。如同其他编程语言一样，Python 在编写代码时需要一个集成开发环境，IDE 就是集成开发环境的缩写，它给用户提供了编程需要的图形化界面、编辑器、编译器、调试器等。

3.5.1 scikit-learn 简介

编写机器学习算法不容易，但使用机器学习算法非常简单，读者只需套用现有的框架即可。现有的框架可以是某种编程环境中的程序包，或者一款成熟的机器学习商业软件。scikit-learn 是一个针对机器学习的强大 Python 程序包，主要用于构建模型，使用诸如 NumPy、SciPy 和 Matplotlib 等程序包构建，对于统计建模技术（如分类、回归、集群等）非常有效。scikit-learn 的特性包括监督式学习算法、非监督式学习算法和交叉验证，官网地址是 http://scikit-learn.org/。scikit-learn 的基本功能主要被分为六大部分：分类、回归、聚类、数据降维、模型选择和数据预处理。

- 数据预处理（Preprocessing）：功能包括转换输入数据，规范化、编码化。模块有 preprocessing、feature_extraction 和 transformer（转换器）。
- 降维（Dimensionality Reduction）：功能包括 Visualization（可视化）和提高机器学习效率。算法有主成分分析（PCA）、非负矩阵分解（NMF）和 feature_selection（特征选择）等。
- 分类（Classification）：用于解决二元分类问题、多分类问题、图像识别等。算法有逻辑回归、SVM、最近邻、随机森林、朴素贝叶斯、神经网络等。
- 回归（Regression）：可应用于药物反应、股票价格预测等应用场景。算法有线性回归、SVR、LARS 等。
- 聚类（Clustering）：可应用于客户细分、分组实验结果等应用场景。算法有 k-Means、spectral clustering（谱聚类）、mean-shift（均值漂移）等。
- 模型选择（Model Selection）：它的功能是通过参数调整提高精度。模块有 pipeline（流水线）、grid_search（网格搜索）、cross_validation（交叉验证）、metrics（度量）和 learning_curve（学习曲线）等。

另外，scikit-learn 还包括模型融合的模块，如：ensemble（集成学习）和一些辅助工具模块，如：exceptions（异常和警告）、dataset（自带数据集）等。

3.5.2 第一个机器学习实例

正如上节中提到的，Python 中有完备的机器学习程序包 sklearn，它整合了众多的机器学习模型，这些模型的算法已经在程序包中编写好，用户无须知道算法的原理，也无须懂得模型

的含义，甚至无须会编程，即可调用程序包，执行并得到我们需要的结果。整个过程不过只是几行代码，就像把大象放到冰箱一样，只需三步。下面我们以 sklearn 为例，来看看它如何解决机器学习经典案例——泰坦尼克（Titanic）沉船生存预测。

在 Kaggle 网站上有一个关于泰坦尼克号幸存预测的案例，很适合用随机森林算法来预测，所以我们就以这个数据集（https://www.kaggle.com/c/titanic）开始讲解。这个数据集包含 3 个文件，我们现在只关心两个文件，一个是训练数据（train.csv），一个是测试数据（test.csv）。另一个文件是提交给 Kaggle 竞赛的样本文件（输出结果文件）。我们利用泰坦尼克号训练数据集，运用随机森林算法根据乘客的不同变量参数特征进行学习，最后用测试数据集得出其他乘客的预测值（是否幸存）。

对于不熟悉泰坦尼克数据集背景的读者，我们简单解释一下。在 1912 年 3 月 15 号，泰坦尼克号在首航之时就因撞击冰山而沉没，首航之时的 2224 名旅客及船员中共有 1502 名死亡。造成伤亡如此惨重的其中一个原因在于船上没有足够的救生船让旅客及船员及时逃离。我们知道在沉船中存活下来需要一些幸运因素，但是对这些人的背景（如性别、舱位、票价、年龄等）进行调查后发现，幸存与否与这些因素也有一定的关系。从 Kaggle 网站上下载训练和测试数据集后，我们打开训练数据集，列出了部分样本数据如表 3-5 所示。

表 3-5　训练数据样本

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Bra…	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cum…	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Hei…	female	26	0	0	STON/O2. 3101282	7.925		S
4	1	1	Fut…	female	35	1	0	113803	53.1	C123	S
5	0	3	All …	male	35	0	0	373450	8.05		S
6	0	3	Mor…	male		0	0	330877	8.4583		Q

从训练集（表 3-5）中可以看到共有 11 个参数，一个标签（Survived），具体如下：

- PassengerId：乘客 ID 号，这个是自动生成的。因为这是一个顺序标号，对我们预测没有影响，所以不考虑选为特征。
- Survived：是否生还（1-是，0-否）。
- Pclass：乘客的舱位（1-一等舱；2-二等舱；3-三等舱）。
- Name：乘客姓名。
- Sex：乘客性别。
- Age：乘客年龄。
- SibSp：兄弟姐妹，伴侣人数。
- Parch：父母人数。
- Ticket：票号。
- Fare：船票价格。
- Cabin：船舱号。

- Embarked：上船地点，分别有 Q、S 等地点。
- NoAge：未提供年龄。

我们得出特征类型如下：

- 数值型（Numerical）特征：Age（连续型）、Fare（连续型）、SibSp（离散型）、Parch（离散型）。
- 分类型（Categorical）特征：Survived、Sex、Embarked、Pclass。
- 文本型特征：Ticket、Cabin。

我们假定特征集 X 是训练数据集中除 Survived 列之外的所有列，而标签数据 Y 是 Survived 列。X'是测试数据集。

```
from sklearn.ensemble import
RandomForestClassifier
```

第一步：读取程序包
这里我们读取的是随机森林（random forest）模型

好比从图书馆中获得需要的工具材料。

```
model = RandomForestClassifier(n_estimators
=12)
```

第二步：声明模型
告诉计算机我们要用的模型是什么，要用哪种方法解决问题

建立了一个模型，叫作 model，这个 model 要用随机森林模型。

```
model = model.fit(X, Y)
```

第三步：训练模型
给声明的模型“喂”训练数据，让模型自主学习数据。这里 X、Y 分别为训练数据的特征和标签

经过这一步，model 学习了数据并获得了预测能力。

```
Y_hat = model.predict(X')
```

第四步：模型预测
让训练好的模型去“完成任务”，即预测新数据（测试数据）的标签

让 model 对未知标签数据进行预测。

对于不熟悉模型的人，只需记住这几行代码，就可以完整地跑出这个模型。给定一个新数据 X'，预测结果就在 Y'中。代码中并没有多少需要理解的部分或者需要预先掌握的知识，基本都是 sklearn 的预设格式。实际上，sklearn 可以用于绝大多数传统机器学习模型，并且只需在这几行代码上稍作改动。在后面的章节，我们会看到使用 sklearn 的更多实例。

3.5.3 Jupyter Notebook

对于机器学习，首先我们需要有数据，有了数据之后，在训练模型之前我们还需要对数据反复进行清洗和准备，这包括对数据集中的数据打上准确的标签。之后就是根据业务特性来尝试不同的机器学习算法。随着云计算的普及，机器学习也经常在云端完成，我们在云端创建

cluster，导入数据并进行训练。训练结束后关闭 cluster。一个流行的开源工具就是 Jupyter Notebook。

Jupyter Notebook 的本质是一个 Web 应用程序，便于创建和共享程序文档，支持实时代码、数学方程、可视化和 Markdown。它可广泛用在数据清理和转换、数值模拟、统计建模、机器学习等领域。有人认为 Jupyter Notebook 是用 Python 做机器学习最好用的 IDE。

那么，什么是 Jupyter Notebook（以下简称 Jupyter）？简单来说，它是一种模块化的 Python 编辑器（现在也支持 R 等多种语言）。在 Jupyter 中，可以把 Python 代码分开每一段来运行。在软件开发中，Jupyter 可能显得并没有那么好用，这个模块化的功能反而会破坏掉程序的整体性；但是当我们在做数据处理、分析、建模、观察结果的时候，Jupyter 模块化的功能不仅会为我们提供更好的体验，更能大大缩小运行代码及调试代码的时间，同时还会让我们整个处理和建模的过程变得更加清晰。也就是说，Jupyter Notebook 将 Python 的交互式特点发挥到了极致。

熟悉 Python 的读者一定对 Python 的交互式功能感触颇深。Python 的易读性和交互性非常好。Java 代码每次都要经过编译，而且每一行的代码没有办法单独运行。与之不同的是，Python 的每一行都像是命令行的命令，简单易懂且有交互性，我们输入一句，它便返回一句的结果。但在一般的 IDE 中，Python 的这一交互功能被极大地限制，通常我们会将程序整段编写之后一起运行。而在 Jupyter Notebook 当中，我们可以每写几行或者每完成一个小的模块便运行一次。也许对于软件工程师们来说，这个功能并没有多大的吸引力。但是对身为机器学习工程师的我们来说，这个功能非常好，我们使用它来对数据进行探索，尝试不同的模型和算法。机器学习的分析和建模是非常碎片化的工作，而每一块的碎片又有着一定的独立性。数据分析和处理的过程往往是一个不断试验的过程，我们需要一次又一次的改变预处理的方式、尝试不同的特征工程处理、一遍又一遍地调整模型参数，等等。每一部分的工作都需要反复试验、反复修改，而下一模块需要用到的只不过是上一模块输出的数据。通过 Jupyter，我们可以以最快的速度知道自己做出的调整是好还是坏，并尽快进入到下一次的试验当中。由于 Jupyter Notebook 可以将输出结果嵌套在 Notebook 中，并且支持 Markdown 语句的操作，这样使得我们可以与同事共享不同阶段的代码和结果。

下面我们来体验一下 Jupyter Notebook。在 www.kaggle.com 上就安装着 Jupyter Notebook。在浏览器上输入 https://www.kaggle.com/c/titanic/notebooks，就进入了为泰坦尼克（Titanic）沉船生存预测案例准备好了的 Notebook 环境，如图 3-13 所示。

在 Public 选项卡下选中任何一个项目，用鼠标单击之即可进入，并在新页面上单击“Copy and Edit”复制按钮，这就会将选中的项目复制到我们自己的环境下，如图 3-13 所示。我们会看到，有多个灰色的代码块，这个叫作 code cell。把鼠标移动到第一个代码块那里，左侧边框线变为蓝色粗线条，这就是命令模式（见图 3-14）。单击蓝色线条左边的[ ]，这个方框变为蓝色箭头，再单击一下，就可以运行选中的代码块。我们再运行下一个代码块（Load Data 部分），运行结果显示在代码后面，如图 3-15 所示。图 3-14 上的“+Markdown”按钮，这是我们输入文本的地方。我们可以在运行代码后添加结论和注释等。单击“+Code”按钮，就可以输入新的代码块。

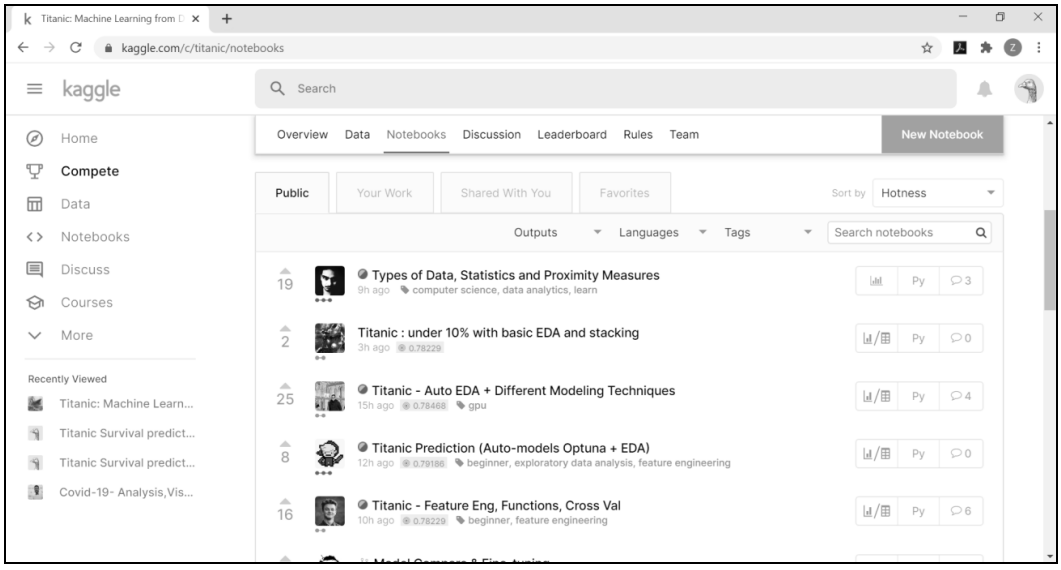


图 3-13　kaggle 上的 Notebook 环境

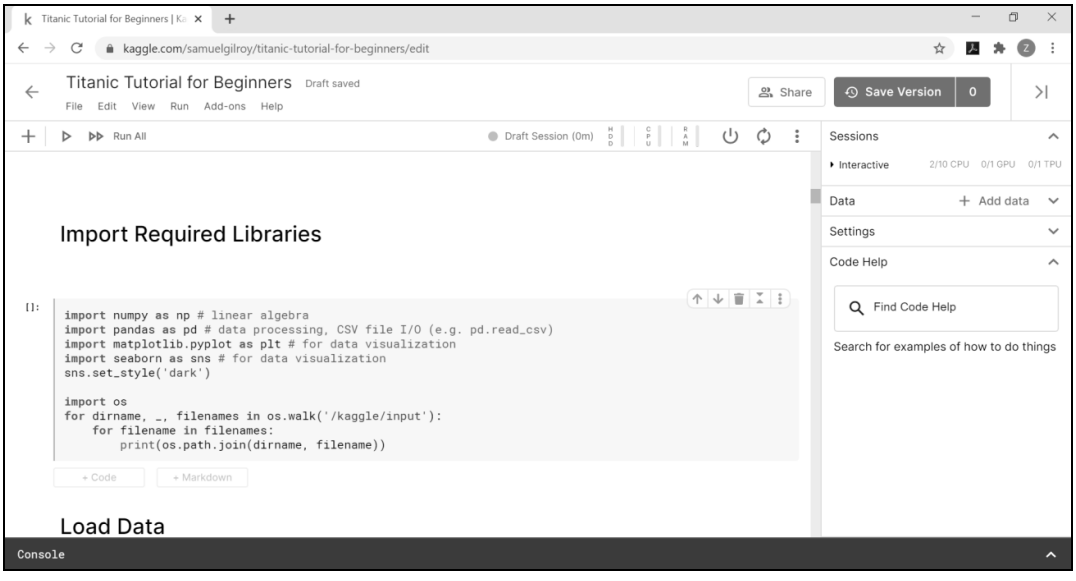


图 3-14　在 Notebook 上编辑和运行代码

在 Notebook 上，运行结果直接显示在代码下面，非常方便查看，如图 3-15 所示。

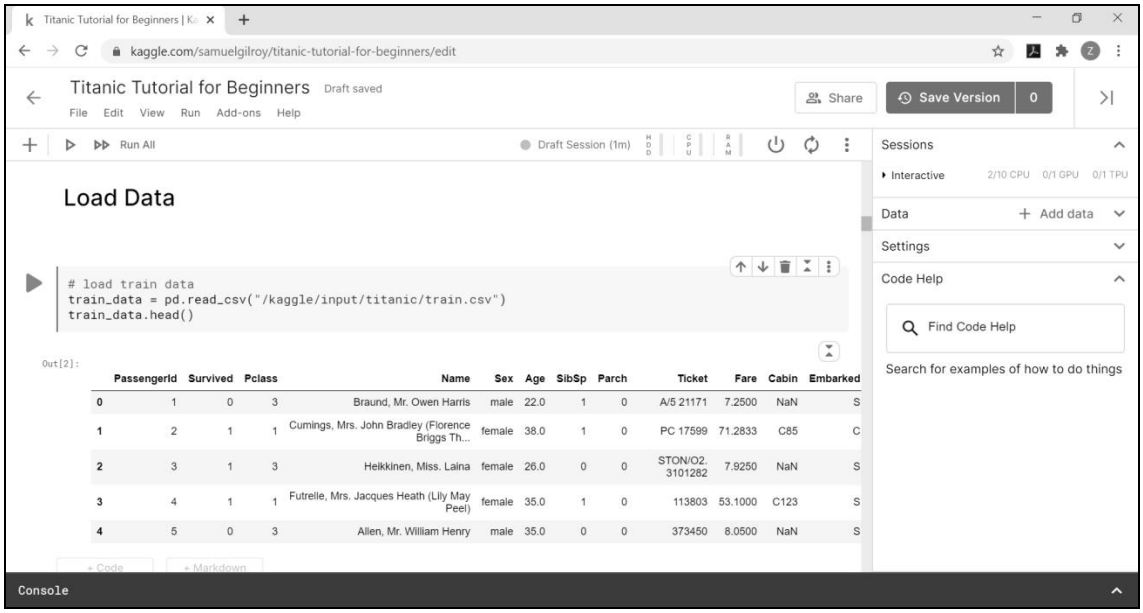


图 3-15　在 Notebook 上运行代码并查看结果

通过上面这个例子，我们看到，Jupyter Notebook 是以网页的形式打开，可以在网页页面中直接编写代码和运行代码，代码的运行结果也会直接在代码块下显示。如在编程过程中需要编写说明文档，可在同一个页面中直接编写，便于完成及时的说明和解释，非常方便。Notebooks 的确是数据科学家手中非常灵活、可交互和强大的工具。

另外，安装 Anaconda 发行版时会自动一同安装好 Jupyter Notebook，所以，我们可以在自己的机器上体验 Jupyter Notebook。在终端中输入命令 Jupyter notebook。这时会在终端中显示一系列 Notebook 的服务器信息，同时浏览器将会自动启动 Jupyter Notebook。

3.5.4　更多实例分析

下面我们以 3.3.1 小节中所讲的探索性分析和数据清洗为例，针对泰坦尼克号幸存预测的案例，来看看怎么使用 sklearn 进行数据的探索性分析和清洗。以下代码都是在 Jupyter Notebook 上运行的。我们先加载数据，然后通过 describe()看一下数据的整体分布：

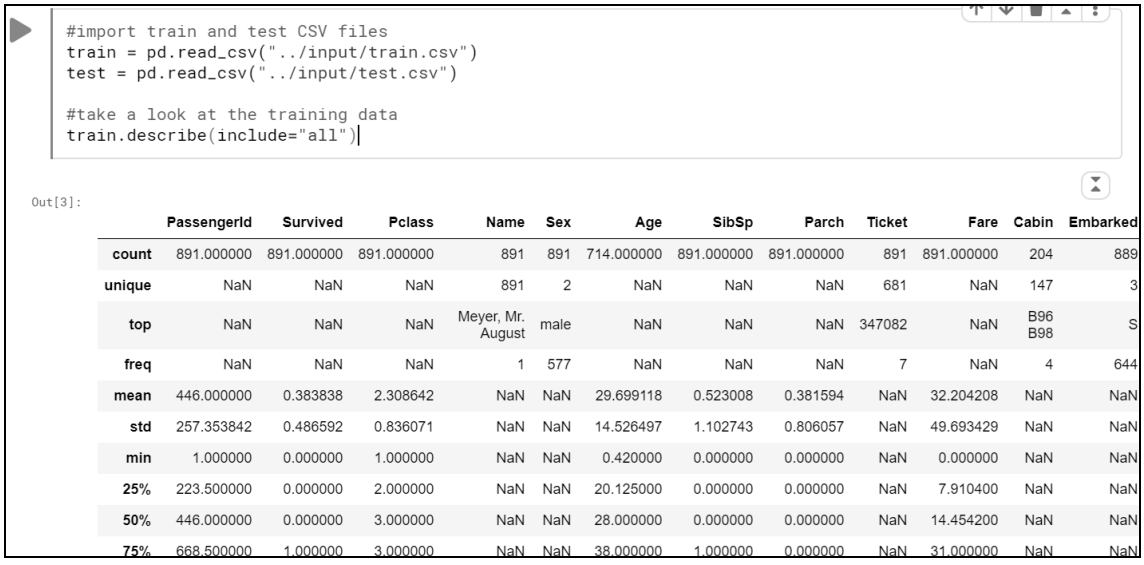
```
#data analysis libraries
import numpy as np
import pandas as pd

#visualization libraries
import matplotlib.pyplot as plt
import seaborn as sns

#import train and test CSV files
train = pd.read_csv("../input/train.csv")
test = pd.read_csv("../input/test.csv")
```

```
#take a look at the training data
train.describe(include="all")
```

结果如图 3-16 所示。



```
#import train and test CSV files
train = pd.read_csv("../input/train.csv")
test = pd.read_csv("../input/test.csv")

#take a look at the training data
train.describe(include="all")
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
count	891.000000	891.000000	891.000000	891	891	714.000000	891.000000	891.000000	891	891.000000	204	889
unique	NaN	NaN	NaN	891	2	NaN	NaN	NaN	681	NaN	147	3
top	NaN	NaN	NaN	Meyer, Mr. August	male	NaN	NaN	NaN	347082	NaN	B96 B98	S
freq	NaN	NaN	NaN	1	577	NaN	NaN	NaN	7	NaN	4	644
mean	446.000000	0.383838	2.308642	NaN	NaN	29.699118	0.523008	0.381594	NaN	32.204208	NaN	NaN
std	257.353842	0.486592	0.836071	NaN	NaN	14.526497	1.102743	0.806057	NaN	49.693429	NaN	NaN
min	1.000000	0.000000	1.000000	NaN	NaN	0.420000	0.000000	0.000000	NaN	0.000000	NaN	NaN
25%	223.500000	0.000000	2.000000	NaN	NaN	20.125000	0.000000	0.000000	NaN	7.910400	NaN	NaN
50%	446.000000	0.000000	3.000000	NaN	NaN	28.000000	0.000000	0.000000	NaN	14.454200	NaN	NaN
75%	668.500000	1.000000	3.000000	NaN	NaN	38.000000	1.000000	0.000000	NaN	31.000000	NaN	NaN

图 3-16　探索训练数据

上述代码会显示数据集的行数和各列的数据缺失情况。相关的函数有：

- head()：获取前五行数据，供快速参考。
- info()：获取总行数、每个属性的类型、非空值的数量。
- value_counts()：获取每个值出现的次数。
- hist()：直方图的形式展示数值型数据。
- describe()：简要显示数据的数字特征。例如：总数、平均值、标准差、最大值/最小值、25%/50%/75%值。

从上面的输出结果看出：

- 总共有 891 名乘客。
- 年龄数据只有 714 个，缺失了部分数据。
- Cabin 数据只有 204 个，缺失的太多了。
- Embarked 也缺失了一点数据。

执行以下代码，确认那些缺失数据的特征列的详细信息：

```
#check for any other unusable values
print(pd.isnull(train).sum())
```

结果如下：

```
PassengerId      0
Survived         0
```

```
Pclass          0
Name            0
Sex             0
Age           177
SibSp           0
Parch           0
Ticket          0
Fare            0
Cabin         687
Embarked        2
dtype: int64
```

下面看一下男女不同性别在生存率上的区别：

```
#draw a bar plot of survival by sex
sns.barplot(x="Sex", y="Survived", data=train)
```

结果如图 3-17 所示。女性的生存率大大高于男性。

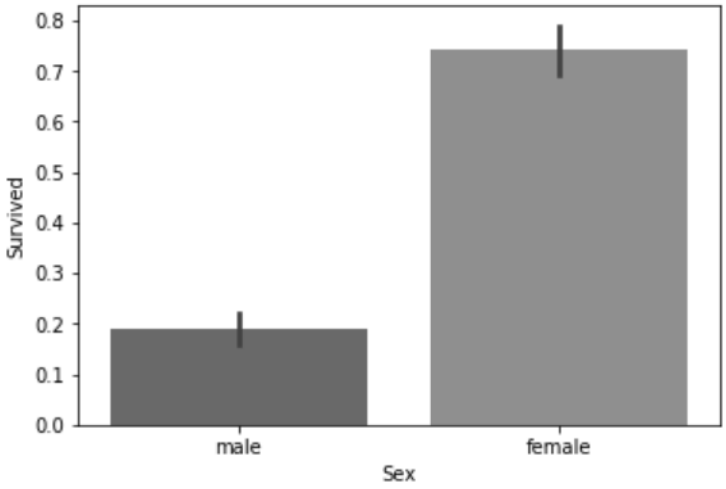


图 3-17　男女性别在生存率上的区别

下面看一下舱位等级在生存率上的区别：

```
#draw a bar plot of survival by Pclass
sns.barplot(x="Pclass", y="Survived", data=train)
```

结果如图 3-18 所示。舱位等级越高，生存率越高。

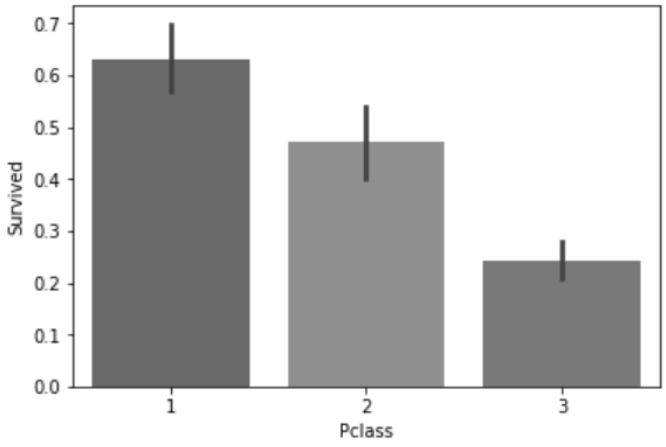


图 3-18　舱位等级在生存率上的区别

我们把年龄分成几个组，然后看看每个组的生存率。为了分组的需要，我们暂时把缺失年龄的数据填上–0.5。

```
#sort the ages into logical categories
train["Age"] = train["Age"].fillna(-0.5)
test["Age"] = test["Age"].fillna(-0.5)
bins = [-1, 0, 5, 12, 18, 24, 35, 60, np.inf]
labels = ['Unknown', 'Baby', 'Child', 'Teenager', 'Student', 'Young Adult', 'Adult',
'Senior']
train['AgeGroup'] = pd.cut(train["Age"], bins, labels = labels)
test['AgeGroup'] = pd.cut(test["Age"], bins, labels = labels)

#draw a bar plot of Age vs. survival
sns.barplot(x="AgeGroup", y="Survived", data=train)
plt.show()
```

结果如图 3-19 所示。婴儿组的生存率最高。

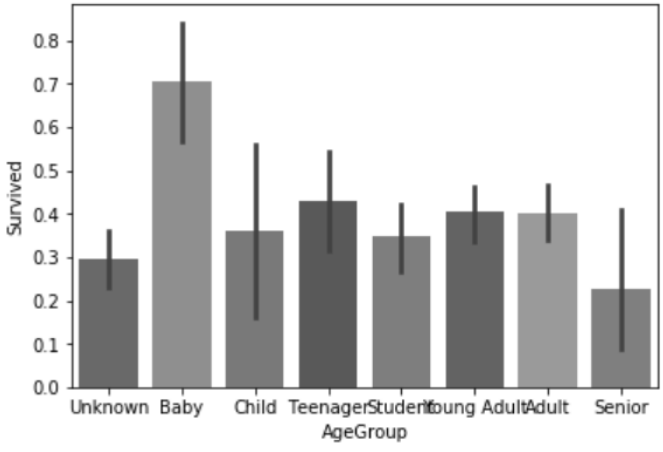


图 3-19　年龄组的生存率分布

如表 3-5 所示，Cabin 特征列好多行都没有值（NaN）。因为太多行缺失这个特征列数据，所以，我们做一个数据清洗，把这个特征列去掉：

```
#drop the Cabin feature since not a lot more useful information can be extracted
from it.
train = train.drop(['Cabin'], axis = 1)
test = test.drop(['Cabin'], axis = 1)
```

然后，我们验证一下这个列是否已经去除，显示一下前几行数据：

```
train.head()
```

结果如图 3-20 所示。

```
train.head()
```

Out[8]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

图 3-20　显示前几行数据

除了 Cabin 特征列，Name 和 Ticket 的信息看上去同生存率的关联不大，这些列也可以去除。在我们完成数据的探索性分析和数据清洗之后，就进入下一步：特征工程。我们在下一章详细阐述特征工程。