第 1 部分

深度学习基础篇

1 概述

当今时代，人工智能（Artificial Intelligence，AI）、机器学习（Machine Learning，ML）、深度学习（Deep Learning，DL）都是耳熟能详的一些概念。机器学习是实现人工智能的一种方式，而深度学习是机器学习的一个分支。NVIDIA 的一张图（如图 1-1 所示）很好地概括了三者之间的关系[1]。人工智能从 20 世纪 50 年代开始兴起，机器学习在 80 年代兴起，而深度学习的流行则晚一些，在 2010 年左右。

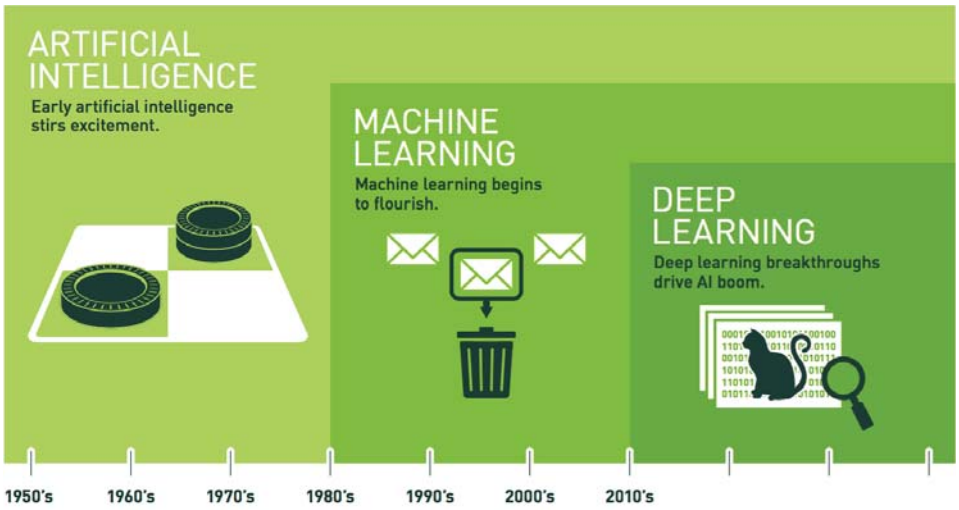

图 1-1　人工智能、机器学习、深度学习三者之间的关系

1.1 人工智能

2016 年 3 月 9 日至 15 日，Google 旗下 DeepMind 公司开发的围棋程序 AlphaGo 与世界围棋冠军、职业九段选手李世石进行人机大战，最终以 4:1 的总比分赢得比赛。

2016 年年末至 2017 年年初，AlphaGo 在中国围棋网站上以 Master 账号与中、日、韩数十位围棋高手过招快棋，连胜 60 局。

2017 年 5 月 23 日至 27 日，AlphaGo 迎战世界冠军柯洁，以 3:0 的比分毫无悬念地赢下比赛。

2017 年 12 月 5 日，DeepMind 宣布，新 AI AlphaZero 只须学习 34 小时即可战胜 AlphaGo。

随着 AlphaGo 的高歌猛进以及媒体的积极炒作，人工智能的概念得到迅速普及，甚至有人开始担心未来人工智能是否会危害到人类。

1.1.1 人工智能的分类

人工智能也称为机器智能，是指人工制造出来的机器或系统展现出来的智能。

人工智能也可以进一步分为以下两种类型。

- 弱人工智能（Weak AI）：通常是指机器通过机器学习之类的技术从大量数据中学到一些规律，这种学习实际是记忆性的，机器本身并无意识，只是执行某些算法或任务的工具。弱人工智能有时也称为狭义人工智能（Narrow AI）。
- 强人工智能（Strong AI）：机器具有意识，能够完全像人类一样思考和具有感情。强人工智能也称为通用人工智能（General AI）或全人工智能（Full AI）。

AlphaGo 虽然看上去比人类还厉害，但依然只是弱人工智能，本身并无意识可言。而部分人担心的可能会危害到人类的人工智能，则可以定义为第三类人工智能——超级智能（Super-Intelligence）——机器具有比人类更强大的智慧，甚至是人类无法理解的智慧。

1.1.2 人工智能发展史

提及人工智能，就不能不介绍计算机科学之父艾仑·麦席森·图灵（Alan Mathison Turing，1912—1954 年），他在 1950 年创作的 *Computer Machinery and Intelligence* 中提出了智能的概念，以及著名的图灵测试——计算机能否在智力行为上表现得和人没有区别。后来英国皇家学会规定的图灵测试标准为：如果机器可以在 5 分钟内回答由人类测试者提出的一系列问

题，且其超过 30% 的回答可以让测试者认为是人类所回答的，则该机器通过图灵测试并认为其具有智能。时至今日，图灵测试一直被广泛用于测试机器是否具有智能。

人工智能比计算机出现得还更早一些，最早出现在 1955 年的一次 10 人研讨会提案中，一般认为，1956 年的达特茅斯会议（Dartmouth Artificial Intelligence Conference）才是人工智能真正诞生的地方，当时参加会议的很多人都是大名鼎鼎的科学家，包括 John McCarthy（人工智能之父，1971 年图灵奖得主，Lisp 语言之父）、Marvin Minsky（人工智能之父，1969 年图灵奖得主）、Nathaniel Rochester（IBM 第一代通用计算机 701 主设计师）、Claude Shannon（信息论之父）。他们提出人工智能的研究目标是设计可以模拟人类的机器，这种机器可以使用语言，具有抽象理解能力。

至今，人工智能这个概念的提出已经半个多世纪了，为何很多人在最近一两年才真正切身感受到的存在？这一切与人工智能的发展历程密切相关。大体上，人工智能的发展可以分为以下几个阶段。

1. 第一个黄金时期（1956 年至 20 世纪 70 年代中）

1956 年达特茅斯会议后的 10 多年是人工智能发展的第一个黄金时期。在这个时期，大家认为逻辑推理能力是计算机具有智能的最重要原因，这个时期也被称为人工智能“推理期”。计算机被广泛用于解决代数题、证明几何定理等，这些成果得到了广泛赞赏，也让当时的研究者信心倍增，甚至很多人认为推理就是智能，有了推理能力就可以制造出完全智能的机器。

在这个时期，人工智能的相关研究也得到了政府的大力支持，获得了大笔的科研资金。1963 年 6 月，新建立的 ARPA（即后来的 DARPA，Defense Advanced Research Projects Agency，美国国防部高级研究计划局）赞助了 MIT 222 万美元经费，用于资助 MAC 工程[2]，其中包括 Marvin Minsky 和 John McCarthy 于 1959 年建立的人工智能实验室。此后 ARPA 每年提供 300 万美元人工智能研究经费，直到 20 世纪 70 年代。1963 年，John McCarthy 在斯坦福大学成立了另一个前瞻性的人工智能实验室。

2. 第一个低谷时期（20 世纪 70 年代中至 80 年代初）

虽然大家信心倍增，也投入了大量的科研资金，但是渐渐地发现当时的计算机运算能力有限，有限的内存和运算速度使得计算机很难处理实际应用中的人工智能问题，之前的盲目承诺无法兑现，光靠推理并没有实现真正的人工智能。人们渐渐失去耐心，批评和怀疑也接踵而至，1973 年有个著名的 Lighthill 报告（具体是指 James Lighthill 所写的论文 *Artificial Intelligence: A General Survey*），深度抨击了人工智能的进程，这导致之后英国的科研经费

大量转向其他方向。YouTube 上还有 1973 年 Lighthill 与 Richard Gregory、John McCarthy、Donald Michie 等人工智能支持者的辩论视频，感兴趣的读者可以搜一下 Lighthill debate。在一定程度上，Lighthill 报告加速了将人工智能打入冷宫的进程。从 1974 年开始，已经很难找到支持人工智能的科研经费。

在这个时期，感知机之类的联结主义遭遇冷落，1969 年 Minsky 和 Papert 出版了著作 *Perceptrons: an introduction to computational geometry*，书中暗示 1958 年由 Frank Rosenblatt（1928—1971 年）提出的感知机具有严重局限，从数学角度证明了单层感知机计算能力有限的根本原因，指出单层感知机甚至连异或（XOR）这样的问题也不能解决，并论证了单层感知机的这些局限性在多层感知机中是不可能被全部克服的。此处详情可参考本书 2.3.2 节。

同样在这个时期，专家系统之父 Edward Albert Feigenbaum（1994 年图灵奖得主）等研究者开始倡导智能机器必须具备知识，据此可以认为，20 世纪 70 年中后期人工智能进入了“知识期”。在这个时期，专家系统代替逻辑推理成为新宠，人们总结出来的大量知识通过规则之类的系统输入计算机，规则的详细程度决定了机器的智能水平，现在人们常说的“有多少人工就有多少智能”是非常适合这个时期的。

3. 第二个黄金时期（20 世纪 80 年代初至 80 年代末）

20 世纪 80 年代，人工智能开始复苏。1981 年，日本经济产业省为支持第五代计算机项目拨款 8.5 亿美元，这种计算机可以与人对话、翻译、理解图像，并能像人一样推理。此后，英国、美国纷纷效仿，也启动了很多相关项目，人工智能迎来了新的发展时期。

20 世纪 80 年代初，另一个令人振奋的事件是 John Hopfield 和 David Rumelhart 使联结主义重获新生。1982 年美国加州理工学院物理学家 John Hopfield 博士提出了 Hopfield 网络，这是一种递归神经网络，从输出到输入增加了反馈连接；1986 年 David Rumelhart 等人发表了 *Parallel Distributed Processing*，文中详细讲解了具有非线性连续变换函数的多层感知机的误差反向传播（Error Back Propagation，BP）算法，时至今日的深度学习，BP 依然是中流砥柱。

4. 第二个低谷时期（20 世纪 80 年代末至 90 年代初）

然而，好景不常在，专家系统慢慢暴露出维护难、不完善等缺点。1987 年，Apple 和 IBM 的普通台式机的性能反而超过了“智能计算机”，专家系统被质疑。20 世纪 80 年代末，DARPA 对人工智能的期望也降低了。1991 年，日本的第五代计算机宣告失败，人工智能再一次跌入低谷。

5. 第三个黄金时期（20 世纪 90 年代中至今）

从 20 世纪 90 年代中期开始，人工智能陆续走向台前，处于第三个黄金发展时期，在这个时期发生的事件包括：

- 1997 年，深蓝击败国际象棋世界冠军卡斯帕罗夫。
- 2011 年，IBM Watson 在美国电视知识抢答竞赛节目“危险边缘 (Jeopardy!)”中击败了史上胜率最高的两位人类冠军。
- 2012 年至今，ImageNet 图像分类大赛年年创新高，并在一定程度上超过了人类的识别能力。
- 2016—2017 年，Google DeepMind 创造的 AlphaGo 围棋系统相继战胜世界冠军李世石、柯洁。

世界大国纷纷布局人工智能，美国在 2013 年 4 月由奥巴马政府宣布投入 1.1 亿美元，并从 2014 年开始，每年各投入 3~5 亿美元，十年总计将投入 45 亿美元。

欧盟在 2013 年年初公布了总投资 12 亿欧元的十年计划，预计在 2018 年之前开发出一个具有意识的智能大脑，同时在 2014 年 6 月启动了机器人研发计划，目标是为欧盟各行各业提供机器人。

日本在 2015 年 12 月开展了第五个科学与技术基础五年计划，预计总投资 26 万亿日元，用来实现一个全球领先的“超级智能社会（Super）”，以及发展信息技术以及人工智能、机器人等相关技术。

韩国在 2013 年 5 月提出了 ExoBrain 十年计划，预计总投资 9000 万美元，计划开发专业领域的人机对话系统。

中国在 2015 年 7 月发布了《国务院关于积极推进“互联网 +”行动的指导意见》，其中人工智能是重点布局的 11 个领域之一；2016 年 5 月发改委公开了《“互联网 +”人工智能三年行动实施方案》；同时，脑科学研究也上升到国家战略高度。

而在这一次人工智能的黄金发展过程中，深度学习起到了至关重要的作用，除上面提到的 ImageNet 图像分类大赛、AlphaGo 与深度学习密切相关外，深度学习使研究者在计算机视觉、语音识别、机器翻译等各个领域都取得了有史以来的最长足进步。深度学习依然是从古老的联结主义发展而来的，是隐层多于一层的神经网络。之所以在这个时期深度学习才得到大力发展，主要原因包括三点：

- 算法——逐层训练初始化模型、分布式并行训练算法等能力的提升。
- 计算——GPU（Graphics Processing Unit）、FPGA（Field-Programmable Gate Array）、TPU（Tensor Processing Unit）等能力的大幅提升。

- 大量的训练数据——例如 ImageNet 千万级别的图像数据。

当然，在这个时期也存在泡沫，有人鼓吹传统行业的从业者即将失业，有人担心自己要被机器取代甚至消灭，甚至很小的深度学习创新也开始被媒体捧到天上去，生怕大家觉得这个创新的影响力不够大。作为技术人来说，刨去这些泡沫，深度学习在很多领域都带来了有史以来最大的技术突破，甚至很多突破远远超出了技术实践者本身的预期。

1.2 机器学习

机器学习（Machine Learning）是一门研究计算机模拟和实现人类行为的科学，通过不断改善知识结构，进而超越人类能力的学科。机器学习算法是从数据中自动分析并获得规律，进而可以对未知数据进行预测的算法。

1.2.1 机器学习的由来

很多时候，人们希望能借助机器的力量来自动完成一些任务，从而将人类从烦琐的事项中解放出来。比如自动监测违规车辆及排查嫌疑车，可以代替交通警察用人眼监控显示屏；自动驾驶，可以选择最佳路线、躲避其他车辆而安全地驾驶；自动人脸识别，可以代替人工完成特定的服务。

概括来说，这个过程涉及了两大步骤。

- 认知这个世界，获取信息。
- 根据信息进行判断和决策。

从人类的角度看，第 2 步显然重要得多。人们进行了种种努力，不断探索如何能排除感性的干扰，做出更加理性的决策。这样的决策在给定信息的情况下，被称为“全局最优”策略。而这一步，在机器看来却轻巧得多，它可以充分调动强大的计算能力，综合各种优化算法，在极短的时间内就能给出最优的答案。

但另一方面——“像人类一样认知这个世界”——却不是它的强项。面对一张图片，它可以告诉你一共有多少个像素点，也可以准确地给出图片上每一个像素点的像素值，但却分辨不出那些像素点组成的脸庞。如图 1-2 所示，十年前机器学习领域还在聚焦于如何能让机器准确地识别类似简单的物体。机器的识别能力甚至比不上一个 3 岁的孩童。不仅如此，在经过训练能认出图片上的物体后，一旦光影变幻、物体遮挡或角度变化，就很可能又会识别失败。

图 1-2　Caltech 101 数据集

人们挠挠头，不知道怎么教机器这个憨憨的学生去“感知”这个世界。于是人们转而看看关于自己大脑的研究，也就是神经科学，希望能获得一些关于“认知”的理论。可惜，当时大脑神经科学也是一片广阔而充满了未解之谜的领域。虽然对神经元的研究、激活和信息传导有了一定的成果，但还不足以解释“认知”这个宏大的课题。尽管如此，人们还是乐观地开始了对机器学习领域中的人工神经网络的研究。经过大半个世纪的坎坷和沉浮，厚积薄发，在 21 世纪开始大放异彩，在各个领域都取得了惊人的进展。

机器学习的领域很广泛，与视觉相关的领域包括：物体识别、图像分割、图像索引、人脸识别、场景识别、场景匹配等。而且还有很多有趣的商业应用，比如谷歌眼镜等。

与听觉相关的领域包括：语音识别、乐曲片段匹配，甚至有自动作曲这样有趣的方向。

与认知相关的领域包括：自然语言处理、专家系统等。

基于对人群偏好的推测而进行的内容推荐，也由于有着广阔的应用场景而成为机器学习中很热门的一个领域，包括诸如网页推荐、广告推荐、购物产品推荐、电影推荐等。

此外，还有很多其他五花八门的方向，比如机器人的相关研究。如图 1-3 所示，斯坦福的 Jackrabbot 就是一个带着领带、风度翩翩的社会化行走机器人。和其他机器人不同，它在人行道上行走时，会特别学习人类的社会习惯，比如行走时注意他人的个人空间、有礼貌地行走，而不是只为了走到目的地而加快步伐地横冲直撞。

这里，我们简单澄清一些与机器学习密切相关且容易混淆的概念。

模式识别（Pattern Recognition）——在一定程度上等同于机器学习，一般认为模式识别最初来自于工业界，而机器学习来自于学术界，机器学习的经典书籍 *Pattern Recognition and Machine Learning* 所讲的就是两者不分家。

图 1-3 Jackrabbot 机器人会遵从社会习俗

统计学习（Statistical Learning）——也是机器学习的近义词，机器学习的很多方法都源自于统计学习，这些方法往往具有优美的数学推导，比如支持向量机（SVM）方法等。总体上，统计学习更偏数学理论一些，而机器学习更偏实践一些。

数据挖掘（Data Mining）——在有些场景中也被等同于机器学习，但更专业的解释应该是机器学习在大数据领域的应用，通过机器学习的方法从大数据中挖掘出规律或知识。

计算机视觉（Computer Vision）——强调机器学习在图像领域的应用。可以说，迄今为止，计算机视觉是机器学习尤其是深度学习最成功应用的领域，没有之一。

语音识别（Speech Recognition）——研究机器听懂人类声音的领域。目前语音识别也取得了长足的进步，有 Siri、语音输入法等大家耳熟能详的应用。

自然语言处理（Natural Language Processing）——研究机器理解人类语言的领域。相比计算机视觉、语音识别的感知问题，自然语言处理尤其是其中的语义理解属于认知问题，相对更难一些。

那么具体来说，机器学习是什么呢？它和本书要讲的神经网络以及深度学习是什么关系呢？下面我们将探讨这些问题。

1.2.2 机器学习发展史

由于机器学习只是实现人工智能的一种方式，所以人工智能的发展史实质上包括了机器学习的发展历程。“推理期”“知识期”“学习期”就是指与机器学习相关的主流时期，感知机、支持向量机、神经网络等又是机器学习具体的模型，在此不再赘述。

可以说，1996 年至今，机器学习在工业界得到广泛应用，从而使机器学习的发展达到一个前所未有的新高度。

比如搜索引擎中的分词、新词挖掘、垃圾网页过滤、网页滤重、Learning to Rank、PageRank、主题模型、摘要提取、特征学习等，大量使用了机器学习中的逻辑回归（Logistic Regression，LR）、支持向量机（Supported Vector Machine，SVM）、GBDT（Gradient Boosting Decision Tree）、Latent Semantic Analysis（LSA）/Probabilistic Latent Semantic Analysis（PLSA）/Latent Dirichlet Allocation（LDA）、概率图、深度学习等模型。

再比如在计算广告点击率预测中广泛使用了机器学习中的 LR（Logistic Regression）、BPR（Bayesian Probit Regression）、FTRL（Follow-The-Regularized-Leader）、Online Learning、深度学习等相关技术。

而计算机视觉、语音识别、机器翻译等更是在近几年被深度学习不断刷新高度。

1.2.3 机器学习方法分类

机器学习方法可以大致分为监督学习、无监督学习、半监督学习、增强学习等几类。

监督学习（Supervised Learning）——通过对标注的训练数据进行学习，得到一个从输入特征到标签的映射模型，再利用这个模型对未知标签的新数据进行预测。比如我们拥有大量正常内容的邮件，同时拥有大量垃圾邮件，那么就可以训练一个监督学习模型来做垃圾邮件分类，最终得到的模型就能鉴定新邮件是否是垃圾邮件。

监督学习又可以进一步分为分类（Classification）和回归（Regression）等类别。如果标签是离散类别的，则一般认为是分类问题，比如前面提到的垃圾邮件分类等；而如果标签是连续数值型的，则一般认为是回归问题，比如房价的预测问题等。

无监督学习（Unsupervised Learning）——不需要对训练数据进行标注，直接对数据进行建模。比如一堆杂乱无章的文字片段或者图片，我们完全可以根据文字或图片本身的内容对其进行大致的归类。

无监督学习比较常见的类别有聚类（Clustering）、密度估计（Density Estimation）和降维（Dimension Reduction）等。其中，聚类是根据样本之间的特征相似度将一组数据聚为一类，使得类内的数据相似度比不同类间的数据相似度更高。密度估计是根据数据集统计推断样本集对应的概率分布。降维，顾名思义，就是降低输入数据的维度。在很多应用中，原始数据具有非常高的维度（比如在广告点击率预测应用中，特征维度往往达到上亿级别），而且有很多特征是冗余或者不相关的，降维算法有助于去除无关特征、合并冗余特征。

半监督学习（Semi-Supervised Learning）——介于监督学习和无监督学习之间的方法。在实际应用中，数据标注往往对模型的学习非常有帮助，但代价也不低，有时候甚至超过了可以忍受的限度，这时候半监督学习就是一种很好的选择。半监督学习的方法非常多，其中

滚雪球式的主动学习（Active Learning）是数据挖掘中非常常用的方法，利用学习算法主动选出最值得标注的数据进行人工标注，标注完成后，新的标注数据和之前的标注数据合在一起继续进行训练，训练完毕后继续用算法甄选性价比最高的数据进行人工标注，如此不断迭代，最后得到的模型效果往往非常好。

增强学习（Reinforcement Learning，也翻译成强化学习）——一种交互式的学习方法，模型根据环境给予的奖励或惩罚不断调整自己的策略，尽量获得最大的长远收益。相关的具体介绍可以参考第 27 章。

1.2.4 机器学习中的基本概念

在机器学习算法中，目前在业界得到较多应用的主要是监督学习，监督学习需要训练数据，其本身由模型、策略和算法[3] 三要素组成。

机器学习的模型是从数据中学到的用来描述数据所在空间的数学模型，可以说是经过了数学抽象的规律。模型是机器学习的最终目的，有了模型，才能对未知数据进行预测或分析。在监督学习中，模型一般指具体的条件概率分布模型或者决策模型。模型的假设空间 $\mathcal{F}$ 是所有可能的条件概率分布或者决策函数。这里为了简单起见，仅以决策函数为例进行说明。假设输入集合为 X，对应的输出集合（标签，Label）为 Y，其假设空间可以表示为：

$$\mathcal{F} = \{f|Y \sim f(X)\}$$

机器学习常用的模型有很多，比如线性模型、逻辑回归、Softmax、神经网络/深度学习、SVM、决策树、随机森林、GBDT、与矩阵分解相关的系列模型等。

由于假设空间对应的函数有很多，对于如何选择就需要引入特定的评估策略，机器学习一般引入损失函数（Loss Function）或代价函数（Cost Function）来评估预测错误的程度。

常见的损失函数如表 1-1 所示，其中 y 表示对应输入样本 x 的 Label，$\hat{y}$ 为函数预测值 $f(x)$，注意这里只针对单个样本计算损失。

表 1-1 常见的损失函数

损失函数	公式
0-1 损失	$L(y,\hat{y}) = \begin{cases} 0, & y = \hat{y} \\ 1, & y \neq \hat{y} \end{cases}$
绝对值损失	$L(y,\hat{y}) = \lvert y - \hat{y} \rvert$

续表

损失函数	公式
平方差损失	$L(y,\hat{y})=(y-\hat{y})^2$
负 Log 似然	$L=-y\log p(y=1\|x)-(1-y)\log(1-p(y=1\|x))$

损失函数在输入输出联合概率分布 $p(x,y)$ 下的期望称为风险函数（Risk Function）或者期望损失（Expected Loss），用 $R(f)$ 表示[3]。

$$R_{\exp}(f)=E_p[L(y,\hat{y})]=\int L(y,\hat{y})p(x,y)\mathrm{d}x\mathrm{d}y$$

机器学习的目标是要选择一个期望损失最小的模型，但是这在现实中不可行，因为联合概率分布 $p(x,y)$ 表征的是所有样本遵循的分布，一般无法求得。只能退而求其次利用训练数据集的平均损失近似表示，这个平均损失称为经验风险（Empirical Risk）或者经验损失（Empirical Loss），记为：

$$R_{\mathrm{emp}}(f)=\frac{1}{N}\sum_{i=1}^{N}L(y_i,\hat{y}_i)$$

其中，N 为样本数量。根据大数定律，当 N 趋近于无穷大时，$R_{\mathrm{emp}}(f)$ 趋近于 $R_{\exp}(f)$[3]。

算法是机器学习的具体学习方法，也可以称为优化算法，包括：梯度下降法、牛顿法、拟牛顿法等[4]。

1.3 神经网络

在人工智能的发展史中，我们已经反复发现了神经网络的踪影。作为机器学习的一个分支，神经网络的发展也是跌宕起伏的。

神经网络（Neural Network）是从人类脑神经元的研究中获得灵感，模拟其神经元的功能和网络结构，来完成认知任务的一类机器学习算法；还有一类机器学习算法，则不局限于神经元，而是尝试将问题从数学上抽象，从而对该简化的数学问题进行研究并做出解答。

而深度学习（Deep Learning），则是指多层神经网络，即隐层大于一层的神经网络。在后面的章节中，我们还会详细地讲一讲网络结构和隐层到底是什么。

1.3.1 神经网络发展史

1. 神经网络的提出与发展（1943—1969 年）

早在 1943 年，人工神经网络就已由 McCulloc 和 Pitts 提出，他们分析了理想化的人工神经元网络，并且指出了它们运行简单逻辑运算的机制。但这仅仅是一种理想化的蓝图。直至将近 15 年后，康奈尔大学的实验心理学家 Frank Rosenblatt 在一台 IBM–704 计算机上模拟实现了一种他发明的叫作“感知机”的神经网络模型，人工神经网络才走进了现实。稍后，伴随着 Frank Rosenblatt 出版的一本名为《神经动力学原理：感知机和大脑机制的理论》的书，感知机迅速获得了人们的关注，并被寄予了极高的期望。

然而，1969 年，一本名为《感知机》的书详细地分析了感知机的适用范围，并明确提出对于简单的异或逻辑问题，感知机都由于其非线性而无法解决，而现实中的问题恰巧大多都不是线性可分的。尽管在 5 年后，Werbos 的博士论文证明了，只要在感知机的网络中多加一层，并且利用“后向传播”的学习方法，就可以解决异或问题，但是人们依然对感知机持悲观的态度。不仅如此，这种看法还扩大到所有的神经网络科学上，以至于对整个神经网络的研究陷入了停滞状态。

为了本书内容的简洁性，以下如不特指，“神经网络”均指代“人工神经网络”。

2. 神经网络的困境与 SVM 的独领风骚（1971—2005 年）

从 20 世纪 70 年代开始，人们对神经网络的研究热情不断下降。与此同时，以 Vapnik 为首的科学家创造性地提出了 VC 维的概念，以及结构风险最小化原则。Vapnik 是研究统计学出身的，数学功底深厚。随着这个理论的深入，并经过 20 年的摸索后，Cortes 和 Vapnik 等人在 1993 年提出了“支持向量机”（Support Vector Machine），成功地将其应用于实际问题中。支持向量机旨在利用核（Kernel）技巧把非线性问题转换成线性问题，解决了感知机所不能解决的问题，一时间独领风骚。其坚实的理论基础和解决现实问题的有效性，使它获得了广泛的认可。而同时，Vapnik 等统计机器学习理论专家从理论的角度怀疑神经网络的泛化能力，学术界对于神经网络的研究也更加趋于悲观。

尽管如此，在这长达半个世纪的冰河期，依然有神经网络学家在坚守着自己的阵地。1982 年，Hopfield 提出了一种新的神经网络，它可以解决一大类模式识别问题，并且可以给出一类组合优化问题的近似解。1986 年，Rummelhart 与 McClelland 再次提出了神经网络的学习算法——后向传播。LeCun 也发明了卷积神经网络，并利用其实现自动提取图像的特征，成功地完成了手写数字的识别。这些都为后来神经网络的再次兴起奠定了坚实的基础。

3. Hinton 引领的神经网络复兴（2006 年）

2006 年，Hinton 提出了深度神经网络（深度学习）。深度神经网络指的是隐层大于一层的网络结构。Hinton 提出首先用 Restricted Boltzmann Machine 经过非监督学习来学习出网络结构，然后再由后向传播算法学习网络内部的参数值。

尽管如此，深度学习仍然广受质疑。于是 Hinton 带领其学生埋头苦干，于 2012 年在计算机视觉领域的著名比赛——ImageNet 分类比赛中，以高出第二名 10 个百分点的战绩高调地夺得第一名。这是深度学习在沉寂半个世纪后，第一次在机器学习领域的比赛中参赛，并且取得了卓著的成绩，震动了整个机器学习界。这一成绩坚实地印证了深度学习的有效性，并随后在各个领域也都迅速拔得头筹。深度学习正式进入了复兴和辉煌的时代。

这一次复兴，离不开 Hinton、LeCun、Bengio（关系如图 1-4 所示）和其他优秀研究者（比如第四巨头 Andrew Ng 等）的努力工作，他们坚信神经网络的有效实用性，并不断摸索真正可行的神经网络道路。21 世纪不断普及的大数据以及高度并行的计算设备——图形处理单元（Graphics Processing Unit，GPU）也为神经网络提供了必不可少的支持。有了这些，才有了深度学习在各个领域遍地开花的今天。

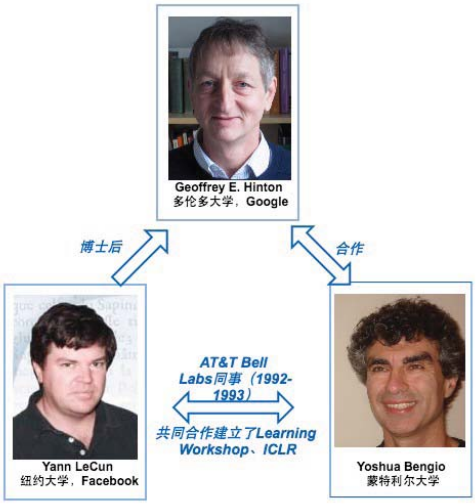

图 1-4　深度学习巨头 Hinton、LeCun 和 Bengio 的关系

回顾神经网络的兴起——衰落——复兴乃至于辉煌的过程，不禁让人唏嘘。如今深度学习研究的大放异彩，离不开大师们近半个世纪的坚守和在质疑中坚定地前行。这不仅需要灵感，还需要魄力，以及一以贯之的坚定的信心。

深度学习在图像识别领域大获成功之后，又被迅速应用到其他问题上。看起来各不相同的问题，一旦理解它们仅仅是特征不同、基于特征都要完成对应的分类问题时，各个问题似乎就有了相似之处。当然，在实践中，能成功地把深度学习应用于各类问题上还是需要相当的想象力、创造力以及对模型的把控力的。图 1-5 列举了部分精彩的实例，有些似乎超出了人们的想象，却都成为了现实。而对于实现这些有趣应用的神秘而强大的深度学习，我们也将揭开它的面纱。

图 1-5 深度学习在各领域遍地开花

参考文献

[1] Deep Learning Explained - NVIDIA. https://www.nvidia.com/content/dam/en-zz/Solutions/deep-learning/home/DeepLearning_eBook_FINAL.pdf.

[2] MAC 工程. http://www.multicians.org/project-mac.html.

[3] 李航. 统计学习方法. 北京：清华大学出版社，2012.

[4] Stephen Boyd and Lieven Vandenberghe. Convex Optimization. Cambridge University Press.