



# 绪 论

## 1.1 人工神经网络发展史

### 1.1.1 人工神经网络的提出

人工智能(artificial intelligence, AI),如同“长生不老”和“星际漫游”一样,是人类最美好的梦想之一。虽然计算机技术已经取得了长足的进步,但是到目前为止,还没有一台计算机能够产生类人智能。1950年,计算机和人工智能的鼻祖——图灵,提出著名的“图灵机”和“图灵测试”。图灵试验的设想,即隔墙对话,如果人类无法辨别谈话对象是人还是计算机,可判定对方具有人工智能特性。这在当时无疑给计算机,尤其是人工智能,预设了一个很高的期望值。随后人工智能的进展,远远没有达到图灵试验的标准,最终造成了人们认为人工智能及相关领域是“伪科学”。

近年来,由于深度学习及人工智能领域发展迅速,使人们对人工智能产生了新的遐想。深度学习是人工神经网络的延伸与发展,而人工神经网络是对生物神经网络基本功能的模拟。人工神经网络起源于人类对人脑和智能的探索,是生物神经元网络互相连接结构的简化数学模型,又被称为连接主义,是人工智能的一个重要分支。

人工神经元数学描述模型是人工神经网络的基础。1943年,Warren McCulloch 和 Walter Pitts 两位科学家(如图 1-1)发表题为 *A Logical Calculus of the Ideas Immanent in Nervous Activity* 的论文,首次提出人工神经元的数学描述——McCulloch-Pitts 神经元模型(简称 M-P 神经元模型),如图 1-2 所示。该模型借鉴了已知的生物神经细胞原理,通过数学方法对其信息处理过程进行描述,是人类历史上第一次对大脑工作原理描述的尝试,也是第一个神经元数学模型。M-P 神经元模型可以看作连接主义的起源、人工神经网络与深度学习的基石。

M-P 神经元模型模拟生物神经元的时空信息整合特性,对神经元的输入信号进行加权求和,并与阈值进行比较,然后经过激活函数的运算,最后产生神经元激活输出。两位科学家从数学上证明了由人工神经元构成的人工神经网络可以计算任何算术和逻辑函数,但由于当时没有找到有效的训练算法,权值需要根据问题进行事先设定,缺乏实用性。

人工神经元权值调整思想何时被提出? 1949年,神经学家 Donald Olding Hebb(如图 1-3)在 *The Organization of Behavior* 一书中对生物神经元之间连接强度的变化进行了

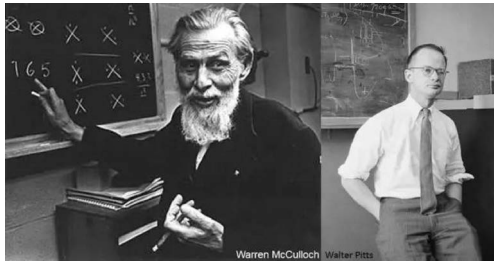


图 1-1 Warren McCulloch 和 Walter Pitts

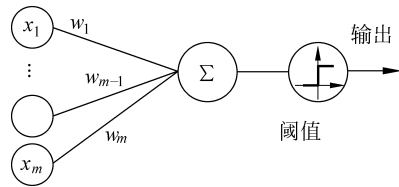


图 1-2 M-P 神经元模型

分析,首次提出一种神经元权值调整方法,称为 Hebb 学习规则,从而将神经网络和机器学习联系起来,并且这一思想与生物神经元的条件反射一致。Hebb 规则为人工神经网络的训练奠定了基础。

人工神经网络的开创者 Frank Rosenblatt,如图 1-4 所示,他也是一名计算机学家和心理学家,早在 1951 年他还是一名研究生的时候,其在暑期作业中便开发出了一个神经网络系统 SNARC。1957 年,就职于康奈尔大学航空实验室的 Frank Rosenblatt,发明了感知器人工神经网络。该网络被视为一种最简单形式的前馈神经网络模型,是一种二元线性分类器,其激活函数采用了符号函数  $\text{sign}(x)$ 。感知器是人工神经网络的第一个实际应用,标志着神经网络进入了新的发展阶段。Frank Rosenblatt 通过软件实现感知器后,开始构建硬件版本的感知器 Mark I。他将 400 个光电设备作为神经元,将可调电位器作为突触权重,并且在学习期间权重的变化由电动机调整电位器实现,构造了硬件感知器,并用于图像识别。由于当时的技术限制,基于物理实现的感知器是比较罕见的。这种用于图像识别的系统受到了各方的关注,并因此从美国海军获得了大量资金。



图 1-3 Donald Olding Hebb



图 1-4 Frank Rosenblatt

这次成功的应用也引起了学者对神经网络的广泛兴趣。1960 年,斯坦福大学教授 Bernard Widrow 和他的研究生 Ted Hoff 开发了 Adaline 网络和最小均方滤波器,Adaline 网络将感知器的阶跃函数替换为线性函数,如图 1-5 所示。

### 1.1.2 人工神经网络的陨落

人工智能的另一个分支——符号主义,在此阶段的发展基本趋于停滞。Dartmouth John McCarthy 和 Marvin Minsky 在 1956 年达特茅斯学院举行的会议中提出了人工智能的概念,用来反对早期控制论里的联结主义。会议指出机器根据输入和输出进行自适应调整

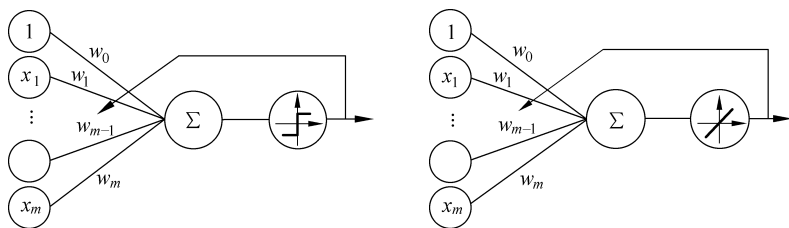


图 1-5 Perceptron(感知器)与 Adaline 对比

是不够的,“符号主义”的目标是用人工定义的规则编制程序算法,并融入计算机系统中,提高操纵系统的智能性。20 世纪 50 年代至 70 年代初,早期 AI 的核心都是推理模型,所以 AI 诞生之初对联结主义的一些观点是排斥的。这个研究领域主要被麻省理工学院的 Minsky Papert、卡内基梅隆大学的 Simon Newell 和斯坦福大学的 McCarthy 等研究人员所主导。1969 年,Marvin Minsky 和 Seymour Papert(图 1-6)撰写了 *Perceptrons: an Introduction to Computational Geometry* 一书,从数学的角度证明了单层感知器网络具有有限的功能,甚至在面对简单的“异或”逻辑问题时也显得无能为力。

虽然多层感知器能够解决异或问题,但当时并没有找到有效的训练算法。这一书的出版,无疑给神经网络的研究浇了一盆冷水。1971 年, Frank Rosenblatt 英年早逝之后,神经网络被抛弃,相关项目的资金资助被停止,神经网络的研究陷入了很长一段时间的低迷期。



图 1-6 Marvin Minsky 与 Seymour Papert

20 世纪 70 年代早期,人工智能的研究进入了它的第一个冬天,联结主义和符号主义的研究都处于停滞状态。两个流派都承诺过多,而结果却遥遥无期。联结主义一边, Frank Rosenblatt 的感知机被过早公之于众,在《激动人心的智能机器》新闻中报道“电子计算机雏形出现,海军希望它能走、说、看、写,制造自己,甚至拥有自我意识”。符号主义一边,以 Herbert Simon 和 Marvin Minsky 为首的研究者们,提出翻译俄语的翻译器、渗透进敌人战线的机器人、坦克和飞机驾驶员的语音指挥系统等宏图壮志,宏图面对的现实是“智能系统”还只是机房里的游戏,不切实际的宣言很快被否定掉。尽管这其中的部分功能在目前技术下,已能够很好地实现,但在当时技术条件下是无法实现的。1966 年,美国研究委员会削减了机器翻译的预算,随后撤回了对人工智能的财务和学术支持,包括 Minsky 和 Papert 在麻省理工学院的 Micromonde 项目、斯坦福大学的 Shakey 机器人、美国国防部高级研究计划局的 SUR 语音识别计划等。1973 年,英格兰在重要的莱特希尔(Lighthill)报告中,劝说人们停止对 AI 的公共资助。至此,人工智能研究陷入低谷。

### 1.1.3 人工神经网络的兴起

Frank Rosenblatt 已发现多层感知器能够解决复杂的非线性问题,但他并没有找到有效的多层感知机训练算法。直到 1974 年,哈佛大学的 Paul Werbos 博士,在其博士论文中提出了影响深远的 BP (back propagation)神经网络学习算法。1982 年,David Parker 重新发现了 BP 神经网络学习算法,但都没有引起人们的重视,并未唤起人们对人工神经网络研究的兴趣。同年,John Hopfield(图 1-7)提出了连续型和离散型的 Hopfield 神经网络模型,

并采用全互联型神经网络模型,实现对复杂的旅行商问题进行求解,向人们展示了神经网络的求解能力,促使神经网络的研究再次进入了蓬勃发展的时期。

1983年,Geoffrey E. Hinton(图 1-8)和 Terrence J. Sejnowski 设计了玻尔兹曼机,首次提出了“隐单元”的概念。在全连接的反馈神经网络中,包含了可见层和一个隐层,这就是玻尔兹曼机。神经网络层数的增加可以提供更大的灵活性,但参数的训练算法一直是制约多层神经网络发展的一个重要瓶颈。而一个沉睡近十年的伟大算法即将被唤醒。1986年,David E. Rumelhart、Geoffrey E. Hinton 和 Ronald J. Williams 发表文章 *Learning representations by back-propagating errors*,重新报道了 BP 神经网络学习算法,引起了人们对 BP 算法的重视。BP 算法引入的可微非线性神经元(如 Sigmoid 激活函数神经元),克服了早期神经元的弱点,为多层神经网络的学习训练与实现提供了一种切实可行的方法。但基于 Sigmoid 激活函数的 BP 网络是一种全局逼近网络,网络收敛速度非常慢。1988年,继 BP 算法之后,David Broomhead 和 David Lowe 将径向基函数引入到神经网络的设计中,构造了径向基神经网络,由于径向基神经网络是一种局部逼近网络,其收敛速度更快。因此,径向基神经网络是神经网络真正走向实用化的一个重要标志。1989年,George Cybenko 证明了“万能逼近定理”,证明多层前馈网络可以近似任意函数。对于具有单隐层、传递函数为 Sigmoid() 的连续型前馈神经网络,只要隐层神经元的个数足够多,网络可以以任意精度逼近任意复杂的连续映射。这样,BP 神经网络凭借其其对复杂连续函数的刻画能力,打开了被 Marvin Minsky 和 Seymour Papert 早已关闭的人工神经网络研究的大门。

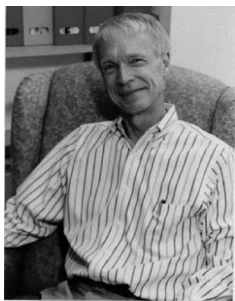


图 1-7 John Hopfield



图 1-8 Geoffrey E. Hinton

由于浅层神经网络容易产生过拟合以及参数训练速度慢等原因,人工神经网络的发展进入平淡期。统计学习理论是一种专门研究小样本情况下机器学习规律的理论,自二十世纪六七十年代已有相关研究,到九十年代中期,随着其理论的不断发展和成熟,也由于神经网络等学习方法缺乏实质性进展,统计学习理论开始受到越来越广泛的重视。同时,在这一理论基础上发展了一种新的通用学习方法——支持向量机,支持向量机表现出很多优于已有方法的性能。值得一提的是,1997年 Sepp Hochreiter 和 Jürgen Schmidhuber 提出长短期记忆模型,为序列信息的建模奠定了基础。

#### 1.1.4 深度学习的提出

随着科技的发展,计算机处理速度、存储能力得到了大幅度提高,同时大数据的广泛应用,为深度学习(deep learning)的提出铺平了道路。2006年,Geoffrey E. Hinton 和他的学生 R. R. Salakhutdinov 在科学杂志上发表题为 *Reducing the Dimensionality of Data with*

*Neural Networks* 的文章,展示了深度神经网络的魅力,掀起了神经网络在学术界和工业界的研究热潮。文章阐述了两个重要观点:一是多隐层的神经网络可以学习到刻画数据本质属性的特征,对数据可视化和分类等任务具有很大帮助;二是可以借助无监督的“逐层预训练”策略,克服深层神经网络在训练上存在的问题。这篇文章是一个分水岭,拉开了深度学习的大幕,标志着深度学习的诞生。因此,2006年以前的神经网络研究常被称为浅层学习,典型的浅层学习模型包括:单层感知器、BP网络、传统隐马尔可夫模型、条件随机场、最大熵模型、集成学习、支持向量机、核回归及仅含单隐层的多层感知器等,而2006年后的神经网络研究被称为深度学习。同年,Hinton又提出了深度信念网络。深度信念网络是基于受限玻尔兹曼机构建的深度网络。玻尔兹曼机由Geoffrey E. Hinton和Terrence J. Sejnowski在1985年提出的,1986年Paul Smolensky命名了受限玻尔兹曼机,但直到Hinton及其合作者在2006年发明快速学习算法后,受限玻尔兹曼机才变得知名。自动编码器早在1986年被提出,2006年之后,Hinton等人利用自动编码器实现了深度自编码器、稀疏自编码器等,以及利用自编码器构造了深度神经网络。

卷积神经网络作为一种典型的深度学习网络结构,已经成为当前图像处理领域研究的主要工具。早在1989年,Yan Lecun等人提出卷积技术,对AT&T贝尔实验室的邮政编码进行识别,使用美国邮政服务数据库训练多层网络,以识别包裹上的邮政编码。这种方法成功应用于银行业(支票金额识别)和邮政行业中。但由于当时的计算机运算能力以及数据量的限制,该网络没有得到有效发展。1998年,Yann LeCun提出了用于字符识别的卷积神经网络LeNet-5,并在小规模手写数字识别中取得了较好的结果。因此,Yann LeCun也被称为卷积网络之父。2012年,Alex Krizhevsky等使用AlexNet卷积神经网络,在ImageNet图像分类竞赛任务中以大幅优势取得了冠军,卷积神经网络在图像分类中取得了巨大成功。随后Alex Krizhevsky,Ilya Sutskever和Geoffrey E. Hinton发表了*ImageNet Classification with Deep Convolutional Neural Networks*的文章,为卷积神经网络在图像处理领域的应用奠定了基础。

2018年被称为深度学习三巨头的Yoshua Bengio,Yann LeCun和Geoffrey E. Hinton由于其在概念和工程上取得的巨大突破,使得深度神经网络成为计算的关键元素,而获得图灵奖。而后,长短期记忆网络(LSTM)之父Jürgen Schmidhuber极力肯定和推广长短期记忆网络在人工神经网络和深度学习领域的巨大作用。2021年,长短期记忆网络提出者和奠基者Sepp Hochreiter获得了IEEE CIS神经网络先驱奖,以表彰他对长短期记忆网络的发展做出的卓越贡献。Geoffrey E. Hinton于1998年,Yann LeCun于2014年,Yoshua Bengio于2019年分别都已获得IEEE神经网络先驱奖,如图1-9所示。

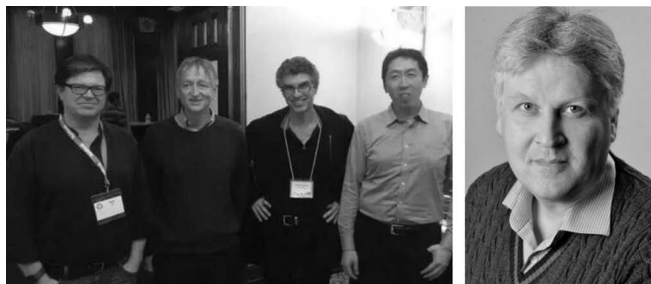


图 1-9 深度学习核心人物(LeCun,Hinton,Bengio 和吴恩达)及 Sepp Hochreiter

## 1.2 人工神经网络学习机理

深度学习的研究已经渗透到生活的各个领域,已成为人工智能技术的重要发展分支。人工智能最终的目的是使机器具备与人相当的归纳能力、学习能力、分析能力和逻辑思考能力,虽然当前的技术离这一目标还很遥远,但是深度学习无疑提供了一种可能的途径,使得机器在单一领域的的能力超越人类。

### 1.2.1 生物学机理

无论是浅层网络还是深度学习,都是受到生物系统启发而提出的。早在1958年,David Hubel和Torsten Wiesel在约翰·霍普金斯大学就开始研究瞳孔区域与大脑皮层神经元的对应关系。研究人员在猫的后脑头骨上,开了一个3毫米的小孔,向孔内插入电极,测量神经元的活跃程度。当在猫的眼前展现各种形状、亮度的物体时,改变物体放置的位置和角度,使猫瞳孔感受不同类型与不同强弱的刺激。试验的目的是去证明一个猜测:位于后脑皮层的不同视觉神经元与瞳孔所受刺激之间存在某种对应关系。实验结果表明,一旦瞳孔受到某一种刺激,后脑皮层的某一部分神经元就会活跃。经历了反复的试验,David Hubel和Torsten Wiesel发现了一种被称为“方向选择性细胞”的神经元细胞。当瞳孔发现眼前物体的边缘,并且边缘指向某个方向时,这种神经元细胞就会活跃。这个发现激发了人们对于神经系统的进一步思考。神经-中枢-大脑的工作过程,或许是一个不断迭代、不断抽象的过程。抽象是指从原始信号,到低级抽象,并逐渐向高级抽象迭代。人类的逻辑思维经常使用高度抽象的概念。例如,从原始信号摄入(瞳孔摄入像素),经过初步处理(大脑皮层某些细胞发现边缘和方向),然后抽象(大脑判定眼前的物体的形状等),然后进一步抽象(进一步判定该目标物体)。而这一系列处理过程便是特征提取与概念抽象的过程。1981年的诺贝尔生理学或医学奖,颁发给了David Hubel、Torsten Wiesel以及Roger Sperry,以奖励其在“视觉系统信息处理”研究中的贡献,即可视皮层是分级的。神经网络的结构设计也都遵循了这一猜想。

关于灵长类是如何识别其他动物,尤其是同类的脸,一直以来众说纷纭。2005年,美国加州大学洛杉矶分校的研究人员通过向被试者们展示许多照片,当饰演过影视剧老友记中“瑞秋”的著名女演员出现时,许多被试者大脑中都有一个相同的神经元被激活,被称为“安妮斯顿细胞”。受此实验启发,其他研究人员还找到了“茱莉娅·罗伯茨细胞”“科比细胞”等。因此,多年以来学界都怀疑,大脑中负责人脸识别的神经元是被“定制”的,也被称为“祖母细胞”。这些专属神经元会被一张特定或相似的人脸快速激活,差别较大的人脸就无法响应。2017年,加州理工学院的神经科学家Doris Tsao和Le Chang的一项研究推翻了这一假设。Doris Tsao和Le Chang研究了两只猕猴的大脑,试图确定动物大脑中脸细胞的位置。研究人员从50个维度对人脸进行表征,将不同的人脸或其他物体的图像展示给这两只猕猴。同时,研究人员将电极植入两只猕猴的大脑,猕猴在观看围绕50个维度有所差异的人类脸部照片时,研究人员记录了猕猴大脑中205个脸部识别神经元对这50个维度的不同反应。研究人员对得到的上百万种反馈信息进行解码,得到了每种反馈的具体含义。每个神经元都会对一些面部参数的特定组合产生响应,这些各有分工的神经元从不同的角度分

析人脸,形成的信息组合到一起,就拼凑出一张完整的人脸,这一发现推翻了人脸由特定细胞识别的假说。猕猴对脸部的识别是由大脑中2个面部补丁区的200多个不同神经元共同编码完成面部重建的,其中一个区域有106个细胞,另一个区域有99个细胞。这一研究成果进一步证实了神经-中枢-大脑的工作过程是一个不断迭代、不断抽象的过程,以及确定了特征学习在神经元活动及脑部认知中的重要性。

### 1.2.2 浅层学习

浅层学习是机器学习的第一次浪潮。20世纪80年代末期,用于人工神经网络的反向传播算法的发明,重振了神经网络领域的研究。BP算法使得人工神经网络模型能够从训练样本中学习统计规律,从而对未知事件进行预测。基于人工神经网络的机器学习方法与人工规则系统相比,在很多方面具有优越性。这时的人工神经网络,虽被称作多层感知器,但实际只是含有一层隐节点的浅层模型。20世纪90年代,各种各样的浅层机器学习模型相继被提出,例如支撑向量机、Boosting方法等。这些模型的结构基本上可以看作带有一层隐节点的网络结构,或是没有隐节点的浅层模型。这些模型无论是在理论分析还是应用中都获得了巨大的成功。相比之下,由于人工神经网络缺乏理论分析,训练方法过度依赖经验和技巧,收敛速度慢等问题,这个时期的浅层人工神经网络相对沉寂。

BP算法作为传统多层人工神经网络的训练算法,实际上对具有两层以上隐含层的神经网络,已无法有效地训练。深度神经网络(涉及多个非线性处理单元层)是一种非凸目标函数,其训练过程中普遍存在局部极小的问题,使得训练更加困难。神经网络训练困难主要源于BP算法的固有缺陷:

- (1) 梯度值被压缩:在误差向输入层的传播过程中,误差校正信号越来越小;
- (2) 收敛到局部最小值:随机值初始化会导致该情况的发生,尤其是远离最优区域时;
- (3) 网络的训练只能使用有标签数据:现实中大部分的数据是没标签的,而大脑可以从无标签数据中学习。

### 1.2.3 深度学习

深度学习是机器的第二次浪潮。2006年,深度学习概念被提出。深度学习是机器学习研究的一种新尝试,是人工神经网络的延伸,其目的在于建立可模拟人脑进行数据分析、学习的神经网络模型,例如图像分析、声音和文本处理等。深度学习的实质是通过构建具有多隐层的机器学习模型,利用海量的训练数据,训练网络的特征提取能力,最终提升模型分类或预测的准确性。因此,“深度学习”是一种手段,而“特征学习”才是目的。深度学习模仿人脑的机制来解释数据,通过组合低层特征形成更加抽象的高层属性或特征,以发现数据的分布式特征表示。深度学习的概念源于人工神经网络的研究,含多隐层的多层感知器就是一种深度网络结构。Hinton提出的基于无监督的网络预训练方法,有效地克服了多层神经网络训练困难的问题。深度神经网络的训练被分为两步:一是通过无监督算法每次训练一层网络,实现网络的预训练;二是对已建立的深度网络模型进行有监督的微调。深度学习训练过程具体步骤如下:

- (1) 使用自下而上非监督训练(从底层网络训练开始,逐层向高层训练):采用无标定数据(有标定数据也可),以及无监督训练算法分层训练各层参数,这是与传统神经网络区别

最大的部分,这一过程可以看作网络特征学习的过程。首先,使用无标定数据训练网络的第 1 层参数,网络的训练可以看作是最小化具有单隐层神经网络输出重构误差的过程,由于模型容量的限制以及稀疏性约束,使得得到的模型能够学习到数据本身的结构,从而获得比输入更具有表示能力的特征信息;然后,利用第 1 隐层的输出作为第 2 隐层的输入,进行网络第 2 层的预训练;同理,在学习得到第  $n-1$  层后,将第  $n-1$  层的输出作为第  $n$  层的输入,预训练第  $n$  层,由此得到各层的预训练参数。

(2) 自顶向下的监督学习(利用带标签的数据进行误差反传,对网络进行微调): 基于第一步得到的各层参数进一步微调整个多层模型的参数,网络的训练是一个有监督训练过程。预训练过程类似神经网络参数的随机初始化过程,由于深度学习的第一步不是随机初始化,而是通过学习输入数据的特征而得到的,因此这个初值更接近全局最优解,从而使得网络能够更有效地训练。深度神经网络的训练很大程度上归功于预训练过程的特征学习。

深度学习有两个主要观点: 一是深层网络具有强大的特征学习能力,多隐层神经网络优异的特征学习能力能够得到对数据有更本质的刻画特征,从而有利于分类; 二是深度神经网络的训练可以通过“逐层初始化”来进行预训练,从而实现网络的有效训练。深度学习与浅层学习的区别在于深度学习强调模型的深度,通常有 5 层甚至更多的隐层,同时深度学习明确突出了特征学习的重要性,通过逐层特征变换,将样本在原空间的特征表示到一个新的特征空间,从而使分类或预测更加容易。另外,深度学习利用大数据训练网络的特征提取能力,能够刻画数据丰富的内在信息。

随着 2006 年深度学习的提出,机器学习领域取得了突破性进展。图灵试验,不再是可望而不可即。深度学习技术不仅依赖于计算平台、云计算等对大数据的并行处理能力,而且依赖于更加有效的算法。借助于深度学习算法,人们解决了“抽象概念”描述这个亘古难题。2012 年 6 月,由斯坦福大学著名的在机器学习方面的教授 Andrew Ng 与在大规模计算机系统方面的世界顶尖专家 Jeff Dean 共同主导了 Google Brain 项目,吸引了公众的广泛关注。项目利用 16000 个中央处理器(CPU)核心的并行计算平台训练一种深度神经网络机器学习模型,网络内部共有 10 亿个节点。虽然无法与人脑中 150 多亿个神经元相提并论(人脑神经元互相连接的突触数非常庞大,曾经有人估算过,如果将一个人大脑中所有神经细胞的轴突和树突依次连接起来,并拉成一根直线长度是地球与月亮间距离的两倍),但对人工神经网络来说已是非常巨大。深度神经网络在图像识别、语音识别等领域获得了巨大的成功。2012 年 11 月,微软在中国天津的一次活动中公开演示了一个全自动的同声传译系统,讲演者使用英文演讲,后台的计算机自动完成语音识别、英中机器翻译和中文语音合成,效果非常流畅。据报道,后台关键技术之一便是深度学习。谷歌基于人工智能技术开发了智能围棋系统——阿尔法狗(AlphaGo),并测试了其围棋水平。2016 年 3 月 AlphaGo 约战了韩国围棋高手李世石,并以 4:1 的成绩取得了胜利; 2017 年 5 月,柯洁对战 AlphaGo 的升级版,最终以 0:3 败下阵来。这标志着专用人工智能取得了突破性进展,它的竞赛性能在测试中已超越人类。而这只是开始。2018 年,在多人对战游戏中,OpenAI 在 5 对 5 的 DOTA2 中战胜人类玩家,DeepMind AI 在多人射击游戏中战胜人类玩家,IBM 举办了人机辩论大赛; 2019 年,人工智能在星际争霸 2 的人机对战中以 10:1 的成绩战胜人类玩家。2022 年年底,美国人工智能研究实验室 OpenAI 推出一种人工智能技术驱动的智能聊天机器人 ChatGPT。该模型利用真实世界中大量的语料数据进行训练,使得 ChatGPT 上知天

文、下知地理,还具有能根据聊天的上下文进行互动的能力,具有与真正人类几乎无异的聊天能力。ChatGPT 不只是一个聊天机器人,还能进行撰写邮件、视频脚本、文案、翻译、代码等任务。ChatGPT 的提出为通用人工智能的发展提供了新的路径。人工智能技术还在不断的快速发展,并有望取得令人瞩目的成绩。

### 1.2.4 特征学习

机器学习是一门专门研究利用计算机模拟或实现人类学习行为的方法,以获取新的知识或技能,并重新组织已有的知识结构使自身性能不断改善的学科。目前,通过机器学习去解决问题的思路如图 1-10 所示(以视觉感知为例):

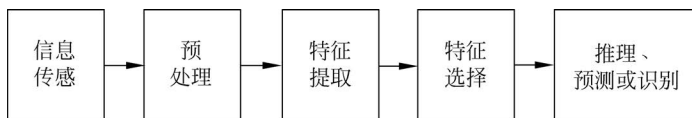


图 1-10 视觉信息处理过程

视觉感知通过视觉传感器获得图像数据,经过预处理、特征提取、特征选择,再到推理、预测或者识别。最后一个环节,也是机器学习的主要研究内容。而中间的三部分,概括起来就是特征表达。良好的特征表达,对算法最终的识别准确度至关重要,而且系统主要的计算和测试工作都在这一部分。例如图像识别、语音识别、自然语言理解、天气预测、基因表达、内容推荐等。在传统的机器学习算法中,特征提取工作一般采用人工设计的方式完成。目前已有许多特征提取方法,例如在视觉信息处理中,好的特征应具有大小、尺度和旋转等不变性与可区分性,如尺度不变特征变换(scale-invariant feature transform, SIFT)特征提取,是局部图像特征描述子研究领域一项里程碑式的工作,由于 SIFT 对尺度、旋转以及一定视角和光照变化等图像变化都具有不变性,并且 SIFT 具有很强的可区分性,能实现良好的特征提取,因此为最终问题的解决奠定了基础。传统机器学习特征提取方法与预测模型是相互独立的,而深度学习强调了特征学习的重要性,并将特征提取与预测模型融合到一起,实现特征提取与预测模型的同时训练,提升了机器学习模型解决问题的效果。

#### 1. 特征表示的粒度

对一张图像来说,像素级的特征是没有价值的。例如图 1-11 中加菲猫的脸,从像素级别根本得不到任何有用的信息,即无法进行猫脸与非猫脸的区分。而如果是一个具有结构性或者说具有语义的特征信息,如是否具有胡须、耳朵的位置等信息,就容易将猫脸与非猫脸进行区分,学习算法才能发挥作用。

#### 2. 初级(浅层)特征表示

1995 年前后, Bruno Olshausen 和 David Field 两位学者任职于康奈尔大学,他们试图同时利用生理学和计算机的手段,研究视觉问题,提出了一种稀疏编码算法。两位学者从收集到的黑白风景图像中提取 400 个  $16 \text{ 像素} \times 16 \text{ 像素}$  的图像块  $S[k]$ , 随机提取图像块  $T$ , 调整图像块  $S[k]$  的系数  $a[k]$ , 实现图像块  $T$  的表示。稀疏编码的生成是一个重复迭代的过程,每次迭代分两步:

- (1) 选择一组  $S[k]$ , 然后调整  $a[k]$ , 使得  $\text{Sum}_k(a[k] * S[k])$  接近  $T$ ;



-1.0	-1.0	-1.0	-1.0	-1.0	-0.9	-0.6	-0.5	-0.5	-0.5	-0.6	-0.6	-0.6	-0.8	-1.0	-1.0	-1.0	-1.0
-1.0	-1.0	-1.0	-1.0	-0.8	-0.4	-0.4	-0.5	-0.7	-0.7	-0.8	-0.8	-0.8	-0.7	-0.8	-1.0	-1.0	-1.0
-1.0	-1.0	-1.0	-0.9	-0.6	-0.6	0.0	0.3	0.1	-0.1	-0.4	-0.6	-0.8	-1.0	-0.8	-0.9	-1.0	-1.0
-1.0	-1.0	-1.0	-0.4	-0.4	0.4	1.0	0.9	0.9	0.7	0.5	0.0	-0.3	-0.6	-1.0	-0.8	-0.9	-1.0
-1.0	-1.0	-0.7	-0.1	0.4	1.0	1.0	0.9	0.8	0.7	0.6	0.3	-0.1	0.1	-0.4	-0.9	-0.6	-1.0
-1.0	-1.0	-0.4	0.2	0.9	1.0	1.0	0.9	0.8	0.7	0.6	0.3	-0.1	-0.2	-0.1	-0.8	-0.5	-0.8
-1.0	-0.9	-0.5	0.2	1.0	1.0	1.0	0.9	0.8	0.6	0.6	0.3	-0.2	-0.4	-0.2	-0.5	-0.6	-0.6
-1.0	-1.0	-0.5	0.6	1.0	1.0	1.0	0.9	0.8	0.6	0.6	0.4	-0.1	-0.3	-0.3	-0.5	-0.8	-0.6
-1.0	-1.0	-0.3	0.9	0.9	0.8	0.4	0.3	0.7	0.6	0.5	0.2	-0.2	-0.3	-0.3	-0.6	-1.0	-0.9
-1.0	-1.0	-0.3	0.9	0.9	0.8	0.3	0.0	0.6	0.7	-0.1	-0.2	-0.2	-0.5	-0.5	-0.6	-0.9	-1.0
-1.0	-0.8	0.2	0.9	0.9	0.8	0.1	-0.1	0.6	0.8	-0.4	-0.3	-0.3	-0.6	-0.6	-0.6	-0.9	-1.0
-1.0	-0.8	0.6	0.9	1.0	1.0	0.8	0.5	0.9	0.8	-0.4	-0.3	-0.2	-0.5	-0.5	-0.5	-0.7	-0.9
-1.0	-1.0	0.5	0.9	1.0	0.9	0.8	0.9	1.0	0.8	-0.4	-0.4	-0.1	-0.3	-0.3	-0.5	-0.5	-0.9
-1.0	-1.0	0.4	1.0	1.0	0.9	0.8	0.8	0.8	0.6	-0.5	-0.4	-0.1	-0.3	-0.5	-0.5	-0.5	-1.0
-1.0	-1.0	0.0	1.0	0.9	1.0	0.8	0.8	0.7	0.3	-0.5	-0.3	-0.1	-0.3	-0.6	-0.5	-0.8	-1.0
-1.0	-1.0	-0.8	0.7	1.0	0.9	0.9	0.9	0.9	0.6	0.0	-0.3	-0.2	-0.5	-0.6	-0.5	-0.9	-1.0
-1.0	-1.0	-1.0	0.5	1.0	0.9	0.6	0.6	0.4	0.2	-0.3	-0.5	-0.4	-0.5	-0.6	-0.8	-1.0	-1.0
-1.0	-1.0	-1.0	0.3	1.0	0.8	0.6	0.8	0.5	0.0	-0.5	-0.6	-0.4	-0.5	-0.6	-0.9	-1.0	-1.0
-1.0	-1.0	-1.0	-0.5	0.8	0.8	0.9	0.9	0.7	0.1	-0.4	-0.5	-0.4	-0.6	-0.8	-1.0	-1.0	-1.0

图 1-11 图形的像素表示

(2) 固定  $a[k]$ , 在 400 个碎片中, 选择其他更合适的碎片  $S'[k]$ , 替代原有的  $S[k]$ , 使得  $\text{Sum}_k(a[k] * S'[k])$  接近  $T$ 。

经过几次迭代后, 最佳的  $S[k]$  组合被遴选出来。令人惊奇的是, 被选中的  $S[k]$  基本上都是照片中不同物体的边缘线, 这些线段形状相似, 区别在于方向。Bruno Olshausen 和 David Field 的算法结果, 与 David Hubel 和 Torsten Wiesel 的生理发现不谋而合。也就是说, 复杂图形往往由一些基本结构组成。

### 3. 结构性特征表示

小块图形可以由基本边缘构成, 更结构化、更复杂的具有概念性的图形如何表示呢? 这就需要更高层次的特征表示。这是一个层次递进的过程, 高层表达由底层表达的组合而成。底层提取的是图像的边缘, 中间层是这些基的组合, 并获得高一层的基。直观上说, 就是找到有效的特征再将其进行组合, 得到更高一层的特征, 递归向上进行特征组合学习。因此, 在利用不同灰度样本做训练时, 所得边缘是相似的, 但利用边缘组合为目标的局部, 再利用部分局部的目标组合完成的目标都是不同的。