

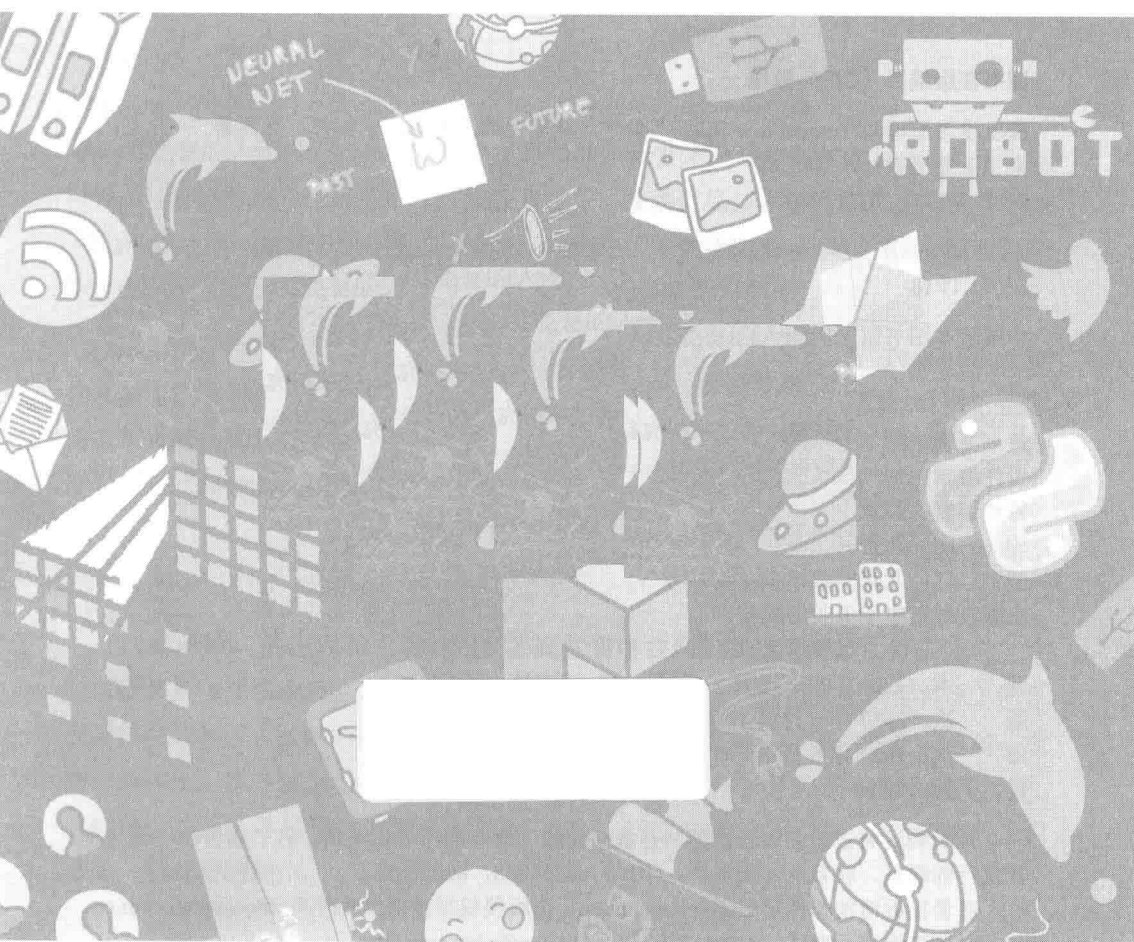
深度学习原理与 TensorFlow实践

喻伊 莫瑜 主编
王琛 胡振邦 高杰 著



深度学习原理与 TensorFlow实践

喻俨 莫瑜 主编
王琛 胡振邦 高杰 著



电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书主要介绍了深度学习的基础原理和 TensorFlow 系统基本使用方法。TensorFlow 是目前机器学习、深度学习领域最优秀的计算系统之一，本书结合实例介绍了使用 TensorFlow 开发机器学习应用的详细方法和步骤。同时，本书着重讲解了用于图像识别的卷积神经网络和用于自然语言处理的循环神经网络的理论知识及其 TensorFlow 实现方法，并结合实际场景和例子描述了深度学习技术的应用范围与效果。

本书非常适合对机器学习、深度学习感兴趣的读者，或是对深度学习理论有所了解，希望尝试更多工程实践的读者，抑或是对工程产品有较多经验，希望学习深度学习理论的读者。

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

图书在版编目 (CIP) 数据

深度学习原理与 TensorFlow 实践 / 喻俨, 莫瑜主编; 王琛, 胡振邦, 高杰著. —北京: 电子工业出版社, 2017.6

ISBN 978-7-121-31298-4

I. ①深… II. ①喻… ②莫… ③王… ④胡… ⑤高… III. ①人工智能—算法—研究
IV. ①TP18

中国版本图书馆 CIP 数据核字(2017)第 071463 号

策划编辑: 刘 皎

责任编辑: 刘 皎

特约编辑: 顾慧芳

印 刷: 北京季蜂印刷有限公司

装 订: 北京季蜂印刷有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 720×1000 1/16 印张: 19 字数: 364.8 千字

版 次: 2017 年 6 月第 1 版

印 次: 2017 年 6 月第 2 次印刷

定 价: 79.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: 010-51260888-819, faq@phei.com.cn。



目 录

1	深度学习简介	1
1.1	深度学习介绍.....	1
1.2	深度学习的趋势.....	7
1.3	参考资料.....	10
2	TensorFlow 系统介绍	12
2.1	TensorFlow 诞生的动机.....	12
2.2	TensorFlow 系统简介.....	14
2.3	TensorFlow 基础概念.....	16
2.3.1	计算图.....	16
2.3.2	Session 会话.....	18
2.4	系统架构.....	19
2.5	源码结构.....	21
2.5.1	后端执行引擎.....	22
2.5.2	前端语言接口.....	24
2.6	小结.....	24
2.7	参考资料.....	25
3	Hello TensorFlow	26
3.1	环境准备.....	26

3.1.1	Mac OS 安装	27
3.1.2	Linux GPU 服务器安装	28
3.1.3	常用 Python 库	32
3.2	Titanic 题目实战	34
3.2.1	Kaggle 平台介绍	34
3.2.2	Titanic 题目介绍	35
3.2.3	数据读入及预处理	38
3.2.4	构建计算图	40
3.2.5	构建训练迭代过程	44
3.2.6	执行训练	46
3.2.7	存储和加载模型参数	47
3.2.8	预测测试数据结果	50
3.3	数据挖掘的技巧	51
3.3.1	数据可视化	52
3.3.2	特征工程	54
3.3.3	多种算法模型	57
3.4	TensorBoard 可视化	58
3.4.1	记录事件数据	58
3.4.2	启动 TensorBoard 服务	60
3.5	数据读取	62
3.5.1	数据文件格式	63
3.5.2	TFRecord	63
3.6	SkFlow、TFLearn 与 TF-Slim	67
3.7	小结	69
3.8	参考资料	69

4

CNN “看懂” 世界	71
-------------------	----

4.1	图像识别的难题	72
4.2	CNNs 的基本原理	74
4.2.1	卷积的数学意义	75
4.2.2	卷积滤波	77
4.2.3	CNNs 中的卷积层	81
4.2.4	池化 (Pooling)	83
4.2.5	ReLU	84

4.2.6	多层卷积.....	86
4.2.7	Dropout.....	86
4.3	经典 CNN 模型.....	87
4.3.1	AlexNet.....	88
4.3.2	VGGNets.....	95
4.3.3	GoogLeNet & Inception.....	98
4.3.4	ResNets.....	106
4.4	图像风格转换.....	109
4.4.1	量化的风格.....	109
4.4.2	风格的滤镜.....	116
4.5	小结.....	120
4.6	参考资料.....	121

5

RNN “能说会道”	123
------------	-----

5.1	文本理解和文本生成问题.....	124
5.2	标准 RNN 模型.....	128
5.2.1	RNN 模型介绍.....	128
5.2.2	BPTT 算法.....	130
5.2.3	灵活的 RNN 结构.....	132
5.2.4	TensorFlow 实现正弦序列预测.....	135
5.3	LSTM 模型.....	138
5.3.1	长期依赖的难题.....	138
5.3.2	LSTM 基本原理.....	139
5.3.3	TensorFlow 构建 LSTM 模型.....	142
5.4	更多 RNN 的变体.....	144
5.5	语言模型.....	146
5.5.1	NGram 语言模型.....	146
5.5.2	神经网络语言模型.....	148
5.5.3	循环神经网络语言模型.....	150
5.5.4	语言模型也能写代码.....	152
5.5.5	改进方向.....	163
5.6	对话机器人.....	164
5.6.1	对话机器人的发展.....	165
5.6.2	基于 seq2seq 的对话机器人.....	169

5.7 小结.....	181
5.8 参考资料.....	182

6

CNN+LSTM 看图说话	183
---------------	-----

6.1 CNN+LSTM 网络模型与图像检测问题.....	184
6.1.1 OverFeat 和 Faster R-CNN 图像检测算法介绍.....	185
6.1.2 遮挡目标图像检测方法.....	187
6.1.3 ReInspect 算法实现及模块说明.....	188
6.1.4 ReInspect 算法的实验数据与结论.....	204
6.2 CNN+LSTM 网络模型与图像摘要问题.....	207
6.2.1 图像摘要问题.....	208
6.2.2 NIC 图像摘要生成算法.....	209
6.2.3 NIC 图像摘要生成算法实现说明.....	214
6.2.4 NIC 算法的实验数据与结论.....	243
6.3 小结.....	249
6.4 参考资料.....	250

7

损失函数与优化算法	253
-----------	-----

7.1 目标函数优化策略.....	254
7.1.1 梯度下降算法.....	254
7.1.2 RMSProp 优化算法.....	256
7.1.3 Adam 优化算法.....	257
7.1.4 目标函数优化算法小结.....	258
7.2 类别采样 (Candidate Sampling) 损失函数.....	259
7.2.1 softmax 类别采样损失函数.....	261
7.2.2 噪声对比估计类别采样损失函数.....	281
7.2.3 负样本估计类别采样损失函数.....	286
7.2.4 类别采样 logistic 损失函数.....	286
7.3 小结.....	287
7.4 参考资料.....	288

1

深度学习简介

1.1 深度学习介绍

深度学习是目前机器学习学科发展最蓬勃的分支,也是整个人工智能领域中应用前景最为广阔的技术。在现如今的生活,不管是在 iPhone 上随手调戏 Siri,还是看着 AlphaGo 赢得围棋世界第一的宝座,都让人们真真切切地感受到人工智能已经不再是停留在科幻小说中的幻想,深度学习的时代已经到来了!

人工智能 (Artificial Intelligence, AI) 是计算机科学中的一个分支学科,早在 20 世纪 50 年代就被提出和确立了。著名的“图灵测试”是 AI 发展的终极目标,如果某种机器运行的逻辑程序可以表现出与人类等价或者无法分辨的智能,则认为机器有了思维,能够进行思考。从实用的角度讲, AI 的目标是要让计算机系统能够自动完成那些需要依靠人类智慧才能完成的工作。

在 AI 发展的早期阶段,随着计算机自动化所取得的成功, AI 的主要方法和思路是将人类总结的知识用一系列规范的、形式化的数学规则来表示,然后通过自动化的程序代替人类处理问题。以知识为基础的专家系统 (knowledge-based expert system) 就是这方面的典型代表,它将某个领域中人类专家的经验通过知识表示方法写成一条

一条规则，系统依照规则推理模拟专家的思维方式。不过，在实际应用中，专家系统都没有取得太大的成功，其最主要的局限性体现在系统明显受到规则数量的限制，规则数量决定了系统对不同情况的适应程度，然而规则是有限的，问题发生时的状况是无限的，用有限的规则处理无限的可能，注定是苦海无涯。

早期在 AI 方面取得成功的项目，多数解决的是具有明确规则和条件的问题，比如西洋跳棋。1997 年 IBM 的“深蓝”计算机在国际象棋上战胜人类世界冠军卡斯帕罗夫就是这个方面最著名的例子。对于人类来说，下象棋当然是很有挑战的项目，但是相比真实世界的复杂程度而言，国际象棋其实只是一个简单问题。棋盘上只有 32 个棋子和 64 个可以落子的位置，走法规则是非常明确的，所有可能的局面组合是有限的，可以被穷举出来，利用计算机的计算能力辅助以启发式搜索等算法，在摩尔定律的作用下，击败人类只是时间问题。在这类问题中，问题的表示通常都不是难题，一个普通程序员也可以在很短时间内完成一个象棋程序。然而许多真实世界的问题却并不都是那么容易能用计算机语言表达清楚的，比如图像识别和语音识别，这些问题有着比国际象棋大得多的问题域，即使对于人类来说，也有很多不能确定、无法选择的时刻，所以用规则来描述问题是不现实的。

因此，“演绎法”的规则推理暂时行不通，“归纳法”就成为了唯一的出路。基于概率统计的机器学习（**machine learning**）逐渐成为人工智能的主流方法。与专家系统不同，机器学习不会在系统中输入任何规则，而是直接在大量真实世界产生的数据中挑选最具有代表性的**样本（samples）**交给算法处理，让算法自动在数据中寻找和学习特定的规律，而这些由数据得来的规律，就是我们本来需要输入的规则。这种从数据中学习规律的过程也叫做**模式识别（pattern recognition）**。

机器学习的基本思路是假设样本数据与真实世界的概率分布相同，这样就可以认为算法从样本数据中归纳所得的规律在一般情况下同样适用。朴素贝叶斯（**naive bayes**）和逻辑回归（**logistic regression**）等算法都是机器学习的经典方法，并且都在实际应用中取得了很好的效果。

机器学习一般分为监督学习和无监督学习两种。监督学习要求每条样本数据都有对应的**标签（label）**，样本数据作为输入，标签作为目标输出，学习的目标是求出输入与输出之间的关系函数 $y = f(x)$ ，使得针对每个输入样本 x ，都得到期望的输出结

果 y 。朴素贝叶斯、逻辑回归和神经网络等都属于监督学习的方法。而对于无监督学习来说，样本数据并没有标签，学习的目标是为了探索样本数据之间是否有隐含的不易被发现的关系，统计样本的分布情况，典型的算法有 **K-means** 等各种聚类算法。

监督学习主要解决两类核心问题，即回归 (**regression**) 和分类 (**classification**)。回归和分类的区别在于强调一个是连续的，一个是离散的。回归的输出可以是任意实数，而分类的目标输出为离散的类别编号，或是布尔类型的二值判断。分类问题可以利用概率模型以回归方式来求解，即认为样本所属的真实类别概率为 100%，样本属于其他类别的概率为 0%。通过将离散的类别编号转化为连续的概率，利用回归方法学习和预测样本属于每个类别的概率，概率最大的类别就是分类结果。

对于朴素贝叶斯和逻辑回归等简单的机器学习算法来说，本质上是要计算样本输入与目标输出之间的相关性。相关性固然是非常重要的，但在处理真实世界问题的时候，判断相关性其实是人们确定了所有影响因素之后的一个后续问题，各个影响因素的表示 (**representation**) 会严重影响对于相关性的判断。比如要辨别一段语音中演讲者是男人、女人还是小孩，简单机器学习算法的判断依据很可能主要来自音量而不是音色和音调。再比如，假设我们想用逻辑回归判断明年北京房价是否继续上涨，如果选定人口净流入数和地铁修建的里程数作为输入的话，会得到人口流入越多房价越高的结论，那么如果明年人口净流入减少，则系统会预测房价下跌。但真实情况是房价与货币增发量相关性更大，如果明年货币继续超发，最终房价还是会上涨，所以预测就可能会出现比较大的误差。

人口流入数、货币增发量这些对于预测房价有影响的因素，是我们从众多维度中提取的**特征 (feature)**。正如在前文例子中描述的那样，设计合适的特征表示在机器学习中是一项极其重要却又非常困难的工作。一方面，特征选取会直接影响预测的稳定性，要得到准确的预测结果就必须选中相关度最高的特征。在使用逻辑回归、朴素贝叶斯等简单机器学习算法的时候，由于特征选取问题而导致模型失效的情况比比皆是。但另一方面，如何设计特征又需要加入许多人类的先验经验才能完成，需要运用人类的智慧和经验分析各种因素所带来的直接或间接的影响。所以通常的做法是首先列举出各种可能的特征，然后通过交叉组合的方式进行穷举验证。在很长一段时间，**特征工程 (feature engineering)** 都是机器学习的重中之重，是每个研究员的必修技能。特征工程对于传统机器学习算法来说是如此重要，但同时也是机器学习应用的最

大束缚,不仅费时费力,还需要由人类提供大量的先验经验以弥补对数据本身挖掘不足的缺陷。若要拓展机器学习的适用范围,必须要降低学习算法对特征工程的依赖性。

深度学习 (Deep Learning, 也曾被叫作 Feature Learning) 是在人工神经网络 (Artificial neural network, ANN) 基础上发展而来的一种表示学习 (Representation Learning) 方法,也是一种机器学习方法,而且是人工智能领域最具发展前景的一个分支,如图 1-1 所示。其主要模型是各种深度神经网络 (Deep Neural Network)。表示学习是近年来机器学习领域发展最迅猛、最受学术界追捧的方向。所谓表示学习,就是要让算法在少量人为先验经验的情况下,能够自动从数据中抽取出合适的特征,完成原本需要通过特征工程才能得到的结果。在表示学习范畴中,深度学习是通过多层非线性变换的组合方式,得到更抽象也更有效的特征表示。原先使用机器学习解决问题的主要工作包括需要大量人工处理的特征工程加上一个可被训练的分类器,而到了深度学习的时代,特征工程已经被各种可训练的特征提取器所取代,在应用效率上有了显著提高。更重要的是,人工智能目标就是让机器有能力理解我们所在的世界,只有当它能学会如何感知和辨别数据背后的各种隐含因素的时候才能达到这个目标。

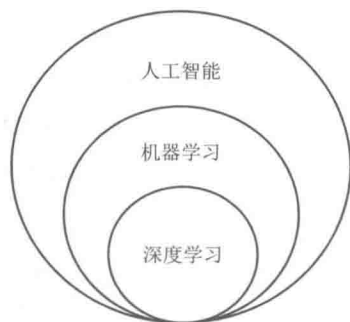


图 1-1 深度学习是一种机器学习方法,是人工智能领域最具前景的分支

真实世界的问题之所以困难,很大程度上是因为我们能够观察到的数据都是种种因素叠加而成的。比如同一朵花在白天和夜晚观察的颜色是不同的,同一辆车在不同角度观察的形状也不一样。表示学习最核心的诉求就是要发现真正重要的特征,舍弃那些并不影响判断的因素。要做到这种程度当然不是一件容易的事,那么深度学习又是如何做到呢?

深度学习是通过构建一个多层的表示学习结构,使用一系列非线性变换操作把从

原始数据中提取的简单的特征进行组合，从而得到更高层、更抽象的表示。在图像识别的场景中，图像在计算机中最基本的表示是一组像素值集合，从像素到物体的映射关系需要经过一个很长的过程，从像素组成细小的边，由边组成基础的纹理基元，纹理基元组合而成图形，图形构成物体的各种组成部分，最后组成物体的整体。同样，对于文本的理解也符合类似的过程，先认识各个字母，再由字母组成单词，单词组成词组，词组组成句子，句子组成段落，段落构成完整的故事。这个过程对于人类来说眨眼之间就已经完成了，但是对于计算机来说是非常复杂的，很难简单直接地一步求得这种映射关系。因此，深度学习模型的结构设计遵循了这种思路，具体做法是将一系列相对简单的非线性映射操作构建成一个多层网络，每一层（layer）都完成一次特征变换。以人脸识别为例，网络以像素表示的图像作为输入，在低级层次中主要学习到代表图像边缘的特征，可能是连续几个像素所组成的某个方向上的线段。中级层次会学习到由边缘线段所组成的局部图案，这些图案实际上是构成目标物体的各种部件，比如眼睛、鼻子、耳朵。在最后的高级层次中，以各种局部部件作为基本单元就可以组合出人脸的抽象表示，比如包括人脸会上会有一个鼻子两只眼睛、眼睛的相对位置在鼻子的两侧，等等。而符合这种抽象表示的图像，就可以被判定为人脸图片。如此多层学习结构中的中间特征如图 1-2 所示。

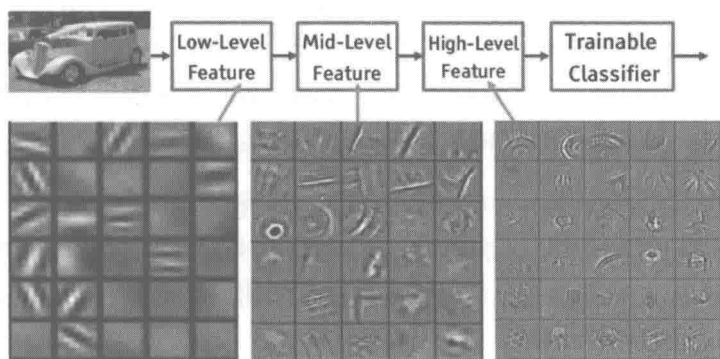


图 1-2 多层学习结构中的中间特征

那什么样的结构才算是“深度”？其实包含三个以上隐层（hidden layer）的神经网络就可以说是一种深度学习模型，但由于在真实场景使用时参数过多、计算量过大，存在梯度消失和梯度爆炸的问题，无法做到稳定收敛，所以一般不会使用。一个网络的深度，可以以网络中串联的计算的层数，或者是非线性变换次数，甚至更加抽

象一些，以不同的计算概念来评估。关键在于，深度学习比传统机器学习模型多了多级特征提取的结构，能够进行表示学习，这样才算是有深度。从这个意义上说，只有两个隐层的神经网络就不能算是有深度的结构，因为它并没有能力分级提取特征。

了解人工神经网络知识的读者可能会有一个疑问，具有一个隐层的人工神经网络已经非常强大了，理论上来说，只要神经元足够多，构成一个很“宽”的网络，它可以拟合任意的函数，那为什么还要更“深”？原因在于：在神经元数量相同的情况下，深层网络结构具有更大的容量，分层组合带来的是指数级的表达空间，能够组合成更多不同类型的子结构，这样可以更容易地学习和表示各种特征。并且，隐层增加则意味着由激活函数（activation function）带来的非线性变换的嵌套层数更多，就能构造更复杂的映射关系。

现如今，深度学习已经在多种应用上取得了突破性进展。卷积神经网络（Convolutional Neural Networks, CNNs）是这一波深度学习浪潮的引领者。2012年 AlexNet 在 ILSVRC 图像识别竞赛中所带来的惊人表现，让学术界看到了深度学习所蕴含的巨大潜力。在随后的几年中，随着 GoogLeNet、VGGNets、ResNets 等模型的提出，CNNs 的识别准确率持续提高，让计算机拥有了超越人类的图像识别能力。关于 CNNs 的应用将在本书第 4 章中详细介绍。同时，深度学习在自然语言处理（Natural Language Processing, NLP）方面同样取得了巨大的成功，不但有 Siri 这样可以与人类正常交流的对话机器人，甚至能够写诗、作曲。本书第 5 章会介绍循环神经网络（Recurrent Neural Networks, RNNs）在处理语音或者文字等问题中的应用。

总结来说，深度学习是一种机器学习方法，同时也是目前最有希望具有处理复杂的真实世界问题的能力的人工智能方法。深度学习的分层结构能够从简单概念的组合中学习到高级抽象的特征表示，让计算机真正具有理解世界的能力。不同类型人工智能系统的流程图如图 1-3 所示。

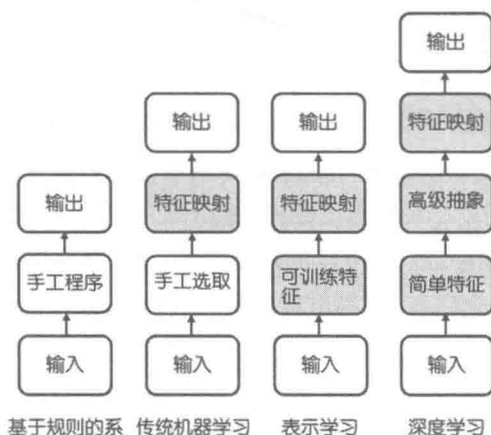


图 1-3 不同类型人工智能系统的流程图。灰色框表示具有从数据中学习的能力

1.2 深度学习的趋势

20 世纪 90 年代，在卷积神经网络应用于文字识别之后，深度学习曾经经历一段时间的沉寂，而 2012 年在图像识别领域的爆发式突破，让人们意识到其巨大的潜力和广阔的应用前景。现如今，以 Google、Facebook、Nvidia 等为首的科技公司全都重兵押注深度学习，而国内的百度、阿里巴巴、腾讯也紧随其后，让人深深感觉到这是一个朝气蓬勃的领域，同时也是一个硝烟弥漫的战场。

很多人会好奇，人工神经网络几十年前就已经存在了，为什么直到今天才认为深度学习是一项突破性的技术？事实上，在 20 世纪 90 年代深度学习就已经有成功的商业应用，但是在那时，人们更多会认为这是一门只有少数几个专家才能掌握的艺术而不是技术。深度学习的成功，依赖于四项基本要素：

- 海量的训练数据；
- 非常灵活的模型；
- 足够的运算能力；
- 足够对抗维度灾难（curse of dimensionality）的先验经验。

机器学习是要从数据中学习知识，正所谓“巧妇难为无米之炊”，即使再厉害的算法，也需要优质的数据集支持。所以在人工智能几十年的发展过程中，数据集的种

类和规模都在逐渐增长。图 1-4 展示了一些经典数据集规模的对比。从中可以看到，样本的规模几乎是在以指数级增长，其中一些著名的数据集（如 ImageNet）的样本数量已经达到了千万级别。原先，机器学习的任务是，要从少量数据中学习特定的规律，然后泛化到更一般的场景中，并且要防止各种欠拟合和过拟合。但随着大数据时代的到来，数据集本身就已经包含了各种可能出现的情况，从数据集中学习的规则可以直接应用在现实场景中。以 2016 年的趋势来看，对于有监督的深度学习算法来说，只要每个类别有 5000 个样本，机器就能达到令人满意的表现，而若有 1000 万以上的样本，机器就能达到甚至超越人类的水平。在互联网已经深入人们生活每一个角落的今天，数据更是会越来越多，各种“独角兽”公司每天产生的数据量是 PB 级别的。当收集数据已经不再是难题，如何更有效地利用数据就成为了新的挑战。

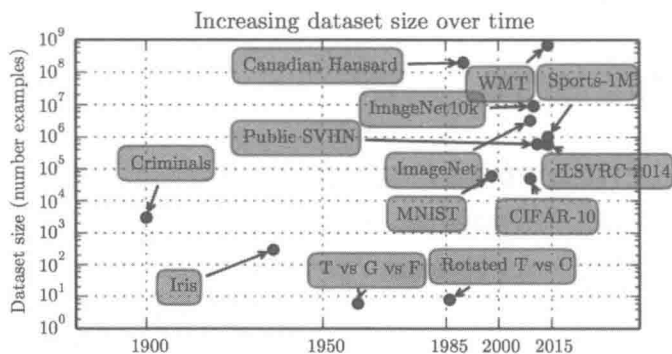


图 1-4 在人工智能学科刚被确立的时候，研究的数据仅在几百条的规模。到了 20 世纪 90 年代，著名的 MNIST 手写数字数据集有 60000 多张 28×28 分辨率的黑白图片。而到了近些年，ImageNet 数据集超过了 100 万张大尺寸彩色照片

深度学习涉及大量的科学计算。计算机硬件的性能提升也为深度学习发展推广提供了基本必要条件。著名 CPU 厂商英特尔（Intel）的创始人之一戈登·摩尔（Gordon Moore）曾提出摩尔定律，其内容为：当价格不变时，集成电路上可容纳的元器件的数目，约每隔 18~24 个月便会增加一倍，性能也将提升一倍。摩尔定律概括了 20 世纪末一段时间 CPU 的发展速度。到 2017 年，家用电脑 CPU 计算频率已经接近当前技术的上限。因此，除了靠提升单个 CPU 的计算频率来提升计算速度外，人们还想到使用多核或多 CPU 集成的方式来提升计算性能。CPU 能够快速处理多种类型的计算，但并不是每种类型的计算耗时都相同的（例如双精度浮点数乘、除法和条件判

断语句)。因此多 CPU 集成, 同样受到频率同步和稳定性的限制。我们可以看到, 市面上好几万元的服务器品牌的 CPU 核心往往是 16~64 个, 而主频一般都在 3.0GHz 以下。而几千元的 PC 主频虽然可以达到 3.0GHz, 而 CPU 核心一般是 2~8 个。

考虑到深度学习的特殊计算需求和 CPU 综合计算性能方面的这些限制, 人们考虑使用 GPU 来解决科学计算问题。20 世纪 90 年代, PC 电脑发展初期 GPU 的主要功能是用于显示交互界面。其画面的显示功能涉及坐标点变换、栅格化、平面渲染等大量的浮点数加、减、乘、除等计算操作。对比来说, CPU 是面向通用计算的产品, 而 GPU 则是天生面向大规模浮点数并行计算的。与 CPU 相比, GPU 的内置计算核心虽然计算频率相对较低, 但内置计算核心数量更多。以著名 GPU 厂商英伟达 (Nvidia) 为例, 近几年推出的 Tesla、Titan X 系列显卡处理器数量都以上百、上千计, 而计算主频在 1GHz 左右。在 GPU 的帮助下, 神经网络的训练效率大幅提高。与同等价位 CPU 相比, 速度提升了约几十倍, 分析大型的深度网络结构的效率因此得到了提升。英伟达公司 Tesla 系列显卡与 CPU 芯片的计算能力比较如图 1-5 所示。

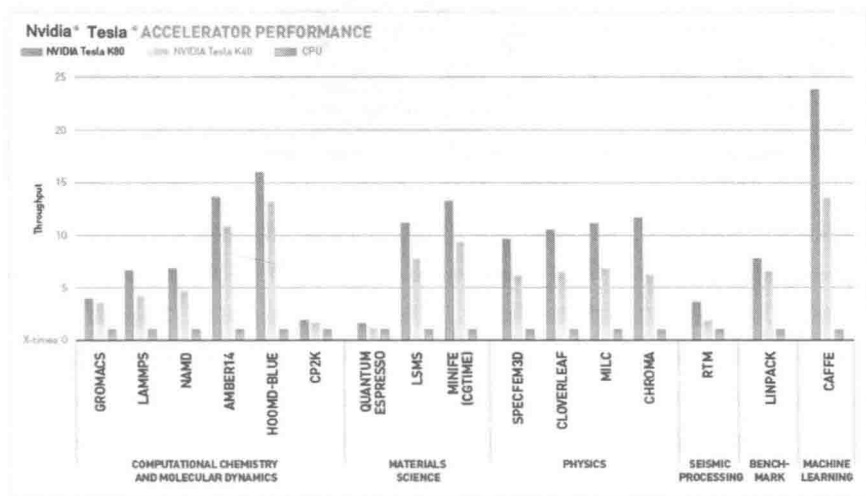


图 1-5 Tesla 系列显卡与 CPU 芯片的算力比较。在科学计算方面 GPU 显卡是 CPU 速度的数倍

随着计算能力的提升和更大规模数据集的出现, 神经网络模型不但结构越来越复杂, 而且规模也在不断增大。尤其自从隐层结构提出以来, 神经网络的神经元数量每 2.4 年翻一番, 如图 1-6 所示。对于深度学习来说, 人类赋予其最重要的一个先验经验就是: 分层组合的表示方式更能准确地描述我们所在世界的规则。因此, 深度学习

模型发展出更深的结构、更大的规模也是必然趋势。更何况实验已经证明，更大的模型能够达到更好的表现。以图像识别为例，拥有 8 个参数层的 AlexNet 在 ImageNet 大规模视觉识别挑战 (ILSVRC) 中将图像分类的错误率降低到 15.3%，而有 19 个参数层的 VGGNet 的分类错误率只有 7.3%，现今最大、最深的 ResNet 有超过 150 个参数层，分类错误率仅为惊人的 3.57%，这已经是超越人类能力的成绩。

可以预见的是，深度学习在未来将能处理越来越复杂的任务，也注定会在我们日常生活中扮演越来越重要的角色。

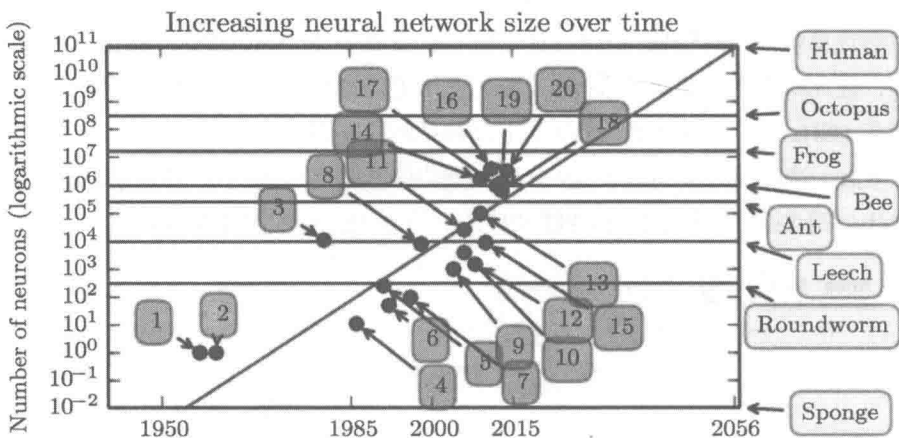


图 1-6 自从隐层结构提出以来，神经网络的神经元数量每 2.4 年翻一番。图中各编号代表的分别为：1. 感知机 (Perceptron)；2. 自适应线性单元 (Adaptive linear element)；3. 神经认知机 (neocognitron)；4. 早期反向传播网络；5. 用于语音识别的递归神经网络；6. 用于语音识别的多层感知机；7. sigmoid 置信网络；8. LeNet-5；9. 回声状态网络 (Echo state Network)；10. 深度置信网络 (Deep belief network)；11. GPU 加速的卷积网络；12. 深度玻尔兹曼机 (Deep Boltzmann machine)；13. GPU 加速的深度置信网络；14. 非监督卷积网络；15. GPU 加速的多层感知机；16. OMP-1 网络；17. 分布式自动编码器 (Distributed autoencoder)；18. 多 GPU 加速的卷积网络；19. COTS HPC 非监督卷积网络；20. GoogLeNet

1.3 参考资料

[1] 本部分内容参考并翻译自：Ian Goodfellow and Yoshua Bengio and Aaron Courville. "Deep learning." An MIT Press book in preparation. Draft chapters