编者按：人工智能(artificial intelligence)是人类最美好的梦想之一，图灵测试给人工智能提出了很高的期望，即隔墙对话的双方并不知道对方是人还是计算机。机器学习(machine learning)作为人工智能的一个重要分支，目标是使系统具备从原始数据中提取特征，从而获得自主学习的能力。从 20 世纪 80 年代以来，机器学习在算法、理论和实际应用方面都取得了巨大的成功，特别是浅层学习(shallow learning)模型在互联网领域得到了广泛应用，如搜索广告系统的广告点击率预估、网页搜索排序、垃圾邮件过滤、基于内容的推荐系统等。

随着大数据时代的来临，数据在 Volumn、Variety、Velocity 和 Value 上的高维特性给浅层学习带来了新的技术挑战。一方面，大数据里含有的丰富的信息维度使得浅层模型处于欠拟合的状态，“大数据＋浅层模型”不能深刻揭示海量数据里所承载的特征结构，从而无法充分发掘出更多有价值的信息和知识。另一方面，浅层模型强调模型主要负责分类或预测，样本的特征是靠人工经验来抽取，在模型选择正确的前提下，特征的好坏就成为整个系统性能的瓶颈。而发现一个好的特征需要许多经验和技巧，要求开发人员反复的摸索，对待解决的问题有着深入的理解，因此，人工确定样本特征不是一个可广泛扩展的途径。

2006 年，机器学习领域专家 Hinton 等人提出的深度学习(deep learning)是解决上述挑战的有力技术。深度学习模仿人脑的分层模型结构，通过构建具有很多隐藏层的深层模型，对输入数据逐级提取从底层到高层的特征，从而能很好的建立从底层信号到高层含义的映射关系，在海量训练数据中学习有用的特征，最终提升分类或预测的准确性。区别于浅层学习，深度学习不仅强调模型结构的深度，还明确突出了特征学习的重要性，通过逐层映射变换，将样本在原空间的特征表示变换到一个新的特征空间，更有利于分类或预测。

近年来，Google、Microsoft、IBM、百度等知名企业纷纷投入大量人力物力资源进行深度学习技术研发，在语音、图像、自然语言、在线广告等领域取得显著进展，开启了深度学习在学术界和工业界的浪潮。鉴于此，本刊组织了本次系列讲座，旨在把握其发展动向，激发新的研究灵感。本次讲座以深度学习理论和研究新进展为主题，具体包括深度学习的理论框架、训练过程、常用模型以及深度学习在语音、图像等领域的应用等。

深度学习理论及其应用专题讲座(一)

第 1 讲　深度学习：开启人工智能的新纪元*

李轶南[1]，张雄伟[2]，李治中[3]，吴海佳[4]，孙久皓[5]

(1. 解放军理工大学指挥信息系统学院研究生 1 队，江苏 南京 210007；2. 解放军理工大学指挥信息系统学院；
3. 解放军理工大学野战工程学院；4. 解放军后勤工程学院，重庆 401311；
5. 中国人民解放军 65040 部队，辽宁 沈阳 110005)

摘　要：深度学习是近年来在人工智能领域颇具有影响和的代表性技术之一。应用该技术所取得的前所未有的突破引领了学术界和工业界的新的发展浪潮。百度、谷歌、微软等一大批国内外高科技公司纷纷注入大量资源对深度学习进行研发，并在语音识别，图像处理，机器翻译等方面获得了显著地进展。肇始于 2006 年的深度学习技术被誉为最近十年内机器学习领域最成功的研究方向。它的出现为处理如今大量增长的海量数据，提供了一

* 收稿日期：2015-10-09；修回日期：2015-10-19
基金项目：国家自然科学基金资助项目(61402519，61471394)、江苏省自然科学基金资助项目(BK20140071，BK20140074，BK2012510)
作者简介：李轶南(1988－)，男，博士生.

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

种新的出路和可能。文章首先探讨了深度学习理论提出的基本动机，回顾了深度学习的发展历程，总结了深度学习在各个领域的应用及进展，最后指出深度学习存在的问题并对未来进行展望。

关键词：人工智能；机器学习；深度学习；语音识别；图像处理

中图分类号：TP18　　**文献标识码**：A　　**DOI**：10.16464/j.cnki.cn32-1289.2015.04.0019

Deep Learning：Path to New Era of Artificial Intelligence

LI Yi-nan[1], *ZHANG Xiong-wei*[2], *LI Zhi-zhong*[3], *WU Hai-jia*[4], *SUN Jiu-hao*[5]

(1. Postgradute Team1 CCIS, PLAUST, Nanjing 210007, China;
2. College of Command Information System, PLAUST; 3. College of Field Engineering, PLAUST;
4. Chongqing Logistical Engineering University, Chongqing 401311, China;
5. Unit 65040 of PLA, Shenyang 110005, China)

Abstract: Deep Learning is one of the most influential and representative technologies of artificial intelligence in recent years. The unprecedented breakthrough in the application of deep learning has led the new trend of development in academic research and industry application. A large number of domestic and foreign high-tech companies such as Baidu, Google and Microsoft have invested a large amount of resources into the R&D of deep learning, leading to significant progress of speech recognition, image progressing, machine translation and other aspects. Deep learning technology since 2006 has been considered as the most successful progress of the machine learning society in the past 10 years. The emergence of deep learning perhaps provides a new paradigm for processing the ever increasing big data. Firstly, the basic motivation of deep learning theory was discussed. Then, the development history of deep learning was reviewed. The application and development of deep learning in various areas were also summarized. Finally, the existing problems of deep learning and the prospect of deep learning were pointed out.

Key words: artificial intelligence; machine learning; deep learning; speech recognition, image processing

大数据时代，以移动互联网为代表的新兴科技革命使得整个人类社会的数据量呈现出爆炸式的增长态势。知名数据分析公司 Domo 在 2015 年 8 月的最新统计数据[1]显示，从 2013 年到现在，全球使用互联网的人口从 24 亿增长到 32 亿，增幅近 20%。目前，在短短的一分钟时间内，平均会新产生超过 300 小时的 YouTube 视频，收到超过 416 万次的 Facebook 点赞，多达 4 310 次的 Amazon 访问量，提供超过 7 万小时的 Netflix 流式视频，下载多达 51 000 个苹果的应用程序，以及超过 11 万次的 Skype 电话拨打量。这些数据从一个侧面刻画揭示了规模空前的数据涌现量。这些以越来越快的速度奔涌而来的大数据正深刻而根本地影响和改变着我们日常生活的方方面面。一方面，大数据给数据的存储和分析造成了空前的压力与困难，其中只有极小比例的数据得以被分析和利用，传统的人工智能算法越来越难以胜任规模空前的海量数据处理任务；另一方面，低成本地获取大量多源、非结构化廉价数据又能更好地帮助我们寻找出潜在的规律和模式，使得我们得以从更高的层次，更广的视野，更多的维度来审视错综复杂的客观世界，从而更科学合理地做出分析与决策。如何扬长避短，实现对大数据的充分合理利用是亟需思考和解决的问题，深度学习正是在这样的背景下应运而生。鉴于它对于学术界和工业界带来的革命性影响，2014 年《麻省理工学院技术评论》(MIT Technology Review)中将其列为 2013 年十大突破性技术之首[2]。

深度学习模拟大脑的分层结构，研究如何从海量数据中自动地获取具有层次性的多层特征表达。该技术在本质上是一种数据驱动的人工智能方法，它采用一系列的非线性变换，构建起具有多个隐层的学习模型，从而从原始数据中提取出由低层到高层逐层抽象的特征表达，实现提升分类或预测准确性的最终目的。

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

本文将分深度学习的动机、深度学习的发展历程、深度学习技术的实际应用以及深度学习的未来发展方向几个方面对深度学习进行介绍。

1 深度学习的动机

机器学习是人工智能的基础，其研究问题的核心和本质是如何利用计算机来模拟或仿照人类的学习行为。20 世纪 80 年代末以来，随着计算能力的大幅提升，机器学习的相关技术得到了蓬勃的发展。按照学习模型的层次结构来对机器学习的发展阶段进行划分，可以大致将其分为浅层学习(shallow learning)和深度学习(deep learning)两个阶段。

浅层结构模型通常包含不超过一层或两层的非线性特征变换[3]。20 世纪 90 年代，各种各样的浅层模型相继涌现，诸如高斯混合模型 GMM(Gaussian Mixture Model)、支持向量机 SVM(Support Vector Machine)、提升方法(boosting)、最大熵方法(比如代表性的 logistic regression)等，具有非常好的理论支撑和实际应用效果，因此获得了巨大的成功。通常来说，浅层学习可以概括为图 1 的几个步骤。

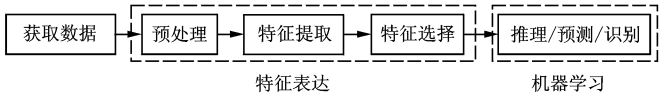
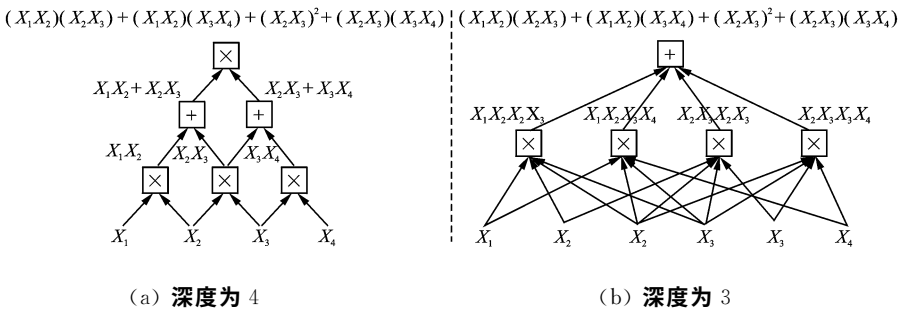


图 1 浅层学习的基本步骤

图 1 中预处理、特征提取、特征选择这三个部分概括起来就是特征表达。特征表达的优劣将直接影响浅层学习的最终性能。然而，特征的构造和选择往往需要非常强的专业知识背景和极高的技巧，找到理想的特征在很大程度上依靠的是经验和运气，且调节特征的参数也需要花费大量的时间。在处理大量样本数据时，浅层学习的特征表达往往会遇到意想不到的问题，因此亟需更具普适性的替代方法，来自行实现特征的抽取与表达，此时，沉寂许久的深度学习方法重新回归到学者们的视野。

特征表达本质上是对输入数据的一种函数映射，函数映射可以通过流向图表示。所谓流向图是指一种可以表示计算的图模型，在该模型中，结点间的连接表示数据流向，结点则表示一种基本运算。流向图的一个重要属性是深度，即从输入到输出的最长路径。如图 2 所示为一个多项式函数的流向图，最底层为输入层，从下往上，偶数层为乘法运算，奇数层为加法运算，其中图 2(a)的深度为 4，图 2(b)的深度为 3。从图中可以看出，对于同样的函数，图 2(a)通过复用中间结果，只需要进行 6 次基本运算(底层乘法运算每个节点需要 1 次基本运算，3 个节点共需要 3 次基本计算；加法运算两个节点各需 1 次基本运算，共 2 次基本运算；顶层乘法需要 1 次基本运算；合计 6 次基本运算)，而图 2(b)却需要进行 15 次基本运算(底层的乘法运算需要计算 4 个输入数据的乘积，每个节点共需 3 次基本运算，4 个乘法器需要 12 次基本运算；顶层的加法运算计算 4 个输入数据之和，需要 3 次基本运算，共计 15 次基本运算)。

理论研究证明，若利用 K 层流向图可简洁地表示某函数，则通过 $K-1$ 层流向图至少需要指数数量级的参数才能进行表示，在实际应用中难以实现，且存在泛化能力不足的问题[4]。以上是浅层学习走向深度学习的一个重要动机。



(a) 深度为 4 (b) 深度为 3

图 2 多项式函数的流向图

另一方面，神经生物学家发现，哺乳动物的大脑皮层在处理信息时也具有类似因式分解的分层处理机制[5]，信息从感知器官传入大脑，经过多层神经网络传递，每一层神经元都会识别出特定的特征。例如视觉识别系统识别外界信息时，首先检测边缘信息，然后将边缘信息抽象为基本的形状信息，再逐层上升，成为更复杂的视觉目标信息。这种逐层抽象，依次递进的特征提取方式激发了深度神经网络的构想，学者们尝试利用深度神经网络的分层表达能力和无监督学习能力，自动获取数据的特征。这是深度学习理论提出的另一

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

动机。

相较于浅层学习，当训练样本数据不充足时，深度学习非常容易出现过拟合、泛化能力差之类的问题。然而样本不足的问题随着大数据时代的到来迎刃而解。并且由于模型的容量显著增加，使得其能够通过多级非线性表示从大数据中抽取出蕴含在其中的高度非线性表示和复杂函数表示，从而实现浅层学习难以企及的效果。

2　深度学习的发展历程

神经网络 NN(Neural Network)诞生至今已有半个多世纪，其发展阶段的划分也是一件仁者见仁、智者见智的事情，本文为了回顾深度学习的发展历程，在此处主要分为三个阶段进行简要梳理。

2.1　神经网络的奠基阶段

模拟神经网络的火种发源于 1943 年，美国神经生物学家 Warren Sturgis McCulloch 和数学家 Walter Pitts 合著的神经网络开山之作《神经活动中内含的逻辑计算》中提出一种极为简单的神经元模型神经元的阈值元件模型，简称 M-P 模型。该模型采用数理模型的方法模拟人脑信息处理，将神经元当作一个功能逻辑器件对待，并使用逻辑运行的网络来模拟大脑。

1949 年，神经心理学家 Donald O. Hebb 出版了《行为组织学》一书。该书中提出的“Hebb 学习规则”指出，学习过程最终发生在神经元之间的突触部位，突触的联结强度随着突触前后神经元的活动而变化，变化的量与两个神经元的活性之和成正比。2000 年诺贝尔医学奖获得者 Eric Kandel 通过动物实验证实了 Hebb 规则。Hebb 对神经网络的发展起到了重大的推动作用，其成果至今仍被广泛引用。

神经网络真正引起重视是在 1958 年，Frank Rosenblatt 通过增加学习机制，推广了 M-P 模型，并首次将神经网络理论付诸工程实现，制造了叫做“感知器”的神经网络计算机，并从理论上证明了单层神经网络处理线性可分分类问题的收敛性[6]。在 1962 年，Rosenblatt 总结了他的研究成果，并出版了《神经动力学原理：感知机和大脑机制的理论》[7]一书。他的研究成果引起了美国军方的极大兴趣。美国军方甚至认为神经网络工程比原子弹工程更加重要，因此不惜投入重金，大力资助了 Rosenblatt 的工作，并掀起了神经网络的首次高潮。

人工智能的另一位奠基人 Minsky 则认为感知器无法解决人工智能问题，他从感知器的局限性入手，从数学上证明了不包含隐层单元的感知器无法处理异或这样的非线性分类问题。1969 年 Minsky 与 Papert 教授合著的《感知机：计算几何学》一书宣告了感知器的失败。神经网络的研究也从此堕入低谷，加州理工学院的 Carver Mead 曾用“二十年大饥荒”来描述这段时间。

2.2　神经网络的再度繁荣

神经网络的低迷并没有影响到神经生物学的发展。哈佛大学的神经生物学家 Hubel 和 Wiesel 基于对猫视网膜和视皮层中枢神经细胞的信息处理机制研究提出了“局部感受野”概念[8]揭示了视觉神经的机理并于 1981 年获得诺贝尔医学奖。他们的发现对后来的卷积神经网络有很大的启发意义[9]。神经网络真正意义上的再度繁荣是以 1982 年美国加州理工学院的生物物理学家 John J. Hopfield 向美国科学院提交关于神经网络的报告为标志的。Hopfield 采用新型的全互连神经网络，利用所定义的计算能量函数，成功给出了 TSP(Travelling Salesman Problem)问题的近似解。此后，研究者致力于拓展 Hopfield 模型使之更接近人脑的功能特性。1986 年 D. Rumelhart 和 J. McClelland 合著出版了被后来神经网络研究者奉为圣经的轰动性著作(Parallel Distributed Processing：Exploration in the Microstructures of Cognition)宣告了神经网络研究热潮的再度到来。

然而神经网络的这次热潮并没能持续多久又很快被淹没了，原因在于神经网络往往需要精细管理，其中又充满了诸多令人难以捉摸的窍门和技巧。在机器学习过程中，需要为系统提供大量有标签的训练实例，在处理多层网络时计算开销将会变得巨大且速度非常慢，学习时间难以衡量并且由于优化问题的非凸性而极

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

易陷入局部极小值，从而使系统性能变得极差。与此同时，以 SVM 算法为代表的统计学习理论，以 Boosting 算法为代表的强化学习理论和以图模型(graphic model)为代表的概率推断理论由于具有良好的理论支撑，优秀的性能表现，成功统治了机器学习和人工智能领域。

2.3 走向深层的神经网络

2006 年，加拿大多伦多大学 Geoffrey E. Hinton 教授和他的学生在学术刊物 Science 上发表了一篇论文[10]，他们使用逐层训练的方法克服了传统方法在训练深层次神经网络时遇到的困难，成功实现了数据降维并获得了良好的效果。同时，该方法还具有良好的可视化性能。该工作的诞生掀起了神经网络走向深层的发展浪潮。

Hinton 教授在文献[11]中提出了深度置信网络 DBN(Deep Belief Network)模型，如图 3 所示。该模型使用对比散度 CD(Contrastive Divergence)算法进行高效训练，且在训练过程中可不依赖于监督信息，这些特性使得该模型日渐成为深度学习的主流框架。在该框架中，一个深度神经网络由若干个受限波尔兹曼机 RBM(Restricted Boltzmann Machine)堆叠起来，训练采用由低到高逐层训练的方法进行。单层 RBM 可通过 CD 算法进行快速训练，因此这一框架将深度网络学习问题简化为多个 RBM 的分层训练问题，解决了深度网络训练中的高复杂度问题。

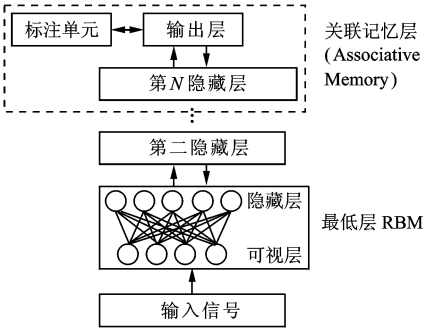


图 3 深度信念网(DBN)

组成 DBN 的核心模块是 RBM，如图 4 所示为一个典型的 RBM，其由 m 个可视结点和 n 个隐藏结点组成，其中可视结点之间是相互独立的，仅与隐藏结点相关，同样隐藏结点之间也是相互独立的，仅与可视结点相关。这一约束简化了对 RBM 的训练。RBM 的隐藏结点为 0/1 随机数，而可视结点的取值有两种情况，分别为 0/1 随机数或者服从高斯分布的随机数，前者称为 Bernoulli-Bernoulli RBM，后者称为 Gaussian-Bernoulli RBM。

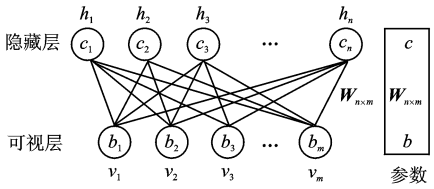


图 4 受限玻尔兹曼机(RBM)

RBM 中有三个参数，$W_{n\times m}$ 是可视层与隐藏层之间的权重矩阵，$b=(b_1, b_2, \cdots, b_m)$ 是可视结点的偏移量，$c=(c_1, c_2, \cdots, c_n)$ 是隐藏结点的偏移量。这三个参数决定了 RBM 如何从一个 m 维的样本中提取 n 维的特征。

除了 DBN 之外，还有很多其他的深度神经网络，文献[12]根据网络的编解码情况(编码器实现输入数据到隐层特征空间的自底向上的映射；解码器则将隐层特征逆向映射，追求与原始输入的尽可能的相似性，需要指出的是，为了防止简单复制重建后的输出，需要增加一定的约束条件)将其划分为前馈深度网络 FFDN(Feed-Forward Deep Network)、反馈神经网络 FBDN(Feed-Back Deep Network)和双向深度网络 BDDN(Bi-Directional Deep Network)3 类，然后又可对深度网络进一步进行划分，如图 5 所示。

FFDN 由多个编码器通过层叠而成，比较有代表性的网络包括多层感知机 MLP(Multi-Layer Perceptron)[13-14]和卷积神经网络 CNN(Convolutional Neural Network)[15]；FBDN 则由多个解码器层叠而成，包括反卷积网络 DN(Deconvolutional Network)[16]、层次稀疏编码网络 HSC(Hierarchical Sparse Coding)[17]等；BDDN 同样具有多个层次，并且每个层次都能实现单独的编码和解码，代表性的有深度玻尔兹曼机(DBM)[18]、栈式自编码器(SAE)[19]等。

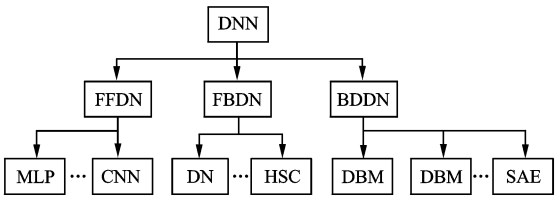


图 5 依照编解码方式划分深度神经网络

神经网络走向深层、各种深度学习算法如雨后春笋般地涌现绝非偶然，这主要得益于三个方面：一是大数据时代的海量数据的到来使得深度神经网络这种大容量复杂模型有了用武之地，大量廉价的训练数据解决了早期深度神经网络由于训练样例不足而常常出现地过拟合、泛化能力差等问题；二是神经网络的训练技

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

巧的大量涌现，包括初值选择、新的非线性映射函数的提出、防止过拟合的随机丢弃技术等；三是计算能力的大幅提升，CPU/GPU 计算能力的大幅提升，多核、集群、大容量内存、并行技术的广泛应用为深度学习提供了硬件平台和技术手段。这些方面的成熟与进展必然助力神经网络走向深层。

3　深度学习的应用进展

2012 年 6 月由斯坦福大学机器学习 Andrew Ng 教授（Ng 是 Berkeley 教授 Michale Jordan 的弟子，并创立了网络公开课程 Coursera）和大规模计算专家 Jeff Dean 共同主导的 Google Brain 项目，将稀疏编码和自编码模型相结合，构建了由 16 000 个 CPU Core 组成的并行计算平台，训练出一种叫做“深度神经网络”DNN（Deep Neural Network）的机器学习模型，该模型包含 10 亿个连接和 9 层神经网络。数据来自 YouTube 上的 1 000 万张 200×200 未标记图片，在无监督地训练一周以后，该模型可以从众多不同的目标中自动识别出猫脸。纽约时报报道了这一进展[20]，这种通过无监督模式使机器形成对猫这一抽象概念的认识引起了业界的轰动和广泛关注[21]。

2012 年 11 月，微软在中国天津的一次活动中公开展示了一套全自动机器同声传译系统。该系统的后台计算机能够连贯地完成了英语语音识别、英中机器翻译和中文语音合成，从而将前台用英文演讲的演讲者所说内容用比较流畅通顺的汉语表达出来。该系统的关键技术同样是 DNN。值得一提的是，微软研究院的语音识别专家 Deng Li 和 Yu Dong 从 2009 年开始与 Hinton 合作，并于 2012 年发表了在语音识别领域颇具颠覆性的研究成果[22]。他们使用大量比音素小很多的被称作 senones 的单元直接对声音进行建模，利用深度神经网络来逐层抽取、发掘出连续多帧语音信号组成的高维特征，替代了在语音识别领域占据垄断地位的 GMM 模型（GMM 通过似然概率来建模，本质上是可视作包含一个隐层的浅层网络），并成功推出了第一个应用于大词汇量语音识别系统的上下文相关的深层神经网络-隐马尔科夫模型（CD-DNN-HMM）。相较于之前最领先的 CD-GMM-HMM 模型，相对误差降低了超过 16%，随后在 Switchboard 标准数据集上的测试结果显示，基准测试字词错误率为 18.5%，相对错误率减少了 33%，是语音识别近 20 年取得的最大突破。此外，深度神经网络在语音合成[23]，机器翻译领域也取得了令人瞩目的成绩。

深度学习在图像处理领域也取得了空前的成就。2012 年底，Geoff Hinton 的博士生 Alex Krizhevsky 和 Ilya Sutskever 首次将卷积神经网络应用于 ImageNet 大规模视觉识别挑战赛 ILSVRC（ImageNet large scale visual recognition challenge）[24]，他们使用一个 8 层的深度卷积神经网络（包括 5 个卷积层和 3 个全连接层，共包含 6 000 万个参数和 65 万个神经元），取得了图像分类和目标定位任务的第 1 名的成绩[25]。其中图像分类任务中，top-5 错误率为 15.3%，远低于第 2 名（使用手动特征设计方法）26.3% 的错误率。在 ILSVRC-2013 上，M. D. Zeiler 采用了去卷积神经网络（deconvolutional network）对中间层特征进行可视化[26]，改进了 Krizhevsky 的方法，将分类错误率进一步降低到 11.2%。在 ILSVRC-2014 比赛中，几乎所有的参赛队都采用了深度卷积神经网络，分类识别错误率降低到 6.7%，这从一个侧面证明了深度神经网络的巨大潜力和潜在性能。

深度学习在人脸识别领域也取得了令人瞩目的成绩，香港中文大学采用 4 层卷积神经网络的 DeepID 项目[27]和 Facebook 采用 5 层卷积神经网络的 DeepFace 项目[28]在 LFW 数据库上人脸识别正确率分别达到 97.45% 和 97.35%，略低于人类 97.5% 的识别正确率。随后的 DeepID2 项目又进一步将识别准确率提升到 99.15%。

在深度学习显示出其巨大发展潜力的同时，各大科技企业纷纷抢夺该方面的人才。Google 竞价超过微软等公司，斥资数千万美元收购了仅仅由 Alex Krizhevsky、Ilya Sutskever 和 Geoff Hinton 3 人组成的刚刚注册几个月的公司。Facebook 则邀请卷积神经网络的创始人 Yann Lecun（此人亦是 Geoff Hinton 的弟子）加入 Facebook AI lab。Yann Lecun 的学生陆续开了几家专注于深度学习的公司，也纷纷被 Twitter、Facebook、Google 所收购。Andrew Ng 离开谷歌，担任百度花费 3 亿美元巨资所筹建的研究实验室的首席科学家。

近年来，深度学习的应用如雨后春笋般涌现，一些有意思的应用也不断涌现出来，2015 年来自马里兰

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

大学和澳大利亚 NICTA 的研究人员通过给机器人观看 YouTube 视频，使其自动转化为机器人的动作指令并加以模仿[29]。系统的底层包括两个卷积神经网络，一个用于分辨手的姿势类型，另一个用于目标识别。高层结构基于视觉序列，用以生成指挥机器人动作的指令。同年，德国科学家则使用深度神经网络学习人类的画风，实现了对于绘画作品的内容与形式分离[30]，可以将绘画作品以某种特定形式的绘画风格表现出来，该工作为理解人类创作艺术品和感知艺术品内在机理提供了新的研究思路。

4 深度学习存在的问题及未来展望

深度学习所取得的史无前例的巨大突破宣告了人工智能的研究迈入了新的纪元，有人甚至预言“大数据＋深度模型”模式将会横扫过去种类繁多的机器学习算法，使其统统成为历史。然而，事实上，深度学习技术并不完美，依然存在诸多问题有待解决。首先是理论研究问题，深度学习所应用的模型都是非凸函数，对其开展理论研究十分困难。使用多少训练样本，调用多少计算资源才能够获得有学习能力的足够好的深度模型是一个难以回答的问题。同时理论支撑的缺乏使得我们难以评价所设计的深度模型是否最优。其次，深度学习众多的参数计算需要消耗大量的计算资源，为了提高计算效率，大规模并行计算势在必行，目前成熟的 DNN 技术大量采用的随机梯度方法难以在多个计算机之间并行，导致训练时间极为漫长，高效的并行算法研究必然是未来深度学习的研究方向。再次，深度学习相较于人脑，规模还不够大，数量级只有人脑的万分之一，发展更大规模更高效率的硬件也是提升识别率的有效方法。最后，我们对于人脑的认知机理还有待进一步研究，历史经验显示对于人类学习认知过程的生理、心理学研究往往能够助力于深度学习的进一步发展。

参考文献：

[1] James Josh. Data never sleeps 3.0[EB/OL]. (2015-08-13) [2015-10-19]. https://www.domo.com/blog/2015/08/data-never-sleeps-3-0/.

[2] MIT Technology Review. 10 breakthrough technologies 2013[EB/OL]. (2013-04-23)[2015-10-19]. http://www.technologyreview.com/lists/breakthrough-technologies/2013/.

[3] Deng L, Yu D. Deep learning: methods and applications[J]. Foundations and Trends in Signal Processing, 2014, 7(3-4): 197-387.

[4] Bengio Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.

[5] Lee T S, Mumford D. Hierarchical Bayesian inference in the visual cortex[J]. Journal of the Optical Society of America, 2003, 20(7): 1434-1448.

[6] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization on the brain[J]. Psychological Review, 1958, 65(6): 386-408.

[7] Rosenblatt F. Principles of neurodynamics: perceptrons and the theory of brain mechanisms[M]. New York: Spartan, 1962.

[8] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of Physiology, 1962, 160(1): 106-154.

[9] Le C Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision[C]//2010 IEEE International Symposium on Circuits and Systems. Paris: IEEE Press, 2010: 249-256.

[10] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[11] Hinton G, Osindero S, The Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[12] 尹宝才，王文通，王立春. 深度学习研究综述[J]. 北京工业大学学报，2015，41(1)：48-59.

[13] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 2528-2535.

[14] Gardner M W, Dorling S R. Artificial neural networks(the multilayer perceptron)—a review of applications in the atmospheric sciences[J]. Atmospheric Environment, 1998, 32(14): 2627-2636.

[15] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE,

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1998,86(11):2278-2324.

[16] Zeiler M D,Krishnan D,Taylor G W,et al. Deconvolutional networks[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco,USA:IEEE Press,2010:2528-2535.

[17] Yu K,Lin L,Lafferty J. Learning image representations from the pixel level via hierarchical sparse coding[C]// Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition. USA:IEEE Press,2011:1713-1720.

[18] Salakhutdinov R,Hinton G. Deep Boltzmann Machines[J]. Proceeding of Conference on Artificial Intelligence & Statistics,2009,5(8):448-455.

[19] Vincent P,Larochelle H,Bengio Y,et al. Extracting and composing robust features with denoising autoencoders[C]// Proceeding of the 25th International Conference on Machine Learning. Helsinki,Finland:ACM Press,2008:1096-1103.

[20] Markoff J. How many computers to identify a cat?[N]. The New York Times,2012-06-25(10).

[21] 余　凯,贾　磊,陈雨强,等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展,2013,50(9):1799-1804.

[22] Dahl G,Yu Dong,Deng Li,et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition[J]. IEEE Transactions on Acoustics Speech and Signal Processing,2012,20(1):30-42.

[23] Zen H,Senior A,Schuster M. Statistical parametric speech synthesis using deep neural networks[C]// International Conference on Acoustic,Speech and Signal Processing. Vancouver,Canada:IEEE Press,2013:7962-7966.

[24] Deng J,Dong W,Socher R,et al. Imagenet:a large-scale hierarchical image database[C]// IEEE Conference on Visi-on and Pattern Recognition. Miami,USA:IEEE Press,2009:248-255.

[25] Krizhevsky A,Sutskever I,Hinton G. Imagenet classification with deep convolutional neural networks[C]// Proceedings of Advances in Neural Information Processing Systems. South Lake Tahoe,USA:IEEE Press,2012:1097-1105.

[26] Zeiler M D,Fergus R. Visualizing and Understanding Convolutional Neural Networks[EB/OL]. (2013-11-28)[2014-05-20]. http://arxiv.org/pdf/1131.2901v3.

[27] Sun Y,Wang X,Tang X. Deep learning face representation from predicting 10,000 classes[C]// Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus,USA:IEEE Press,2014:1891-1898.

[28] Tang Y,Yang M,Ranzato M A,et al. Deepface:closing the gap to human-level performance in face verification[C]// Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE Press, 2014: 1701-1708.

[29] Ya Yezhou,Li Yi,Fermuller C,et al. Robot learning manipulation action plans by "watching" unconstrained videos from the world wide web[EB/OL]. (2015-01-25)[2015-10-19]. http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9286/9673.

[30] Leon A G,Alexander S E,Matthias B. A neural algorithm of artistic style[EB/OL]. (2015-08-26)[2015-09-02]. http://arxiv.org/abs/1508.06576.

(上接第 33 页)任何的译码开销的条件下,其误码率便可低于 10^{-3}。所以信宿端接收很少的数据包便可以高概率译码,大大提高了DTN网络数据传输的吞吐量。

参考文献:

[1] Li Yun,Wang Xiaoying,Liu Zhanjun,et al. Analysis of link interruption characteristics in the DTN[J]. Journal on Communications,2008,29(11):232-236.

[2] 周晓波,周　健,卢汉成,等. DTN网络的延时模型分析[J]. 计算机研究与发展,2008,45(6):960-966.

[3] Zeng D,Guo S,Hu J. Reliable Bulk-Data Dissemination in Delay Tolerant Networks[J]. IEEE Transactions on Parallel and Distributed Systems. 2014,25(8):2180-2189.

[4] 黄佳庆. 网络编码原理[M]. 北京. 国防工业出版社,2012:1-37.

[5] Wang Y,Jain S,Martonosi M,et al. Erasure-coding based routing for opportunistic networks[J]. Proceeding of ACM Workshop on Delay-tolerant Networking,2005,33(8):229-236.

[6] Luby M. LT Codes[C]// Proceedings of the 43rd Symposium on Foundations of Computer Science. Vancouver:IEEE Computer Society. 2002:271-282.

[7] Shokrollahi A. Raptor codes[J]. IEEE Transactions on Information Theory,2004. 52(6):2551-2567.

[8] 陆维阳. 喷泉码的优化设计及其应用[D]. 北京:北京邮电大学,2013.

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net