

DOI: 10.11992/tis.201611021
网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20170116.1503.004.html>

大数据与深度学习综述

马世龙, 乌尼日其其格, 李小平
(北京航空航天大学 软件开发环境国家重点实验室, 北京 100191)

摘要: 大数据时代改变了基于数理统计的传统数据科学, 促进了数据分析方法的创新, 从机器学习和多层神经网络演化而来的深度学习是当前大数据处理与分析的研究前沿。从机器学习到深度学习, 经历了早期的符号归纳机器学习、统计机器学习、神经网络和 20 世纪末开始的数据挖掘等几十年的研究和实践, 发现深度学习可以挖掘大数据的潜在价值。本文给出大数据和深度学习的综述, 特别是, 给出了各种深层结构及其学习算法之间关联的图谱, 给出了深度学习在若干领域应用的知名案例。最后, 展望了大数据上深度学习的发展与挑战。

关键词: 大数据; 机器学习; 深层结构; 深度学习; 神经网络; 人工智能; 学习算法; 派生树

中图分类号: TP311 **文献标志码:** A **文章编号:** 1673-4785(2016)06-0728-15

中文引用格式: 马世龙, 乌尼日其其格, 李小平. 大数据与深度学习综述[J]. 智能系统学报, 2016, 11(6): 728-742.
英文引用格式: MA Shilong, WUNIRI Qiqige, LI Xiaoping. Deep learning with big data: state of the art and development[J]. CAAI Transactions on Intelligent Systems, 2016, 11(6): 728-742.

Deep learning with big data: state of the art and development

MA Shilong, WUNIRI Qiqige, LI Xiaoping
(State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China)

Abstract: As the era of the big data arrives, it is accompanied by profound changes to traditional data science based on statistics. Big data also pushes innovations in the methods of data analysis. Deep learning that evolves from machine learning and multilayer neural networks are currently extremely active research areas. From the symbolic machine learning and statistical machine learning to the artificial neural network, followed by data mining in the 90s, this has built a solid foundation for deep learning (DL) that makes it a notable tool for discovering the potential value behind big data. This survey compactly summarized big data and DL, proposed a generative relationship tree of the major deep networks and the algorithms, illustrated a broad area of applications based on DL, and highlighted the challenges to DL with big data, as well as identified future trends.

Keywords: big data; machine learning; deep network; deep learning; neural network; artificial intelligence; learning algorithm; derivation tree

大数据不仅为企业带来丰厚的利润, 也开启了科学研究的第四范式, 即数据密集型科学发现^[1]。学术界和产业界对大数据的认识正逐步清晰化并形成共识。大数据时代同时也改变了基于概率论和数

理统计的传统数据科学, 促进了新的数据分析方法的创新, 从机器学习和多层神经网络演化而来的深度学习是当前大数据处理和分析方法的研究前沿。

1 大数据及其挑战

大数据 (big data) 的概念自 1996 年由 John Mashey^[2]提出以来, 经历了一段时间的众说纷纭, 带着产业界的事实数据, 不断进入学术界的研究领域, 引领了一个时代^[3]的到来。

收稿日期: 2016-11-15.
基金项目: 国家自然科学基金项目 (61003016, 61300007, 61305054); 科技部基本科研业务费重点科技创新类项目 (YWF-14-JSXY-007); 软件开发环境国家重点实验室自主探索基金项目 (SKLSDE-2012ZX-28, SKLSDE-2014ZX-06).
通信作者: 李小平. E-mail: lee.rex@163.com.

1.1 大数据特点和界定

从 21 世纪初开始,产业界开始意识到数据产生的规模和速度可能会对基础设施特别是存储设备造成压力^[2]。根据 Martin Hilbert 的一项统计人类信息总量的研究^[4],在 2000 年,数字存储信息只占全球数据量的 1/4,其他 75%的信息尚都存储在报纸、胶片、黑胶唱片和盒式磁带这类媒介上;但经过短短的几年时间,到了 2007 年,人类大约存储了 300 EB 的数据,其中只有 7%是存储在报纸、书籍、图片等媒介上的模拟数据,其余全部是数字数据。数字数据的总量以每 40 个月翻一番的速度积累。注:PB(PetaBytes 拍字节)=1 024 TB=2⁵⁰字节,EB(ExaBytes 艾字节)=1 024PB=2⁶⁰字节,ZB(ZettaBytes 泽字节)=1 024 EB=2⁷⁰字节,YB(Yotta-Bytes 尧字节)=1 024 ZB=2⁸⁰字节。

然而有了规模和速度就是大数据吗?研究人员在不同时期对大数据的特点进行了总结。首当其冲的是 2001 年,META 集团分析师 Doug Laney 给出大数据的 3V 特征^[5],分别为规模性(Volume)、多样性(Variety)和高速性(Velocity)。10 年后,IDC 在此基础上又提出第 4 个特征,即数据的价值(Value)^[6]。2012 年 IBM 则认为大数据的第 4 个特征是指真实性(Veracity)^[7-8]。后来,有人将上述所有特征合起来称为大数据的 5V 特征,也有人从不同的应用视角和需求出发,又提出了粘性(Viscosity)、邻近性(Vicinity)、模糊性(Vague)等多种不同的特征,形成了 3+xV 的大数据特征^[9]。这些特征的具体含义,如表 1。

表 1 大数据特征
Table 1 Big data characteristics

名称	含义
规模性(Volume)	规模可从数百 TB 到数十数百 PB、甚至到 EB 规模
多样性(Variety)	包括各种格式和形态的数据,如文本、图像、音频、视频
实效性(Velocity)	需要在一定的时间限度下得到及时处理
价值密度(Value)	价值密度低,需要通过分析挖掘和利用产生商业价值
真实性(Veracity)	采集的数据的质量影响分析结果的准确性
易变性(Variability)	指数据流的格式变化多样
粘性(Viscosity)	指数据流之间的关联性是否强
邻近性(Vicinity)	获取数据资源的距离
传播性(Virality)	数据在网络中传播的速度
有效性(Volatility)	数据的有效性及存储期限
模糊性(Vague)	因采集手段的多样性和局限性,获取的数据具有模糊性

随着应用的不同研究人员给出了众多的大数据

特点,但最早的 3 个 V 仍被视作大数据应具备的 3 个特征,贯穿于大数据生命周期中的各个阶段^[9],从而也形成了学术界和产业界认同趋于一致的大数据定义,如表 2 所示。

表 2 大数据定义
Table 2 Big data definitions

机构	定义
Gartner	需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产 ^[10]
麦肯锡	一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合,具有海量数据规模、快速数据流转、多样数据类型和价值密度低等四大特征 ^[11]
Wiki	大数据是指利用常用软件工具捕获、管理和处理数据所耗时间超过可容忍时间的数据集 ^[12]

1.2 大数据存储与处理现状

早在 2000 年,美国斯隆数字巡天项目启动,其位于新墨西哥州的望远镜,在短短几周内收集到的数据比天文学历史上总共收集的数据还要多。云平台 DOMO 公司从 2010 年开始做过一份有趣的统计——“数据从不休息”,将不同社交网络每分钟产生的数据进行比较,并形象地给出了它们的数据总量^[13]。据统计,某社交网络产生的评论数在 2010—2011 年期间为每分钟 60 余万条,而在 2013—2015 期间迅速增长为每分钟 410 余万条。据另一个统计,2015 年美国股市每天的成交量可高达 70 亿股。Google 每天要处理超过 24 PB 的数据,这个数据处理量是美国国家图书馆所有纸质出版物所含数据量的千倍之多^[14]。由此说明,除了互联网之外,物联网、移动终端乃至传统的产业都在迅速产生大量的数据。研究人员对大数据的产生方式进行了分类^[15],并指出大数据产生方式经历了被动生成、主动生成、自动生成 3 个阶段^[15],如表 3 所示。

表 3 大数据产生阶段
Table 3 Big data generating

阶段	方式	特点
1	被动	运营式系统阶段,伴随一定的运营活动数据被动产生并保存用户原创内容阶段,尤其是在 Web2.0 时代,社交网络的出现以及以智能手机、平板电脑为代表的新型移动设备的使用率上升,爆炸式地、主动地产生 UGC (user generated content) 数据
2	主动	感知式系统阶段,随着带有处理功能的传感器广泛布置于社会各处,源源不断地、自动地产生新的数据
3	自动	

大数据产生如此之快,对存储提出更高的要求。产业界纷纷投入建设规模庞大的数据存储基础设施。到 2012 年为止,Google 部署在全球的数据中心就有 36 个,单个数据中心的计算机节点将达到数百万个;微软在全球建设超过 20 个数据中心,仅在 2009 年底,就在芝加哥建成当时世界最大的模块化数据中心,包括 220 个集装箱,每个集装箱有 1 800~2 500 台机器不等^[16]。IDC 统计显示,全球的数据储量仅在 2011 年就达到 1.8 ZB,2015 年全球大数据储量达到 8.61 ZB,而今后 10 年,用于存储数据的全球服务器总量还将增长 10 倍^[17]。

大数据不仅需要如此妥善的存储和监管,更需

要进行及时、有效地分析和处理,进而挖掘其中的价值。M.Viktor^[14]指出,利用大数据进行分析时,与以往不同的是:1) 要使用全生命周期数据,即不是样本而是全体;2) 与精确的数据分析相比,更加关注对全量复杂多样数据的分析;3) 与传统的因果关系分析相比,更加关注事物之间的相关关系;4) 基于机器学习的大数据分析结果将减少传统决策中的主观因素。

大数据处理在一般流程上包括数据存储、数据抽取与集成、数据分析、数据解释和用户等几个层次,如文献[15]中图 1 所示。

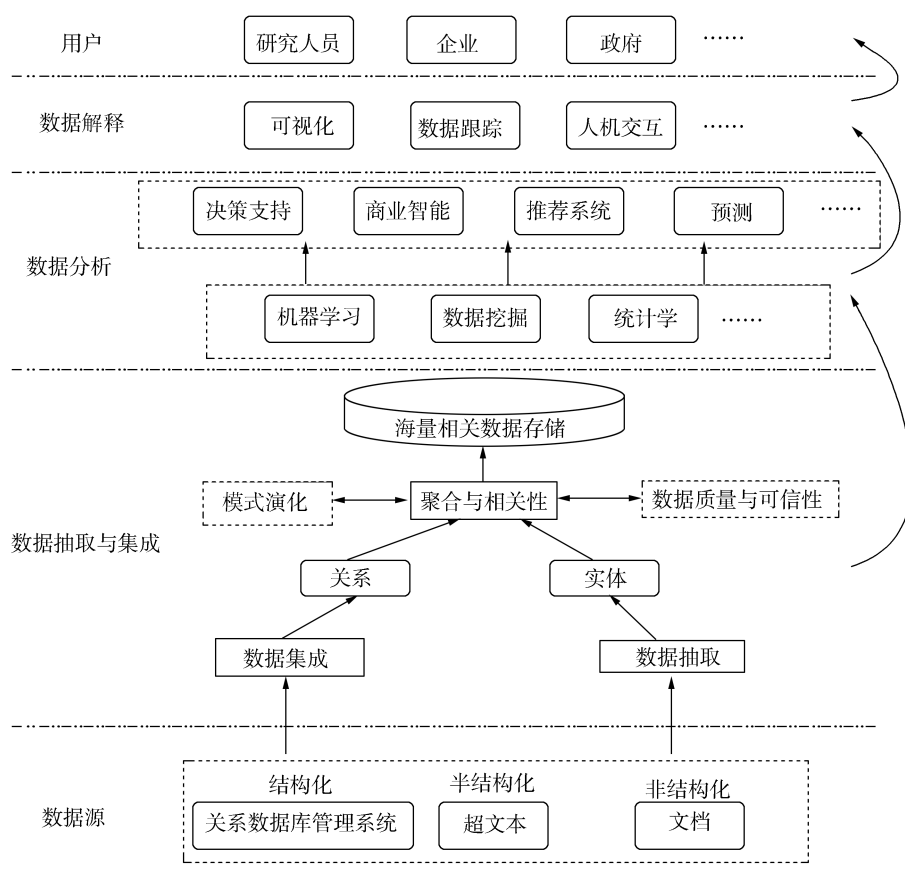


图 1 大数据处理一般流程

Fig.1 Basic framework of big data processing

其中,数据分析是整个大数据处理流程的核心。在这一层所处理的是从异构数据源抽取和集成的数据,也称为分析的原始数据。根据应用的不同需求,可以从中取舍,利用全部或部分数据进行分析。大数据的价值正是产生于此处^[15]。

大数据处理在形式上又分为三大类型,分别为静态数据批量处理、在线数据实时处理和图形图像数据综合处理。其中,对在线数据的实时处理又可以划分为流式数据的处理和交互数据的处理。文献

[18]中对上述几种大数据处理形式从所处理数据的特征、典型应用以及相关经典处理系统等几个维度给出了系统而全面的总结。

可以说,从大数据产生到现在的 10 余年里,关于大数据分析与处理的技术呈现百家争鸣、百花齐放的状态,也取得了显著的成绩。2003 年,人类第 1 次破译人体基因密码时,用了 10 年才完成 30 亿对碱基对的排序,而 10 年后,世界范围内的基因仪 15 min 就可以完成同样的工作。美国股市日交易量的 2/3 都由

建立在数学模型和算法之上的计算机程序自动完成,这些程序运用海量数据,能够预测利益和降低风险。2009 年 Google 公司一家就为美国政府贡献了 540 亿美元的利润。大数据已经成为企业、政府、机构决策的重要源泉,基于大数据分析的应用也成为人们衣食住行必不可少的工具。

1.3 大数据面临的挑战

如果说大数据产生之初所面临的挑战主要表现在如何及时收集和合理存储上^[2],那么 10 余年后的今天,大数据所面临的更多是如何有效地分析大数据^[19-20]。大数据分析是指大数据内容上的分析与计算。由于大数据的众多特点,诸多传统方法如数据挖掘不能直接应用于大数据集场景,大数据分析面临新的挑战,包括^[9, 15, 18]:

1) 传统算法主要基于内存,随着数据规模的空前扩大,它们的时空开销(计算复杂度)变得难于容忍。如何应对大批量的数据,将其装入内存并高效运行成为新的挑战。

2) 为了支持全数据量的实时数据处理,由于有时无法永久化存储,同时数据使用环境持续变化,使得无法掌握数据整个生命周期的特征。如何通过传统批量算法,基于历史数据构建无偏训练集成为新的挑战。

3) 在大数据环境下,数据生产和采集的能力日益增强,这导致数据在规模增大的同时呈现出新的特点:属性数量大而稀疏、数据体量大而高噪声、数据高维而复杂多样。如何处理高维、高噪声、稀疏数据成为新的挑战。

4) 虽然机器学习善于处理非确定性的复杂问题,但是对于大数据处理与分析的场景,由于大数据的复杂多样性,机器学习在统计分析、学习目标和学习效率方面遇到了新的挑战。

大数据分析所涉及的关键技术包括深度学习、知识计算和可视化等^[18]。其中深度学习是一种基于机器学习、数据挖掘技术以及神经网络理论,分析大数据潜在价值的过程。本文,后续将围绕深度学习进行综述,并最后给出这两个热门研究领域在未来的一些可能性。

2 从机器学习到深度学习

2.1 机器学习发展历程

如果说人的学习是通过观察获得某种技能的过程,那么机器学习就是在用计算机模仿这一过程^[21]。机器学习被认为是计算机拥有智能的根本途径,在人工智能发展的早期阶段占据了重要地

位^[22]。机器学习的研究从其热衷度上大致分为 3 个阶段。

首先是 20 世纪五六十年代,机器学习的萌芽时期。1959 年,Arthur Samuel 设计了一个具有学习能力的下棋程序,可以通过一次次的对弈改善程序自身的棋艺。该程序不仅在 4 年后,战胜其设计者本人,7 年后还战胜了美国一位保持 8 不败战绩的冠军^[23]。1956 年, Frank Rosenblatt 提出了一种基于神经网络的数学模型—感知机,采用线性优化的方法模拟人类学习的神经系统^[24]。同期, Widrow 提出最小均方误差(least mean square, LMS) 算法开启了对自适应元素的训练^[25]。这些探索使得机器学习第一次成为热门研究。

然而,1969 年人工智能之父 Marvin Minsky 指出单层感知机无法处理线性不可分问题,如异或(同或)的分类;以及基于“黑箱”原理无法将模型与现实世界直接对应等问题^[26-27],使得机器学习的研究一度进入低谷。虽然 Widrow 和 Winter 提出的 Madaline 算法通过分段线性化的思想能够解决异或的分类问题,但仍然不能彻底解决感知机所面临的挑战。但是,这一尝试却开启了研究人员基于符号归纳的机器学习和集成机器学习的探索^[27]。同时,20 世纪 70 年代随着有限样本统计理论引入机器学习,涌现了基于人工神经网络(artificial neural network, ANN)上的众多统计机器学习算法,最著名包括支持向量机(SVM)^[28]、高斯混合模型(GMM)^[29]以及逻辑回归(LR)。从而机器学习在一段冷静时期之后第 2 次成为研究热点。

进入 20 世纪 80 年代后期, Rumelhart 提出多层感知机从而解决了线性不可分的问题^[30]。但由于数据产生速度的不断提升,多层感知机也变得对其无法适应,机器学习的算法亟待改进^[26]。与此同时,在神经网络领域 Paul Werbos 提出反向传播(back propagation, BP)算法使线性不可分的问题在复杂神经网络上也能得以解决^[31]。从此,机器学习进入第 3 个阶段,即快速发展时期。在这一阶段,不断涌现出更优秀的算法^[32],推动了人工智能在语音识别、图形图像处理以及自然语言处理方面的进展。

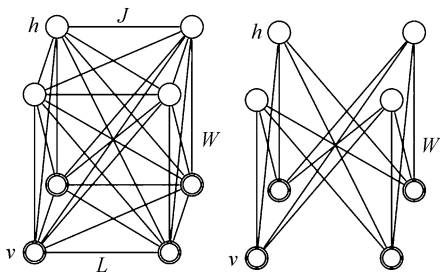
但机器学习在几十年的发展,仍有很多问题未能解决。其中包括:BP 算法随着神经网络层次的加深,参数优化效果无法传递到前层,从而导致容易出现局部最优解^[33]或过拟合问题^[34];此外,由于机器学习在实际应用中需要手工设计特征和线性分类器,它不仅依赖领域专家的知识,还需要人在学习过程中参与这使得学习耗时耗力。而且,这种机器学

习无法很好地处理自然数据(无标签数据),同时不易应用于深层的网络^[35]。然而解决这些问题,促成了机器学习新的分支——深度学习的研究。

2.2 多层结构和深度学习

深度学习也叫无监督特征学习(unsupervised feature learning),即可以无需人为设计特征提取,特征从数据中学习而来。深度学习实质上是多层表示学习(representation learning)方法的非线性组合。表示学习是指从数据中学习表示(或特征),以便在分类和预测时提取数据中 useful 信息^[36]。深度学习从原始数据开始将每层表示(或特征)逐层转换为更高层更抽象的表示,从而发现高维数据中错综复杂的结构^[35]。

深度学习的发展不仅源于机器学习的丰厚积累,同时也受到统计力学的启发。1985 年 D. H. Ackley 等基于玻尔兹曼分布,提出了一种具有无监督学习能力的神经网络玻尔兹曼机(Boltzmann machine, BM)^[37]。该模型是一种对称耦合的随机反馈型二值单元神经网络,由可视单元和多个隐藏单元组成,用可视单元和隐单元表示随机网络与随机环境的学习模型,用权值表示单元之间的相关性。通过该模型能够描述变量之间的相互高阶作用,但其算法复杂,不易应用^[38]。次年 P. Smolensky 基于他本人所提出的调和论给出了一种受限的玻尔兹曼机模型(RBM)。该模型将 BM 限定为两层网络,一个可视单元层和一个隐藏单元层。并且进一步限定层内神经元之间相互独立,无连接,层间的神经元才可以相互连接。如图 2 所示^[39]。



(a) 一般玻尔兹曼机 (b) 受限的玻尔兹曼机

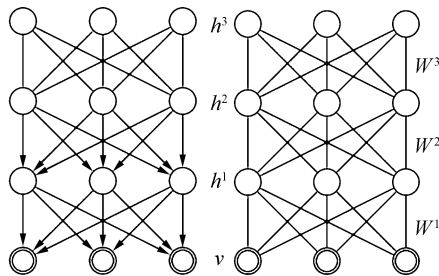
图 2 一般玻尔兹曼机和受限的玻尔兹曼机

Fig.2 A general BM and a RBM

RBM 通过两层马尔可夫随机场,从训练样本得到的隐藏层中神经元状态,并估计独立于数据的期望值。该模型由于大幅度提高了 BM 的学习效率^[40],被众多研究人员所借鉴,从而开启了深度学习的研究热潮。

其中最为典型的深度结构包括深度置信网络(deep belief network, DBN)和深度玻尔兹曼机(deep

Boltzmann machine, DBM)。DBN 是由 GE Hinton 于 2006 年提出的一种串联堆叠 RBM 形成的深层模型^[41]。该模型在训练阶段将一层 RBM 的输出作为另一层 RBM 的输入,由此逐步训练隐藏层的高阶数据相关性,最后采用 BP 对权值进行微调。而 DBM 是一种特殊的 BM。不同的是,除了有一个可视层之外,它具有多个隐藏层,并且只有相邻隐藏层的单元之间才可以有连接。它们之间的对比如图 3^[39]。



(a) 深度置信网络 (b) 深度玻尔兹曼机

图 3 深度置信网络和深度玻尔兹曼机

Fig.3 A DBN and a DBM

对于一个 RBM,如果把隐藏层的层数增加,可以得到一个 DBM;如果在靠近可视层的部分使用贝叶斯信念网络(即有向图模型),而在最远离可视层的部分使用 RBM,则可以得到一个 DBN。由于 RBM 的训练中采用对比散度 CD 算法^[42]可以快速得到训练,使得深度置信网络的训练速度也大幅度提升。

在深度学习发展的 10 年中,基于上述两种网络结构的深度学习算法不时涌现,使其成为一门广袤的学科。根据文献[43-45],本文对现有深度学习算法之间的派生关系进行梳理,形成如图 4 所示的树形结构。

该图通过节点和有向边展示了不同深层结构之间的派生关系。派生关系表示深度网络是在网络结构或训练算法上的微调或改进。有些深度学习算法是在原有某个深度学习算法的基础上对其网络结构进行了调整而形成,例如堆叠自动编码器就是受 DBN 启发,将其中的 RBM 替换为 AE 而形成,如图 4 中 2007 年 Yoshua 等提出的 SAE。而有些深度学习算法则是结合了多种已有深度学习算法派生而来,例如堆叠卷积自动编码器就是在卷积网络中采用了自动编码器 AE 而形成,如图 4 中 2011 年 Masci 提出的 SCAE。采用有向边将文中所涉及的深度学习算法相连接,可以形成一棵深度网络派生树。在深度学习的整个发展过程中 DBN、DBM、AE 和 CNN 构成了早期的基础模型。后续的众多研究则是在此基础上提出或改进的新的学习模型。关于更多派生模型的详细信息如表 4 所示。

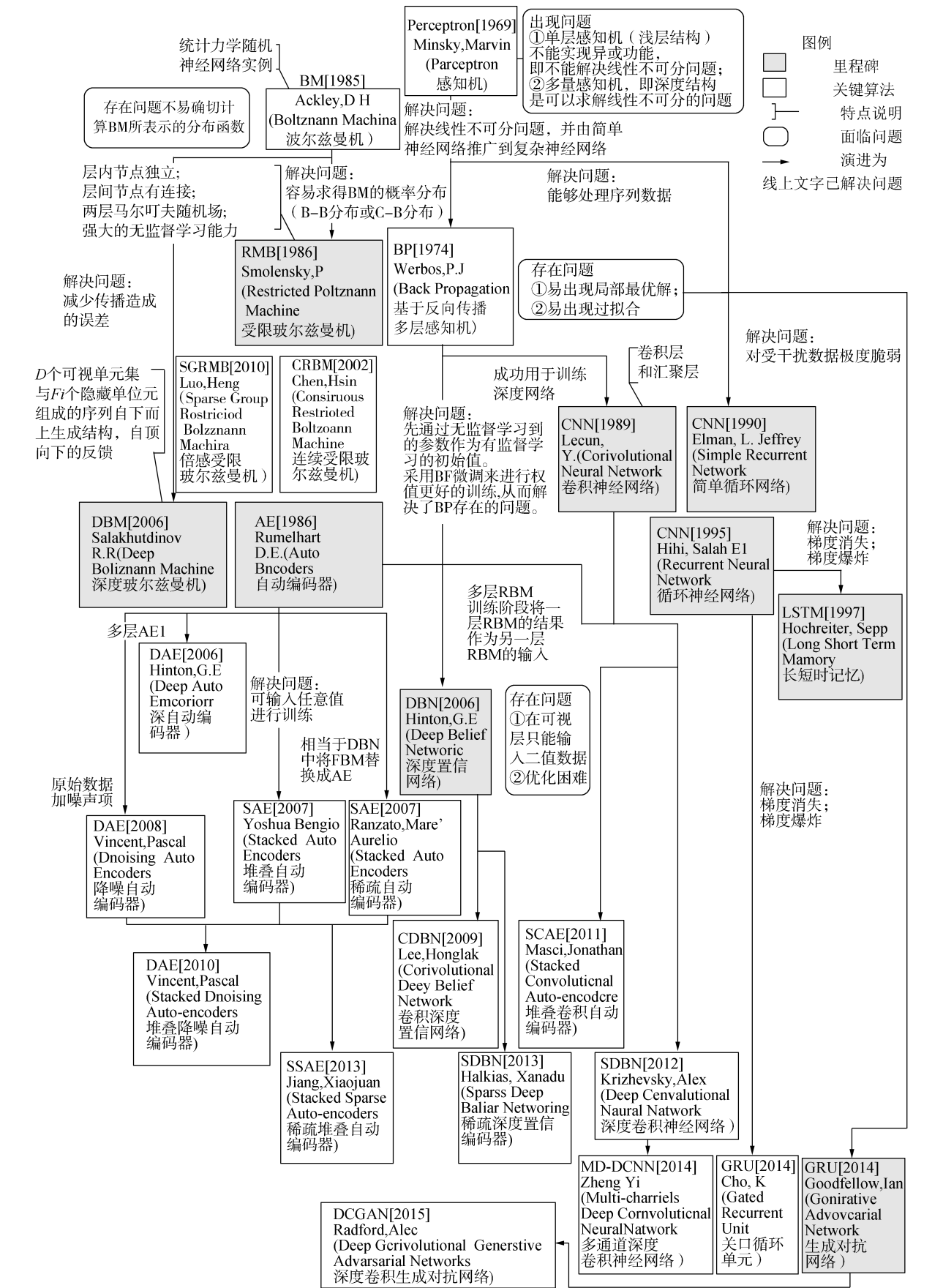


图 4 深层网络派生树

Fig.4 The derivation tree of the deep networks

表 4 深层网络及其算法
Table 4 Deep networks and algorithms

年份	深度模型	提出者	网络结构	相关训练算法	模型特点及解决问题	存在问题
1969	Perceptron ^[26] 感知机	M.Minsky S.Papert	—	—	线性可分问题	线性不可分问题
1974	BP ^[31] 反向传播	P.J.Werbos	—	链式积分法	线性不可分问题;从简单神经网络到复杂神经网络的推广	局部最优解问题;过拟合问题
1985	BM ^[37] 玻尔兹曼机	D.H.Ackley G.E.Hinton T.J.Sejnowski	多层	—	统计力学中一种能量模型,随机神经网络实例	难于计算分布
1986	RBM ^[40] 受限玻尔兹曼机	P.Smolensky	2 层 无向边	对比散度法	容易求得 BM 的概率分布;具有无监督学习能力	效率低
2002	CRBM ^[46] 连续受限玻尔兹曼机	H.Chen A.Murray	2 层	MCD;BP	能够对连续数据建模	参数调优困难
2010	SGRBM ^[47] 稀疏组受限玻尔兹曼机	Luo Heng Shen Ruimin Niu Cahngyong	2 层	稀疏惩罚对数似然;BP	稀疏表示符合生物学特征惩罚隐藏单元的损失	隐藏单元分组方式和依据尚不明确
2006	DBN ^[41] 深度置信网络	G.E.Hinton R.R.Salakhutdinov	多层 有/无向边 全连接	贪心逐层训练算法;BP	RBM 的堆叠;以无监督学习到的参数作为有监督学习的初始值,从而解决了 BP 的问题	可视层只能接收二值数值;优化困难
2006	DBM ^[39] 深度玻尔兹曼机	R.R.Salakhutdinov G.E.Hinton	多层 无向边 全连接	BP	BM 的特殊形式;自下而上生成结构;减少传播造成的误差	效率低
2009	CDBN ^[48] 卷积深度置信网络	Lee Honglak Grosse Roger Ranganath Rajesh Ng Andrew Y	多层 无向边 部分连接	贪心逐层训练算法;BP	采用概率最大池能够对高维图像做全尺寸衡量,并对输入的局部变换具有不变性	优化困难
2013	SDBN ^[49] 稀疏深度置信网络	X.Halkias S.Paris H.Glotin	多层 有/无向边 全连接	混淆范数;BP	混淆范数作为稀疏约束的 DBN 结构	优化困难
1986	AE ^[50] 自动编码器	D.E.Rumelhart G.E.Hinton R.J.Williams	3 层	贪心逐层训练算法	通过编码器和解码器工作完成训练;通过损失函数最小化求出网络的参数;	不能用于分类
2006	DAE ^[41] 深自动编码器	G.E.Hinton R.R.Salakhutdinov	多层	贪心逐层训练算法;BP	无监督逐层贪心训练算法完成对隐含层的预训练;并通过 BP 微调,显著降低了性能指数;	隐藏层数量和神经元的数量增多导致梯度稀释

续表 4

年份	深度模型	提出者	网络结构	相关训练算法	模型特点及解决问题	存在问题
2007	SAE ^[51] 堆叠自动编码器	B.Yoshua L.Pascal P.Dan Hugo Larochelle	多层 有/无向边 全连接	梯度下降 算法;BP	将 DBN 中的 RBM 替换为 AE 后的生成模型;通过将第一层的贝努力分布的输入改为高斯分布,扩展成可输入任意值进行训练	同上
2007	SAE ^[52] 稀疏自动编码器	M.Ranzato , Y.Boureau Y.Lecun	3 层	梯度下降算法;BP	降维,学习稀疏的特征表达	同上
2008	dAE ^[53] 降噪自动编码器	P.Vincent H.Larochelle Y.Bengio A.Manzagol	3 层	梯度下降算法;BP	在破损数据的基础上训练;使训练得到的权重噪声较小,从而提高鲁棒性	同上
2010	SDAE ^[54] 堆叠消噪自动编码器	P.Vincent H.Larochelle I.Lajoie Y.Bengio 等	多层	梯度下降 算法;BP	将多个 dAE 堆叠起来形成深度网络结构,用来提取特征表达	同上
2013	SSAE ^[55] 稀疏堆叠自动编码器	Jiang Xiaojuan Zhang Yinghua Zhang Wensheng 等	多层	梯度下降 算法;BP	在 SAE 的损失函数上加入稀疏惩罚值形成的深层网络	同上
1998	CNN ^[56] 卷积神经网络	Y.Lecun L.Bottou Y.Bengio P.Haffner	多层 无向边 局部连接 共享权值	梯度下降 算法;BP	包含卷积层和子采样层;可以接受 2D 结构的输入;具有较强的畸变鲁棒性	要求较高计算能力的资源
2011	SCAE ^[57] 堆叠卷积自动编码器	Masci Jonathan Meier Ueli Dan Cireşan 等	多层	梯度下降算法;BP	堆叠的 CAE 结构,每层采用没有正则项的传统梯度下降算法进行训练	同上
2012	DCNN ^[58] 深度卷积神经网络	A.Krizhevsky I.Sutskever G.Hinton	多层 局部连接 共享权值	梯度下降 算法;BP	CNN 的深层结构,采用纯监督学习广泛应用于图像识别	同上
1990	SRN ^[59] 简单循环网络	J.L.Elman	3 层	BPTT;梯度 下降算法	时间维度上的深层结构;上一时刻的输出是下一时刻的输入	长时间依赖问题
1995	RNN ^[60] 循环神经网络	S.E.Hihi M.Q.Hc-J Y.Bengio	多层	BPTT;梯度 下降算法	多层的时间维度上的深层结构;能够处理序列数据	梯度消失或 梯度爆炸
1997	LSTM ^[61] 长短是记忆	S.Hochreiter J.Schmidhuber	多层	BPTT;梯度 下降算法	通过为每一个神经元引入 gate 和存储单元,能够解决 RNN 所面临的梯度消失或爆炸问题 由于具有记忆功能,能够处理较为复杂的序列数据	训练复杂度较高、解码时延较高

续表 4

年份	深度模型	提出者	网络结构	相关训练算法	模型特点及解决问题	存在问题
2014	GRU ^[62] 关口循环单元	K.Cho B.Van Merrienboer D.Bahdanau Y.Bengio	多层	BPTT;梯度下降算法	相比于 LSTM,只设置一个更新关口,运行比 LSTM 更快,更容易	表达能力相对弱
2014	Attention ^[63] 注意力机制	V.Mnih N.Heess A.Graves K.Kavukcuoglu	—	—	受人类的注意力机制的启发,每次处理注意力部分的数据,减少任务复杂度	增加了存储开销
2014	GAN ^[64] 生成对抗网络	Goodfellow Ian Pougetabadie Jean Mirza Mehdi Xu Bing Wardefarley David Ozair Sherjil Courville Aaron Bengio Yoshua	多层 无向边 局部连接 共享权值	BP;dropout	由不同网络组成,成对出现,协同工作 一个网络负责生成内容,另一个负责对内容进行评价多以前馈网络和卷积网络的结合为主	训练较难;训练过程不稳定
2015	DCGAN ^[65] 深度卷积生成对抗网络	A.Radford L.Metz S.Chintala	多层 无向边 局部连接 共享权值	BP;梯度下降算法	GAN 基于 CNN 的扩展,可以从训练数据中学习近似的分布情况	训练过程不稳定

注:‘—’表示尚不明确或不适用。

派生树有助于理解众多神经网络结构之间的关系,而对应的表 4 则对每个深度学习算法的网络结构、训练算法以及解决问题或存在问题给出了简要总结。两组信息合起来,能够成为该领域学习参考的一个索引。

深度学习的网络结构因网络的层数、权重共享性以及边的特点不同而有所不同。因此,在表 4 中给出了每一种深度学习算法的网络结构特点。其中,绝大多数深度学习算法体现为空间维度上的深层结构,且属于前向反馈神经网络;而以循环神经网络(RNN)为代表的 SRN、LSTM 和 GRU 等深度学习算法,通过引入定向循环,具有时间维度上的深层结构,从而可以处理那些输入之间有前后关联的问题。根据对标注数据的依赖程度,深度学习算法中 DBN、AE 及其派生分支体现为以无监督学习或半监督学习为主;CNN、RNN 及其派生分支则以有监督学习为主。此外,根据学习函数的形式,机器学习算法又可以分为生成模型和判别模型^[66]。表 4 列出的深度学习模型中,DBN 及其派生的深度学习模型(如 CDBN)属于生成模型,AE 及其派生的深度学习

模型则属于判别模型,而像 DBM、GAN 等深度学习模型既包括生成模型也包括判别模型。

在作者所收集信息的范围内,派生树涵盖了目前主流的 20 余种模型,然而研究人员实际提出的深层网络模型并不仅限于此。而且,也有新的机制在被引入到已有的深度学习模型中,例如 2014 年 Mnih, Volodymyr 等受人类注意力机制的启发,提出了 Attention 模型(也叫注意力机制)。这一新的机制,通过与 RNN 或 CNN 结合,降低了数据处理任务复杂度的同时,进一步拓宽了深度学习模型的应用领域。作者认为新的模型是对原有基础框架的扩展或改进,该树仍在不断生长。

2.3 深度学习的应用案例

派生树(图 4)揭示了近几年的时间里深度学习算法如雨后春笋不断涌现。这些算法在计算机图像识别、语音识别、自然语言处理以及信息检索等领域在不断刷新历史记录。本节根据不同应用领域对相关的知名案例进行简述。

1) 图像识别

图像识别过去依赖人工设置的特征,特别适合

于提取低等级边界信息的 SIFT(scale invariant feature transform,尺度不变特征转换)或 HOG(Histogram of Oriented Gradients,方向梯度直方图)等^[43]。然而,随着小样本无法真实反映实际复杂的情况,研究人员开始在大数据集上进行实验。

2006年,GE Hinton 用一个拥有3个隐藏层,170多万权重的 DBN 在 MNIST 手写特征识别的数据集上进行训练,在没有对样本进行预处理的情况下,在1万左右的测试集上,错误率为1.25%,低于反向传播网络的最好成绩1.5%,以及支持向量机(SVN)的1.4%^[67]。

1995年 Yann Lecun 推出 LeNet-5 系统,该系统通过2个卷积层、2个子采样层和2个全连接层,形成典型的 CNN 网络。在 MNIST 数据集上的实验得到了0.9%的错误率,20世纪90年代用于银行手写支票识别^[68]。这也是 CNN 成功案例之一。

2010年,A.Krizhevsky 用 DCNN 算法在 LSVRC-2010 数据集上,对1000个不同类别的120万个高分辨率图像进行分类,测试结果在 top1 和 top5 上的错误率分别为37.5%和17.0%,刷新了当时的记录,而他们所采用的是具有5个卷积层、6000万个参数和65万个神经元的深层网络^[58]。该方法在 ILS-VRC-2012 数据集上 top5 的测试错误率为15.3%。

2011年,在 Google 的 xLab,A.Y. Ng 和 J. Dean 建立了全球最大的神经网络——“深度神经网络”(deep neural networks, DNN),也叫“谷歌大脑”,并进行了一个猫脸识别的实验。该实验从 YouTube 视频中取出1000万张200×200的静态图片,让系统自动学习并判断哪些是猫的图片。实验结果表明,他们所建立的深层网络及其算法,在 ImageNet 数据集1万类图中效果提升15%,2.2万类图中效果提升70%^[69]。而据纽约时报^[70],该网络是一个用16000个CPU并行计算平台训练内部拥有10亿个节点的机器学习模型。

2014年,Sun Yi 等提出 Deep-ID 应用于人脸识别,在 LFW 上取得97.45%的准确率^[71]。同年,Zheng Yi 提出的多通道深度卷积神经网络(multi-channels deep convolutional neural networks, MC-DCNN)在 BID-MC 数据集上获得最好的准确率为94.67%。

2016年初,谷歌旗下最强大脑(DeepMind)公司推出人工智能机器人 AlphaGo,创下了围棋人工智能领域的诸多世界纪录。包括:在不让子的情况下,第一次在完整的围棋竞技中击败专业选手(比分5:0);在中国围棋规则下,成功挑战围棋世界冠军李世石

(比分4:1)^[72]。AlphaGo 的关键技术有深度学习、强化学习和蒙特卡洛树搜索。在其有监督学习策略和强化学习价值网络中采用了 CNN 结构^[73]。

2) 语音识别

在过去语音识别一直采用 GMM-HMM 模型。然而,2012年 G.E.Hinton 等考虑了语音数据内部原有的结构特征以后,将传统的模型中的高斯混合模型 GMM 替换为 DBN 进行实验,结果表明在 TIMIT 核心测试集上,错误率降到20.7%,有明显提升^[74]。

无独有偶,其他研究人员也尝试将 GMM-HMM 中的 GMM 替换为其他深度神经网络,也得到了良好的效果。例如,2014年 IBM 沃特森研究中心 T.N. Sainath 证实 DNN 比过去的 GMM-HMM 提升8%~15%^[43]。2012年 Pan Jia 提出将 GMM 改为上下文相关的 DNN,并对320h的英文接线员的语音记录和70~700h不等的3个中文语音记录进行试验,结果表明错误率可降低3%^[75]。

微软公司的语音视频检索系统(Microsoft audio video indexing service, MAVIS)^[76],也是在 CD-DNN-HMM 深度模型的基础上进行开发,其在 RT03S 数据集上单词错误率从27.4%降低到18.5%,相当于效果提升33%^[77]。2012年,微软在天津公开演示了 MAVIS 系统对现场讲演者的英文演讲,进行后台的语音识别、英中机器翻译和中文语音合成等一系列处理,效果流畅。

3) 自然语言处理

自然语言处理(NLP)传统处理方法的缺陷为采用浅层结构;使用线性分类器且需要人工设计大量较好的特征进行预处理;特征在分离的任务中被串联导致传播误差增大^[43]。

2003年, Bengio 等提出词向量(word embedding 或 word representation)方法^[78],采用神经网络构建语言模型。之后,研究人员在此基础上提出了不同的词向量训练模型,包括:2008年 Collobert, Ronan 等提出的 C&W 模型用以完成自然语言处理中的词性标注、命名实体识别、短语识别、语义角色标注等任务,由其在语义角色标签的问题上进行无人工设计特征参与的训练,错误率降至14.3%,刷新了纪录^[79];同年, Mnih 和 Hinton 提出的 HLBL 模型^[80],将 Bengio 的词向量方法中最后隐藏层到输出层的矩阵乘积替换为一个层级结构,在算法复杂度上得到显著提升;由于人类自然语言具有递归特性,即任何语言中的句子,事实上可以由词、短语递归组合而成,因此,将循环神经网络(recurrent neural network, RNN)引入 NLP 成为一种趋势。从2010年开

始,来自 Google 的 T.Mikolov 一直从事该领域的研究,提出了 RNNLM(循环神经网络语言模型),在语言模型的训练速度、准确率以及困惑度(Perplexity)上得到了改善^[81-84];2012年由 Eric H. Huang 对 Bengio 的模型进行改进,加入了语义信息,能够对一词有多个表示,性能高于 C&W 模型^[85]。此外,T.Mikolov 还提出了 Word2Vec 模型^[86-87],并将相应的工具开源,成为最为典型的词向量案例。

4) 信息检索

深度学习在信息检索上的应用主要通过提取有用的语义特征进行子序列文档排序^[43, 88]。

2009年 R.Salakhutdinov 等指出当时的信息检索系统的主要问题表现在:①在词计数空间中直接计算文档的相似性,使得在大词汇量下计算效率低;②没有充分利用词汇之间的语义相似性。为此,他们从词计数向量出发通过 DAE 模型得到紧凑编码使之映射到内存,并将相似内存地址的文档进行归类,使检索的效率提高的同时,计算效率也得到了提升^[43]。

纵观深度学习在人工智能不同细分领域中的应用,在计算机视觉、语音识别和自然语言处理中成绩相对显著;在信息检索领域虽然仍有待突破,但依然不时涌现新的深度结构及其算法。

3 大数据上深度学习展望

深度学习横跨计算机科学、工程技术和统计学等多个学科并应用于政治、金融、天文、地理以及社会生活等广泛的领域^[89]。深度学习的优点在于模型的表达能力强,能够处理具有高维稀疏特征的数据,而大数据所面临的挑战亟待引入深度学习的思想、方法和技术进行及时有效地解决。如何将深度学习应用于大数据分析,发现数据背后的潜在价值成为业界关注的热点。

3.1 大数据上深度学习所面临的问题

尽管深度学习已经取得令人瞩目的成绩,但这一研究领域尚处在初期阶段,仍然面临诸多问题^[90-91]:

1) 深层结构的深度、隐藏层数和隐藏节点数如何决定,例如,“谷歌大脑”的内部神经元个数(10亿)相比于人类大脑的神经元个数(150亿)还差一个数量级;

2) 梯度下降法如何进行改进以达到更好的局部极值点甚至是全局最优点;

3) 如何评价深度学习获得的特征是优是劣,以及如何解释所学到的特征;

4) 深度学习在大规模网络中取得了一定的成

绩但仍处于萌芽阶段,如何处理大数据的规模所带来的大样本、高维属性和数据类型的多样性。

3.2 大数据分析上的深度学习展望

谷歌 AlphaGo 的亮相,让更多的人关注深度学习。基于大数据分析的决策系统已成为国家“十三五”规划中各行业计划中的建设项目。未来大数据上深度学习的发展可能包括^[35]:

1) 在计算机视觉领域,深度学习应该能够更好地处理目标和行为识别这类复杂的问题,应该能够学习更为复杂的函数关系^[90]。因此,在这个领域可以期待更多的深度结构和深度学习算法,以解决更为复杂的问题。特别是卷积网络 CNN 与利用强化学习的 RNN 相结合,期望能够学习人类视觉的主动性。

2) 在自然语言处理方面,期待利用 RNN 模型的更多新算法,可以更好地理解自然语言中的句子或整个文档,同时期待将 GAN 与 RNN 相结合,增强学习的鲁棒性。

3) 虽然在语音识别和手写识别等领域已经将深度学习与简单推理紧密结合,但在未来,可以期待深度学习与复杂推理的有效结合,以期接近人工智能的终极目标。

4) 大数据的规模之大,并不意味着训练样本数量的充足。由于目前深度学习模型仍然需要大量标注样本进行训练,因此,如何基于少量样本训练深度学习模型的研究令人期待。

5) 面对大数据特征所带来的环境变化应充分利用计算资源,进一步提高效率使之在真实的大数据分析上走向实用。

6) 目前深度学习框架在不需要人工提取特征的情况下,能够实现端到端(end-to-end)的学习,但人类文明发展至今积累了丰富的先验知识,如何利用现存的人类知识库,将其融入现有的深度学习框架将会成为重要的研究点。

参考文献:

- [1] TOLLE K M, TANSLEY D, HEY A J G. The fourth paradigm: data-intensive scientific discovery[J]. Proceedings of the IEEE, 2012, 99(8): 1334-7.
- [2] MASHEY J R. Big Data and the next wave of infra stress [D]. Berkeley: University of California, 1997.
- [3] MAYER-SCHÖNBERGER V, CUKIER K. A big data: a revolution that will transform how we live, work, and think [M]. Boston: Eamon Dolan, 2013.
- [4] HILBERT M, LÓPEZ P. The world's technological capacity to store, communicate, and compute information[J]. Science, 2011, 332(6025): 60-65.

- [5] LANEY D. 3D data management: controlling data volume, velocity, and variety [R]. META Group Research Note, 2001.
- [6] IDC. IIS: the “four vs” of big data [EB/OL]. [2016-11-11]. <http://www.computerworld.com.au/article/396198/iis-is-four-vs-big-data/>.
- [7] SCHROECK M J, SHOCKLEY R, SMART J, et al. Analytics: the real-world use of big data [R]. Oxford: IBM, 2012.
- [8] IBM. The four v's of big data [EB/OL]. 2014 [2016-11-11]. <http://www.ibmdatahub.com/infographic/four-vs-big-data>.
- [9] 郭平, 王可, 罗阿理, 等. 大数据分析中的计算智能研究现状与展望[J]. 软件学报, 2015, 26(11): 3010-3025.
GUO Ping, WANG Ke, LUO Ali, et al. Computational intelligence for big data analysis: current status and future prospect[J]. Journal of software, 2015, 26(11): 3010-3025.
- [10] Gartner. Big data [EB/OL]. [2016-11-11]. <http://www.gartner.com/it-glossary/big-data/>.
- [11] MANYIKA J, CHUI M, BROWN B, et al. Big data: the next frontier for innovation, competition, and productivity [R]. Analytics; McKinsey & Company, 2011.
- [12] Wikiprdia. Big data [EB/OL]. 2009. [2016-11-11]. https://en.wikipedia.org/wiki/Big_data.
- [13] JAMES J. How much data is created every minute? [EB/OL]. [2016-11-11]. <https://www.domo.com/blog/how-much-data-is-created-every-minute/>.
- [14] 维克托·迈尔·舍恩伯格, 周涛. 大数据时代生活、工作与思维的大变革[M]. 周涛, 译. 杭州: 浙江人民出版社, 2013: 136-136.
- [15] 孟小峰, 慈祥. 大数据管理: 概念、技术与挑战[J]. 计算机研究与发展, 2013, 50(1): 146-169.
MENG Xiaofeng, CI Xiang. Big data management: concepts, techniques and challenges[J]. Journal of computer research and development, 2013, 50(1): 146-169.
- [16] 王意洁, 孙伟东, 周松, 等. 云计算环境下的分布存储关键技术[J]. 软件学报, 2012, 23(4): 962-986.
WANG Yijie, SUN Weidong, ZHOU Song, et al. Key technologies of distributed storage for cloud computing[J]. Journal of software, 2012, 23(4): 962-986.
- [17] GANTZ J, REINSEL D. Extracting value from chaos [R]. Idcenc2 Report, 2011.
- [18] 程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述[J]. 软件学报, 2014, 25(9): 1889-1908.
CHENG Xueqi, JIN Xiaolong, WANG Yuanzhuo, et al. Survey on big data system and analytic technology [J]. Journal of software, 2014, 25(9): 1889-1908.
- [19] STUART D. The data revolution: big data, open data, data infrastructures and their consequences[J]. Online information review, 2015, 39(2): 272.
- [20] LABRINIDIS A, JAGADISH H V. Challenges and opportunities with big data [J]. Proceedings of the vldb endowment, 2012, 5(12): 2032-2033.
- [21] ABU-MOSTAFA Y S, MAGDON-ISMAIL M, LIN H T. Learning from data: a short course [M]. Chicago: Aml-book, 2012.
- [22] 洪家荣. 机器学习——回顾与展望[J]. 计算机科学, 1991, 18(2): 1-8.
HONG Jiarong. Machine learning - review and vision [J]. Computer science, 1991, 18(2): 1-8.
- [23] SAMUEL A L. Some studies in machine learning using the game of checkers. II—recent progress [J]. Annual review in automatic programming, 1969, 6: 1-36.
- [24] ROSENBLATT F. The perceptron—a perceiving and recognizing automaton [R]. Ithaca, NY: Cornell Aeronautical Laboratory, 1957.
- [25] WIDROW B, LEHR M A. 30 years of adaptive neural networks: perceptron, Madaline, and backpropagation [J]. Proceedings of the IEEE, 1990, 78(9): 1415-1442.
- [26] MINSKY M, PAPERT S A. Perceptrons: an introduction to computational geometry, expanded edition [M]. Cambridge, Mass: MIT Press, 1988: 449-452.
- [27] 王珏, 石纯一. 机器学习研究[J]. 广西师范大学学报: 自然科学版, 2003, 21(2): 1-15.
WANG Jue, SHI Chunyi. Investigations on machine learning [J]. Journal of Guangxi normal university: natural science edition, 2003, 21(2): 1-15.
- [28] CORTES C, VAPNIK V. Support-vector networks [J]. Machine learning, 1995, 20(3): 273-297.
- [29] REYNOLDS D A, ROSE R C, SMITH M J T. A mixture modeling approach to text-independent speaker identification [J]. Journal of the acoustical society of america, 1990, 87(S1): 109.
- [30] RUMELHART D E, MCCLELLAND J L. Parallel distributed processing: explorations in the microstructure of cognition: foundations [M]. Cambridge, Mass: MIT Press, 1987.
- [31] WERBOS P J. Backpropagation through time: what it does and how to do it [J]. Proceedings of the IEEE, 1990, 78(10): 1550-1560.
- [32] WU Xindong, KUMAR V, QUINLAN J R, et al. Top 10 algorithms in data mining [J]. Knowledge and information systems, 2008, 14(1): 1-37.
- [33] GORI M, TESI A. on the problem of local minima in back-propagation [J]. IEEE transactions on pattern analysis and machine intelligence, 1992, 14(1): 76-86.
- [34] FLETCHER L, KATKOVNIK V, STEFFENS F E, et al. Optimizing the number of hidden nodes of a feedforward artificial neural network [C]//Proceedings of 1998 IEEE In-

- ternational Joint Conference Neural Networks. Anchorage, AK; IEEE, 1998, 2: 1608–1612.
- [35] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436–444.
- [36] BENGIO Y, COURVILLE A, VINCENT P. A courville and P vincent, representation learning: a review and new perspectives[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(8): 1798–1828.
- [37] ACKLEY D H, HINTON G E, SEJNOWSKI T J. A learning algorithm for boltzmann machines[J]. Cognitive science, 1985, 9(1): 147–169.
- [38] 刘建伟, 刘媛, 罗雄麟. 玻尔兹曼机研究进展[J]. 计算机研究与发展, 2014, 51(1): 1–16.
LIU Jianwei, LIU Yuan, LUO Xionglin. Research and development on boltzmann machine[J]. Journal of computer research and development, 2014, 51(1): 1–16.
- [39] SALAKHUTDINOV R, HINTON G. Deep boltzmann machines[J]. Journal of machine learning research, 2009, 5(2): 1997–2006.
- [40] SMOLENSKY P. Information processing in dynamical systems: foundations of harmony theory[M]. Cambridge, Mass: MIT Press, 1986: 194–281.
- [41] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504–507.
- [42] HINTON G E. Training products of experts by minimizing contrastive divergence[J]. Neural computation, 2002, 14(8): 1771–800.
- [43] 张建明, 詹智财, 成科扬, 等. 深度学习的研究与发展[J]. 江苏大学学报: 自然科学版, 2015, 36(2): 191–200.
ZHANG Jianming, ZHAN Zhicai, CHENG Keyang, et al. Review on development of deep learning[J]. Journal of Jiangsu university: natural science edition, 2015, 36(2): 191–200.
- [44] 孙志远, 鲁成祥, 史忠植, 等. 深度学习研究与进展[J]. 计算机科学, 2016, 43(2): 1–8.
SUN Zhiyuan, LU Chengxiang, SHI Zhongzhi, et al. Research and advances on deep learning[J]. Computer science, 2016, 43(2): 1–8.
- [45] SCHMIDHUBER J. Deep learning in neural networks: an overview[J]. Neural networks, 2015,
- [46] CHEN H, MURRAY A. A continuous restricted boltzmann machine with a hardware-amenable learning algorithm[J]. Lecture notes in computer science, 2002, 2415: 358–363.
- [47] LUO Heng, SHEN Ruimin, NIU Changyong. Sparse group restricted boltzmann machines[C]//Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence. San Francisco, California, Usa: AAAI Press, 2010.
- [48] LEE H, LARGMAN Y, PHAM P, et al. Unsupervised feature learning for audio classification using convolutional deep belief networks[C]//Advances in Neural Information Processing Systems 22: Conference on Neural Information Processing Systems 2009. Vancouver, British Columbia, Canada, 2009.
- [49] HALKIAS X, PARIS S, GLOTIN H. Sparse penalty in deep belief networks: using the mixed norm constraint[J]. Computer science, 2013.
- [50] POUGETABADIE J, MIRZA M, XU Bing, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 3: 2672–2680.
- [51] YOSHUA Bengio, PASCAL Lamblin, DAN Popovici, et al. Greedy layer-wise training of deep networks[C]//Proceedings of the Nips, Canada, 2006: 153–160.
- [52] RANZATO M A, BOUREAU Y L, LECUN Y. Sparse feature learning for deep belief networks[J]. Advances in neural information processing systems, 2007, 20: 1185–1192.
- [53] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders[C]//Proceedings of the International Conference, F, 2008.
- [54] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. Journal of machine learning research, 2010, 11(12): 3371–408.
- [55] JIANG Xiaojuan, ZHANG Yinghua, ZHANG Wensheng, et al. A novel sparse auto-encoder for deep unsupervised learning[C]//Proceedings of Sixth International Conference on Advanced Computational Intelligence. Hangzhou: IEEE, 2013: 256–261.
- [56] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–324.
- [57] WANG Wei, OOI B C, YANG Xiaoyan, et al. Effective multi-modal retrieval based on stacked auto-encoders[J]. Proceedings of the VLDB endowment, 2014, 7(8): 649–660.
- [58] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25(2): 2012.
- [59] ELMAN J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179–211.
- [60] HIHI S E, HC-J M Q, BENGIO Y. Hierarchical recurrent neural networks for long-term dependencies[J]. Advances in neural information processing systems, 1995, 8(493–9.
- [61] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735–1780.

- [62] CHO K, MERRIENBOER B V, BAHDANAU D, et al. On the properties of neural machine translation: encoder-decoder approaches[J]. Computer science, 2014,
- [63] MNH V, HEES N, GRAVES A, et al. Recurrent models of visual attention[J]. Computer science, 2014, 3(2204-12).
- [64] GOODFELLOW I, POUGETABADIE J, MIRZA M, et al. Generative adversarial Nets[J]. Advances in neural information processing systems, 2014, 2672-80.
- [65] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. Computer science, 2015,
- [66] XUE J H, TITTERINGTON D M. Comment on "on discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes"[J]. Neural processing letters, 2008, 2(3): 169-87.
- [67] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
- [68] LECUN Y, JACKEL L D, BOTTOU L, et al. Learning algorithms for classification: a comparison on handwritten digit recognition[M]//OH J H, CHO S. Neural Networks: The Statistical Mechanics Perspective. Singapore: World Scientific, 1995.
- [69] LE Q V. Building high-level features using large scale unsupervised learning[C]//Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vancouver, BC: IEEE, 2013: 8595-8598.
- [70] NYTIMES. In a big network of computers evidence of machine learning[EB/OL]. [2016-11-11].<http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html?pagewanted=all>.
- [71] SUN Yi, WANG Xiaogang, TANG Xiaou. Deep learning face representation from predicting 10,000 classes[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE, 2014.
- [72] BBC. Artificial intelligence: Google's AlphaGo beats Go master lee Se-dol[EB/OL]. 2016. [2016-11-11]. <http://www.bbc.com/news/technology-35785875>.
- [73] SILVER D, HUANG J, MADDISON C J, et al. Mastering the game of go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.
- [74] MOHAMED A R, DAHL G E, HINTON G. Acoustic modeling using deep belief networks[J]. IEEE transactions on audio, speech, and language processing, 2012, 20(1): 14-22.
- [75] PAN Jia, LIU Cong, WANG Zhiguo, et al. Investigation of deep neural networks (DNN) for large vocabulary continuous speech recognition: why DNN surpasses GMMs in acoustic modeling[C]//Proceedings of the 8th International Symposium on Chinese Spoken Language Processing (ISCSLP). Kowloon: IEEE, 2012: 301-305.
- [76] Microsoft. Microsoft audio video indexing service[EB/OL]. [2016-11-11]. <https://www.microsoft.com/en-us/research/project/mavis/>.
- [77] SEIDE F, LI Gang, YU Dong. Conversational speech transcription using context-dependent deep neural networks[C]//INTERSPEECH 2011, Conference of the International Speech Communication Association. Florence, Italy, 2011.
- [78] MORIN F, BENGIO Y. Hierarchical probabilistic neural network language model[C]//Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics, 2005.
- [79] COLLOBERT R, WESTON J. A unified architecture for natural language processing: deep neural networks with multitask learning[C]//Proceedings of the 25th International Conference on Machine Learning (ICML). NEC Laboratories America, Inc, 2008.
- [80] MNH A, HINTON G. A scalable hierarchical distributed language model[C]. Proceedings of the Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2008.
- [81] MIKOLOV T, KOMBRINK S, ČERNOCKÝ J, et al. Extensions of recurrent neural network language model[C]//Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Prague: IEEE, 2011.
- [82] MIKOLOV T, DEORAS A, POVEY D, et al. Strategies for training large scale neural network language models[C]//Proceedings of 2011 IEEE Workshop on Automatic Speech Recognition and Understanding. Waikoloa, HI: IEEE, 2011.
- [83] MIKOLOV T, ZWEIG G. Context dependent recurrent neural network language model[C]//Proceedings of 2012 IEEE Spoken Language Technology Workshop (SLT). Miami, FL: IEEE, 2012.
- [84] MIKOLOV T, KARAFIÁT M, BURGET L, et al. Recurrent neural network based language model[C]//Proceedings of the INTERSPEECH 2010, 11th Conference of the International Speech Communication Association. Makuhari, Chiba, Japan, 2010.
- [85] HUANG E H, SOCHER R, MANNING C D, et al. Improving word representations via global context and multiple word prototypes[C]//Proceedings of the Meeting of the Association for Computational Linguistics: Long Papers, F, 2012.
- [86] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient es-

timization of word representations in vector space[J]. Computer science, 2013,

[87] BAEZA-YATES R A, RIBEIRO-NETO B. Modern information retrieval: the concepts and technology behind search[M]. 2nd ed. New York: Addison Wesley, 2011: 26-28.

[89] HARRINGTON P. Machine learning in action[M]. Shelter Island, N.Y.: Manning Publications Co, 2012.

[90] 郑胤, 陈权崎, 章毓晋. 深度学习及其在目标和行为识别中的新进展[J]. 中国图象图形学报, 2014, 19(2): 175-184.

ZHENG Yin, CHEN Quanqi, ZHANG Yujin. Deep learning and its new progress in object and behavior recognition [J]. Journal of image and graphics, 2014, 19(2): 175-184.

[91] CHEN Xuewen, LIN Xiaotong. Big data deep learning: challenges and perspectives[J]. IEEE access, 2014, 2: 514-525.

作者简介:



马世龙,男,1953年生,教授、博士生导师、中国人工智能学会常务理事、中国人工智能学会人工智能基础专业委员会主任。主要研究方向为海量信息处理的计算模型、自动推理、软件工程。近年来获得 2012 年度国防科技进步二等奖等奖项,在国内外学术刊物和学术会议发表论文 160 多篇。



乌尼日其其格,女,1979年生,博士研究生,主要研究方向为云计算与大数据、计算机软件形式化方法。



李小平,男,1979年生,博士研究生,主要研究方向为云计算与大数据、计算机软件形式化方法。

第十二届中国生物特征识别大会

2017 Chinese Conference On Biometric Recognition

中国生物特征识别大会(Chinese Conference on Biometric Recognition)是由中国人工智能学会(CAAI)主办的国内生物特征识别领域的学术盛会。自 2000 年始,CCBR 已经在北京、杭州、西安、广州、济南、沈阳、天津和成都等地成功举办了 11 届,有力促进了国内本领域的学术和技术发展。

第十二届中国生物特征识别大会(CCB2017)将于 2017 年 10 月 28-29 日在深圳举行,由深圳大学计算机与软件学院和哈尔滨工业大学(深圳)计算机科学与技术学院联合承办。本届会议将汇聚国内从事生物特征识别理论与应用研究的广大科研工作者,并邀请国际同行,共同分享我国生物特征识别研究的最新理论和技术成果,为大家提供精彩的学术盛宴。

征文范围包括(不限于):

生物特征获取装置;

生物特征信号质量评价与增强;

基于生物特征的情感计算;

人脸检测、识别与跟踪;

指纹、掌纹、静脉识别;

虹膜识别;

说话人识别;

笔迹(含签名)识别;

步态识别;

其他生物特征的识别与处理;

多模态生物识别与信息融合;

生物特征数据库建设与合成;

大规模生物特征识别系统;

生物特征识别系统防伪与安全;

生物特征识别系统评估及应用。

会议网站:<http://cv.szu.edu.cn/ccbr2017/>