

深度学习研究进展

郭丽丽¹ 丁世飞^{1,2}

(中国矿业大学计算机科学与技术学院 徐州 221116)¹

(中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)²

摘 要 深度学习(Deep Learning)是一个近几年备受关注的研究领域,在机器学习中起着重要的作用。如果说浅层学习是机器学习的一次浪潮,那么深度学习作为机器学习的一个新领域,将掀起机器学习的又一次浪潮。深度学习通过建立、模拟人脑的分层结构来实现对外部输入的数据进行从低级到高级的特征提取,从而能够解释外部数据。首先介绍了深度学习的由来,分析了浅层学习存在的弊端;其次列举了深度学习的经典方法,主要以监督学习和无监督学习来展开介绍;然后对深度学习的最新研究进展及其应用进行了综述;最后总结了深度学习发展所面临的问题。

关键词 机器学习,浅层学习,深度学习,卷积神经网络,深度置信网

中图分类号 TP18 文献标识码 A DOI 10.11896/j.issn.1002-137X.2015.5.006

Research Progress on Deep Learning

GUO Li-li¹ DING Shi-fei^{1,2}

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)¹

(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)²

Abstract Deep learning plays an important role in machine learning. If shallow learning is a wave of machine learning, as a new field of machine learning, the deep learning will set off another wave of machine learning. Deep learning establishes and simulates the human brain's hierarchical structure to extract the external input data's features from lower to higher, which can explain the external data. Firstly, this paper discussed the origin of deep learning. Secondly, it described the common methods of deep learning illustrated by the example of supervised learning and unsupervised learning. Then it generalized deep learning's recent research and applications. Finally, it concluded the problems of development.

Keywords Machine learning, Shallow learning, Deep learning, CNNs, DBNs

1 引言

机器学习在人工智能领域具有举足轻重的地位,如果一个智能系统不具有学习能力,那它就不能称为一个真正的智能系统^[1]。机器学习通过算法,让机器可以从外界输入的大量的数据中学习规律,从而进行识别判断。机器学习的发展经历了浅层学习和深度学习两次浪潮。深度学习可以理解为神经网络的发展,神经网络是对人脑或生物神经网络基本特征进行抽象和建模,可以从外界环境中学习,并以与生物类似的交互方式适应环境^[2]。神经网络是智能学科的重要部分,为解决复杂问题和智能控制提供了有效的途径。神经网络曾一度成为机器学习领域备受关注的方向。

20 世纪 80 年代,美国学者 Rumelhart、Hinton 和 Williams 等完整地提出了用于人工神经网络的反向传播算法(Back Propagation 算法, BP 算法)^[3],给机器学习的研究注入了新的血液。研究人员发现,一个人工神经网络模型利用 BP 算法可以从大量的训练样本中学习统计到规律,从而便可以

对未知的事件进行预测。与过去基于人工规则的系统相比,这种基于统计的机器学习方法在很多方面都表现出优越性^[4]。但是 BP 算法在神经网络的层数增多时很容易陷入局部最优,也很容易过拟合。20 世纪 90 年代,各种浅层机器学习模型相继被提出,支持向量机(Support Vector Machines, SVM)就是这个时期产生的,无论是理论分析还是应用都很成功。然而这个时期的多层神经网络由于理论分析不够成熟,而且训练方法经验不足,因此对它的研究相对沉寂。这一现象最终于 2006 年被打破,加拿大多伦多大学教授、机器学习领域的领军人物 Hinton 和他的学生 Salakhutdinov 在著名学术刊物《科学》上发表的文章中提出了深度网络和深度学习概念^[5],开启了深度学习的研究热潮。

本文主要对深度学习进行了详细介绍,针对应用现状提出了一些问题,希望在以后的研究中有更大的突破。本文第 2 节通过列举现在算法的局限性,说明深度学习存在的必要性;第 3 节介绍了深度学习的常用方法;第 4 节总结了深度学习的研究进展及应用;最后对深度学习的研究进行了总结与展望。

到稿日期:2014-02-17 返修日期:2014-04-30 本文受国家重点基础研究发展规划(973 计划)(2013CB329502),国家自然科学基金(61379101)资助。

郭丽丽(1990—),女,硕士生,主要研究方向为神经计算、深度学习;丁世飞(1963—),男,教授,博士生导师,CCF 高级会员,主要研究方向为智能信息处理、人工智能与模式识别、机器学习与数据挖掘、粗糙集与软计算、粒度计算等, E-mail: dingsf@cumt.edu.cn。

2 深度学习的由来

2.1 人脑视觉机理启示

人类每时每刻都面临着大量的感知数据,但大脑总能很容易地捕获重要的信息。人工智能的核心问题就是模仿大脑这种高效准确地表示信息的能力。通过近些年的研究,我们对大脑机理已有了一些了解,这些都推动了人工智能的发展。神经学研究表明,人的视觉系统的信息处理是分级的,从低级的 V1 区提取边缘特征,到 V2 区的形状,再到更高层^[6,7]。人类大脑在接收到外部信号时,不是直接对数据进行处理,而是通过一个多层的网络模型来获取数据的规律^[8-10]。这种层次结构的感知系统使视觉系统需要处理的数据量大大减少,并保留了物体有用的结构信息。

2.2 现有机器学习的局限性

深度学习与浅层学习相对。现在很多的学习方法都是浅层结构算法,它们存在一定的局限性,比如在样本有限的情况下表示复杂函数的能力有限,针对复杂的分类问题其泛化能力受到一定制约^[11]。而深度学习可通过学习一种深层非线性网络结构,实现复杂函数逼近,表征输入数据分布式表示,并且能在样本集很少的情况下去学习数据集的本质特征^[12]。

例如多项式 $\prod_{i=1}^n \sum_{j=1}^m a_{ij} x_j$ 计算方式不同,计算复杂度会有很大的不同。如果计算和的积,计算复杂度为 $O(mn)$;如果计算积的和,计算复杂度为 $O(n^m)$ 。参数多的结构不仅训练复杂,训练时间长,而且泛化性也很差,还容易产生过拟合问题^[13]。虽然浅层学习的应用也很广泛,但它只对简单的计算才有效,并不能到达人脑的反应效果,这就需要深度的机器学习。这些都表明浅层学习网络有很大的局限性,激发了我们对深度网络建模的研究。

深度机器学习是数据分布式表示的必然结果。有很多学习结构的学习算法得到的学习器是局部估计算子,例如,由核方法构造的学习器, $f(x) = b + \sum_{i=1}^n a_i K(x, x_i)$, 是由对模板的匹配度加权构成的。对于这样的问题,通常我们有合理的假设,但当目标函数非常复杂时,由于需要利用参数进行描述的区域数目也是巨大的,因此这样的模型泛化能力很差。在机器学习和神经网络研究中分布式表示可以处理维数灾难和局部泛化限制^[14]。分布式表示不仅可以很好地描述概念间的相似性,而且合适的分布式表示在有限的数据下能体现出更好的泛化性能^[15]。理解和处理接收到的信息是人类认知活动的重要环节,由于这些信息的结构一般都很复杂,因此构造深度的学习机器去实现一些人类的认知活动是很有必要的。

2.3 特征提取的需要

机器学习通过算法,让机器可以从外界输入的大量数据中学习规律,从而进行识别判断。机器学习在解决图像识别、语音识别、自然语言理解等问题时的大致流程如图 1 所示。

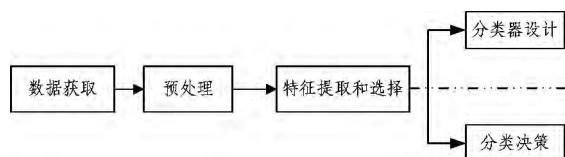


图 1 模式识别流程图

首先通过传感器来获得数据,然后经过预处理、特征提取、特征选择,再到推理、预测和识别。良好的特征表达影响着最终算法的准确性,而且系统主要的计算和测试工作都在这一环节。这个环节一般都是人工完成的,靠人工提取特征是一种非常费力的方法,不能保证选取的质量,而且它的调节需要大量的时间。然而深度学习能自动地学习一些特征,不需要人参与特征的选取过程。

深度学习是一个多层次的学习,如图 2 所示,用较少的隐含层是不可能达到与人脑类似的效果的。这需要多层的逐层学习并把学习的知识传递给下一层,通过这种方式,就可以实现对输入信息进行分级表达^[16]。深度学习的实质就是通过建立、模拟人脑的分层结构,对外部输入的声音、图像、文本等数据进行从低级到高级的特征提取,从而能够解释外部数据。与传统学习结构相比,深度学习更加强调模型结构的深度,通常含有多层的隐层节点,而且在深度学习中,特征学习至关重要,通过特征的逐层变换完成最后的预测和识别。

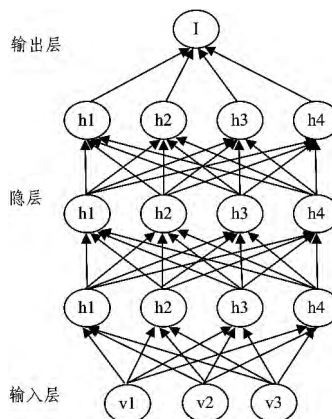


图 2 含多个隐层的深度学习示意图

3 深度学习的经典算法

深度学习作为机器学习的一个分支,其学习方法可以分为监督学习和无监督学习。两种方法都具有其独特的学习模型:多层感知机(Multilayer Perceptron)、卷积神经网络(Convolutional Neural Networks, CNNs)等属于监督学习;深度置信网(Deep Belief Nets, DBNs)、自动编码器(Auto Encoders)、去噪自动编码器(Denoising Autoencoders)^[17]、稀疏编码(Sparse Coding)等属于无监督学习。文中分别以卷积神经网络和深度置信网为例进行介绍。

3.1 监督学习:卷积神经网络

20 世纪 60 年代,Hubel 和 Wiesel 通过对猫视觉皮层细胞的研究,提出了感受野(receptive field)的概念^[18]。受此启发,Fukushima 提出神经认知机(neocognitron)可看作是 CNNs 的第一个实现网络,也是感受野概念在人工神经网络领域的首次应用^[19]。随后 LeCun 等人设计并采用基于误差梯度的算法训练了卷积神经网络^[20,21],并且其在一些模式识别任务中展现出了相对于当时其它方法的领先性能。现代生理学关于视觉系统的理解也与 CNNs 中的图像处理过程相一致^[22],这为 CNNs 在图像识别中的应用奠定了基础。CNNs 是第一个真正成功地采用多层次结构网络的具有鲁棒性的深度学习方法,通过研究数据在空间上的关联性,来减少训练

参数的数量。目前来看,在图像识别领域,CNNs 已经成为一种高效的识别方法^[23]。

CNNs 是一个多层的神经网络,如图 3 所示,每层由多个二维平面组成,每个平面又由多个独立的神经元组成^[24]。上一层中的一组局部单元作为下一层邻近单元的输入,这种局部连接观点最早起源于感知器^[25]。外界输入图像通过可训练的滤波器加偏置进行卷积,卷积后在 C1 层会产生 3 个特征映射图;然后特征映射图中每组像素分别进行求和加偏置,再通过 Sigmoid 函数得到 S2 层的特征映射图;这些映射图再通过滤波器得到 C3 层;C3 与 S2 类似,再产生 S4;最后,这些像素值被光栅化,并且连接成向量输入到神经网络,从而便得到了输出。一般地,C 层为特征提取层,每个神经元的输入与前一层的局部感受野相连,并提取该局部特征,根据局部特征来确定它与其他特征空间的位置关系;S 层是特征映射层,特征映射具有位移不变性,每个特征映射为一个平面,平面上所有神经元的权值是相等的,因而减少了网络自由参数的个数,降低了网络参数选择的复杂度。每一个特征提取层(C 层)都会跟着一个用于求局部平均及二次提取的计算层(S 层),这便构成了两次特征提取的结构,从而在对输入样本识别时,网络有很好的畸变容忍能力^[26]。对于每一个神经元,都定义了对应的接受域^[27,28],其只接受从自己接受域传来的信号。多个映射层组合起来可以获得层之间的关系和空域上的信息,从而方便进行图像处理。

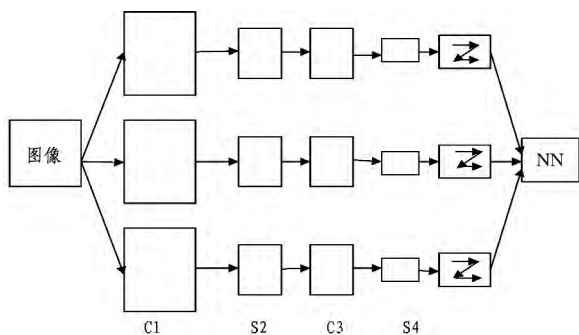


图 3 卷积神经网络结构示意图

CNNs 是人工神经网络的一种,其适应性强,善于挖掘数据局部特征。它的权值共享网络结构使之更类似于生物神经网络,降低了网络模型的复杂度,减少了权值的数量,使得 CNNs 在模式识别中的各个领域得到应用并取得了很好的结果^[29,30]。CNNs 通过结合局部感知区域、共享权重、空间或时间上的降采样来充分利用数据本身包含的局部性等特征,优化网络结构,并且保证一定程度上的位移的不变性。由 Le-Cun 提出的 LeNet 模型^[20]在应用到各种不同的图像识别任务时都取得了不错的效果,被认为是通用图像识别系统的代表之一。通过这些年的研究工作,CNNs 的应用越来越多,如人脸检测^[31]、文档分析^[32]、语音检测^[33]、车牌识别^[34]等方面。2006 年 Kussul 等人提出的采用排列编码技术的神经网络在人脸识别、手写数字识别^[35]和小物体识别等识别任务上都取得了与一些专用分类系统相当的性能表现;并且在 2012 年,研究人员把视频数据里连续的帧当作卷积神经网络的输入数据,这样就可以引入时间维度上的数据,从而识别人体的动作^[36]。

3.2 无监督学习:深度置信网

DBNs 是目前研究和应用都比较广泛的深度学习结构,

它由多个受限玻尔兹曼机(Restricted Boltzmann Machines, RBM)累加而成。RBM 结构如图 4 所示,分为可视层即输入数据层(v)和隐藏层(h),每一层的节点之间没有连接,但层和层之间彼此互连。相比传统的 sigmoid 信念网络,RBM 易于连接权值的学习。Hinton 等人^[37]认为,如果一个典型的 DBN 有 1 个隐含层,那么可以用联合概率分布来描述输入数据 v 和隐含向量的关系:

$$p(v, h^1, \dots, h^l) = \left(\prod_{k=1}^{l-2} p(h^k | h^{k+2}) \right) p(h^{l-2}, h^l) \quad (1)$$

其中, $v = h^0$, $p(h^k | h^{k+2})$ 是条件概率分布。DBN 学习的过程中,所要学习的就是联合概率分布,在机器学习的领域中,联合概率分布的意义就是对象的生成。

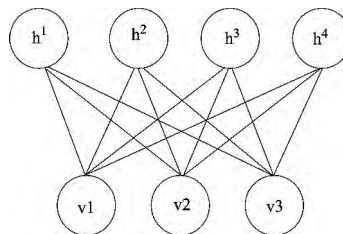


图 4 限制玻尔兹曼机结构图

传统的 BP 算法在经典的网络结构中广泛应用,但对于深度学习的训练却遇到了很多困难:第一,BP 算法是监督学习,训练需要有标签的样本集,但实际能得到的数据都是无标签的;第二,BP 算法在多隐层的学习结构中,学习过程较慢;第三,不适当的参数选择会导致局部最优解。为了获取生成性权值,预训练采用非监督贪婪逐层算法,非监督贪婪逐层训练算法被 Hinton 证明是有效的^[38]。

非监督贪婪逐层训练算法的核心思想是:把 DBN 分层(见图 5),每一层进行无监督学习,每次只训练一层,将其结果作为高一层的输入,最后用监督学习调整所有层。在这个训练阶段,首先,在可视层会产生一个向量 v ,通过它将值映射给隐单元;然后,可视层的输入会被随机地选择,以尝试去重构原始的输入信号;最后,这些新可视单元再次映射给隐单元,获得新的隐单元 h 。执行这种反复步骤叫做吉布斯(Gibbs)采样^[39]。隐层激活单元和可视层输入之间的相关性差别就作为权值更新的主要依据。在最高两层,权值被连接到一起,从而更低层的输出将会提供一个参考的线索或者关联给顶层,这样顶层就会将其联系到它的记忆内容。预训练结束后,DBN 可以利用带标签的数据及 BP 算法去调整网络结构的性能。DBNs 的 BP 算法只需要对权值参数空间进行一个局部的搜索,这相比前向神经网络来说,训练的时间会显著减少,训练 RBM 是 Gibbs 有效的随机抽样技术^[40]。在贪婪的学习算法过程中,采用了 Wake-Sleep 算法的基本思想^[41],算法在 wake 阶段,利用学习得到的权重,按照自底向上的顺序为下一层的训练提供数据;在 sleep 阶段,按照自顶向下的顺序利用权重对数据进行重组。

DBNs 是目前研究和应用都比较广泛的深度学习结构,由于灵活性很好,因此比较容易拓展,例如卷积 DBNs(Convolutional Deep Belief Networks, CDBNs)就是 DBNs 的一个拓展,给语音信号处理问题带来了突破性的进展。DBNs 作为一个新兴的生成模型,已广泛应用到了对象建模、特征提取、识别等领域。

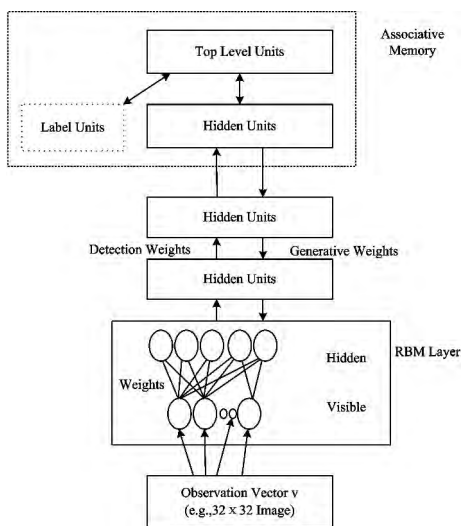


图 5 DBN 结构

4 深度学习的研究进展应用

4.1 研究近况

随着深度学习的广泛应用,很多新的算法被提出。2010 年提出的一种新颖的半监督学习算法即判别深度置信网(Discriminative Deep Belief Networks, DDBNs),被成功地应用到可视化数据分类^[42]。通过结合非监督学习的泛化能力和监督学习的判别能力,DDBN 在合成数据集以及真实世界的数据集中展示出了令人印象深刻的学习表现。文献^[43]提出一种新的深度学习算法即深凸网络(Deep Convex Network, DCN),用来解决语音识别中可扩展的挑战。2013 年国内学者又开发了一种半监督学习算法,称为卷积深度网络(Convolutional Deep Networks, CDN),用来解决深度学习中图像的分类问题^[44]。提取符合数据分布结构的特征一直是模式识别领域的热点问题,孙志军等在预训练阶段采用非监督正则化,并利用边际 Fisher 准则进一步约束提取的特征^[45],提出了基于深度学习的边际 Fisher 分析特征提取算法 DMFA(Deep Marginal Fisher Analysis)^[46],其提升了识别率。

用深层学习的算法去破译个人的思维和想法是我们面临的下一个挑战,要实现这种理解能力,首先是构建可以理解人们感情的算法,然后建立能理解多维度情感的算法。为解决自然领域的情感分析问题,许多新的算法相继被提出,例如:文献^[47]提出了一种新颖的被称为活跃深度网络(Active Deep Network, ADN)的半监督学习算法,用来解决在标记数据不足的基础上进行情感分类的问题。最近,斯坦福大学的研究生 Richard Socher 和 Andrew Ng(Google 深度学习项目工程师之一)等人共同研究开发了一个深度学习的新算法,即 Neural Analysis of Sentiment(NaSent)^[48]。NaSent 算法从人脑中得到灵感,目的是改善当前书面语言的分析方法,从而确定字里行间流露的感情。与其他算法相比,NaSent 可以更好地理解书面语言。

NaSent 旨在开发一种可在无人监督的情况下运行的算法。由于词义会随语境的不同而变化,就算是语言专家也难以准确地定义语言中的感情。深度学习模型就是为了解决这些问题。目前,应用最广的情绪分析是“词袋(bag of words)”

模型^[49],它并没有将词序列入考虑范围。词袋中的词汇有正面和负面之分,通过计数来推断整个段落含义是正面还是负面。不过有人指出将词汇单独分析的方法并不准确,必须将其放入到越来越大的结构中。Socher 和他的团队从影评网站 Rotten Tomatoes 抽取了 12000 个句子,并将其粗略分割为 214000 个词组,每个词组以数字的形式标记为负面、中立或正面,计算机科学家称这些数字化的表述为“特征表示”,这与人脑理解概念和定义的方式类似。NaSent 算法的核心就是分析和组织这些被标记的数据。NaSent 将准确率从 80% 提高到了 85%。

4.2 深度学习的应用

语音是神经网络最早的应用之一,虽然中间神经网络的研究有过沉寂,但现在神经网络在语音识别领域已取得了突破性的成果。2010 年前后,微软和 Google 的语音组都招聘了 Hinton 教授的学生去实习,他们丢弃传统的 MFCC/PLP 特征,用深度学习去学习语音信号里的特征,并且用深度学习技术进行声学模型建模,最后在标准数据集 TIMIT 上取得了很好的效果。2010 年,微软雷德蒙研究院的 Deng Li 博士与 Hinton 发现深层网络可显著提高语音识别的精度^[50]。微软亚洲研究院进一步深化这项成果,建立了一些巨大的深度神经网络(deep neural networks, DNN),这是语音识别研究史上最大的同类模型。该模型在 Switchboard 标准数据集上的识别错误率比最低错误率降低了 33%。据报道,2012 年 11 月,微软在中国天津演示了一个全自动同声传译系统,其后面支撑的关键技术也是 DNN^[51]。

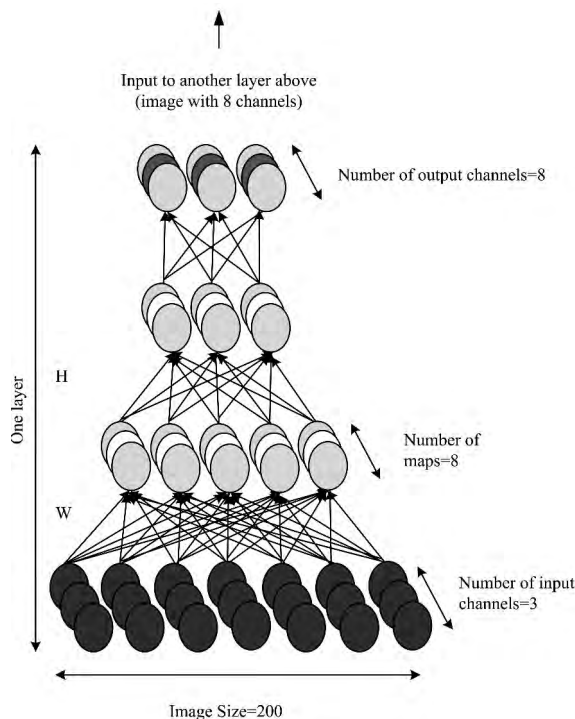


图 6 一个深度自编码器的子网络

从 2006 年深度学习被提出开始,其在目标识别领域的应用主要集中在 MNIST 手写图像问题上,并打破了 SVM 在这个数据集的霸权地位,将 1.4%^[52] 的错误率刷新到 0.27%。在最近几年,深度学习将目光从数字识别移到自然图像的目标识别^[53]。例如,谷歌研究院也投入到深度学习的研究中来,2012 年 6 月《纽约时报》报道了谷歌的 Google Brain 项

目^[54]。Google Brain 项目的指导思想是将计算机科学与神经科学相结合,这是在人工智能领域从来没有实现过的。这个项目是由著名的斯坦福大学人工智能实验室主任 Andrew Ng 和在大规模计算机系统方面的世界顶尖专家 Dean 共同主导,用 16000 个 CPU Core 的并行计算平台训练一种称为“深度神经网络”的机器学习模型,如图 6 所示,共 10 亿个参数需要学习,其可以称为最大的神经网络,使机器自主学习识别出了猫,在图像识别领域取得了惊人的效果,在 ImageNet 评测上将错误率从 26%降低到 15%。虽然,Andrew Ng 的深度学习算法在精确度和灵活性上还远不及人类的大脑,但相信总有一天会达到我们预期的效果。

除了上述应用,深度学习在自然语言处理、多任务学习等领域也有很多应用,但总体来说,没有在语音、图像上取得的进展大,但假以时日,一定会有系统性的突破。在中国,众多企业也纷纷加入深度学习的研究。2013 年 1 月,在百度的年会上,创始人兼 CEO 李彦宏宣布要成立百度研究院,其中第一个重点方向就是深度学习,并为此而成立 Institute of Deep Learning(IDL),这是百度成立 10 多年以来第一次成立研究院^[55]。2013 年 4 月,MIT Technology Review 杂志将深度学习列为 2013 十大突破性技术之首^[56]。与此同时,南非的神经科学家 Henry Markman 与科学家们合作,希望可以通过在一台计算机上进行数千次实验来模拟人脑。2013 年 10 月据国外媒体报道,Facebook 正在试图通过启用称作 Deep Learning 的新的人工智能技术来帮助理解它的用户和相关数据。Deep Learning 甚至可以仅仅基于那些并未有明确暗示的帖子来理解用户的心情或者预测将要发生的事。总之,深度学习得到了学术界和工业界的广泛关注。

结束语 本文对深度学习进行了全面分析,包括其发展缘由、算法模型、研究进展及应用等。在许多领域,深度学习都表现出了巨大的潜力,但深度学习作为机器学习的一个新领域,现在仍处于发展阶段,还有很多问题值得我们进一步深入研究:

(1)人脑的神经网络是有前馈连接和反馈连接的,但现在我们研究的深度网络中还没有加入反馈连接;而且反馈神经网络比较复杂,没有统一的训练算法,这些都给深度学习的研究带来了严峻的挑战。

(2)模型训练时间过长,如果把训练合并,并提高训练的速度,那么深度学习的实用性会大大提高。

(3)是否可以提出新的更加有效且更容易做理论分析的深度学习算法模型,探索新的特征提取模型是值得深入研究的内容;另外针对具体的问题,是否可以建立一个通用的深度模型也是值得进一步研究的。

(4)深度学习的思想就是输入信息的逐层压缩表示,即这一层的输出作为下一层的输入,通过这种方式,就可以实现输入信息的分级表达。但层数是否存在最优值以及如何确定该值,需要我们继续研究。

尽管深度学习的研究还存在许多问题,但它对机器学习领域产生的影响是不容小觑的。更加复杂且更加强大的深度模型能深刻揭示大数据里所承载的信息,并对未来和未知事件作更精准的预测。总之,深度学习是一个值得研究的领域,在未来的几年一定会更加的成熟。

参 考 文 献

- [1] 丁世飞. 人工智能[M]. 北京:清华大学出版社,2011
- [2] 史忠值. 神经网络[M]. 北京:高等教育出版社,2009
- [3] Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536
- [4] 余凯,贾磊,陈雨强. 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
- [5] Hinton G, Salakhutdinov R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507
- [6] Ding Shi-fei, Zhang Yan-an, Chen Jin-rong, et al. Research on Using Genetic Algorithms to Optimize Elman Neural Networks [J]. Neural Computing and Applications, 2013, 23(2): 293-297
- [7] Ding Shi-fei, Jia Wei-kuan, Su Chun-yang, et al. Research of Neural Network Algorithm Based on Factor Analysis and Cluster Analysis[J]. Neural Computing and Applications, 2011, 20(2): 297-302
- [8] Lee T S, Mumford D. Hierarchical Bayesian inference in the visual cortex[J]. Optical Society of America, 2003, 20(7): 1434-1448
- [9] Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(3): 411-426
- [10] Lee T S, Mumford D, Romero R, et al. The role of the primary visual cortex in higher level vision[J]. Vision Research, 1998, 38(15): 2429-2454
- [11] Bengio Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127
- [12] Bengio Y, LeCun Y. Scaling learning algorithms towards AI [M]// Bottou L, Chapelle O, Decoste D, et al. Large-Scale Kernel Machines. Cambridge: MIT Press, 2007: 321-358
- [13] 李海峰,李纯果. 深度学习结构和算法比较分析[J]. 河北大学学报:自然科学版, 2012, 32(5): 538-544
- [14] Hinton G E. Learning distributed representations of concepts [C]//Proc. of the 8th Annual Conference of the Cognitive Science Society, 1986: 1-12
- [15] 孙志军,薛磊,许阳明. 深度学习研究综述[J]. 计算机应用研究, 2012, 29(8): 2806-2810
- [16] Bengio Y, Delalleau O. On the expressive power of deep architectures[C]//Proceedings of the 22nd International Conference on Algorithmic Learning Theory, Berlin Heidelberg, 2011: 18-36
- [17] Vincent P, Larochelle H, Jajoe 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): 3371-3408
- [18] Hubel D H, Wiesel T N. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex [J]. Journal of Physiology, 1962, 160: 106-154
- [19] Fukushima K. Neocognition: A Self-Organizing Neural Network Model for a mechanism of Pattern Recognition Unaffected by Shift in Position[J]. Biological Cybernetics, 1980, 36: 193-202
- [20] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of IEEE, 1988, 86(11): 2278-2324

- [21] LeCun Y, Boser B, Denker J S, et al. Backpropagation Applied to Handwritten Zip Code Recognition[J]. *Neural Computation*, 1989, 1(4): 541-551
- [22] Serre T, Keriman G, Kouch M. A Quantitative Theory of Immediate Visual Recognition” *Progress in Brain Research, Computational Neuroscience*[J]. *Theoretical Insights into Brain Function*, 2007, 165: 33-56
- [23] LeCun Y, Kavukcuoglu U K, Farabe C. Convolutional networks and applications in vision[Z]. *International Symposium on Circuits and Systems*, Paris, 2010
- [24] Kwolek B. Face Detection Using Convolutional Neural Networks And Gabor Filters[J]. *Artificial Neural Networks: Biological Inspirations*, 2005, 3699: 551-556
- [25] Rosenblatt F. The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain[J]. *Psychological Review*, 1958, 65: 386-408
- [26] Neubauer C. Shape, position and size invariant visual pattern recognition based on principles of neocognitron and perception in artificial neural networks[M]. *Netherlands: North Holland*, 1992: 833-837
- [27] Laserson J. From neural networks to deep learning: zeroing in on the human brain[J]. *ACM Crossroads Student Magazine*, 2011, 18(1): 29-34
- [28] LeCun Y, Botton L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of IEEE*, 1998, 86(11): 2278-2324
- [29] Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders[C]// *Proceedings of the 25th International Conference on Machine Learning (ICML'2008)*. New York: ACM Press, 2008: 1096-1103
- [30] Huang Fu-jie, LeCun Y. Large-scale learning with SVM and convolutional for generic object categorization[C]// *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington DC: IEEE Computer Society, 2006: 284-291
- [31] 聂仁灿, 姚绍文, 周冬明. 基于简化脉冲耦合神经网络的人脸识别[J]. *计算机科学*, 2014, 41(2): 297-301
- [32] Simard P Y, Steinkraus D, Platt J C. Best Practice for Convolutional Neural Networks Applied to Visual Document Analysis [C]// *Seventh International Conference on Document Analysis and Recognition*. 2003: 963-985
- [33] Sukittanon S, Surendran A C, Burges C J C, et al. Convolutional Networks for speech Detection[OL]. http://131.107.65.14/pubs/68033/convnet_speechdetect.pdf
- [34] Chen Y, Han C, Wang C, et al. The application of a convolution neural network on face and license plate detection[C]// *18th International Conference on Pattern Recognition (ICPR)*. Hong Kong, China: IEEE Computer Society, 2006: 552-555
- [35] 赵元庆, 吴华. 多尺度特征和神经网络相融合的手写体数字识别[J]. *计算机科学*, 2013, 40(8): 316-318
- [36] Ji S, Xu W, Yang M, et al. 3D Convolutional Neural Networks for Human Action Recognition[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 221-231
- [37] Hinton G E. Distributed representations[R]. *Tech. Report*, University of Toronto, 1984
- [38] Hinton G E, Osindero S. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006, 18: 1527-1554
- [39] Lawrence C E, Altschul S F, Boguski M S, et al. Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment[J]. *Science*, 1993, 262: 208-214
- [40] Bishop C M. *Pattern recognition and machine learning*[M]. New York: Springer, 2006
- [41] Hinton G E, Dayan P, Frey B, et al. The wake-sleep algorithm for unsupervised neural network[J]. *Science*, 1995, 268: 1158-1161
- [42] Liu Yan, Zhou Shu-sen, Chen Qing-cai. Discriminative deep belief networks for visual data classification[J]. *Pattern Recognition*, 2010, 44(10): 2287-2296
- [43] Yu Dong, Deng Li. Deep convex net: a scalable architecture for speech pattern classification[C]// *Proc. of the 12th Annual Conference of International Speech Communication Association*. 2011: 2296-2299
- [44] Zhou Shu-sen, Chen Qing-cai, Wang Xiao-long. Convolutional Deep Networks for Visual Data Classification [J]. *Neural Process Lett*, 2013, 38: 17-27
- [45] Wong W K, Sun M M. Deep learning regularized fisher mappings [J]. *IEEE Transactions on Neural Networks*, 2011, 22(10): 1668-1675
- [46] 孙志军, 薛磊, 许阳明. 基于深度学的边际 Fisher 分析特征提取算法[J]. *电子与信息学报*, 2013, 35(4): 805-811
- [47] Zhou Shu-sen, Chen Qing-cai, Wang Xiao-long. Active deep learning method for semi-supervised sentiment classification[J]. *Neurocomputing*, 2013, 120: 536-546
- [48] Tom A. Stanford algorithm analyzes sentence sentiment, advances machine learning [N]. *Stanford University*, 2013
- [49] Li Fei-fei, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories[J]. *Computer Vision and Image Understanding*, 2004, 106(1): 59-70
- [50] Hinton G E, Li Deng, Dong Yu, et al. Deep neural networks for acoustic modeling in speech recognition [J]. *IEEE Signal Processing Magazine*, 2012, 29(6): 82-97
- [51] Markoff J. Scientists see promise in deep-learning programs [N]. *The New York Times*, 2012-11-23
- [52] Bach F, Jenatton R, Obozinski G. Structured sparsity through convex optimization[OL]. <http://arxiv.org/pdf.1109.2397/pdf>
- [53] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1798-1828
- [54] Markoff J. How many computers to identify a cat? [N]. *The new York Times*, 2012-06-25
- [55] 李彦宏. 2012 百度年会主题报告: 相信技术的力量[R]. 北京: 百度, 2013
- [56] 10 Breakthrough Technologies 2013 [N]. *MIT Technology Review*, 2013-04-2