

DOI:10.11992/tis.201511028

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20160715.1353.002.html>

深度学习研究方法研究新进展

刘帅师,程曦,郭文燕,陈奇

(长春工业大学 电气与电子工程学院,吉林 长春 130000)

摘 要:本文依据模型结构对深度学习进行了归纳和总结,描述了不同模型的结构和特点。首先介绍了深度学习的概念及意义,然后介绍了 4 种典型模型:卷积神经网络、深度信念网络、深度玻尔兹曼机和堆叠自动编码器,并对近 3 年深度学习在语音处理、计算机视觉、自然语言处理以及医疗应用等方面的应用现状进行介绍,最后对现有深度学习模型进行了总结,并且讨论了未来所面临的挑战。

关键词:深度学习;卷积神经网络;深度信念网络;深度玻尔兹曼机;堆叠自动编码器

中图分类号:TP18 **文献标志码:**A **文章编号:**1673-4785(2016)05-0567-10

中文引用格式:刘帅师,程曦,郭文燕,等.深度学习研究方法研究新进展[J].智能系统学报,2016,11(5):567-577.

英文引用格式:LIU Shuaishi, CHENG Xi, GUO Wenyan, et al. Progress report on new research in deep learning[J]. CAAI transactions on intelligent systems, 2016,11(5):567-577.

Progress report on new research in deep learning

LIU Shuaishi, CHENG Xi, GUO Wenyan, CHEN Qi

(College of Electrical and Electronic Engineering, Changchun University of Technology, Changchun 130000, China)

Abstract: Deep learning has recently received widespread attention. Using a model structure, this paper gives a summarization and analysis on deep learning by describing and reviewing the structure and characteristics of different models. The paper firstly introduces the concept and significance of deep learning, and then reviews four typical models: a convolutional neural network; deep belief networks; the deep Boltzmann machine; and an automatic stacking encoder. The paper then concludes by reviewing the applications of deep learning as regards speech processing, computer vision, natural language processing, medical science, and other aspects. Finally, the existing deep learning model is summarized and future challenges discussed.

Keywords: deep learning; convolutional neural network; deep belief networks; deep Boltzmann machine; automatic stacking encoder

深度学习是机器学习的一个分支,属于人工智能的新领域。深度学习的本质是特征提取,即通过组合低层次的特征形成更加抽象的高层表示,以达到获得最佳特征的目的^[1]。它主要通过神经网络来模拟人的大脑的学习过程,希望实现对现实对象或数据(图像、语音及文本等)的抽象表达,整合特征抽取和分类器到一个学习框架下^[2]。目前,深度学习在许多领域取得了广泛的关注,成为当今的研

究热点。

2006 年,机器学习大师 Hinton 等^[3]在《科学》上发表的一篇论文,开启了深度学习的浪潮。他提出了深度信念网的概念,成功地利用贪心策略逐层训练由限制玻尔兹曼机组成的深层架构,解决了以往深度网络训练困难的问题。此后,Hinton、Lecun、Bengio 等大量科研人员对深度学习的模型构建、训练方式等做出了杰出的贡献。文献[4]对近几年新兴的深度学习的初始化方法、模型结构、学习算法等进行了详细的分析。2014 年余滨等^[5]从训练方式的角度对深度学习进行了总结。文献[6]依据数据流向对深度学习进行不同分类,本文将重点放在模

收稿日期:2015-11-27. 网络出版日期:2016-07-15.

基金项目:吉林省科技厅青年科研基金项目(20140520065JH, 20140520076JH);长春工业大学科学研究发展基金自然科学基金项目(2010XN07).

通信作者: E-mail:刘帅师. E-mail:liu-shuaishi@126.com.

型结构,以深度学习的核心模型结构而展开。文献[7]从深度学习结构进行展开,本文大量增加了近3年新的研究成果,因此能够更准确地反映该领域的最新研究进展。

本文依据模型结构,着重介绍4种典型的深度学习模型,即卷积神经网络、深度信念网络、深度玻尔兹曼机、堆叠自动编码器。下面对这些模型进行描述。

1 深度学习典型模型

1.1 卷积神经网络

1.1.1 历史

卷积神经网络(convolutional neural networks, CNN)最早在20世纪80年代就已被提出,是由于在研究猫脑皮层时受到的启发。它的典型模型LeNet-5^[8]系统,在MNIST上得到了0.9%的错误率,并在20世纪90年代就已用于银行的手写支票识别^[7]。但是,由于在大尺寸图像上没有好的效果,一度被人忽视。随着高效的GPU计算的兴起,直到2012年Hinton在ImageNet问题的成功,才使它在近几年流行起来^[9]。如今卷积神经网络已经成为众多科学领域的研究热点之一,特别是在图像识别领域。由于该网络避免了对图像的复杂前期预处理,可以直接输入原始图像,因而得到了广泛的应用。

1.1.2 结构

卷积神经网络是前馈神经网络的一种。卷积神经网络的模型如图1所示,它是一个多层的神经网络,每层由多个二维平面组成,而每个平面由多个独立神经元组成。传统的神经网络层与层之间神经元采取全连接方式,而卷积神经网络采用稀疏连接方式,即每个特征图上的神经元只连接上一层的一个小区域的神经元连接。

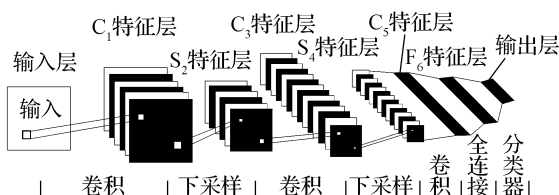


图1 卷积神经网络模型

Fig.1 Convolutional neural network model

卷积神经网络的低隐含层是由卷积层和最大池采样层交替组成,高层通常是全连接层作为分类器

使用。

首先,为了降低网络的复杂性,卷积神经网络采用权重共享方式,即同一个特征图,卷积核是一样的^[10];其次,对得到的特征输入给一个非线性函数,比如ReLU等;最后,再采取下采样方法,比如最大池化等。下采样的作用是把语义上相似的特征合并起来,这是因为形成一个主题的特征的相对位置不太一样^[11]。

1.1.3 训练方式

卷积神经网络的训练采用有监督训练方式。首先是向前传播,即输入 X 经过卷积神经网络后变为输出 O ,再将 O 与标签进行比较,然后以向后传播的方式,到将所得误差传播到每个节点,根据权值更新公式,更新相应的卷积核权值^[4,6]。

此外,以卷积神经网络为核心的深度学习网络还有3-D卷积神经网络(3-D convolutional neural networks, 3D-CNN)^[11],光谱网络(spectral networks, SN)^[12],金字塔卷积神经网络(pyramid convolutional neural networks, PCNN)^[13],多级金字塔卷积神经网络(multi level pyramid convolutional neural networks, MLPCNN)^[14]等。

1.2 受限制玻尔兹曼机为核心的深度网络

受限制玻尔兹曼机为核心的深度网络有2种:深度信念网和深度玻尔兹曼机。现简要分析二者的区别。

1.2.1 受限制玻尔兹曼机

受限制玻尔兹曼机(restricted Boltzmann machine, RBM)是一类无向图模型^[4],由可视层和隐含层组成,与玻尔兹曼机(Boltzmann machine, BM)不同,层内无连接,层间有连接。这种结构更易于计算隐含层单元与可视层单元的条件分布^[5]。受限制玻尔兹曼机的训练方式通常采用对比散度方(contrastive divergence, CD)。常见的玻尔兹曼机的演变模型如图2所示。

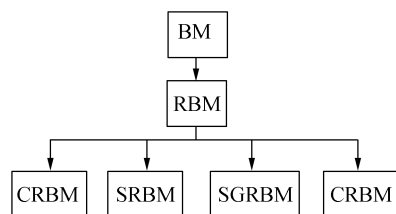


图2 玻尔兹曼机演变模型

Fig.2 The evolution model of boltzmann machine

受限制玻尔兹曼机的演变模型有卷积受限制玻尔兹曼机 (convolutional restricted Boltzmann machine, CRBM)^[15]、稀疏受限制玻尔兹曼机 (sparse restricted Boltzmann machine, SRBM)^[16]、稀疏组受限制玻尔兹曼机 (sparse group restricted Boltzmann machine, SGRBM)^[17]、分类受限制玻尔兹曼机 (class restricted Boltzmann machine, CRBM)^[18] 等。更详细的描述内容参见文献[19]。

1.2.2 深度信念网络

深度信念网络 (deep belief networks, DBN) 是由多个受限制玻尔兹曼机 (RBM) 叠加而成的深度网络。深度信念网络的典型结构如图 3 所示, 它通过无监督预训练和有监督微调来训练整个深度信念网络^[7]。预训练时用无标签数据单独训练每一层受限制玻尔兹曼机, 通过自下而上的方式, 将下层受限制玻尔兹曼机输出作为上层受限制玻尔兹曼机输入。当预训练完成后, 网络会获得一个较好的网络初始值, 但这还不是最优的^[20]。再采用有标签数据去训练网络, 误差自顶向下传播, 一般采用梯度下降法对网络进行微调。深度信念网络的出现是深度学习的转折点, 目前深度信念网络已应用于语音、图像处理等方面, 尤其是在大数据方面^[10]。

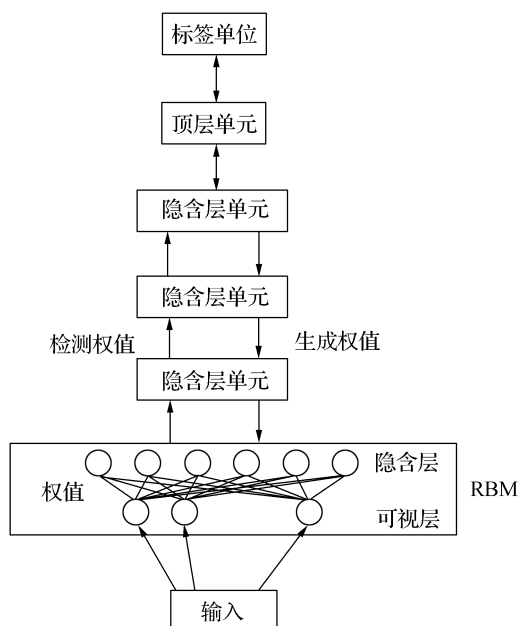


图3 深度信念网络典型结构

Fig.3 The illustration of deep belief network framework

深度信念网络的变种模型有卷积深度信念网 (convolutional deep belief networks, CDBN)^[21]、稀疏深度信念网 (sparse deep belief networks, SDBN)^[22]、判别深度信念网 (discriminative deep belief networks,

DDBN)^[23] 等。

1.2.3 深度玻尔兹曼机

深度玻尔兹曼机 (deep Boltzmann machine, DBM) 与深度信念网络相似, 都是以受限制玻尔兹曼机叠加而成。但是, 与深度信念网络不同, 层间均为无向连接, 省略了由上至下的反馈参数调节。训练方式也与深度信念网络相似, 先采用无监督预训练方法, 得到初始权值, 再运用场均值算法, 最后采用有监督微方式进行微调。

1.3 堆叠自动编码器

1.3.1 自动编码器

自动编码器 (auto encoder, AE) 由编码器与解码器组成, 其原理如图 4 所示。核心思想是将输入信号进行编码, 使用编码之后的信号重建原始信号, 目的是让重建信号与原始信号相比重建误差最小^[24]。编码器将输入数据映射到特征空间, 解码器将特征映射回数据空间, 完成对输入数据的重建。

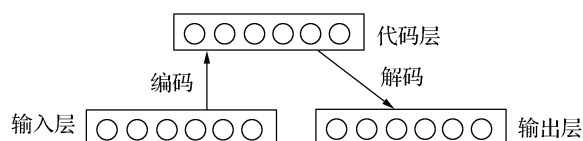


图4 自动编码器原理图

Fig.4 The illustration of autoencoder

自动编码器演化的模型如图 5 所示, 自动编码器演化的模型有去噪自动编码器 (denoising auto encoder, DAE)^[25]、稀疏自动编码器 (sparse auto encoder, SAE)^[26]、收缩自动编码器 (contractive auto encoder, CAE)^[27]、卷积自动编码器 (convolutional auto encoder, CAE)^[28] 等。更详细的描述内容参见文献[29-31]。

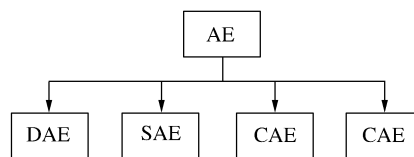


图5 自动编码器的演变模型

Fig.5 The evolution model of autoencoder

1.3.2 堆叠自动编码器原理

堆叠自动编码器 (stacked auto encoders, SAE) 与深度信念网络类似, 其结构如图 6 所示, 都是由简单结构叠加起来的深层网络。简单来说, 就是将 DBN 中的 RBM 替换成 AE 就得到了 SAE。自动编码器的训练过程也是使用贪心逐层预训练算法, 但

因为是通过重构误差来进行训练,相比较而言比受限玻尔兹曼机训练容易^[7]。

以自动编码器及其变种模型为核心的深度网络称之为深度自动编码(deep auto encoders, DAE)。文献[32]对深度自动编码器进行了详细的描述。堆叠自动编码器就是一种典型的深度自动编码。类似的还有堆叠去噪自动编码器(stacked denoising auto-encoders, SDAE)^[25]、堆叠稀疏自动编码器(stacked sparse auto encoders, SSAE)^[33]等。

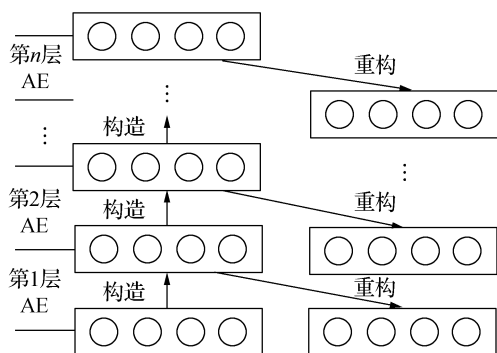


图 6 堆叠自动编码器结构

Fig.6 The illustration of stacked Autoencoder framework

2 深度学习应用

2.1 语音处理

长期以来,语音识别技术普遍采用的是声学模型混合高斯模型。但这种混合高斯模型本质上是一种浅层网络建模,不能充分描述特征的状态空间分布^[34]。2011年微软^[35]将深度学习引入语音识别领域,提出深度神经网络 DNN,本质上是把混合高斯模型替换成了深度神经网络模型大大提高识别率。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%。2014年 Van 等^[36]在网络音乐平台 Spotify 使用深度卷积神经网络做基于内容的音乐推荐,以及实现依靠音频信号预测听众的收听喜好,然后采用 WMF(weighted matrix factorization)模型进行评分预测。百度的深度学习语音识别系统 DeepSpeech^[37]可以在饭店等嘈杂环境下实现将近 81% 的辨识准确率。2015年 Chan 等^[38]提出了 LAS(listen, attend and spell)系统。该系统利用金字塔式双向的 RNN 网络。不同于传统的模型,不需要完整的端对端的 CTC(connectionist temporal classification),实现跳过音素直接把语音识别为字符,合成了约 4 万小时音频。当不依赖语音词典和

语言模型时在谷歌语音搜索任务词错率达到 14.2%,当结合语言模型时词错率达到 11.2%。

2.2 计算机视觉

2012年 Hinton^[39]和他的 2 个学生在著名的 ImageNet 问题上用更深的卷积神经网络取得世界最好结果,赢得了冠军,使得图像识别大踏步前进。随后,深度学习在图像处理方面取得突破性进展,如物体定位^[40]、脸部识别^[41]和人体姿势估计^[42]等。2015年深海团队^[43]利用卷积神经网络对 3 万个例子进行 121 种浮游生物分类。Denton 等^[44]利用卷积神经网络通过用户的性别、年龄、城市和图片等特征进行整合。在 Facebook 上对用户上传的图片进行标签、分类。

2.2.1 图像语义分割

图像包括很多层信息,例如这幅图像是否有特定的物体(如汽车)。所谓的图像语义分割,就是描述图片中包含哪些物体、包括街头的场景分割、三维扫描、对 3-D 人体解剖分割定位等。因为是基于像素分类方法,所以这个问题会带来巨大的计算量。2015年 Long 等^[45]提出了利用全卷积网络(fully convolutional networks, FCN,)的概念去进行图像语义分割。同年 Behnke 等^[46]提出了语义 RGB-D 感知器的概念,使得基于深度学习的图像语义分割得到进一步发展。

2.2.2 人脸识别

Linkface 开发了基于深度学习的人脸检测创新算法。无论场景中是单人还是多人,是侧脸、半遮挡还是模糊等情景中,均能进行精准检测。据全球最具权威的人脸检测评测平台 Fddb 最新数据,Linkface 的人脸检测算法达到了世界领先的水平。2014年 Facebook 开发一种叫 DeepFace^[41]技术。其贡献在于对人脸对齐和人脸表示环节的改进。通过革新的 3-D 人脸建模勾勒出脸部特征,然后通过颜色过滤做出一个刻画特定脸部元素的平面模型。Facebook 建立了一个来自于 4 030 个人的 440 万张标签化的人脸池,Facebook 称这是迄今为止最大规模的人脸池。它是一个拥有 9 层的深度卷积神经网络,网络有超过 1.2 亿个参数。该技术在 LFW 数据集上取得了 97.25% 的平均精度,已经接近人类的识别水平。2015年 Google 提出 FaceNet^[47]进行人脸验证。它直接学习图像到欧式空间上点的映射,然后基于这个编码再做人脸识别、人脸验证和人脸聚

类等。其中两张图像所对应的特征欧式空间上的点之间的距离直接对应着两个图像是否相似。FaceNet 并没有像 DeepFace 和 DeepID 那样需要对齐。FaceNet 得到最终表示后不用像 DeepID 那样需要再训练模型进行分类,直接计算距离就可以,简单而有效。在 Youtube 数据集上测试准确率为95.12%。

目前,传统人脸识别技术主要集中在可见光谱的范畴,对于跨模态人脸识别问题尚无好的解决方法。2015 年 Sarfraz 等^[48]利用深度神经网络,成功将红外热图像与可见光图像进行匹配,实现了跨模态人脸匹配。该网络可以在短短35 ms的时间内,能够将红外热图像匹配到其可见光图像,可以实现实时运行。

2.2.3 表情识别

目前,大部分研究者把卷积神经网络应用在表情识别上。例如,2013 年 Liu 等^[49]提出了构建一个新的深层结构(AU-aware deep networks, AUDN),基于卷积神经网络进行特征提取,连接 SVM 做表情分类器。2014 年 Ouellet 等^[50]使用卷积神经网络对电脑前的游戏玩家进行实时表情识别。Song 等^[51]利用了一种5层卷积神经网络,实现了每幅图像在服务器的预测时间为50 ms,每个图像的往返时间小于100 ms,在智能手机上实现实时表情识别。Ijina 等^[52]用 Kinect 深度传感器得到的图片作为表情识别的对象,并在卷积神经网络进行表情识别取得了较好的效果。Byeon 等^[53]使用3D卷积神经网络去识别视频人脸表情。文献^[54]证明在实时表情识别系统,卷积神经网络比深层神经网络具有更好的效果。

还有一部分研究者利用其他深度学习模型进行表情识别。例如,McLaughlin 等^[55]提出一种基于深度信念网络的实时表情识别系统,但只能检测4种表情。2013 年 He 等^[56]利用深度玻尔兹曼机对红外热图像进行表情识别。

此外,一些研究者们将多种深度学习模型结合起来进行表情识别。例如,2014 年 LYU 等^[57]将深度信念网络与自编码器相结合来进行识别。2015 年 Jung 等^[58]将卷积神经网络与深度神经网络合起来。Kahou 等^[59]提出一种视频表情识别系统 EmoNets。卷积神经网络捕捉视频信息,检测人脸。深度信念网络捕捉音频信息,自编码器捕捉人肢体行为。该理论赢得了2013 EmotiW 挑战赛,在2014

的数据集上准确率达到47.67%。

2.3 自然语言处理

Sasihithlu 等^[60]采用递归自编码方法(recursive auto encoders, RAE)来解决较为复杂的情感分析问题。Johnson 等^[61]提出一种基于卷积神经网络直接在词袋模型(BoW)上用做文本分类任务。2015 年谷歌的 Good 等利用深度神经网络开发了字镜头(word lens)实时视频翻译性能和通话实时翻译功能。它可以实现拿着手机摄像头对着实物,实物中的文字就可被即时识别出,并被翻译成目标语言,目前该技术可支持20多种语言的即时视觉翻译。更重要的是即使它在不联网的状态下也能进行工作,所有深度学习的庞大计算都是在手机上完成的。李婷等^[62]利用堆叠去噪自动编码器(stack denoising auto encoder, SDAE)识别盲文。

2.4 医疗应用

Deep Genomics 公司开始把基因组和深度学习结合起来,Deep Genomics 已经推出了他们的第一款产品 SPIDEX。只需将测试结果和细胞类型导入,SPIDEX 便可分析出某一变异对 RNA 剪切的影响,并计算出该变异与疾病之间的关系。Kozioł 等^[63]利用一种受限玻尔兹曼机用于肝细胞癌的分类。2015 年 Fauw 等^[64]利用20多层的卷积神经网络检测糖尿病视网膜病变的眼底图像。

3 模型总结及面临的挑战

3.1 深度学习模型

本文对深度学习模型进行分类、概括,在此以模型的结构为序,对深度学习模型进行总结如表1~表3所示。

1)模型结构。目前,大部分的深度学习模型都是以卷积神经网络、深度信念网络、深度玻尔兹曼机、堆叠自动编码器等几种基本模型为基础演变而来。除此之外,还有像递归神经网络(recurrent neural networks, RNN)^[74]、深度凸形网络(deep convex net, DCN)^[75]等其他类型的新型深度模型。

2)训练方式。深度学习模型的训练方式主要有有监督学习和无监督学习2种。训练方式因模型结构而异,一般以卷积神经网络为核心的模型一般采取有监督训练方式。而以受限制玻尔兹曼机与自动编码器为核心的模型,大部分采用无监督学习方式预训练,配合有监督微调模式进行参数训练。

表 1 典型深度学习模型
Table 1 Typical deep learning model

序号	模型名称	模型描述	常用训练方式	适用范围	作者,文献,年份
1	卷积神经网络 (CNN)	多个卷积层和下采样层交替组成,以及一个全连接层	有监督,运用前向传播和后向传播(BP)相结合方式	图像处理,自然语言处理,视频分析	Lecun Y, [65], 1990
	3D 卷积神经网络 (3D-CNN)	适用于视频中,通过堆叠多个连续的帧组成一个立方体,然后在立方体中运用 3D 卷积核,捕捉在时间和空间维度都具有区分性的特征	有监督,通过随机对角 Levenberg-Marquardt 法来优化训练	视频分析	Ji S, [66], 2013
	金字塔卷积神经网络 (PCNN)	与 CNN 不同,是 Siamese 网络,由多个金字塔组成	有监督逐层训练	图像处理	Fan H, [13], 2014
	深度信念网络 (DBN)	多个 RBM 累加而成的深度网络	无监督贪婪逐层预训练,有监督微调(BP)	语音识别,图像处理,自然语言处理	Hinton G E, [67], 2006
2	卷积深度信念网络 (CDBN)	将 DBN 中 RBM 替换成 CRBM	无监督贪婪逐层预训练,有监督微调	图像处理,语音信号处理	Lee H, [21], 2011
	稀疏深度信念网络 (SDBN)	将 DBN 中 RBM 替换成 SRBM	无监督贪婪逐层预训练,有监督微调	图像处理	Halkias X C, [22], 2014
3	深度玻尔兹曼机 (DBM)	多个 RBM 累加而成,与 DBN 不同层间均为无向连接,省略了由上至下的反馈参数调节	无监督贪婪逐层预训练,再进行场均值算法,有监督微调(BP)	图像处理	Salakhutdinov R, [68], 2009
	卷积深度玻尔兹曼机 (CDBM)	将 DBM 中 RBM 替换成 CRBM	无监督贪婪逐层预训练,再进行场均值算法,有监督微调	图像处理	杜骞, [69], 2014
4	堆叠自动编码器 (SAE)	多个 AE 累加而成	无监督贪婪逐层预训练,再将网络展开,有监督微调(BP)	图像处理	Bengio Y, [70], 2009
	堆叠去噪自动编码器 (SDAE)	将 SAE 中 AE 替换成 DAE	无监督贪婪逐层预训练,再将网络展开,有监督微调(BP)	图像处理,自然语言处理	Vincent P, [25], 2008

表 2 玻尔兹曼机及其演化模型
Table 2 Boltzmann machine and its evolution model

序号	模型名称	模型描述	作者,文献,年份
1	玻尔兹曼机(BM)	随机神经元组成的二值随机机器,一种随机递归神经网络	Hinton G E,[71],1986
2	受限制玻尔兹曼机(RBM)	与 BM 结果相同,只是没有层间的连接	Smolensky P,[72],1986
3	卷积受限制玻尔兹曼机(CRBM)	三层两部无向图模型,由可视层、检测层、池层组成	Lee H,[15],2009
4	稀疏受限制玻尔兹曼机(SRBM)	在 RBM 的对数似然目标函数上,增加了一个稀疏惩罚项	Lee H,[16],2008
5	稀疏租受限制玻尔兹曼机(SGRBM)	将组稀疏方法应用到 RBM	罗恒,[17],2011
6	分类受限制玻尔兹曼机(CRBM)	利用包含二值随机变量的隐单元来拟合输入特征与类标签的联合分布	Larochelle H,[18],2008

表 3 自动编码器及其演化模型
Table 3 Automatic encoder and its evolution model

序号	模型名称	模型描述	作者,文献,年份
1	自动编码器(AE)	多层神经网络,从输入层到隐藏层为编码器,从隐藏层到输出层为解码器	Rumelhart D E,[73],1986
2	去噪自动编码器(DAE)	在 AE 的输入层引入随机噪声	Vincent P,[25],2008
3	稀疏自动编码器(SAE)	AE 加入稀疏性限制	Bengio Y,[26],2007
4	收缩自动编码器(CAE)	对 AE 的升维和降维过程加以限制	Rifai S,[27],2011
5	卷积自动编码器(CAE)	AE 每个隐含层的节点都是用来做卷积	Masci J,[28],2011

3)应用领域。深度学习在语音处理、计算机视觉的应用已十分广泛,许多技术已用于商用。但是,在自然语言处理的应用尚不成熟^[1]。一些研究者尝试用递归神经网络去解决这一问题。文献[74-78]对递归神经网络在文本生成和机器翻译的应用做出了详细的描述。目前,递归神经网络的变种模型长短时记忆模型(long short-term memory,LSTM)被证明比传统的递归神经网络更加有效^[79]。

3.2 面临的挑战

1)模型结构创新。自 Hinton 提出深度学习的思想以来,已经涌现出大量的深度学习模型,然而大部分的模型的构建依旧停留在以简单模型(如 AE,RBM 等)叠加而成的深度网络,或是几种深度学习模型简单相叠加,来构建深度学习模型。这种形式的模型往往不能发挥深度学习的优势,是否存在其他有效的深度学习模型,是否可以让深度学习与其他方法进行融合,这是今后要研究的问题。

2)训练方式的改进。深度学习已经在各个领域取得了突破性的成果,大部分深度学习模型均采用无监督学习方式。但是,离完全的无监督学习还有一定的距离。目前的深度学习模型在无监督预训练后,仍然需要有监督的微调,并没有做到完全意义

上的无监督学习。因此,如何做到完全意义上的无监督学习是未来研究的重点。

3)减少训练时间。当待解决的问题过于复杂,使深度学习模型参数增加时,会导致模型的训练时间逐渐上升,是否可以在不改变硬件性能的条件下,对算法进行改进,在保证精度的同时,提高训练速度。所以,减少训练时间,仍是深度学习需要努力的研究方向。

4)实现在线学习。目前,深度学习的算法大多采用无监督预训练与有监督微调配合的方式进行。然而,一旦在线环境下引入全局微调,会使结果陷入局部最小值。因此,这种训练算法不利于在线学习。是否可以改进算法进而将深度学习应用于在线环境,这是未来要思考的问题。

5)克服对抗样本。通过稍微修改实际样本,而构造出的合成样本,会使一个分类器以高置信度认为它们属于错误的分类,这就是深度学习对抗样本问题^[80-82]。研究如何克服它们可以帮助我们避免潜在的安全问题。然而,目前为止并没有好的方法出现。一些研究人员尝试使用常见的正则化方法(包括均化多重模型、均化图像多采样观测等)去解决这一问题,但是并没有取得良好的进展。因此,深

度学习的对抗样本问题仍然是待解决的难题之一。

4 结束语

本文详细描述了儿种典型的深度学习模型的构造原理,以及训练方法。并且,对近 3 年深度学习在各个领域的应用进行了概括。最后,在现有深度学习模型的基础上讨论了深度学习面临的挑战。

深度学习自提出以来已经在许多领域取得了突破性的进展。但是,在深度学习实际应用的过程中,往往为了要构造合适的深度学习模型而大费周章。因为目前的深度学习模型大部分是为了解决某一特定问题,而量身定做的。如果用于解决其他问题,效果往往不尽如人意。今后研究者们能否通过改进结构与算法,研究出一种可以应对大部分问题深度学习算法,这是未来要思考的难题。

参考文献:

- [1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [2] 林妙真. 基于深度学习的人脸识别研究[D]. 大连: 大连理工大学, 2013.
LIN Miaozhen. Research on face recognition based on deep learning[D]. Dalian, China: Dalian University of Technology, 2013.
- [3] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [4] 刘建伟, 刘媛, 罗雄麟. 深度学习研究进展[J]. 计算机应用研究, 2014, 31(7): 1921-1930, 1942.
LIU Jianwei, LIU Yuan, LUO Xionglin. Research and development on deep learning[J]. Application research of computers, 2014, 31(7): 1921-1930, 1942.
- [5] 余滨, 李绍滋, 徐素霞, 等. 深度学习: 开启大数据时代的钥匙[J]. 工程研究-跨学科视野中的工程, 2014, 6(3): 233-243.
YU Bin, LI Shaozi, XU Suxia, et al. Deep learning: a key of stepping into the era of big data[J]. Journal of engineering studies, 2014, 6(3): 233-243.
- [6] 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.
YIN Biaocai, WANG Wentong, WANG Lichun. Review of deep learning[J]. Journal of Beijing university of technology, 2015, 41(1): 48-59.
- [7] 张建国, 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 editions, 2015, 36(2): 191-200.
- [8] LECUN Y, JACKEL L D, BOTTOU L, et al. Learning algorithms for classification: a comparison on handwritten digit recognition[M]//OH J H, KWON C, CHO S. Neural Networks: The Statistical Mechanics Perspective. Singapore: World Scientific, 1995: 261-276.
- [9] 陈先昌. 基于卷积神经网络的深度学习算法与应用研究[D]. 杭州: 浙江工商大学, 2014.
CHEN Xianchang. Research on algorithm and application of deep learning based on convolutional neural network[D]. Hangzhou, China: Zhejiang Gongshang University, 2014.
- [10] 李卫. 深度学习在图像识别中的研究及应用[D]. 武汉: 武汉理工大学, 2014.
LI Wei. The research and application of deep learning in image recognition[D]. Wuhan: Wuhan University of Technology, 2014.
- [11] JI Shuiwang, XU Wei, YANG Ming, et al. 3D convolutional neural networks for human action recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(1): 221-231.
- [12] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[EB/OL]. Eprint Arxiv: Arxiv, 2013. [2014-10-10] <http://120.52.73.79/arxiv.org/pdf/1312.6203v3.pdf>.
- [13] FAN Haoqiang, CAO Zhimin, JIANG Yuning, et al. Learning deep face representation[EB/OL]. Eprint Arxiv: Arxiv, 2014. [2014-10-10] <http://120.52.73.80/arxiv.org/pdf/1403.2802v1.pdf>.
- [14] 王冠皓, 徐军. 基于多级金字塔卷积神经网络的快速特征表示方法[J]. 计算机应用研究, 2015, 32(8): 2492-2495.
WANG Guanhao, XU Jun. Fast feature representation method based on multi-level pyramid convolution neural network[J]. Application research of computers, 2015, 32(8): 2492-2495.
- [15] LEE H, GROSSE R, RANGANATH R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]//Proceedings of the 26th Annual International Conference on Machine Learning. New York, NY, USA, 2009: 609-616.
- [16] LEE H, EKANADHAM C, NG A Y. Sparse deep belief net model for visual area V2[C]//Advances in Neural Information Processing Systems 20: 21st Annual Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada, 2007: 873-880.
- [17] 罗恒. 基于协同过滤视角的受限玻尔兹曼机研究[D]. 上海: 上海交通大学, 2011.
LUO Heng. Restricted Boltzmann machines: a collaborative filtering perspective[D]. Shanghai, China: Shanghai Jiao

- Tong University, 2011.
- [18] LAROCHELLE H, BENGIO Y. Classification using discriminative restricted Boltzmann machines [C]//Proceedings of the 25th International Conference on Machine Learning. New York, NY, USA, 2008: 536-543.
- [19] 张春霞, 姬楠楠, 王冠伟. 受限波尔兹曼机[J]. 工程数学学报, 2015, 32(2): 159-173.
- ZHANG Chunxia, JI Nannan, WANG Guanwei. Restricted Boltzmann machines [J]. Chinese journal of engineering mathematics, 2015, 32(2): 159-173.
- [20] 刘银华. LBP 和深度信念网络在非限制条件下人脸识别研究[D]. 江门: 五邑大学, 2014.
- LIU Yinhua. The research of face recognition under unconstrained condition via LBP and deep belief network [D]. Jiangmen: Wuyi University, 2014.
- [21] LEE H, GROSSE R, RANGANATH R, et al. Unsupervised learning of hierarchical representations with convolutional deep belief networks [J]. Communications of the ACM, 2011, 54(10): 95-103.
- [22] HALKIAS X C, PARIS S, GLOTIN H. Sparse penalty in deep belief networks: using the mixed norm constraint [EB/OL]. [2014-05-08]. <http://arxiv.org/pdf/1301.3533.pdf>.
- [23] LIU Yan, ZHOU Shusen, CHEN Qingcai. Discriminative deep belief networks for visual data classification [J]. Pattern recognition, 2011, 44(10/11): 2287-2296.
- [24] 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.
- [25] 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. New York, NY, USA, 2008: 1096-1103.
- [26] BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy layer-wise training of deep networks [C]//Advances in Neural Information Processing Systems 19: 20th Annual Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada, 2006: 153-160.
- [27] RIFAI S, VINCENT P, MULLER X, et al. Contractive auto-encoders: explicit invariance during feature extraction [C]//Proceedings of the 28th International Conference on Machine Learning. Bellevue, WA, USA, 2011.
- [28] MASCI J, MEIER U, CIREŞAN D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction [C]//Proceedings of the 21st International Conference on Artificial Neural Networks, Part I. Berlin Heidelberg, Germany, 2011: 52-59.
- [29] 王雅思. 深度学习中的自编码器的表达能力研究[D]. 哈尔滨: 哈尔滨工业大学, 2014.
- WANG Yasi. Representation ability research of auto-encoders in deep learning [D]. Harbin: Harbin Institute of Technology, 2014.
- [30] 李远豪. 基于深度自编码器的人脸美丽吸引力预测研究[D]. 江门: 五邑大学, 2014.
- LI Yuanhao. A study for facial beauty attractiveness prediction based on deep autoencoder [D]. Jiangmen: Wuyi University, 2014.
- [31] 林洲汉. 基于自动编码器的高光谱图像特征提取及分类方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2014.
- LIN Zhouhan. Hyperspectral image feature extraction and classification based on autoencoders [D]. Harbin: Harbin Institute of Technology, 2014.
- [32] 曲建岭, 杜辰飞, 邸亚洲, 等. 深度自动编码器的研究与展望[J]. 计算机与现代化, 2014(8): 128-134.
- QU Jianling, DU Chenfei, DI Yazhou, et al. Research and prospect of deep auto-encoders [J]. Jisuanji yu xiandaihua, 2014(8): 128-134.
- [33] 林少飞, 盛惠兴, 李庆武. 基于堆叠稀疏自动编码器的手写数字分类[J]. 微处理机, 2015(1): 47-51.
- LIN Shaofei, SHENG Huixing, LI Qingwu. Handwritten digital classification based on the stacked sparse autoencoders [J]. Microprocessors, 2015(1): 47-51.
- [34] 陈硕. 深度学习神经网络在语音识别中的应用研究[D]. 广州: 华南理工大学, 2013.
- CHEN Shuo. Research of deep learning neural networks applications in speech recognition [D]. Guangzhou, China: South China University of Technology, 2013.
- [35] 郭丽丽, 丁世飞. 深度学习研究进展[J]. 计算机科学, 2015, 42(5): 28-33.
- GOU Lili, DING Shifei. Research progress on deep learning [J]. Computer science, 2015, 42(5): 28-33.
- [36] VAN DEN OORD A, DIELEMAN S, SCHRAUWEN B. Deep content-based music recommendation [M]//Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems. Lake Tahoe, 2013: 2643-2651.
- [37] HANNUN A, CASE C, CASPER J, et al. Deep speech: scaling up end-to-end speech recognition [EB/OL]. Eprint Arxiv: Arxiv, 2014. [2014-12-19] <https://arxiv.org/pdf/1412.5567v2.pdf>.
- [38] 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.
- YU Kai, JIA Lei, CHEN Yuqiang. Deep learning: yesterday, today, and tomorrow [J]. Journal of computer research and development, 2013, 50(9): 1799-1804.

- [39] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, 2014: 580-587.
- [40] TAIGMAN Y, YANG Ming, RANZATO M A, et al. DeepFace: closing the gap to human-level performance in face verification[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, 2014: 1701-1708.
- [41] TOSHEV A, SZEGEDY C. DeepPose: human pose estimation via deep neural networks [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, 2014: 1653-1660.
- [42] DIELEMAN S. Classifying plankton with deep neural networks[EB/OL]. (2015-03-17) [2015-05-30]. <http://benanne.github.io/2015/03/17/plankton.html>.
- [43] DENTON E, WESTON J, PALURI M, et al. User conditional hashtag prediction for images [C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA, 2015: 1731-1740.
- [44] LONG J, SHELHAMER E, DARRELL T Fully convolutional networks for semantic segmentation[J]. IEEE Conference on Computer Vision & Pattern Recognition. 2015, 79(10):1337-1342.
- [45] SCHWARA M, SCHULZ H, BEHNKE S. RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features[C]//Proceedings of the 2015 IEEE International Conference on Robotics and Automation. Seattle, WA, 2015: 1329-1335.
- [46] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [C]// Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015: 815-823.
- [47] SARFRAZ M S, STIEFELHAGEN R. Deep perceptual mapping for thermal to visible face recognition.[EB/OL] Eprint Arxiv: Arxiv, 2015. [2015-12-23]. <http://120.52.73.80/arxiv.org/pdf/1507.02879v1.pdf>.
- [48] LIU Mengyi, LI Shaoxin, SHAN Shiguang, et al. Au-aware deep networks for facial expression recognition[C]// Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Shanghai, China, 2013: 1-6.
- [49] OUELLET S. Real-time emotion recognition for gaming using deep convolutional network features[EB/OL]. Eprint Arxiv: Arxiv, 2014. [2014-7-16]. <https://arxiv.org/pdf/1408.3750v1.pdf>.
- [50] SONG I, KIM H J, JEON P B. Deep learning for real-time robust facial expression recognition on a smartphone[C]// Proceedings of the 2014 IEEE International Conference on Consumer Electronics. Las Vegas, NV, 2014: 564-567.
- [51] IJJINA E P, MOHAN C K. Facial expression recognition using kinect depth sensor and convolutional neural networks [C]//Proceedings of the 2014 13th International Conference on Machine Learning and Applications. Detroit, MI, 2014: 392-396.
- [52] BYEON Y H, KWAK K C. Facial expression recognition using 3D convolutional neural network [J]. International journal of advanced computer science and applications, 2014, 5(12): 107-112.
- [53] JUNG H, LEE S, PARK S, et al. Development of deep learning-based facial expression recognition system [C]// Frontiers of Computer Vision (FCV), 2015 21st Korea-Japan Joint Workshop on 2015: 1-4.
- [54] MCLAUGHLIN T, MAI L, BAYANBAT N. Emotion recognition with deep-belief networks [EB/OL]. 2008 <http://cs229.stanford.edu/proj2010/McLaughlinLeBayanbat-RecognizingEmotionsWithDeepBeliefNets.pdf>.
- [55] HE Shan, WANG Shanfei, LAN Wuwei, et al. Facial expression recognition using deep Boltzmann machine from thermal infrared images[C]//Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. Geneva, 2013: 239-244.
- [56] LV Yadan, FENG Zhiyong, XU Chao. Facial expression recognition via deep learning [C]//Proceedings of the 2014 International Conference on Smart Computing. Hong Kong, China, 2014: 303-308.
- [57] JUNG H, LEE S, PARK S, et al. Deep temporal appearance-geometry network for facial expression recognition [EB/OL]. Eprint Arxiv: Arxiv, 2015. [2015-6-5]. <http://120.52.73.75/arxiv.org/pdf/1503.01532v1.pdf>.
- [58] KAHOU S E, BOUTHILLIER X, LAMBLIN P et al. EmoNets: Multimodal deep learning approaches for emotion recognition in video[J]. Journal on Multimodal User Interfaces, 2015, 10(2): 1-13.
- [59] SASHIHITHLU S, SOMAN S S. Complex sentiment analysis using recursive autoencoders [EB/OL]. Core. ac. Uk: CiteSeerX, 2013. [2015-9-30]. <https://core.ac.uk/display/23426251>.
- [60] JOHNSON R, ZHANG Tong. Effective use of word order for text categorization with convolutional neural networks [EB/OL]. Eprint Arxiv: Arxiv, 2014. [2014-10-10]. <http://120.52.73.79/arxiv.org/pdf/1412.1058.pdf>.
- [61] 李婷. 基于深度学习的盲文识别方法[J]. 计算机与现代化, 2015(6): 37-40.
- LI Ting. A deep learning method for braille recognition[J]. Jisuanji yu xiandaihua, 2015(6): 37-40.
- [62] KOZIOL J A, TAN E M, DAI Liping, et al. Restricted Boltzmann machines for classification of hepatocellular car-

- cinoma[J]. Computational biology journal, 2014, 2014: 418069.
- [63] FAUW J D. Detecting diabetic retinopathy in eye images [EB/OL]. 2015 [2015-07-28]. <http://jeffreydf.github.io/diabetic-retinopathy-detection>.
- [64] CUN Y L, BOSER B, DENKER J S, et al. Handwritten digit recognition with a back-propagation network [C]//Advances in Neural Information Processing Systems 2. San Francisco, CA, USA, 1990: 396-404.
- [65] JI Shuiwang, XU Wei, YANG Ming, et al. 3D convolutional neural networks for human action recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(1): 221-231.
- [66] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural computation, 2006, 18(7): 1527-1554.
- [67] SALAKHUTDINOV R, HINTON G E. Deep Boltzmann machines [C]//Proceedings of the 12th International Conference on Artificial Intelligence and Statistics. Clearwater, Florida, USA, 2009: 448-455.
- [68] 杜骞. 深度学习在图像语义分类中的应用[D]. 武汉: 华中师范大学, 2014.
- DU Qian. Application of deep learning in image semantic classification [D]. Wuhan: Central China Normal University, 2014.
- [69] BENGIO Y. Learning deep architectures for AI [J]. Foundations and trends in machine learning, 2009, 2(1): 1-127.
- [70] HINTON G E, SEJNOWSKI T J. Learning and relearning in Boltzmann machines [M]//Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MA, USA: MIT Press, 1986.
- [71] SMOLENSKY P. Information processing in dynamical systems: foundations of harmony theory [M]//Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MA, USA: MIT Press, 1986.
- [72] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [J]. Nature, 1986, 323(6088): 533-536.
- [73] MIKOLOV T, KARAFIÁT M, BURGET L, et al. Recurrent neural network based language model [C]//Proceedings of the Interspeech 2010 11th Annual Conference of the International Speech Communication Association. Makuhari, Chiba, Japan, 2010: 1045-1048.
- [74] DENG Li, YU Dong. Deep convex net: a scalable architecture for speech pattern classification [C]//Proceedings of the 12th Annual Conference of the International Speech Communication Association. Florence, Italy, 2011: 2296-2299.
- [75] MIKOLOV T, KOMBRINK S, BURGET L. Extensions of recurrent neural network language model [C]//Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing. Prague, 2011: 5528-5531.
- [76] LIU Shujie, YANG Nan, LI Mu, et al. A recursive recurrent neural network for statistical machine translation [C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, USA, 2014: 1491-1500.
- [77] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks [EB/OL]. 2014. [2014-10-12]. <http://120.52.73.79/arxiv.org/pdf/1409.3215v3.pdf>.
- [78] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks [C]//Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, 2013: 6645-6649.
- [79] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks [EB/OL]. Eprint Arxiv: Arxiv, 2014. [2014-10-12]. <http://120.52.73.78/arxiv.org/pdf/1312.6199v4.pdf>.
- [80] NGUYEN A, YOSINSKI J, CLUNE J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images [C]//Computer Vision and Pattern Recognition. IEEE, Boston, USA, 2015: 427-436.
- [81] LIPTON Z C. (Deep learning's deep flaws)'s deep flaws [EB/OL]. [2015-07-28]. <http://www.kdnuggets.com/2015/01/deep-learning-flaws-universal-machine-learning.html>.

作者简介:



刘帅师,女,1981年生,讲师,博士,主要研究方向为模式识别、计算机视觉。



程曦,男,1989年生,硕士研究生,主要研究方向为模式识别、机器学习。



郭文燕,女,1991年生,硕士研究生,主要研究方向为模式识别、机器学习。