

自然语言处理中的深度学习：方法及应用

林奕欧, 雷航, 李晓瑜, 吴佳

(电子科技大学信息与软件工程学院 成都 610054)

【摘要】该文围绕特征表示和模型原理,以神经网络语言模型与词向量作为深度学习与自然语言处理结合的切入点,概述了当前主要深度神经网络的模型原理和相关应用。之后综述了当前研究人员在自然语言处理热点领域上所使用的最新深度学习方法及所取得的成果。最后总结了深度学习方法在当前自然语言处理研究应用中所遇到的瓶颈,并对未来可能的研究重点做出展望。

关键词 深度学习; 深度神经网络; 语言模型; 自然语言处理; 词向量

中图分类号 TP391 **文献标志码** A **doi**:10.3969/j.issn.1001-0548.2017.06.021

Deep Learning in NLP: Methods and Applications

LIN Yi-ou, LEI Hang, LI Xiao-yu, and WU Jia

(School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract With the rise of deep learning waves, the full force of deep learning methods has hit the Natural Language Process (NLP) and ushered in amazing technological advances in many different application areas of NLP. In this article, we firstly present the development history, main advantages and research situation of deep learning. Secondly, in terms of both feature representation and model theory, we introduces the neural language model and word embedding as the entry point, and present an overview of modeling and implementations of Deep Neural Network (DNN). Then we focus on the newest deep learning models with their wonderful and competitive performances related to different NLP tasks. At last, we discuss and summarize the existing problems of deep learning in NLP with the possible future directions.

Key words deep learning; deep neural networks; language models; nature language process; word embedding

作为机器学习和人工智能领域发展最为迅速的研究方向,深度学习受到学术界和工业界的高度关注。深度学习是基于特征自学习和深度神经网络(DNN)的一系列机器学习算法的总称。目前深度学习的研究有了长足的发展,在传统特征选择与提取框架上取得了巨大突破,对包括自然语言处理、生物医学分析、遥感影像解译在内的诸多领域产生越来越重要的影响,并在计算机视觉和语音识别领域取得了革命性的成功。

当前,如何应用深度学习技术解决自然语言处理(NLP)相关任务是深度学习的研究热点。NLP作为计算机科学与人工智能交叉领域中的重要研究方向,综合了语言学、计算机科学、逻辑学、心理学、人工智能等学科的知识与成果。其主要研究任务包括词性标注、机器翻译、命名实体识别、机器问答、

情感分析、自动文摘、句法分析和共指消解等。自然语言作为高度抽象的符号化系统,文本间的关系难以度量,相关研究高度依赖人工构建特征。而深度学习方法的优势恰恰在于其强大的判别能力和特征自学习能力,非常适合自然语言高维数、无标签和大数据的特点。为此,本文将对当前深度学习如何应用在NLP领域展开综述性讨论,并进一步分析其中的应用难点和未来可能的突破方向。

1 深度学习研究现状

深度学习源于人工神经网络的研究。人工神经网络(artificial neural network, ANN)作为计算工具是由文献[1]引入。之后,Hebb自组织学习规则、感知机模型、Hopfield神经网络、玻尔兹曼机、误差反向传播算法和径向基神经网络等也相继被提出。文献

收稿日期:2016-07-04;修回日期:2017-03-27

基金项目:国家自然科学基金(61502082);中央高校基本科研业务费(ZYGX2014J065)

作者简介:林奕欧(1991-),男,博士生,主要从事自然语言处理和深度学习方面的研究。

[2]利用逐层贪心算法初始化深度信念网络,开启了深度学习的浪潮,指出深度学习的本质是一种通用的特征学习方法,其核心思想在于提取低层特征,组合形成更高层的抽象表示,以发现数据的分布规律。文献[2]的方法有效地缓解了DNN层数增加所带来的梯度消失或者梯度爆炸问题。随后文献[3]使用自动编码器取代深度信念网络的隐藏层,并通过实验证明了DNN的有效性。同时,研究发现人类信息处理机制需要从丰富的感官输入中提取复杂结构并重新构建内部表示,使得人类语言系统和感知系统都具有明显的层结构^[4],这从仿生学的角度,为DNN多层网络结构的有效性提供了理论依据。

此外,深度学习的兴起还有赖于大数据和机器学习性能的提升。大数据是具有大量性、多样性、低价值密度性的数据的统称,深度学习是处理大数据常用的方法论,两者有紧密的联系。以声学建模为例,其通常面临的是十亿到千亿级别的训练样本,实验发现训练后模型处于欠拟合状态,因此大数据需要深度学习^[5]。另外,随着图形处理器(graphics processing unit, GPU)的发展,有效且可扩展的分布式GPU集群的使用大大加速了深度模型的训练过程,极大地促进了深度学习在业界的使用。

目前,NLP应用逐渐成为深度学习研究中又一活跃热点。2013年,随着词向量word2vec^[6]的兴起,各种词的分布式特征相关研究层出不穷。2014年开始,研究者使用不同的DNN模型,例如卷积网络,循环网络和递归网络,在包括词性标注、情感分析、句法分析等传统NLP应用上取得重大进展。2015年后,深度学习方法开始在机器翻译、机器问答、自动文摘、阅读理解等自然语言理解领域攻城略地,逐渐成为NLP的主流工具。在未来几年,深度学习将持续在自然语言理解领域做出巨大影响^[7]。

2 分布式特征表示

分布式特征表示(distributional representation)是深度学习与NLP相结合的切入点,这些分布式特征是通过神经网络语言模型学习得到的。

2.1 神经网络语言模型

语言模型是计算任意词序在文本中出现的概率的统计模型,是NLP的基础性课题,对语音识别、词性标注、机器翻译、句法分析等研究都有至关重要的作用。神经网络生成的语言模型称为神经网络语言模型,由文献[8]提出。之后文献[9]等对神经网络语言模型进行了深入研究,其成果引起广泛关注。

文献[9]构建的神经网络语言模型如图1所示,仅有一层隐藏层,使用softmax计算词序的概率,并通过随机梯度上升法最大惩罚似然估计以训练网络参数。

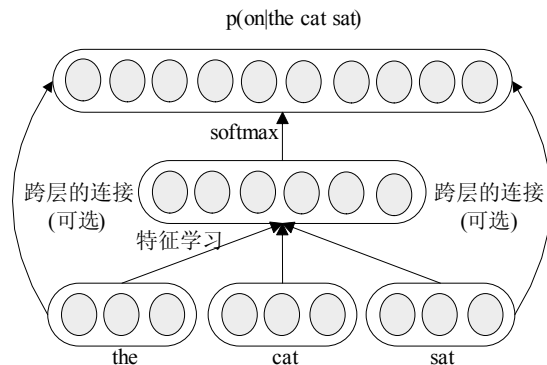


图1 基于神经网络的语言模型

在文献[9]的基础上,文献[10]从语料中自动学习出词的层次结构,并结合受限玻尔兹曼机提出了HLBL模型。文献[11]又在文献[10]的基础上提出SENNA模型,通过对一个句子的合理性进行打分以训练词向量,在多项NLP任务上成功应用。之后文献[12]复现了HLBL模型和SENNA模型,并比较了二者的优劣。文献[13]在SENNA的基础上,对每个词向量的训练加入了相应全局信息,并用多个词向量对应一个词以解决词的多义性问题。不同于其他人的工作,文献[6]使用了循环神经网络来训练语言模型,并在2013年开源word2vec。从HLBL模型开始,研究语言模型的目的不再是获得真正的语言模型而在于获得可用的词向量。

2.2 词向量

词向量通常指通过语言模型学习得到的词的分布式特征表示,也被称为词编码,可以非稀疏的表示大规模语料中复杂的上下文信息。目前最为人所熟知的有以下6种公开发表的词向量:HLBL、SENNA、Turian's、Huang's、word2vec、Glove^[14]。文献[15]发现word2vec词向量间具有语义上的联系,即词向量的加减存在明显的语义关系,并在SemEval 2012 task上取得超过Turian词向量的结果,证明了word2vec的高可用性。文献[16]证明了当word2vec词向量使用skip-gram模型配合负采样技术训练时,与基于SVD的共现矩阵分解的词向量具有相同的最优解。同年文献[14]提出了Glove词向量,并证明基于矩阵的词向量可以取得远比word2vec优异的性能,但根据文献[17]提出的测评指标显示word2vec在大部分测评指标优于Glove和SENNA。除了使用词向量解决当前自然语言测评任务外,也有许多学

者对词向量进行了其他广泛而深入地研究, 如文献[18]研究了语种的差异对词向量的影响, 文献[19]就如何生成更好的词向量进行深入讨论, 文献[20]利用词向量计算了文档的相似度。

3 深度神经网络模型

3.1 深度信念网络和堆叠自编码

深度信念网络(deep belief nets, DBN)是由受限玻尔兹曼机(restricted Boltzmann machine, RBM)堆叠而生成的一种模型。DBN通过训练网络的权重, 使网络具有还原输入层训练数据的能力。DBN采用的训练步骤如下:

- 1) 当前层RBM为可见层则接收原始数据输入, 否则接收上一层RBM的输出, 并训练当前层RBM;
- 2) 网络总层数满足要求则执行步骤4), 否则置下一层RBM为当前层;
- 3) 重复步骤1)和步骤2);
- 4) 微调网络, 使用有监督学习算法将模型收敛到局部最优解。

文献[21]讨论了RBM和DBN网络的层数设置、网络泛化能力以及可能的扩展, 并使用自编码器(auto-encoder, AE)取代DBN网络中每一层的RBM, 由此简单堆叠数个AE得到的神经网络在文献[3]中被称为堆叠自编码网络(stacked auto-encoders, SAE)。目前SAE有两种典型的改进: 1) 在隐藏神经元加入稀疏性限制, 使网络中大部分神经元处于抑制状态, 形成稀疏自编码网络^[22]; 2) 在SAE网络的编码过程加入噪音, 增加SAE网络的抗噪性, 形成堆叠降噪自编码网络^[23]。SAE网络由于强大的特征学习能力^[24], 被广泛使用在多模态检索^[25]、图像分类^[26]、情感分析^[27]等诸多领域中。

3.2 循环神经网络与长短期记忆模型

循环神经网络(recurrent neural networks, RNN)是隐藏层和自身存在连接的一类神经网络。相较于前馈神经网络, 循环神经网络可将本次隐藏层的计算结果用于下次隐藏层的计算, 因此可以用来处理时间序列问题, 比如文本生成^[28]、机器翻译^[29]和语音识别^[30]。循环神经网络的优化算法为BPTT算法(backpropagation through time)^[31]。由于梯度消失的原因, 循环神经网络的反馈误差往往只能向后传递5~10层, 因此文献[32]在循环神经网络的基础上提出长短期记忆模型(long-short term memory, LSTM)。LSTM使用Cell结构记忆之前的输入, 使得网络可以学习到合适的时机重置Cell结构。LSTM有诸多结构

变体, 文献[33]给出了其中8种流行变体的比较。文献[34]则在超过1万种循环网络架构上进行了测试, 发现并列举在某些任务上可能比LSTM更好的架构。

循环神经网络和LSTM具有许多NLP应用。文献[35]将门控循环网络用于情感分析, 在IMDB等影评数据集上较SVM和CNN方法在准确率上有5%左右的提升。文献[36]使用双向LSTM网络结合卷积神经网络和条件随机场解决词性标注和命名实体识别问题, 分别取得97.55%和91.21%的最好结果。

3.3 递归神经网络和卷积神经网络

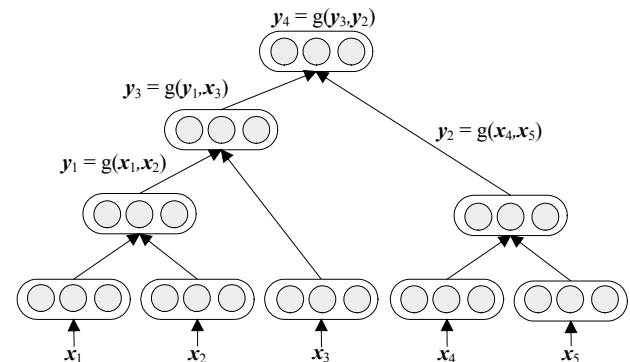


图2 递归神经网络模型结构图

递归神经网络(recursive neural networks)是利用树形神经网络结构递归构造而成, 用于构建句子语义信息的深度神经网络^[37]。递归神经网络所用的树形结构一般是二叉树, 典型的递归神经网络如图2所示。对于句法分析任务, 图2中 x 表示词向量, y 表示合并而成的子树向量。定义全局参数 W , b 和 U , 以及子树合理性评分变量 s , 对于合并节点($x_1, x_2 \rightarrow y_1$), 有:

$$y_1 = \tanh \left(W \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b \right) \quad (4)$$

$$s_1 = U^T y_1 \quad (5)$$

随机初始全局参数, 使用贪心算法, 相邻的叶子节点(或子树)两两合并成子树并计算评分, 取分最高者合并, 直到最终形成句法树。句法分析任务是有标定的, 即存在一个正确的句法树构造, 因此训练目标是优化网络参数使得整个网络的评分损失最小。除了句法分析, 递归神经网络还可用于关系分类^[38]和情感分析^[39]中。

卷积神经网络(convolutional neural networks, CNN)是由文献[40]提出并由文献[41]改进的深度神经网络。在一般前馈神经网络中, 输入层和隐藏层之间采用全连接结构, 而在CNN中每一个卷积层节点只与一个固定大小的区域有连接, 连接的权重矩

阵称为卷积核。池化(pooling)是CNN所采用的另一项关键技术,在固定大小的区域使用平均值或最大值代替原有的矩阵区域,既减少了特征数目又增加了网络的鲁棒性。

目前CNN在NLP领域的应用有许多新的尝试。文献[6]将CNN用于语义角色标注,文献[42]使用字符作为语义特征,采用大规模文本语料训练CNN模型用于本体分类、情感分析和文本分类,所用的CNN模型如图3所示。

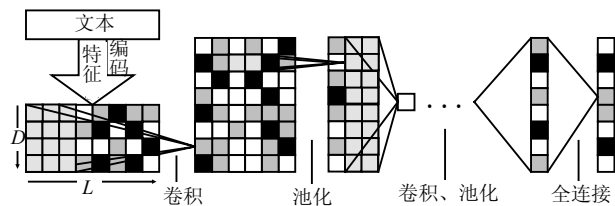


图3 用于文本分类的卷积网络结构图

4 NLP应用研究进展

深度学习方法在诸多NLP领域中得到广泛应用。在机器翻译领域,文献[43]将DNN和词编码用于机器翻译,困惑度下降15%。文献[44]利用双语词向量作为特征,对应的BLEU值提升了0.48。同时用于解决机器翻译问题(如词语对齐^[45]、语序问题^[46])的循环神经网络也被广泛用于文本生成。文献[28]只使用字符序列训练循环神经网络文本生成器,效果接近加入了大量人工规则的文本生成系统。关于如何使用循环网络生成文本,文献[47-48]提供了丰富有趣的案例。近来,关于图片的解释性文本生成也受到广泛关注^[49]。

词性标注、组块分析、语义角色标注和命名实体识别在SENN系统^[11]中给出了统一的解决框架,即基于词向量特征的深度网络判别模型。在此基础上,文献[50]将字符特征加入词性标注任务中,准确率略有提升(0.3%)。组块分析与词性标注任务类似,算法准确率的提升也已陷入瓶颈^[51]。对于语义角色标注任务,文献[52]提出使用半监督学习训练词编码,有效地提升了准确率。命名实体(人名、地名、时间和数字等的识别)是许多NLP应用的基础,受到众多学者关注。相较于传统的特征工程方法^[53],文献[54]使用了词和字符作为特征,并将LSTM与条件随机场相结合,采用dropout策略,在实体识别上取得了更好的识别率。

在结构句法分析上,文献[55]对比了不同递归神经网络对结构句法分析的影响,发现递归神经网络句法分析的最好效果也略逊于斯坦福大学的句法分

析工具(Stanford Parser)。因此文献[38]将Stanford Parser原理与递归神经网络相结合,使得算法准确率进一步提升(0.855~0.904)。至于文本分类任务,使用字符特征和词编码的卷积网络^[11]、基于张量组合的递归网络模型^[39]和基于树形结构的循环网络模型^[56]是当前卓有成效的混合深度网络。

机器问答是一项极其困难的NLP任务。文献[57]给出了使用神经网络求解机器问答的一般流程。文献[58]提出了记忆神经网络,以经过语义分析和人为筛选的先验事实文本为输入,有监督学习循环神经网络权重。近年来针对图像内容的多模态问答任务也受到了广泛的关注^[59]。

在国内,将深度学习方法应用于NLP的研究也越来越多的受到学者的关注。文献[60]将自适应递归神经网络用于情感分析,文献[61]将情感信息直接嵌入词向量并用于情感分类,文献[62]将DBN用于代词指代消歧。国内研究团队也不约而同地使用深度学习解决NLP热点问题。华为诺亚方舟实验室将CNN用于机器翻译^[63]和多模态问答^[64]。微软亚洲研究院致力于利用不同的深度网络实现机器问答、机器翻译和聊天机器人^[65-67]。清华大学自然语言处理与社会人文计算实验室将深度学习方法用于机器翻译,关系抽取以及知识的分布式表示中^[68-70],苏州大学自然语言组则侧重于中文信息抽取和多语情感分析^[71-72]。哈工大、中科院、北京大学等高校的自然语言组也屡次在国际会议上发表高水平学术论文^[73-75],越来越多的中国学者对深度学习结合NLP领域的研究做出了卓越的贡献。

5 结束语

尽管深度学习已经在诸多应用领域取得巨大成功,但深度学习作为一项正在蓬勃发展的新兴技术,仍然有许多研究难点需要攻克。其中最大的瓶颈在于,除了仿生学的角度,目前深度学习的理论依据还处于起步阶段,大部分的研究成果都是经验性的,没有足够的理论来指导实验,研究者无法确定网络架构、超参数设置是否已是最优的组合。除此之外,目前仍没一种通用的深度网络或学习策略可以适用于大多数的应用任务,因此深度学习领域的研究者正在不断尝试新的网络架构和学习策略,以提升网络的泛化性能。

目前,深度学习用于NLP领域的主要步骤可以归结为如下3步:1) 将原始文本作为输入,自学习得到文本特征的分布表示。2) 将分布式向量特征作

为深度神经网络的输入。3) 针对不同的应用需求, 使用不同的深度学习模型, 有监督的训练网络权重。目前深度学习结合NLP的应用前景及其广泛。深度学习模型在语法分析和信息抽取等研究的基础上, 被灵活地运用在多语言机器翻译、机器问答、多模态应用、聊天机器人等一系列自然语言任务上。

然而深度学习在NLP研究上尚未取得像语音识别和计算机视觉那样巨大的成功。本文认为深度学习方法在NLP应用上的难点和可能的突破统一于以下4个方面: 1) 可广泛适用于不同NLP任务的通用语义特征。2) 超参数设置相关研究。3) 新型网络架构和学习策略的提出和研究(如注意力模型^[76])。4) 基于自然语言的逻辑推理和多模态应用。前者将提升机器的“智能”, 后者扩展“智能”的应用领域。

综上, 本文认为深度学习方法在NLP领域已经有许多很有价值的尝试, 在不久的将来, 将取得更大的成功。但未来依旧充满了挑战, 值得更多的研究者进行广泛而深入地研究。

参 考 文 献

- [1] LANDAHL H D, MCCULLOCH W S, PITTS W. A statistical consequence of the logical calculus of nervous nets[J]. *Bulletin of Mathematical Biology*, 1943, 5(4): 135-137.
- [2] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.
- [3] BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy layer-wise training of deep networks[C]//*Proceedings of NIPS*. Vancouver, Canada: MIT Press, 2007:153-160.
- [4] MATSUGU M, MORI K, MITARI Y, et al. Subject independent facial expression recognition with robust face detection using a convolutional neural network[J]. *Neural Networks*, 2003, 16(5): 555-559.
- [5] 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, 9: 1799-1804.
YU Kai, JIA Lei, CHEN Yu-qiang, et al. Deep learning: Yesterday, today, and tomorrow[J]. *Journal of Computer Research and Development*, 2013, 9: 1799-1804.
- [6] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[C]//*Proceedings of ICLR*. Scottsdale, Arizona, USA: arXiv Press, 2013: 1301.3781.
- [7] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [8] XU W, RUDNICKY A I. Can artificial neural networks learn language models?[C]//*Proceedings of International Conference on Speech and Language Processing*. Beijing, China: Speech Communication Press. 2000.
- [9] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model[J]. *Journal of Machine Learning Research*, 2003, 3: 1137-1155.
- [10] MNH A, HINTON G E. A scalable hierarchical distributed language model[C]//*Proceedings of NIPS*. New York: Curran Associates Inc, 2008.
- [11] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch[J]. *The Journal of Machine Learning Research*, 2011, 12: 2493-2537.
- [12] TURIAN J, RATINOV L, BENGIO Y. Word representations: a simple and general method for semi-supervised learning[C]//*Proceedings of ACL*. Uppsala, Sweden: ACL Press, 2010: 384-394.
- [13] HUANG E H, SOCHER R, MANNING C D, et al. Improving word representations via global context and multiple word prototypes[C]//*Proceedings of ACL*. Jeju, Korea: ACL Press, 2012: 873-882.
- [14] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]//*Proceedings of EMNLP*. Doha, Qatar: ACL Press, 2014, 14: 1532-1543.
- [15] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//*Proceedings of NIPS*. Nevada, USA: MIT Press, 2013: 3111-3119.
- [16] LEVY O, GOLDBERG Y. Neural word embedding as implicit matrix factorization[C]//*Proceedings of NIPS*. Montreal, Quebec: MIT Press, 2014: 2177-2185.
- [17] SCHNABEL T, LABUTOV I, MIMNO D, et al. Evaluation methods for unsupervised word embeddings[C]//*Proceedings of EMNLP*. Lisbon, Portugal: ACL Press, 2015.
- [18] AL-RFOU R, PEROZZI B, SKIENA S. Polyglot: Distributed word representations for multilingual nlp[C]//*Proceedings of CoNLL*. Sofia, Bulgaria: ACL Press, 2013: 183.
- [19] LAI S, LIU K, XU L, et al. How to generate a good word embedding?[J]. *IEEE Intelligent Systems*, 2016, 31(6): 5-14.
- [20] KUSNER M, SUN Y, KOLKIN N, et al. From word embeddings to document distances[C]//*Proceedings of ICML*. Lille, France: Omni Press, 2015: 957-966.
- [21] Le ROUX N, BENGIO Y. Representational power of restricted Boltzmann machines and deep belief networks[J]. *Neural Computation*, 2008, 20(6): 1631-1649.
- [22] LEE H, EKANADHAM C, NG A Y. Sparse deep belief net model for visual area V2[C]//*Proceedings of NIPS*. New York, USA: ACM Press, 2008: 873-880.
- [23] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders[C]//*Proceedings of ICML*. New York, USA: ACM Press, 2008: 1096-1103.
- [24] GEHRING J, MIAO Y, METZE F, et al. Extracting deep bottle-neck features using stacked auto-encoders[C]//*Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, BC, Canada: IEEE, 2013: 3377-3381.
- [25] WANG W, OOI B C, YANG X, et al. Effective multi-modal retrieval based on stacked auto-encoders[J]. *Proceedings of the Vldb Endowment*, 2014, 7(8): 649-660.

- [26] XIE J, XU L, CHEN E. Image denoising and inpainting with deep neural networks[C]//Proceedings of NIPS. Nevada, USA: MIT Press, 2012: 341-349.
- [27] GLOROT X, BORDES A, BENGIO Y. Domain adaptation for large-scale sentiment classification: a deep learning approach[C]//Proceedings of ICML. Bellevue, Washington, USA: ACM Press, 2011: 513-520.
- [28] SUTSKEVER I, MARTENS J, HINTON G E. Generating text with recurrent neural networks[C]//Proceedings of ICML. Bellevue, Washington, USA: ACM Press, 2011, 1017-1024.
- [29] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation[C]//Proceedings of EMNLP Processing. Doha, Qatar: ACL Press, 2014: 1724-1734.
- [30] GRAVES A, JAITLEY N. Towards end-to-end speech recognition with recurrent neural networks[C]//Proceedings of ICML. Beijing, China: [s.n.], 2014: 1764-1772.
- [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] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [33] GREFF K, SRIVASTAVA R K, KOUTNÍK J, et al. LSTM: a search space odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 99: 1-11.
- [34] JOZEFOWICZ R, ZAREMBA W, SUTSKEVER I. An empirical exploration of recurrent network architectures [C]//Proceedings of ICML. Lille, France: Omni Press, 2015: 2342-2350.
- [35] TANG D, QIN B, LIU T. Document modeling with gated recurrent neural network for sentiment classification[C]//Proceedings of EMNLP. Lisbon, Portugal: ACL Press, 2015: 1422-1432.
- [36] MA X, HOVY E. End-to-end sequence labeling via bi-directional lstm-cnns-crf[C]//Proceedings of ACL. Berlin, Germany: ACL Press, 2016:1064-1074.
- [37] LIU S, YANG N, LI M, et al. A recursive recurrent neural network for statistical machine translation[C]//Proceedings of EMNLP. Doha, Qatar: ACL Press, 2014: 1491-1500.
- [38] SOCHER R, HUVAL B, MANNING C D, et al. Semantic compositionality through recursive matrix-vector spaces [C]//Proceedings of the EMNLP-CoNLL. Jeju Island, Korea: ACL Press, 2012: 1201-1211.
- [39] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//Proceedings of Advances in Neural Information Processing Systems. Nevada, USA: MIT Press, 2013: 926-934.
- [40] FUKUSHIMA K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological Cybernetics, 1980, 36(4): 193-202.
- [41] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [42] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification[C]//Proceedings of NIPS. Montreal, Quebec, Canada: MIT Press, 2015: 649-657.
- [43] SCHWENK H. Continuous space translation models for phrase-based statistical machine translation[C]//Proceedings of COLING. Mumbai, India: ACL Press, 2012: 1071-1080.
- [44] ZOU W Y, SOCHER R, CER D, et al. Bilingual word embeddings for phrase-based machine translation[C]//Proceedings of EMNLP. Seattle, USA: ACL Press, 2013: 1393-1398.
- [45] YANG N, LIU S, LI M, et al. Word alignment modeling with context dependent deep neural network[C]//Proceedings of ACL. Sofia, Bulgaria: ACL Press, 2013: 166-175.
- [46] ZHANG J, LIU S, LI M, et al. Mind the gap: Machine translation by minimizing the semantic gap in embedding space[C]//AAAI Conference on Artificial Intelligence. Québec, Canada: AAAI Press, 2014.
- [47] KARPATHY A. The unreasonable effectiveness of recurrent neural networks[EB/OL]. [2015-05-21]. <https://karpathy.github.io/2015/05/21/rnn-effectiveness/>.
- [48] 郑伟秩. 让神经网络会做唐诗[EB/OL]. [2016-02-01]. <http://zhengwy.com/neural-network-for-tangshi/>. ZHENG Wei-Zhi. Let the neural network write poetry of Tang Dynasty[EB/OL]. [2016-02-01]. <http://zhengwy.com/neural-network-for-tangshi/>.
- [49] KARPATHY A, LI Fei-fei. Deep visual-semantic alignments for generating image descriptions[C]//Proceedings of CVPR. Boston, USA: IEEE, 2015.
- [50] SANTOS C D, ZADROZNY B. Learning character-level representations for part-of-speech tagging[C]//Proceedings of ICML. Beijing, China, 2014: 1818-1826.
- [51] DAHL G E, ADAMS R P, LAROCHELLE H. Training restricted boltzmann machines on word observations[C]//Proceedings of ICML. Edinburgh, UK: Omni Press, 2012: 679-686.
- [52] WESTON J, RATLE F, MOBAHI H, et al. Deep learning via semi-supervised embedding[M]//Neural Networks: Tricks of the Trade. [S.l.]: Springer Heidelberg, 2012: 639-655.
- [53] TKACHENKO M, SIMANOVSKY A. Named entity recognition: Exploring features[C]//Proceeding of KONVENS. Vienna, Austria: Wien, 2012: 118-127.
- [54] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[C]//Proceedings of NAACL. San Diego, USA: ACL Press, 2016: 260-270.
- [55] SOCHER R, MANNING C D, NG A Y. Learning continuous phrase representations and syntactic parsing with recursive neural networks[C]//Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop. British Columbia, Canada: MIT Press, 2010: 2550-2558.
- [56] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long

- short-term memory networks[C]//Proceedings of ACL. Beijing, China: ACL Press, 2015: 1556-1566.
- [57] SAGARA T, HAGIWARA M. Natural language neural network and its application to question-answering system[J]. Neurocomputing, 2014, 142: 201-208.
- [58] WESTON J, CHOPRA S, BORDES A. Memory networks [C]//Proceedings of ICLR. San Diego, California, USA: arXiv Press, 2015.
- [59] ANDREAS J, ROHRBACH M, DARRELL T, et al. Learning to compose neural networks for question answering[C]//Proceedings of NAACL-HLT. San Diego California, USA: ACL Press, 2016: 1545-1554.
- [60] DONG L, WEI F, ZHOU M, et al. Adaptive multicompositionality for recursive neural models with applications to sentiment analysis[C]//AAAI Conference on Artificial Intelligence. Québec, Canada: AAAI Press, 2014: 1537-1543.
- [61] TANG D, WEI F, QIN B, et al. Coooolll: a deep learning system for Twitter sentiment classification[C]//Proceedings of the 8th International Workshop on Semantic Evaluation. Dublin, Ireland: ACL Press, 2014: 208-212.
- [62] 奚雪峰, 周国栋. 基于deep learning的代词指代消解[J]. 北京大学学报(自然科学版), 2014, 50(1): 100-110.
XI Xue-feng, ZHOU Guo-dong. Pronoun resolution based on deep learning[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2014, 50(1): 100-110.
- [63] MENG F, LU Z, WANG M, et al. Encoding source language with convolutional neural network for machine translation[C]//Proceedings of ACL. Beijing, China: ACL Press, 2015.
- [64] MA L, LU Z, LI H. Learning to answer questions from image using convolutional neural network[C]//AAAI Conference on Artificial Intelligence. Phoenix, USA: [s.n.], 2016.
- [65] DONG L, WEI F, ZHOU M, et al. Question answering over freebase with multi-column convolutional neural networks[C]//Proceedings of ACL. Beijing, China: ACL Press, 2015: 260-269.
- [66] FENG S, LIU S, YANG N, et al. Improving attention modeling with implicit distortion and fertility for machine translation[C]//Proceedings of COLING. Osaka, Japan: ACL Press, 2016: 3082-3092.
- [67] YAN Z, DUAN N, BAO J, et al. DocChat: an information retrieval approach for chatbot engines using unstructured documents[C]//Proceedings of ACL. Berlin, Germany: ACL Press, 2016: 516-525.
- [68] CHENG Y, XU W, HE Z, et al. Semi-supervised learning for neural machine translation[C]//Proceedings of ACL. Berlin, Germany: ACL Press, 2016: 1965-1974.
- [69] LIN Y, SHEN S, LIU Z, et al. Neural relation extraction with selective attention over instances[C]//Proceedings of ACL. Berlin, Germany: ACL Press, 2016, 1: 2124-2133.
- [70] 刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. 计算机研究与发展, 2016, 2: 247-261.
LIU Zhi-yuan, SUN Mao-song, LIN Yan-kai, et al. Knowledge representation learning: a review[J]. Journal of Computer Research and Development, 2016, 2: 247-261.
- [71] LI P, ZHOU G. Joint argument inference in Chinese event extraction with argument consistency and event relevance[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(4): 612-622.
- [72] WANG Z, ZHANG Y, LEE S Y M, et al. A bilingual attention network for code-switched emotion prediction [C]//Proceedings of COLING. Osaka, Japan: ACL Press, 2016: 1624-1634.
- [73] GUO J, CHE W, WANG H, et al. Revisiting embedding features for simple semi-supervised learning[C]// Proceedings of EMNLP. Doha, Qatar: ACL Press, 2014: 110-120.
- [74] LI X, ZHANG J, ZONG C. Towards zero unknown word in neural machine translation[C]//Proceedings of IJCAI. New York, USA: AAAI Press, 2016: 2852-2858.
- [75] PEI W, GE T, CHANG B. An effective neural network model for graph-based dependency parsing[C]// Proceedings of ACL. Beijing, China: ACL Press, 2015.
- [76] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]// International Conference on Learning Representations. San Diego, California, USA: arXiv Press, 2015: 1409.0473V7.

编辑 叶芳