



## 基于深度学习的文本分类技术研究进展

何力<sup>1</sup>, 郑灶贤<sup>2</sup>, 项凤涛<sup>1</sup>, 吴建宅<sup>1</sup>, 谭林<sup>3</sup>

(1. 国防科技大学 智能科学学院, 长沙 410073; 2. 南方电网数字电网研究院有限公司, 广州 510555;  
3. 湖南天河国云科技有限公司, 长沙 410073)

**摘要:** 随着深度学习技术的快速发展, 许多研究者尝试利用深度学习来解决文本分类问题, 特别是在卷积神经网络和循环神经网络方面, 出现了许多新颖且有效的分类方法。对基于深度神经网络的文本分类问题进行分析, 介绍卷积神经网络、循环神经网络、注意力机制等方法在文本分类中的应用和发展, 分析多种典型分类方法的特点和性能, 从准确率和运行时间方面对基础网络结构进行比较, 表明深度神经网络较传统机器学习方法在用于文本分类时更具优势, 其中卷积神经网络具有优秀的分类性能和泛化能力。在此基础上, 指出当前深度文本分类模型存在的不足, 并对未来的研究方向进行展望。

**关键词:** 深度学习; 文本分类; 卷积神经网络; 循环神经网络; 注意力机制

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 何力, 郑灶贤, 项凤涛, 等. 基于深度学习的文本分类技术研究进展[J]. 计算机工程, 2021, 47(2): 1-11.

**英文引用格式:** HE Li, ZHENG Zaoxian, XIANG Fengtao, et al. Research progress of text classification technology based on deep learning[J]. Computer Engineering, 2021, 47(2): 1-11.

## Research Progress of Text Classification Technology Based on Deep Learning

HE Li<sup>1</sup>, ZHENG Zaoxian<sup>2</sup>, XIANG Fengtao<sup>1</sup>, WU Jianzhai<sup>1</sup>, TAN Lin<sup>3</sup>

(1. College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China;  
2. Digital Grid Research Institute, China Southern Power Grid, Guangzhou 510555, China;  
3. Hunan Tianhe Guoyun Technology Co., Ltd., Changsha 410073, China)

**[Abstract]** With the rapid development of deep learning technology, many researchers have tried to utilize deep learning to solve the text classification problems. Especially in terms of the Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), many novel and effective classification methods have been proposed. This paper analyzes the problem of text classification based on Deep Neural Network (DNN) and introduces the application and development of the CNN, RNN, attention mechanism and other methods in text classification. Then it analyzes the characteristics and performance of multiple text classification methods based on deep learning, and compares the basic network structures in terms of accuracy and elapsed time, and shows that DNN methods outperform the traditional machine learning methods, and among them, CNN provides better classification performance and generalization ability. On this basis, this paper points out the deficiencies of the existing deep models for text classification, and prospects the direction of future research.

**[Key words]** deep learning; text classification; Convolutional Neural Network (CNN); Recurrent Neural Network (RNN); attention mechanism

DOI: 10.19678/j.issn.1000-3428.0059099

### 0 概述

文本分类技术经历了从专家系统到机器学习再到深度学习的发展过程。在20世纪80年代以前, 基于规

则系统的文本分类方法需要领域专家定义一系列分类规则, 通过规则匹配判断文本类别。基于规则的分类方法容易理解, 但该方法依赖专家知识, 系统构建成本高且可移植性差。20世纪90年代, 机器学习技术逐渐

**基金项目:** 国家自然科学基金(U1734208)。

**作者简介:** 何力(1984—), 男, 讲师、博士, 主研方向为自然语言处理、机器学习; 郑灶贤, 高级工程师、硕士; 项凤涛、吴建宅, 讲师、博士; 谭林, 助理研究员、博士。

**收稿日期:** 2020-07-29 **修回日期:** 2020-09-30 **E-mail:** lihe@nudt.edu.cn

走向成熟,出现了许多经典的文本分类算法,如决策树<sup>[1]</sup>、朴素贝叶斯<sup>[2]</sup>、支持向量机<sup>[3]</sup>、最大熵<sup>[4]</sup>、最近邻<sup>[5]</sup>等,这些方法部分克服了上述缺点,一定程度上实现了分类器的自动生成,被广泛应用于各个领域。然而,机器学习方法在构建分类器之前通常需要繁杂的人工特征工程,这限制了其进一步发展。

2012年之后,深度学习算法引起了研究者的广泛关注。深度学习为机器学习建模提供了一种直接端到端的解决方案,可避免复杂的特征工程。GolVe<sup>[6]</sup>和word2vec<sup>[7]</sup>等词向量模型的提出,使深度学习算法成功地应用到文本处理领域,随后出现了各种基于深度神经网络(Deep Neural Network, DNN)的文本分类方法。这些方法主要采用卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)和注意力机制等深度学习技术,并且取得了比传统方法更为出色的性能。近年来,图卷积网络(Graph Convolutional Network, GCN)、区域嵌入和元学习等一些新的深度学习方法也被应用于文本分类领域。

本文对基于深度神经网络的文本分类技术进行介绍和分析,阐述卷积神经网络、循环神经网络和注意力机制等方法在文本分类中的应用和发展情况,总结各类方法的特点及区别,并对不同方法的性能表现和适用场景进行比较,讨论在应用深度学习方法处理文本分类任务时应当注意的问题。在此基础上,指出针对该技术未来的研究方向。

## 1 基于深度学习的文本分类方法

传统机器学习方法采用词袋(Bag-of-Words, BoW)模型表示文本,导致文本特征数据出现高维、稀疏问题,影响文本分析的效率和性能。随着深度学习技术的发展,出现了更为有效的文本表示方法。文献[8]提出的神经网络语言模型(Neural Network Language Model, NNLM)将单词映射到低维稠密空间,采用词向量度量单词之间的语义相关性。文献[9]提出的word2vec模型实现了词向量的高效计算,被广泛应用于自然语言处理领域。

基于深度学习的文本分类方法以词向量对词语进行语义表示,然后通过语义组合的方式获得句子和文档的语义表示<sup>[10]</sup>。神经网络语义组合方法主要包括卷积神经网络、循环神经网络和注意力机制等,这些方法采用了不同的组合方式将词级别的语义组合到句子和文档级别的语义。本节阐述卷积神经网络、循环神经网络和注意力机制在深度学习文本分类中的应用情况,介绍相应分类方法,同时探讨其他类型的深度学习分类模型。

### 1.1 卷积神经网络

卷积神经网络最初是为图像处理而构建的,其结构类似于视觉皮层,之后被有效用于文本分类。卷积神经网络为文本分类提供了一种直接端到端的解决方案,相比传统机器学习方法,其不需要开发者准备先验信息,并且能够避免复杂的人工特征工程。文献[11-13]将卷积神经网络应用于文本分类,分别使用了两层卷积网络、基于单词向量的单层卷积网络和基于字符向量的多层卷积网络。

文献[11]提出基于两层卷积网络结构的分类模型DCNN。该模型采用一维卷积,不改变输入序列宽度,其中池化层采用动态池化策略返回 $k$ 组最大值。通过一维卷积和动态池化,DCNN可以实现连续多层卷积。该研究验证了在情感分类和问题分类等任务上,神经网络方法性能优于朴素贝叶斯、支持向量机等传统分类方法。

文献[12]提出单卷积层结构的句子分类模型TextCNN。该模型对文本进行整行的卷积计算,从而使每次卷积的行数恰好对应 $n$ -gram语言模型中的词语个数 $n$ 。相比在输入向量的单一维度或部分维度上进行卷积计算(例如DCNN),整行卷积保持了每个单词的语义完整性,更具可理解性。因此,在自然语言处理中,单层CNN结构使用较多,此处单层包含一对卷积层和池化层。与传统机器学习方法以及RNN、DCNN等深度学习方法相比,TextCNN在一系列分类任务上取得了优异的结果,说明浅层神经网络能够在文本分类任务上实现准确分类。

文献[13]提出字符级的卷积神经网络分类模型char-CNN。该模型是一个采用字符级的独热(one-hot)向量输入的6层卷积网络模型,其以词向量表示字符,以字符向量取代单词向量作为模型输入,将多卷积层网络连接到一个双向循环层。char-CNN在短文本分类上取得了与单词向量方法相近的性能,但在新闻和百科分类上,单词向量方法性能更好。

将卷积神经网络直接应用到文本分类可取得较好的效果,但相关研究并未止步于此。针对深度分类模型复杂、深度网络梯度消失和短文本内容简短等问题,文献[14-16]分别提出了相应的解决方案。

文献[14]提出了fastText模型。该模型将输入词序列投射到词嵌入空间,然后通过一个平均池化层计算文档向量的值,再连接到一个全连接层得到分类结果。相比一般的卷积网络结构,fastText避免了卷积操作,可以看作是一个特殊的卷积网络模型。该模型还

采用  $n$ -gram 扩充特征词,以增加文本的时序特征。fastText 模型结构简单,但在多数分类任务上取得了与 VDCNN<sup>[17]</sup>和 char-CNN 等深度模型相当的分类性能,进一步说明浅层学习在文本分类问题上能够获得深度模型的学习效果。

文献[15]提出深度卷积网络分类模型 DPCNN。该模型参考深度残差网络(ResNet)<sup>[18]</sup>,采用跳层连接技术以缓解在增加模型深度时出现的梯度消失问题,其通过固定特征图的数量,采用步长为2的最大池化操作,使每个卷积层的数据规模和相应的计算时间减半,从而形成一个金字塔。实验结果表明,在用户评论和新闻数据分类任务上,DPCNN 的准确率超过了 char-CNN 和 VDCNN 等分类方法。

由于短文本内容简短,因此会导致分类模型缺少足够的特征进行类别判断。对此,文献[16]提出结合知识的 CNN 短文本分类模型 KPCNN。该模型将知识库丰富的短文本信息与每篇短文和知识库中的相关概念进行结合,使预训练的词和概念的嵌入相互连接,形成对应的词-概念嵌入作为 CNN 的输入。该模型同时引入了字符级语义特征,将字符嵌入输入一个单独的 CNN,通过两个 CNN 子网络分别提取词-概念和字符的信息,并将两个子网络的输出相互连接,送入输出层得到分类结果。实验结果表明,结合知识库的概念信息和字符信息能够有效提升 CNN 的分类精度。

增加卷积神经网络模型的深度能否带来文本分类性能的提升,这是一个重要的问题。对此,文献[19]分别将基于字符输入的 CNN 模型和基于单词输入的 CNN 模型应用于5个标准文本分类和情感分析任务中进行实验分析,结果表明:当以字符序列作为文本输入时,深度模型能够提供比浅层模型更好的性能,但是提升效果不显著;当以单词序列作为文本输入时,简单的浅层网络 TextCNN 的性能可以达到甚至优于 DenseNet<sup>[20]</sup>等深层模型。此外,单词级浅层网络性能要优于深层字符级网络,这也证明文本分类中单词观察比字符特征提取更有效。

卷积神经网络在文本分类任务中得到了成功应用,特别是浅层卷积网络模型,更适合处理文本分类任务。然而,卷积神经网络需要开发者指定精确的模型结构并设置相应的超参数,这要求开发者必须以自己的经验确定哪些参数应花费精力调优,哪些参数设置对性能没有影响,并且找出独立于特定数据集的最佳设置。下文以 TextCNN 为例,分析使用卷积神经网络训练文本分类模型时应当注意的问

题。根据上文的实验结论和文献[21]对 TextCNN 的敏感度分析,对影响卷积神经网络性能的各项参数进行总结,如表1所示。

表1 卷积神经网络分类模型参数说明

Table 1 Parameters description of classification model based on convolutional neural network

参数	说明
输入	可以采用预先训练好的词向量初始化输入
卷积窗口大小	卷积窗口大小对性能影响较大,在具体任务中应对这一参数进行调优;组合使用多个不同窗口大小的卷积核,相比采用单一大小的卷积核能够获得更优性能
特征图数量	特征图数量的最佳取值依赖于具体数据集,文本任务可在50~600之间搜索特征图数量的最佳值
卷积层激活函数	Identity、ReLU 和 tanh 性能较好,一般可选择 ReLU 作为激活函数
池化方法	在 1-max、k-max 和局部平均池化等方法中,1-max 池化性能最优
正则化策略	相比 dropout、L1 和 L2 正则化方法,BN 算法防止过拟合的效果不显著

输入文本的特征表示在文本分类中非常重要,传统机器学习一般采用 one-hot 模型表示文本特征。但因为句子的信息简短,不能向高维编码提供足够的信息,所以采用 one-hot 模型表示的 CNN 不适合文本分类,而词嵌入将词映射到稠密特征空间,更适合神经网络计算。词向量可以采用随机值初始化,也可以采用预先训练好的词向量进行初始化,常用的方法有 word2vec 和 GloVe。在实际应用中,采用预先训练好的词向量初始化,可以加快训练迭代过程。对于词嵌入维度大小,一般可以在 50~200 之间寻找最佳取值。对于文本词典大小,并不是越大越好,词典过大反而可能降低文本分类的准确率。

CNN 超参数的设定非常重要,其中包括卷积窗口大小的设置。由于文本卷积一般由整行卷积计算构成,因此只需要确定每次卷积的行数即可,这个行数可以对应  $n$ -gram 模型中  $n$  的取值,例如按照 2、3、4 这样来取值。不同的数据集有着不同的最优窗口大小,一般文本长度越长,窗口应该越大。针对目标数据集,首先应找到单一卷积窗口的最佳值,然后在最佳值附近选取多个不同窗口大小的卷积核,再进行不同卷积核的组合,找到最优组合。特征图的数量设置也很重要,合适的特征图数量能够提高分类性能,但找到一个合适值非常耗时,增大特征图数量会增加训练时间。为了防止过拟合,模型通常在全连接层加入失活率,从而随机地舍弃一部分连接。



## 1.2 循环神经网络

卷积神经网络在很多任务中有不错的表现,但其需要固定卷积窗口的大小,导致无法建模更长的序列信息,而循环神经网络恰好可以避免这一问题。在文本分类任务中,实际使用较多的长短期记忆(Long Short Term Memory, LSTM)网络,从某种意义上其原理可以理解为捕获变长的 $n$ 元语言信息。RNN能够处理变长文本,而CNN一般则要按照固定长度对文本进行截取,相比RNN, CNN超参调节过程也很繁琐。因此, RNN是自然语言处理领域的常用模型。本节首先介绍LSTM和门控循环单元(Gated Recurrent Unit, GRU)的应用,然后介绍RNN在文本分类中应用的一些新方法,最后分析比较CNN和RNN的文本分类性能。

LSTM网络由HOCHREITER等人于1997提出<sup>[22]</sup>,其是一种特殊的RNN,能够学习长期依赖关系。此后,研究者对其进行了改进和推广,在处理很多问题时都取得了成功。GRAVES于2012年给出LSTM的一个具体实现方案<sup>[23]</sup>,本文对该文提出的模型进行了分类测试,发现其分类性能不如TextCNN。LSTM的另一个变体是GRU,由CHO于2014年提出<sup>[24]</sup>。GRU将输入门和遗忘门合并为一个单一的更新门,同时合并了细胞状态和隐含状态,因而其结构较标准LSTM模型更简单。在一些分类任务中,GRU性能超过了LSTM。

对于自然语言处理任务,如客户评论情感分析,循环神经网络模型要解析整条评论内容后方能作出结论,然而在实践中,对于容易分类的评论不需要读取整个输入。对此,文献[25]提出一种快速阅读分类方法,其受人类阅读技巧的启发,通过一个智能循环代理评估当前片段的重要性,以决定是进行预测还是跳过某些文本,或者重读部分语句。该方法的具体模型由一个LSTM和一个策略模块构成,其中策略模块控制模型的阅读和决策。在同一精度水平下,该方法能够有效加速模型计算。文献[26]提出基于信息共享机制的循环神经网络模型Multi-RNN,其采用多层LSTM结构,利用多任务数据训练模型,能够通过多个任务的训练数据提高单一任务的分类性能。

相比CNN, RNN能够方便地实现多层叠加,构成多层神经网络,例如多层LSTM或者多层GRU,因此,在序列标注、命名实体识别、seq2seq模型等很多场景都有应用。对于分类任务,文献[26]对CNN和RNN进行了测试,其中RNN采用文献[23]提出的LSTM实现, CNN分别采用TextCNN<sup>[12]</sup>和DCNN<sup>[11]</sup>实现。测试结果表明,在电影评论情感分类任务上, RNN的性能不如TextCNN、DCNN等卷积网络方

法,说明对于文本分类任务,卷积网络方法更适合构建文本的语义表示。此外,对于基于词向量的循环神经网络模型,在网络深度超过两层后,分类性能提升会很小,但深层模型能给字符型输入的网络(如char-CNN、char-RNN等)带来性能提升。

## 1.3 CNN-RNN组合模型

卷积操作检测特征具有平移不变性,但序列数据中的长期依赖关系不能由卷积神经网络直接获得,而是需要组合多个卷积层计算得到。虽然循环神经网络能够从时序数据中学习长期依赖关系,但是在处理长文本时,可能出现模型训练不稳定的情况,在一些文本分类任务上性能不如CNN<sup>[27]</sup>。针对这两种网络在文本处理中的不足,一些研究尝试结合CNN和RNN,以期获得性能更好的神经网络模型。根据卷积层和循环层在网络架构中的先后关系,可以将这些模型分为两类,即先循环层后卷积层的RCNN和先卷积层后循环层的CRNN。

在RCNN方面,文献[28]针对卷积神经网络需要固定卷积核大小且可能导致上下文语义缺失的问题,提出一个循环卷积神经网络模型。该模型由1层LSTM和1层CNN组成,首先采用双向循环网络学习文本表示,利用循环神经网络学习中心词左侧和右侧的词表示,然后将这两个词表示连接起来作为中心词的表示,再输入到卷积层。该模型同时利用了CNN和RNN的优点,即利用RNN学习词序信息,利用CNN获取文本主要成分。

在CRNN方面,文献[29]提出字符型输入模型char-CRNN。该模型同样是将卷积神经网络与双向LSTM相结合,但顺序不同,其首先使用多层卷积网络处理输入文本的字符序列,然后通过池化函数缩短序列的长度,最后将生成的特征提供给一个双向LSTM层进行处理。char-CRNN采用与图像二维卷积相同的计算方式,可进行特征图的组合卷积,实现字符序列的连续多层卷积。

因为卷积层和池化层的局部性,现有的多数深度模型需要多层卷积才能捕获更高层次的平移不变性特征。在此过程中,池化层往往只捕获句子中最重要的特征,导致局部细节信息丢失。对此,文献[30]提出一个CRNN模型,以RNN代替CNN模型中的池化层。该模型首先使用一个卷积层来提取文档特征,然后以LSTM层代替池化层捕获长期依赖关系,从而通过去除池化层有效减少参数数量。

在多标签文本分类方面,针对现有方法在局部语义信息提取和标签相关性建模方面存在的不足,

文献[31]提出一个组合 CNN 和 LSTM 的多标签文本分类模型。该模型利用以词向量为输入的 CNN 提取文本特征,将 CNN 的输出向量作为 LSTM 的输入,通过 LSTM 对标签之间的相关关系进行建模,用于预测输入文档的标签序列。

#### 1.4 注意力机制

传统的编码器-解码器框架在编码时,可能会对不完全相关的信息进行编码。当输入很长或者输入信息丰富时,编码选择会变得更加困难。在文本摘要和机器翻译等任务中,输入文本和输出文本之间存在一定的对齐关系,这意味着每个生成步骤都与输入文本的特定部分高度相关,这种直觉激发了注意力机制。注意力机制试图通过允许编码器回顾输入序列来缓解上述问题,即编码器会基于一个上下文向量来调节输出<sup>[32]</sup>。

深度学习方法在处理文本数据时,通常采用求平均或者求和等方式将词汇合成为语句及文档的语义表示。实际上,文档的不同部分对文本信息的影响是不同的,因此,通过引入注意力机制,深度学习在逐层合成文本语义过程中,就会像人类阅读文章一样,对文本中不同句子以及句子中的不同词汇给予不同的关注度,从而实现更合理的自然语言建模。

文献[33]模型将注意力机制应用于机器翻译,提高了翻译性能。该模型利用多层感知器,根据解码器的最后一个隐藏状态确定注意力信号并将其作用到输入隐藏状态序列。注意力机制使得模型在预测一个目标单词时能够自动选择原句子的相关部分作为解码器的输入。此后,注意力机制也被应用到图像描述生成<sup>[34]</sup>、机器阅读理解<sup>[35]</sup>等其他自然语言处理任务中。

在文本分类任务上,文献[36]在注意力机制的基础上,考虑到文本是层次式结构,可以用词向量表示句子,再由句子向量表示文档,即采用“词-句子-文章”的层次化结构来表示一篇文本,提出层次化注意力机制模型 HAN。该模型包含两个层次的注意力机制,分别存在于词层次和句子层次,具有对文本中重要性不同的句子和词给予不同“注意力”的能力,其在提高分类性能的同时,还可以选择出含有丰富信息的词语和句子。

现有方法大多采用粗粒度注意力机制,当目标单词和上下文很长时会引入信息损失。文献[37]提出一种细粒度注意力机制来捕获单词级别的目标和上下文之间的交互,从而减少粗粒度注意力机制造成的损失,在此基础上,结合细粒度和粗粒度注意力形成多粒度注意力网络(Multi-Grained Attention Network, MGAN)。

在情感分类测试中, MGAN 性能优于基本循环神经网络和多种注意力机制方法,说明细粒度的交互可以为分类带来额外的有用信息。

在文本分析过程中,注意力机制能够根据目标对象重点关注与之密切关联的词语和句子,而面向目标的情感分析任务主要是分析句子中所描述对象的情感极性,因此,越来越多的研究将注意力机制应用到情感分析任务中。文献[38-39]使用注意力机制增强 LSTM,从而为分类提供额外的支持,实现了基于目标的情感分析,说明利用注意力机制的 RNN 能够较好地解决这个问题。文献[40]分析注意力机制和 CNN 在完成这一任务时存在的不足,提出一种利用 CNN 进行目标情感分析的模型,并且设计针对目标的变换单元,使目标信息能够被更好地表示。在该模型中,变换单元位于 RNN 层和 CNN 层之间,一方面将带有上下文信息的词特征结合目标信息以实现目标调整机制,另一方面保留上下文信息以实现上下文保存机制。文献[41]提出一种基于记忆网络的多跳注意力方法,将注意力机制应用于记忆网络。该方法采用多注意力计算层改进记忆网络中的信息区域查找过程,有效提高了情感分类性能。

#### 1.5 其他深度学习分类方法

近年来,深度学习在图像领域的应用进展领先于自然语言处理领域,当前文本分类的研究借鉴了深度学习在图像处理领域应用的研究成果,例如图卷积网络、区域嵌入、元学习等。此外,一些研究人员在利用深度学习解决文本分类问题时,还考虑使用强化学习和集成学习等方法。

##### 1.5.1 图卷积网络

CNN 和 RNN 模型考虑文本的顺序信息和局部信息,能够很好地捕获连续词序列中的语义信息,但忽略了全局的词共现,而词共现中携带了不连续及长距离的语义信息。基于图网络的神经网络模型能够更灵活地表征节点与节点之间的关系。因此,文献[42]将图卷积网络应用于文本分类,提出文本图卷积网络模型 Text GCN。该模型采用两层图网络,构建了单词和文档两种节点,相应地有两种类型边,边的权重以 TF-IDF 和 PMI 来表征,并利用 softmax 分类。Text GCN 能够获取全局词共现信息并使用有限标注的文档来执行所需任务,在多个基准数据集上表现出色。

##### 1.5.2 区域嵌入

短语的识别和表示在文本处理中非常有用,传统的 n-gram 模型可以看作是该方法的近似,但会受



到数据稀疏性的影响。针对多数 CNN、RNN 等深度学习模型忽略建模上下文信息重要性的问题,文献[43]提出区域嵌入方法 LRE,其原理是一个词的语义表示由自身含义及周围词的含义共同确定,分别对应词嵌入和本地上下文单元,然后使用词和相应的本地上下文单元生成区域嵌入。该方法在完成文本分类任务的同时,还能捕捉到文本中显著的短语表达。文献[44]指出 LRE 仅能捕获固定长度区域内的上下文信息,并且生成本地上下文单元需要消耗大量的内存空间,进而提出一种新的区域嵌入方法。该方法采用自适应上下文单元替换原方法中的本地上下文单元,通过为每个区域生成一个上下文单元,使每个单词与相应的上下文单元交互,然后产生区域嵌入。与原有的区域嵌入模型 LRE 相比,新模型的参数更少,能够利用一个小参数空间取得更优的分类性能,同时避免词语歧义。

### 1.5.3 元学习

元学习在计算机视觉领域表现出优越的性能,因为在计算机视觉中,低级模式可以在学习任务之间转移。然而将元学习直接应用于文本面临挑战,因为单词与任务具有相关性,在一类任务中起重要作用的单词,可能在另一类任务中就不重要了。文献[45]将动态路由算法和典型的元学习框架相结合,模拟人类的归纳能力,提出一个针对小样本文本分类的神经网络模型。该模型结合分布签名来学习单词编码,通过元学习框架将这些签名映射到注意力分数,然后计算单词的词汇表示。在小样本分类和关系分类上,该模型性能优于在词汇知识上学习的原型网络方法<sup>[46]</sup>。

### 1.5.4 强化学习

文本分类在很大程度上依赖于表示学习,文本分类的表示模型主要包括词袋模型、序列表示模型、基于注意力的模型和结构化表示模型。结构化表示模型能够抽取句子结构,但需要预先指定解析树,而在实际中事先往往不知道具体的句子表示结构。对此,文献[47]提出一种无需明确结构注释即可识别任务相关句子结构的强化学习方法。该方法使用一个结构化表示模型学习文本表示,包括信息提取 LSTM 和分层结构 LSTM 两部分。其中:信息提取 LSTM 能够删除不相关的单

词并提取一个句子中与任务相关的单词;分层结构 LSTM 能够发现与任务相关的结构并构建结构化句子表示。该表征方法通过识别重要单词或与任务相关的结构来学习文本表示,从而获得有竞争力的分类性能,为文本分类的表示学习提供了一个新思路。

### 1.5.5 集成深度学习

文献[48]提出一个随机多模型深度学习方法,其使用 DNN、CNN 和 RNN 3 种深度学习架构,随机生成各个模型的隐藏层个数和神经元节点数,然后由所有这些随机深度学习模型进行多数表决得出预测结果。该方法通过融合多种深度学习模型提高模型鲁棒性和准确性,能够处理包括文本、视频、图像在内的多种类型的输入。相比采用单一深度学习模型的方法,集成深度学习采用并行学习体系结构构建不同的深度神经网络分类模型,然后通过集成学习方法组合各个 DNN、RNN 和 CNN 模型,最后形成统一的分类框架,从而获得更高的分类精度,并可用于多种数据类型。在实际应用中,往往不是采用单一的深度模型,而是采用模型融合的方法建立多种深度模型,以此获得更好的性能表现。

## 2 深度学习文本分类方法的性能分析和比较

下文通过分析比较现有的深度学习分类方法,为进一步开展深度学习在文本分类中的应用提供借鉴。首先,总结多种深度学习方法在标准数据集上的分类性能。为了便于对各个方法进行比较,采用文献[13]建立的公开数据集,如表 2 所示,其中:平均长度指每篇文档平均包含的单词个数;AG 是一个新闻语料集,包括世界、体育、商业和科技 4 个类别;Sogou 是搜狗新闻数据集,笔者结合拼音包和分词工具将中文文本转换为拼音序列,这样英文模型就可以直接应用到该数据集;Dbpedia 是从维基百科中抽取的数据,包含 14 个类别;Yahoo! Answers 是从雅虎问答系统中抽取的问题答案数据集,包含 10 个主题类别;Yelp 是 Yelp 数据挑战赛的评论数据集;Amazon 是亚马逊用户的产品评价数据,用 Ama 表示;.P 表示极性二分类(正/负);.F 表示星级评价分类。数据细节和数据预处理过程此处不再详述。

表 2 数据集信息

Table 2 Information of datasets

比较项	AG	Sogou	Dbpedia	Yelp.P	Yelp.F	Yahoo!Answers	Ama.F	Ama.P
训练文档数	120K	450K	560K	560K	650K	1.4M	3M	3.6M
测试文档数	7.6K	60K	70K	38K	50K	60K	650K	400K
类别数	4	5	14	2	5	10	5	2
平均长度	45	578	55	153	155	112	93	91

表3按照时间顺序列出了各个方法的分类准确率,由于TextCNN在原文采用了不同的实验数据,因此此处引用文献[19]对该模型的测试结果。

1)逻辑回归。BoW、BoW TFIDF、n-gram以及n-gram IFIDF方法均使用逻辑回归进行文本分类,区别在于所用的特征计算方法不同。BoW和BoW TFIDF选择出现最频繁的50 000个单词作为特征,分别采用词频和词频逆文档频率计算特征值。n-gram选择出现最频繁的500 000个n元特征序列作为特征,特征值计算方法与BoW相同。

2)支持向量机(SVM)。SVM+Unigram同BoW一样使用词袋模型表示文本,然后用SVM训练分类器。SVM+Bigram采用二元语言模型表示文档。SVM+AverageSG采用word2vec词嵌入表示单词,然后用所有单词的词向量平均值表示文档,最后用支持向量机分类。

3)神经网络方法。表3中的神经网络方法在上节均有介绍,包括深度模型和浅层模型,分别记为Y和N。

对典型深度学习文本分类方法的特点和适用场景等进行分析,如表4所示。

表3 分类准确率比较

Table 3 Comparison of classification accuracies

分类方法	深层/浅层	Yelp.P	Yelp.F	Ama.P	Ama.F	AG	Sogou	Yahoo!Answers	DBP
BoW <sup>[13]</sup>	—	82.2	58.0	90.4	54.4	88.8	92.8	68.9	96.6
BoW TFIDF <sup>[13]</sup>	—	83.7	59.9	91.0	55.3	89.6	93.4	71.0	97.4
n-gram <sup>[13]</sup>	—	95.6	56.3	92.0	54.3	92.0	97.1	68.5	98.6
n-gram TFIDF <sup>[13]</sup>	—	95.4	54.8	91.5	52.4	92.4	97.2	68.5	98.7
SVM+Unigram <sup>[49]</sup>	—	—	61.1	—	—	—	—	—	—
SVM+Bigram <sup>[49]</sup>	—	—	62.4	—	—	—	—	—	—
SVM+AverageSG <sup>[49]</sup>	—	—	56.8	—	—	—	—	—	—
TextCNN <sup>[12]</sup>	N	95.9	64.9	—	—	92.2	—	73.0	98.7
char-CNN <sup>[13]</sup>	Y	94.7	62.0	94.5	59.6	87.2	95.1	71.2	98.3
char-CRNN <sup>[29]</sup>	Y	94.5	61.8	94.1	59.2	91.4	95.2	71.7	98.6
VDCNN <sup>[17]</sup>	Y	95.7	64.7	95.7	63.0	91.3	96.8	73.4	98.7
HN-ATT <sup>[36]</sup>	Y	—	71.0	—	63.6	—	—	75.8	—
FastText <sup>[14]</sup>	N	95.7	63.9	94.6	60.2	92.5	96.8	72.3	98.6
DPCNN <sup>[15]</sup>	Y	97.4	69.4	96.7	65.2	—	—	76.1	—
ARE <sup>[44]</sup>	N	96.6	65.9	95.9	62.6	93.1	97.5	74.9	99.1

表4 典型深度学习文本分类方法比较

Table 4 Comparison of typical text classification methods based on deep learning

分类方法	输入	网络结构	实验性能及重要结论
DCNN <sup>[11]</sup>	随机初始化词向量	2层卷积神经网络	在情感分类和问题分类任务上性能优于朴素贝叶斯、支持向量机、递归神经网络等传统分类方法
TextCNN <sup>[12]</sup>	采用Google预先训练的word2vec词向量	1层卷积神经网络	在SST和TREC数据集上与DCNN效果相当,在TREC上甚至略优于DCNN;采用预先训练好的word2vec结果初始化词向量,相比随机初始化能有效提高CNN方法的准确率;在目前主流的文本分类任务中,合理的浅层学习能够达到深度学习的效果
char-CNN <sup>[13]</sup>	采用字符级的one-hot向量表示输入	6层卷积神经网络	取得了同基于词向量CNN方法相当的性能;对于英文文本分类问题,采用n-gram特征的分类方法效果很好,特别对于传统的新闻、百科主题分类任务;相比词向量CNN,在新闻、百科主题分类任务(规则文本)上性能较差,在用户评论情感分类任务(不规则文本)上性能较优
FastText <sup>[14]</sup>	采用词向量表示输入,利用n-gram扩充特征词	1层嵌入层+平均池化+1层全连接层	取得与对比神经网络方法相近的性能,但计算时间更少
Multi-RNN <sup>[26]</sup>	采用维基百科数据训练word2vec向量	多层LSTM	在电影评论情感分类任务上性能不如TextCNN和DCNN
RCNN <sup>[28]</sup>	采用维基百科数据训练word2vec向量	1层循环神经网络+1层卷积神经网络	对于多数分类任务,神经网络方法(RecursiveNN、CNN、RCNN)性能优于采用词袋表示的传统分类方法;卷积神经网络方法性能优于其他神经网络方法,更适合构建文本的语义表示
char-CRNN <sup>[29]</sup>	采用字符的词向量表示	多层卷积神经网络+1层双向LSTM	在百科、新闻、网络评论等大规模文本分类任务中取得了同卷积神经网络相近的性能,更适合处理类别数目巨大、语料数量较少的任务

由表3和表4可以看出:

1)在传统机器学习方面:对于大多数文本分类任务,神经网络方法性能优于传统的机器学习方法,例如在情感分类和问题分类任务上,表3中的神经网络方法性能基本都超过了逻辑回归和支持向量机这两个经典的分类模型。

2)在浅层模型方面:单卷积层的TextCNN采用预先训练好的word2vec词向量初始化,相比随机初始化可以提高CNN方法的准确率;FastText采用n-gram扩充特征词,然后将词向量输入到一个全连接网络,取得了同基于词向量CNN方法相当的性能;TextCNN和FastText方法的较优性能,说明对于文本分类任务,合理的浅层学习能够达到深度学习的效果;ARE使用区域嵌入方法捕获单词上下文信息,类似于FastText采用的n-gram模型,同样取得了很好的分类效果。

3)在深度模型方面:char-CNN和char-CRNN的较优性能,说明对于大规模文本数据的分类任务,字符级的深度模型能够取得同词向量卷积神经网络相近的性能;HN-ATT使用注意力机制,通过对输入词

或句子进行差分评分来建立表示,相比单纯采用序列表示模型的深度CNN或RNN取得了更好的分类效果,说明基于注意力机制的表征学习方法适用于自然语言处理;同深度残差网络在图像识别上的优秀表现一样,DPCNN在文本分类问题上同样表现出色。

为更好地分析神经网络方法的分类性能,本文对CNN、LSTM、GRU、MLP等基本网络进行性能测试,其中CNN分类采用了经典的TextCNN实现。此外,还增加了一个组合模型CRNN和一个浅层模型FastText。所有模型均调用Keras库实现,未做特别优化,输入均采用随机初始化向量。Keras是一个由Python编写的开源人工神经网络库,可以作为Tensorflow的高阶应用程序接口进行深度学习模型的设计、调试、评估、应用和可视化。IMDB情感分类数据集是Stanford整理的一套影评情感数据,含有25 000个训练样本和25 000个测试样本,可用于文本情感分类。表5比较了不同分类模型的网络结构、精度和运行时间,其中运行时间包括模型的训练时间和分类测试时间。

表5 深度学习基础网络结构IMDB分类测试结果  
Table 5 Classification test results of deep learning basic network on IMDB

分类模型	网络结构	精度	运行时间(训练+测试)
CNN	1层卷积神经网络	0.881	21 min 9 s+42 s
双向LSTM	1层双向循环神经网络	0.850	371 min 2 s+287 s
GRU	2层门控循环单元	0.859	7 min 47 s+33 s
MLP	2层全连接神经网络	0.836	8 min 23 s+9 s
CRNN	1层卷积神经网络+1层LSTM	0.869	18 min 9 s+34 s
FastText	1层嵌入层+平均池化+1层全连接层	0.888(unigram)	2 min 21 s+4 s
		0.904(bigram)	103 min 50 s+147 s

测试结果表明,对于短文本分类任务,卷积神经网络性能优于循环神经网络和多层感知器,比较适合构建文本的语义表示。在常用的循环神经网络中,相比LSTM网络,门控循环单元分类效果更好。FastText先提取n-gram特征得到特征序列,再对每个特征做词嵌入操作,将句子中的所有词向量相加求平均,得到句子的向量表示。采用unigram时不增加特征词。同其他方法的特征词集合一样,FastText取得了与CNN相当的性能,但训练时间更短。通过引入bigram特征和增加二元特征序列,FastText的分类性能甚至超过了CNN,但需要更大的计算开销,此时FastText模型的训练时间为采用unigram特征的50倍左右,为TextCNN的5倍左右。

### 3 未来研究方向

#### 3.1 新的文本处理深度架构

神经网络方法在文本分类任务中得到了较多应用,然而从已有研究和本文分析来看,深度模型未被证明比浅层模型更为有效。当前文本分类所采用的深度模型,来源于最初为图像处理开发的深度模型,其中图像被表示为真实稠密数据。不同于图像,文本是离散稀疏数据,同样结构的深度网络应用到不同类型的数据上,可能产生不同的结果<sup>[19]</sup>。例如,在卷积神经网络中,图像在经过卷积层后输出依然是二维矩阵,而文本在经过卷积层后一般会变成一维向量。此外,还应在其他自然语言处理任务以及更



多数据集,特别是中文文本集上,验证深度模型的有效性。

### 3.2 结合知识的深度模型

当前的深度模型分类方法同统计机器学习算法原理一样,模型没有利用已有知识,不能像人一样理解文本的语义。结合知识,特别是利用HowNet、WordNet等知识库进行自然语言处理,将会给深度学习模型带来新的提升。例如,文献[50]在中文微博情感分类中,利用HowNet情感词集对句子进行词性标注,将输入句子的词性向量和句子内容层面的词向量结合作为卷积神经网络的输入,对结合知识建立深度模型进行了初步尝试。未来,结合知识库和知识图谱建立自然语言处理的深度模型,会是打破现有深度学习方法性能瓶颈的可能途径。

在许多实际的文本分析场景中有大量的未标记数据,但缺少足够的标记数据训练深度模型。虽然目前已出现一些无监督或半监督方法,例如针对小样本类别的零样本学习策略,但这些方法仍处于研究探索中,需要设计能够充分利用未标记数据的深度学习方法。利用大量的未标记数据构建知识库或知识图谱,然后结合知识建立深度模型将是未来的一个研究方向。此外,采用强化学习方法同样可以减少数据标记工作。

### 3.3 对抗学习

对抗生成网络(Generative Adversarial Network, GAN)<sup>[51]</sup>为传统机器学习提供了一种新的学习模式,生成网络和判别网络利用“对抗”的思想学习模型,生成网络的目标是生成尽量真实的数据,例如文本、图像、视频,判别网络的目标是尽量区分生成对象的真假,这就像人与人之间通过竞争而不断进步的情况。GAN可被广泛应用于自然语言处理领域,近年来在图像描述(图像到文本)<sup>[52]</sup>、机器翻译(文本到文本)<sup>[53]</sup>、图像生成(文本到图像)<sup>[54-56]</sup>等任务中得到了成功应用,其通过对抗学习增强人工神经网络的预测和生成能力,为自然语言处理和文本分析提供了一个新的技术途径,例如可以采用对抗学习方法,以大量无监督句子和少量带标签的句子训练生成模型,模型最终可以生成特定情感倾向的句子。利用该模型可以进行样本自动生成,从而解决深度学习中样本数据获取困难的问题。

## 4 结束语

本文阐述深度学习在文本分类中的应用和研究进展,对现有多种基于深度学习的文本分类方法进

行分析和比较,并指出未来的研究方向。深度学习为文本分类提供了一个新的解决方案,相比传统的文本分类方法,深度学习可以直接从输入中学习特征表示,避免复杂的人工特征工程,简化文本预处理工作,使研究者更专注于数据挖掘算法本身。深度学习模型通过改进表征学习方法和模型结构,能够获得比传统统计学习模型更好的分类性能,特别是卷积神经网络,其在实际应用中具有较好的分类性能和泛化能力。现有研究多针对英文文本建立深度模型,同时采用英文数据集进行性能评测,而这些方法对中文文本是否同样有效,需要做进一步的研究和验证。

### 参考文献

- [1] LI H, YAMANISHI K. Text classification using ESC-based stochastic decision lists [C]//Proceedings of the 8th International Conference on Information and Knowledge Management. New York, USA: ACM Press, 1999: 122-130.
- [2] KIM S B, HAN K S, RIM H C, et al. Some effective techniques for naive Bayes text classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(11): 1457-1466.
- [3] JOACHIMS T. Text categorization with support vector machines: learning with many relevant features [M]. New York, USA: Springer US, 1998.
- [4] KAZAMA J, TSUJII J. Maximum entropy models with inequality constraints: a case study on text categorization [J]. Machine Learning, 2005, 60(1/2/3): 159-194.
- [5] SOUCY P, MINEAU G W. A simple KNN algorithm for text categorization [C]//Proceedings of 2001 IEEE International Conference on Data Mining. Washington D. C., USA: IEEE Press, 2001: 647-648.
- [6] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation [C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. San Diego, USA: ACL, 2014: 1532-1543.
- [7] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [EB/OL]. (2013-01-16) [2020-08-10]. <https://arxiv.org/pdf/1301.3781.pdf>.
- [8] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model [J]. Journal of Machine Learning Research, 2003, 3: 1137-1155.
- [9] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2013: 3111-3119.
- [10] LAI Siwei. Research on word and document semantic

- vector representation methods based on neural network[D]. Beijing: University of Chinese Academy of Sciences, 2016. (in Chinese)
- 来斯惟. 基于神经网络的词和文档语义向量表示方法研究[D]. 北京: 中国科学院大学, 2016.
- [11] KALCHBRENNER N, GREFFENSTETTE E, BLUNSON P. A convolutional neural network for modelling sentences[EB/OL]. (2014-04-08)[2020-08-10]. <https://arxiv.org/pdf/1404.2188.pdf>.
- [12] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. San Diego, USA: ACL, 2014: 1746-1751.
- [13] ZHANG X, ZHAO J B, LECUN Y. Character-level convolutional networks for text classification[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2015: 649-657.
- [14] JOULIN A, GRAVE E, BOJANOWSKI P, et al. Bag of tricks for efficient text classification[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers). San Diego, USA: ACL, 2017: 427-431.
- [15] JOHNSON R, ZHANG T. Deep pyramid convolutional neural networks for text categorization[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). San Diego, USA: ACL, 2017: 562-570.
- [16] HONG S, OH J, HAN B, et al. Learning transferrable knowledge for semantic segmentation with deep convolutional neural network[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 3204-3212.
- [17] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for natural language processing[EB/OL]. (2016-06-06)[2020-08-10]. <https://arxiv.org/pdf/1606.01781v1.pdf>.
- [18] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [19] LE H, CERISARA C, DENIS A. Do convolutional networks need to be deep for text classification? [C]//Proceedings of AAAI Workshop on Affective Content Analysis. [S. l.]: AAAI, 2018: 29-36.
- [20] HUANG G, LIU Z, MAATEN V, et al. Densely connected convolutional networks[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 770-778.
- [21] ZHANG Y, WALLACE B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification[EB/OL]. (2015-10-13)[2020-08-10]. <https://arxiv.org/pdf/1510.03820.pdf>.
- [22] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [23] GRAVES A. Supervised sequence labelling with recurrent neural networks[M]. Berlin, Germany: Springer, 2012.
- [24] CHO K, MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [EB/OL]. (2014-06-03)[2020-08-10]. <https://arxiv.org/pdf/1406.1078.pdf>.
- [25] YU K Y, LIU Y, SCHWING A, et al. Fast and accurate text classification: skimming, rereading and early stopping[C]//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: [s. n.], 2018: 1-5.
- [26] LIU Pengfei, QIU Xipeng, HUANG Xuanjing. Recurrent neural network for text classification with multi-task learning[C]//Proceedings of International Joint Conference on Artificial Intelligence. [S. l.]: AAAI, 2016: 2873-2879.
- [27] DAI A, LE Q V. Semi-supervised sequence learning[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2015: 3079-3087.
- [28] LAI Siwei, XU Liheng, LIU Kang, et al. Recurrent convolutional neural networks for text classification[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2015: 2267-2273.
- [29] XIAO Y J, CHO K. Efficient character-level document classification by combining convolution and recurrent layers[EB/OL]. (2016-02-01)[2020-08-10]. <https://arxiv.org/pdf/1602.00367.pdf>.
- [30] HASSAN A, MAHMOOD A. Convolutional recurrent deep learning model for sentence classification[J]. IEEE Access, 2018, 6: 13949-13957.
- [31] CHEN Guibin, YE Deheng, XING Zhenchang, et al. Ensemble application of convolutional and recurrent neural networks for multi-label text categorization[C]//Proceedings of 2017 International Joint Conference on Neural Networks. Washington D. C., USA: IEEE Press, 2017: 2377-2383.
- [32] TOM Y, DEVAMANYU H, SOUJANYA P, et al. Recent trends in deep learning based natural language processing [J]. IEEE Computational Intelligence Magazine, 2018, 13(3): 55-75.
- [33] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]//Proceedings of the 3rd International Conference on Learning Representations. Banff, Canada: [s. n.], 2014: 1-5.
- [34] XU K, BA J, KIROUS R, et al. Show, attend and tell: neural image caption generation with visual attention [C]//Proceedings of the 32nd International Conference on Machine Learning. [S. l.]: JMLR, 2015: 2048-2057.
- [35] HERMANN K, KOISK T, GREFFENSTETTE E, et al. Teaching machines to read and comprehend [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2015: 1693-1701.

- [36] YANG Z C, YANG D Y, DYER C, et al. Hierarchical attention networks for document classification [C]//Proceedings of 2016 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies; Human Language Technologies. San Diego, USA: ACL, 2016: 1480-1489.
- [37] FAN Feifan, FENG Yansong, ZHAO Dongyan. Multi-grained attention network for aspect-level sentiment classification [C]//Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. San Diego, USA: ACL, 2018: 3433-3442.
- [38] WANG Yequan, HUANG Minlie, ZHU Xiaoyan, et al. Attention-based LSTM for aspect-level sentiment classification [C]//Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. San Diego, USA: ACL, 2016: 606-615.
- [39] MA Y K, PENG H Y, CAMBRIA E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM [C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2018: 5876-5883.
- [40] LI X, BING L D, LAM W, et al. Transformation networks for target-oriented sentiment classification [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. San Diego, USA: ACL, 2018: 946-956.
- [41] TANG Duyu, QIN Bing, LIU Ting. Aspect level sentiment classification with deep memory network [C]//Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. San Diego, USA: ACL, 2016: 214-224.
- [42] YAO Liang, MAO Chengsheng, LUO Yuan. Graph convolutional networks for text classification [C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2019: 7370-7377.
- [43] QIAO Chao, HUANG Bo, NIU Guocheng, et al. A new method of region embedding for text classification [C]//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018: 6-10.
- [44] XIANG Liuyu, JIN Xiaoming, YI Lan, et al. Adaptive region embedding for text classification [C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2019: 7314-7321.
- [45] BAO Yujia, WU Menghua, CHANG Shiyu, et al. Few-shot text classification with distributional signatures [C]//Proceedings of the 8th International Conference on Learning Representation. Addis Ababa, Ethiopia: [s. n.], 2020: 1-5.
- [46] JAKE S, KEVIN S, ZEMEL R. Prototypical networks for few-shot learning [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 4077-4087.
- [47] ZHANG Tianyang, HUANG Minlie, ZHAO Li. Learning structured representation for text classification via reinforcement learning [C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2018: 6053-6060.
- [48] KOWSARI K, HEIDARYSAFA M, BROWN D E, et al. RMDL: random multimodel deep learning for classification [C]//Proceedings of the 2nd International Conference on Information System and Data Mining. New York, USA: ACM Press, 2018: 19-28.
- [49] TANG Duyu, QIN Bing, LIU Ting. Document modeling with gated recurrent neural network for sentiment classification [C]//Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing. San Diego, USA: ACL, 2015: 1422-1432.
- [50] CHEN Ke, LIANG Bin, KE Wende, et al. Chinese micro-blog sentiment analysis based on multi-channels convolutional neural networks [J]. Journal of Computer Research and Development, 2018, 55(5): 945-957. (in Chinese)  
陈珂, 梁斌, 柯文德, 等. 基于多通道卷积神经网络的中文微博情感分析 [J]. 计算机研究与发展, 2018, 55(5): 945-957.
- [51] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2014: 2672-2680.
- [52] DAI B, FIDLER S, URTASUN R, et al. Towards diverse and natural image descriptions via a conditional GAN [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2989-2998.
- [53] YANG Zhen, CHEN Wei, WANG Feng, et al. Improving neural machine translation with conditional sequence generative adversarial nets [C]//Proceedings of 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, USA: ACL, 2018: 1346-1355.
- [54] ZHANG Han, XU Tao, LI Hongsheng, et al. StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 5908-5916.
- [55] XU Tao, ZHANG Pengchuan, HUANG Qiuyuan, et al. AttnGAN: fine-grained text to image generation with attentional generative adversarial networks [C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 1316-1324.
- [56] QIAO Tingting, ZHANG Jing, XU Duanqing, et al. MirrorGAN: learning text-to-image generation by redescription [C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 1505-1514.