计算机研究与发展 DOI:10.7544/issn1000-1239.2018.20160947
Journal of Computer Research and Development 55(5):933-944,2018

基于双向分层语义模型的多源新闻评论情绪预测

张 莹 王 超 郭文雅 袁晓洁
(南开大学计算机与控制工程学院 天津 300350)
(yingzhang@nankai.edu.cn)

Multi-Source Emotion Tagging for Online News Comments Using Bi-Directional Hierarchical Semantic Representation Model

Zhang Ying, Wang Chao, Guo Wenya, and Yuan Xiaojie
(*College of Computer and Control Engineering, Nankai University, Tianjin* 300350)

Abstract With the rapid growth of news services, users can now actively respond to online news by expressing subjective emotions. Such emotions can help understand the preferences and perspectives of individual users, and thus may facilitate online publishers to provide users with more relevant services. Research on emotion tagging has obtained promising achievement, but there are still some problems: Firstly, traditional methods regard a document as a flow or bag of words, which cannot extract the logical relationship features among sentences appropriately. Therefore, these methods cannot express the semantic of the document properly when there exists logical relationship among the sentences in the document. Secondly, these methods use only the semantic of the document itself, while ignoring the accompanying information sources, which can significantly influence the interpretation of the sentiment contained in documents. In order to solve these problems, this paper proposes a hierarchical semantic representation model of news comments using multiple information sources, called bi-directional hierarchical semantic neural network (Bi-HSNN), which not only captures the sentiment among words in sentences, but also applies a bottom-up way to learn the logical relationship among sentences in the document. This paper tackles the task of emotion tagging on comments of online news by exploiting multiple information sources including comments, news articles, and the user-generated emotion votes. A series of experiments on real-world datasets have demonstrated the effectiveness of the proposed model.

Key words online news comments; emotion tagging; hierarchical semantic representation; multiple information sources; neural networks

摘 要 随着在线新闻服务的迅猛发展,用户在阅读新闻后可以非常方便地表达自己的主观情绪,有效分析和预测用户的情绪有助于新闻服务提供商为新闻用户提供更好的服务.情绪标注研究已经取得了

收稿日期:2016-12-12;**修回日期**:2017-08-09

基金项目:国家自然科学基金项目(61402243);天津市自然科学基金项目(16JCQNJC00500);国家"八六三"高技术研究发展计划基金项目(2015AA015401);教育部-中国移动科研基金项目(MCM20150507)
This work was supported by the National Natural Science Foundation of China (61402243), the Natural Science Foundation of Tianjin (16JCQNJC00500), the National High Technology Research and Development Program of China (863 Program) (2015AA015401), and the Research Fund of Ministry of Education and Chinese Mobile Communications Corporation (MCM20150507).

通信作者:袁晓洁(yuanxj@nankai.edu.cn)

很多优秀的成果,但仍然存在着一些问题:1)传统的方法将整个文档看作单词流或词袋,不能对句子间的逻辑关系进行建模,在文档中的句子间包含逻辑关系时,这些方法无法适当地表达文档的语义;2)这些方法只用了文档本身的语义,忽略了与该文档相关的其他信息源中信息,而这些信息源对该文档的语义表达也有一定的影响.为了解决这些问题,提出了一种基于多信息源的在线新闻评论双向分层语义表示模型,称为双向分层语义神经网络(bi-directional hierarchical semantic neural network, Bi-HSNN),该模型既捕获句子中词语所表达的情感,又自底向上地学习文档中句子间的逻辑关系,并利用评论、新闻和用户投票等多种信息源对在线新闻评论的情绪进行标注.在真实数据集上的一系列实验,验证了该模型的有效性.

关键词 在线新闻评论;情绪标注;分层语义表示;多信息源;神经网络

中图法分类号 TP301.6

随着互联网技术和各种类型互联网服务的迅速发展,人们获取信息的途径发生了根本性的转变.其中,在线新闻服务成为了一种新型信息传递形式,每天吸引着数以亿计的在线用户,并允许用户在阅读新闻后发表评论表达自己的观点.由于在线新闻信息的传播速度快、受众范围广,有效分析用户的情绪有助于在线新闻服务商为用户提供更好的服务,也有助于政府及时了解舆情,对互联网内容进行更加有效的监管.因此,自动判别在线新闻评论的情绪具有重要的理论意义和应用价值.

一种很自然的判别在线新闻评论的方法是根据用户撰写的评论内容,将评论划分到多类情绪标签中的某一类中,即将该任务看作多分类问题.近看来,情绪标注的研究已经取得了诸多成果,但仍然存在一定的问题:

1) 传统的情绪标注主要采用机器学习方法,使用人工的方式抽取特征,并使用分类器进行分类,此类方法的性能依赖于特征的选取[1],不能与分析需求很好地结合,也不能准确地表达高层语义;

2) 现有工作大多没有充分考虑文本中句子之间的逻辑关系,句子之间的因果关系、转折关系等对整个文本的语义有重要的影响,转折之后的内容相较于转折之前的部分对整个文本的语义有更大的影响,在进行情绪标注时,充分利用句子之间的逻辑关系才能得到更好的效果;

3) 现有工作没有充分利用其他相关信息源,目前大部分工作只用研究对象的文本内容的一种信息源进行挖掘,以在线新闻评论的情绪标注为例,新闻评论固然是体现读者阅读情绪的直接信息,但新闻的文本内容也是影响用户情绪的直接原因,另外,新闻网站提供情绪投票服务,用户对该新闻的情绪投票也可帮助预测该评论所蕴含的情绪,因此新闻评论的文本内容、新闻文档的文本内容以及用户情绪投票结果均可从不同角度帮助用户情绪的预测.

为了解决上述问题,本文提出了一种基于多信息源的双向分层语义表示模型,称为双向分层语义神经网络(bi-directional hierarchical semantic neural network, Bi-HSNN).该模型采用深度学习的方法,从数据中自动学习特征,解决了传统机器学习中人工抽取特征的问题;构建分层语义表示模型,既抽取文本的语义关系,又自底向上地学习文档中句子间的逻辑关系;另外,利用新闻评论文本、与评论相关的新闻文本、新闻的用户情绪投票3个信息源,对在线新闻评论进行情绪标注.

本文描述了在新浪新闻和腾讯新闻2个真实数据集上的一系列实验结果和相关分析,比较了各单信息源和不同信息源的组合方案,验证了利用多信息源进行在线新闻情绪标注的必要性;在真实数据集上的实验表明,本文提出的Bi-HSNN模型相较于已有基于传统机器学习的情绪标注方法和基于其他神经网络模型的方法具有更好的情绪标注效果.

1 相关工作

情绪标注(emotion tagging)是意见挖掘和情绪分析(opinion mining and emotion analysis)领域最重要的子课题之一,近几年受到了越来越多学者的关注.1992年Hearst[2]首次提出了基于文本的情绪分析和意见挖掘问题,此后学术界出现了大量关于文本情绪分析和意见挖掘的研究工作[3].情绪标注是意见挖掘和情绪分析领域重要的子课题之一[4],该子课题要求给出一个能够预测出一个文本所对应的情绪(如产品评价[5-7]、新闻文档[8-10]、新闻评论[11-12]等)的解决方案.本文主要关注在线新闻评论的情绪标注问题.

机器学习技术广泛应用于情绪标注问题上,目前大多数的研究将重心放在特征模式的设计上,通过特征工程技术预先人工定义的特征模式抽取出相关的特征,使用支持向量机(support vector machine, SVM)等分类器将文本特征分类到某一情绪类别中,这些方法一般只使用单词级别的特征对文本进行分类.与此不同,前期工作[8-9,13]作出假设,认为相较于文本中具体的某一单词或关键字,文本的情绪应该和文本所描述的话题(topic)更具有相关性,进而提出了一系列情绪-话题模型.

自2006年Hinton等人[14]将深度学习方法的研究成果发表于《Science》,深度学习方法在情绪标注领域得到了很好地应用.

目前基于深度学习的文本情绪分析方法主要通过自然语言处理相关技术对文本的情绪和语义进行分析. Ranzato等人[15]提出基于深度学习的文本表达学习模型,利用深度多层自动编码网络,逐层学习文本的高层语义特征.同时,Weston等人[16]提出深度嵌入的方式学习特征的层次表达,同时利用不同层次的特征实现不同特征层之间的特征共享,利用辅助任务更好的学习特征.一些研究工作基于语义合成性原则和自然语言的可递归性,提出了一系列基于递归神经网络的文本情绪分析模型. Socher等人基于递归自动编码机(recursive autoencoder)加入了情绪类别相关的信息,并引入重构损失的概念,提出了一种半监督式的递归自动编码机模型.该模型在通过文本所包含单词的词向量生成文本向量时,可以在其中保留更多情绪相关的信息. Socher等人[17]在之后又基于语义合成性原则进一步提出了矩阵-向量递归神经网络(matrix-vector recurrent neural network, MV-RNN). Tang等人[6]提出一种改进的词向量学习方法,将推特短文本中的表情符号作为次要情绪标注,并根据其中的情绪信息进行词向量的学习,在此基础之上,将包含情绪信息的词向量作为词汇的特征,用于构建大规模的情绪词典.

本文采用深度学习的方法解决在线新闻评论的情绪标注问题,提出一种新型神经网络模型,在生成文档的特征表示时不仅考虑到句子中词汇之间的关系,而且能兼顾到文档中句子之间的逻辑关系,从而生成更具代表性的文本分层语义表示,更充分地运用文本中所蕴含的语义信息来对文本进行情绪标注.

另一方面,上述方案只使用了文档自身的信息,忽略了其他相关信息源中的信息,而这些信息源通常对文档语义的表达具有显著的影响.本文在进行新闻评论文本情绪标注时,使用了多种信息源来提升准确度,包括新闻评论文本、新闻文档文本以及用户情绪投票信息.本文扩展了之前的初步研究工作[18],引入了改进的双向长短记忆模型对在线新闻评论进行了多信息源情绪分析工作,捕获文档更加完整的语义特征,从而提高情绪标注的效果.

2　基于多信息源的双向分层语义神经网络

2.1　问题定义

情绪标注问题可以形式化为一个多分类问题,目标是为一段用户撰写的文本分配某一个情绪.本文提出的在线新闻评论情绪标注问题的目标是根据一系列用户在阅读新闻后撰写的评论,以及评论所对应的新闻文档、大量用户对新闻做出的情绪投票,标注在线新闻评论蕴含的情绪.

现对在线新闻评论情绪标注问题定义如下:给定一个新闻评论集合 C 和新闻文档集合 D,每一条新闻评论 $c\in C$ 都对应一个 $d\in D$,表示 c 是在用户阅读新闻 d 之后做出的评论.此外定义情绪集合 $E=\{e_1,e_2,\cdots,e_K\}$,为新闻评论标注的情绪是该集合中的一个元素.每一个新闻文档 d 都对应一个用户情绪投票 $M_d=\{\mu_1,\mu_2,\cdots,\mu_K\}$,其中 $\mu_k\in\mathbb{R}$ 是投票中对于情绪 e_k 的投票计数.在此情况下,情绪标注工作需要做的是根据评论文本 c、评论对应的新闻文档 d 以及 d 对应的用户情绪投票 M_d 的信息,将评论 c 分配至情绪集合中的一个 e_k.

本文提出的基于多信息源的双向分层神经网络模型可分为3个部分:

1) 文档的分层语义神经网络,学习句子中单词间的语义关系和句子间的逻辑关系;

2) 考虑新闻评论内容、新闻内容和用户情绪投票信息的多信息源分层语义神经网络,充分融合多信息源进行特征提取;

3) 适当的分类方法,将抽取的特征分配到某个具体的情绪.

2.2　文档的双向分层神经网络

本节主要介绍了文档的双向分层神经网络,将可变长度的文本表示成定长的特征向量.

单词是句子的基本组成部分,而句子又在结构上和语义上组成了文档.语义合成性原则[19]指出,任意表达式(如一个句子或一篇文档)的语义源自于它每一个组成部分各自的语义以及它们组合在一起

所使用的规则.因此本文提出的方法在计算文档表示时可以分为2步:1)从单词的表示(如词嵌入)生成句子级别的语义表示;2)再从句子的表示最终生成整篇文档的语义表示.

2.2.1 句子语义表示模块

在获取句子的语义表示之前,本文首先使用词向量(word embedding)来表示每一个单词.词向量技术是自然语言处理领域中的一种特征学习技术,它将一个维度为词典中所有单词数量的高维空间嵌入到一个维数低得多的连续向量空间中,词典中的每个单词被表示为实数域上的向量,因此词向量也被称为词嵌入技术,也可以看作一种降维技术.

根据词向量技术,词典中每一个单词被表示为一个低维的、连续的、实数值的向量,所有的向量都存储在矩阵 $\boldsymbol{L}\in\mathbb{R}^{dim\times|V|}$ 中,其中 dim 为预定义的词向量的维度,V 表示文本的词典.词向量的值可以被随机初始化,然后作为一个参数随着深度学习的过程和神经网络一起训练[20-21],或者通过成熟的词向量算法预先求得[22-23].本文采取第2种方式,使用成熟的词嵌入技术求得词向量来更好地利用单词的语义和语法上的联系.

本文引入卷积神经网络来计算句子的语义表示,图1为模型示意图,模型主要包括4层:

1) 输入层.输入层的输入为一个句子中的所有词.根据词向量矩阵,输入层将句子中的词映射为相应的词向量,将整个句子重新组织为该句子中所有词所对应的词向量按照原词出现的顺序排列之后的连接.

2) 卷积层.卷积层使用不同的卷积核对输入进行扫描,抽取得到输入中感受野内的词与词之间局部特征,得到特征矩阵.需要注意的是,每一个卷积核都将产生一个独立的特征矩阵.

3) 池化层.通过对特征矩阵进行平均池化操作得到句子的核心特征,并将不定长的句子输入转化为定长的特征输出.

4) 折叠层.卷积核不同,卷积层所得到的特征矩阵不同,池化层所得到的句子特征也不相同.折叠层将不同卷积核得到的不同句子特征进行归并,生成最终的句子特征.

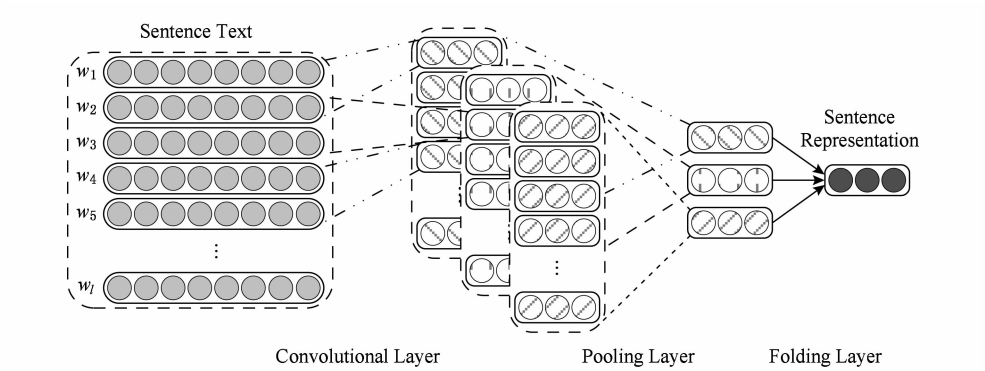

Fig. 1 The overview of our bi-directional hierarchical semantic representation model using multiple information sources

图1 多信息源双向分层语义表示模型示意图

句子语义特征表示模块中的卷积神经网络部分不包含损失层,这是因为句子模块并不是独立训练的模块,而是和文档语义特征表示模块一起训练,因此会在总体模型的最上层拥有统一的损失函数层.

输入层将输入句子中的单词映射为较低维度的词向量,将句子重新组织为词向量的串联.给定一个包含 n 个单词的句子 $\{w_1,w_2,\cdots,w_n\}$,句子中的每一个单词 w_i 都通过经由语料库预先求得的词向量矩阵 $\boldsymbol{L}\in\mathbb{R}^{dim\times|V|}$ 映射为该单词所对应的词向量 $\boldsymbol{we}_i\in\mathbb{R}^{dim}$,其中,$dim$ 为词向量的维度,V 为词典,$|V|$ 即为词典的大小.则卷积神经网络的输入层的输出,即卷积层的输入为

$$\boldsymbol{I}=(\boldsymbol{we}_1,\boldsymbol{we}_2,\cdots,\boldsymbol{we}_n)\in\mathbb{R}^{dim\times n}. \tag{1}$$

本文在卷积层使用不同窗口大小的卷积核对句子向量进行卷积操作,捕捉不同粒度的上下文语义特征,并在此基础上生成句子的语义表示特征.这种方式已经被证明在情绪分析和情绪标注领域具有很好的效果.本文使用窗口为3,4,5的卷积核,用来捕获3,4,5-gram的语义特征.

令卷积核为 cf,l_{cf} 为卷积核 cf 的窗口宽度,$\boldsymbol{W}_{cf}$ 和 $\boldsymbol{b}_{cf}$ 分别为卷积核 cf 线性部分的共享参数.该卷积核作用于感受野内的词向量上,并由此生成所对应的局部特征.感受野的大小即为卷积核的窗口大小.

因此卷积层中,卷积核 cf 在每一个感受野上的输入,为卷积核窗口中从 $\boldsymbol{we}_i$ 开始共 l_{cf} 个单词所对应的词向量的串联:

$$\boldsymbol{I}_{cf}=(\boldsymbol{we}_i,\boldsymbol{we}_{i+1},\cdots,\boldsymbol{we}_{i+l_{cf}-1})\in\mathbb{R}^{dim\times l_{cf}},\quad(2)$$

则卷积层中卷积核 cf 对于该感受野内的输入,求得局部特征:

$$\boldsymbol{O}_{cf}=\tanh(\boldsymbol{W}_{cf}\cdot\boldsymbol{I}_{cf}+\boldsymbol{b}_{cf}),\quad(3)$$

$$\boldsymbol{W}_{cf}\in\mathbb{R}^{locf\times(dim\times l_{cf})},\quad(4)$$

其中,$\boldsymbol{W}_{cf}$ 为卷积操作的权重矩阵;$\boldsymbol{b}_{cf}\in\mathbb{R}^{locf}$,为卷积操作的偏置量;$locf$ 是卷积层输出的长度;tanh 是分线性的双曲正切函数,为卷积操作增加非线性特性.

每一个卷积核对窗口大小为 l_{cf} 的词向量序列进行卷积,得到该部分的局部特征值,顺序求得每一部分的局部特征后得到整个句子的特征矩阵:

$$\boldsymbol{O}_{cf}=(\boldsymbol{O}_{cf,0},\boldsymbol{O}_{cf,1},\cdots,\boldsymbol{O}_{cf,n-l_{cf}}),\quad(5)$$

其中,n 为句子中所包含的词向量的总数.

卷积层之后,池化层将对卷积操作所得到的特征矩阵进行池化操作.本文中使用平均池化操作,在特征矩阵上取平均值,该平均值即为该卷积核所得到的句子语义特征表示:

$$\hat{\boldsymbol{O}}_{cf}=avg(\boldsymbol{O}_{cf,i}),\quad(6)$$

其中,$i=0,1,2,\cdots,n-l_{cf}$.此外,池化操作也可以保证卷积神经网络对于不定长的输入,也能产生特定长度的句子特征.

在此之后,本文通过一个平均折叠层将所有的池化层输出进行归并,得到最终的句子语义表示特征:

$$\hat{\boldsymbol{O}}=avg(\hat{\boldsymbol{O}}_{cf1},\hat{\boldsymbol{O}}_{cf2},\hat{\boldsymbol{O}}_{cf3}),\quad(7)$$

其中,$cf1,cf2,cf3$ 分别为上文所述窗口大小为 3,4,5 的卷积核.

2.2.2　文档语义表示模块

本节介绍利用改进的双向长短记忆模型(bi-directional long-short term memory model, Bi-LSTM)从句子的语义表示进一步生成文档的语义表示的过程.

给定一组句子的语义表示并由此生成文档语义表示,最简单和自然的策略是对所有的句子表示取平均值或最值.但这种方法完全没有考虑到句子之间的语序以及逻辑联系,显然不能捕捉句子间的因果或转折等复杂关系.神经网络方法常使用循环神经网络(recurrent neural network, RNN)解决这一问题,循环神经网络善于处理序列类型的数据,不同于全连接的前馈神经网络,循环神经网络通过循环递归的结构使得自身具有一定程度的记忆功能,能够有效地利用序列自身所携带的信息进行分析预测.然而,循环神经网络由于天然的结构缺陷,导致实践中无法学习输入数据中距离较远的逻辑关系.长短期记忆模型将循环神经网络的循环结构替换成长短期记忆单元,解决远距离逻辑关系的学习问题[24].

本文使用一种经过改进的双向长短记忆模型(Bi-LSTM)来获得文档的语义表示,Bi-LSTM 转化函数如下.

1) 向前推算(forward pass):

$$\boldsymbol{f}_{t,fwd}=\delta(\boldsymbol{W}_{f,fwd}\cdot(\boldsymbol{h}_{t-1,fwd},\boldsymbol{x}_t)+\boldsymbol{b}_{f,fwd}),\quad(8)$$

$$\boldsymbol{i}_{t,fwd}=\delta(\boldsymbol{W}_{i,fwd}\cdot(\boldsymbol{h}_{t-1,fwd},\boldsymbol{x}_t)+\boldsymbol{b}_{i,fwd}),\quad(9)$$

$$\boldsymbol{o}_{t,fwd}=\delta(\boldsymbol{W}_{o,fwd}\cdot(\boldsymbol{h}_{t-1,fwd},\boldsymbol{x}_t)+\boldsymbol{b}_{o,fwd}),\quad(10)$$

$$\boldsymbol{C}'_{t,fwd}=\tanh(\boldsymbol{W}_{C,fwd}\cdot(\boldsymbol{h}_{t-1,fwd},\boldsymbol{x}_t)+\boldsymbol{b}_{C,fwd}),\quad(11)$$

$$\boldsymbol{C}_{t,fwd}=\boldsymbol{f}_{t,fwd}\odot\boldsymbol{C}_{t-1,fwd}+\boldsymbol{i}_{t,fwd}\odot\boldsymbol{C}'_{t,fwd},\quad(12)$$

$$\boldsymbol{h}_{t,fwd}=\boldsymbol{o}_{t,fwd}\odot\boldsymbol{C}_{t,fwd},\quad(13)$$

其中,$\boldsymbol{x}_t$ 为 Bi-LSTM 模型在第 t 步的输入向量,本文中为文档中第 t 个句子的句子语义表示;$\boldsymbol{f}_{t,fwd}$,$\boldsymbol{i}_{t,fwd}$,$\boldsymbol{o}_{t,fwd}$,$\boldsymbol{W}_{f,fwd}$,$\boldsymbol{W}_{i,fwd}$,$\boldsymbol{W}_{o,fwd}$,$\boldsymbol{b}_{f,fwd}$,$\boldsymbol{b}_{i,fwd}$,$\boldsymbol{b}_{o,fwd}$ 分别负责隐含层向量(hidden vector)和输入向量的遗忘和更新操作;$\boldsymbol{W}_{C,fwd}$ 和 $\boldsymbol{b}_{C,fwd}$ 用于生成候选向量 $\boldsymbol{C}'_t$;$\boldsymbol{h}_{t-1,fwd}$ 为 Bi-LSTM 的隐含层向量,用来表示历史信息,保存前 $t-1$ 步所积累的知识;$\boldsymbol{C}_{t-1,fwd}$ 和 $\boldsymbol{C}'_{t,fwd}$ 分别为第 t 步时神经元的原始状态向量和候选向量;$\odot$ 为 2 个向量按元素相乘运算.

2) 向后推算(backward pass):

$$\boldsymbol{f}_{t,bwd}=\delta(\boldsymbol{W}_{f,bwd}\cdot(\boldsymbol{h}_{t+1,bwd},\boldsymbol{x}_t)+\boldsymbol{b}_{f,bwd}),\quad(14)$$

$$\boldsymbol{i}_{t,bwd}=\delta(\boldsymbol{W}_{i,bwd}\cdot(\boldsymbol{h}_{t+1,bwd},\boldsymbol{x}_t)+\boldsymbol{b}_{i,bwd}),\quad(15)$$

$$\boldsymbol{o}_{t,bwd}=\delta(\boldsymbol{W}_{o,bwd}\cdot(\boldsymbol{h}_{t+1,bwd},\boldsymbol{x}_t)+\boldsymbol{b}_{o,bwd}),\quad(16)$$

$$\boldsymbol{C}'_{t,bwd}=\tanh(\boldsymbol{W}_{C,bwd}\cdot(\boldsymbol{h}_{t+1,bwd},\boldsymbol{x}_t)+\boldsymbol{b}_{C,bwd}),\quad(17)$$

$$\boldsymbol{C}_{t,bwd}=\boldsymbol{f}_{t,bwd}\odot\boldsymbol{C}_{t+1,bwd}+\boldsymbol{i}_{t,bwd}\odot\boldsymbol{C}'_{t,bwd},\quad(18)$$

$$\boldsymbol{h}_{t,bwd}=\boldsymbol{o}_{t,bwd}\odot\boldsymbol{C}_{t,bwd}.\quad(19)$$

向后推算部分公式与向前推算部分类似,具体符号意义这里不再赘述.需要注意的是,向前推算从段落中第 1 句扫描至最后 1 句,共设有 n 句话,则 t 从 1 遍历至 n;而向后推算部分是从最后一句开始,即 t 从 n 遍历至 1.

在经典 Bi-LSTM[25] 中,最后 1 层隐含层的输出通常作为最终的文本表示,如图 2 所示.本文对 Bi-LSTM 进行了简单的扩展,提出了 Avg Bi-LSTM 模型,该模型将所有隐含层的输出取平均值作为最终的文本语义表示,如图 3 所示.这样,最终获得带有不同间隔的句子间的语义和情绪的逻辑关系的文本特征表示.

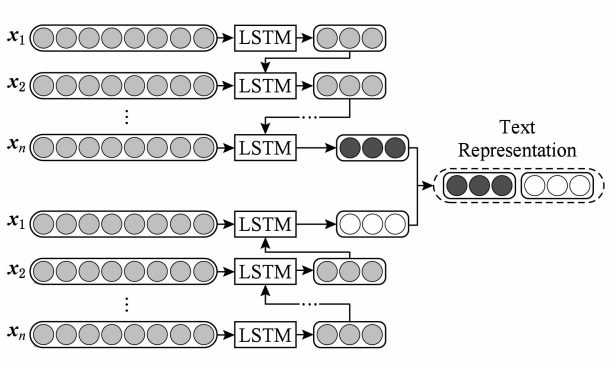

Fig. 2 Classical Bi-LSTM

图 2 经典 Bi-LSTM 模型示意图

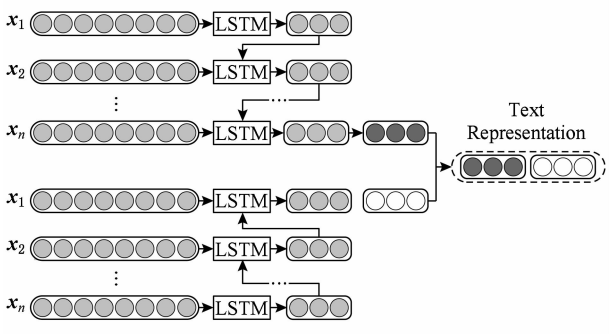

Fig. 3 Avg Bi-LSTM

图 3 Avg Bi-LSTM 模型示意图

2.3 多信息源的双向分层神经网络

本节介绍多信息源在线新闻评论情绪标注方法,需要考虑的信息源包括用户评论,在线新闻,读者情绪投票.图 4 为基于多信息源的双向分层神经网络的示意图,可以看出,在本模型中,新闻评论的多源分层语义表示特征由 3 部分组成:新闻评论文本的分层语义表示特征、新闻文档文本的分层语义表示特征以及用户情绪投票特征.

前 2 种信息源为用户评论和新闻文档的分层语义表示.用户评论是用户对于新闻文档的反应,能够直接体现用户的情绪,另一方面,评论的情绪很显然会直接受到新闻文档的影响,因此本文将新闻文档纳入考虑的范围内.为了对用户评论和新闻文档进行建模,本文利用 2.2 节的分层语义表示模型获得每一条评论 c 和新闻文档 d 的特征特征向量表示 $\boldsymbol{c}'\in\mathbb{R}^{dimc}$ 和 $\boldsymbol{d}'\in\mathbb{R}^{dimd}$,其中 $dimc$ 和 $dimd$ 分别为评论和新闻文档的特征向量的维度.

最后一种信息源通过新闻文档的用户情绪投票得到.当为 1 条评论标注情绪时,可以根据归一化处理的用户情绪投票来判断该评论可能属于哪一个情绪类别.根据用户情绪投票信息,新闻文档 d 的某条评论 c 被标注为情绪类别 e_i 的可能性,用 μ_i' 表示.通过归一化处理,将投票向量 $\boldsymbol{M}_d$ 重新表示为

$$\boldsymbol{M}_d' = (\mu_1', \mu_2', \cdots, \mu_K'),$$

$$\mu_i' = \frac{\mu_i}{\sum_{j=1}^{K}\mu_j}. \tag{20}$$

在此基础上,单条评论 c 的最终情绪表示的特征向量为 $\boldsymbol{rep}_c=(\boldsymbol{c}', \boldsymbol{d}', \boldsymbol{M}_d')$.

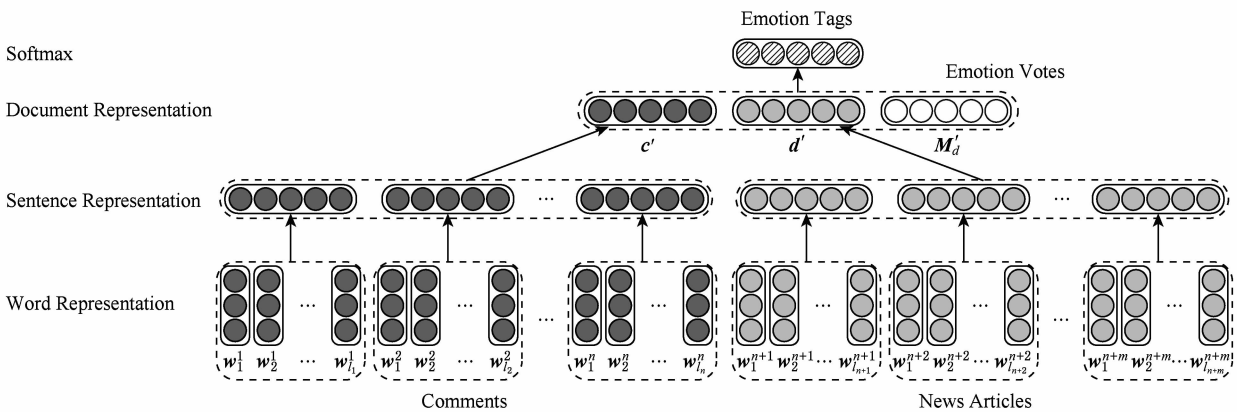

Fig. 4 The overview of our bi-directional hierarchical semantic neural network using multiple information sources

图 4 多信息源双向分层神经网络示意图

2.4 情绪分类方法

本节使用 2.3 节中得到的新闻评论的多源分层语义表示特征作为分类的特征向量,代替传统机器学习中人工抽取的特征进行情绪分类.

如图 4 所示,多信息源的在线新闻评论情绪标注工作最终抽取的特征是在线新闻、在线新闻评论和用户情绪投票特征向量的串联,在情绪分析过程中,本文引入 Softmax 分类器,将新闻评论的特征表示转换为条件概率来表示评论被分到每一个情绪类别的概率.

对于给定语料库中第 i 个评论 c_i,c_i 的情绪标注结果为情绪类别集合 e_k($k=1,2,\cdots,K$)中一种

情绪的条件概率可通过函数 $Softmax$ 计算得出:

$$P(e_k|c_i)=P(e_k|\boldsymbol{rep}_i)=\frac{\exp(\boldsymbol{\omega}_k^{\mathrm{T}}\boldsymbol{rep}_i)}{\sum_{e_j\in E}\exp(\boldsymbol{\omega}_j^{\mathrm{T}}\boldsymbol{rep}_i)},\quad(21)$$

其中,c_i 为第 i 个评论,$\boldsymbol{rep}_i$ 为评论 c_i 的多信息源分层语义表示特征,$\boldsymbol{\omega}$ 为将表示 $\boldsymbol{rep}_i$ 转换为 $|E|$ 维实数向量的矩阵,E 为情绪集合,$\boldsymbol{\omega}_j$ 可以看作 $\boldsymbol{rep}_i$ 中每一项对于情绪 e_j 的结合系数.

由于训练集中每一条评论都有与之对应的真实情绪标签,本文通过有监督的方式对模型进行训练.本文将真实的情绪分布与预测得到的情绪分布之间的交叉熵误差作为目标损失函数:

$$J(\theta)=-\sum_{c\in C}\sum_{e\in E}P^g(e|c)\times\log_a(P(e|c)),\quad(22)$$

其中,$c\in C$ 为语料库中的一条评论,e 为情绪集合 E 中的一种情绪,a 可以是任何一个合法的对数函数的底数,$P(e|c)$ 表示模型预测的评论 c 属于情绪 e 的概率,$P^g(e|c)$ 表示真实情绪分布中评论 c 属于情绪 e 的概率分布.需要注意的是,在真实的情绪分布中,只有评论真正属于的情绪类别的概率是 1,其余全被置为 0.

之后将交叉熵误差作为目标损失函数,通过反向传播算法训练,由随机梯度下降法更新模型的参数.

本文使用 dropout[10,26] 方法进行参数的正则化,用以消除模型过拟合的影响.dropout 方法的核心思想是在训练阶段(training)随机地移除神经网络中的部分节点,而在测试阶段(testing)使用全部节点.dropout 引入一个超参数 p,表示神经网络中每一个节点在每一次迭代中都会以 $1-p$ 的概率被移除,或以 p 的概率被保留.因此在每次迭代中,神经网络只有被保留的部分节点所组成的"子网络"会被训练,也就只有这些保留的节点中的参数会被更新,而其他节点的参数依然保留上次迭代的结果.这样不仅能加速训练过程,而且能大大解决过拟合问题所带来的影响.

具体来说,对于模型中的 CNN 模块,在对输入向量 $\boldsymbol{I}_{cf}$ 数据进行卷积操作之前,先将其与一个 dropout 向量相乘得到新的输入向量 $\boldsymbol{I}'_{cf}$:

$$\boldsymbol{I}'_{cf}=\boldsymbol{I}_{cf}\odot\boldsymbol{m},\quad(23)$$

$$m(i)\sim Bernoulli(p),\quad(24)$$

其中,$\boldsymbol{m}$ 为 dropout 向量,与 $\boldsymbol{I}_{cf}$ 具有相同的维度,$m(i)$ 表示 $\boldsymbol{m}$ 中的第 i 个元素.$\boldsymbol{m}$ 对于不同的 $\boldsymbol{I}_{cf}$ 都要重新计算,以保证每一次移除操作都是完全独立的.

对于模型中的 Bi-LSTM 模块的隐含层向量也要进行相似的 dropout 处理,这里不再赘述.

3　实验与分析

本节首先介绍相关的实验设置,包括实验所使用的数据集、评价指标以及用于比较的基准算法,之后给出实验结果并进行分析和讨论.

3.1　数据集

新浪新闻(Sina news)和腾讯新闻(QQ news)是中国用户最多的新闻网站之一.从新浪新闻的社会频道(society channel)① 和腾讯新闻的娱乐频道(entertainment channel)② 中收集了 2011 年 1 月到 6 月点击量较多的热点新闻,以及它们的热点评论和用户情绪投票信息用于实验.在数据集中随机采样出部分新闻以及它们排名前 20 的热点评论(如果评论数量不足 20,则选择全部评论)以及用户情绪投票作为训练集和测试集,命名为新浪数据集(Sina dataset)和腾讯数据集(QQ dataset).新浪数据集中共有 5 185 条评论、369 条新闻、83 634 个投票,腾讯数据集中共有 5 414 条评论、372 条新闻、993 089 个投票.每一条评论都与它所对应的新闻相对应.相关统计信息如表 1 所示:

Table 1　The Statistics of Labeled Comments of Datasets

表 1　数据集评论标注统计信息

Statistical Items	Sina Dataset	QQ Dataset
Number of Comments	5 185	5 414
Number of News Articles	369	372
Number of Emotion Votes	83 634	993 089

为了评估预测效果,对 2 个数据集情绪分类进行人工标注.在新浪新闻和腾讯新闻中,虽然用户可以使用网站自带的投票系统表达情绪,但是投票系统和评论系统是独立的,投票不能与评论一一对应,也就不能将用户的投票作为评论所表达情绪的标注向量.所以本文只借用了新浪新闻和腾讯新闻投票系统中的情绪类别,并人工对每一条的情绪进行标注.由于工作量的原因,每个数据集只分配了 3 个标注人员.新浪数据集和腾讯数据集的 8 类情绪的标注结果如表 2 和表 3 所示:

① http://news.sina.com.cn/society/

② http://ent.qq.com/

Table 2 The Statistics of Labeled Comments of Sina Dataset

表 2 新浪数据集各类情绪标注评论统计信息

Emotion	Number of Comments	Rate/%
Touched	905	17.45
Sympathetic	614	11.84
Bored	336	6.48
Angry	1752	33.79
Amused	408	7.87
Sad	654	12.61
Surprised	196	3.78
Fervent	320	6.17

Table 3 The Statistics of Labeled Comments of QQ Dataset

表 3 腾讯数据集各类情绪标注评论统计信息

Emotion	Number of Comments	Rate/%
Touched	1619	29.90
Sympathetic	139	2.57
Bored	641	11.84
Angry	1639	30.27
Amused	563	10.40
Sad	355	6.56
Surprised	85	1.57
Fervent	373	6.89

为了验证标注的质量,从 2 个数据集中各自随机抽取了 100 条评论,并请 1 名审阅人(不是标注者)对评论进行标注. 新浪数据集中,审阅人的判断与标注结果有 91 条相同,腾讯数据集中有 94 条相同,因此本文认为标注的质量满足实验要求.

本文只使用了中文的数据集,因为目前没有找到类似的带有用户情绪投票信息的英文在线新闻服务,但本文提出的模型可适用于任何语言的新闻服务.

3.2 评价指标

本文使用 2 种指标用于评价算法的效果:

1) 平均倒数排名(mean reciprocal rank, *MRR*)

平均倒数排名表示评论真实情绪标签在预测的经过排序的预测情绪分布中的位置. 给定一个评论 $c\in C$ 以及它对应的真实的情绪标注 e'_c、算法预测的评论 c 的情绪排序列表 L_c,令 $rank_{L_c}(e'_c)$ 表示 e'_c 在 L_c 中的位置,则 *MRR* 可被定义为

$$MRR=\frac{1}{|C|}\sum_{c\in C}\frac{1}{rank_{L_c}(e'_c)}. \tag{25}$$

2) 准确率 $accu@m$

给定一个评论 $c\in C$ 以及它对应的真实情绪标注 e'_c、算法预测的评论 c 的情绪排序列表 $L_c@m$,该列表只包含 L_c 的 top-m 项,则 $accu_c@m$ 可被定义为

$$accu_c@m=\begin{cases}1, & e'_c\in L_c@m,\\ 0, & e'_c\notin L_c@m,\end{cases} \tag{26}$$

则对于整个数据集的 $accu@m$ 可被定义为

$$accu@m=\frac{\sum_{c\in C}accu_c@m}{|C|}. \tag{27}$$

3.3 对比方法

本文采用 10 折交叉验证的方式对提出的 Bi-HSNN 算法和下列方法在 2 个数据集上进行了实验,并将结果进行了比较和分析.

1) SVM+n-grams. 将评论文本的切分成 n-grams,并将其作为评论文本的特征,然后使用 LIBLINEAR[27] 训练 SVM 分类器. 实验中 n-grams 中 n 的取值分别为 1,2,3.

2) WE(word-emotion)[13]. WE 是一种基于情绪词典的生成模型,它首先建立词级别和话题级别的情绪词典,然后据此预测评论的情绪分类.

3) RPWM (reader perspective weighted model)[28]. 在 RPWM 中,单独计算每条评论的情绪熵,并据此为评论分配不同的权重. 权重不同,评论对情绪预测结果的影响不同.

4) 标准 CNN[8] 与 Bi-LSTM[25]. 作为目前情绪和语义分析领域最前沿的算法,CNN 和 Bi-LSTM 在本文中也被作为基准算法进行效果比较. 特别地,在对比实验中,标准 CNN 和 Bi-HSNN 的 CNN 模块一样,使用了窗口大小分别为 3,4,5 的卷积核.

5) CM(content-based model)与 MCM(meta classification model)[11]. 这 2 种模型建立有监督的固定组合分类模型,并使用传统机器学习方法对评论的情绪进行标注.

6) Bi-HSNN 为本文提出的算法.

3.4 实验效果对比

本节进行了 3 组对比实验,分别探究了本论文提出的算法在单信息源和多信息源情况下与 SVM+n-grams,WE,RPWM 等基准算法的性能差异及原因,并探究了多信息源的情绪标注是否能有效提高情绪标注的效果.

3.4.1 单信息源算法效果对比

本节实验用来比较只使用用户评论一种信息源时 Bi-HSNN 的效果. 实验结果如表 4 和表 5 所示.

通过结果可以看出,基于 SVM 的方法即使几乎没有使用任何语义信息,仍然拥有很好的表现,几乎是除了 Bi-HSNN 之外效果最好的方法. 但随着 n

的增大,n-grams 特征变得越来越稀疏,对于新闻评论这种包含文字较少的短文本,数据稀疏问题更加明显.这也是用 trigrams 特征的 SVM 分类器在 3 个 SVM 分类器中效果最差的原因.本文尝试只考虑情绪词的方式来对特征进行降维,但结果并没有显著提高.

Table 4　Performance on Sina Dataset

表 4　新浪数据集上各方法的评论情绪标注性能

Methods	*MRR*	*accu*@1	*accu*@2	*accu*@3
SVM+unigrams	0.629 8	0.445 5	0.630 8	0.761 5
SVM+bigrams	0.590 1	0.405 7	0.573 4	0.699 0
SVM+trigrams	0.549 7	0.352 8	0.523 7	0.662 7
WE	0.5687	0.365 0	0.558 7	0.705 2
RPWM	0.534 7	0.3356	0.497 3	0.651 2
Standard CNN	0.616 6	0.422 5	0.666 8	0.764 2
Standard Bi-LSTM	0.643 9	0.458 7	0.692 3	0.801 2
CM	0.623 1	0.436 7	0.632 9	0.751 4
Bi-HSNN	0.689 5	0.550 6	0.769 1	0.832 2

Table 5　Performance on QQ Dataset

表 5　腾讯数据集上各方法的评论情绪标注性能

Methods	*MRR*	*accu*@1	*accu*@2	*accu*@3
SVM+unigrams	0.615 3	0.425 6	0.613 0	0.754 9
SVM+bigrams	0.602 8	0.408 1	0.609 4	0.733 0
SVM+trigrams	0.547 7	0.308 4	0.591 3	0.711 8
WE	0.534 0	0.336 5	0.507 7	0.639 5
RPWM	0.543 8	0.363 8	0.515 6	0.620 6
Standard CNN	0.632 6	0.4400	0.617 2	0.779 7
Standard Bi-LSTM	0.669 7	0.448 7	0.6362	0.810 5
CM	0.622 5	0.447 2	0.620 2	0.730 6
Bi-HSNN	0.736 9	0.519 0	0.741 8	0.871 9

WE 通过建立情绪词典的方式对评论的情绪进行预测并取得了一定的效果,但是它也只是将评论文档看作词袋模型,并没有进一步考虑语义信息.RPWM 是 WE 的一种改进,一方面通过 LDA 模型将情绪和话题结合起来;另一方面为每一条评论计算情绪熵,并将其作为权重赋予评论,评论根据其权重不同对预测结果的影响也不同,用这种方式来减小噪声评论对于预测的影响.然而实验结果表明:RPWM 的结果并没有很明显的提升,原因在于本文所使用的数据集中,评论与评论之间并没有很大的差异.

CM 使用评论文本中的情绪词作为特征,并使用 L2 逻辑斯蒂回归模型进行情绪标注.因为 CM 模型只考虑了文本中更有可能表现出情绪的词汇,因此它比上述算法拥有更好的效果.此外通过 CM 和 WE,RPWM 的对比可以看出,在情绪标注中,判别模型确实比生成模型拥有更好的效果.

标准 CNN 和 Bi-LSTM 的性能要大大优于上述大部分算法,因为它们考虑了文本的语义.实验结果说明:语义合成性原则确实对于语义和情绪的理解具有重要的意义.然而这 2 种方法在对句子间逻辑关系建模等方面仍有很大的改进空间.

本节所使用的 Bi-HSNN 为只使用评论文本情况下的版本.实验结果显示:Bi-HSNN 性能优于上述所有对比算法,因为它不仅通过对句子内部的语义进行建模,而且对句子间的逻辑关系进行了建模.这使得 Bi-HSNN 拥有理解文档中复杂语义的能力.此外通过 Bi-HSNN 与 CNN/Bi-LSTM 的对比可以看出,句子间的逻辑关系确实能够帮助对整个文档的语义和情绪进行理解.

显著性检验表明:Bi-HSNN 在置信级别为 0.95 的情况下性能优于其他所有基准算法.

3.4.2　多信息源算法效果对比

本节实验用来比较使用所有 3 种信息源的情况下,Bi-HSNN 与 SVM+n-grams,WE,RPWM 等基准算法对比的效果.实验结果如表 6 和表 7 所示:

Table 6　Performance on Sina Dataset

表 6　新浪数据集上各方法的评论情绪标注性能

Methods	*MRR*	*accu*@1	*accu*@2	*accu*@3
SVM+unigrams	0.684 7	0.506 7	0.686 6	0.817 1
SVM+bigrams	0.646 2	0.467 8	0.633 4	0.758 3
SVM+trigrams	0.609 2	0.415 4	0.591 7	0.732 6
Standard CNN	0.654 4	0.490 9	0.671 7	0.775 2
Standard Bi-LSTM	0.698 4	0.537 8	0.728 9	0.821 5
MCM	0.6493	0.481 4	0.645 6	0.758 5
Bi-HSNN	0.752 3	0.601 8	0.825 1	0.923 4

Table 7　Performance on QQ Dataset

表 7　腾讯数据集上各方法的评论情绪标注性能

Methods	*MRR*	*accu*@1	*accu*@2	*accu*@3
SVM+unigrams	0.671 5	0.486 2	0.674 9	0.817 7
SVM+bigrams	0.677 4	0.474 8	0.673 7	0.818 6
SVM+trigrams	0.596 7	0.363 4	0.647 8	0.765 9
Standard CNN	0.671 9	0.498 4	0.710 1	0.811 9
Standard Bi-LSTM	0.729 6	0.541 2	0.728 8	0.850 1
MCM	0.677 7	0.510 4	0.693 9	0.800 2
Bi-HSNN	0.770 8	0.574 1	0.748 4	0.939

SVM,CNN,Bi-LSTM 根据 Bi-HSNN 的思想修改为多信息源模型,即将 3 种信息源的特征串联输入到分类器中. MCM 本身就是 CM 的多信息源版本(但只使用了 2 种信息源),因此本文只对它进行了很小的改动. WE 和 RPWM 为生成模型,因此无法被改造成多信息源版本,因此未被考虑.

Bi-HSNN 为使用全部 3 种信息源时的版本.

实验结果表明:在使用全部 3 种信息源的情况下,Bi-HSNN 与所有基准算法相比,仍然拥有最佳的效果. 这再次印证了 Bi-HSNN 的多层次结构赋予它从句子内语义和句间语义 2 个角度对文本进行建模的能力. 与其他方法相比,Bi-HSNN 展现出其极强的语义表示与情绪挖掘能力.

显著性检验表明:Bi-HSNN 在置信级别为 0.95 的情况下,性能优于其他所有基准算法.

3.4.3 多信息源效果分析

本节实验用于:1)证明是否每一种信息源对于用户评论的情绪标注都有帮助;2)评价使用不同信息源的 Bi-HSNN 的效果差异,实验结果如表 8 和表 9 所示:

Table 8 Performance on Sina Dataset

表 8 新浪数据集上不同信息源的评论情绪标注性能

Methods	*MRR*	*accu*@1	*accu*@2	*accu*@3
CC	0.689 5	0.550 6	0.769 1	0.832 2
CN	0.637 7	0.408 7	0.495 5	0.699 7
UEV	0.633 8	0.403 2	0.555 3	0.716 6
CC+CN	0.714 9	0.524 9	0.754 0	0.863 7
CC+UEV	0.766 6	0.596 4	0.808 0	0.907 9
CN+UEV	0.709 1	0.532 8	0.749 2	0.871 1
CC+CN+UEV	0.752 3	0.601 8	0.825 1	0.923 4

Table 9 Performance on QQ Dataset

表 9 腾讯数据集上不同信息源的评论情绪标注性能

Methods	*MRR*	*accu*@1	*accu*@2	*accu*@3
CC	0.736 9	0.519	0.741 8	0.871 9
CN	0.612 8	0.364 9	0.457 3	0.696 6
UEV	0.613 4	0.373 7	0.528 6	0.712 5
CC+CN	0.730 1	0.503 6	0.737 8	0.893 1
CC+UEV	0.763 8	0.543 6	0.747 1	0.902 6
CN+UEV	0.688 1	0.500 3	0.696 5	0.849 4
CC+CN+UEV	0.770 8	0.574 1	0.748 4	0.939 0

CC,CN,UEV 表示 Bi-HSNN 分别只使用评论文本、新闻文本和用户情绪投票单信息源进行情绪标注. CC+CN,CC+UEV,CN+UEV 使用 2 种信息源的组合作为特征;CC+CN+UEV 使用全部 3 种信息源.

实验结果显示:CC 在单信息源版本中拥有最优的效果,表明评论文本对于评论情绪的预测是最可靠最有效的信息源,这是因为评论就是本文所要进行情绪标注的客体. UEV 的效果是第二优的,证明用户情绪投票可以为评论的情绪标注提供更多的信息,本文认为这是因为用户的投票比新闻文档本身能更直观地反映大众读者在阅读之后的情绪倾向.

通过实验结果也能看出,多信息源版本的 Bi-HSNN 性能普遍优于单信息源版本,说明不同信息源的组合确实比单一信息源更有效,而且每一种信息源都或多或少地对于评论的语义和情绪的理解有一定帮助.

最后 CC+CN+UEV 拥有最优的效果,这证明使用全部 3 种信息源确实比只使用一种或 2 种信息源更有效,每一种信息源都会为情绪标注提供一个不同的视角,分别在不同程度上帮助模型得到一个更好的预测结果.

显著性检验表明:CC+CN+UEV 在置信区间为 0.95 的情况下,性能优于其他所有算法.

3.4.4 dropout 效果分析

本文使用 dropout 方法进行参数的正则化,用以消除模型过拟合的影响. 本节实验用于探索 dropout 比率大小对情绪标注效果的影响.

由于在现实使用中 dropout 通常取值为 $p=0.5$[26],因此本文将 $p=0.5$ 当作基准,其他 dropout 率的效果通过 *accu*@1 相比 $p=0.5$ 时的变化来衡量. 实验表明,*MRR* 与 *accu*@2,*accu*@3 的变化趋势和曲线与 *accu*@1 相同,因此实验结果只展示 *accu*@1 的变化. 图 5 和图 6 中分别展示了本文单信息源和多信息源部分的 dropout 效果.

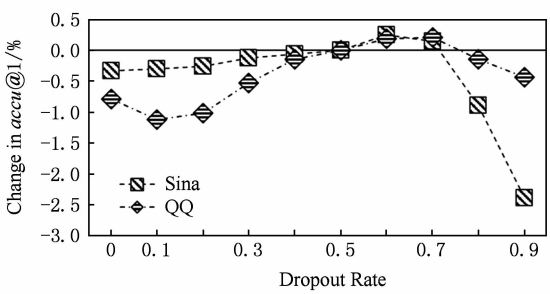

Fig. 5 Effect of dropout rate on single information source

图 5 单信息源不同 dropout 率的效果比较

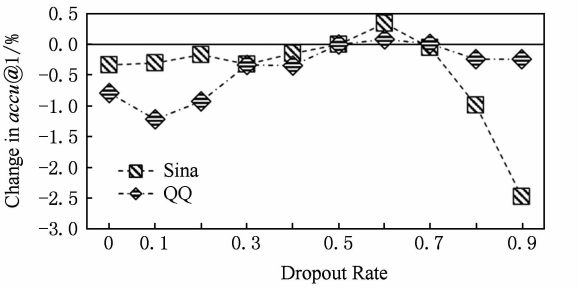

Fig. 6　Effect of dropout rate on multiple information sources

图 6　多信息源不同 dropout 率的效果比较

由图 5 和图 6 可以看出,根据数据集的不同,dropout 率取 0.1～0.7 时都会对情绪标注的效果有一定的提升[10].在本文的实验中,dropout 率选取为 $p=0.6$.

4　总结与未来工作

本文提出在线新闻评论情绪标注问题,考虑在线新闻、新闻评论和用户投票信息,构建了基于多信息源的双向分层神经网络.建立能够对句内语义关系和句间逻辑关系建模的文档表示模型,并提出了多信息源的情绪标注方法,实验证明:本文提出的方法能够有效提高在线新闻评论情绪标注的准确度.

未来可以对更多的信息源进行分析,如新闻中的图片同样属于对读者情绪有重要影响的信息,或者对每一名用户的阅读习惯和情绪倾向进行建模,以提升预测的准确度.

参　考　文　献

[1] Domingos P. A few useful things to know about machine learning [J]. Communications of the ACM, 2012, 55(10): 78-87

[2] Hearst M A. Direction-based text interpretation as an information access refinement [G] //Text-based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval. Mahwah, NJ: Lawrence Erlbaum Associates, 1992: 257-274

[3] Zhao Yanyan, Qin Bing, Liu Ting. Sentiment analysis on text [J]. Journal of Software, 2010, 21(8): 1834-1848 (in Chinese)
(赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848)

[4] Liu Bing. Opinion mining and sentiment analysis [J]. Synthesis Lectures on Human Language Technologies, 2011, 2(2): 459-526

[5] Tang Duyu, Qin Bing, Liu Ting. Document modeling with gated recurrent neural network for sentiment classification [C] //Proc of the 2015 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1422-1432

[6] Tang Duyu, Qin Bing, Liu Ting. Learning semantic representations of users and products for document level sentiment classification [C] //Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1014-1023

[7] Turney P D. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews [C] //Proc of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA: ACL, 2002: 417-424

[8] Bao Shenghua, Xu Shengliang, Zhang Li, et al. Joint emotion-topic modeling for social affective text mining [C] //Proc of the 9th IEEE Int Conf on Data Mining. Piscataway, NJ: IEEE, 2009: 699-704

[9] Bao Shenghua, Xu Shengliang, Zhang Li, et al. Mining social emotions from affective text [J]. IEEE Trans on Knowledge & Data Engineering, 2012, 24(9): 1658-1670

[10] Wu Haibing, Gu Xiaodong. Towards dropout training for convolutional neural networks [J]. Neural Networks, 2015, 71(11): 1-10

[11] Zhang Ying, Fang Yi, Quan Xiaojun, et al. Emotion tagging for comments of online news by meta classification with heterogeneous information sources [C] //Proc of the 35th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2012: 1059-1060

[12] Zhang Ying, Zhang Ning, Si Luo, et al. Cross-domain and cross-category emotion tagging for comments of online news [C] //Proc of the 37th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2014: 627-636

[13] Rao Yanghui, Lei Jingsheng, Liu Wenyin, et al. Building emotional dictionary for sentiment analysis of online news [J]. World Wide Web, 2014, 17(4): 723-742

[14] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507

[15] Ranzato M A, Szummer M. Semi-supervised learning of compact document representations with deep networks [C] //Proc of the 25th Int Conf on Machine Learning. New York: ACM, 2008: 792-799

[16] Weston J, Ratle F, Collobert R. Deep learning via semi-supervised embedding [C] //Proc of the 25th Int Conf on Machine Learning. New York: ACM, 2012: 1168-1175

[17] Socher R, Pennington J, Huang E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions [C] //Proc of the 2011 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2011: 151-161

[18] Wang Chao, Zhang Ying, Jie Wei, et al. Hierarchical semantic representations of online news comments for emotion tagging using multiple information sources [C] // Proc of the 22nd Int Conf on Database Systems for Advanced Applications. Berlin: Springer, 2017: 121-136

[19] Frege G. Sense and reference [J]. Philosophical Review, 1948, 57(3): 238-241

[20] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences [OL]. (2014-08-08) [2016-01-23]. https://arxiv.org/abs/1404.2188

[21] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C] //Proc of the 2013 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2013: 1631-1642

[22] Mikolov T, Sutskever I, Chen Kai, et al. Distributed representations of words and phrases and their compositionality [C] //Proc of the 2013 Conf on Neural Information Processing Systems. Lake Tahoe: NIPS, 2013: 3111-3119

[23] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation [C] //Proc of the 2014 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1532-1543

[24] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780

[25] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural Networks, 2005, 18(5): 602-610

[26] Pham V, Bluche T, Kermorvant C, et al. Dropout improves recurrent neural networks for handwriting recognition [C] // Proc of the 14th Int Conf on Frontiers in Handwriting Recognition (ICFHR). Piscataway, NJ: IEEE, 2014: 285-290

[27] Fan Rongen, Chang Kaiwei, Hsieh C J, et al. LIBLINEAR: A library for large linear classification [J]. Journal of Machine Learning Research, 2008, 9(9): 1871-1874

[28] Li Xin, Rao Yanghui, Chen Yanjia, et al. Social emotion classification via reader perspective weighted model [C] // Proc of the 30th AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2016: 4230-4231

Zhang Ying, born in 1986. PhD, associate professor. Her main research interests include sentiment analysis, data mining, and information retrieval.

Wang Chao, born in 1989. PhD. His main research interests include sentiment analysis, information retrieval and machine learning.

Guo Wenya, born in 1994. Master. Her main research interests include sentiment analysis and data mining.

Yuan Xiaojie, born in 1963. PhD, professor, PhD supervisor. Her main research interests include data management and data mining.