

基于交叉注意力机制和新闻正文的评论情感分类

王启发 王中卿 李寿山 周国栋

苏州大学计算机科学与技术学院 江苏 苏州 215006

(20185227011@stu.suda.edu.cn)



摘要 目前,新闻评论已成为新闻的重要衍生数据,新闻评论表达了评论者对新闻事件的观点、立场及个人情感,通过对新闻评论的情感倾向性分析有助于了解社会舆情及动向,因此,新闻评论的情感研究备受广大学者的青睐。通常的新闻评论情感分析只考虑评论文本本身的信息,而新闻评论文本信息和新闻正文信息往往是紧密相关的,基于此,文中提出一种基于交叉注意力机制并结合正文的新闻评论情感分类方法。首先通过双向长短时记忆网络模型分别对新闻正文与评论文本进行特征表示;然后基于交叉注意力机制进一步捕获评论与正文的关系,得到两个更新后的新闻正文与评论文本的向量表示;再将两者拼接得到的语义表示输入全连接层,使用 sigmoid 激活函数进行分类预测,从而实现新闻评论的情感分类。结果表明,提出的基于交叉注意力机制和新闻正文的评论情感分类模型可以有效地提升新闻评论情感分类的准确率,该模型的 F1 值较 3 个基准模型分别提高了 1.72%,3.24% 和 6.21%。

关键词: 注意力机制;双向长短时记忆网络;情感分类

中图法分类号 TP391.1

Comment Sentiment Classification Using Cross-attention Mechanism and News Content

WANG Qi-fa, WANG Zhong-qing, LI Shou-shan and ZHOU Guo-dong

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract At present, news comment has become important news derived data. News comment expresses commentators' views, positions and personal feelings on news events. Through the analysis of sentiment orientation of news comment, it is helpful to understand the social public opinion and trend. Therefore, the sentiment research of news comment is favored by many scholars. The usual news comment sentiment analysis only considers the information of the comment text itself. However, news comment text information is often closely related to news content information. Based on this, this paper proposed a comment sentiment classification method using cross-attention mechanism and combined with news content. Firstly, the bi-directional long short-term memory network model is used to characterize the news content and the comment text respectively. Then, the cross-attention mechanism is used to further capture important information, and obtain the vector representation of the two updated news content texts and comment texts. And then the semantic representation obtained by splicing them together is input into the full connection layer, and sigmoid activation function is used for classification prediction, so as to realize the sentiment classification of news comments. The results show that the model of comment sentiment classification using cross-attention mechanism and news content can effectively improve the accuracy of sentiment classification of news comment, and this model improves by 1.72%, 3.24% and 6.21% on F1 respectively compare with the three benchmark models.

Keywords Attention mechanism, Bi-directional Long Short-Term Memory Network, Sentiment Classification

1 引言

随着互联网技术的飞速发展和移动终端的快速普及,新闻的传播速度越来越快,人们获取的信息量越来越多,加之网络平台具有开放性、透明性、实时性、互动性和低门槛性,为人

们表达意见和交流信息提供了媒介,从而衍生出更多的数据信息,例如新闻评论数据,就是新闻的重要衍生数据,现在几乎每一条新闻都伴有或多或少的评论,尤其对于一些热点新闻,其评论更是千姿百态,每天仅新闻方面就会产生不计其数的信息数据。在大数据时代,每一份数据都具有一定的价值,

到稿日期:2019-09-25 返修日期:2019-12-06 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金青年科学基金项目(61806137,61702518);江苏省高等学校自然科学研究面上项目(18KJB520043)

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (61806137,61702518) and Natural Science Foundation of the Jiangsu Higher Education Institutions of China (18KJB520043).

通信作者:王中卿(wangzq@suda.edu.cn)

新闻评论数据自然也不例外,于是吸引了自然语言处理、语言学等领域的一大批学者从不同角度对其进行分析研究。

目前,虽然有相关学者针对新闻评论做了一系列情感分析的工作,但很多都是对评论做单独的考虑和处理,如文献[1-5],无论是基于传统机器学习方法还是基于深度学习方法,都只是针对评论文本本身进行情感词统计或是特征提取。然而,评论信息往往与新闻正文信息有着密切的联系,即新闻正文内容会影响评论的情感表达。如表1的样例,从评论的字面理解,我们会认为评论作者表达的是“支持”态度,但结合新闻正文来看,评论作者支持的是“一胎”,而非新闻中的“二胎政策”,所以该条评论最终表达的情感应为“不支持”。

表1 新闻正文及评论样例

Table 1 Sample of news content and comment

新闻	【国家正式开放二胎的政策下来啦!赞!】你支持二胎政策吗?
评论	一胎多好!支持一胎!

基于此,本文将同时考虑新闻正文信息和新闻评论信息对评论情感极性判断的影响。首先基于长短时记忆神经网络模型构建新闻正文和评论文本的表示,由于注意力机制在计算机视觉和自然语言处理领域都有相对优异的表现,因此,为了聚焦新闻正文和评论之间的相互联系,我们采用本文提出的交叉注意力机制对文本进一步捕获重要信息,并得到两个更新后的新闻正文与评论文本向量表示,然后对二者进行向量拼接,从而实现新闻正文与评论信息融合后的情感分类。通过设计相关实验,验证了将新闻正文信息与评论信息融合以及引入交叉注意力机制对评论情感极性的判断具有明显的提升作用。

本文第2节主要介绍了情感分类的相关工作;第3节介绍了实验相关数据、实验的具体方法流程以及训练和优化策略,重点描述了本文构建的模型;第4节介绍了实验数据的采集处理、实验参数设置、评估准则、基准系统以及实验结果分析;最后,对研究工作进行总结,同时提出了下一步的研究方向。

2 相关工作

情感分类是情感分析的任务之一,一直以来都是自然语言处理领域的研究热点。情感分类是指根据文本所表达的含义和情感信息将文本划分成褒扬的或贬义的两种或几种类型,是对文本作者的倾向性、观点和态度的划分,因此有时也被称作倾向性分析(opinion analysis)。情感分类作为一种特殊的分类问题,既有一般模式分类的共性特征,也有其特殊性,如情感信息表达的隐蔽性、多义性和极性不明显等。

近年来,随着神经网络的快速发展,很多神经网络技术也被应用到情感分类当中。2011年,Zhou等[6]提出了一种半监督情感分类的主动深度网络(Active Deep Networks,ADN)。首先,对基于标记的评论和大量未标记的评论进行无监督学习,然后通过具有指数损失函数的基于梯度下降的监督学习构造的结构进行微调,最后对5个情感分类数据集进行测试。实验证明ADN优于经典的半监督学习算法。随着卷积神经网络(Convolutional Neural Networks,CNN)[7]在

计算机视觉领域取得了实质性的进展,Kim等[8]在预处理的词向量上训练卷积神经网络,进行句子级分类任务的一系列实验,其中就包括情感分类任务。他们提出的基于CNN的句子分类模型在多个公共数据集上都取得了较好的分类效果。Cao等[9]发现,卷积神经网络与支持向量机(Support Vector Machine,SVM)分类器相比,虽然情感分类的准确度得到了提升,但是不能有效地执行非线性分类,因此,他们提出了一种将CNN与SVM结合起来的模型。实验表明,该模型能达到较高的情感分类准确率。Tang等[10]提出一种基于门控递归神经网络的文档建模方法,对来自IMDB的4个大型review数据集进行文档级情感分类。实验结果表明,门控递归神经网络在情感分类文档建模方面明显优于标准递归神经网络。

随着深度学习的迅速发展,长短时记忆网络(Long Short-Term Memory Network,LSTM)[11]作为一种具有更复杂计算单元的递归神经网络,因具有优良的序列信息保存能力,逐渐被应用于各种序列建模任务,并取得了很好的效果。Wang等[12]将LSTM模型应用到Tweets的情感分类任务中,并取得了不错的效果。Tai等[13]将传统的LSTM结构推广到树状网络拓扑结构来进行情感分类任务。之后,国内外学者发现,将注意力机制[14-16]引入文本情感分类任务中有明显的效果,而且注意力机制在自然语言处理的其他任务中表现也很优异。Yang等[17]针对文本分类提出了一个层次化的注意力机制,分别就词和句子构建两个层次的注意力机制,使得文本中重要性不同的句子和词被给予了不同的“注意力”能力,实验结果相较以往的模型效果提升显著。Li等[18]提出了一种用于跨域情感分类的分层注意转移网络(Hierarchical Attention Transfer Network,HATN)。HATN提供了分层的注意力转移机制,可以通过自动捕获枢轴和非枢轴来转移跨域的情绪注意力。Zhou等[19]提出了一种面向双语的基于分层注意力机制和LSTM的情感分类模型,该模型在以英语为源语言、中文为目标语言的基准数据集上取得了较好的效果。

本文深受以上相关工作的启发,尝试将长短时记忆神经网络和注意力机制相结合,并应用到新闻评论的情感分类中。由于新闻正文是评论者表达观点、立场以及情感的重要依据,因此,我们设想用双向LSTM学习新闻正文和新闻评论的文本特征,用交叉注意力机制进一步捕获重要信息,最后将二者信息融合进行情感极性分类。根据相关工作分析,我们的设想是可行的。因此,本文开展了基于交叉注意力机制和新闻正文的评论情感分类的研究工作。

3 基于交叉注意力机制和新闻正文的评论情感分类

本文首先利用双向长短时记忆网络(Bi-directional Long Short-Term Memory Network,BiLSTM)分别对新闻文本和评论文本进行特征表示。为了进一步捕获新闻正文和评论之间相互作用的重要信息,本文采用交叉注意力机制让新闻正文和评论相互作用,从而得到更准确的新闻正文和评论各自的特征向量表示,然后对两个向量进行拼接,得到新闻正文和评论信息融合后的向量表示,从而实现融合新闻正文信息的

评论情感分类。本文整体模型如图 1 所示。

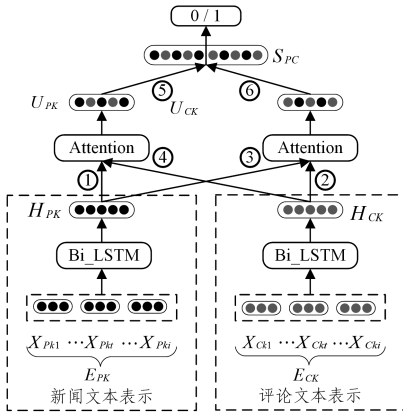


图 1 基于交叉注意力机制和新闻正文的评论情感分类模型

Fig. 1 Model of comment sentiment classification based on cross-attention mechanism and news content

3.1 评论信息的表示

新闻评论是广大网民对新闻事件发表的个人观点。本文先将每一条评论文本进行预处理,采用 jieba 分词技术对其进行中文分词,得到评论序列,根据词频统计后得到的词汇表对评论序列中的每一个元素编码为 0~9842 的正整数,第 k 条评论可以表示为 $C_k = [\omega_{Ck1}, \omega_{Ck2}, \omega_{Ck3}, \dots, \omega_{Cki}]$,其中 $\omega_{Cki} \in R$ (R 表示 0~9842 的正整数集合, ω_{Cki} 即为该词在词汇表 V 中对应的编号)。由于本实验中限制了输入句子的最大长度为 30,于是对输入的评论文本向量长度进行统一,长度超过 30 的句子从前端截断,长度不足的在句子前端补 0。

通过以上方法得到的句子向量 C_k 经过 Embedding 词嵌入表示为向量:

$$E_{Ck} = [X_{Ck1}, X_{Ck2}, X_{Ck3}, \dots, X_{Ckt}, \dots, X_{Cki}] \quad (1)$$

本文使用 BiLSTM 来学习评论的深层文本信息的表示。BiLSTM 对句子的嵌入表示进行编码,得到相应的隐含向量表示,其具体过程如下:

$$\vec{h}_{Ckt} = \overrightarrow{LSTM}(x_{Ckt}) \quad (2)$$

$$\overleftarrow{h}_{Ckt} = \overleftarrow{LSTM}(x_{Ckt}) \quad (3)$$

$$h_{Ckt} = [\vec{h}_{Ckt}, \overleftarrow{h}_{Ckt}] \quad (4)$$

$$H_{Ck} = [h_{Ck1}, h_{Ck2}, \dots, h_{Ckt}, \dots, h_{Cki}] \quad (5)$$

其中, $\vec{h}_{Ckt}$ 和 $\overleftarrow{h}_{Ckt}$ 分别为句子 C_k 中 t 时刻的输入通过前向和反向 LSTM 神经网络后得到的状态表示, h_{Ckt} 表示句子 C_k 中 t 时刻的输入通过 BiLSTM 得到的状态表示。

3.2 新闻正文信息的表示

本文研究的是新闻评论的情感分析,不难想到,新闻正文信息或多或少会对评论情感分析产生一定的影响。因此,实验中考虑加入新闻正文文本信息,如同评论文本的预处理方法,先对新闻正文文本进行中文分词,得到新闻正文序列,再根据词汇表对该序列中的每一个元素编码,于是第 k 条新闻正文文本可以表示为 $P_k = [\omega_{Pk1}, \omega_{Pk2}, \omega_{Pk3}, \dots, \omega_{Pki}]$,其中 $\omega_{Pki} \in R$ (R 表示 0~9842 的正整数集合, ω_{Pki} 即为该词在词汇表 V 中对应的编号)。采用与评论表示相同的方法对输入的新闻文本向量长度进行统一。

以上得到的句子向量 P_k 经过 Embedding 词嵌入表示为向量:

$$E_{Pk} = [X_{Pk1}, X_{Pk2}, X_{Pk3}, \dots, X_{Pkt}, \dots, X_{Pki}] \quad (6)$$

本文使用 BiLSTM 来学习评论的深层文本信息的表示。BiLSTM 对句子的嵌入表示进行编码,得到相应的隐含向量表示 H_{Pk} ,具体过程如下:

$$\vec{h}_{Pkt} = \overrightarrow{LSTM}(x_{Pkt}) \quad (7)$$

$$\overleftarrow{h}_{Pkt} = \overleftarrow{LSTM}(x_{Pkt}) \quad (8)$$

$$h_{Pkt} = [\vec{h}_{Pkt}, \overleftarrow{h}_{Pkt}] \quad (9)$$

$$H_{Pk} = [h_{Pk1}, h_{Pk2}, \dots, h_{Pkt}, \dots, h_{Pki}] \quad (10)$$

其中, h_{Pkt} 表示句子 P_k 中 t 时刻的输入通过 BiLSTM 得到的状态表示。

3.3 基于交叉注意力机制和新闻正文的评论情感分类

注意力机制的作用是聚焦重要信息,本文研究的是基于新闻正文的评论情感分类。为了将新闻正文文本信息与评论信息进行有效的融合,本文提出了一种基于交叉注意力机制的文本信息融合方法,即让新闻正文同评论互相作用,得到对当前任务更重要、更准确的特征表示,从而实现新闻正文与评论信息融合后的情感分类,具体流程如下。

1) 新闻评论作用于新闻正文的注意力机制:用新闻评论信息为其对应的新闻正文信息添加注意力,从而对新闻正文文本信息中与评论文本信息联系密切的信息给予更大的关注度,如图 1 中①、④的箭头指向所示,最后得到一个新的新闻正文文本信息的表示,其中 Attention 机制的计算公式如下:

$$U_{Pk} = \sum_{t=1}^T \partial_{Ct} h_{Pkt} \quad (11)$$

$$\partial_{Ct} = \frac{\exp(u_t)}{\sum_{j=1}^T \exp(u_j)} \quad (12)$$

$$u_t = \text{Tanh}(\mathbf{W}^T h_{Pkt} + b) \quad (13)$$

其中, h_{Pkt} 表示新闻正文向量 H_{Pk} 中的第 t 个单词的向量表示, ∂_{Ct} 表示评论向量 H_{Ck} 到单词 h_{Pkt} 的注意力权重, Tanh 是一种非线性激活函数, $\mathbf{W}$ 是中间矩阵, b 是偏移量,它们都在训练期间被随机初始化和更新。

2) 新闻正文作用于新闻评论的注意力机制:用新闻正文的信息为其对应的新闻评论信息添加注意力,从而对评论文本信息中与新闻正文文本信息联系密切的信息给予更大的关注,如图 1 中②、③的箭头指向所示,注意力机制的方式与上述 H_{Ck} 作用于 H_{Pk} 的注意力机制方式相同,最后,得到一个新的评论文本信息的表示 u_{Ck} ;

3) 将 u_{Pk} 与 u_{Ck} 行拼接,如图 1 中⑤、⑥的箭头指向所示,得到新闻文本信息与评论信息相融合的文本表示 S_{PC} ;

4) 将信息融合后的文本表示输入到一个全连接层,使用 sigmoid 激活函数,最终输出一个 0~1 之间的概率值,本文规定概率值大于 0.5 的文本归为正类,小于 0.5 的文本归为负类。

4 实验与分析

4.1 实验数据集

4.1.1 数据采集

本文使用的数据主要来源于新浪微博。通过网络爬虫程

序爬取新闻正文及其评论,并进行数据清洗,去除重复评论、回复评论、含非法字符评论和句子长度过短评论,最终筛选出评论长度大于 10 个字符、评论条数大于 150 的新闻,共计 200 个。

4.1.2 数据标注

本文实验是对新闻评论进行情感分类,因此需要对所采集的数据进行人工标注,每一条新闻挑选大约 50 条内容相似度小于 80% 的评论进行情绪分类标注,情绪共分为 5 类: Happiness, Anger, Sadness, Fear, Surprise。最终共标注 200 个新闻,10 077 条评论。

4.1.3 数据预处理

由于部分情绪类别语料较少,不宜做多分类任务。本文主要做二分类任务,将 Happiness 归为正类,其余 4 类归为负类,然后分别对标注数据的情感标签(label)、新闻正文和新闻评论这 3 部分进行切分、编码、分词,最终得到情感标签序列、新闻正文序列和新闻评论序列。为了保证神经网络学到的信息的均衡性,本实验对训练样本做了均衡处理,分别选取正类样本 1 400 条,负类样本 1 400 条,共计 2 800 条作为训练集,再对训练集用直接计数统计法进行词频统计,根据词频由高到低进行倒排序,再将排序好的词表依次从 0 开始编号,词为键,编号为值,随后得到一个具有 9 843 个键值对的词汇字典;然后根据词汇字典对所有的新闻正文序列和评论序列进行编码,即将文本序列的每个词转换成词汇字典中该词相对应的编号,于是得到新闻正文句子向量和评论句子向量。实验数据集的分布情况如表 2 所列。

表 2 数据集的相关信息

Table 2 Related information of dataset

数据类别	数量统计
新闻总数	200
评论总数	10 077
正例样本数	2 409
负例样本数	7 438
训练集样本数	2 800(1 400/类)
验证集样本数	400(200/类)
测试集样本数	3 286
词汇表词汇总数	9 843
新闻正文最大长度	315
评论最大长度	255

4.2 实验参数设置

为了获取模型的最优参数集合,本实验中需要调节的参数如表 3 所列。

表 3 模型参数调节列表

Table 3 Model parameter adjustment list

参数	取值
Optimizer	Adam
Loss	Binary_crossentropy
Learn_rate	0.000 1
Batch_size	25
Embed_size	128
Hidden_size	64
Max_len	30
Epoch	5
Dropout	0.2

4.3 评估准则

本实验主要采用综合评价指标(F1 值)来衡量分类的准确率,相关公式如下:

$$F1=\frac{2\times P\times R}{P+R}$$

(14)

$$P=\frac{TP}{TP+FP}$$

(15)

$$R=\frac{TP}{TP+FN}$$

(16)

其中,TP 表示将正例预测为正例的个数,FP 表示将反例预测为正例的个数,FN 表示将正例预测为反例的个数。

4.4 基准系统

我们选用以下方法进行实验的对比分析,主要分为 3 种方式。

(1)只考虑评论信息

LSTM_C 模型:仅输入评论文本,采用 LSTM 对评论文本进行特征向量表示,然后经过一个全连接层,使用 sigmoid 激活函数,输出 0~1 之间的概率值。

(2)新闻正文信息与评论信息相融合

LSTM_PC 模型:新闻正文和评论均采用 LSTM 进行文本表示,然后将两个 LSTM 层输出表示向量进行拼接,经过一个全连接层,使用 sigmoid 激活函数,输出 0~1 之间的概率值。

(3)新闻标题与评论信息相融合

LSTM_TC 模型:新闻标题和评论均采用 LSTM 进行文本表示,然后将两个 LSTM 层输出表示向量进行拼接,经过一个全连接层,使用 sigmoid 激活函数,输出 0~1 之间的概率值。

4.5 实验结果与分析

4.5.1 与基准系统结果的对比

本文提出了基于交叉注意力机制和新闻正文的评论情感分类的模型——BiLSTM_P2CAtt+C2PAtt 模型:新闻正文和评论均采用双向 LSTM 进行文本表示,并引入交叉注意力机制得到两个新的正文和评论向量表示,然后将二者进行拼接,经过一个全连接层,使用 sigmoid 激活函数,输出 0~1 之间的概率值。与以上几种模型相比,该模型性能最优。

为了体现本文提出的融合交叉注意力机制和新闻正文信息的评论情感分类的有效性,将本文模型与基准系统进行对比,实验结果如表 4 所列。

表 4 本文模型与基准系统的对比

Table 4 Comparison of proposed model and baseline

模型	F1 值
LSTM_C	0.733 2
LSTM_TC	0.762 9
LSTM_PC	0.778 1
BiLSTM_P2CAtt+C2PAtt	0.795 3

由表 4 可以得出以下几点结论:

1)从 LSTM_C,LSTM_TC,LSTM_PC 3 个模型的对比实验可以看出,LSTM_C 的 F1 值小于 LSTM_TC 和 LSTM_PC 的 F1 值,这说明仅仅考虑评论文本自身的信息对评论进

行情感分类是不够的,因为网络新闻评论文本相对简短且口语化,有很多重要的信息可能被省略,而这些被省略的信息大多来自新闻标题和正文,因此,我们将评论分别与新闻标题、新闻正文相结合进行情感分类实验,结果表明,融合新闻标题或新闻正文信息均能提升新闻评论情感分类的准确率。

2)对比 LSTM_TC 和 LSTM_PC 两个模型,从实验结果可以看出,将新闻评论和新闻正文信息融合的实验效果要优于将新闻评论和新闻标题信息融合的实验效果,这是由于新闻正文文本中包含的信息远远多于新闻标题的信息,因此通过这两组实验可以证明本文将新闻正文与评论信息融合进行评论情感分类是有效的。由于新闻正文比标题更有效,本文在后面的研究中主要基于新闻正文。

3)通过 LSTM_PC 和 BiLSTM_P2CAtt+C2PAtt 两个模型的对比实验可以发现,直接用 LSTM 对新闻正文和新闻评论进行学习表示然后进行向量拼接的实验效果较差,而引入本文提出的交叉注意力机制后,实验效果得到了显著的提升,这是由于新闻正文信息中包含一些重要的信息,通过交叉注意力机制可以有效地增强模型对于文本信息的捕获能力,进而提高了评论情感分类的性能。由此证明,本文提出的基于交叉注意力机制和新闻正文的评论情感分类的模型确实可以有效地提升评论情感分类的准确率。

4.5.2 不同因素对实验结果的影响

为了验证模型的有效性,本文考虑了不同因素对实验的影响,主要分为以下 3 个方面。

(1)注意力机制的影响

LSTM_PC 模型:新闻正文和评论均采用 LSTM 进行文本表示,然后将两个 LSTM 层输出表示向量进行拼接,经过一个全连接层,输出 0~1 之间的概率值。

LSTM_PC+ SelfAtt 模型:在 LSTM_PC 模型的基础上,对两个 LSTM 层输出表示向量拼接后的文本表示向量引入自注意力机制。

(2)网络结构的影响

BiLSTM_P2CAtt 模型:新闻正文和评论均采用双向 LSTM 进行文本表示,并引入正文作用于评论的注意力机制,得到新的评论向量表示,然后经过一个全连接层,输出 0~1 之间的概率值。

BiLSTM_P2CAtt+C2PAtt 模型:新闻正文和评论均采用双向 LSTM 进行文本表示,并引入交叉注意力机制得到两个新的正文和评论向量表示,然后将二者进行拼接,经过一个全连接层,输出 0~1 之间的概率值。

(3)辅助特征的影响

BiLSTM_T2CAtt+C2TAtt 模型:该模型与本文提出的基于交叉注意力机制和新闻正文的评论情感分类的模型网络结构完全一致,只是将新闻正文作为新闻评论的情感分析辅助特征改成了将新闻标题作为辅助特征,即新闻标题和评论均采用双向 LSTM 进行文本表示,并引入交叉注意力机制得到两个新的正文和评论向量表示,然后将二者进行拼接,经过一个全连接层,输出 0~1 之间的概率值。

对比实验的结果如表 5 所列。

表 5 不同模型的最优 F1 值的对比

Table 5 Comparison of optimal F1 values of different models

模型	F1 值
LSTM_PC	0.7781
LSTM_PC+ SelfAtt	0.7906
BiLSTM_P2CAtt	0.7817
BiLSTM_T2CAtt+C2TAtt	0.7835
BiLSTM_P2CAtt+C2PAtt	0.7953

由表 5 可以得出以下几点结论:

1)通过 LSTM_PC 模型与 LSTM_PC+ SelfAtt 模型的对比实验可以看出,注意力机制对实验效果具有一定的影响。可以发现,单纯地将新闻正文的表示向量与评论的表示向量进行拼接来预测评论的情感极性效果较差,而对拼接后的向量,即新闻正文与评论信息融合后的向量引入注意力机制,可显著提升实验效果,这是由于注意力机制可以聚焦当前任务中相对重要的信息,即进一步捕获新闻正文与评论中的重要信息,使得特征向量表示更加准确,从而提升评论情感分类的准确率。

2)将 BiLSTM_P2CAtt+C2PAtt 模型与 BiLSTM_P2C-Att 模型做对比实验,从结果可以看出,前者的 F1 值高于后者,即本文提出的 BiLSTM_P2CAtt+C2PAtt 模型的实验效果更好。前者只是新闻正文对评论作用的注意力机制,因此只是对评论文本信息中与新闻正文文本信息联系密切的信息给予了更大的关注,最终得到的只是捕获了部分重要信息的新的评论表示向量,所以预测的效果较差。而后者增加了评论对新闻正文作用的注意力机制,这样既可以得到捕获了重要信息的新的评论表示向量,也可以得到捕获了重要信息的新的新闻正文的表示向量,然后将二者信息进行融合,从而得到更加全面、准确的特征表示向量,因此该模型能够有效地提高评论情感分类的准确率。

3)通过对比 BiLSTM_T2CAtt+C2Tatt 和 BiLSTM_P2C-Att+C2PAtt 两个模型,由实验结果可知,后者的 F1 值高于前者,即在同样的神经网络框架下,将新闻正文信息作为评论情感分类的辅助特征的实验效果要优于将新闻标题信息作为评论情感分类的辅助特征的实验效果,这是由于新闻正文文本中包含的信息要远远多于新闻标题包含的信息,所以通过这两组实验可以证明本文将新闻正文与评论信息融合进行评论情感分类是有效的。

结束语 本文针对收集和标注的新闻评论及新闻正文数据展开了基于交叉注意力机制和新闻正文的评论情感分类实验,通过认真分析实验结果,不断调整模型结构,反复修改模型的各项参数,使得所提模型的性能逐渐提升。最终,实验证明,提出的基于交叉注意力机制和新闻正文的评论情感分类模型相比其他实验模型而言,有显著的效果提升。通过实验可以发现,新闻正文的信息确实对评论的情感分类有一定的影响,将新闻正文和评论的信息进行融合可以有效地提升评论情感分类的准确率,而再引入注意力机制,实验效果还会得到进一步提高。

参 考 文 献

[1] FAN W,SUN S. Sentiment classification for online comments on

- Chinese news[C]//International Conference on Computer Application & System Modeling. IEEE,2010.
- [2] NING Y,ZHU T,WANG Y. Affective-word based Chinese text sentiment classification[C]//International Conference on Pervasive Computing & Applications. IEEE,2011.
- [3] SU Z,HUA X,ZHANG D,et al. Chinese sentiment classification using a neural network tool—Word2vec[C]//International Conference on Multisensor Fusion & Information Integration for Intelligent Systems. 2014.
- [4] ZHANG D,XU H,SU Z,et al. Chinese comments sentiment classification based on word2vec and SVMperf[J]. Expert Systems with Applications,2015,42(4):1857-1863.
- [5] ABDULLIN Y B,IVANOV V V. Deep learning model for bilingual sentiment classification of short texts[J]. Naučno-tehnički Vestnik Informacionnih Tehnologij,2017,17(1):129-136.
- [6] ZHOU S,CHEN Q,WANG X. Active deep networks for semi-supervised sentiment classification[C]//Proceedings of the 23rd International Conference on Computational Linguistics:Posters. Association for Computational Linguistics,2010:1515-1523.
- [7] 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.
- [8] KIM Y. Convolutional Neural Networks for Sentence Classification[C]//Proceedings of the 2014 conference on empirical methods in natural language processing. 2014:1746-1751.
- [9] CAO Y,XU R,CHEN T. Combining Convolutional Neural Network and Support Vector Machine for Sentiment Classification [C]//Proceedings of the Meeting of the Chinese National Conference on Social Media Processing. 2015:144-155.
- [10] TANG D,QIN B,LIU T. Document modeling with gated recurrent neural network for sentiment classification[C]//Proceedings of the 2015 conference on empirical methods in natural language processing. 2015:1422-1432.
- [11] HOCHREITER S,SCHMIDHUBER,JÜRGEN. Long Short-Term Memory[J]. Neural Computation,1997,9(8):1735-1780.
- [12] WANG X,LIU Y,CHENGJIE S U N,et al. Predicting polarities of tweets by composing word embeddings with long short-term memory[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1:Long Papers). 2015,1:1343-1353.
- [13] TAI K S,SOCHER R,MANNING C D. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks [J]. Computer Science,2015,5(1):36.
- [14] MNIH V,HEESS N,GRAVES A. Recurrent models of visual attention [J]. Advances in Neural Information Processing Systems,2014(3):2204-2212.
- [15] BAHDANAU D,CHO K,BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate [C]//Proceedings of the ICLR. 2015:1-15.
- [16] LUONG M,MANNING C. D. Effective Approaches to Attention-based Neural Machine Translation [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015:1412-1421.
- [17] YANG Z,YANG D,DYER C,et al. Hierarchical Attention Networks for Document Classification [C] // Proceedings of the North American Chapter of the Association for Computational Linguistics;Human Language Technologies. 2017:1480-1489.
- [18] LI Z,WEI Y,ZHANG Y,et al. Hierarchical attention transfer network for cross-domain sentiment classification[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [19] ZHOU X,WAN X,XIAO J. Attention-based LSTM network for cross-lingual sentiment classification[C] // Proceedings of the 2016 conference on empirical methods in natural language processing. 2016:247-256



WANG Qi-fa, born in 1994, postgraduate, is a member of China Computer Federation. His main research interests include natural language processing and emotion analysis.



WANG Zhong-qing, born in 1987, Ph.D., postgraduate supervisor, is a member of China Computer Federation. His main research interests include natural language understanding and information extraction.