山西大学学报(自然科学版)47(4):751—760,2024
Journal of Shanxi University(Nat. Sci. Ed.)

DOI:10.13451/j.sxu.ns.2023067

结合交叉注意力的双通道恶意评论识别方法

张琳钰,卢益清*
(北京信息科技大学 信息管理学院,北京 100192)

摘　要:恶意评论识别本质上是一个文本分类的问题。相较传统的文本分类,恶意评论往往伴随着表达方式更微妙且随意的特点,使得传统文本分类网络识别精度不高、识别效果不好,无法达到需求。为解决上述问题,本文提出一种结合交叉注意力机制的双通道文本分类网络(Two-channel text classification network combined with cross-attention mechanism, CA2TC)。该模型同时使用图卷积神经网络(Graph Convolutional Network,GCN)和双向长短期记忆网络(Bidirectional Long Short-term Memory, BiLSTM)获得两种不同的文本上下文特征信息,两种不同的特征信息可以从多个角度更好表达文本的含义。提出的交叉注意力机制对双通道提取的文本特征进行精炼并融合。最后将精炼特征拼接后经全连接层再送入softmax进行分类。本文采用微博收集的恶意评论数据对提出的方法进行实验验证。实验结果表明,与一些主流的分类模型相比,提出的模型识别效果更优,分类精度较主流分类模型相比提高1.06%至2.89%。CA2TC模型能够充分提取恶意评论文本特征,从而有效识别恶意评论。

关键词:恶意评论识别;双通道;图卷积神经网络;双向长短期记忆网络;交叉注意力机制
中图分类号:TP391　　**文献标志码**:A　　**文章编号**:0253-2395(2024)04-0751-10

Two-channel Malicious Comment Recognition Method Combined with Cross-attention Mechanism

ZHANG Linyu, LU Yiqing*
(*School of Information Management, Beijing Information Science and Technology University, Beijing* 100192*, China*)

Abstract: The detection of malicious comments is essentially a text classification problem. Compared to typical text classification, malicious comments are often accompanied by more subtle and unpredictable expressions, which results in low identification accuracy, poor recognition effect, and inability to satisfy demands. To tackle the aforementioned issues, this paper proposes a two-channel text classification network combined with cross-attention mechanism (CA2TC), which employs graph convolutional network (GCN) and bidirectional long short-term memory (BiLSTM) to generate two distinct texts. Contextual feature information, as well as two distinct feature information, may be used to better explain the meaning of the text from numerous viewpoints. The suggested cross-attention approach improves and combines text characteristics gathered from two channels. Finally, the corrected features are concatenated and transmitted to softmax through the fully connected layer for classification. The malicious comment data acquired from Weibo is utilized in this study to validate the suggested strategy. The experimental results show that, compared to some mainstream classification models, the proposed model has a better recognition effect, with classification accuracy increasing by 1.06% to 2.89%. The CA2TC model can fully extract the text features of malicious comments, leading effectively identify malicious comments.

Key words: malicious comment recognition; two-channel; GCN; BiLSTM; cross-attention mechanism

收稿日期:2023-02-24;**接受日期**:2023-03-28
基金项目:国家自然科学基金(U1936111)
作者简介:张琳钰(1997-),女,山西长治人,硕士研究生,研究方向为自然语言处理。E-mail:13116149862@163.com
* **通信作者**:卢益清(LU Yiqing),E-mail:luyiqing@126.com
引文格式:张琳钰,卢益清.结合交叉注意力的双通道恶意评论识别方法[J].山西大学学报(自然科学版),2024,47(4):751-760. DOI:10.13451/j.sxu.ns.2023067

0 引言

由于网络中的信息日趋庞大,人们对其难辨真伪。有现象表明,网民由于内心的道德感容易因为一些未经证实的信息对某个人或某个群体进行言语上的攻击而对其造成一定的伤害。这种对人的身心健康具有严重负面影响的文本称为恶意评论。恶意评论在社交网络中大量存在会引起更多不知情的人民群众对当事人做出偏离事实的评价,进而引起网络暴力。因此,在社交网络中将恶意评论识别并处理是十分重要的。由于社交网络中评论信息的庞大性,组织人工去筛选恶意评论显然已不能满足营造良好网络生态的需求,一种识别恶意评论的自动化手段亟需被实现。

目前,有学者针对恶意评论识别做了相关研究。杨金灵[1]将词向量集成技术和数据增强技术相结合,提出了集成词向量与数据增强的恶意文本分类模型(ENSVEC-DA)模型,以此来解决恶意评论识别任务中存在的词向量单一和样本数据集有限的问题。吴浩等[2]提出了基于 BERT (Bidirectional Encoder Representations from Transformers)和 RCNN(Recurrent Convolutional Neural Networks)的检测中文违规评论的模型(Recognition Model of Illegal Comments Based on BERT-RCNN,RBR)来对中文违规评论进行识别。周娅等[3]为了解决类不平衡和样本重叠问题,提出了一种基于分层欠采样和 Bi-GRU(Bidirectional Gated Recurrent Unit)的恶意行为检测模型(SSU-BG)。针对恶意评论识别的方法大多侧重于从数据维度来提升恶意评论识别的精度,而没有考虑模型对恶意评论识别的影响。运用深度学习模型来对恶意评论进行检测的研究中,只对恶意评论的单一特征进行提取,而没有考虑恶意评论文本之间的关系,即整个文档的空间信息。因此对提取恶意文本的特征信息方面仍然具有提升空间。

究其本质,恶意评论识别是文本分类的一个分支任务。文本分类是将一组文本自动分为预定义的类别的任务,是自然语言处理(Natural Language Processing, NLP)的一个重要任务,已被应用于推荐系统[4]、垃圾邮件过滤系统[5]等领域。传统文本分类的方法研究多专注于基于规则和基于统计的文本分类方法,如基于情感词典的分类、朴素贝叶斯分类。但基于规则和基于统计的文本分类方法研究多集中于一类特定的短语或句子,而不能考虑词语与词语之间、句子与句子之间的内在联系。尤其是传统的基于情感词典的分类方法偏向于计算文本出现的情感词的情感极性来判断文本的极性。而如今的文本数据并不是将显而易见的特殊词进行堆积,句子的意义隐藏在更深层的表达中,因此应考虑上下文语义而不是单独去看每个词语来对文本进行分类。所以,传统的分类方法虽利用文本中显著的特点而被成功应用于各个领域,但已经不适用于现今的文本分类。

深度学习[6]近年来在文本分类、机器翻译等多个领域取得了显著成果。在文本分类任务中,对于深度学习方法的研究主要分为卷积神经网络[7](convolutional neural network, CNN)、循环神经网络[8](Recurrent Neural Network, RNN)与图卷积神经网络[9](Graph Convolutional Network, GCN)三种。

CNN 可以处理高维的数据,但无法对时间序列上的变化进行建模,而文本本身是一种时序数据,所以时序对文本分类来说至关重要。相比 CNN,RNN 可以解决文本的时序问题,在文本分类领域使用较多。当前,比较主流的方法是长短记忆网络[10](Long Short-term Memory, LSTM),其有效解决了 RNN 在训练过程中容易出现的梯度消失和梯度爆炸的问题,而且还能捕获文本的长距离依赖关系。双向长短期记忆网络[11](Bidirectional Long Short-term Memory, BiLSTM)是 LSTM 的进一步发展,BiLSTM 可以访问前向上下文和后向上下文,比 LSTM 可以更好地解决顺序建模任务。目前,LSTM 和 BiLSTM 已应用于文本分类,并取得了一定的成果。文本也可以表示为图结构,因此,GCN 也被用于处理文本分类。相比 CNN、RNN,GCN 通过对词与词之间、词与文本之间建立边来构图,从而获取词与词、词与文档之间的关系。GCN 构成的图结构包含了整个数据集中的信息。不同于 RNN 只能对文本按顺序来逐词处理,通过 GCN 得到的文本特征信息可以提取数据集中包含的空间信息,尤

其是恶意评论之间的相关性信息,从而更好地对文本进行分类。近年来,GCN在文本分类领域得到很多应用。

图卷积神经网络通过不断整合节点与节点之间的信息,一个节点可以得到与该节点周围的邻接节点的信息,因此整合文档信息的能力强大,但将文本构建为图后就会缺失文本所具有的时序性的特点。BiLSTM对文本顺序建模可以提取文本的时序特征,并且解决了文本的长距离依赖问题。但文本高维向量作为LSTM的输入会导致网络参数急剧增加而使网络结构难以优化的情况。卷积运算可以在提取特征的同时用降维的方式解决这个问题。因此,BiLSTM需要与卷积运算结合来对向量进行降维。GCN、BiLSTM虽然可以获得文本的大量特征信息,但不可能将重点放在所获得的上下文信息中的重要信息上。关注重要信息将提高分类的准确性。注意力机制可以通过设置不同的权重来突出上下文信息中的重要信息。将GCN、BiLSTM与注意力机制相结合,可以进一步提高分类精度。

针对传统分类网络对恶意评论识别精度不高、识别效果不好的问题,本文提出了一种结合交叉注意力机制的双通道文本分类网络(Two-channel text classification network combined with cross-attention mechanism, CA2TC)。该网络包含词嵌入层、GCN层、与卷积结合的BiLSTM层、交叉注意力机制层、拼接层、全连接层以及softmax。GCN和BiLSTM从不同的角度提取恶意评论文本的上下文信息。BiLSTM用于从卷积层输出的提取了句子不同位置的n-gram特征并降维的文本向量中再次提取文本的上下文信息。本文还设计了独特的交叉注意力机制来对恶意信息进行聚焦。交叉注意力机制可以突出双通道提取的关键信息并进行有效融合。最后将经过精炼的特征拼接在一起,并经过全连接层由softmax进行分类。为验证所提方法的性能,本文采用微博收集的恶意评论数据集选取6个主流的文本分类模型作对比。实验表明,与一些主流的分类模型相比,提出的模型在分类精度上表现更好。

本文的其余部分组织如下。第1节介绍了LSTM和GCN及两者结合模型在文本分类领域的应用。第2节介绍本文所提出的结合交叉注意力机制的双通道文本分类网络CA2TC。第3节中讨论并分析实验结果。最后,在第4节中给出了一些结论和未来研究的可能路径。

1 相关工作

在深度学习方法中,LSTM模型用于处理序列数据。随着自然语言处理的发展,LSTM的应用范围也逐渐扩大。为了满足各种任务中出现的变长序列问题,许多学者对LSTM进行改进,提出了许多处理方法。目前,LSTM及其变体已经在各种任务中得到了充分的应用。Shi等[12]通过将LSTM记忆细胞中的神经元替换为震荡神经元的方式提出了CNO-LSTM模型。与NLP任务的主流基线模型相比,CNO-LSTM表现更优。Huang等[13]通过利用历史信息提出一种带补偿的长短期记忆方法LSTM-Com。Abdul等[14]提出了一种混合LSTM和CNN与GloVe来执行性别暴力的多分类。丁锋等[15]提出了BiLSTM-CRF模型来抽取消极情绪意见目标。

目前,GCN被认为能提取文档的空间特征而被应用在文本分类领域。学者们为进一步将GCN应用于文本分类中做了许多研究。2019年,Yao等[16]首次将图卷积神经网络应用在文本分类任务中,提出了Text GCN模型。Zhang等[17]将门控循环单元(Gated Recurrent Unit, GRU)与图卷积神经网络结合提出了TextING模型,该模型利用图门控卷积神经网络来对文本分类。Cui等[18]提出一种基于GCN的半监督短文本分类自训练方法来解决由于稀疏性和标记数据有限而使半监督短文本分类困难的问题。Li等[19]提出了一种多流图卷积网络(MS-GCN),用于通过代表字文档挖掘进行文本分类。尽管使用单一的模型在文本分类领域取得了进展,但仍然有提升空间。

研究证明,将LSTM与GCN结合进行文本分类可以提升模型效果。Yang等[20]基于CNN和LSTM的组合方法用于捕获局部特征以丰富特征信息来对GCN模型增强,并使用权重值调整增强的强度。Liu等[21]提出了TensorGCN模

型,该模型引入了长短记忆网络(LSTM)和单词间的句法依赖,用于表达单词间的语义与句法关系。Gao等[22]提出了一种改进的图卷积神经网络和递归神经网络的混合结构。Tang等[23]引入了BiLSTM网络、词性信息和依赖关系来解决GCN不能够解决上下文依赖和词汇多义的问题,提出一种改进的GCN网络(IGCN)。LSTM与GCN结合使模型能够关注到时序和空间两个维度的信息,但无法使两种信息得到有效的充分结合。

目前,注意力机制已成为一种选择重要信息以获得较优结果的有效方法。因此,将注意力机制与GCN、LSTM相结合可以获得更好的效果。Chen等[24]引入了双重注意力机制并提出了具有知识驱动注意力的深度短文本分类模型(Short Text Classification with Knowledge powered Attention,STCKA)。Liu等[25]将注意力机制、卷积与BiLSTM结合,提出AC-BiLSTM文本分类模型,该模型有效提高了文本分类性能。有研究[26]将注意力机制引入GCN模型提出图注意力网络(Graph Attention Network, GAT),为不同的节点赋予不同的权重以表示不同的节点拥有不同的影响力,并实现了引文分类任务,该网络增加了模型的解释性。Hu等[27]提出HGAT(Heterogeneous Graph Attention Networks)模型,该模型使用图注意力网络来对短文本进行分类,解决了语义稀疏、歧义性等在短文本中的问题。注意力机制能突出特征信息,但常见的注意力机制通常以分配不同的权重的方式对信息进行加权,然后再以相加或直接拼接的方式进行融合。这种方式很可能会因为权重的分配而丢掉一部分被认为不重要的信息造成信息丢失影响分类精度,仍有待改进。

可以看出,为了提高LSTM与GCN的性能,学者对其基本结构进行了大量的研究,使得LSTM与GCN在文本分类方面取得了突出的成绩。上述研究为CA2TC的基础,因此对其进行简要阐述。

2 结合交叉注意力机制的双通道文本分类网络

为了提高恶意评论文本分类模型的性能,本文尝试将与卷积结合的BiLSTM、GCN和交叉注意力机制集成在一起,设计一种结合交叉注意力机制的双通道文本分类网络(CA2TC)。CA2TC模型分别由词嵌入层、GCN层、与卷积结合的BiLSTM层、交叉注意力机制层、拼接层、全连接层与softmax组成,其总体框架如图1所示。

CA2TC模型处理文本的流程如下:

1)文本首先经过词嵌入层转换为向量表示。

2)GCN层从文本中提取空间特征,与卷积层结合的BiLSTM从文本中提取n-gram特征与上下文特征进行句子建模。

3)交叉注意力机制对得到的文本特征信息进行结合。CA2TC中的交叉注意力机制对GCN和与卷积层结合的BiLSTM得到的文本特征向量分别进行特征融合更新。

4)由交叉注意力机制处理的特征被连接在一起并输入到全连接层,最后送入softmax分类器。

2.1 词嵌入层

传统的词嵌入方法,如one-hot编码,有两个主要的问题:词序丢失和维度过大。而分布式的词嵌入方法,如word2vec,依赖于词库来对文本的词进行转换,不能很好地根据上下文来对文本的语义准确表达。因此,本文选用预训练模型BERT[28]作为文本词嵌入的方法,来更好地关注文本的上下文信息准确表达文本原本的意思。在本文中,每个词向量的维度为768。

2.2 GCN层

本文使用Text GCN来对文本构建一个包含词节点和文档节点的异构图。其中,词节点和词节点与词节点和文档节点之间分别基于频率-逆文档频率(TF-IDF)和点互信息(PMI)来计算边的权值:

$$A_{i,j}=\begin{cases}PMI(i,j), & i和j都是词,PMI(i,j)>0\\ TF\text{-}IDF_{i,j}, & i为文档,j为词\\ 1, & i=j\\ 0, & 其他\end{cases} \tag{1}$$

在Text GCN中,使用单位矩阵作为初始节点特征,然后将其送入GCN模型迭代训练。第i个GCN层的输出特征矩阵$L^{(i)}$为:

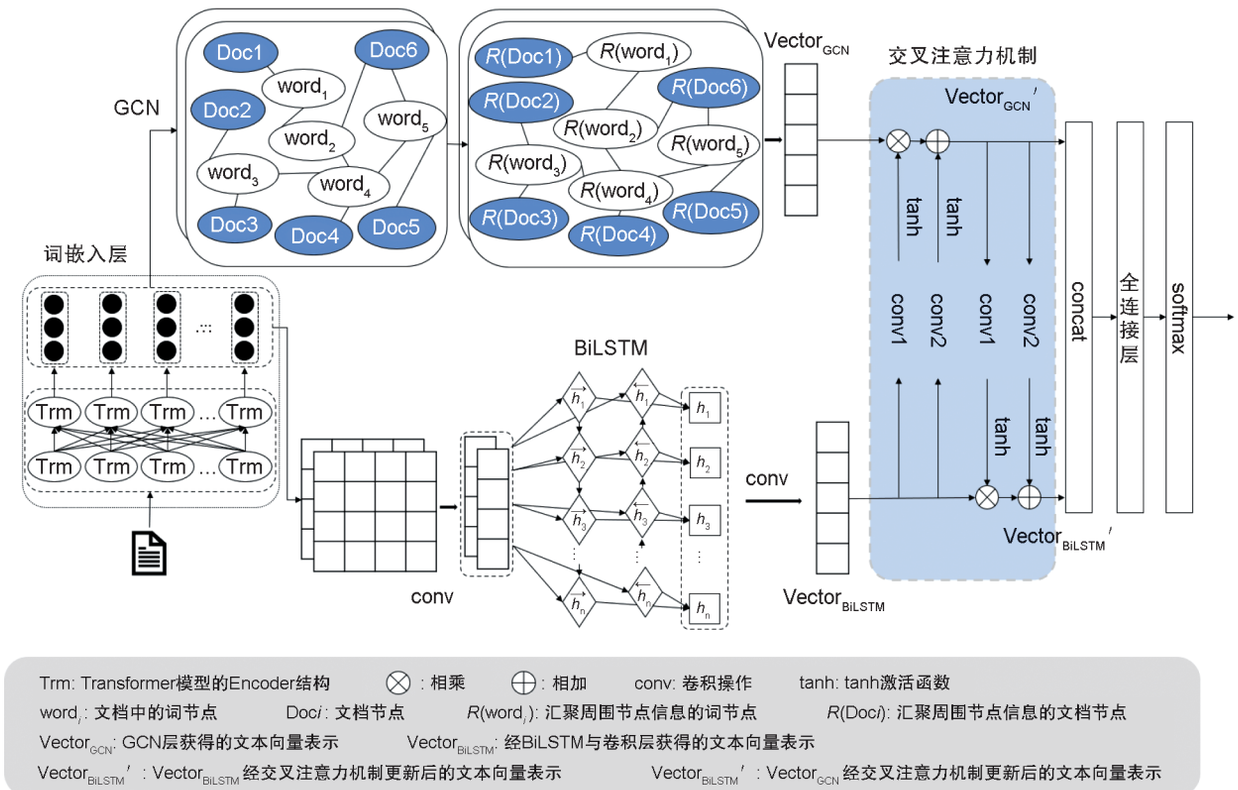

图1 CA2TC模型总体框架图

Fig. 1 Overall framework of the CA2TC model

$$L(i)=\rho(\tilde{A}L^{(i-1)}W^{(i)}), \tag{2}$$

其中ρ为激活函数ReLU,$\tilde{A}$是归一化邻接矩阵,$L^{(i-1)}$为第$i-1$层的输出,$W^{(i)}$为第i层的权值矩阵。

本文使用目前实验效果最好的两层GCN模型进行堆叠,则第二层GCN层的输出如公式(3)所示,输出大小为256维。

$$output_{\mathrm{GCN}}=\rho(\tilde{A}L^{(1)}W^{(1)})。 \tag{3}$$

2.3 与卷积结合的BiLSTM层

在本层中,使用单卷积层获取序列信息并降低数据的输入维度。卷积层的卷积运算在一维中进行。在卷积层,100个窗口大小为3的过滤器在文本表示上移动来提取特征。随着过滤器的移动,生成了许多捕捉句法和语义特征的序列。但由卷积获得的特征序列并不包含时序信息,使用BiLSTM来进一步从卷积层获得的特征序列中提取上下文信息。

为保证与GCN层的输出大小一致,文本经BiLSTM得到的向量会再通过一个大小为3×256,卷积核数量为256的二维卷积层将其转换为256维的向量。

2.4 交叉注意力机制层

注意力机制可以关注关键词的特征,减少非关键词对文本分类的影响。常用的注意力机制,如自注意力机制,通常是充当一个全连接层的作用,类似于一个softmax。本文提出了一种交叉注意力机制,不同于自注意力机制,交叉注意力机制是将独立的两通道特征进行精炼并融合,使两者的信息分布更合理,有足够的冗余信息,其结构如图2所示。

交叉注意力机制由两个输入$input_1$、$input_2$,两个相同的卷积层以及一个输出$output$组成。它的功能是通过$input_2$对$input_1$的一系列计算来更新$input_1$,从而达到融合两个输入的特征的目的。其计算方法如下:

$$output=\beta\cdot input_1+\alpha。 \tag{4}$$

α和β分别是$input_2$经过$conv1$、$conv2$两个卷积层提取关键信息并经tanh激活后得到,其中,$conv1$与$conv2$都是卷积核数目为256,大小为1×256的一维卷积。α和β计算公式如下:

$$\alpha=\tanh\left(conv1\left(input_2\right)\right), \tag{5}$$

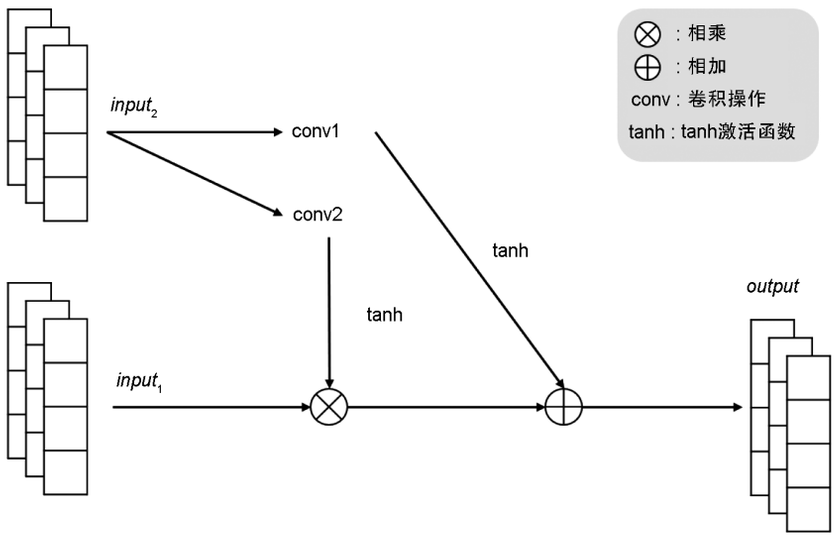

图2 交叉注意力机制结构图

Fig. 2 Structural diagram of the cross-attention mechanism

$$\beta=\tanh(conv2(input_2))。\tag{6}$$

交叉注意力机制可以更好地整合不同维度的特征信息。具体来说,虽然诸如自注意力机制等在几乎所有最先进的文本分类模型中都作为用来聚焦重点信息的必不可少的一部分,但其只能关注到本身的信息。相比之下,交叉注意力机制可以将本身的信息与文本经过另一种网络提取出的特征结合起来,可以得到整合不同维度特征的文本向量。因此,交叉归一化可以在保留自身信息的前提下整合不同维度的特征信息。

将交叉注意力机制层运用在CA2TC中的目的是要对GCN和与卷积结合的BiLSTM得到的向量相互进行更新:

$$\boldsymbol{output}'_{\mathrm{GCN}}=\tanh\left(conv2\left(output_{\mathrm{BiLSTM}}\right)\right)\boldsymbol{output}_{\mathrm{GCN}}+\tanh\left(conv1\left(output_{\mathrm{BiLSTM}}\right)\right),\tag{7}$$

$$output'_{\mathrm{BiLSTM}}=\tanh\left(conv2\left(\boldsymbol{output}'_{\mathrm{GCN}}\right)\right)output_{\mathrm{BiLSTM}}+\tanh\left(conv1\left(\boldsymbol{output}'_{\mathrm{GCN}}\right)\right)。\tag{8}$$

首先利用与卷积结合的BiLSTM层的输出$output_{\mathrm{BiLSTM}}$通过交叉注意力机制对GCN得到的向量$\boldsymbol{output}_{\mathrm{GCN}}$更新,更新后的向量$\boldsymbol{output}'_{\mathrm{GCN}}$是将BiLSTM模型提取的时序信息与GCN获取的空间特征信息整合后的结构,但更偏重于GCN提取的空间信息。然后再利用$\boldsymbol{output}'_{\mathrm{GCN}}$通过交叉注意力机制对与卷积结合的BiLSTM层的输出$output_{\mathrm{BiLSTM}}$进行更新,得到的$output'_{\mathrm{BiLSTM}}$同样整合了GCN模型提取的空间信息与BiLSTM提取的时序信息,但更侧重与卷积结合的BiLSTM提取的时序信息。这样做可以在不损失各自模型提取的信息的同时,整合自身缺乏的特征。

为对文本向量进行进一步的分类,CA2TC将经过交叉注意力机制的两输出进行拼接,得到综合的上下文表示*output*,然后经过全连接层再送入softmax进行分类。*output*计算公式如下:

$$output=concat\left(\boldsymbol{output}'_{\mathrm{GCN}},output'_{\mathrm{BiLSTM}}\right)。\tag{9}$$

CA2TC的主要贡献和独创性在于:1)将GCN与BiLSTM结合。GCN层可以根据构建异构图的方法来提取文档的空间特征,但由于图结构并没有时序性特征,GCN并不能获取到文本的时序特征。BiLSTM根据其前向和后向的隐藏层可以充分获取到文本的上下文信息,但BiLSTM可以处理序列数据,而不能捕获文档之间的空间信息。将两者结合可以弥补模型本身的不足,获取到不同的特征信息;2)卷积层从原始文本中提取低级语义特征,用于降维。对于文本分类,整个文档的向量表示通常是高维的向量。当使用BiLSTM捕获整个文档的语义时,BiLSTM的参数将显著增加。但是过多的网络参数会增加网络优化的难度。直接降低文本向量的维数会丢失大量信息,降低分类的准确性。卷积层被认为是善于提取输入的

健壮和抽象的特征。此外,卷积层还可以降低输入数据的维数。因此,一维卷积层可以提取文本向量的特征信息,同时降低文本向量的维数;3)使用交叉注意力机制来充分整合GCN与BiLSTM获取到的时间与空间信息。设计独特的交叉注意力机制来对由GCN和BiLSTM捕获的特征信息进行整合,通过互相更新的方式,将时序信息整合进GCN的输出中,也将空间信息整合进与卷积结合的BiLSTM的输出,既不损失相互的信息,又将缺少的特征整合,使文本的向量可以在充分获取上下文信息的同时关注到空间特征,使文本语义理解更加准确。因此,本文的方法有效地提高了分类精度。

3 实验

为了全面评估和分析CA2TC模型的性能,本文将CA2TC模型与6个基线模型进行了性能对比。

3.1 实验数据与环境

由于研究课题的特殊性,本文利用爬虫技术在微博社交平台上讨论比较热门的微博下爬取了微博评论2.1万条,其中恶意评论8 737条,非恶意评论14 785条,用作二分类。用于实验的训练集有18 804条,测试集有2 350条。

本文利用正则表达式等方法对数据集进行了相关预处理,包括去除对恶意评论分类无意义的词汇、相关字母转换为小写、表情转文字等操作,以使实验效果发挥最大作用。

本实验使用内存为32 GB带Ubuntu系统的服务器GeForce RTX 3090进行训练。训练代码使用Python 3.8.5,Torch版本为1.7.1。

3.2 实验参数

CA2TC模型部分超参数设置如表1所示。

3.3 评价指标

本文主要的评价指标有准确率(Accuracy)、精确率(Precision)、召回率(Recall)和$F1$值($F1$ Score)4项指标对于CA2TC模型性能进行评估。衡量模型性能的指标主要成分如表2所示。

表1 CA2TC模型超参数设置

Table 1 Super parameter setting of CA2TC model

模型	参数	值
CA2TC	学习率	2×10^{-5}
	batch_size	32
	epoch	12

表2 模型评价指标主要成分

Table 2 Main components of the evaluation index

真实标签	预测结果	
	正例(0)	负例(1)
正例(0)	真正例(TP)	假负例(FN)
负例(1)	假正例(FP)	真负例(TN)

表2中TP表示正样本能够被模型预测为正的数目;同样地,TN表示负样本被模型预测为负的数目。FP表示负样本被模型预测为正的数目,FN则为正样本被模型预测为负的数目。则有:

Accuracy,即准确率,表示模型预测正确的样本的占比;Precision,即精确率,表示的是模型预测为正的样本数中实际为正的样本在其中的占比;Recall,即召回率,表示的是模型正确预测为正的样本占样本集中正样本数的比例。Accuracy、Precision、Recall的计算公式分别如下所示:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN},\tag{10}$$

$$\text{Precision}=\frac{TP}{TP+FP},\tag{11}$$

$$\text{Recall}=\frac{TP}{TP+FN}。\tag{12}$$

但是在实验过程中,往往会出现Precision和Recall值不平衡的现象,因此引入$F1$值来均衡精确率和召回率对模型的影响。通过对精确率和召回率的加权平均来有效解决这个现象。一般地,定义F_β分数为:

$$F_\beta=\frac{(1+\beta^2)\text{Precision}\times\text{Recall}}{(\beta_2\times\text{Precision})+\text{Recall}},\tag{13}$$

则$F1$为:

$$F1=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}。\tag{14}$$

3.4 对比实验

为了验证CA2TC模型在恶意评论检测上的有效性,本文在相同实验环境中比较了6种深度神经网络模型,实验结果如表3所示。

1)BERT:基于BERT预训练模型得到文本的字向量表示,再经过全连接层与Softmax进行

分类。

2) BERT-CNN:文献[29]提出的BERT-CNN模型,使用BERT预训练模型得到文本的字向量表示,送入CNN网络提取特征,再经过Softmax进行分类。

3) RBR:基于文献[4]提出的RBR模型,使用BERT预训练模型得到文本的字向量表示,送入RCNN网络提取特征,再经过Softmax进行分类。

4) DeBERTa(Decoding-enhanced BERT with disentangled attention):基于文献[30]提出的DeBERTa模型,并使用其进行分类。

5) ALBERT(A Lite BERT):基于文献[31]提出的ALBERT模型,并使用其进行分类。

6) BiLSTM+Self-Attention:文献[32]使用skip-gram算法训练文本字向量再送入加入Self-Attention的BiLSTM网络提取特征,再经Softmax进行分类,为使对比实验与本文契合,在此使用BERT预训练模型训练文本字向量。

表3 对比实验结果

Table 3 Comparison of experimental results

模型	指标			
	Accuracy	Precision	Recall	*F*1
BERT	83.70%	82.86%	81.82%	82.27%
BERT-CNN	84.09%	83.02%	82.76%	82.89%
RBR	83.91%	82.91%	82.39%	82.63%
DeBERTa	85.53%	84.96%	83.68%	84.23%
ALBERT	84.85%	83.99%	83.28%	83.60%
BiLSTM+Self-Attention	85.02%	85.18%	82.29%	83.71%
CA2TC(本文)	**86.59%**	**85.94%**	**85.09%**	**85.47%**

从表3中可以看出,CA2TC模型在此恶意评论数据集上取得了比其他方法更好的结果。在上述7种方法中,CA2TC在准确率、精确率、召回率以及*F*1值上都优于其他6种模型。CA2TC模型相较于其他6种对比模型表现最好的DeBERTa模型在准确率、精确率、召回率和*F*1值分别提高1.06%、0.98%、1.41%和1.24%。实验表明,与现今比较流行的深度学习模型相比,CA2TC并不是简单的模型拼接,而是考虑了两种模型的优缺点,再有交叉注意力机制对两输入的特征融合,使得模型能够获取更多的特征信息,从而性能优于其他对比模型。

3.5 消融实验

本文设置了4个消融实验来验证CA2TC模型各个模块结合的合理性,实验结果如表4所示。

1) BiLSTM:文本经BERT预训练模型得到文本词向量,再经BiLSTM模型提取特征,最后经过全连接层送入softmax进行分类;

2) GCN:文本经BERT预训练模型得到文本词向量,再经GCN模型提取特征,最后经过全连接层送入softmax进行分类;

3) 去掉交叉注意力机制的CA2TC:文本经过BERT预训练模型得到文本词向量,再分别经过GCN和与卷积结合的BiLSTM模型,得到的两个输出直接拼接,之后经过全连接层送入softmax进行分类;

4) 换成自注意力机制的CA2TC:文本经过BERT预训练模型得到文本词向量,再分别经过GCN和与卷积结合的BiLSTM模型,得到的两个输出直接拼接,之后经过自注意力机制和全连接层送入softmax进行分类。

表4 消融实验结果

Table 4 Results of ablation experiment

模型	指标			
	Accuracy	Precision	Recall	*F*1
BERT+BiLSTM	83.62%	82.49%	82.32%	82.40%
BERT+GCN	73.11%	71.94%	68.18%	68.93%
去掉交叉注意力机制的CA2TC	86.09%	85.90%	83.91%	84.70%
换成自注意力机制的CA2TC	83.79%	82.84%	82.13%	82.49%
CA2TC(本文)	**86.59%**	**85.94%**	**85.09%**	**85.47%**

从表4可以看出,交叉注意力机制层、BiLSTM和GCN对CA2TC的性能有很大的影响。在上述架构中,CA2TC的效果最好。与去掉交叉注意力机制的CA2TC相比,CA2TC在*F*1值上提高了0.77%。结果表明,去掉交叉注意力机制,CA2TC的性能显著下降。在CA2TC中,交叉注意力机制可以对GCN和与卷积结合的BiLSTM得到的信息进行融合,充分考虑了时序信息和空间信息,而简单的拼接并不能让两种特征信息的作用完全表达。与换成自注意力机制的CA2TC相比,CA2TC在*F*1值上提高了2.98%。自注意力机制只是分配了两种信息的权重,相较于交叉注意力机制将两种信息融合,自注意力机制在对两向量分配权重时损失一部分相关信息,而导致结果低于去掉交叉注

意力机制的CA2TC。与BiLSTM和GCN相比，CA2TC相对提高了3.07%～16.54%。结果表明，去除交叉注意力机制层和BiLSTM或GCN后，CA2TC在去掉交叉注意力机制的基础上性能进一步下降。这说明去掉对重要信息的关注与融合后，单一的特征并不能使模型进行很好的分类。综上所述，特征信息提取和融合的方式会影响模型的性能，因此，本文将GCN和与卷积结合的BiLSTM结合以提取侧重不同维度的信息，再经交叉注意力机制进行特征融合以获得最优的性能。

4 结论

本文提出了一种结合交叉注意力机制的双通道文本分类网络CA2TC，该网络同时使用GCN和BiLSTM获得文本的时序和空间信息，利用提出的交叉注意力机制对两种特征进行精炼与融合，最后经拼接全连接层后送入softmax进行分类。在收集的微博恶意评论数据集上对提出的方法进行实验，结果表明，与6种当前流行的深度学习模型对比，提出的模型能够更准确地识别恶意评论，在4种评价指标上都优于其他模型。

未来的工作重点是关注机制的研究和网络架构的设计。未来的工作主要包括以下几个部分：1)利用其他注意力机制进一步完善本文的方法；2)考察注意力机制对方法性能的影响；3)设计新的网络架构；4)将本文的方法落地，能更好地应用到现实网络中。

参考文献：

[1] 杨金灵. 基于词向量集成与数据增强的恶意评论分类模型[J]. 科学技术创新, 2022(22): 76-81. DOI: 10.3969/j.issn.1673-1328.2022.22.020.
YANG J L. Toxic Comments Classification Model Based on Word Vector Ensemble and Data Augmentation[J]. *Sci Technol Innov*, 2022(22): 76-81. DOI: 10.3969/j.issn.1673-1328.2022.22.020.

[2] 吴浩, 潘善亮. 基于BERT-RCNN的中文违规评论识别研究[J]. 中文信息学报, 2022, **36**(1): 92-103. DOI: 10.3969/j.issn.1003-0077.2022.01.011.
WU H, PAN S L. Research on Recognition of Chinese Illegal Comments Based on BERT-RCNN[J]. *J Chin Inf Process*, 2022, **36**(1): 92-103. DOI: 10.3969/j.issn.1003-0077.2022.01.011.

[3] 周娅, 李赛. 基于分层欠采样和Bi-GRU的恶意行为检测模型[J]. 计算机工程与设计, 2022, **43**(2): 413-419. DOI: 10.16208/j.issn1000-7024.2022.02.016.
ZHOU Y, LI S. Malicious Behavior Detection Model Based on Hierarchical Undersampling and Bi-GRU[J]. *Comput Eng Des*, 2022, **43**(2): 413-419. DOI: 10.16208/j.issn1000-7024.2022.02.016.

[4] WATANABE A, SASANO R, TAKAMURA H, *et al*. Generating Personalized Snippets for Web Page Recommender Systems[C]//2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). Washington: IEEE, 2014, **2**: 218-225. DOI: 10.1109/WI-IAT.2014.101.

[5] ALMEIDA T A, SILVA T P, SANTOS I, *et al*. Text Normalization and Semantic Indexing to Enhance Instant Messaging and SMS Spam Filtering[J]. *Knowl Based Syst*, 2016, **108**(C): 25-32. DOI: 10.1016/j.knosys.2016.05.001.

[6] SCHMIDHUBER J. Deep Learning in Neural Networks: an Overview[J]. *Neural Netw*, 2015, **61**: 85-117. DOI: 10.1016/j.neunet.2014.09.003.

[7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. *Commun ACM*, 2017, **60**(6): 84-90. DOI: 10.1145/3065386.

[8] FUNAHASHI K I, NAKAMURA Y. Approximation of Dynamical Systems by Continuous Time Recurrent Neural Networks[J]. *Neural Netw*, 1993, **6**(6): 801-806. DOI: 10.1016/s0893-6080(05)80125-x.

[9] KIPF T N, WELLING M. Semi-supervised Classification with Graph Convolutional Networks[EB/OL]. arXiv Preprint: 1609.02907, 2016. https://arXiv.org/abs/1609.02907.

[10] HOCHREITER S, SCHMIDHUBER J. Long Short-term Memory[J]. *Neural Comput*, 1997, **9**(8): 1735-1780. DOI: 10.1162/neco.1997.9.8.1735.

[11] ALEX, GRAVES, . Framewise Phoneme Classification with Bidirectional LSTM and other Neural Network Architectures[J]. *Neural Netw*, 2005, **18**(5/6): 602-610. DOI: 10.1016/j.neunet.2005.06.042.

[12] SHI N B, CHEN Z H, CHEN L, *et al*. CNO-LSTM: a Chaotic Neural Oscillatory Long Short-term Memory Model for Text Classification[J]. *IEEE Access*, 2022, **10**: 129564-129579. DOI: 10.1109/ACCESS. 2022.3228600.

[13] HUANG W, LIU M Y, SHANG W Q, *et al*. LSTM with Compensation Method for Text Classification[J]. *Int J Wirel Mob Comput*, 2021, **20**(2): 159-167. DOI:

10.1504/ijwmc.2021.114139.

[14] ISMAIL A A, YUSOFF M. An Efficient Hybrid LSTM-CNN and CNN-LSTM with GloVe for Text Multi-class Sentiment Classification in Gender Violence[J]. *Int J Adv Comput Sci Appl*, 2022, **13**(9), DOI: 10.14569/ijacsa.2022.0130999.

[15] 丁锋, 孙晓. 基于注意力机制和BiLSTM-CRF的消极情绪意见目标抽取[J]. 计算机科学, 2022, **49**(2): 223-230. DOI: 10.11896/jsjkx.210100046.
DING F, SUN X. Negative-emotion Opinion Target Extraction Based on Attention and BiLSTM-CRF[J]. *Comput Sci*, 2022, **49**(2): 223-230. DOI: 10.11896/jsjkx.210100046.

[16] YAO L, MAO C S, LUO Y. Graph Convolutional Networks for Text Classification[J]. *Proc AAAI Conf Artif Intell*, 2019, **33**(1): 7370-7377. DOI: 10.1609/aaai.v33i01.33017370.

[17] ZHANG Y, YU X, CUI Z, *et al*. Every Document Owns Its Structure: Inductive Text Classification *via* Graph Neural Networks[EB/OL]. arXiv Preprint: 2004.13826, 2020. https://arXiv.org/abs/2004.13826.

[18] CUI H Y, WANG G K, LI Y X, *et al*. Self-training Method Based on GCN for Semi-supervised Short Text Classification[J]. *Inf Sci Int J*, 2022, **611**(C): 18-29. DOI: 10.1016/j.ins.2022.07.186.

[19] LI M, CHEN S Y, YANG W F, *et al*. Multi-stream Graph Convolutional Networks for Text Classification *via* Representative-word Document Mining[J]. *Int J Comp Intel Appl*, 2022, **21**(4): 2250028. DOI: 10.1142/s1469026822500286.

[20] YANG C L, GUO Y C, LI X W, *et al*. A Novel Method Using Local Feature to Enhance GCN for Text Classification[C]//2021 11th International Conference on Intelligent Control and Information Processing (ICICIP). IEEE, 2021: 59-65.DOI: 10.1109/ICICIP53388.2021. 9642171.

[21] LIU X E, YOU X X, ZHANG X, *et al*. Tensor Graph Convolutional Networks for Text Classification[J]. *Proc AAAI Conf Artif Intell*, 2020, **34**(5): 8409-8416. DOI: 10.1609/aaai.v34i05.6359.

[22] GAO L C, WANG J K, PI Z X, *et al*. A Hybrid GCN and RNN Structure Based on Attention Mechanism for Text Classification[J]. *J Phys: Conf Ser*, 2020, **1575**(1): 012130. DOI: 10.1088/1742-6596/1575/1/012130.

[23] TANG H L, MI Y, XUE F, *et al*. An Integration Model Based on Graph Convolutional Network for Text Classification[J]. *IEEE Access*, 2020, **8**: 148865-148876. DOI: 10.1109/ACCESS.2020.3015770.

[24] CHEN J D, HU Y Z, LIU J P, *et al*. Deep Short Text Classification with Knowledge Powered Attention[J]. *Proc AAAI Conf Artif Intell*, 2019, **33**(1): 6252-6259. DOI: 10.1609/aaai.v33i01.33016252.

[25] LIU G, GUO J B. Bidirectional LSTM with Attention Mechanism and Convolutional Layer for Text Classification[J]. *Neurocomputing*, 2019, **337**: 325-338. DOI: 10.1016/j.neucom.2019.01.078.

[26] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, *et al*. Graph Attention Networks[EB/OL]. arXiv Preprint: 1710.10903, 2017. https://arXiv.org/abs/1710.10903.

[27] HU L M, YANG T C, SHI C, *et al*. Heterogeneous Graph Attention Networks for Semi-supervised Short Text Classification[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 4821-4830. DOI: 10.18653/v1/d19-1488.

[28] DEVLIN J, CHANG M, LEE K, *et al*. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. arXiv Preprint: 1810.04805, 2018. https://arXiv.org/abs/1810.04805.

[29] DONG J C, HE F J, GUO Y C, *et al*. A Commodity Review Sentiment Analysis Based on BERT-CNN Model [C]//2020 5th International Conference on Computer and Communication Systems (ICCCS). IEEE, 2020: 143-147. DOI: 10.1109/ICCCS49078.2020.9118434.

[30] HE P, LIU X, GAO J, *et al*. DeBERTa: Decoding-enhanced BERT with Disentangled Attention[EB/OL]. arXiv Preprint: 2006.03654, 2020. https://arXiv.org/abs/2006.03654.

[31] LAN Z, CHEN M, GOODMAN S, *et al*. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations[EB/OL]. arXiv Preprint: 1909.11942, 2019. https://arXiv.org/abs/1909.11942.

[32] 吴小华, 陈莉, 魏甜甜, 等. 基于Self-Attention和Bi-LSTM的中文短文本情感分析[J]. 中文信息学报, 2019, **33**(6): 100-107. DOI: 10.3969/j. issn. 1003-0077.2019.06.015.
WU X H, CHEN L, WEI T T, *et al*. Sentiment Analysis of Chinese Short Text Based on Self-attention and Bi-LSTM[J]. *J Chin Inf Process*, 2019, **33**(6): 100-107. DOI: 10.3969/j.issn.1003-0077.2019.06.015.