Journal of Computer Applications
计算机应用 2023, 43(S1):49 - 55
ISSN 1001-9081
CODEN JYIIDU
2023-06-30
http://www.joca.cn

文章编号:1001-9081(2023)S1-0049-07
DOI:10.11772/j.issn.1001-9081.2022101606

基于图卷积神经网络的虚假新闻检测

倪铭远,邓宏涛,高　望*
(江汉大学 人工智能学院,武汉 430056)
(*通信作者电子邮箱gaow@jhun.edu.cn)

摘　要:当前检测虚假新闻的方法往往依赖于人工设计的特征,并且需要提供大量如用户信息、传播路径等不宜获取的隐私信息,使得模型泛化性较差。针对上述问题,提出一种基于图卷积网络(GCN)和预训练ALBERT(A-Lite-Bidirectional Encoder Representations from Transformers)构建的新闻检测模型GCN-ALBERT。首先,利用GCN捕获文本全局信息,提取新闻文本的全局语义信息;其次,利用自注意力机制融合ALBERT生成的局部信息与全局信息;最后,建立包含局部信息和全局信息的分类表示,从而实现虚假新闻检测。实验结果表明,所提模型在两个真实的英文数据集上与预训练语言模型BERT(Bidirectional Encoder Representations from Transformers)相比,宏F1值分别提高了3.0%和4.2%。所提模型能够有效融合新闻文本的全局信息和局部信息,准确率更高。

关键词:虚假新闻检测;图卷积网络;ALBERT;自注意力机制;预训练模型
中图分类号:TP391.1　　**文献标志码:**A

Fake news detection based on graph convolutional neural network

NI Mingyuan, DENG Hongtao, GAO Wang*
(*College of Artificial Intelligence, Jianghan University, Wuhan Hubei* 430056, *China*)

Abstract: Current methods for fake news detection often rely on artificial designed features, and need to provide a large amount of privacy information, such as user information and communication paths, making the model less generalized. Aiming at the above problems, a new news detection model based on Graph Convolution Network (GCN) and pre-trained ALBERT (A-Lite-Bidirectional Encoder Representations from Transformers) called GCN-ALBERT was proposed. Firstly, GCN was used to capture global information of text and extract the global semantic features of news text. Secondly, self-attention mechanism was used to fuse the local information learnt by ALBERT with the global semantic features. Finally, a classified representation containing local information and global information was established to detect fake news. Experimental results show that the macro F1 value of the proposed model is improved by 3.0% and 4.2% respectively compared with the pre-trained language model BERT (Bidirectional Encoder Representations from Transformers) on two real English datasets. The proposed model can effectively integrate the global information and local information of news text, and has the higher accuracy.

Key words: fake news detection; Graph Convolutional Network (GCN); ALBERT (A-Lite-Bidirectional Encoder Representations from Transformers); self-attention mechanism; pre-trained model

0　引言

信息时代下社交媒体在人们的日常生活中不可或缺,用户可以在社交媒体上表达自我、浏览新闻、转发信息并与他人互动[1]。信息可通过社交网络进一步传——对不同事件的看法和情绪可以通过用户的参与和交互反映。社交网络的便利性和低成本带来了更多元的信息传播渠道,同时也带来了负面的影响,即虚假新闻等误导信息的传播。

虚假新闻指一种在网络媒体上广泛传播虚假信息的新闻,其文字内容与事实不符,用于欺骗读者或造谣生事。这些新闻的传播会使公众接收大量的错误信息,对公众造成重大误导并为某些不法群体带来经济利益[2]。社交媒体平台往往通过用户或少量维护人员向平台主动举报的方式,获得疑似虚假新闻的信息。然后,这些平台再通过人工审核的方式判定被举报的信息是否为虚假新闻。以上方法可以在一定程度上遏制虚假新闻的传播,但这种方法过度依赖人工审查,对审查人员的专业知识有较高要求,并且虚假新闻可能在人工审查阶段就已经广泛传播。

目前已有很多学者使用数据挖掘和机器学习等技术检测虚假新闻。典型的方法是根据新闻的文本内容提取特征,例如*n*-gram和词袋特征[3],并利用监督学习方法(如随机森林和支持向量机(Support Vector Machine, SVM))将虚假新闻检测任务转变为二分类问题[4]。此外,还有部分学者利用多模态上下文信息识别虚假新闻,例如用户配置文件、转发频率和传

收稿日期:2022-10-26;**修回日期:**2023-01-09;**录用日期:**2023-01-11。
基金项目:武汉市教育局市属高校产学研项目(CXY202208);江汉大学学科特色专项项目(2022XKZK10)。
作者简介:倪铭远(1995—),男,湖北武汉人,硕士研究生,主要研究方向:自然语言处理、谣言检测;　邓宏涛(1972—),男,湖北武汉人,教授,硕士,主要研究方向:大数据分析与挖掘;　高望(1983—),男,湖北武汉人,讲师,博士,CCF会员,主要研究方向:自然语言处理、信息检索。

播路径等。

然而，现有基于内容的方法往往使用新闻标题等短文本，存在严重的数据稀疏问题[5]，并且文本特征也需要研究人员专门设计。此外，一些先进的模型则依赖于一系列社交关联信息，如转发数、分享数、评论数等[6]。一方面，由于商业信息壁垒和用户隐私保护问题，各大社交媒体公司对用户数据严格保密，而虚假新闻的编造者往往使用新注册的虚假账号，缺乏有效的用户信息，因此获取新闻发布者的用户信息和转发的传播结构成本较高[7]。另一方面，虚假新闻的转发路径需要在该新闻传播一段时间后才能够被获取，在这个过程中，该新闻可能已经被大量用户阅读并传播；并且社交媒体上的大多数用户倾向于简单的转发，很少留下评论，许多用户甚至选择隐藏或删除社交互动记录。因此，若能够仅使用文本信息对新闻的真实性进行有效鉴别，在虚假新闻传播的早期阶段进行处理，对于防范虚假新闻带来的负面影响将起到极大的作用。

近年来，深度学习模型通过在学习特征表示中嵌入语义和句法信息的方式表现出极佳的性能。得益于深度神经网络强大的特征学习能力，许多研究者关注于如何利用深度神经网络从新闻内容中自动学习更多有用的特征，而非由人工刻意选取[8]。然而，这类方法在学习文本的远距离语义信息方面受到限制。自注意力机制的使用有助于缓解这个问题，但问题仍然存在。例如，一则真实新闻文本如下所示。

“You can’t give a child an aspirin in school without permission. You can’t use any kind of medication, but we can secretly take the child off and have an abortion.”

这句话含蓄地表达了对事件的讽刺。如果没有将文本与这个含义更明确地联系起来，分类器可能会低估这个包含讽刺意味的观点，反而因为其中包含有大量意图误导公众、吸引注意力的虚假新闻特征[9]，如“can’t”“secretly”“take the child off”等带有强烈煽动性和情绪性的词语，使得检测模型可能会赋予这些词语较高的权重，从而将它们错误地判断为虚假信息。

近期，不少研究者提出了一些新方法综合考量词语和文本之间的全局信息，最具代表性的工作是图卷积网络（Graph Convolutional Network, GCN）[10]。这些方法将文本中的词语构建为词汇图，图中每个节点表示通过多次聚合更新操作将包含所有邻居节点的特征，从而能学习到文本的全局信息。但是只考虑全局语义信息的GCN难以抽取局部语义特征（如词序等），而这些局部语义特征对于理解语句的语义也非常重要[11]。

受GCN和预训练语言模型中自注意力机制的启发，本文结合两种机制的优势，提出了GCN-ALBERT（A-Lite-Bidirectional Encoder Representations from Transformers）模型。GCN-ALBERT模型首先基于词的共现信息在词汇图上构造一个图卷积网络，对语句的全局信息进行编码；然后，将图向量和词向量连接在一起，并输入ALBERT[12]的自注意力编码器中，词向量和图向量在模型训练迭代的同时，通过自注意力机制相互促进；最后，GCN-ALBERT利用该机制融合文本的局部信息和全局信息，从而更好地识别虚假新闻文本。与机器学习或神经网络等传统方法相比，GCN-ALBERT仅依赖文本信息，不需要专门手工提取复杂特征，也不需要收集用户信息和传播路径等不易获取的信息数据，适合在虚假新闻上传到网络的早期阶段根据文本内容判断新闻的真实性。

1 相关工作

1.1 自注意力机制

注意力机制，特别是Vaswani等[13]提出的自注意力机制，极大地提高了文本分类任务的性能。注意力机制的特点是赋予重点程度不同的信息不同的权重，重要信息会被分配较大的权重，而非重要信息则被分配较小的权重。通过自注意力机制获得的词语表示，可以通过融合语句中其他词的表征，将该词与其他词之间的关系结合起来[14]。例如，骆文莉等[15]利用不同扩张率的空洞卷积设计的多层次特征融合模块和注意力机制有效地实现跨通道间的信息交互，提取更能代表文本的特征，提升文本检测的准确率；Ma等[16]采用循环神经网络（Recurrent Neural Network, RNN）顺序处理谣言传播序列以获得虚假新闻的传播特征。为了更进一步提高性能，许多研究者通过Transformer模型抽取更多的远距离依赖关系特征。例如，BERT（Bidirectional Encoder Representations from Transformers）是文本分类中最成功的预训练语言模型之一[17]。该模型通过多层多头自注意力机制和位置嵌入，使编码器通过对文本不同部分进行多头注意力计算，来提取输入文本具有上下文信息的语义表示。

上述方法仅关注局部上下文信息，对于包含上下文窗口的词语，会生成一个包含上下文权重的语义表示，然而，整个语句的全局语义信息却被忽视。

1.2 图卷积网络

图神经网络（Graph Neural Network, GNN）将词语之间的全局关系表示为一个图，其中词语是节点，词语间的关系用边表示。利用这种图结构，GNN更善于捕捉语句中词语的全局信息[18]。因此，很多学者提出各种基于GNN的模型，并成功应用于文本分类任务。其中，Shu等[19]考虑了新闻发布关系、传播关系以及用户关系构建异质信息网络，使用矩阵分解的方式获得各个新闻节点的嵌入表示，从而进行虚假新闻检测；Han等[20]利用GNN获取Twitter上假新闻的远程监督关系抽取，使它能够处理非结构化文本输入的关系推理。

基于GNN的研究，Kipf等[21]创造性地提出了基于图谱方法的图卷积网络（GCN）。GCN首先根据给定的关系图构建一个对称的邻接矩阵，然后通过卷积神经网络（Convolutional Neural Network, CNN）对所构建的图进行特征提取和邻域节点的特征聚合，从而得到文本的特征表示。王昕岩等[22]通过将事件之间联系的紧密程度描述为连边权重，提出一种基于加权图卷积网络（Weighted-Graph Convolutional Network, W-GCN）的模型。Yao等[10]提出用于文本分类的TextGCN，可以视为GCN的一个特例。TextGCN与GCN的区别是该模型基于一个异构图，其中的词语和文档都是图中的节点。Mehta等[23]设计了一种用于迭代表示学习的推理框架，应用不同的推理操作学习图中的隐藏关系，不断改进社会背景表征，从而实现虚假新闻检测。

GCN及其变体的优势是利用卷积抽取图中的全局语义特征，但是它们处理图结构中的不确定性能力有限，并且没有考虑到词语之间的顺序等局部信息。当词序和其他局部信息在任务中起到关键作用时，GCN的表现往往不够理想。之前的研究已经证明GNN在社交媒体事件检测方面的有效性，但未

充分考虑文本上下文信息对检测性能的影响[24]。因此,本文将GCN与捕获局部信息的模型相结合,有利于提升模型的性能。

1.3 融合方法

很多方法通过融合文本的各种特征来提高模型的性能。例如,Wang等[25]提出了一个事件对抗神经网络模型,利用多模态信息识别虚假信息。该模型首先利用卷积神经网络融合输入的文本和视觉特征,然后通过对抗神经网络去除事件特定的特征,并保留可转移的特征。Shu等[6]提出了一种虚假新闻检测框架dEFEND(Explainable FakE News Detection),该框架融合重要用户的评论以及新闻内容和评论之间的相关性来学习语义特征,并利用注意力机制增强虚假新闻检测模型的可解释性。Ruchansky等[26]提出一种混合深度模型检测虚假新闻。该模型通过融合回复文本和用户个人信息,并根据用户的社交互动判断新闻的真实性。黄皓等[27]在其基础上通过添加EXTRACT模块和门控循环单元(Gate Recurrent Unit, GRU)模块获取传播路径特征和文本反馈特征从而在准确性和时效之间获得更好的平衡。Yuan等[28]通过将用户信息、转发信息和源推文之间的全局关系建模为一个异构图,将全局结构信息与局部语义信息联合编码用于谣言检测。Zhou等[29]从假新闻、欺骗类型和点击诱饵之间的关系考察新闻内容,并依靠社会心理学和法医心理学的成熟理论实现仅针对于新闻内容的可解释虚假新闻检测。

本文认为语句中局部和全局语义信息之间的融合同样重要,并且可以使下游任务受益。例如,Levy等[30]指出在大规模语料库上训练的词向量,例如Word2Vec(Word to Vector)、GloVe(Global Vectors)、FastText可以学习到语句中词语之间的远距离依赖关系。然而,在词向量的训练过程中,词语和其对应的上下文通常被限制在一个较小的文本窗口(5个词)中,从而难以学习到更长的全局语义关系。与之相比,本文所提出的GCN-ALBERT模型将ALBERT学习到的局部信息融合到GCN抽取的全局特征中,有利于提高模型的语义理解能力,并首次将GCN和ALBERT融合并应用于虚假新闻检测的研究工作。

2 GCN-ALBERT

文本的全局语义信息可以使用多种方式获取。GCN-ALBERT使用的方式是利用语句中词语之间的关系。具体来说,GCN-ALBERT利用文档中词语间的共现关系来构建词汇图。该模型首先将输入语句转换为向量表示,并选择词汇图的相关部分表示全局信息,而输入文本的局部信息可由预训练语言模型ALBERT学习得到。如图1所示,本文使用多层注意力机制融合输入文本的局部信息和全局信息。

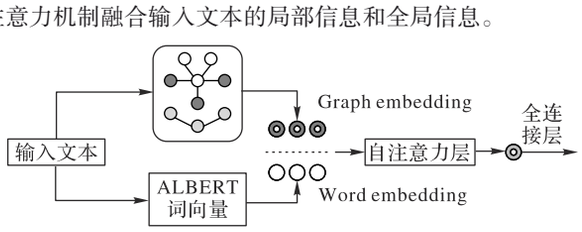

图1 GCN-ALBERT的结构

2.1 词汇图构建

对于语料中的所有词汇,本文使用标准化的点对互信息(Normalized Pointwise Mutual Information, NPMI)构建一个词汇图,该值的取值范围为[-1,1],计算方式如式(1)所示:

$$f_{\mathrm{NPMI}}(\alpha,\beta)=-\frac{1}{\ln p(\alpha,\beta)}\ln\frac{p(\alpha,\beta)}{p(\alpha)p(\beta)} \tag{1}$$

$$p(\alpha,\beta)=\#N(\alpha,\beta)/\#N \tag{2}$$

$$p(\alpha)=\#N(\alpha)/\#N \tag{3}$$

其中:α和β分别表示两个不同的词语;$\#N(*)$表示含有一个词语(例如α)或一对词语(例如α和β)的滑动窗口的数量;$\#N$表示滑动窗口的总数。

为了学习词语的长距离语义依赖关系,本文将滑动窗口的大小设置为整个语句的长度。当NPMI值大于0时,表示两个词语之间有较强的语义相关性;当NPMI值小于0时,则表示两个词语的语义相关性较差。在GCN-ALBERT模型中,当两个词语的NPMI值大于指定阈值时,则会在它们之间建立一条连边。通过实验结果可以发现,当阈值设置为0.2时,模型的性能较好。

2.2 词汇图卷积网络

传统的图神经网络通过聚合、更新、循环迭代这3个过程进行节点表示学习。然而,在聚合过程中,使用节点平均的方式进行聚合不太合理。这是因为不同节点的重要性程度不一样,节点的"度"不同,因此需要引入一种类似于"注意力机制"的方式来对各节点传递信息的权重进行计算。为了克服图神经网络的不足之处,图卷积网络增加了针对每个节点"度"的归一化操作。

标准的图卷积网络是一个多层(通常为两层)的神经网络,并且直接针对图进行卷积计算。然后,根据节点的邻域属性训练出节点的向量表示。形式上可表述为,假设词汇图用$G=(V,E)$表示,其中V和E分别是节点的集合和边的集合。对于GCN的单个卷积层,更新后n个节点的向量表示$\boldsymbol{H}\in\mathbb{R}^{n\times h}$可通过式(4)计算:

$$\boldsymbol{H}=\tilde{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W} \tag{4}$$

其中:$\boldsymbol{X}\in\mathbb{R}^{n\times m}$表示具有$n$个节点和$m$维特征的输入矩阵;$\boldsymbol{W}\in\mathbb{R}^{m\times h}$是一个权重矩阵。归一化对称邻接矩阵$\tilde{\boldsymbol{A}}$通过式(5)表示:

$$\tilde{\boldsymbol{A}}=\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}} \tag{5}$$

在图卷积模型中,为了有效缓解数值不稳定和梯度爆炸、梯度消失等问题,对$\boldsymbol{A}$进行归一化操作。通过式(6)计算:

$$\boldsymbol{D}_{ii}=\sum_{j}\boldsymbol{A}_{ij} \tag{6}$$

在不同任务中,GCN中节点的形式各不相同。例如针对文本分类任务,图中节点是需要分类的文档表示。这要求训练集、验证集和测试集中所有文档都要出现在图中,对于新的检测文档才不会出现节点缺失的问题。如此一来,在许多预测任务中,GCN的使用受到诸多限制。原因是在这些任务中,测试集中的文档数据在训练时无法使用。

针对虚假新闻检测任务,本文的目标是通过对相关词语进行卷积操作获取全局信息,而不是语料中的文档。因此,GCN-ALBERT所构建的词汇图卷积网络是基于文档中的词语。对于某个文档,令文档由其词语组成的行向量$\boldsymbol{d}$表示,卷积运算如式(7)所示:

$$\boldsymbol{h}=(\tilde{\boldsymbol{A}}\boldsymbol{d}^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{d}\tilde{\boldsymbol{A}}\boldsymbol{W} \tag{7}$$

其中:$\tilde{\boldsymbol{A}}^{\mathrm{T}}=\tilde{\boldsymbol{A}}$代表词汇图;$\boldsymbol{d}\tilde{\boldsymbol{A}}$提取与输入句子$\boldsymbol{d}$相关的词图部分;$\boldsymbol{W}\in\mathbb{R}^{v\times h}$表示单个文档隐藏状态向量的权重,$v$表示词汇

表的大小。对于mini-batch中的多个文档$\boldsymbol{D}$,式(8)中的卷积计算可以表示为:

$$\boldsymbol{H}=\boldsymbol{D}\tilde{\boldsymbol{A}}\boldsymbol{W} \tag{8}$$

两层带有激活函数ReLU的词汇图卷积网络可表示为:

$$V_{\mathrm{GCN}}=\mathrm{ReLU}(\boldsymbol{D}_{s,v}\tilde{\boldsymbol{A}}_{v,v}\boldsymbol{W}_{v,h})\boldsymbol{W}_{h,l} \tag{9}$$

其中:s表示mini-batch中新闻文档的个数;v表示词汇表的大小;h表示隐藏层的维度;l表示语句向量的维度或类别的个数;$\tilde{\boldsymbol{A}}_{v,v}$代表$\tilde{\boldsymbol{A}}$的矩阵维度为$v\times v$。$\boldsymbol{D}_{s,v}$中的每一行可以是词袋特征或是ALBERT词向量,表示对应新闻文档的特征表示。如式(9)所示,GCN-ALBERT通过$\boldsymbol{D}_{s,v}\tilde{\boldsymbol{A}}_{v,v}$获得词汇图中与输入新闻文档相关的部分。通过两层卷积计算,该模型将输入语句的词语和词汇表中相关词语进行结合。

2.3 GCN与ALBERT融合方法

和BERT模型类似,预训练语言模型ALBERT是基于Transformer编码器结构构建的,但ALBERT模型是BERT模型的轻量版。在不影响模型性能的前提下,ALBERT在因式分解、层间参数共享、段落间连贯性这3个方面进行了改进。

1)对嵌入参数进行因式分解。ALBERT将词嵌入层映射到低维空间进行降维,再映射到隐藏层。

2)ALBERT采用跨层参数共享的策略,避免参数量随着网络深度的增加而增多。

3)ALBERT使用语序预测任务(Sentence-Order Prediction, SOP)代替BERT中效果较差的下一句预测任务(Next Sentence Prediction, NSP)作为训练任务,解决了NSP造成的句间连贯性损失问题。

综合以上3个特点,使得ALBERT模型参数量仅为BERT的1/18,而训练速度却是BERT的1.7倍[12]。由此可见,ALBERT在显著减少参数量的同时仍保持了较高的性能。

将ALBERT应用于虚假新闻检测时,模型结构包括三个主要部分:第一部分是包含词语位置信息的词语嵌入模块;第二部分是采用多层多头自注意叠加的Transformer模块;第三部分是使用语句特征表示进行分类的全连接层。

其中,注意力的值是通过对键$\boldsymbol{K}$、值$\boldsymbol{V}$和查询$\boldsymbol{Q}$进行如式(10)计算实现的:

$$V_{\mathrm{Attention}}=(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\boldsymbol{K}}}}\right)\boldsymbol{V} \tag{10}$$

其中:分母是一个比例因子,作用是平衡注意力值的大小;$d_{\boldsymbol{K}}$表示键$\boldsymbol{K}$、查询$\boldsymbol{Q}$向量表示的维度。利用自注意力机制,模型可以计算出每个词语的加权向量表示,对上下文信息进行编码。

GCN-ALBERT并不是仅将新闻语句的词向量输入到ALBERT编码层中,而是将式(9)中得到的图向量和词向量序列同时作为ALBERT编码层的输入。如此一来,GCN-ALBERT不仅可以编码语句的位置信息,还通过GCN获取相关全局背景知识。将由式(10)得到的局部信息嵌入和全局信息嵌入输入到自注意力编码层中,通过12层自注意力编码器相互交互使得两者相互融和。GCN用于融合的语句特征向量$\boldsymbol{G}_{\mathrm{output}}$可通过式(11)表示为:

$$\boldsymbol{G}_{\mathrm{output}}=\mathrm{ReLU}(\boldsymbol{D}_{s,c,v}\hat{\boldsymbol{A}}_{v,v}\boldsymbol{W}_{v,h})\boldsymbol{W}_{h,t} \tag{11}$$

其中:c表示词向量的维度;t表示图向量的维度。在图卷积网络中原本用于分类的$\boldsymbol{G}_{\mathrm{output}}$转变为ALBERT编码层输入的一部分,含有语句的全局信息。最终输入到ALBERT编码层的向量$\boldsymbol{F}_{\mathrm{embedding}}$如式(12)所示:

$$\boldsymbol{F}_{\mathrm{embedding}}=\mathrm{Concat}(\boldsymbol{G}_{\mathrm{output}},\boldsymbol{F}_{\mathrm{ALBERT}}) \tag{12}$$

其中:Concat()表示级联操作;$\boldsymbol{F}_{\mathrm{ALBERT}}$表示ALBERT预训练的词向量。最终,GCN-ALBERT利用最上层的全连接网络对新闻本文内容进行分类。本文使用如式(13)所示交叉熵损失函数对模型进行训练优化:

$$L=-\frac{1}{N}\sum_{i}\left[y_i\cdot\mathrm{lb}(p_i)+(1-y_i)\cdot\mathrm{lb}(1-p_i)\right] \tag{13}$$

其中:y_i表示第i个样本的标签;p_i表示第i个样本预测为虚假新闻的概率。

3 实验与结果分析

为了评估GCN-ALBERT模型的性能,本文在两个真实世界的虚假新闻数据集上进行了大量的实验;并且,本文将GCN-ALBERT与基线模型进行了比较,以验证本文方法是否能够有效利用局部信息和全局信息获得更优的虚假新闻检测效果。

3.1 基线方法

除GCN-ALBERT模型外,将基于机器学习的虚假新闻检测模型、基于神经网络的虚假新闻检测模型以及基于预训练语言模型三类在各自类别具有优异性能的虚假新闻检测模型作为基线方法进行对比。本文均使用原始论文中的默认参数设置。

1)基于机器学习的虚假新闻检测模型。

①基于SVM的模型PTK(Propagation Tree Kernel)[31]。一种基于图内核的SVM分类器,通过评估传播树结构之间的相似性来捕获高阶模式用于区分虚假信息。

②基于随机森林的模型PES(Periodic External Shocks)[32]。通过重复进行两倍交叉验证从特征集中依次减少特征得出最重要的特征用于识别虚假新闻。

2)基于神经网络的虚假新闻检测模型。

①基于卷积神经网络的TCNN-URG(Two-level Convolutional Neural Network, User Response Generator)[33]。通过一个用户响应生成器捕获用户响应信息,一个两级CNN捕获语义信息,从而提高模型的检测能力。

②基于长短期记忆(Long Short-Term Memory, LSTM)网络的CSI(Capture-Score-Integrate)[26]。一种利用LSTM捕获用户文本和时间特征,学习用户信息表示,从而实现更准确的虚假新闻检测的混合模型。

3)基于预训练语言模型。

①BERT[17]。本文使用bert-base-uncased(https://huggingface.co/bert-base-uncased)版本的预训练语言模型BERT检测虚假新闻。

②RoBERTa[34]。本文使用RoBERTa-base(https://huggingface.co/roberta-base)版本的预训练语言模型RoBERTa检测虚假新闻。

3.2 数据集和预处理

本文的实验在两个虚假新闻检测数据集上进行,数据集的简要介绍如下。

1)MediaEval2015[35]。该数据集用于检测社交媒体上的多模态虚假内容。在MediaEval2015数据集中,训练集包含与17个事件相关的9 000条虚假信息和6 000条真实信息,而测

试集包含与35个事件相关的2 000条新闻数据。数据集原始信息中包含文本内容、相关的图片视频内容和社交信息。在本文实验中,仅使用MediaEval2015数据集中的文本内容。

2)Twitter[31]。该数据集用于检测Twitter上的虚假信息。Twitter数据集中不仅包含文本内容,还含有推文的传播路径等相关信息。本文仅使用其中的文本内容进行实验。

在数据预处理阶段,本文仅使用数据集当中的"虚假"和"真实"两类文本数据,并删除了数据集中URL字符串和Emoji表情等内容。在两个数据集中,本文删除了含有非英文语言的相关推文,仅保留英文文本内容。此外,本文使用NLTK(Natural Language ToolKit)的TweetTokenizer对数据进行词干还原,并全部转换为小写字母。经过预处理后数据集的统计结果如表1所示。

表1 数据集统计信息表

数据集	新闻总数	真实数	虚假数
MediaEval2015	11 850	6 841	5 009
Twitter	1 154	579	575

3.3 评估指标

本文使用加权平均F1值和宏F1值这两个常用指标评估模型的性能,计算方式如式(14)~(15)所示:

$$V_{\text{F1\_w\_avg}}=\sum_{i=1}^{C}F_{1\,ci}\times W_{ci} \tag{14}$$

$$V_{\text{F1\_macro}}=\frac{1}{C}\sum_{i=1}^{C}F_{1\,ci} \tag{15}$$

其中:C表示类别的数量;$F_{1\,ci}$表示第i类数据的F1值,该值为精确率和召回率的调和平均数,$F_1=2\times\frac{precision\times recall}{precision+recall}$,其中$precision$表示精确率,$recall$代表召回率;$W_{ci}$表示属于第$i$类数据的数量和数据集总数量的比值。

3.4 参数设置

对于批大小和学习率设置,本文在两个数据集上使用不同的超参数组合对GCN-ALBERT进行训练。实验结果如表2所示。

表2 两个数据集上的实验结果

批大小	学习率	MediaEval2015		Twitter2015	
		加权平均F1值/%	宏F1值/%	加权平均F1值/%	宏F1值/%
8	1E-5	88.46	88.07	78.95	77.13
	2E-5	80.33	79.87	38.67	38.19
	3E-5	87.04	86.70	31.40	31.33
	4E-5	87.94	87.59	39.59	38.92
	5E-5	88.64	88.34	31.76	30.75
16	1E-5	88.68	88.41	**92.07**	**91.93**
	2E-5	47.38	42.43	36.97	35.87
	3E-5	84.36	83.98	69.01	68.90
	4E-5	87.33	87.02	43.51	42.93
	5E-5	85.32	84.98	88.59	88.45
32	1E-5	88.81	88.50	87.56	86.85
	2E-5	89.03	88.70	42.78	41.88
	3E-5	64.25	61.63	63.80	63.58
	4E-5	**89.43**	**89.17**	42.57	41.85
	5E-5	81.02	80.73	40.82	39.93

本文使用ALBERT Tokenizer对输入新闻文本进行切分,因此GCN词汇表是ALBERT词汇表的子集。在计算NPMI的过程中,整个输入语句作为词汇图构建的窗口,以获得更长的语义依赖关系。对于两个数据集,NPMI的阈值都设置为0.2,以过滤词语之间无意义的关系。本文将图卷积神经网络输出图向量维度设置为128。对ALBERT,本文使用albert-base-v2版本(https://huggingface.co/albert-base-v2),并将最大输入序列长度设置为200。dropout概率设置为0.2,以防止模型过拟合。

从表2中可以看出,当批大小设置为32,学习率设置为4E-5时,GCN-ALBERT在MediaEval2015数据集上取得最优的性能。在Twitter数据集上,当批大小设置为16,学习率设置为1E-5时,该模型性能最佳。此外,在Twitter数据集上的学习率不宜设置过大,这可能是因为该数据集是由推文这种短文本组成的,推文中的文本存在大量网络用语,口语化、不规范、数据稀疏问题严重[36]。因此,较大的学习率使得模型误差波动较大,难以快速收敛。

在实验中,本文使用交叉熵作为GCN-ALBERT的损失函数,使用Adam作为模型的训练优化器。对于标签分布不均匀的情况,本文使用scikit_learn的comput_class_weight函数计算模型的加权损失。具体来说,类别c的权重W_c可通过式(16)计算:

$$W_c=N_d/(N_l*N_c) \tag{16}$$

其中:N_d是数据集中新闻的总数;N_l表示类别数;N_c表示属于类c新闻的总数。

3.5 实验结果及分析

GCN-ALBERT和基线模型在两个数据集上的实验结果如表3所示。

表3 MediaEval2015和Twitter数据集上实验结果 单位:%

模型类别	模型	MediaEval2015		Twitter	
		加权平均F1值	宏F1值	加权平均F1值	宏F1值
基于机器学习	PTK	80.14	78.93	81.13	80.26
	PES	82.06	79.87	83.03	82.76
基于神经网络	TCNN-URG	87.42	87.02	86.68	86.62
	CSI	85.11	84.91	89.25	89.11
基于预训练语言模型	BERT	87.29	86.57	88.52	88.26
	RoBERTa	87.32	86.17	84.65	84.23
GCN-ALBERT		**89.43**	**89.17**	**92.07**	**91.93**

从表3的实验结果可以看出,GCN-ALBERT模型在MediaEval2015和Twitter两个数据集上的加权平均F1值和宏F1值都要优于TCNN-URG、CSI、BERT、RoBERTa这些经典基线模型,相较于BERT模型,宏F1值分别提高了3.0%和4.2%。而传统的机器学习模型由于需要先进行特征选择,因此实验结果容易受到模型所选特征的影响,难以获得更高维、更复杂的特征数据从而限制了模型性能。GCN-ALBERT在MediaEval2015数据集上,与效果次优的TCNN-URG模型相比,加权平均F1值和宏F1值上分别提高了2.3%和2.5%,而在Twitter数据集上,与效果次优的CSI模型相比,GCN-ALBERT的加权平均F1值和宏F1值都提高了3.2%。实验结果验证了本文所提出方法在虚假新闻检测任务上的有效性。表明了GCN-ALBERT模型通过融合全局信息和局部信息的方式能有效提升模型对复杂文本的理解能力,有助于模型性能的提高。此外,GCN-ALBERT仅需要使用新闻文本内容进行识别,适用于虚假新闻早期检测应用场景。

从实验结果中，本文还发现基于Transformer的BERT模型在两个数据集上的平均性能要优于基于神经网络的TCNN-URG和CSI模型。这可能是因为Transformer使用自注意力和多头注意力机制对输入新闻文本进行建模，有助于模型学习到蕴含更丰富语义的文本特征表示。另一方面，BERT和RoBERTa的结果表明在大规模外部语料上进行预训练得到的通用语言表示，能够提高虚假新闻检测任务的性能。

接下来，为了验证模型不同部分各自发挥的作用，本文通过消融实验验证GCN和ALBERT两个部分对本文所提出方法的贡献。实验结果如表4所示。

表4 GCN-ALBERT消融实验结果 单位：%

模型	MediaEval2015		Twitter	
	加权平均F1值	宏F1值	加权平均F1值	宏F1值
GCN	79.34	78.86	77.11	76.32
ALBERT	87.58	86.94	78.42	77.43
GCN-ALBERT	89.43	89.17	92.07	91.93

从表4中可以看出，GCN-ALBERT的两个部分在虚假新闻检测任务中都发挥了重要作用。当数据集较大、文本内容较规范时，ALBERT的性能要优于图卷积网络；在较稀疏，规范程度较差的Twitter数据集上，GCN抽取的全局语义特征起到了关键作用。在两个数据集上，GCN-ALBERT的加权平均F1值和宏F1值都要优于ALBERT和GCN。实验结果表明GCN-ALBERT利用自注意力融合GCN和ALBERT，使得局部信息和全局信息相互促进，能够提高模型识别虚假新闻的性能。

4 结论

本文提出了一种基于图卷积神经网络的新模型GCN-ALBERT，该模型将图卷积神经网络与ALBERT结合进行虚假新闻检测。GCN-ALBERT首先利用GCN学习新闻文本的全局信息，并将全局文本特征和ALBERT提取的局部信息进行融合。然后，GCN-ALBERT利用自注意力层使两种类型的信息能够相互促进。在两个虚假新闻检测数据集上的实验结果表明，GCN-ALBERT的性能要优于基于神经网络的模型和基于预训练语言模型的基线方法。本文提出的方法仅使用新闻文本内容就能判断其真实性。在虚假信息早期检测领域，该模型的应用前景非常广阔，未来将利用提示学习训练GCN-ALBERT，提升模型的小样本学习能力，进一步提升检测性能。

参考文献（References）

[1] GAO W, LI L, TAO X, et al. Identifying informative tweets during a pandemic via a topic-aware neural language model[J]. World Wide Web, 2023, 26: 55-70.

[2] 毛震东，赵博文，白嘉萌，等. 基于传播意图特征的虚假新闻检测方法综述[J]. 信号处理, 2022, 38(6):1155-1169.

[3] WANG X, YU C, BAUMGARTNER S, et al. Relevant document discovery for fact-checking articles[C]// Companion Proceedings of the Web Conference 2018. [S. l.]: International World Wide Web Conferences Steering Committee, 2018: 525-533.

[4] 毕蓓，潘慧瑶，陈峰，等. 基于异构图注意力网络的微博谣言监测模型[J]. 计算机应用, 2021, 41(12): 3546-3550.

[5] 薛海涛，王莉，杨延杰，等. 基于用户传播网络与消息内容融合的谣言检测模型[J]. 计算机应用, 2021, 41(12): 3540-3545.

[6] SHU K, CUI L, WANG S, ET AL. dEFEND: Explainable fake news detection [C]// Proceedings of the 2019 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2019: 395-405.

[7] LI C T, LIN Y J, YEH M Y. Forecasting participants of information diffusion on social networks with its applications[J]. Information Sciences, 2018, 422: 432-446.

[8] CHEN T, LI X, YIN H, et al. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection [C]// Proceedings of the 2018 Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer, 2018: 40-52.

[9] BIYANI P, TSIOUTSIOULIKLIS K, BLACKMER J. "8 amazing secrets for getting more clicks": detecting clickbaits in news streams using article informality[C]// Proceedings of the 30th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016:94-100.

[10] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification[C]// Proceedings of the 33th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 7370-7377.

[11] LU Z, DU P, NIE J-Y. VGCN-BERT: Augmenting BERT with graph embedding for text classification[C]// Proceedings of the 2020 European Conference on Information Retrieval. Cham: Springer, 2020: 369-382.

[12] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: a lite BERT for self-supervised learning of language representations [C]// Proceedings of the 2020 International Conference on Learning Representation. New Orleans: ICLR, 2020: 1-17.

[13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[14] GAO W, LI L, ZHU X, et al. Detecting disaster-related tweets via multi-modal adversarial neural network[J]. IEEE Multimedia, 2020, 27(4): 28-37.

[15] 骆文莉，吴秦. 多层次特征融合与注意力机制的文本检测[J]. 小型微型计算机系统, 2022, 43(4):815-821.

[16] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence. California: ijcai. org, 2016: 3818-3824.

[17] DEVLIN J, CHANG M-W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics ,2019: 4171-4186.

[18] CAI H, ZHENG V W, CHANG K C-C. A comprehensive survey of graph embedding: problems, techniques, and applications[J]. IEEE Transactions on Knowledge and Data Engineering. 2018, 30(9): 1616-1637.

[19] SHU K, WANG S, LIU H. Beyond news contents: The role of social context for fake news detection[C]// Proceedings of the 12th ACM International Conference on Web Search and Data Mining. New York: ACM ,2019: 312-320.

[20] HAN Y, KARUNASEKERA S, LECKIE C. Graph neural networks with continual learning for fake news detection from social media [EB/OL]. [2020-08-14]. https://arxiv. org/pdf/2007. 03316v2. pdf.

[21] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks [C]// Proceedings of the 2017 International Conference on Learning Representation. New Orleans:: ICLR, 2017:1-14.

[22] 王昕岩, 宋玉蓉, 宋波. 一种加权图卷积神经网络的新浪微博谣言检测方法[J]. 小型微型计算机系统, 2021, 42(8):1780-1786.

[23] MEHTA N, PACHECO M L, GOLDWASSER D. Tackling fake news detection by continually improving social context representations using graph neural networks [C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, ,2022: 1363-1380.

[24] GAO W, FANG Y, Li L, et al. Event detection in social media via graph neural network [C]// Proceedings of the 22nd International Conference on Web Information Systems Engineering. New York: ACM, 2021: 370-384.

[25] WANG Y, MA F, JIN Z, et al. EANN: event adversarial neural networks for multi-modal fake news detection [C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM ,2018: 849-857.

[26] RUCHANSKY N, SEO S, LIU Y. CSI: a hybrid deep model for fake news detection [C]// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. New York: ACM ,2017: 797-806.

[27] 黄皓, 周丽华, 黄亚群, 等. 基于混合深度模型的虚假信息早期检测[J]. 山东大学学报(工学版), 2022, 52(4): 89-98, 109.

[28] YUAN C, MA Q, ZHOU W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection [C]// Proceedings of the 2019 IEEE International Conference on Data Mining. Piscataway: IEEE, 2019: 796-805.

[29] ZHOU X, JAIN A, PHOHA V V, et al. Fake news early detection: an interdisciplinary study [EB/OL]. [2020-09-16]. https://arxiv. org/pdf/1904. 11679. pdf.

[30] LEVY O, GOLDBERG Y. Neural word embedding as implicit matrix factorization [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. New York: ACM, 2014: 2177-2185.

[31] MA J, GAO W, WONG K-F. Detect rumors in microblog posts using propagation structure via kernel learning [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2017:708-717.

[32] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media [C]// Proceedings of the 2013 IEEE 13th International Conference on Data Mining. Piscataway: IEEE, 2013: 1103-1108.

[33] QIAN F, GONG C, SHARMA K, et al. Neural user response generator: fake news detection with collective user intelligence [C]// Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California: IJCAI, 2018: 3834-3840.

[34] LIU Y, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized bert pretraining approach [EB/OL]. [2019-07-26]. https://arxiv. org/pdf/1907. 11692. pdf.

[35] BOIDIDOU C, ANDREADOU K, PAPADOPOULOS S, et al. Verifying multimedia use at mediaeval 2015 [J]. MediaEval. 2015, 3(3): 1-3.

[36] GAO W, PENG M, WANG H, et al. Incorporating word embeddings into topic modeling of short texts [J]. Knowledge and Information Systems, 2019, 61(2): 1123-1145.