文章编号:1671-6833(2024)04-0070-09

基于图卷积网络的多特征融合谣言检测方法

关昌珊, 邴万龙, 刘雅辉, 顾鹏飞, 马洪亮

(石河子大学 信息科学与技术学院,新疆 石河子 832003)

摘 要: 目前,大部分谣言检测工作主要基于 Twitter 或新浪微博原文本内容、传播结构和传播文本内容进行谣言检测,忽略了原文本特征与其他特征的有效融合,以及传播用户在谣言传播过程中的作用。针对以上问题,提出了一种基于图卷积网络的多特征融合模型 GCNs-BERT,模型同时融合了原文本特征、传播用户特征和传播结构特征。首先,基于传播结构和传播用户构建传播图,将多个用户属性的组合作为传播节点特征;其次,利用多个图卷积网络学习在不同用户属性组合的情况下传播图的表达,同时采用 BERT 模型学习原文本内容特征表达,最终与图卷积网络学习的特征相融合用于检测谣言。利用公开的新浪微博数据集进行的大量实验表明:GCNs-BERT 模型明显优于基线方法。此外,在新冠疫情数据集上进行 GCNs-BERT 模型泛化能力实验,此数据集训练样本大小仅有新浪微博数据集的 1/5,仍然取得了 92.5%的准确率,证明模型具有较好的泛化能力。

关键词:谣言检测; 图卷积网络; 传播图; 传播用户; 特征融合

中图分类号: TP18 **文献标志码**:A **doi**:10.13705/j.issn.1671-6833.2024.01.011

随着互联网的飞速发展,社交媒体已成为获取信息、交换意见的主要平台。根据中国互联网络信息中心发布的第 52 次《中国互联网发展状况统计报告》[1],截至 2023 年 6 月,中国网民规模达 10.79 亿,互联网普及率达 76.4%。社交媒体平台,尤其是新浪微博(以下简称微博)月活用户量已达 5.99 亿,其在传播正常消息的同时也会滋生大量谣言,并在短时间内迅速传播,从而对个人或社会产生恶劣的影响。因此,针对谣言造成的潜在危害,高效、准确地开展谣言检测任务,对积极、正面的舆情引导具有重要意义。

早期的谣言检测研究基于机器学习的方法,手动提取一组特征使用决策树、支持向量机、随机森林等方法进行谣言检测。这些方法取得了一定的成果,但是过分依赖于手工构建特征,且特征提取类别较为单一,无法获得上下文语义等重要信息。

相比于机器学习,深度学习方法能够从大量数据中自动学习出需要的特征,研究者们开始尝试采用深度学习的方法进行谣言检测。梁兆君等[2]使用 BERT 模型将原文本向量化,分别在 Twitter15 和 Twitter16 数据集中取得了 77.5% 和 78.6% 的准确率。一部分研究除使用原文本之外,针对评论文本内容[3-4]学习上下文信息之间的关系来检测谣言,取得了不错的进展。原文本蕴含丰富的语义信息。因此,另一部分研究者们融合原文本和传播信息,将谣言的传播结构构建为传播树[5-6]或传播图[7-9],原文本和传播文本经过 TF-IDF 编码后作为传播树或传播图的节点特征,基于传播结构和内容考虑全局转发关系进行谣言检测,也获得了较好的检测效果。然而,以上方法仍存在两点不足:原文本特征利用不足,以及原文本特征和其他特征的有效融合方式,尤其与传播结构的融合还需要进一步探究;忽略了传播用户在谣言传播过程中的作用,孟青等[10]指出,当微博用户在浏览微博内容时,页面中显示的其他用户发布的转发评论会对用户造成一定的干扰,使用户受影响从而有可能对某些博文进行转发或评论。因此,本文针对存在的不足主要研究如何更好地融入原文本内容特征,以及如何将多个用户属性同传播结构特征有机地结合起来用于谣言检测,进一步提高谣言检测的准确率。针对这两方面,本文

收稿日期:2023-09-10;**修订日期**:2023-10-12

基金项目:国家自然科学基金资助项目(62062060);石河子大学高层次人才科研启动项目(RCZK2018C11, RCZK2018C38)

通信作者:刘雅辉(1979—),女,新疆石河子人,石河子大学副教授,博士,主要从事网络空间安全、机器学习研究,E-mail:lyh@shzu.edu.cn。

引用本文:关昌珊,邴万龙,刘雅辉,等. 基于图卷积网络的多特征融合谣言检测方法[J]. 郑州大学学报(工学版),2024,45(4):70-78. (GUAN C S, BING W L, LIU Y H, et al. Multi-feature fusion rumor detection method based on graph convolutional network[J]. Journal of Zhengzhou University (Engineering Science),2024,45(4):70-78.)

提出了 GCNs-BERT 模型,同时融合了原文本特征、传播用户特征和传播结构特征。首先,基于传播结构构建传播树;其次,提取多个用户属性特征并筛选出多个用户属性特征的组合作为传播节点的特征,组成传播图;其次,利用多个图卷积网络(graph convolutional network,GCN)学习在不同用户属性特征组合的情况下传播图的特征表达;最后,考虑到 BERT(bidirectional encoder representations from transformer)[11]模型在文本深层语义特征提取时的良好表现,利用 BERT 模型学习原文本内容特征,最终与图卷积网络学习的特征融合起来检测谣言。

1　相关工作

1.1　机器学习方法

早期的研究者们基于手动提取的特征,利用机器学习中的分类方法进行谣言检测。一些工作利用谣言与非谣言传播时序和传播结构的特征差异,来提高仅使用原文本特征检测谣言的准确率。例如,Ma 等[12]提出了基于整个谣言生命周期的时间序列特征,开发了一个动态序列-时间结构的 SVM 分类器(SVM_{all}^{DSTS}),在新浪微博数据集上取得了 86.1% 的精确度和 85.4%的召回率。Liu 等[13]从消息的传播中提取了结构特征、时间特征、用户特征和内容特征 4 种类型的特征,并将这些特征和 Wu 等[14]提取的特征相结合,在新浪微博转发数大于 100 的数据集中,识别准确率提高到了 94.3%。这些方法虽然在谣言检测任务中取得了一定的甄别效果,但需要手动提取特征,过程费时费力,使用的特征并不全面,也未能利用上下文等的重要背景信息。

1.2　深度学习方法

深度学习方法弥补了机器学习方法中存在的不足,能自动学习特征,提高了谣言检测效率。相关工作可以分为基于文本内容的谣言检测方法以及基于传播结构和文本内容的谣言检测方法。

1.2.1　基于文本内容的谣言检测方法

基于文本内容的方法是将微博文本内容作为研究对象进行谣言检测,这类任务通常针对原文本数据和转发评论数据。Ma 等[4]首次基于传播内容利用循环神经网络学习微博文本的表达,捕捉文本上下文信息随时间的变化来检测谣言,在新浪微博数据集中识别准确率达到 91%。Wang 等[15]基于文本内容构建了图神经网络模型(SemSeq4FD)用于早期谣言检测,该模型考虑了谣言文本中句子之间的全局语义关系、局部顺序特征和全局顺序特征,将文本建模为一个完整的图,并通过具有自注意力机制的图卷积网络学习全局句子表达进行谣言检测。Ma 等[16]受到对抗性学习的启发,提出了一种基于生成对抗的谣言检测方法,生成器用于产生不确定或冲突的特征,使判别器从极具挑战性的样本中学习到更具代表性的谣言特征表达。

1.2.2　基于传播结构和文本内容的谣言检测方法

Sharma 等[17]发现,谣言发布者通常会根据真实信息的表达特点,故意效仿、捏造信息从而躲避检测。然而仅基于谣言内容的检测方法不能够高效检测谣言,因此,部分研究者们尝试将传播结构和文本内容相结合来检测谣言。

Ma 等[5]基于传播结构和文本内容,利用递归神经网络分别学习传播树中自顶向下的传播方向和自底向上的扩散方向上各节点的隐藏表达来检测谣言,将谣言分为非谣言、假谣言、真实谣言和未经证实的谣言 4 类,分别在 Twitter15 和 Twitter16 数据集中取得了 72.3%和 73.7%的准确率。然而,基于传播树的方法仅关注学习序列化特征,忽略了帖子上下文之间的全局转发关系。

近年来,由于图卷积网络模型强大的学习能力,研究者们尝试利用图卷积网络模型[18]学习传播结构的表达。首先将谣言的传播过程构建为一张图,进而将谣言分类问题转化为图分类任务,取得了不错的检测效果。Song 等[19]融合传播结构、文本内容和发布时间信息构建了连续的动态扩散网络,利用图神经网络从时间交互的角度捕捉谣言传播的动态演变模式,在新浪微博数据集中分类准确率达到了 96.8%。Bian 等[7]提出了双向图卷积网络(Bi-GCN),该方法用自顶向下的图表示谣言的传播信息,同时用自底向上的图表示谣言的扩散信息,通过图卷积网络学习图中的结构信息进行谣言分类,在新浪微博数据集中分类准确率达到了 96.1%。然而,这些方法仅关注了传播过程中传播关系的表达,忽略了传播节点,即用户属性对谣言检测的影响。

此外,谣言中用户的基本特征和在线社交网络中的行为特征[20]与非谣言有一定的差异,研究者们尝试将用户相关特征融入谣言检测任务中。Lu 等[21]基于原文本和传播用户提出了图感知协同注意网络(GCAN)谣言检测方法,首先从原文本中学习单词嵌入,其次提取用户特征,并使用卷积和递归神经网络来学习基于用户特征的转发传播的表达,构建了图结构来对用户之间的潜在交互进行建模,并使用图卷积网络来学习用户交互的图感知表达,分别在 Twitter15 和 Twitter16 数据集中取得了

87.6%和 90.8%的准确率。

可以看出,深度学习方法的准确率普遍比机器学习方法高。综合分析可得,内容特征、用户特征和传播特征相融合对检测谣言效果更好。因此,本文基于原文本特征、传播结构特征和传播用户特征提出了基于图卷积网络的多特征融合的谣言检测方法,将传播用户和传播结构结合起来构建传播图,学习谣言与非谣言在传播过程中的用户和结构差异,同时充分利用原文本信息进行多特征融合,进一步提高了谣言检测的效果。

2 问题定义

定义 1 谣言。谣言是指在社交媒体中传播的与真实信息不符或故意伪造的信息,如错误信息和虚假信息[22]等。

定义 2 传播图。在社交平台上,一条信息的所有转发用户之间的转发路径形成了树状结构,这种结构通常被称为传播树。本文利用传播树结构和相应用户特征构建了传播图。其中,根节点表示发布博文的用户,其他节点表示后续转发博文的用户。

给定一个谣言检测数据集 $M=\{m_1,m_2,\cdots,m_p\}$,其中 m_i 为第 i 个博文的消息集,p 为数据集中博文的个数,$m_i=\{s_i,u_i,u_1^{\,i},u_2^{\,i},\cdots,u_j^{\,i},G_i\}$,其中 s_i 为博文 m_i 的原发博文,u_i 为发布博文的用户,u_j^i 为原发博文 s_i 的第 j 个转发用户,j 为博文 m_i 中的转发用户数,$G_i=<V_i,E_i>$为对 s_i 的转发关系构建的传播图。

谣言检测将消息分为谣言和非谣言两类,每个博文 m_i 可标注为类别标签 $y_i\in Y(0,1)$,其中 Y 代表博文的类别标签集合。谣言检测任务可以转化为学习一个函数 $f:f(m_i)\rightarrow y_i$。

3 GCNs-BERT 模型

本文提出的谣言检测模型 GCNs-BERT 的总体架构如图 1 所示,主要分为 3 个模块:传播图特征表达模块、原文本特征表达模块和预测模块。

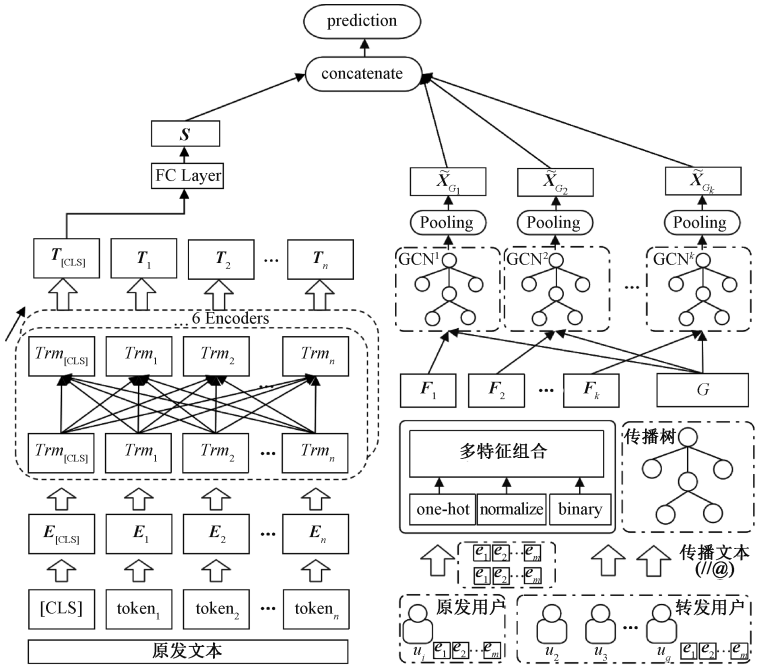

图 1 GCNs-BERT 模型架构图

Figure 1 GCNs-BERT model architecture diagram

3.1 传播图特征表达模块

图卷积神经网络的主要思想是通过学习节点间的信息传播来更新节点的特征表达,对所有节点迭代地聚合自身节点和邻居节点的信息,最终生成节点新的特征表达。

3.1.1 传播图的结构构建

根据博文 m_i 的传播关系构建传播图 $G_i=<V_i,E_i>$,其中 G_i 为无向图,图卷积神经网络通过对邻居节点进行聚合的方式来捕获传递信息;$V_i=\{u_i,u_1^i,u_2^i,\cdots,u_j^i\}$ 表示原博文发布用户 u_i 和它对应的 j 个转发用户的集合;$E_i=\{e_{qv}\mid q=0,1,\cdots,k;v=0,1,\cdots,k\}$ 表示传播图中的所有边集。如图 2 所示为博文 m_i 转发消息所对应的传播关系图,其中用户 u_1 和 u_2 相继转发了原博文 s_i,在集合 E_i 中则包含 e_{01}、e_{10}、e_{02} 和 e_{20};u_5 转发了 u_1,则集合 E_i 中则包含 e_{15} 和 e_{51}。设 $\boldsymbol{A_i}\in\{0,1\}^{k_i\times k_i}$ 为邻接矩阵,其元素如

下所示:

$$a_{qv}^{i}=\begin{cases}1, & \text{若 } e_{qv}^{i}\in E_i;\\ 0, & \text{其他。}\end{cases} \tag{1}$$

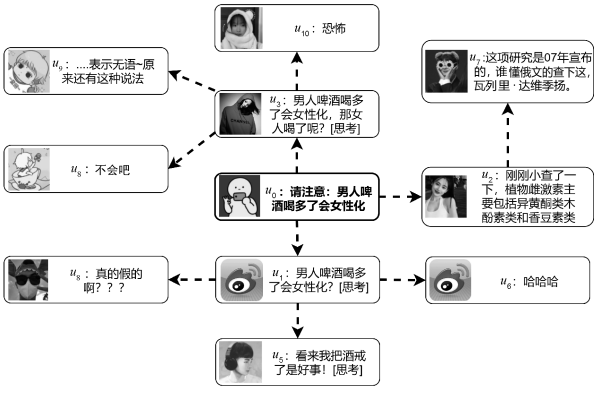

图 2　社交媒体场景下传播图的构造方法图

Figure 2　Construction method diagram of propagation graph in social media scenario

3.1.2　传播图的节点特征选择

发布微博的用户以及传播用户其自身特征对谣言检测具有一定的辅助作用,用户的认证类型和信用情况等属性一定程度上代表了用户的影响力,有影响力的用户能促进观点、行为、创新和产品在社交网络中的传播。因此,本文传播图节点的特征是由转发用户特征组成。对转发用户的 18 个特征进行筛选,最终获得了 8 个比较有甄别度的特征分别为用户性别、认证类型、动态数(用户发布博文总数)、粉丝数、注册天数、活跃度(动态数/注册天数)、是否写个人简介、点赞数。

考虑到图卷积神经网络的训练效率和过拟合问题,针对上述 8 个特征再次进行组合筛选,最终获得了最具有甄别力的 6 组用户特征组合 $\boldsymbol{F}$,如表 1 所示。Yang 等[23]研究发现男性相比于女性,发布谣言的概率更低,说明用户性别特征在检测谣言中具有一定辅助作用。因此,性别特征与其他多个用户特征组合均获得了较好的检测效果。针对不同类型的用户特征做数据预处理,认证类型和会员类型这类离散型数据,采用 one-hot 编码的形式表达;对连续的数值型特征进行 min-max 标准化表达;对是否写个人简介的布尔类型数据做二值化表达,最终得出传播图中每个节点的特征表达。

3.1.3　传播图的特征表达

构建好传播图以及选择好特征之后,为了充分利用传播图中的传播结构信息,使传播图中的各个节点能更好地聚合邻居信息以获得更好的特征表达,引入了图卷积神经网络。

GCN 通过相邻节点信息来更新该节点的隐藏层信息,输入为特征矩阵 $\boldsymbol{X}\in\mathbf{R}^{n\times n}$,邻接矩阵 $\boldsymbol{A}_i\in\mathbf{R}^{n\times n}$。若 GCN 模型有多层时,会聚合更多的邻居节点特征,因此在 GCN 中 l 个隐藏层的特征矩阵 $\boldsymbol{X}^{l+1}$ 为

$$\boldsymbol{X}^{l+1}=\mathrm{ReLU}(\tilde{\boldsymbol{A}}\boldsymbol{X}^{l}\boldsymbol{W}^{l})。 \tag{2}$$

式中:$\boldsymbol{X}^{l+1}\in\mathbf{R}^{n\times m}$ 为图卷积操作后的特征矩阵;$\mathrm{ReLU}(\cdot)$为激活函数;$\boldsymbol{W}^{l}\in\mathbf{R}^{n\times m}$ 为可学习的参数;l 为图卷积操作的层数;$\tilde{\boldsymbol{A}}$ 代表对邻居所传播的信息进行标准化后的邻接矩阵[24]:

$$\tilde{\boldsymbol{A}}=\tilde{\boldsymbol{D}}^{-\frac{1}{2}}(\boldsymbol{A}+\boldsymbol{I})\tilde{\boldsymbol{D}}^{-\frac{1}{2}}。 \tag{3}$$

式中:$\tilde{\boldsymbol{D}}$ 为传播图对应的度矩阵;$\tilde{D}_{ii}=\sum j\tilde{A}_{jj}$。

GCN 在更新自身节点时,通常会添加自连接,把自身特征和邻居特征结合起来更新节点:

$$X_i^{*}=\sum_{j\in N}A_{ij}X_j+X_j。 \tag{4}$$

通过 5 层图卷积操作后,得到特征矩阵 $\boldsymbol{X}^5$,为防止过拟合缩小参数矩阵的尺寸,在图卷积操作之后加入一个平均池化层。更新后的特征向量表示为

$$\boldsymbol{X}_G=\mathrm{average\_pooling}(\boldsymbol{X}^5)。 \tag{5}$$

表 1　用户特征组合

Table 1　User characteristic combination

特征	特征组合
$\boldsymbol{F}_1$	活跃度、粉丝数
$\boldsymbol{F}_2$	性别、注册天数
$\boldsymbol{F}_3$	性别、认证类型
$\boldsymbol{F}_4$	性别、动态数
$\boldsymbol{F}_5$	是否写个人简介、注册天数
$\boldsymbol{F}_6$	性别、点赞数

3.2　原文本特征表达模块

谣言的原博文中包含着重要且丰富的原始信息,充分利用原博文信息能够提高谣言检测的性能。由于 BERT 能双向提取语义信息从而得到更充分、更隐蔽的特征,本文利用预训练 BERT 模型[11]来学习原文本中的上下文语义信息。

3.2.1　词嵌入层

预训练 BERT 的输入共包含 3 个嵌入层,分别是词嵌入层(token embeddings)、句子嵌入层(segment embeddings)和位置嵌入层(position embeddings)。长度为 n 的微博原文本经过分词后得到 $W=\{[\mathrm{CLS}],\mathrm{token}_1,\mathrm{token}_2,\cdots,\mathrm{token}_n\}$,其中,[CLS]表示用于后续分类的 token 标志,token_i 表示 W 中的第 i 个词。将 $\boldsymbol{W}$ 输入 3 个嵌入层分别得到词向量 $\boldsymbol{W}_{\mathrm{token}}$、文本向量 $\boldsymbol{W}_{\mathrm{segment}}$ 和位置向量 $\boldsymbol{W}_{\mathrm{position}}$,将三者进行叠加得到新向量 $\boldsymbol{W}=\{\boldsymbol{W}_{[\mathrm{CLS}]},\boldsymbol{W}_1,\boldsymbol{W}_2,$

$\cdots, \boldsymbol{W}_n\}$,如图 1 所示。其中,$\boldsymbol{W}$ 也可表示为

$$\boldsymbol{W} = \boldsymbol{W}_{\text{token}} + \boldsymbol{W}_{\text{segment}} + \boldsymbol{W}_{\text{position}}。\tag{6}$$

3.2.2 编码层

编码层的任务是将词嵌入层转换后的向量编码成具有丰富的上下文语义信息的序列向量。如图 1 所示,BERT 的内部结构使用到 Transformer 的编码器(encoder)部分,通过编码器的多头注意力机制,使得每个 token 的编码表达(Trm_i)之间均有一条相互关注的边,可以实现对不同距离的词语之间具有同等的关注程度。BERT 内部共串联堆叠了 6 个相同结构的编码器,以更好地学习输入文本的上下文语义信息。其中,多头注意力可表示为

$$\text{MultiHead}(Q,K,V) = \text{Concat}(\boldsymbol{head}_1, \boldsymbol{head}_2, \cdots, \boldsymbol{head}_n)\boldsymbol{\omega};\tag{7}$$

$$\boldsymbol{head}_i = \text{Attention}(\boldsymbol{Q}\boldsymbol{\omega}_i^Q, \boldsymbol{K}\boldsymbol{\omega}_i^K, \boldsymbol{V}\boldsymbol{\omega}_i^V)。\tag{8}$$

式中:n 为多头注意力机制头的数量;$\boldsymbol{head}_i$ 为第 i 个头的输出;$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 由输入特征矩阵线性变换而得;$\boldsymbol{\omega}^Q$、$\boldsymbol{\omega}^K$、$\boldsymbol{\omega}^V$ 分别为训练后学习到的 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 的参数矩阵。

最后将分类标记[CLS]学习到的特征向量输入全连接前馈神经网络层,得到博文的语义表达 $\boldsymbol{S}$。

3.3 预测模块

本文采用多特征的混合融合方法进行谣言预测,如图 3 所示。为提高整个网络的鲁棒性,首先在图卷积网络中加入两层全连接层得到最终特征向量,将表 2 中的 6 组用户特征分别输入到 6 个 GCN 中训练,共得到 6 组传播图特征向量:

$$\hat{\boldsymbol{X}} = \text{Concate}(\tilde{\boldsymbol{X}}_{G_1}, \tilde{\boldsymbol{X}}_{G_2}, \tilde{\boldsymbol{X}}_{G_3}, \tilde{\boldsymbol{X}}_{G_4}, \tilde{\boldsymbol{X}}_{G_5}, \tilde{\boldsymbol{X}}_{G_6})。\tag{9}$$

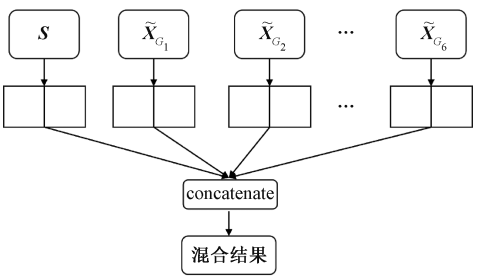

图 3 多特征混合融合方法图

Figure 3 Multi-feature hybrid fusion method

表 2 公开数据集统计信息

Table 2 Expose dataset statistics

统计项目	非谣言	谣言
博文数量	2 291	2 229
传播树数量	2 291	2 229
转发帖子数量	1 711 420	2 086 693
每条博文的平均帖子数	747	936

再将传播图特征向量和原文本特征向量拼接:

$$\boldsymbol{C} = \text{Concate}(\hat{\boldsymbol{X}}, \boldsymbol{S})。\tag{10}$$

然后将谣言检测问题转化为分类任务,基于最终特征融合表达,通过输入全连接层和 softmax 预测该博文属于某类谣言的概率:

$$\hat{y} = \text{softmax}(\boldsymbol{W}_c + \boldsymbol{b}_c),\tag{11}$$

式中:$\boldsymbol{W}_c$ 和 $\boldsymbol{b}_c$ 均为通过训练可学习的参数。模型通过最小化交叉熵损失函数来降低真实标签与预测标签之间的分类误差:

$$L(\theta) = -\sum_{i=0}^{k-1} y_i \log \hat{y}_i。\tag{12}$$

式中:k 为分类的类别数;θ 为模型的参数;$y_i \in \{0, 1\}$ 为真实标签值。

4 实验

4.1 数据集

本文在 2 个数据集上对模型进行了实验。首先,在 2016 年 Ma 等[4]公开的微博数据集上评价了提出的 GCNs-BERT 模型的有效性。为提取更丰富的特征提高谣言检测的性能,过滤了转发数小于 30 的博文,数据集中包含原博文和转发帖,以及用户的性别、粉丝数等属性,数据集统计信息如表 2 所示。考虑到公开数据集的时效性,同时为了验证模型在新事件谣言检测上的泛化能力,本文爬取新浪微博平台与新冠疫情相关的微博数据,与部分公开数据集组合成新的实验数据集来评估 GCNs-BERT 模型的泛化能力。新数据集的训练集包含 1 000 条公开数据集中的谣言和非谣言数据,以及 400 条新冠疫情谣言和非谣言数据;测试集包含 200 条新冠疫情谣言和非谣言数据。

4.2 GCNs-BERT 模型有效性实验

4.2.1 参数设置

验证实验中将图卷积网络层数设置为 5 层,每层节点的隐藏向量维度 d 均设为 256,即 $d_1 = d_2 = d_3 = d_4 = d_5 = 256$。为了防止过拟合,模型各层的随机失活 $Dropout = 0.5$,训练过程中 $Batch\_Size$ 大小为 128。采用 Adam 算法优化模型,学习率设置为 0.005,迭代次数设置为 500 并设置提前结束。

4.2.2 基线方法

选取基于机器学习的谣言检测方法和基于深度学习的谣言检测方法作为基线方法,与本文的 GCNs-BERT 进行对比。

(1)DTC[20]:基于人工设计的统计特征构建决策树分类模型判断信息的可信度。

(2)SVM[13]:从扩散角度提出了基于传播结构、传播时间和传播用户的 11 个特征,并将其与 Wu

等[14]所提取的特征相结合,使用混合核函数的 SVM 分类器检测谣言,此基线本文只使用基于 RBF 的 SVM 分类器检测谣言。

(3)LSTM[25]:基于长短期记忆网络和最大池化结合,通过建模相关帖子的传播结构的动态变化来检测谣言。

(4)BERT:本文利用 Devlin 等[26]提出的 BERT 预训练模型,基于微博原博文学习上下文语义进行谣言分类。

(5)Bi-GCN[5]:使用原文本、传播结构和传播文本,利用双向图卷积网络来检测谣言。

(6)GCNs:本文基于传播结构和传播用户构建传播图,根据 6 组用户特征组合构建 6 个 GCN 网络以捕获丰富的传播结构和传播用户特征来检测谣言。

4.2.3　结果分析

使用准确率、精确度、召回率和 $F1$ 值来验证模型的谣言检测性能。表 3 展示了在公开数据集中本文提出的模型 GCNs-BERT 以及所有基线方法的实验结果。

表 3　谣言检测实验结果

Table 3　Rumor detection experiment results

方法	类别	准确率	精确度	召回率	$F1$
DTC	1	0.885	0.834	0.955	0.890
	0		0.949	0.817	0.878
SVM	1	0.913	0.901	0.923	0.912
	0		0.924	0.902	0.913
LSTM	1	0.928	0.963	0.898	0.920
	0		0.893	0.962	0.926
BERT	1	0.938	0.935	0.940	0.938
	0		0.942	0.936	0.939
Bi-GCN	1	0.946	0.944	0.949	0.943
	0		0.950	0.940	0.942
GCNs	1	0.947	0.940	0.952	0.946
	0		0.953	0.941	0.947
GCNs-BERT	1	**0.956**	**0.964**	0.945	0.955
	0		0.949	**0.966**	**0.957**

从表 3 中可以看出,深度学习的方法性能要比机器学习的更好,本文的 GCNs-BERT 与 DTC 相比,准确率提高了 7.1 百分点,与 SVM 相比提高了 4.3 百分点。原因是机器学习依赖于手工提取特征,而深度学习能够自动捕获到更深层次的特征以及特征之间的关联,有助于提高谣言检测的性能。

GCNs-BERT 模型效果优于 LSTM,准确率提升了 2.8 百分点,说明虽然 LSTM 能够捕捉传播内容、传播用户和传播结构在整个传播过程中的一些动态变化,但可能会在时间序列中丢失一些整体的传播信息,影响最终的预测结果。针对谣言的传播特点,使用图结构来学习整个传播过程中的综合特征对检测谣言更有效。

GCNs-BERT 模型比 BERT 模型准确率提升了 1.8 百分点,说明图卷积网络模型能够有效学习到谣言和非谣言在传播结构和传播用户特征上的差异,能够进一步提升谣言检测的准确率。

虽然 Bi-GCN 使用了双向的图卷积网络对传播结构进行了建模,同时增强传播树中的原节点特征表达,但其忽略了传播过程中传播用户的重要影响。本文模型中将传播用户特征作为传播树的节点特征,学习谣言在传播过程中传播用户的影响,准确率与 Bi-GCN 相比提升了 1.0 百分点,说明传播用户特征对谣言检测任务有较好的辅助效果。

GCNs-BERT 模型的检测准确率比 GCNs 提升了 0.9 百分点,原因是 GCNs 未加入微博原文本特征,说明原文本中包含着丰富且重要的信息,充分利用原文本内容特征能有效提升模型的性能。

4.3　GCNs-BERT 模型泛化能力实验

在新冠疫情数据集中进行 GCNs-BERT 模型泛化能力实验,其参数设置参考 4.2.1 节。由于这部分数据量较小且考虑到模型的复杂度将每个 GCN 的隐藏向量维度设置为 128,隐藏层数设为 3,并在 DTC、SVM、BERT、GCNs、GCNs-BERT 模型上进行实验,结果如表 4 所示。

表 4　新冠疫情谣言识别实验结果

Table 4　Experimental results of COVID-19 rumor identification

方法	类别	准确率	精确度	召回率	$F1$
DTC	1	0.871	0.823	0.892	0.880
	0		0.882	0.797	0.861
SVM	1	0.896	0.874	0.883	0.895
	0		0.898	0.869	0.905
BERT	1	0.905	0.885	0.930	0.907
	0		0.926	0.880	0.902
GCNs	1	0.865	0.796	0.980	0.878
	0		0.974	0.750	0.847
GCNs-BERT	1	**0.925**	0.912	**0.940**	**0.926**
	0		**0.938**	0.910	0.923

从表 3 和表 4 中看出,表 4 的结果均低于表 3 的结果。其主要原因是新冠疫情数据受新浪微博平台管制导致微博传播数量受限,影响了模型识别的效果。GCNs-BERT 与 DTC 相比,准确率提高了 5.4 百分点,与 SVM 相比提高了 2.9 百分点,与 BERT 模型相比提高了 2 百分点。结果表明 GCNs-BERT 泛化能力强,在新的数据集上仍然能获得较好的谣

言检测效果。

4.4 消融实验

4.4.1 用户特征组合分析

为了证明特征组合的有效性和必要性,并避免原文本语义特征对实验的干扰,去除了 BERT 模型的部分,仅使用传播结构特征、8 个用户特征直接拼接以及 8 个用户特征组合 3 种方案进行对比实验,结果如表 5 所示。

表 5 特征分析结果对比

Table 5 Comparison of characteristic analysis results

特征	类别	准确率	精确度	召回率	$F1$
GCN(仅用传播结构)	1	0.887	0.888	0.880	0.884
	0		0.885	0.893	0.889
GCN(8 个用户特征拼接)	1	0.934	0.920	0.948	0.933
	0		0.948	0.920	0.934
GCNs	1	0.947	0.940	0.952	0.946
	0		0.953	0.941	0.947

从表 5 中可以看出,GCNs 比仅使用传播结构的 GCN 提升了 6 百分点的准确率,说明仅用传播结构不能充分地学习到用户间的转发特征,加入传播用户特征后能明显提高谣言检测的准确率。其次 GCNs 比使用 8 个用户特征直接拼接的 GCN 提高了 1.3 百分点的准确率,说明将用户特征组合后采用多个 GCN 能更有效地学习到用户和传播结构特征,进一步提升了谣言检测的效果,证明了在模型中用户特征组合的有效性。

4.4.2 传播用户不同特征组合的影响分析

本文使用不同的传播用户特征来训练 GCNs-BERT 模型。传播用户特征结果对比如表 6 所示,表中的(-)表示包含了除当前特征组之外的 5 组传播用户特征。

表 6 传播用户特征结果对比

Table 6 Comparison of propagation user characteristics results

特征	类别	准确率	精确度	召回率	$F1$
$(-)\boldsymbol{F}_1$	1	0.954	0.960	0.946	0.953
	0		0.949	0.962	0.955
$(-)\boldsymbol{F}_2$	1	0.955	0.963	0.944	0.954
	0		0.947	0.965	0.956
$(-)\boldsymbol{F}_3$	1	0.955	0.962	0.945	0.953
	0		0.948	0.964	0.956
$(-)\boldsymbol{F}_4$	1	0.955	0.964	0.944	0.954
	0		0.947	0.962	0.955
$(-)\boldsymbol{F}_5$	1	0.954	0.960	0.945	0.952
	0		0.947	0.962	0.955
$(-)\boldsymbol{F}_6$	1	**0.952**	0.960	0.941	0.950
	0		0.944	0.961	0.953

结果表明,包含第 6 组特征的效果影响最显著,即用户性别和点赞数影响了模型 0.4 百分点的准确率。用户点赞数越多说明此条微博的影响力越大,被转发的可能性也会越大。其次,包含第 1 组和第 5 组特征对模型结果也有较好的效果,即活跃度、粉丝数、是否写个人简介和注册天数,影响了模型 0.2 百分点的准确率。网络水军与正常用户有明显差异,通常相对有权威或影响力的微博用户都会填写个人简介,因此粉丝数量比较多,同时注册天数也比较久。而注册微博有一定目的性的用户,例如网络水军一般极少填写个人简介,粉丝数量和注册天数都比较少,更有可能发布或转发谣言。

4.4.3 GCN 层数对模型性能的影响分析

由于在图卷积神经网络模型中堆叠较少的层数后,网络就能达到最好效果,继续增加图卷积层数反而会导致结果变得更差。因此分析了图卷积网络层数对 GCNs-BERT 模型最终性能的影响。假设层数 $l=\{1,2,3,4,5,6\}$,在公开数据集上训练模型层数,对模型最终结果的影响如图 4 所示。

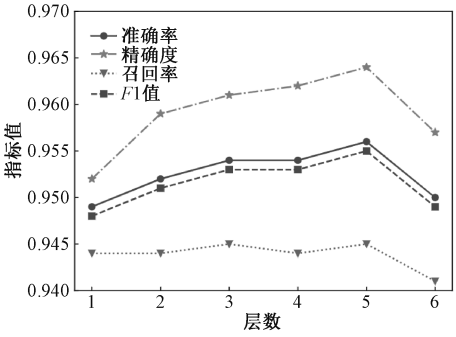

图 4 GCN 层数对模型性能的影响

Figure 4 Effect of the number of GCN layers on model performance

可以看出,当层数为 5 层 时,GCNs-BERT 模型的准确率和 $F1$ 值达到最高值,这证实了在实验中选择 5 层是合理的。随着层数的增加,图卷积网络参数变多,分类效果反而下降。

5 结论

本文研究了基于原文本内容、传播结构和传播用户相融合的谣言检测任务,提出了基于图卷积网络的多特征融合的谣言检测模型 GCNs-BERT。该模型将传播结构特征与传播用户特征相融合,使用图卷积网络来学习传播用户和传播结构的表达,使用 BERT 模型学习原文本内容的表达,最后进行多特征融合来检测谣言。相比于机器学习的基线方法,本文模型能够更好地学习文本的上下文特征以及传播结构特征;相比于循环神经网络基线方法,本

文模型能够学习传播过程的整体特征;相比于图卷积网络的基线方法,本文模型充分利用了传播用户的特征组合。同时,为了验证 GCNs-BERT 模型的泛化能力,在新冠疫情数据集中进行实验,得到了较好的谣言检测效果。此外,本文还进行了用户特征组合的消融实验,实验发现传播用户的性别和点赞数 2 个特征组合对谣言检测有较好的甄别效果。本文所构建的谣言检测模型 GCNs-BERT 很好地融合了原文本特征、传播结构特征和传播用户特征,获得了较好的谣言检测效果,进一步提升了谣言的检测性能。

在未来的工作中,将尝试以下两方面的工作:尝试采集原博文对应的图像数据集,提取图像中的信息,与原博文内容信息结合构建多模态检测模型,以尽可能早且准确地检测出谣言,进一步提升早期谣言检测性能;进行传播用户的影响力分析,并对用户影响力建模,学习传播用户更准确的表达,进一步提升传播用户特征的甄别力来更好地辅助预测谣言。

参考文献:

[1] 中国互联网络信息中心. 第 52 次《中国互联网络发展状况统计报告》[R/OL]. (2023-08-23)[2023-09-01]. https://www.cnnic.net.cn/n4/2023/0828/c88-10829.html.
China Internet Network Information Center. The 52nd China statistical report on internet development[R/OL]. (2023-08-23)[2023-09-01]. https://www.cnnic.net.cn/n4/2023/0828/c88-10829.html.

[2] 梁兆君, 但志平, 罗衍潮, 等. 基于 BERT 模型的增强混合神经网络的谣言检测[J]. 计算机应用与软件, 2021, 38(3): 147-152, 189.
LIANG Z J, DAN Z P, LUO Y C, et al. Rumor detection of improved hybrid neural network based on BERT model[J]. Computer Applications and Software, 2021, 38(3)147-152, 189.

[3] LI J W, NI S W, KAO H Y. Meet the truth: leverage objective facts and subjective views for interpretable rumor detection[EB/OL]. (2021-07-21)[2023-09-01]. https://arxiv.org/abs/2107.10747.

[4] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]//Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: IJCAI, 2016: 3818-3824.

[5] MA J, GAO W, WONG K F. Rumor detection on twitter with tree-structured recursive neural networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg:ACL,2018:1980-1989.

[6] 胡斗, 卫玲蔚, 周薇, 等. 一种基于多关系传播树的谣言检测方法[J]. 计算机研究与发展, 2021, 58(7): 1395-1411.
HU D, WEI L W, ZHOU W, et al. A rumor detection approach based on multi-relational propagation tree[J]. Journal of Computer Research and Development, 2021, 58(7): 1395-1411.

[7] BIAN T, XIAO X, XU T Y, et al. Rumor detection on social media with bi-directional graph convolutional networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 549-556.

[8] 杨延杰, 王莉, 王宇航. 融合源信息和门控图神经网络的谣言检测研究[J]. 计算机研究与发展, 2021, 58(7): 1412-1424.
YANG Y J, WANG L, WANG Y H. Rumor detection based on source information and gating graph neural network[J]. Journal of Computer Research and Development, 2021, 58(7): 1412-1424.

[9] BAI N, MENG F R, RUI X B, et al. Rumour detection based on graph convolutional neural net[J]. IEEE Access, 2021, 9: 21686-21693.

[10] 孟青, 刘波, 张恒远, 等. 在线社交网络中群体影响力的建模与分析[J]. 计算机学报, 2021, 44(6): 1064-1079.
MENG Q, LIU B, ZHANG H Y, et al. Multi-relational group influence modeling and analysis in online social networks[J]. Chinese Journal of Computers, 2021, 44(6): 1064-1079.

[11] 张铭泉, 周辉, 曹锦纲. 基于注意力机制的双 BERT 有向情感文本分类研究[J]. 智能系统学报, 2022, 17(6): 1220-1227.
ZHANG M Q, ZHOU H, CAO J G. Dual BERT directed sentiment text classification based on attention mechanism[J]. CAAI Transactions on Intelligent Systems, 2022, 17(6): 1220-1227.

[12] MA J, GAO W, WEI Z Y, et al. Detect rumors using time series of social context information on microblogging websites[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York: ACM, 2015: 1751-1754.

[13] LIU Y H, JIN X L, SHEN H W, et al. Do rumors diffuse differently from non-rumors? a systematically empirical analysis in Sina Weibo for rumor identification[J]. Advances in Knowledge Discovery and Data Mining, 2017, 10234: 407-420.

[14] WU K, YANG S, ZHU K Q. False rumors detection on Sina Weibo by propagation structures[C]//2015 IEEE 31st International Conference on Data Engineering. Piscataway: IEEE, 2015: 651-662.

[15] WANG Y H, WANG L, YANG Y J, et al. SemSeq4FD: integrating global semantic relationship and local sequen-

tial order to enhance text representation for fake news detection[J]. Expert Systems with Applications, 2021, 166: 114090.

[16] MA J, GAO W, WONG K F. Detect rumors on Twitter by promoting information campaigns with generative adversarial learning[C]//WWW '19: The World Wide Web Conference. New York: ACM, 2019: 3049-3055.

[17] SHARMA K, QIAN F, JIANG H, et al. Combating fake news: a survey on identification and mitigation techniques [J]. ACM Transactions on Intelligent Systems and Technology, 10(3): 1-42.

[18] PHAN H T, NGUYEN N T, HWANG D. Fake news detection: a survey of graph neural network methods[J]. Applied Soft Computing, 2023, 139: 110235.

[19] SONG C G, SHU K, WU B. Temporally evolving graph neural network for fake news detection[J]. Information Processing & Management, 2021, 58(6): 102712.

[20] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on Twitter[C]//International World Wide Web Conference Committee (IW3C2). New York: ACM, 2011: 1-10.

[21] LU Y J, LI C T. GCAN: graph-aware co-attention networks for explainable fake news detection on social media [EB/OL]. (2020-04-24)[2023-09-01]. https://arxiv.org/abs/2004.11648.

[22] 刘雅辉, 靳小龙, 沈华伟, 等. 社交媒体中的谣言识别研究综述[J]. 计算机学报, 2018, 41(7): 1536-1558. LIU Y H, JIN X L, SHEN H W, et al. A survey on rumor identification over social media[J]. Chinese Journal of Computers, 2018, 41(7): 1536-1558.

[23] YANG J, COUNTS S, MORRIS M R, et al. Microblog credibility perceptions: comparing the USA and China [C]//Proceedings of the 2013 conference on Computer Supported Cooperative Work. New York: ACM, 2013: 575-586.

[24] GONG S Z, SINNOTT R O, QI J Z, et al. Fake news detection through graph-based neural networks: a survey [EB/OL]. (2023-07-24)[2023-09-01]. https://arxiv.org/abs/2307.12639.

[25] LIU Y H, JIN X L, SHEN H W. Towards early identification of online rumors based on long short-term memory networks[J]. Information Processing & Management, 2019, 56(4): 1457-1467.

[26] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. (2019-05-24)[2023-09-01]. https://arxiv.org/abs/1810.04805.

Multi-feature Fusion Rumor Detection Method Based on Graph Convolutional Network

GUAN Changshan, BING Wanlong, LIU Yahui, GU Pengfei, MA Hongliang

(School of Information Science and Technology, Shihezi University, Shihezi 832003, China)

Abstract: At present, most rumor detection work mainly based on the original text content, communication structure and communication text content of Twitter or Weibo. However, these methods ignored the effective integration of original text features with other features, as well as the role of propagating users in the process of rumor propagation. Aiming at the shortcomings of the existing work, a multi-feature fusion model GCNs-BERT based on graph convolutional network was proposed, which combined the features of the original text, the propagating user and the propagating structure. Firstly, a propagation graph was constructed based on the propagation structure and the propagation users, and the combination of multiple user attributes was used as the propagation node feature. Then, multiple graph convolutional networks were used to learn the representation of the propagation graph with different user attribute combinations, and BERT model was used to learn the feature representation of the original text content. Finally, the fusion with the features learned by the graph convolutional network was used to detect rumors. A large number of experiments using publicly available Weibo data sets showed that the GCNs-BERT model was significantly better than the baseline method. In addition, the generalization ability experiment of GCNs-BERT model was conducted on the novel coronavirus epidemic data set. The training sample size of this data set was only 1/5 of that of the public Weibo data set, and the accuracy rate was still 92.5%, which proved that the model had good generalization ability.

Keywords: rumor detection; graph convolutional network; propagation graph; propagation user; feature fusion