



情报杂志
Journal of Intelligence
ISSN 1002-1965, CN 61-1167/G3

《情报杂志》网络首发论文

题目：基于推文传播模式与跨模态特征的网络谣言检测研究
作者：彭竞杰，顾益军，张岚泽
网络首发日期：2024-03-09
引用格式：彭竞杰，顾益军，张岚泽. 基于推文传播模式与跨模态特征的网络谣言检测研究[J/OL]. 情报杂志. <https://link.cnki.net/urlid/61.1167.G3.20240307.1239.014>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于推文传播模式与跨模态特征的网络 谣言检测研究*

彭竞杰 顾益军 张岚泽

(中国人民公安大学信息安全学院 北京 100038)

摘要: [研究目的] 为了有效治理网络谣言,减少网络谣言对社会稳定带来的威胁,提出充分整合帖子的多模态信息和传播模式信息对谣言进行精准识别。[研究方法] 提出融合推文传播模式信息与跨模态特征的网络谣言检测模型(PPCMRD)。在推文传播特征挖掘方面,首先通过推断潜在连接补全推文传播图,接着采用双向标签图注意力模块编码推文的多个传播模式,然后通过传播模式信息融合模块捕获模式特征间的互补信息,得到帖子的传播特征;在整合多模态特征方面,该模型将帖子的文本、图像和推文传播特征集成在一起,采用跨模态共同注意力机制捕捉不同模态信息间的互补关系,得到帖子的最终嵌入表示,判断是否是谣言。[研究结论] 在两个公开数据集上的实验结果表明,PPCMRD模型能够有效地检测谣言,并优于当前的基线模型。

关键词: 网络谣言;谣言检测;推文传播模式信息;跨模态特征融合;双向标签图注意力机制

中图分类号: G252

Research on Online Rumor Detection Based on Tweet Propagation Patterns and Cross-Modal Features

Peng Jingjie Gu Yijun Zhang Lanze

(Department of Information and Network Security, People's Public Security University of China,
Beijing 100038)

Abstract: [Research purpose] To effectively manage online rumors and reduce the threat of online rumors to social stability, we propose to fully integrate the multimodal information of tweets and the propagation pattern information to accurately identify rumors. [Research method] We propose a rumor detection model (PPCMRD) that integrates tweet propagation pattern information and cross-modal features. In terms of tweet propagation feature mining, the first step is to complement the tweet propagation graph by inferring potential connections, followed by encoding multiple propagation patterns of tweets using the bidirectional signed graph attention module, and then capturing the complementary information between pattern features through the propagation pattern information fusion module to obtain the propagation features of the tweet. In terms of integrating the multimodal features, this model integrates the text, image, and tweet propagation features of the tweet, and employs the cross-modal co-attention mechanism to capture the complementary relationship between different modal information and get the final embedding representation of the tweet to determine whether it is a rumor or not. [Research conclusion] The experimental results on two public datasets demonstrate that the proposed approach could effectively detect rumors and outperforms the current baseline models.

Key words: online rumor; rumor detection; tweet propagation patterns; cross-modal features fusion; bidirectional signed graph attention mechanism

基金项目:本文系项目“2023JC02”和“中央高校基本科研业务费专项资金资助项目”(编号:2023JKF01ZK14)研究成果。

作者简介:彭竞杰,男,1999年生,硕士研究生,研究方向:复杂网络、谣言检测、深度学习等;顾益军,男,1968年生,博士,教授,博士生导师,研究方向:复杂网络、深度学习;张岚泽,男,1998年生,硕士研究生,研究方向:复杂网络、深度学习等。

通信作者:顾益军

0 引言

随着互联网技术的发展,网络社交媒体上的帖子不再只限于纯文本形式传播,演变为包含文本、图像等多种内容的多模态帖子。有研究表明,带有图片的推文转发次数是纯文本推文的 11 倍以上^[1],大大增加了谣言信息的误导性与迷惑性,严重危害着个人生活、社会稳定等各个方面,因此迫切需要能够高效准确地检测网络谣言的方法。

以往的谣言检测方法主要基于推文内容信息或传播结构信息,并结合使用特征工程或深度学习的方法对推文进行检测^[2],例如 SAFE^[3]和 FakeBERT^[4]分别使用 TextCNN^[5]和 BERT^[6,7,8]编码推文的文本信息进行检测;MVAE^[9]和 EANN^[10]基于 DNN 编码推文的文本和图像信息进行检测;GCNFN^[11]和 GNNCL^[12]使用 GCN^[13]编码推文的传播结构进行检测。帖子的多模态数据蕴含着丰富的指向性信息^[1,14,15],不同模态的信息之间会起到互补的作用^[16,17],然而,上述方法仅使用了推文某一方面的特征过于局限。

此外,谣言的深度传播模式^[14]和广度散布模式^[15]现已被证实是社交媒体谣言的重要特征,可以有效辅助检测谣言。现有的基于传播特征的谣言检测方法,如 Bi-GCN^[18]和 EBGCN^[19],从自顶向下的深度传播模式和自底向上的广度散布模式对推文传播图进行编码,但它们对两个传播模式特征仅采用直接拼接的方式进行融合,限制了对不同模式特征之间的互补关系进行充分捕捉,未能从全局角度挖掘谣言与非谣言之间的结构特征差异,影响了谣言检测模型的性能。

为了有效解决上述问题,本文提出了融合推文传播模式信息与跨模态特征的网络谣言检测模型(Integrating Tweet Propagation Patterns Information and Cross-Model Features for Rumor Detection Model, PPC-MRD),通过有效建模融合推文传播模式并充分整合推文多模态内容信息,以准确地识别网络谣言。

1 相关研究

当前的谣言检测方法按照其所关注的特征不同可划分为基于推文消息内容的方法和基于推文传播结构的方法。

1.1 基于推文消息内容的谣言检测方法

社交媒体的帖子中包含了大量的内容信息,对于谣言内容特征的挖掘与提取一直以来都是学者们关注的重点。早期的机器学习谣言检测方法主要通过手工方式从推文消息中提取特征进行检测。Yang 等人^[20]通过提取新闻文本、转发次数和地理位置等特征信息,训练了使用径向基核函数的 SVM 分类器进行谣言检

测。Kwon 等人^[21]考虑了推文发布的特殊时间节点的特征,提出了基于时间特征的随机森林检测模型。Ma 等人^[22]利用时间序列建模技术,通过考虑话题事件的生命周期构建融合了社会背景信息的谣言检测模型。Zhao 等人^[23]通过挖掘谣言中的关键字特征,并结合 K-means 聚类和 SVM 分类器进行检测。上述方法虽然取得了一定进展,但过于依赖手工的特征工程,从而导致模型泛化能力不足,效率低。

随着深度学习技术的不断发展,一些学者开始将深度学习应用于谣言检测研究中。Ma 等人^[24]出了基于循环神经网络(RNN)的谣言检测模型,利用推文的时间特征来进行谣言检测。Liu 等人^[25]将谣言的转发与评论内容信息视做一个时间序列,利用 RNN 和 CNN 建模该序列进行谣言检测。Lu 等人^[26]提出利用源推文的短文本内容、转发用户序列以及用户资料进行谣言检测。随着图神经网络的进一步发展, Yuan 等人^[27]利用注意力机制联合编码事件的局部和全局语义信息,构建异质图进行谣言检测。Ren 等人^[28]提出基于异质图注意力网络 HGAT 的谣言检测模型,经过节点级和语义级的注意力机制对推文、发布者和主题之间的成对关系进行建模。Zheng 等人^[29]构建了一个多模态特征增强网络,将推文文本特征、传播结构特征等融合进一个框架中进行谣言检测。然而,上述大部分方法使用的推文内容信息较为局限,仅使用了文本信息,而现在大多数谣言都是文本、图片等内容结合的多模态形式。有学者指出,不同模态的信息之间会起到互补的作用^[16,17]。上述谣言检测方法尚未有效整合推文的内容信息和传播信息,导致丢失了很多重要有效的特征,对谣言不同模态信息之间的关联关系挖掘不够充分。

1.2 基于推文传播结构的谣言检测方法

社交媒体中的转发、评论以及回复均有着明确的指向对象,源帖与评论帖、评论帖与评论帖之间的转发与回复关系明确体现了推文传播结构,因此,一些学者从推文传播建模的角度对谣言与非谣言事件进行了深入挖掘。Ma 等人^[30]基于双向树形递归神经网络来捕获谣言的语义信息和传播特征信息。Bian 等人^[18]利用双向图卷积神经网络同时学习谣言自顶向下的传播模式和自底向上的散布模式,捕捉谣言的传播结构信息。Yang 等人^[31]利用图神经网络分析推文之间的结构相似性进行谣言检测。Luo 等人^[32]通过将源推文的传播结构嵌入到向量空间中,使用预训练的 Transformer 提取源推文的上下文语义特征进行谣言检测。Wei 等人^[19]通过研究推文传播过程中的随机性,设计了基于边一致性训练框架的谣言检测模型。Ran 等人^[33]提出了多通道图注意力网络谣言检测方法,通过

并行构建三个子图学习推文传播结构。Wu 等人^[34]基于图卷积神经网络对推文的多个传播关系进行编码,构建了基于图感知的多关系图卷积神经网络谣言检测模型。然而,上述方法中大部分只考虑了推文消息单一方向的深度传播,一定程度上忽略了其在传播范围内广泛散布的特性,以及推文的转发、评论对源推文的隐性反馈,导致获得的推文传播特征质量一般。此外,现有的研究方法尚未有效整合推文传播过程中的不同传播模式特征,从全局角度挖掘谣言与非谣言之间的区别。例如,Bian 等人^[18]提出的双向图卷积神经网络虽然能够同步编码推文的深度传播模式和广度散布模式,但最终只通过直接拼接的方式得到推文的传播模式特征,而没有考虑两个传播模式下相同位置节点之间的模式信息的互补关系。

2 模型设计

为同时解决传统谣言检测方法对推文多模态信息挖掘不充分、结合不完善,以及对不同的推文传播模式信息整合能力不强的问题,本文提出了融合推文传播模式信息与跨模态特征的网络谣言检测模型 (PPCMRD),模型整体框架如图 1 所示。该模型将推文的文本特征、图像特征和传播特征集成在一个统一的框架中,采用跨模态共同注意力机制捕获不同模态信息之间的互补关系。同时,在推文传播特征挖掘方面,设计了双向标签图注意力模块和推文传播模式信息融合模块来编码不同的推文传播模式,以充分捕捉不同传播模式之间的互补信息。本章将分模块重点介绍该模型。

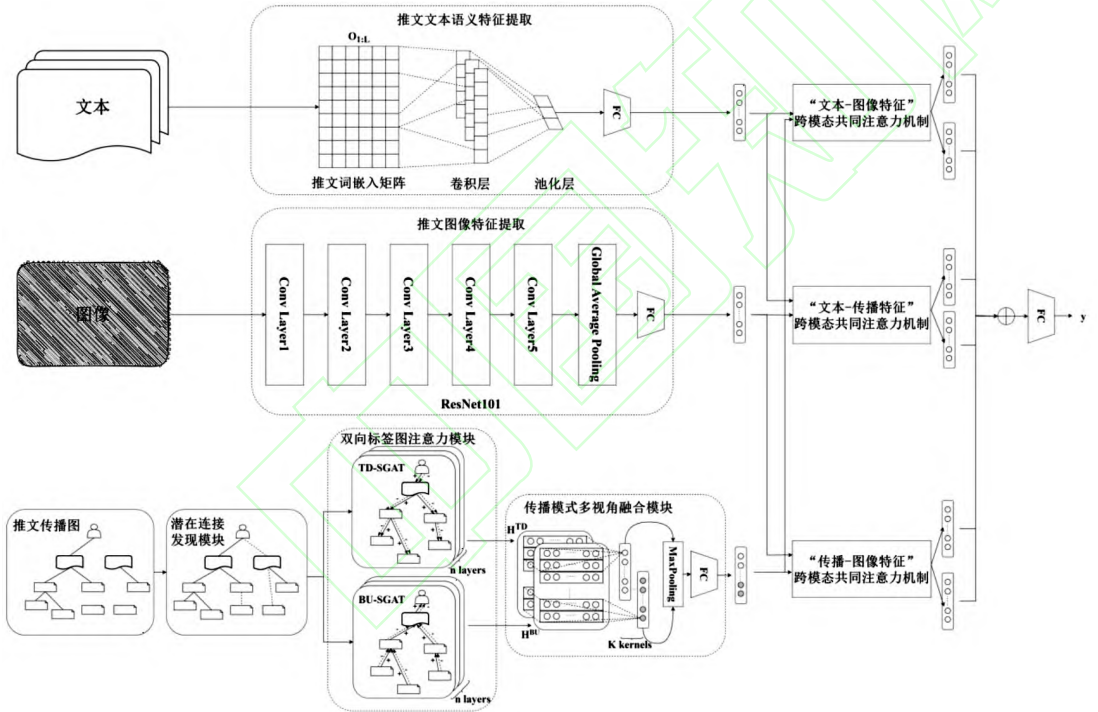


图 1 PPCMRD 模型结构

2.1 谣言检测任务定义

设推文数据集 $P = \{p_1, p_2, \dots, p_n\}$ 是网络社交媒体上一组既有文字又有图像的多媒体帖子,对于每个帖子 $p_i \in P$, $p_i = \{t_i, v_i, u_i, c_i\}$, 其中 t_i , v_i 和 u_i 分别表示发布帖子的推文文本、图像和用户。 $c_i = \{c_i^1, c_i^2, \dots, c_i^l\}$ 表示推文 p_i 的评论集。为了描述推文在社交媒体上的传播行为,本文构建推文传播图 $G = \{V, A, E\}$, 其中 V 表示节点集合,包括用户节点、推文文本节点和评论节点, $A \in \{0, 1\}^{|V| \times |V|}$ 表示节点之间的邻接矩阵,用于描述节点之间的关联关系,包括发布、评论等关系, E 表示边的集合。我们将谣言检测任务定义为一个二分类任务, $y \in \{0, 1\}$ 表示类别标签,其中 $y=1$ 表示谣言, $y=0$ 表示非谣言。融合推文传播模式信息与

跨模态特征的网络谣言检测任务是学习一个函数 $F(p_i) = y$ 来预测给定帖子 p_i 的标签。

2.2 推文文本语义特征提取

本文采用卷积神经网络结合池化操作来提取推文文本的语义特征,该处理过程如图 2 所示。首先,针对每个帖子 p_i , 将文本 t_i 进行填充或截断,使包含不同单词数量的文本具有相同的长度 L , 处理后的文本表示如公式(1)所示:

$$O_{1:L} = \{o_1^i, o_2^i, \dots, o_L^i\} \quad (1)$$

其中 $o \in \mathbb{R}^d$, d 是单词向量的嵌入维度, o_j^i 表示 t_i 的第 j 个单词的嵌入表示。然后,本文在词嵌入矩阵 $O_{1:L}$ 上应用卷积层进行卷积,得到对应的特征图 s_j , 其中 k 表示感受野的大小,本文将处理得到的特征图用

$s^i = \{s_{i1}, s_{i2}, \dots, s_{i(L-k+1)}\}$ 表示。接着对 s_i 进行最大池化操作, 得到 $\hat{s}^i = \max(s^i)$ 。本文使用具有不同感受野大小 $k \in \{3, 4, 5\}$ 的卷积核以获取推文文本不同粒度的语义特征, 最后将所有卷积核的输出拼接在一起输入到全连接层中, 得到推文文本 t_i 的语义特征表示 R_t^i , 如公式(2)所示:

$$R_t^i = \sigma(W_t [\hat{s}_{k=3}^i \parallel \hat{s}_{k=4}^i \parallel \hat{s}_{k=5}^i] + b_t) \quad (2)$$

其中, \parallel 表示拼接操作, W_t 和 b_t 是全连接层可学习得到的参数, $\sigma(\cdot)$ 表示 sigmoid 激活函数。

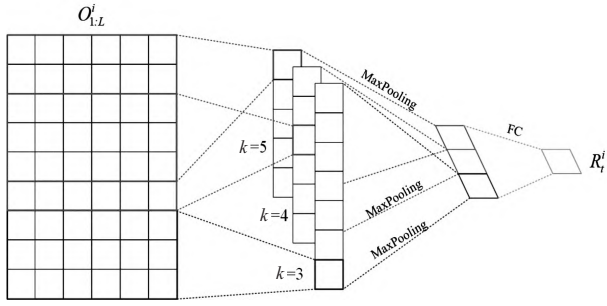


图2 推文文本语义特征提取流程图

2.3 推文图像特征提取

本文使用在 ImageNet 数据集上预训练得到的 ResNet101 模型^[35]来提取图像 v_i 的特征。首先, 本文提取 ResNet101 模型处理图像 v_i 得到的倒数第二层的输出 V_r^i , 接着将其输入到一个全连接层得到帖子的图像特征向量 R_v^i , 该特征向量维度与文本特征向量一致。该处理过程如公式(3)所示:

$$R_v^i = \sigma(W_v V_r^i + b_v) \quad (3)$$

其中, W_v 和 b_v 是全连接层可学习得到的参数, $\sigma(\cdot)$ 表示 sigmoid 激活函数。

2.4 推文传播特征提取

2.4.1 推文传播图潜在连接发现

由于用户隐私问题或平台上的数据爬取限制, 可用的推文传播图中很可能缺少一些节点间的重要连接, 因此需要补全推文传播图上的潜在连接以实现更精准的谣言检测。本文基于图自编码器^[36]对推文传播图进行补全。具体来说, 本文给定推文传播图的邻接矩阵 A 和节点特征嵌入矩阵 $X \in \mathbb{R}^{|V| \times d}$, 其中 d 是每个节点特征向量的维度大小。在 X 中包含推文文本、评论和用户三种类型的节点, 本文首先使用在领域特定的推文语料库上预训练好的 300 维词向量^[27]作为推文文本节点和评论节点的初始特征向量; 其次, 为了挖掘并建模对谣言检测具有明确指向性的用户传播偏好信息, 本文使用用户历史发布的推文文本特征向量的平均值作为用户节点的初始特征向量。

图自编码器 GAE 首先使用图卷积神经网络 GCN^[13]来计算每个节点 n_i 的节点表特征表示 z_i , 然后通过计算得到节点 n_i 和 n_j 之间存在边连接的概率 \hat{a}_{ij} ,

该计算过程如公式(4)-(6)所示:

$$Z = \text{GCN}(X, A) \quad (4)$$

$$z_i = Z_{i,:} \quad (5)$$

$$\hat{a}_{ij} = \sigma(z_i^T z_j) \quad (6)$$

其中, Z 是图卷积神经网络 GCN 输出的节点特征嵌入矩阵, Z 的第 i 行是节点 n_i 的特征向量 z_i , \hat{a}_{ij} 表示节点 n_i 和 n_j 之间存在边连接的概率, σ 表示 sigmoid 函数。

接着, 若节点 n_i 和 n_j 之间的边连接存在概率 \hat{a}_{ij} 值大于 0.5, 则推断节点 n_i 和 n_j 之间存在潜在的边连接, 如公式(7)所示:

$$e_{ij} = \begin{cases} 0, & \text{if } \hat{a}_{ij} \leq 0.5 \\ 1, & \text{otherwise} \end{cases} \quad (7)$$

然后, 本文通过推断得到的潜在边连接补全初始的推文传播图邻接矩阵 $A \in \mathbb{R}^{|V| \times |V|}$ 。其中 a_{ij} 表示邻接矩阵 A 中的元素, 其中 $a_{ij} = 1$ 表示节点 n_i 和 n_j 之间存在边连接, 反之 $a_{ij} = 0$ 表示节点 n_i 和 n_j 之间不存在边连接。补全后的推文传播图邻接矩阵 A' 中的元素 a'_{ij} 定义如公式(8)所示:

$$a'_{ij} = \begin{cases} 0, & \text{if } e_{ij} = 0 \text{ and } a_{ij} = 0 \\ 1, & \text{otherwise} \end{cases} \quad (8)$$

然而, 基于节点的原始特征向量进行的潜在连接发现工作可能会受到伪装特征的影响, 致使后续生成的邻接矩阵质量受到影响。因此, 本文针对该潜在连接发现过程进行强化学习, 并将该学习过程中的损失 L_l 引入最终的损失函数。该训练过程通过最小化重建的邻接矩阵和真实邻接矩阵之间的交叉熵, 以不断迭代得到质量更高的推文传播邻接矩阵, 如公式(9)所示:

$$L_l = \sum_{i \in V, j \in V} (-a_{ij} \log \hat{a}_{ij} - (1 - a_{ij}) \log(1 - \hat{a}_{ij})) \quad (9)$$

2.4.2 基于双向标签图注意力机制的邻域信息聚合

与传统的图注意力网络^[37]不同, 本文构建了一个新型的双向标签图注意力网络来捕捉不同的推文传播模式信息。本文通过构建自顶向下 (Top-Down, TD) 和自底向上 (Bottom-Up, BU) 两个传播模式下的推文传播子图, 同时编码推文在深度传播和广度散布两个方向上的重要传播模式特征, 并引入带有正负符号标签的注意力机制捕捉目标节点与其邻域节点之间的正负关联关系, 以获得更好的节点传播特征表示。

图注意力网络方法的关键在于邻域信息聚合操作, 对于节点 n_i 及其邻域节点集 $N_i = \{n_1, n_2, \dots, n_{|N_i|}\}$, 本文首先计算节点 n_i 与其邻域节点集 N_i 中的每个节点之间的注意力权重系数, 得到注意力权重系

数集合 $\varepsilon_i = \{e_{i1}, e_{i2}, \dots, e_{i|N_i|}\}$, 注意力权重系数计算方法如公式(10)所示:

$$e_{ij} = \text{LeakyReLU}(\hat{a} [W_1 x_i \parallel W_2 x_j]) \quad (10)$$

其中, \parallel 表示拼接操作, \hat{a} 、 W_1 和 W_2 是可学习得到的参数, x_i 和 x_j 分别表示节点 n_i 和 n_j 的嵌入向量, 其中 $j \in \{1, 2, \dots, |N_i|\}$ 且 $n_j \in N_i$ 。

通常来说, 图注意力网络使用 softmax 函数对注意力权重系数进行归一化处理。事实上, 节点之间的注意力权重系数可能包含正负关系, 节点 n_i 和节点 n_j 之间的注意力权重系数 e_{ij} 有可能是一个负值, 其值在经过 softmax 函数归一化处理后会变得很小, 直接使用 softmax 函数进行操作可能会忽略负值带来的影响。例如对于特定节点 n_i , 设其与邻域节点之间的注意力权重系数为 $\varepsilon_i = \{0.3, 0.8, -0.9, -0.2\}$, 经过 softmax 函数归一化处理后, 该权重系数变为 $\varepsilon_i' = \{0.28, 0.46, 0.08, 0.17\}$, 由此可见, 对应权重系数为“-0.9”的邻居节点对目标节点聚合输出的贡献最小。然而, “-0.9”表示两个节点的特征向量方向是相反的, 这种负关系对于谣言检测工作也具有意义, 反映了推文传播过程中的某种伪装行为, 例如对某个非谣言推文帖子发表持反对意见的评论的行为^[38]。然而, 现有的图注意力网络忽略了图中可能存在的同构性假设失灵问题, 模糊了上述关键的呈负相关的高频信号。

为了解决上述问题, 本文受 QSAN 方法^[39]启发, 设计了双向标签图注意力网络 Bi-SGAT (Bidirectional Signed Graph Attention Network)。首先, Bi-SGAT 构建自顶向下 (Top-Down, TD) 和自底向上 (Bottom-Up, BU) 两个传播模式下的推文传播子图, 两个推文传播子图上的边方向相反, 二者的邻接矩阵互为转置, 如图 3 所示。

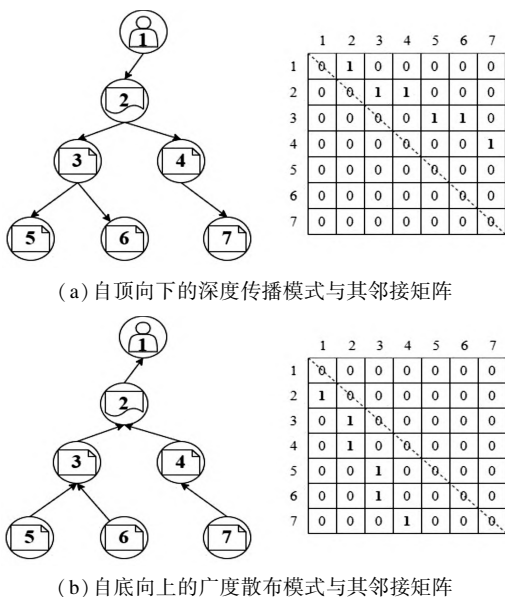


图3 不同传播模式下的消息传播图与其邻接矩阵

接着, Bi-SGAT 使用带有正负符号标签的注意力机制来编码节点之间的正负关联关系。具体而言, 对于自顶向下传播模式中的节点 n_i , 本文将其与在该传播模式下指向节点 n_i 的邻域节点 N_i^{TD} 之间的注意力权重系数的相反数记为 $\widehat{\varepsilon}_i^{TD} = -\varepsilon_i^{TD}$, 接着使用 softmax 函数计算 ε_i^{TD} 和 $\widehat{\varepsilon}_i^{TD}$ 归一化后的权重系数 $\varepsilon_i^{TD'}$ 和 $\widehat{\varepsilon}_i^{TD'}$, 计算过程如公式(11)-(12)所示:

$$\varepsilon_i^{TD'} = \text{softmax}(\varepsilon_i^{TD}) \quad (11)$$

$$\widehat{\varepsilon}_i^{TD'} = \text{softmax}(\widehat{\varepsilon}_i^{TD}) \quad (12)$$

为了有效捕捉节点之间的正负关联关系, 本文使用 $\varepsilon_i^{TD'} \in \mathbb{R}^{|N_i^{TD}|}$ 和 $-\widehat{\varepsilon}_i^{TD'} \in \mathbb{R}^{|N_i^{TD}|}$ 分别计算目标节点 n_i 与其在自顶向下传播模式下的邻域节点 N_i^{TD} 在两组注意力权重系数下的特征向量加权和。然后, 将上述计算得到的两个向量进行拼接, 同时, 为了有效利用节点原始特征向量中蕴含的指示性信息, 拼接上节点的原始特征向量, 并通过一个全连接层。经过两层 Bi-SGAT 网络得到自顶向下传播模式下的节点嵌入表示 \hat{x}_i^{TD} 。以节点 n_i 为例, 其嵌入向量 \hat{x}_i^{TD} 的计算方式如公式(13)-(15)所示:

$$x_{i1}^{TD} = \sigma((\varepsilon_i^{TD'} * X_j^{TD} \parallel -\widehat{\varepsilon}_i^{TD'} * X_j^{TD} \parallel x_i) W_1^{TD}) \quad (13)$$

$$x_{i2}^{TD} = \sigma((\varepsilon_i^{TD'} * X_{j1}^{TD} \parallel -\widehat{\varepsilon}_i^{TD'} * X_{j1}^{TD} \parallel x_{i1}^{TD}) W_2^{TD}) \quad (14)$$

$$\hat{x}_i^{TD} = x_{i2}^{TD} \quad (15)$$

其中, $W_1^{TD} \in \mathbb{R}^{2d \times d}$ 和 $W_2^{TD} \in \mathbb{R}^{2d \times d}$ 是两个全连接层的权重矩阵, $\sigma(\cdot)$ 表示激活函数 ReLU , $X_j^{TD} \in \mathbb{R}^{|N_i^{TD}| \times d}$ 和 $X_{j1}^{TD} \in \mathbb{R}^{|N_i^{TD}| \times d}$ 分别表示在自顶向下传播模式下指向节点 n_i 的邻域节点集 N_i^{TD} 的初始特征嵌入矩阵和经过一层 Bi-SGAT 网络处理后的特征嵌入矩阵。同理, 节点 n_i 经过与上述流程相似的操作可得自底向上传播模式下的节点嵌入表示 \hat{x}_i^{BU} , 如公式(16)-(20)所示:

$$\varepsilon_i^{BU'} = \text{softmax}(\varepsilon_i^{BU}) \quad (16)$$

$$\widehat{\varepsilon}_i^{BU'} = \text{softmax}(\widehat{\varepsilon}_i^{BU}) \quad (17)$$

$$x_{i1}^{BU} = \sigma((\varepsilon_i^{BU'} * X_j^{BU} \parallel -\widehat{\varepsilon}_i^{BU'} * X_j^{BU} \parallel x_i) W_1^{BU}) \quad (18)$$

$$x_{i2}^{BU} = \sigma((\varepsilon_i^{BU'} * X_{j1}^{BU} \parallel -\widehat{\varepsilon}_i^{BU'} * X_{j1}^{BU} \parallel x_{i1}^{BU}) W_2^{BU}) \quad (19)$$

$$\hat{x}_i^{BU} = x_{i2}^{BU} \quad (20)$$

其中, $X_j^{BU} \in \mathbb{R}^{|N_i^{BU}| \times d}$ 和 $X_{j1}^{BU} \in \mathbb{R}^{|N_i^{BU}| \times d}$ 分别表示在

自底上传播模式下指向节点 n_i 的邻域节点集 N_i^{BU} 的初始特征嵌入矩阵和经过一层 Bi-SGAT 网络处理后的特征嵌入矩阵。

2.4.3 推文传播模式信息融合

经过双向标签图注意力网络 Bi-SGAT 处理后,推文传播图中的每个节点分别具有自顶向下传播模式下的嵌入向量 \hat{x}^{TD} 和自底上传播模式下的嵌入向量 \hat{x}^{BU} ,推文文本节点 n_i 与其邻域节点 N_i 构成的推文传播子图可以看作是一张彩色图像,每个方向的推文传播模式可以看作是一个通道,每一行节点特征向量可以看作是一个像素。因此,推文文本节点 n_i 与其邻域节点 N_i 构成的推文传播子图的节点嵌入矩阵 X_i 如公式 (21) 所示:

$$X_i = \begin{bmatrix} \hat{x}_1^{TD}, \hat{x}_2^{TD}, \dots, \hat{x}_n^{TD} \\ \hat{x}_1^{BU}, \hat{x}_2^{BU}, \dots, \hat{x}_n^{BU} \end{bmatrix} \in \mathbb{R}^{2 \times n \times d} \quad (21)$$

将给定的节点嵌入矩阵 X_i 作为输入,构建有 K 个卷积核 $W_k \in \mathbb{R}^{2 \times t \times d}$ 的卷积神经网络,在包含 t 个节点的滑动窗口上进行卷积操作,生成新的特征嵌入表示,如公式 (22) 所示:

$$r_l^k = \sigma(W_k * X_{l:l+t-1} + b_k) \quad (22)$$

其中, l 表示卷积核 W_k 进行每次卷积操作的初始位置, W_k 和 b_k 是可学习得到的参数, $*$ 表示多通道卷积操作, $\sigma(\cdot)$ 表示激活函数 ReLU。本文将 K 个卷积核应用于节点嵌入矩阵 X_i 中所有包含 t 个节点的滑动窗口中。其中,在 W_k 卷积核下得到的特征图如公式 (23) 所示:

$$r^k = [r_1^k, r_2^k, \dots, r_{n-t+1}^k] \quad (23)$$

然后,本文对上述每个卷积核下得到的特征图进行最大池化操作,获取与该卷积核相对应的最重要的特征 $\hat{r}^k = \max\{r^k\}$ 。接着将所有卷积核对应的特征图的最大值拼接在一起,并通过一个全连接层处理,得到融合推文传播模式信息的推文文本节点 n_i 的最终嵌入表示 R_g^i ,具体过程如公式 (24) 所示:

$$R_g^i = \sigma((\hat{r}^1 \parallel \hat{r}^2 \parallel \dots \parallel \hat{r}^K) W_r + b_r) \quad (24)$$

其中, W_r 和 b_r 是全连接层可学习得到的参数, $\sigma(\cdot)$ 表示激活函数 ReLU。

2.5 基于共同注意力机制的推文传播模式信息与跨模态特征融合

本文基于帖子的推文文本特征 R_t^i 、图像特征 R_v^i 和推文传播特征 R_g^i 三种信息设计了基于共同注意力机制^[40]的跨模态特征融合方法,对帖子不同模态信息之间的关联关系进行充分挖掘。该方法通过学习不同模态特征之间的注意力权重系数以增强跨模态的特征嵌入表示。具体来说,对于每一个模态,本文首先使用多头自注意力机制^[41]来增强单一模态的特征表示。

例如,对于推文文本特征 R_t^i ,本文通过 $Q_t^i = R_t^i W_t^Q$, $K_t^i = R_t^i W_t^K$ 和 $V_t^i = R_t^i W_t^V$ 分别计算得到其查询矩阵 Q_t^i 、键矩阵 K_t^i 和值矩阵 V_t^i 。其中, W_t^Q , W_t^K , $W_t^V \in \mathbb{R}^{d \times \frac{d}{h}}$ 分别表示查询权重矩阵、键权重矩阵和值权重矩阵, h 表示注意力头的数量。推文文本模态 R_t^i 的多头自注意力特征表示 Z_t^i 的计算方法如公式 (25) 所示:

$$Z_t^i = \left(\big\|_{h=1}^H \text{softmax}\left(\frac{Q_t^i K_t^{i^T}}{\sqrt{d}}\right) V_t^i \right) W_t^O \quad (25)$$

其中, h 表示第 h 个注意力头, $W_t^O \in \mathbb{R}^{d \times d}$ 表示输出线性变换矩阵,本文对图像特征 R_v^i 和推文传播特征 R_g^i 执行与上述相同的操作,得到相应的多头自注意力特征表示 Z_v^i 和 Z_g^i 。

然后,本文使用共同注意力机制生成增强的跨模态特征嵌入表示。举例来说,本文对帖子 p_i 的推文文本特征和图像特征应用“文本-图像”共同注意力机制,与上述自注意力机制的操作类似,但是将 R_t^i 替换为 Z_v^i 以获得查询矩阵 Q_v^i ,将 R_t^i 替换为 Z_t^i 以获得键矩阵 K_t^i 和值矩阵 V_t^i ,从而得到跨模态的增强特征嵌入表示 Z_{vt}^i ,计算方法如公式 (26) 所示:

$$Z_{vt}^i = \left(\big\|_{h=1}^H \text{softmax}\left(\frac{Q_v^i K_t^{i^T}}{\sqrt{d}}\right) V_t^i \right) W_{vt}^O \quad (26)$$

Z_{vt}^i 表示基于推文文本特征和图像特征的关联关系,融合图像特征增强后的文本特征嵌入表示。与上述共同注意力处理方法相似,通过将 R_v^i 替换为 Z_t^i 以获得查询矩阵 Q_t^i ,将 R_v^i 替换为 Z_v^i 以获得键矩阵 K_v^i 和值矩阵 V_v^i ,从而可以获得融合文本特征增强的图像特征嵌入表示 Z_{tv}^i 。

本文在 R_t^i 、 R_v^i 和 R_g^i 三种类型的特征之间两两执行上述的跨模态共同注意力机制,最终得到 6 个跨模态增强的特征表示: Z_{vt}^i 、 Z_{tv}^i 、 Z_{gt}^i 、 Z_{tg}^i 、 Z_{gv}^i 和 Z_{vg}^i 。接着,将其拼接起来作为帖子 p_i 最终的多模态特征嵌入表示,如公式 (27) 所示:

$$Z^i = \text{concat}(Z_{vt}^i, Z_{tv}^i, Z_{gt}^i, Z_{tg}^i, Z_{gv}^i, Z_{vg}^i) \quad (27)$$

2.6 推文帖子分类

本文将帖子 p_i 的多模态特征嵌入表示 Z^i 输入到一个全连接层中,以预测 p_i 是否是谣言,如公式 (28) 所示:

$$\hat{y}_i = \text{softmax}(W_c Z^i + b_c) \quad (28)$$

其中, W_c 和 b_c 表示全连接层可训练学习到的参数, \hat{y}_i 表示预测帖子 p_i 是谣言的概率。接着,本文使用负对数似然法来优化帖子 p_i 的预测标签与其真实标签之间的损失函数,其中 Y_{train} 表示训练数据的真实标签。该过程如公式 (29) 所示:

$$L_c = - \sum_{y_i \in Y_{train}} (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (29)$$

将 2.4.1 节中邻接矩阵强化学习过程的损失函数 L_l 引入可得到模型最终的损失函数 L , 如公式 (30) 所示:

$$L = \lambda_c L_c + \lambda_l L_l \tag{30}$$

其中, λ_c 和 λ_l 表示用于平衡损失 L_c 和 L_l 的参数。

3 实验与分析

3.1 实验数据集

本文在 Weibo^[29,42] 和 PHEME^[29,43] 两个公开的谣言检测数据集上验证本文模型的效果。其中, Weibo 数据集是从新浪微博不实信息举报平台爬取到的中文谣言数据集。PHEME 数据集由 Twitter 上关于五个突发事件新闻事件的推文帖子组成。上述两个数据集中的帖子包含推文文本、图像和评论三类数据, 移除了其中没有任何文本或图像的帖子。两个数据集的统计信息如表 1 所示:

表 1 数据统计

数据集	推文总数	非谣言推文数量	谣言推文数量	图片数量	用户数量	评论数量
PHEME	2018	1428	590	2018	894	7388
Weibo	1467	877	590	1467	985	4534

3.2 对比实验设计

为了验证 PPCMRD 模型的性能, 本文选取基于推文文本特征、图像特征和传播模式特征的几种经典模型作为基线方法同本文模型进行对比:

QSAN^[39]: 一种基于量子概率的符号注意力虚假信息检测模型, 将量子驱动的文本编码和一种符号注意力机制集成在一个统一的框架中进行虚假信息检测。

SAFE^[3]: 一种联合利用文本和图像多模态信息和模态相关性信息的虚假新闻检测模型。

MVAE^[9]: 一种基于双峰变分自编码器的多模态虚假新闻检测模型。

EANN^[10]: 一种基于生成对抗网络的虚假新闻检测模型, 通过学习不同新闻之间的共同特征, 对新突发事件中的虚假新闻进行检测。

GLAN^[27]: 一种联合编码局部语义和全局结构信息的异质图谣言检测模型。

KMGCN^[44]: 一种基于知识驱动的多模态图卷积神经网络谣言检测模型。

Bi-GCN^[18]: 一种基于双向图卷积神经网络的谣言检测模型, 根据谣言传播的两个方向使用图卷积神经网络提取传播特征进行谣言检测。

EBGCN^[19]: 一种通过贝叶斯方法自适应地剔除推文传播结构中冗余关系的谣言检测模型。

PPCMRD: 本文提出的融合推文传播模式信息与

跨模态特征的网络谣言检测模型。

其中, QSAN 模型仅基于文本数据进行谣言检测, SAFE、MVAE 和 EANN 模型基于文本和图像数据进行谣言检测, GLAN、KMGCN、BiGCN 和 EBGCN 模型基于推文的传播特征进行谣言检测, 本文提出的 PPCMRD 模型创新性地将文本特征、图像特征和推文传播模式特征集成在一个统一的框架中进行谣言检测。

3.3 实验设置

本文提出的模型基于 Pytorch 框架实现, 按照 7:1:2 的比例将数据集划分为训练集、验证集和测试集, 实验中使用五折交叉验证, 并通过准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1 值 (F1-score) 等指标评估模型的性能。推文传播模式信息融合模块中的卷积滑动窗口大小 t 设置为 1, 卷积核数量 K 设置为 3, 跨模态共同注意力机制特征融合模块中的注意力头数 H 设置为 8。在模型训练阶段, 设置学习率为 0.002, batchsize 为 64, dropout 为 0.5, epoch 为 30。同时采用早停法, 使用最小化交叉熵损失函数和 Adam 算法^[45] 对训练参数进行优化, 并使用不固定的随机种子进行多次实验计算实验结果的平均值。

3.4 实验结果与分析

PPCMRD 模型与其他基线模型在两个数据集上的实验结果如表 2、表 3 所示。

表 2 PHEME 数据集实验结果

模型名称	Accuracy	Precision	Recall	F1-score
QSAN	0.7513	0.6997	0.6580	0.6687
SAFE	0.8149	0.7988	0.7950	0.7968
MVAE	0.7762	0.7349	0.7225	0.7277
EANN	0.7713	0.7139	0.7007	0.7044
GLAN	0.8332	0.8125	0.7713	0.7914
KMGCN	0.8123	0.7753	0.7532	0.7641
Bi-GCN	0.8471	0.8403	0.8343	0.8373
EBGCN	0.8299	0.8131	0.7929	0.8029
PPCMRD	0.8987	0.8820	0.8676	0.8748

表 3 Weibo 数据集实验结果

模型名称	Accuracy	Precision	Recall	F1-score
QSAN	0.7101	0.7102	0.6754	0.6758
SAFE	0.8495	0.8498	0.8495	0.8496
MVAE	0.7167	0.7052	0.7021	0.7034
EANN	0.8096	0.8019	0.7968	0.7987
GLAN	0.8244	0.8245	0.8087	0.8165
KMGCN	0.8616	0.8642	0.8561	0.8601
Bi-GCN	0.8672	0.8663	0.8661	0.8662
EBGCN	0.8312	0.8546	0.8176	0.8357
PPCMRD	0.9186	0.9165	0.9093	0.9129

实验结果表明, 本文提出的 PPCMRD 模型在谣言检测的准确率和精度上都显著优于其他基线模型, 对

实验结果的具体分析如下：

首先,由实验结果可知,基于图拓扑的谣言检测方法 (GLAN、KMGCN、Bi-GCN、EBGCN 和 PPCMRD) 效果普遍优于传统的深度学习方法 (QSAN、SAFE、MVAE 和 EANN)。推文传播特征提供了消息传播的上下文关联关系和消息流向信息,这使得模型能够更加全面地理解谣言和非谣言消息的传播模式,相比之下,传统方法缺少对这一关键特征的学习。因此,基于图拓扑的方法表现出了更优秀的效果。

其次,与不考虑推文传播模式的图神经网络模型 (GLAN 和 KMGCN) 相比,基于推文传播模式的谣言检测模型 (Bi-GCN、EBGCN 和 PPCMRD) 在两个数据集上的准确率平均高出 3.59% 和 2.94%。推文的传播模式和扩散模式提供了更加全面的传播信息,能够更好地捕捉包括推文发布者、源推文、推文传播路径以及对源推文的评论反馈等多个方面的信息,这种全面性有助于更好地理解谣言和非谣言的传播模式,从而提高对谣言的检测能力。

再次,融合推文传播模式信息的方法 (PPCMRD) 的效果明显优于直接拼接传播特征向量的方法 (Bi-GCN 和 EBGCN)。从不同方向对推文传播模式进行建模,就像在立体几何中通过整合物体的三视图才能准确识别物体形状一样。因此,基于卷积神经网络的推文传播模式融合方法能够从全局各个方向有效捕捉推文的传播模式信息,提高模型的检测能力。

最后,基于多模态信息的谣言检测方法 (PPCMRD) 明显优于其他基于单模态或多模态信息挖掘不

充分的方法。文本数据提供了丰富的语义信息,图像数据提供了视觉信息,PPCMRD 模型通过深入挖掘帖子的文本、图像和传播模式等特征,采用跨模态共同注意力机制对上述特征进行有效整合,有效捕捉到了不同模态之间的互补性信息,提高了谣言检测的准确性。

3.5 消融实验分析

为了进一步验证 PPCMRD 模型中各模块对模型性能的贡献,本文进行了一系列消融实验,主要包括 5 个部分:

PPCMRD w/o Image: 该模型表示移除对帖子的图像的编码过程,仅使用推文文本和推文传播模式信息进行谣言检测。

PPCMRD w/o Graph: 该模型表示移除对推文传播模式信息的编码过程,仅使用推文的文本和图像信息进行谣言检测。

PPCMRD w/o PED: 该模型表示移除推文传播图潜在连接发现模块,使用未经连接增强的推文传播图提取推文传播特征进行谣言检测。

PPCMRD w/o Bi-SGAT: 该模型表示移除双向标签图注意力模块,不考虑推文在不同方向下的传播模式以及推文传播图中节点的同构性假设失灵问题,使用普通的图卷积神经网络提取推文的传播特征进行谣言检测。

PPCMRD w/o MP: 该模型表示移除推文传播模式信息融合模块,将推文在自顶向下和自底向上方向下的传播模式特征拼接作为推文的最终传播特征表示进行谣言检测。

表 4 消融实验结果对比

模型	PHEME		Weibo	
	Accuracy	F1-score	Accuracy	F1-score
PPCMRD w/o Image	0.8649	0.8443	0.8847	0.8787
PPCMRD w/o Graph	0.8623	0.8436	0.8719	0.8689
PPCMRD w/o PED	0.8727	0.8491	0.8949	0.8895
PPCMRD w/o Bi-SGAT	0.8701	0.8452	0.8915	0.8882
PPCMRD w/o MP	0.8779	0.8523	0.8983	0.8922
PPCMRD	0.8987	0.8748	0.9186	0.9129

消融实验结果如表 4 所示,分析实验结果可以得出如下结论:

首先,移除图像信息后模型的准确率在 PHEME 和 Weibo 数据集上分别下降了 3.38% 和 3.39%,移除推文传播信息后分别下降了 3.64% 和 4.67%。这主要是由于图像数据提供的视觉信息有助于模型更全面地理解帖子的内容,而推文传播信息则补充了消息间的传播路径和关联关系,对于模型深入理解帖子的传播模式具有重要作用。因此,模型在综合建模帖子多模态信息的情况下,能够有效提高谣言检测的准确度。

其次,移除潜在连接预测模块后的模型在 PHEME

和 Weibo 数据集上的准确率分别下降了 2.60% 和 2.37%,证明了潜在连接预测模块的有效性。这主要是因为该模块通过预测推文传播图中的潜在连接,有效补充了推文传播图上的缺失连接,为模型提供了更准确、更全面的推文传播模式。

此外,移除双向标签图注意力模块后的模型在 PHEME 和 Weibo 两个数据集上的准确率分别下降了 2.86% 和 2.71%,证明了双向标签图注意力模块的有效性。这主要是因为双向标签图注意力模块同时编码谣言的深度传播和广度散布两个重要的传播模式特征,充分考虑了推文的转发、评论对源推文的隐性反

馈,并有效捕捉到目标节点与其邻域节点之间的正负关联关系,从而获得了更好的推文传播特征表示。

最后,移除推文传播模式信息融合模块后的模型在 PHEME 和 Weibo 两个数据集上的准确率分别下降了 2.08% 和 2.03%,证明了推文传播模式信息融合模块的有效性。这主要是因为该模块将推文传播图作为一个图像,每个节点的特征向量作为一个像素,多个传播模式作为一组图像通道,经过该模块的卷积神经网络处理有效捕捉到了不同模式间的互补信息,从而获得了更好的推文传播特征表示。

3.6 参数灵敏度分析

本文通过对比模型在不同学习率、batchsize、ResNet 层数以及不同滑动窗口大小 t 下模型的准确率对相关参数进行进一步分析。

学习率大小对于提升模型性能起着较为重要的作用^[46],在保持其他实验设置不变的情况下,本文在 0.001 和 0.005 之间的 5 组数据上进行实验探究学习率大小对模型的影响,结果如图 4 所示。

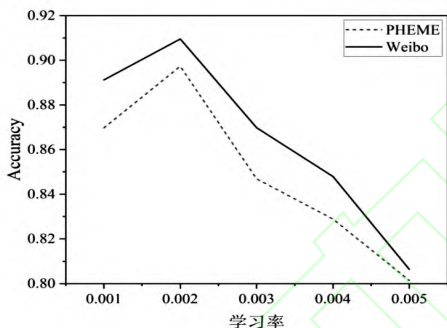


图4 模型准确率和学习率的关系

通过分析图 4 可知:随着学习率的增加,模型的准确率呈现先上升后下降的趋势,说明随着学习率的提高,模型能够更快地收敛到局部最优解,从而使准确率增加。然而,当学习率过高时,模型可能会陷入局部最优解附近的震荡状态,使准确率下降。因此,合适的学习率能够有效提高模型的检测性能。

当前,有研究发现合适的 batchsize 也有助于提升模型的性能^{[47][48]},为了验证 batchsize 对 PPCMRD 模型的影响,本文分别将 batchsize 设置为 16、32、64、128 进行实验,结果如图 5 所示。

通过分析图 5 可知:随着 batchsize 的增加,两个数据集上的准确率均呈现先上升后下降的趋势,在 batchsize 为 64 时,准确率均达到最高点。这可能是因为当 batchsize 过小时数据分布不均匀,导致模型的泛化能力下降,而 batchsize 过大时使用了更多的数据,可能导致一些重要信息被忽略掉。

此外,本文探讨了不同的 ResNet 网络层数对于模型性能的影响,本文选取 5 个当前最常用的 ResNet 网

络层数来探究网络层数大小对于模型的影响,结果如图 6 所示。

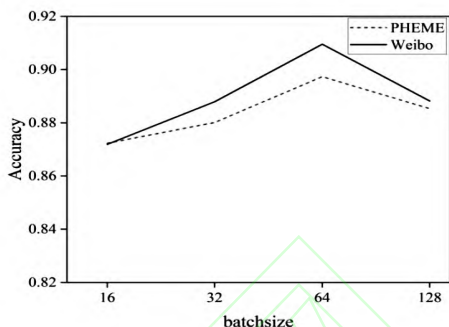


图5 模型准确率和 batchsize 的关系

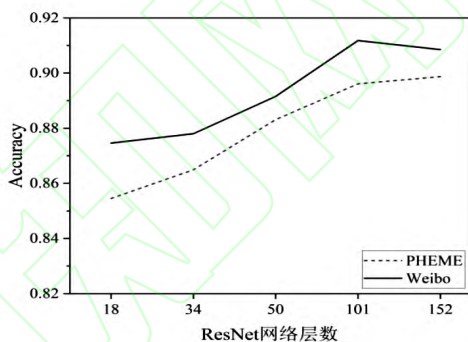


图6 模型准确率和 ResNet 网络层数的关系

通过分析图 6 可知:随着 ResNet 网络层数的增加,模型在两个数据集上的准确率整体上呈上升趋势,这是因为层数更深的 ResNet 具有更多的参数,有助于模型更好地适应训练数据,提高拟合能力。但是当网络层数从 101 增加到 152 时,模型在 PHEME 数据集上的准确率上升趋势不明显,在 Weibo 数据集上的准确率甚至有所下降,这可能是因为 ResNet 模型随着深度的增加变得更加复杂,出现了过拟合的情况。

最后,本文探讨了不同的滑动窗口大小 t 对于模型的影响,本文分别将 t 设置为 1、2、3、4、5 进行实验,结果如图 7 所示。

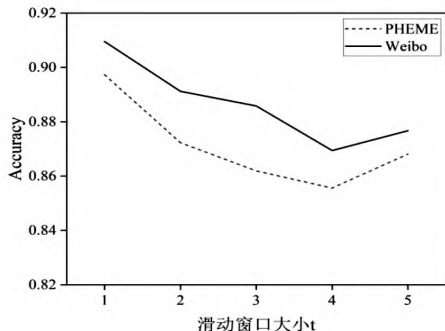


图7 模型准确率和滑动窗口大小 t 的关系

通过分析图 7 可知:当滑动窗口大小为 1 时,本文模型在两个数据集上都获得了最优的准确率。然后随着滑动窗口大小的增加,准确率先下降而后短暂上升,这是因为与图像中相邻像素之间具有紧密的关联性不

同,推文节点的邻域节点之间可能没有直接的关联关系,因此当滑动窗口大小大于1时,模型会学习到很多影响性能的噪声。但随着滑动窗口大小的进一步增加,覆盖范围扩大,节点之间的关联关系有所增强,因此模型的准确率也有所提高。

3.7 基于案例的PPCMRD模型谣言检测分析

本文抽取了Weibo数据集中一则关于“城管暴力执法”的谣言案例与一则关于“亚洲美食节”的非谣言案例进行对比分析。

在推文消息内容检测分析方面,推文的文本和图像蕴含着丰富的指示事件真实性的信息,谣言事件为了达到吸引人注意从而快速广泛传播的目的往往大量使用夸张、惊叹等调动人情绪的词语。相比之下,非谣言事件一般不带有特别的传播目的,因此所用词汇较为中立,从而导致二者在推文内容特征方面存在明显的差异,这是判断事件是否为谣言的关键因素之一。本文模型基于二者之间的差异,通过对推文文本、图片等进行编码,充分捕捉推文在消息内容方面的特征。

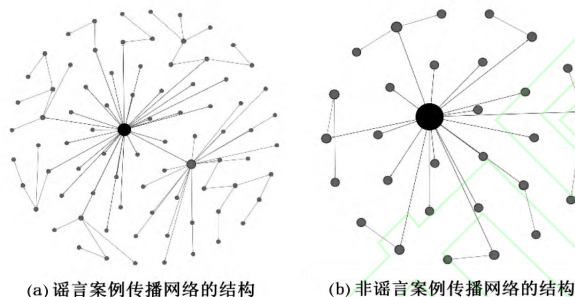


图8 Weibo典型案例传播网络结构

在推文传播模式检测分析方面,图8展示了经模型处理后的两个样本事件网络结构。可以看出,由于谣言与非谣言事件在消息内容方面存在差异且传播目的不同,导致二者在传播网络结构方面也有明显区别。谣言案例的传播树呈现双中心结构,两个中心的一层传播节点数目分别占整体节点总数的35.8%和22.2%,在非谣言的信息传播树中,非谣言案例呈现以根节点为中心的单中心结构,其一层传播节点数占总体节点数的58.1%,信息传播的核心节点相对单一。此外,案例中的谣言网络传播链长度是非谣言的2倍,说明其传播深度比非谣言案例更深。由此可见,二者在传播网络结构上的显著差异也是判断事件是否为谣言的关键因素之一。本文模型在推文传播模式研究方面既考虑了推文消息单一方向的深度传播,也关注其在传播范围内广度散布的特性,并引入带有正负符号标签的注意力机制捕捉目标节点与其邻域节点之间的正负关联关系,这使得谣言源推文节点能够更好地捕捉到其邻域内另一中心节点对其带来的隐性反馈,从而使谣言与非谣言事件获得不同的传播特征表示。

最后,该模型采用跨模态共同注意力机制充分整合上述的推文消息内容特征和推文传播模式特征,捕捉不同模态信息间的互补关系,从而实现对谣言事件与非谣言事件的判别。

4 结 语

本文提出了融合推文传播模式信息与跨模态特征的网络谣言检测模型(PPCMRD)以同时解决传统谣言检测方法对推文多模态信息挖掘不充分、结合不完善,以及对不同的推文传播模式信息整合能力不强的问题。在推文传播特征挖掘方面,首先通过推断潜在连接完善推文传播图;然后,通过双向标签图注意力模块编码多个方向下的推文传播模式;接着,利用推文传播模式信息融合模块捕获不同方向下传播模式特征之间的互补性信息,得到推文传播特征。在整合多模态特征方面,该模型将推文文本、图片和传播特征集成在一个统一框架中,采用跨模态共同注意力机制捕获不同模态信息之间的互补关系,得到帖子的最终嵌入表示判断是否为谣言。两个数据集上的实验结果也验证了该模型的有效性。

未来将从三个方面开展进一步研究:(1)在大语言模型文本生成迅速发展的背景下,尝试对社交媒体上可能存在的大语言模型生成的谣言进行检测,不断优化更新本文模型在未来时代背景下的适用性。(2)本文模型的复杂性可能导致模型在更大规模的谣言检测数据集上的训练和推理速度较慢,下一步将尝试进一步优化模型结构以提高模型效率。(3)部分谣言事件极具伪装性,其在传播初期与非谣言事件的传播结构差异并不显著,从而导致模型将二者误判为同一类的事件,下一步将考虑在推文传播建模研究过程中引入推文传播的时间特征来分析事件的演变过程,从而更准确地判断出谣言事件。

参 考 文 献

- [1] Jin Z, Cao J, Zhang Y, et al. Novel visual and statistical image features for microblogs news verification[J]. IEEE transactions on multimedia, 2016, 19(3): 598-608.
- [2] Chandra S, Mishra P, Yannakoudakis H, et al. Graph-based modeling of online communities for fake news detection[J]. arXiv preprint arXiv:2008.06274, 2020.
- [3] Zhou X, Wu J, Zafarani R. : Similarity-Aware Multi-modal Fake News Detection[C]//Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11 - 14, 2020, Proceedings, Part II. Cham: Springer International Publishing, 2020: 354-367.
- [4] Kaliyar R K, Goswami A, Narang P. FakeBERT: Fake news detection in social media with a BERT-based deep learning ap-

- proach[J]. *Multimedia tools and applications*, 2021, 80(8): 11765–11788.
- [5] Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification[J]. *arXiv preprint arXiv:1510.03820*, 2015.
 - [6] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv:1810.04805*, 2018.
 - [7] Sun L, Hashimoto K, Yin W, et al. Adv-bert: Bert is not robust on misspellings! generating nature adversarial samples on bert[J]. *arXiv preprint arXiv:2003.04985*, 2020.
 - [8] Sun L, Xia C, Yin W, et al. Mixup-transformer: dynamic data augmentation for NLP tasks[J]. *arXiv preprint arXiv:2010.02394*, 2020.
 - [9] Khattar D, Goud J S, Gupta M, et al. Mvae: Multimodal variational autoencoder for fake news detection[C]//The world wide web conference. 2019; 2915–2921.
 - [10] 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 & data mining. 2018; 849–857.
 - [11] Monti F, Frasca F, Eynard D, et al. Fake news detection on social media using geometric deep learning[J]. *arXiv preprint arXiv:1902.06673*, 2019.
 - [12] Han Y, Karunasekera S, Leckie C. Graph neural networks with continual learning for fake news detection from social media[J]. *arXiv preprint arXiv:2007.03316*, 2020.
 - [13] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. *arXiv preprint arXiv:1609.02907*, 2016.
 - [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. IEEE, 2015; 651–662.
 - [15] Thomas S A. Lies, damn lies, and rumors: an analysis of collective efficacy, rumors, and fear in the wake of Katrina[J]. *Sociological Spectrum*, 2007, 27(6): 679–703.
 - [16] 王雷全. 基于图模型的多模态社交媒体分析[D]. 北京邮电大学, 2016.
 - [17] 何俊, 张彩庆, 李小珍等. 面向深度学习的多模态融合技术研究综述[J]. *计算机工程*, 2020, 46(05): 1–11.
 - [18] Bian T, Xiao X, Xu T, et al. Rumor detection on social media with bi-directional graph convolutional networks[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(01): 549–556.
 - [19] Wei L, Hu D, Zhou W, et al. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection[J]. *arXiv preprint arXiv:2107.11934*, 2021.
 - [20] Yang F, Liu Y, Yu X, et al. Automatic detection of rumor on sina weibo[C]//Proceedings of the ACM SIGKDD workshop on mining data semantics. 2012; 1–7.
 - [21] Kwon S, Cha M, Jung K, et al. Prominent features of rumor propagation in online social media[C]//2013 IEEE 13th international conference on data mining. IEEE, 2013; 1103–1108.
 - [22] Ma J, Gao W, Wei Z, 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. 2015; 1751–1754.
 - [23] Zhao Z, Resnick P, Mei Q. Enquiring minds: Early detection of rumors in social media from enquiry posts[C]//Proceedings of the 24th international conference on world wide web. 2015; 1395–1405.
 - [24] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks[J]. 2016.
 - [25] Liu Y, Wu Y F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks[C]//Proceedings of the AAAI conference on artificial intelligence. 2018, 32(1).
 - [26] Lu Y J, Li C T. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media[J]. *arXiv preprint arXiv:2004.11648*, 2020.
 - [27] Yuan C, Ma Q, Zhou W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection[C]//2019 IEEE international conference on data mining (ICDM). IEEE, 2019; 796–805.
 - [28] Ren Y, Zhang J. Fake news detection on news-oriented heterogeneous information networks through hierarchical graph attention[C]//2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021; 1–8.
 - [29] Zheng J, Zhang X, Guo S, et al. MFAN: Multi-modal Feature-enhanced Attention Networks for Rumor Detection[C]. *IJ-CAI*, 2022.
 - [30] Ma J, Gao W, Wong K F. Rumor detection on twitter with tree-structured recursive neural networks[C]. *Association for Computational Linguistics*, 2018.
 - [31] Yang Y. COVID-19 fake news detection via graph neural networks in social media[C]//2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2021; 3178–3180.
 - [32] Luo Z, Li Q, Zheng J. Deep feature fusion for rumor detection on twitter[J]. *IEEE Access*, 2021, 9: 126065–126074.
 - [33] Ran H, Jia C, Zhang P, et al. MGAT-ESM: Multi-channel graph attention neural network with event-sharing module for rumor detection[J]. *Information Sciences*, 2022, 592: 402–416.
 - [34] Wu Y, Yang J, Zhou X, et al. Exploring Graph-aware Multi-View Fusion for Rumor Detection on Social Media[J]. *arXiv preprint arXiv:2212.02419*, 2022.
 - [35] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016; 770–778.
 - [36] Kipf T N, Welling M. Variational graph auto-encoders[J]. *arXiv preprint arXiv:1611.07308*, 2016.
 - [37] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. *arXiv preprint arXiv:1710.10903*, 2017.
 - [38] Yang X, Lyu Y, Tian T, et al. Rumor detection on social media with graph structured adversarial learning[C]//Proceedings of the twenty-ninth international conference on international joint

- p>conferences on artificial intelligence. 2021; 1417–1423.
- [39] Tian T, Liu Y, Yang X, et al. QSAN: A quantum–probability based signed attention network for explainable false information detection[C]//Proceedings of the 29th ACM international conference on information & knowledge management. 2020; 1445–1454.
- [40] Lu J, Batra D, Parikh D, et al. Vilbert: Pretraining task–agnostic visiolinguistic representations for vision–and–language tasks [J]. Advances in neural information processing systems, 2019, 32.
- [41] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [42] Song C, Yang C, Chen H, et al. CED: credible early detection of social media rumors [J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(8); 3035–3047.
- [43] Zubiaga A, Liakata M, Procter R. Exploiting context for rumour detection in social media[C]//Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13–15, 2017, Proceedings, Part I 9. Springer International Publishing, 2017; 109–123.
- [44] Wang Y, Qian S, Hu J, et al. Fake news detection via knowledge–driven multimodal graph convolutional networks [C]//Proceedings of the 2020 international conference on multimedia retrieval. 2020; 540–547.
- [45] Kingma D P, Ba J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv:1412.6980, 2014.
- [46] Fagbohunge O, Qian L. Impact of learning rate on noise resistant property of deep learning models [C]//Proceedings of the Future Technologies Conference. Cham: Springer Nature Switzerland, 2023; 14–30.
- [47] Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis[J]. arXiv preprint arXiv:1809.11096, 2018.
- [48] Cai G, He L, Zhou M C, et al. Learning smooth representation for unsupervised domain adaptation [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021.