·舆情研究·

ParallelGAT: 网络谣言检测方法*

吴 越[1,2] 温 欣[1] 袁 雪[1]

(1.西华大学 计算机与软件工程学院 成都 610039;
2.电子科技大学 计算机科学与工程学院 成都 611730)

摘 要: [**研究目的**]社交媒体平台在促进多元信息交互的同时,也助推了谣言的快速传播。如何准确、及时地发现谣言,已成为多领域学者共同关注的热点问题。最新的谣言检测研究表明,基于谣言传播结构的方法能够捕捉丰富的谣言传播特征,提升谣言检测准确率,而基于外部证据推理的方法可以在传播数据不充分的情况下判别谣言真假,提高谣言检测的时效性。[**研究方法**] 为实现谣言检测准确率和时效性的同步提升,本研究结合这两种方法的优势,提出了基于并行图注意力网络的谣言检测方法 ParallelGAT。ParallelGAT 由两个图注意力网络模型 BiGAT 和 MlGAT 并行构成。其中,BiGAT 模型通过引入注意力机制以捕捉重要的谣言传播和散布特征; MlGAT 模型通过在外部知识中增加多头注意力机制以获取关键的外部句子级证据知识和词语级证据知识; BiGAT 和 MlGAT 的输出特征向量最终通过聚合模块生成谣言检测标签。[**研究结论**] 在公开数据集上的实验结果显示,该文提出的模型优于现有的方法。

关键词: 社交媒体平台; 网络谣言; 谣言检测; 图注意力网络; 传播结构; 证据推理; ParallelGAT
中图分类号: TP391 **文献标识码**: A **文章编号**: 1002-1965(2023)05-0094-08
引用格式: 吴 越,温 欣,袁 雪.ParallelGAT: 网络谣言检测方法[J].情报杂志,2023,42(5):94-101,93.
DOI: 10.3969/j.issn.1002-1965.2023.05.014

ParallelGAT: Network Rumor Detection Method

Wu Yue[1,2] Wen Xin[1] Yuan Xue[1]

(1.School of Computer and Software Engineering, Xihua University, Chengdu 610039;
2.School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611730)

Abstract: [**Research purpose**] Social media platforms not only help the interaction of multiple information, but also incur the rapid spread of rumors. How to detect rumors accurately and timely has become a hot issue in many fields. The latest research on rumor detection shows that the method based on rumor propagation structure can capture rich rumor propagation characteristics and improve the accuracy of rumor detection, and the method based on external evidence reasoning can distinguish the truth and rumor in the case of insufficient propagation data and improve the timeliness of rumor detection. [**Research method**] In order to improve the accuracy and timeliness of rumor detection synchronously, this paper proposed a rumor detection method based on parallel graph attention network, called ParallelGAT. ParallelGAT is consisted of two graph attention network models in parallel, called BiGAT and MlGAT separately. BiGAT uses attention mechanism to capture the important rumor propagation and dissemination characteristics. MlGAT adds multi head attention mechanism to obtain key external sentence level evidence knowledge and word level evidence knowledge of external knowledge. The output eigenvectors of BiGAT and MlGAT finally generate rumor detection tags through the aggregation module. [**Research conclusion**] Experimental results on public open datasets show that the proposed model is superior to the existing methods.

Key words: social media platform; network rumor; rumor detection; graph attention network; propagation structure; reasoning evidence; ParallelGAT

收稿日期: 2022-07-14 修回日期: 2022-09-27
基金项目: 国家自然科学基金项目“微博热点隐话题发现及其时序特性研究”(编号: 61602389) 研究成果。
作者简介: 吴 越,女,1987 年生,博士,副教授,研究方向: 网络舆情分析; 温 欣,男,1997 年生,硕士研究生,研究方向: 谣言早期检测; 袁雪,女,1998 年生,硕士研究生,研究方向: 谣言早期检测。

0　引　言

社交媒体平台在为人们提供信息交互便利的同时,也带来了谣言散布难控的问题。自新冠疫情爆发以来,Facebook 和 Instagram 就删除了超过 2000 万条谣言[1],新浪微博则在 2021 年度处理了 66251 条谣言[2]。谣言就像这流行病一般疯狂蔓延在网络环境中,对民众的生活造成了极大困扰[3]。民众的恐慌、焦虑等情绪更是加速了谣言的扩散,引发了一系列群体性事件[4],如 2020 年初,伊朗有超过 700 人因相信“高浓度酒精能治愈新冠”的谣言,导致酒精中毒身亡[5]。谣言传播不仅会影响人们的心理健康,还会损害人们的身体健康,同时造成经济损失和社会动荡。因此,谣言检测成为亟待解决的热点研究问题。

目前,谣言检测研究已经受到了计算机科学、心理学、社会学、传播学和复杂系统科学等领域学者的广泛关注[6]。人们普遍认为谣言是“未经官方证实的,让人们相信的一种主张”[7],谣言检测任务旨在确定谣言的真实性[8]。从谣言检测语料集 Twitter15[8] 和 PHEME[9] 所提供的标签可以进一步看出,谣言检测本质上是一个细粒度的分类任务[10],目标是把消息分为真实、虚假和未经证实等不同类别。谣言检测较普通的分类任务而言更具挑战性,其困难在于谣言的内容极具迷惑性且传播规律并不固定[11],检测的准确率难以保证。同时,由于谣言具有比真相传播得更快更远的特性[12],这就要求谣言检测在具备高准确率的同时还要有很强的时效性。为了又快又准地检测谣言,研究者们做了大量的相关工作。

近期研究结果表明,谣言传播模式与真实消息传播模式存在很大差异,有效利用传播特征可以帮助我们提升检测准确率[10]。为了获取谣言传播的高阶全局结构特征,有研究者开始应用图卷积网络(Graph Convolution Network, GCN)[13]模型。2020 年 Bian 等人首次将 GCN 应用于谣言检测任务,提出了谣言检测模型 BiGCN[14]。该模型使用两个 GCN 分别从谣言自顶向下和自底向上的传播方向上挖掘其传播和扩散模式,并在 GCN 的每一层中加入源帖信息,以增强谣言根源的影响。实验结果表明,基于 GCN 的方法优于目前最高水平的其他模型。

虽然利用 GCN 捕捉消息传播结构特征的谣言检测方法能够获得较高的准确率,但前提是提供足量的传播信息,因此,难以满足谣言检测的时效性需求。要提升谣言检测的时效性,实现准确的谣言早期检测,还需借助外部知识进行证据推理。基于证据推理的谣言检测模型最早由 Vlachos 等人[15]提出。现有基于证据推理的谣言检测模型主要分为文档检索、证据提取和谣言验证三个子模块,其关键在于谣言验证模块。近期有研究尝试将谣言数据构建为图结构,通过推理实现谣言验证目标,取得了不错的效果。如 Liu 等人通过在图结构中应用两种不同粒度的核,提出了谣言检测模型 KGAT[16],该模型使用节点核计算证据节点的重要程度,使用连边核在图注意力网络中进行细粒度的信息传递,达到了最优的谣言检测效果。

为提升谣言早期检测的准确率,本文结合传播结构分析和证据推理方法,以 BiGCN[14] 和 KGAT[16] 为基础模型,设计了并行图注意力网络 ParallelGAT,在谣言检测的早期阶段实现了准确率 1.8~6.1%的提升。

1　相关工作

已有的谣言自动检测方法大致可分为: 基于传统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两大类。

1.1　基于传统机器学习的谣言检测方法

基于传统机器学习的谣言检测效果主要依赖人工选取的特征,尤其是消息显式特征、隐式特征和事件特征[17]。

a.消息显式特征是可以直接提取、无需额外计算的消息特征[18]。如消息的文本长度[19]、标点和表情符号、超链接、点赞数量、粉丝数量[20]、消息发布时间和地点等。

b.消息隐式特征是无法直接提取、需要通过额外计算的消息特征。如用户的可信度[21]、群体对于信息的质疑率、消极情绪比例、原始信息和转发信息的时间差、信息的传播树大小[22]等。

c.事件特征不仅包含同一事件下的消息特征,还包含消息之间的关系特征。事件、子事件和消息所构成的可信度网络能更全面地反映事件的整体语义和可信度传播过程[23]。

基于消息特征的谣言检测方法相对简单,但由于选取的特征通常只针对某条特定的消息,忽略了不同消息间的潜在联系,准确率偏低。相较而言,基于事件特征的谣言检测方法能够捕捉更丰富的语义特征,效果更优于基于单条消息的分析方法[18]。总体上,基于传统机器学习的谣言检测方法普遍存在有效特征提取困难、易出现过拟合等问题。因此,近期出现许多基于深度学习的谣言检测方法。

1.2　基于深度学习的谣言检测方法

相较于传统机器学习,深度学习方法的特征学习能力更强,在谣言检测任务中的效果也普遍更好。目前用于谣言检测的主流深度学习方法包括: 循环神经网络(Recurrent Neural Network,RNN)[24]、卷积神经网络(Convolutional Neural Networks,CNN)[25]和图神经

网络(Graph Neural Networks,GNN)。

a.RNN能够很好地处理句子等变长时间序列,所以能够适用于具有时序特性的谣言检测。2016年Ma等人[26]首次将RNN应用于谣言检测,通过RNN获取谣言随时间变化的隐性特征,并引入长短期记忆和门控循环单元以解决梯度爆炸问题,提升准确率。此后,出现了一系列改进的RNN谣言检测模型,如自编码器变体与RNN结合的无监督学习谣言检测模型[27],引入注意力机制的RNN谣言检测模型等[3]。

b.与RNN关注序列全局特征不同,CNN关注每次移动卷积核内部的信息,能更全面地提取谣言传播的多种局部特征。Chen等人[28]利用CNN提取谣言时序段中的关键特征,以提升谣言早期检测效果。Ajao等人[29]将CNN与LSTM结合,以检测Twitter中虚假新闻的立场,都取得了不错的效果。

c.近期有研究者尝试将谣言传播过程或者证据关系表示为图结构,并将谣言检测问题转换为图分类问题,建立基于图神经网络的谣言检测方法,取得了较高的准确率。传播结构方面:Bian等人[14]利用GCN获取自顶向下和自底向上的消息传播结构特征;Yang等人[30]在图结构上引入对抗训练方法来提高对谣言传播的图表示学习能力;胡斗等人[10]通过建模消息之间的多种依赖关系,同时增强重要消息的影响力,以捕获更全面的消息传播结构特征用于谣言检测,效果提升明显。证据推理方面:Zhou等人[31]首次将谣言验证转化为图推理任务,并使用了两种不同的注意力;Liu等人[16]将声明和证据作为节点内容构建全连接图,通过聚合图中相邻节点的信息来预测声明的验证结果,准确率达到了当时的最优效果。

从现有研究来看,基于图神经网络的谣言检测方法是当前总体效果最优的方法,无论在传播结构特征捕捉,还是在证据关系提取方面都表现出了良好的性能。本文的研究建立于图神经网络架构基础上,主要是综合传播与证据特征进行早期谣言检测,与本研究最相关的模型是Bian等人[14]和Liu等人[16]分别提出的BiGCN和KGAT模型。本文的贡献在于:不仅改进了两个原生模型,而且实现了两个模型的并行集成,同步提升了谣言检测的准确率和时效性。

2 问题定义

2.1 谣言传播图定义

令 $C=\{c_1,c_2,\cdots,c_m\}$ 表示谣言集合,其中 m 表示谣言总数,$c_i=\{o_i,r_1^i,r_2^i,\cdots,r_{n_i-1}^i,G_i\}$,$o_i$ 作为根节点表示谣言 c_i 传播图中的源消息,r_j^i 表示对源消息的第 j 条相关的回复,n_i 为 c_i 中的帖子数,G_i 表示该谣言的传播结构。$G_i=<V_i,E_i>$,其中 $V_i=\{o_i,r_1^i,r_2^i,\cdots,r_{n_i-1}^i\}$ 表示图中的节点集,每个节点代表一个帖子,$E_i=\{e_{st}^i \mid b,t=0,1,\cdots,n_i-1\}$ 表示图中的边集,每条边表示帖子之间的转发或评论关系。图1为源消息 o_i 的传播示意图。

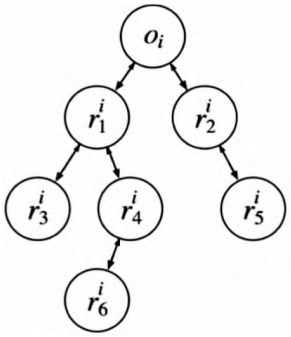


图1 源消息 o_i 传播示意图

2.2 证据结构图定义

令 $\{c_i,e_1^i,e_2^i,\cdots,e_n^i\}$ 表示谣言与证据集合,谣言与证据拼接后作为节点,构成全连接图,如图2所示。图2是谣言 c_i 的证据结构示意图,其中 c_i 为谣言,e_j^i 为与之对应的证据,n 为证据数量,$\|$ 表示数据拼接。

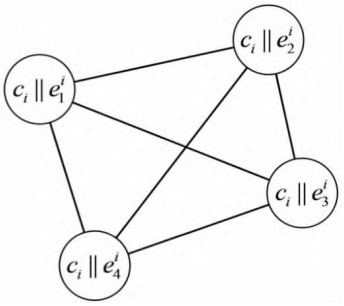


图2 谣言 c_i 证据结构示意图

2.3 谣言分类定义

每个谣言 c_i 都有对应的标签 $y_i \in \{T,F,U\}$,分别表示真消息、假消息和真实性无法验证的谣言。谣言检测的最终目标是训练出一个能够有效对谣言真实性进行判别的分类器 $f: C \rightarrow Y$。

3 ParallelGAT模型

为了同时捕捉谣言的传播结构特征和相关证据特征,本文设计了基于并行图注意力网络的谣言检测模型ParallelGAT。ParallelGAT由两个图注意力网络模块BiGAT和MlGAT并行构成,最终通过特征聚合模块输出检测结果,模型的整体架构如图3所示。

3.1 BiGAT模块

BiGAT模块以BiGCN[13]为基础,进行改进。同BiGCN一样,BiGAT也对谣言数据进行了编码并构建了自顶向下和自底向上两种传播图。不同的地方在于,BiGAT使用了注意力机制去计算每个节点与其邻居节点间的注意力权重,并据此传递信息。具体地,通过设置多层图注意力层以聚合不同注意力分配下的节

点信息,进而得到目标节点的特征表示向量。其中,给定第 l 层的初始节点特征表示向量 $X^{(l)}=[x_0^{(l)},x_1^{(l)},\cdots,x_{n-1}^{(l)}]$,$x_i^{(l)}\in\mathbb{R}^{d_0}$,每个节点信息通过与之相邻的节点信息进行更新的方法如公式 1 所示:

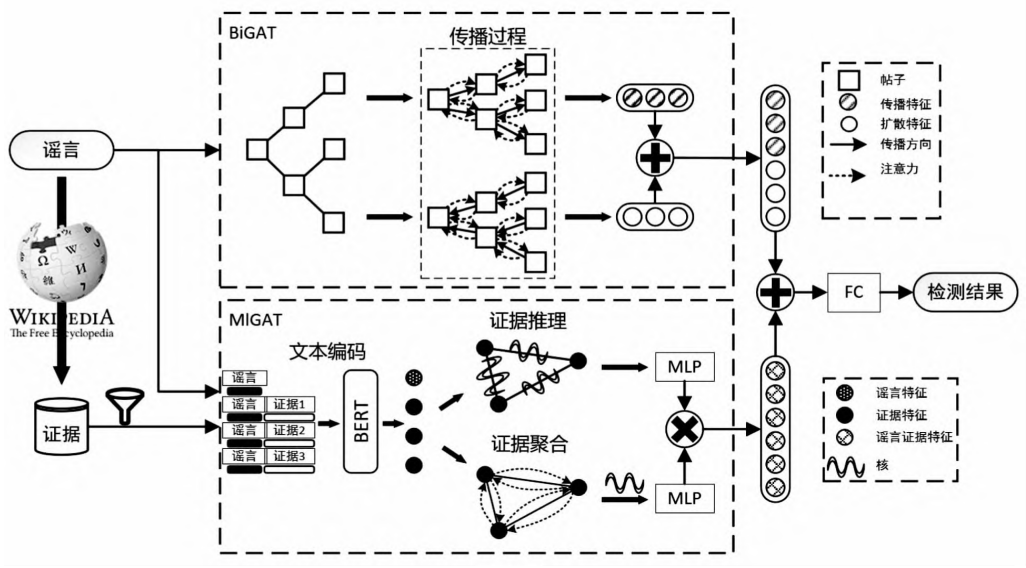


图 3　ParallelGAT 模型架构图

$$\boldsymbol{x}_i'=\alpha_{i,i}\boldsymbol{\Theta}\boldsymbol{x}_i+\sum_{j\in N(i)}\alpha_{i,j}\boldsymbol{\Theta}\boldsymbol{x}_j \tag{1}$$

其中,$\boldsymbol{\Theta}\in\mathbb{R}^{d_i\times d_{i+1}}$ 为权重矩阵,$N(i)$ 表示图中与节点 i 相邻的节点集合,$\alpha_{i,j}$ 为节点 j 到 i 的注意力系数,具体通过一个单层的前馈神经网络计算得到,计算方法如公式 2 所示,其中 $\boldsymbol{a}\in\mathbb{R}^{2d_{i+1}}$ 为权重矩阵,$\cdot^{\mathrm{T}}$ 表示转置,$\|$ 表示向量拼接。

$$\alpha_{i_j}=\frac{\exp(\boldsymbol{a}^{\mathrm{T}}\,\mathrm{LeakyReLU}(\boldsymbol{\Theta}[x_i\,\|\,\boldsymbol{x}_j]))}{\sum_{k\in N(i)\cup\{i\}}\exp(\boldsymbol{a}^{\mathrm{T}}\,\mathrm{LeakyReLU}(\boldsymbol{\Theta}[\boldsymbol{x}_i\,\|\,\boldsymbol{x}_k]))} \tag{2}$$

在经过多轮的信息交换后,对图中所有节点特征进行聚合和拼接,得到综合了自顶向下和自底向上两张传播图信息的特征表示。

3.2　MlGAT 模块

MlGAT 模块以 KGAT[15] 为基础进行改进。MlGAT 的流程和 KGAT 基本一致,先根据谣言文本从外部知识库(如维基百科)中检索并筛选证据,再将证据与谣言拼接送入预训练模型 BERT 生成特征向量,并以特征向量为节点表示构建全连接图,然后利用句子级的注意力在图中传递节点的信息并更新节点的特征表示,利用词语级的注意力选择证据中的重要词语。与 KGAT 不同的是,MlGAT 在应用词语级注意力的同时还使用了多头注意力,以多种不同的注意力分配方式对证据中词语的重要程度进行全面评估。具体地,使用多头注意力层对节点表示进行更新,每个头的计算方法如公式 3 所示。

$$\mathrm{head}_i=f(\boldsymbol{W}_i^q q,\boldsymbol{W}_i^k k,\boldsymbol{W}_i^v v) \tag{3}$$

其中,$\boldsymbol{W}_i^q$、$\boldsymbol{W}_i^k$ 和 $\boldsymbol{W}_i^v$ 均为权重矩阵,q、k、v 均为节点的特征表示向量,函数 $f(\,)$ 代表注意力的计算方法,如公式 4 所示,其中,$\boldsymbol{Q}$ 为 Query,$\boldsymbol{K}$ 为 Key,$\boldsymbol{V}$ 为 Value,d 为输入的 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的长度。

$$f(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}})\,\boldsymbol{v} \tag{4}$$

之后对每个头的计算结果进行聚合,得到节点新的特征表示向量 H^i,如公式 5 所示。

$$H^i=\mathrm{concat}(\mathrm{head}_1,\mathrm{head}_2,\cdots,\mathrm{head}_n)\;\boldsymbol{W}_2 \tag{5}$$

3.3　特征聚合模块

ParallelGAT 模型的最后部分是对 BiGAT 和 MlGAT 所输出的图特征表示向量进行聚合,生成新的图的特征表示向量。具体地,先将谣言检测部分得到的特征表示向量 H_{RD} 和事实验证部分得到的特征表示向量 H_{FV} 进行拼接,得到新的特征表示向量 H_{concat},如公式 6 所示。

$$H_{concat}=\mathrm{concat}(H_{RD},H_{VF}) \tag{6}$$

再利用多层感知机完成对 H_{concat} 内部信息的聚合,并使用 Softmax 函数计算最终的分类结果,如公式 7 所示。

$$P=\mathrm{Softmax}(\mathrm{MLP}(H_{concat})) \tag{7}$$

4　实验设置

4.1　数据集

本文将在公开数据集 PHEME 和 FEVER 上分别测试 BiGAT 和 MlGAT 模块。PHEME 数据集由 Zubiaga 等人[9] 创建,收集了来自九个公共事件的共计 2402 条谣言,分别标记为真实消息、虚假消息和未经证实的消息。由于 PHEME 数据集本身不包含能够验证谣言的证据,本文使用 Hanselowski 等人[33] 提出的

文档检索和证据选择方法,从维基百科中搜索了相关文档进行证据支撑,使 PHEME 数据集中约 6%的样本具备证据。FEVER 数据集由 Thorne 等人[32] 于 2018 年发布,之后 Thorne 等人举办了多次以 FEVER 数据集为基础的竞赛,同时形成了完善的评价体系。FEVER 共包含有 18.5 万条消息以及对应的证据,其内容为非结构化文本,贴近现实社交网络中所使用的语言。PHEME 数据集和 FEVER 数据集的统计信息如表 1 所示。

表 1 谣言检测数据集统计信息

数据集	PHEME	FEVER
样本总数	2402	185455
虚假消息数	638	43107
真实消息数	1067	93367
未经证实的消息数	697	48971

4.2 基线模型

本文选取了一些具有代表性的基于深度学习的谣言检测方法作为基线方法,与本文提出的 ParallelGAT 模型中的两个并行图神经网络 BiGAT 和 MlGAT 模块分别进行对比。

与 BiGAT 模块对比的基线方法包括:

a.RvNN,该方法由 Ma 等人[34] 于 2018 年提出,是一种基于谣言传播结构,使用 GRU 单元进行谣言特征表示向量学习的循环神经网络谣言检测方法;

b.BiGCN,该方法由 Bian 等人[14] 于 2020 年提出,是一种基于图卷积网络,从谣言的传播和散布两个角度进行建模分析,并将谣言分类转化为图分类的谣言检测方法;

c.EBGCN,该方法由 Wei 等人[35] 于 2021 年提出,通过在 BiGCN 中添加边权重推断模块改进而来。

与 MlGAT 模块对比的基线方法包括:

a.UKP-Athene,该方法由 Hanselowski 等人[33] 于 2018 年提出,通过聚合由 ESIM 模型输出的特征表示向量,并利用注意力机制计算得出最终的预测结果;

b.UNC-NLP,该方法由 Nie 等人[36] 于 2019 年提出,在 FEVER 事实验证竞赛中取得了最好成绩。该方法利用一种神经语义匹配网络来联合地解决文档检索、证据选择和谣言验证全部三个子任务;

c.GEAR,该方法由 Zhou 等人[31] 于 2019 年提出,利用图注意力网络从证据中提取特征,并利用注意力神经层对特征进行聚合得到最终谣言预测标签;

d.KGAT,该方法由 Liu 等人[16] 于 2019 年提出,利用节点核和连边核对图注意力网络中节点信息进行更新,并利用多级软匹配对图中所有节点信息进行聚合得出谣言预测标签。

4.3 实验参数设置

BiGAT 模块参考基线模型的参数设置[34],采用维度 $d_0 = 5000$ 的 TF-IDF 特征初始化节点的输入特征表示向量。使用双层的图注意力网络模型进行训练,每层中隐藏向量维度设置为 64,并设置 dropout 概率为 0.5。模型采用 Adam 算法进行训练,学习率设置为 0.001,权值衰减设置为 0.0001,迭代次数设置为 200,并设置当验证集的 loss 值在 10 个迭代次数内不再下降时提前结束训练。

MlGAT 模块参考基线模型的参数设置[31],将谣言证据对的文本长度限制为 130,对于长度未达到 130 的,使用 0 进行补全。BERT 模型的隐藏向量长度设置为 768。使用 BertAdam 优化器进行优化,为避免过拟合,设置 Dropout 层的 dropout 比率为 10%,学习率设置为 5×10^{-5},同时为加快模型运行前期的学习效率,预热学习率设置为 0.1。

4.4 评价指标

在 PHEME 数据集上的谣言检测,本文参考文献[10],使用了准确率 Acc、宏平均 F1 值(mF1,macro-averaged F1) 和加权平均 F1 值(wF1,weighted-averaged F1) 作为评价指标。准确率 Acc 的计算方法为

$$Acc = \frac{TP + FN}{TP + FP + TN + FN} \tag{8}$$

其中,TP(True Positive) 表示被模型预测正确的正样本数,FP(False Positive) 表示被模型预测正确的负样本数,FN(False Negative) 表示被模型预测错误的正样本数,TN(True Negative) 表示被模型预测错误的负样本数。宏平均 F1 值指先对每个类别下的 F1 值求和,再计算算术平均值,计算方法为

$$mF1 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{F1}_i \tag{9}$$

其中,n 表示需要预测出的类别总数。加权平均 F1 值是在不同类别的样本数量不平衡情况下的评价指标,其计算方法为

$$wF1 = \sum_{i=1}^{n} p_i \, \mathrm{F1}_i \tag{10}$$

其中,p_i表示不同类别的样本占全部样本的比例。

在 FEVER 数据集上的谣言检测,本文使用 FEVER 共享任务中通用的 FEVER 分数和分类准确率 Acc 作为评价指标。FEVER 分数是在为每个谣言提供了至少一组能够完全能够证明其真实性的证据后,模型所计算出的分类准确率。

5 实验结果及分析

5.1 BiGAT 模块实验结果

BiGAT 模块在 PHEME 数据集上的谣言检测结果如表 2 所示,RvNN、BiGCN 和 EBGCN 的实验结果引

用自文献[35]。

表 2　BiGAT 谣言检测结果(PHEME 数据集)

方法	Acc	mF1	wF1
RvNN	34.1	26.4	-
BiGCN	56.9	48.3	66.8
EBGCN	71.5	57.5	79.1
BiGAT*	82.6	77.7	80.0

从表 2 可以看出,与基线模型相比,BiGAT 取得了更好的谣言检测效果,说明利用双向图注意力网络对谣言传播结构建模进行谣言检测具有合理性。本文认为主要原因是: 在对谣言的传播特征和散布特征进行建模的过程中,BiGCN 赋予所有连边相同的信息传递权重,难以从重要节点中获取充分的信息量,而 EBGCN 虽考虑了根节点和最大转发节点的影响力,但其自适应权重分配算法略逊于注意力机制的注入。而 BiGAT 能够利用注意力机制自主学习重要节点信息,同时也放大根节点的影响。因此,对整张谣言传播图的特征表示向量学习更全面,更有助于检测出谣言消息。

为探究不同方向的传播图和使用注意力机制寻找重要节点这两点改进对谣言检测产生的潜在影响,通过考虑不同方向的传播图,可以得到 4 种传播图的构建方案: 分别为不考虑传播方向的无向传播图 UD、考虑自顶向下传播方向的传播图 TD、考虑自底向上传播方向的传播图 BU 和同时考虑两种方向的双向传播图 TD+BU。在这 4 种传播图构建方案下,为探究不同的特征表示向量的聚合方式对实验结果的影响,对两种不同的聚合方式进行了对比,分别为最大值聚合方式 Max-agg 和平均值聚合方式 Mean-agg。

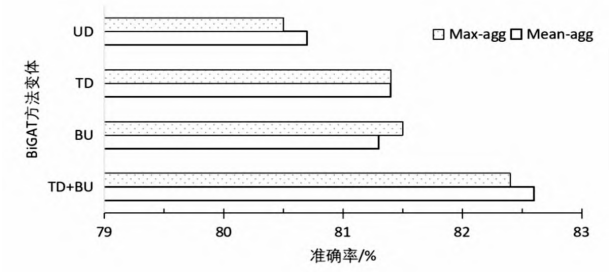
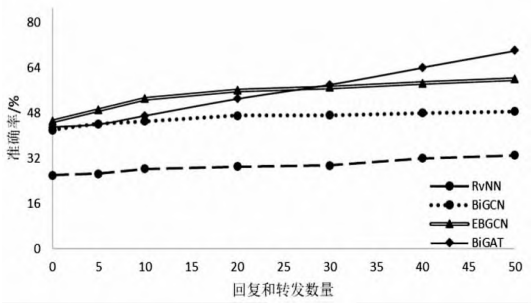


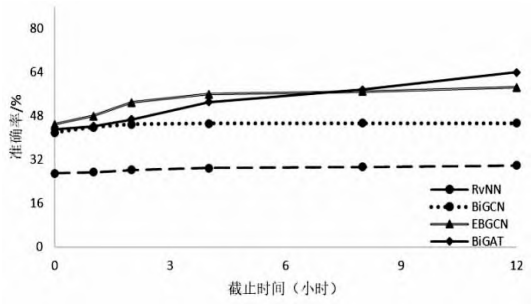
图 4　BiGAT 消融实验结果(PHEME 数据集)

实验结果如图 4 所示。从图 4 中可以看出: a.在使用相同的聚合策略时,根据基于不同方向的传播图结构得到的 4 种变体,其谣言检测性能从低到高依次为: 无向图 UD 变体方法、自顶向下图 TD 变体方法、自底向上图 BU 变体方法和双向图 TD+BU 变体方法。由此可见,同时考虑谣言的传播特征和散布特征,能够提取出更丰富的结构特征,帮助模型更好识别谣言。b.基于平均值聚合的聚合策略在绝大部分情况下能够获得更好的性能表现,平均值聚合也是本文主要采用的聚合策略,说明模型在综合利用图中节点信息时能够获取更全面的特征表示,也说明了整体的节点信息在大部分情况下能够更好地表示整张图的信息。

在早期谣言检测方面,本文参考文献[8],将谣言传播的早期阶段定义为回复和转发数量较少的阶段或者原帖发布后 24-48 小时的阶段。本节将分别依据谣言的回复和转发数量,以及其发布后经历的时间长短,对不同模型的早期谣言检测效果进行评估,实验结果如图 5 所示。



(a) 基于回复和转发数量



(b) 基于距离原帖发布时间

图 5　BiGAT 早期谣言检测结果(PHEME 数据集)

谣言检测准确率随原帖回复数和转发量的变化规律如图 5(a) 所示; 随距离原帖发布时间间隔的变化规律如图 5(b) 所示。从图 5 不难看出: a.随着原帖回复数量和转发数量的增长,四个模型的谣言检测准确率都有所上升,说明谣言检测模型的准确率依赖于可获得的信息量,信息量越多,谣言检测越准确。b.相较于以 GRU 为基础模块的 RvNN 模型,其他三个图卷积神经网络模型的准确率更高,说明在谣言检测方面,图模型较序列模型的效果更好。这可能是由于图模型能够更有效地聚合邻居节点信息,因此能够学习到准确的节点特征表示向量。c.从图 5(a) 可以看出,当回复和转发数量低于 30 时,准确率最高的是 EBGCN 模型; 而当回复和转发数量高于 30 时,BiGAT 模型的谣言检测准确率稳步提升,且明显高于其他三个基线模型。从图 5(b) 可以看出类似的规律,即在谣言发布的最早期,EBGCN 模型的准确率最高,但随着时间的增长,距离原帖发布时间超过 8 个小时以后,BiGAT 模型的优势就体现出来了。这说明 BiGAT 可以通过注意力机制捕捉到关键谣言特征,从而提升谣言检测的准确率,

但前提是需要有足够的信息。d.由于 BiGAT 在信息量极少的情况下进行谣言检测的效果还有待提升,因此,本文考虑在 BiGAT 的基础上加入基于外部知识的证据推理模块。

5.2 MlGAT 模块实验结果

MlGAT 模块在 FEVER 数据集上的谣言检测结果见表 3,表中数据均为模型在开发集上的测试结果。

表 3 MlGAT 模块实验结果(FEVER 数据集)

模型	分类准确率	FEVER 分数
UKP-Athene	68.49	64.74
UNC-NLP	69.72	66.49
GEAR	74.84	70.69
KGAT	75.51	71.61
MlGAT*	78.15	76.05

上述添加引用的模型实验结果均来自于其论文中公布的结果,其中为保证实验环境一致,使用 ESIM 模型进行证据选择,使用 BERT Base 预训练模型进行编码。从表 3 可以看出,在实验环境一致的情况下,MlGAT 在开发集上分类准确率为 78.15%,FEVER 分数为 76.05,性能优于其他 4 个基线模型。本文认为,这主要是由于多头注意力机制能够更好地综合证据中包含的多种隐含信息,对证据起到了更全面的评估作用。

本文还研究了在必要证据数量变化的情况下,MlGAT 模型的性能表现。必要证据是指模型在对谣言做出正确的验证结果时,所需要的最少证据。必要证据的数量越多,模型对谣言验证的难度也越大。FEVER 数据集中谣言必要证据统计信息如表 4 所示。

表 4 FEVER 数据集中谣言的必要证据统计信息

类型	单条必要证据的谣言数量	多条必要证据的谣言数量
训练集	89715	20095
开发集	11372	1960

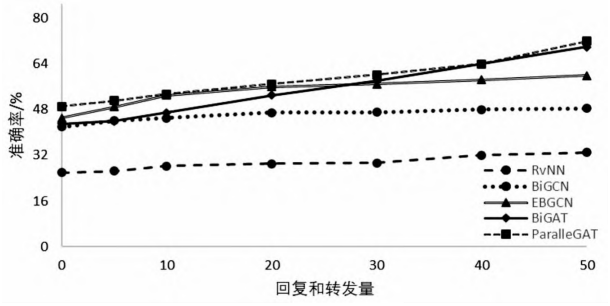
如表 5 所示,在仅有单条必要证据的情况下,MlGAT 模型不需要进行复杂的证据推理,谣言检测准确率明显高于需要多条必要证据的情况。这是因为在单条必要证据的情况下,主要考察的是模型的降噪能力。而在多条必要证据时,考察的是模型从多条证据句中提取细微线索并将其聚合的能力,相比单条必要证据的情况要更加复杂。因此,在进行证据选择时如果能够尽量将必要证据集中到一条句子中,模型的推理效果将得到更大提升。

表 5 MlGAT 在不同必要证据数量下的性能比较

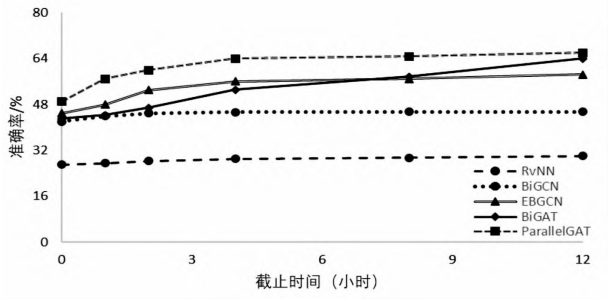
必要证据数量	分类准确率	FEVER 分数
多条	65.84	59.3
单条	81.21	80.1

5.3 ParallelGAT 模型实验结果

ParallelGAT 模型通过结合 MlGAT 模块和 BiGAT 模块的优势,以提升早期谣言检测的准确率,如图 6 所示。



(a) 基于回复和转发量



(b) 基于距离原帖发布时间

图 6 ParallelGAT 模型早期谣言检测结果

在实验 5.1 中,我们进行了 BiGAT 和其他三个基线模型的对比实验,结果显示,BiGAT 在谣言爆发最初阶段的检测效果还有待提升。对此,我们在 BiGAT 的基础上加入了 MlGAT 模块,构成了更强大的早期谣言检测模型 ParallelGAT。从图 6 可以明显看出,即使在谣言爆发的最初阶段,ParallelGAT 较其他基线模型的准确率也有不同幅度的提升。从图 6(a)可以看出,当回复和转发量低于 40 时,ParallelGAT 相比于 BiGAT 准确率提升明显,最高幅度达到 6.1%;从图 6(b)也可以看出,在谣言爆发 12 小时内,ParallelGAT 的准确率较 BiGAT 提升显著,提升幅度最高达到了 16.7%。即使相较于在谣言爆发早期阶段表现最好的 EBGCN 模型,ParallelGAT 的准确率也提升了 4%以上。本文认为 ParallelGAT 之所以能够在谣言早期检测任务中表现突出,主要是由于其在 BiGAT 的基础上融合了证据推理方法的优势,利用外部知识对谣言进行检测,弥补了 BiGAT 在早期谣言检测时因传播信息不足而导致的准确率低的缺陷,从而进一步提升了早期谣言检测性能,达到了准确率和时效性的同步提升。

6 结 论

本文研究了基于文本内容和传播结构信息的谣言检测任务,提出一种基于并行图注意力网络的谣言检测方法 ParallelGAT。该方法通过双向图注意力网络对谣言的传播和扩散特征进行建模,同时利用外部知

识对谣言文本内容的真实性进行推理验证。在 PHEME 公开数据集上的实验表明,本文提出的方法具有比其他基线方法更高的谣言检测性能,并且对处于传播早期阶段的谣言也有较高的检测准确率。由于目前既包含谣言内容、传播结构,又包含谣言相关证据的公开数据集非常稀缺,未来在更广泛数据集上进行实验测试,同时通过设计更有效的特征利用策略,以进一步提升模型的时效性。

参 考 文 献

[1] Covid19 Infodemics Observatory [EB/OL]. [2021-04-11]. https://covid19obs.fbk.eu/#/.

[2] 微博辟谣.@微博辟谣 的个人主页 [CP/OL]. [2021-04-11]. https://weibo.com/u/1866405545.

[3] 刘　妃,曹　敏,左美云.辟谣平台的评价指标体系构建及实证研究 [J].情报杂志,2021,40(9):95-100,94.

[4] 吕　途,陈　昊,林　欢,等.突发公共事件下网络谣言治理策略对谣言传播意愿的影响研究 [J].情报杂志,2020,39(7):87-93.

[5] Forrest A. Coronavirus: 700 dead in Iran after drinking toxic methanol alcohol to 'cure Covid-19' [N/OL]. Independent, 2020.04.28 [2021-04-11]. https://www.independent.co.uk/news/world/middle-east/coronavirus-iran-deaths-toxic-methanol-alcohol-fake-news-rumours-a9487801.html

[6] 陈燕方,李志宇,梁　循,等.在线社会网络谣言检测综述 [J].计算机学报,2018,41(7): 1648-1676.

[7] Allport G W, Postman L. The psychology of rumor [J]. 1947. New York: Henry Holt and Co.

[8] 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. Vancouver, Association for Computational Linguistics, 2017, 708-717.

[9] Zubiaga A, Liakata M, Procter R, et al. Analysing how people orient to and spread rumours in social media by looking at conversational threads [J]. PloS One, 2016, 11(3): e0150989.

[10] 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.

[11] 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.

[12] Vosoughi S, Roy D, Aral S. The spread of true and false news online [J]. Science, 2018, 359(6380): 1146-1151.

[13] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [C] //Proceedings of International Conference on Learning Representations, Toulon, France, ICLR, 2017

[14] 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(1): 549-556.

[15] Vlachos A, Riedel S. Fact checking: Task definition and dataset construction [C] //Proceedings of the ACL 2014 workshop on language technologies and computational social science, 2014: 18-22.

[16] Liu Z, Xiong C, Sun M, et al. Fine-grained fact verification with kernel graph attention network [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL, 2020, 7342-7351.

[17] 高玉君,梁　刚,蒋方婷,等.社会网络谣言检测综述 [J].电子学报,2020,48(7): 1421-1435.

[18] Castillo C, Mendoza M, Poblete B. Information credibility on twitter [C] //Proceedings of the 20th international conference on World wide web, 2011: 675-684.

[19] Qazvinian V, Rosengren E, Radev D, et al. Rumor has it: Identifying misinformation in microblogs [C] //Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011: 1589-1599.

[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, New York, Association for Computing Mechinery, 2012: 1-7.

[21] Cao J, Guo J, Li X, et al. Automatic rumor detection on microblogs: A survey [J]. arXiv preprint arXiv: 1807.03505, 2018.

[22] Wu K, Yang S, Zhu K Q. False rumors detection on sina weibo by propagation structures [C] //proceedings of the 2015 IEEE 31st international conference on data engineering, IEEE, 2015: 651-662.

[23] Huang W, Yan H, Liu R, et al. F-score feature selection based Bayesian reconstruction of visual image from human brain activity [J]. Neurocomputing, 2018, 316: 202-209.

[24] Schuster M, Paliwal K K. Bidirectional recurrent neural networks [J]. IEEE transactions on Signal Processing, 1997, 45(11): 2673-2681.

[25] Chen Y. Convolutional neural network for sentence classification [D] University of Waterloo, 2015.

[26] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks [C] //International Joint Conference on Artificial Intelligence, 2016.

[27] Chen W, Zhang Y, Yeo C K, et al. Unsupervised rumor detection based on users' behaviors using neural networks [J]. Pattern Recognition Letters, 2018, 105: 226-33.

[28] Chen Y, Sui J, Hu L, et al. Attention-residual network with CNN for rumor detection [C] //Proceedings of the 28th ACM international conference on information and knowledge management. New York, ACM, 2019: 1121-1130.

[29] Ajao O, Bhowmik D, Zargari S. Fake news identification on twitter with hybrid cnn and rnn models [C] //Proceedings of the 9th international conference on social media and society. New York, ACM, 2018: 226-230.

[30] 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 Conferences on Artificial Intelligence, 2021: 1417-1423.

(下转第 93 页)

体系不得不面对和解决的问题。为此,必须在法治轨道上充分利用区块链,以技术赋能的方式参与碳数据安全治理。区块链技术能够对碳数据实现链上治理,解决碳数据自身安全风险和利用安全风险,并促进其流通利用,充分释放数据价值。但是,我国目前并未针对区块链的实践应用构建完整融贯的法律体系。因而,应当从宏观维度的规制理念、规制体系和规制路径以及微观维度的具体规则设置两个层面形塑区块链赋能碳数据安全治理的法律体系,最终实现区块链赋能碳数据安全治理的目标。

参 考 文 献

[1] Gatteschi V,Lamberti F,Demartini C,et al. To blockchain or not to blockchain: That is the question[J].IT Professional, 2018,20(2):62-74.

[2] Jiang J Y. Regulating blockchain? A retrospective assessment of China's blockchain policies and regulations[J].Tsinghua China Law Review,2020,12(2):313-364.

[3] Slaughter M,McCormick D. Data is power: Washington needs to craft new rules for the digital age[J].Foreign Affairs,2021,100(3):54-63.

[4] Hofmann E,Rüsch M. Industry 4.0 and the current status as well as future prospects on logistics[J].Computers in Industry,2017(89):23-34.

[5] Senate and House of Representatives of the U.S.A.Federal Information Security Modernization Act of 2014[EB/OL].[2022-10-20]. https://www.congress.gov/113/plaws/publ283/PLAW-113publ283.pdf.

[6] 国家标准化管理委员会. GB/T39477-2020 信息安全技术 政务信息共享 数据安全技术要求[S].北京: 中国标准出版社, 2022: 1.

[7] 刘金瑞.数据安全范式革新及其立法展开[J].环球法律评论,2021,43(1):5-21.

[8] 生态环境部.关于做好全国碳排放权交易市场数据质量监督管理相关工作的通知[EB/OL].[2022-06-27].http://www.mee.gov.cn/xxgk2018/xxgk/xxgk06/202202/t20211025_957707.html.

[9] 生态环境部.生态环境部公开中碳能投等机构碳排放报告数据弄虚作假等典型问题案例[EB/OL].[2022-06-20].https://www.mee.gov.cn/ywgz/ydqhbh/wsqtkz/202203/t20220314_971398.shtml.

[10] 高富平.数据流通理论—数据资源权利配置的基础[J].中外法学,2019,31(6):1405-1424.

[11] Ginsberg D. Blockchain: Web 3.0 or Web 3.No? [J].AALL spectrum,2017,22(1):36-39.

[12] 华为区块链技术开发团队.区块链—技术及其应用[M].北京: 清华大学出版社,2019: 21-27.

[13] 袁　勇,王飞跃.区块链技术发展现状与展望[J].自动化学报,2016,42(4):481-494.

[14] 何　蒲,于　戈,张岩峰,等.区块链技术与应用前瞻综述[J].计算机科学,2017,44(4):1-7.

[15] 沈　鑫,裴庆祺,刘雪峰.区块链技术综述[J].网络与信息安全学报,2016,2(11):11-20.

[16] 谢　晖.论法律价值与制度修辞[J].河南大学学报(社会科学版),2017,57(1):1-27.

[17] 陈爱飞.区块链共谋的反垄断监管[J].现代法学,2022,44(4):145-157.

[18] 陆宇峰.“自创生”系统论法学: 一种理解现代法律的新思路[J].政法论坛,2014,32(4):154-171.

[19] 雷　磊.法教义学的方法[J].中国法律评论,2022,9(5):77-93.

[20] 高奇琦.主权区块链与全球区块链研究[J].世界经济与政治,2022,36(10):50-71.

[21] 张　红.监管沙盒及与我国行政法体系的兼容[J].浙江学刊,2018,56(1):77-86.

[22] 夏庆锋.智能合约的法律性质分析[J].东方法学,2022,15(6):33-43.

[23] 吴志攀.“互联网+”的兴起与法律的滞后性[J].国家行政学院学报,2015,16(3):39-43.

(责编: 王育英; 校对: 刘影梅)

(上接第 101 页)

[31] Zhou J,Han X,Yang C,et al.GEAR: Graph-based evidence aggregating and reasoning for fact verification [C] //Proceedings of ACL 2019,Florence: ACL,2019

[32] Thorne J,Vlachos A,Christodoulopoulos C,et al.Fever: A large-scale dataset for fact extraction and verification [C] //Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. New Orleans: ACL,2018: 809-819

[33] Hanseilowski A,Zhang H,LI Z,et al.Ukp-athene: Multi-sentence textual entailment for claim verification [C] //Proceedings of the First Workshop on Fact Extraction and VERification (FEVER).Brussels: ACL,2018: 103-108

[34] 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.Melbourne: ACL,2018: 1980-1989

[35] Wei L,Hu D,Zhou W,et al.Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection [C] //Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing.Online: ACL,2021: 3845-3854

[36] Nie Y,Chen H,Bansal M.Combining fact extraction and verification with neural semantic matching networks [C] //Proceedings of the AAAI Conference on Artificial Intelligence.AAAI Press, 2019: 6859-6866.

(责编/校对: 贺小利)