

# 基于多视图表征的虚假新闻检测

张新有, 孙峰, 冯力, 邢焕来

(西南交通大学计算机与人工智能学院, 成都 611756)

**摘要:** 社交网络已经成为人们日常生活中获取和分享信息的主要渠道, 同时也为虚假新闻的传播提供了捷径。如今, 针对网络虚假新闻的检测问题受到学术界的广泛关注, 但目前的检测方法缺乏基于新闻多个视角的深度探索或忽视了新闻中不同信息传播方向不同的问题, 有待改进。文章提出一种基于新闻内容、用户信息和新闻传播3种视角的多视图表征和检测的模型 MVRFD (Multi-View Representations for Fake News Detection), 为虚假新闻检测任务提供更全面的视角。首先, 利用协同注意力机制表征新闻内容中的多模态信息, 使用具有不同方向的图神经网络聚合新闻传播过程中的用户信息和观点信息; 然后, 利用双协同注意力机制实现多个视角间的信息交互; 最后, 将新闻内容特征和新闻上下文特征进行融合。在公开数据集上的实验结果表明, 文章所提出的模型实现了 96.7% 的准确率和 96.8% 的  $F1$  值, 优于主流的文本处理模型以及基于单视角的检测模型。

**关键词:** 虚假新闻检测; 图神经网络; 多模态表征; 注意力机制; 多视图表征

**中图分类号:** TP309 **文献标志码:** A **文章编号:** 1671-1122 (2024) 03-0438-11

中文引用格式: 张新有, 孙峰, 冯力, 等. 基于多视图表征的虚假新闻检测 [J]. 信息安全, 2024, 24 (3): 438-448.

英文引用格式: ZHANG Xinyou, SUN Feng, FENG Li, et al. Multi-View Representations for Fake News Detection[J]. Netinfo Security, 2024, 24(3): 438-448.

## Multi-View Representations for Fake News Detection

ZHANG Xinyou, SUN Feng, FENG Li, XING Huanlai

(School of Computer and Artificial Intelligence, Southwest Jiaotong University, Chengdu, 611756, China)

**Abstract:** Social networks have become a major channel for people to access and share information in their daily lives, while also providing shortcuts for the spread of fake news. Nowadays, the detection of online fake news has been widely concerned and studied by the academic community, but the current methods lack in-depth exploration based on multiple perspectives of news or ignore the different directions of different information in news. In order to provide a more comprehensive perspective for fake news detection task, this paper proposed a multi-view representations for fake news detection (MVRFD) model based on three

收稿日期: 2023-12-15

基金项目: 国家自然科学基金 [62172342]

作者简介: 张新有 (1971—), 男, 河南, 副教授, 博士, 主要研究方向为分布式计算与应用、网络安全; 孙峰 (1999—), 男, 安徽, 硕士研究生, 主要研究方向为深度学习、网络信息安全; 冯力 (1974—), 男, 四川, 教授, 博士, 主要研究方向为人工智能和网络安全; 邢焕来 (1984—), 男, 河北, 副教授, 博士, CCF 会员, 主要研究方向为人工智能和网络安全。

通信作者: 孙峰 sun.feng@my.swjtu.edu.cn

perspectives: news content, user information and news propagation. Firstly, the co-attention mechanism was used to represent the multimodal information in news content, and the graph neural network with different directions was used to aggregate user information and views in the process of news transmission. Then the dual-co-attention mechanism was used to realize the information interaction between multiple perspectives. Finally, the features of news content and the news context were integrated. Experiments on the publicly available dataset show that the proposed model achieves 96.7% accuracy and 96.8% *F1* score, which are better than the mainstream text processing models and single-view-based detection models.

**Key words:** fake news detection; graph neural network; multimodal representation; attention mechanism; multi-view representation

## 0 引言

如今,网络假新闻泛滥的现象愈发严重,并且随着人们获取信息的媒介方式逐渐由传统媒体转向网络媒体,假新闻造成的影响也在逐渐扩大。特别是在一些重大舆情事件发生时,假新闻通过网络社交平台快速、广泛地传播,已经严重危害社会的稳定。目前,已有许多研究表明,通过一个分类器实现自动识别假新闻的方法是可行的,但面对日新月异的网络环境,针对虚假信息的研究也不断受到挑战。

早期的研究主要从社交平台帖子中手工提取特征来训练分类器。YANG<sup>[1]</sup>等人基于新浪微博平台使用支持向量机(Support Vector Machine, SVM)模型进行谣言检测。类似的,其他分类器如决策树<sup>[2]</sup>、贝叶斯<sup>[3]</sup>等也被用来进行相关研究。随着神经网络在自然语言处理领域的发展,一些深度学习方法也被广泛用于挖掘新闻文本中的特征,例如,循环神经网络(Recurrent Neural Network, RNN)及其变体<sup>[4]</sup>能够很好地从新闻中挖掘出更高层次的语义信息,而面对社交网络中愈发普遍的多模态信息,卷积神经网络(Convolutional Neural Network, CNN)常被用于处理图像数据。在多模态表征学习领域,大规模预训练模型,如经过充分训练的卷积模型和BERT(Bidirectional Encoder Representations from Transformers)<sup>[5]</sup>模型,已经逐渐替代了传统的手工表征<sup>[6]</sup>。因此,在虚假新闻检测领域,研究重点之一是实现新闻中不同模态间的信息互补,从而获取更全面的新闻特征。

上述研究主要关注的是新闻内容,但在实际网络

环境中新闻风格易发生变化,并且不同领域的新闻之间也有较大差异,因此新闻传播过程中产生的大量上下文信息就显得尤为重要。例如,用户针对新闻帖子的评论或回复中包含大量线索,同时考虑社交网络中存在的虚假或恶意用户会对这些信息产生干扰,因此在虚假新闻的检测过程中,用户的可信度也经常被考虑在内。实际上,一些研究<sup>[1]</sup>很早就开始关注用户在社交平台上的信息并将其作为用户特征,但此类手工提取的特征信息量有限,很难挖掘出更深层次的特征。近几年,一些研究<sup>[7-10]</sup>开始采用深度学习的方法从新闻的上下文中提取特征,图神经网络(Graph Neural Network, GNN)得到广泛应用。这种基于新闻传播特征的方法不会受新闻内容变化的影响,因此被认为具有更好的稳定性,但往往忽略了新闻内容所发挥的主体作用。

综上所述,虽然不同视角下的新闻特征都能够有效分离出真实新闻与假新闻,但为了从更全面的视角来表征新闻并实现新闻内容和上下文信息之间的互补,本文提出了一种基于多视图表征的假新闻检测模型 MVRFD(Multi-View Representations for Fake News Detection)。首先,利用预训练模型获取新闻文本和图片中丰富的基本语义信息,并通过注意力机制融合多模态特征;然后,充分挖掘新闻在传播过程中产生的用户信息和转发者的反馈信息,并分别作为用户视角和传播视角特征;最后,通过注意力机制融合多个视角的表征。

本文的主要贡献包括4个方面:1)综合考虑了新

闻内容、用户和传播3个视角的特征,同时将3个视角完全分离表征并进行充分融合,使获得的新闻表征具有全面的视角;2)为了更好地挖掘上下文视角的特征,分别采用不同的有向图聚合新闻传播过程中的用户信息和观点信息;3)为了探索不同领域信息的重要性,使用注意力机制融合文本和图像间的信息,以及内容视角和上下文视角之间的信息;4)在公开数据集上对本文模型进行实验,证明其有效性,与当前现有模型相比,本文模型表现出更好的性能。

## 1 相关工作

本文按照新闻的不同视角将目前的相关研究分为基于多模态的方法(内容视角)、基于新闻上下文的方法和基于多个视角的方法3类。其中,在大多数场景下用户和传播信息联系较为紧密,且多数研究者会同时考虑新闻传播过程中的用户信息和传播行为,因此本文将两者合并总结为上下文信息,同时进行介绍。

目前,假新闻检测领域中的多模态方法主要针对的是新闻中的文本和图片信息。SONG<sup>[11]</sup>等人通过交叉注意力机制使其中任一模态能够从其他模态中提取出相关信息,并将通过自注意力计算后的原始文本信息和融合了视觉模态的文本信息作为卷积网络不同通道的输入,这种做法减少了模态交叉融合过程中可能产生的噪声信息的影响。XIONG<sup>[12]</sup>等人同样关注多模态特征融合时产生的噪声,因此在图像-文本共注意融合模块之前构建了一个语义一致性评分模块,其目的是基于一致性程度来控制融合强度,并将融合后的特征与文本模态再次融合以增强文本信息的优势。还有一种策略是通过实体识别来发掘文本—图像关联<sup>[13]</sup>,例如,文本图像存在不一致性就是假新闻的有力佐证。而CHEN<sup>[14]</sup>等人认为当单模态特征很明显时,多模态融合反而会降低性能,因此通过自适应地聚合单模态特征和跨模态交互信息来提高假新闻检测的准确性,即在模态之间相似度较弱时依赖单模态特征,反之,在模态之间相似度较强时依赖跨模态交互信息。

本文所关注的新闻上下文信息主要基于参与新闻

传播的用户信息以及传播过程中用户发布的观点信息。MA<sup>[15]</sup>等人将事件相关的所有帖子按时间顺序构建为序列数据,并通过RNN从中挖掘传播特征。而面对社交网络中更丰富的关系型数据,在处理时面临着更大的挑战,目前常见的做法是将其建模为树或图,例如,BIAN<sup>[8]</sup>等人将帖子的转发链构建为传播树,并通过双向的图卷积网络分别挖掘谣言的深度传播和广度传播特征。WEI<sup>[16]</sup>等人提出消息在实际传播过程中会发生节点或关系的丢失,因此提出了一种高斯传播估计方法重建潜在的传播结构,并使用重建后的结构表征与原始结构表征共同检测假新闻。考虑现实世界的信息传播网络是动态变化的,为了捕获缺失的动态传播信息,SONG<sup>[9]</sup>等人使用离散时间下的动态图对新闻传播网络进行建模,并在每个静态图中捕获局部结构信息。类似的,还有针对连续时间下的动态图的研究<sup>[10]</sup>。SUN<sup>[17]</sup>等人指出传统方法很少从全局视角学习新闻和用户之间的联系,为此通过构建一个超图(Hypergraph)来捕获用户的全局行为,并提出了一个不依赖新闻内容和用户信息的学习模型。DOU<sup>[18]</sup>等人的做法是通过提取用户的历史帖子生成用户的偏好信息,并将其作为新闻源帖信息的补充。

不同于上述研究,基于多视图的方法侧重于根据新闻现有信息尽可能全面地挖掘新闻多个方面的特征。其中一些研究将传播过程中的用户、推文等作为节点,并构建异构图捕获节点间的交互。例如,RAN<sup>[19]</sup>等人从一个源帖—转发帖—词元—用户异构图中提取出3个子图,再通过一个多通道图注意力神经网络充分学习3个子图隐含的语义信息和传播模式。类似的,韩晓鸿<sup>[20]</sup>等人构建了一个推文—词—用户异质图。BAZMI<sup>[21]</sup>等人从用户、新闻内容和新闻来源3个视角对新闻进行表征,这3个视角分别对应3个编码器,并同样采用注意力机制进行视角间的交互。为了明确各视角间传递的信息,YING<sup>[22]</sup>等人训练了4个单视角(图像模式、图像语义、文本、多模态)分类器,并根据预测分数对各个视角进行重新加权以得到最终的表征,该模型还基于每



个视角给出了置信度分数,直观地展示各个视角在假新闻检测中的贡献,但该方法只着重于对新闻内容进行分解,缺少新闻在传播过程中的信息。

针对现有多视图方法缺少对多个视角信息的深度探索和融合的问题,本文提出一个基于新闻内容、用户和传播视角的多视图检测模型。首先,通过注意力机制实现模态间的信息交互,从而获取新闻内容视角的表征;然后,将构造出的新闻传播图分解为代表用户视角和传播视角的两个子图,并利用双向的图卷积网络分别挖掘不同视角的表征;最后,通过协同注意力机制实现对各视角表征的有效融合。

## 2 基于多视图表征的假新闻检测方法

### 2.1 问题定义

图1为新闻信息抽取过程,设本文所使用的新闻数据集中的每个新闻事件为 $N$ ,将每个新闻事件传播过程中产生的所有相关信息构建为一个新闻传播树。对于任意一个新闻事件可以将其表示为集合,即 $N=\{T,I,G\}$ , $T$ 和 $I$ 分别表示原始新闻中的文本和图片,图 $G=(V,E)$ 表示该新闻的传播过程,其中 $V$ 和 $E$ 分别为图的节点集和边集。通常情况下,新闻事件的传播由以新闻源帖为起点的多条转发链构成,除了根节点外,图中每个节点对应一个转发帖。即 $V=\{r_0,r_1,\dots,r_n\}$ , $r_0$ 为新闻传播树的根节点, $n$ 为除新闻源帖外与该新闻事件相关的所有帖子数量,对于每个节点 $r_i=\{u_i,p_i\}$ , $u_i$ 表示转发者, $p_i$ 为该转发帖中的用户观点,通常由一条较为简短的文本组成。图的边集 $E=\{e_{st}|s,t=0,\dots,n\}$ , $e_{st}$ 表示 $r_s$ 和 $r_t$ 之间存在转发关系。每个新闻事件都对应一个标签 $y=\{ture,false\}$ ,本文的任务为学习一个分类器将一个新闻事件映射到对应的分类标签 $y$ 上,即 $f:(T,I,G)\rightarrow y$ 。

图2展示了本文提出的MVRFD模型框架,该模型分为多视图表征、多视图特征提取、多视图融合及检测3个阶段。在模型的前两个阶段,根据关注的不同视角可分为两个部分,即内容视图模块和新闻上下文模块,其中后者包含了用户观点和用户信息两个视

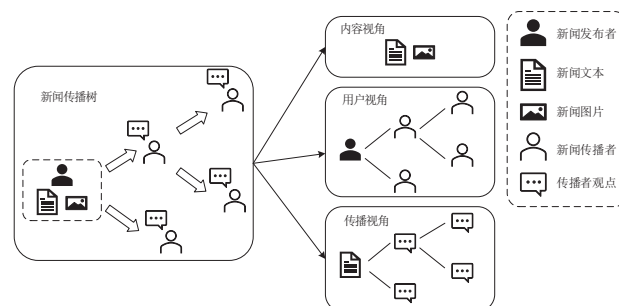


图1 新闻信息抽取过程

图的信息。经由上述两个模块生成的新闻多视图表征经过最后的多视图融合模块融合后即可得到完整的新闻表征,并输入到多视图检测模块进行虚假新闻分类。下文将从上述3个模块的角度详细介绍本文提出的模型。

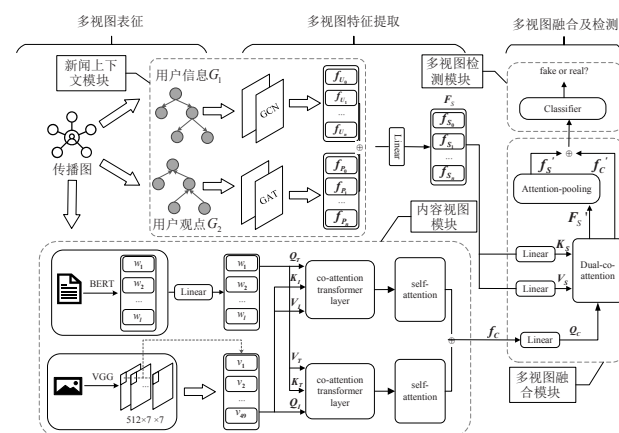


图2 MVRFD 模型框架

### 2.2 新闻内容模块

本文认为新闻的文本与图像中都包含了揭露假新闻的线索,而新闻中的文本和图像对于新闻的语义有着不同的理解模式。针对信息的浅层特征,首先,使用预训练模型分别提取文本与图像的基础语义信息;然后,通过注意力机制完成文本与图像模态之间的信息交互;最后,通过自注意力使单模态中的重要信息得到增强,使最后融合的特征拥有更全面的视角。

具体来说,新闻文本内容 $T$ 也为根节点 $r_0$ 中的 $p_0$ ,通过使用预训练BERT<sup>[5]</sup>表征文本的初级语义,使用VGG<sup>[23]</sup>表征新闻图片 $I$ 的视觉语义,如公式(1)和公式(2)所示。

$$F_T = \text{WordEmbed}_{\text{BERT}}(T) \quad (1)$$

$$F_I = \text{ImgEmbed}_{\text{VGG}}(I) \quad (2)$$

其中,本文参考文献[13]舍弃了VGG网络最后的铺平操作,而是将 $7 \times 7 \times 512$ 的多通道输入中各通道相同位置的值得拼接作为该区域的图像表征,因此得到了一个长度为49的视觉表征序列 $F_I = \{v_1, v_2, \dots, v_{49}\}$ , $v_i$ 表示图像的第 $i$ 个区域。文本表征 $F_T = \{w_1, w_2, \dots, w_l\}$ , $w_i$ 为第 $i$ 个词元, $l$ 为文本的有效长度。

注意力机制可以对信息的不同区域施加不同的重要性。在初步表征文本与图像后,本文使用基于Transformer的双流协同注意力模块<sup>[24]</sup>完成多模态信息交互,如图3所示。

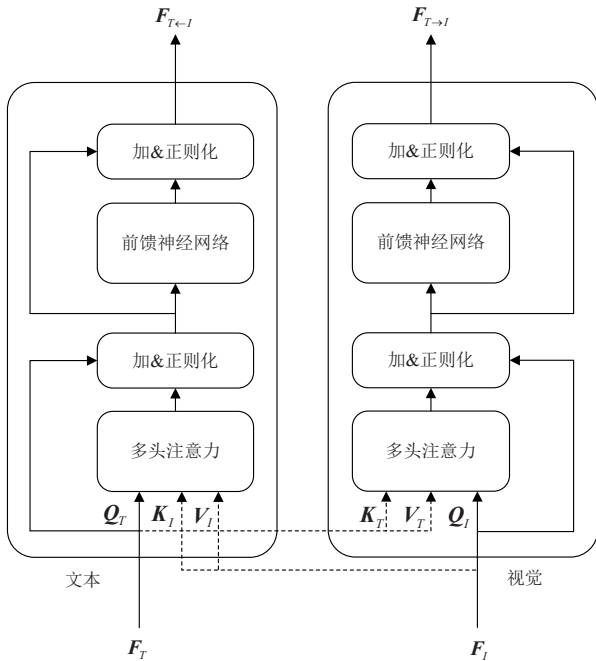


图3 协同注意力模块

在该模块中,每个模态的键和值来自另一个模态,例如,文本模态的查询、键和值分别定义为 $Q_T = F_T \times W_{Q_T}$ , $K_I = F_I \times W_{K_I}$ , $V_I = F_I \times W_{V_I}$ ,经过线性变换的查询、键和值被分为 $h$ 个注意力头,使模型关注不同方面的信息,注意力信息与原来的模态信息进行残差连接得到目标输出,然后将该输出序列的所有位置经过同一个前馈网络(Feed Forward Network, FFN)后使每个位置实现相同的变换,FFN计算过程

如公式(3)所示。

$$\text{FFN}(X) = \text{Linear}_2(\text{ReLU}(\text{Linear}_1(X))) \quad (3)$$

ReLU为使用的激活函数,此外,在注意力层和FFN层后均使用了一个正则化层LN加速模型的收敛,从视觉模态提取信息的文本模态特征 $F_{T \leftarrow I}$ 计算过程如公式(4)~公式(6)所示。

$$A_{T_i} = \text{softmax}(Q_{T_i} \times K_{I_i}^T / \sqrt{d_k / h}) V_{I_i} \quad (4)$$

$$A_T = \text{concat}(A_{T_1}, A_{T_2}, \dots, A_{T_h}) W_o \quad (5)$$

$$F_{T \leftarrow I} = \text{LN}(\text{FFN}(\text{LN}(A_T + F_T)) + \text{LN}(A_T + F_T)) \quad (6)$$

其中 $Q_{T_i}$ , $K_{I_i}$ , $V_{I_i}$ 均是基于 $Q_T$ , $K_I$ , $V_I$ 最后一个维度平均分割得到,分别用于表示第 $i$ 个注意力头的查询、键、值信息。同理,可以得到融合了文本特征的视觉特征 $F_{I \rightarrow T}$ 。模型在协同注意力模块后使用两个自注意力网络增强单个模态中的关键信息,如公式(7)和公式(8)所示。

$$\text{SelfAtt}_T(F_{T \leftarrow I}) = \{w'_1, w'_2, \dots, w'_l\} \quad (7)$$

$$\text{SelfAtt}_I(F_{I \rightarrow T}) = \{v'_1, v'_2, \dots, v'_{49}\} \quad (8)$$

为了使所有新闻的表征融合为一个固定长度的向量,各模态特征进行平均汇聚,因为图像被表征为一个固定长度的序列,所以此处计算平均值时除以固定值49,得到向量表征的文本特征 $f_T$ 以及视觉特征 $f_I$ ,如公式(9)和公式(10)所示。

$$f_T = \left( \sum_{i=1}^l w'_i \right) / l \quad (9)$$

$$f_I = \left( \sum_{i=1}^{49} v'_i \right) / 49 \quad (10)$$

因为新闻的文本内容长度不统一,所以上述注意力值计算过程中忽略了用于填充的词元。此外,对于不包含图片的新闻,使用与视觉表征相同维度的0作为填充,以使模型在实际应用中拥有更稳定的性能。最终,将文本特征与视觉特征连接作为新闻的内容视觉特征向量 $f_C$ ,如公式(11)所示。

$$f_C = f_T \oplus f_I \quad (11)$$

### 2.3 新闻上下文模块

新闻的上下文信息主要由新闻在社交网络传播的

过程中产生,如图1所示,其包含除新闻内容外的所有信息,也可归纳为用户信息(用户视角)和用户观点(传播视角)。BIAN<sup>[8]</sup>等人的研究考虑新闻传播的方向问题,即通过自上而下和自下而上的角度探索新闻的特征,但本文认为需要考虑新闻的不同视角,即不同视角的信息应该具有不同的传播方向。

### 2.3.1 用户视角

社交网络中存在许多容易被假新闻误导的用户,这些用户推动了假新闻的广泛传播,为了探究这些恶意用户或“易感染”用户对假新闻检测的影响,本文将参与新闻转发的所有用户构建为新闻传播图 $G$ 的一个有向子图 $G_1=(V_1, E_1)$ ,并定义为用户子图。如图2所示,子图节点集合 $V_1=\{u_0, u_1, \dots, u_n\}$ ,  $u_0$ 为新闻发布者。由于人们更倾向于追随与自己观点相似的用户,同时转发者与转发者往往是追随者与被迫追随者的关系,因此在用户视角中,被转发者向转发者施加的影响更为突出,即用户之间的影响自上而下传播。因此,定义该子图的边集 $E_1=\{e_{s \rightarrow t} | s, t=0, \dots, n\}$ ,  $e_{s \rightarrow t}$ 表示第 $t$ 个用户转发第 $s$ 个用户的帖子产生的一条有向边,该方向表示信息传递的方向。对于该用户子图,本文使用图卷积网络(Graph Convolutional Network, GCN)<sup>[25]</sup>实现节点间的消息传递。节点的特征 $X_1=\{x_1^0, x_1^1, \dots, x_1^n\}$ 从用户的粉丝数、好友数、关注数以及用户的验证信息、位置信息、描述信息等信息中抽取。其中,使用归一化处理数值类型信息,使用数值0或1表征布尔类型信息。

特别地,将用户是否只发表“转发微博”或不做任何表述的行为作为该用户的布尔类型信息之一,原因是这表示该用户完全赞同被转发者的观点,但仅从文本角度往往很难提取出这一含义。最终,传播过程中的每个用户都被表征为一个13维的向量(包括8个数值型信息和5个布尔类型信息)。子图 $G_1$ 的邻接矩阵 $A_1 \in \{0, 1\}^{(n+1) \times (n+1)}$ ,矩阵中所有位置元素 $a_{st}$ 如公式(12)所示。

$$a_{st} = \begin{cases} 1, & e_{s \rightarrow t} \in E_1 \\ 0, & \text{其他} \end{cases} \quad (12)$$

本文模型使用两层图卷积层,每层的消息传递过程如公式(13)所示。

$$H_i = \sigma(\hat{A}H_{i-1}W) = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H_{i-1}W\right) \quad (13)$$

其中,  $\sigma(\cdot)$ 为激活函数;  $\tilde{A} = A_1 + I$ , 即为图中每个节点加上自循环; 对角矩阵 $\tilde{D}$ 由子图中每个节点的度的数量得到, 即 $\tilde{D}_{ii} = \sum_j A_{ij}$ , 通过 $\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ 对 $\tilde{A}$ 进行归一化, 但在用户子图中, 节点只接收来自父节点的信息, 即此处除根节点外所有节点的度均为1;  $W$ 为权重矩阵, 表示第 $i$ 层的可学习参数。经过公式(13)计算得到第 $i$ 次传递后的节点特征 $H_i$ , 在开始时,  $H_0$ 被初始化为 $X_1$ 。每经过一层图卷积网络, 子图中每个节点都会聚合来自其父节点用户的信息, 本文使用两层图卷积网络挖掘用户特征, 最终所有的用户特征被表征为 $F_U = H_2 = \{f_{u_0}, f_{u_1}, \dots, f_{u_n}\}$ 。

### 2.3.2 传播视角

新闻在网络平台的传播主要是基于转发这一行为, 在这个过程中, 转发者往往会针对被转发者提供的信息发表自己的观点, 这些观点对假新闻的检测也有着重要的参考意义, 例如, 一个大部分转发者都对其持有消极态度的转发帖(或源帖)就应该是不可信的。本文定义另一个与用户子图 $G_1$ 有相同结构的有向子图 $G_2=(V_2, E_2)$ 作为观点子图, 子图中每个节点记录转发时产生的观点信息, 即 $V_2=\{p_0, p_1, \dots, p_n\}$ 。因为转发者发表的观点一般是针对被转发者的, 所以 $E_2=\{e_{s \leftarrow t} | s, t=0, \dots, n\}$ 为所有转发帖节点指向被转发帖节点的有向边的集合, 即与子图 $G_1$ 的方向完全相反。这些节点代表的用户观点是新闻在社交网络中传播后的反馈, 这些信息与传播图的结构共同构成了新闻传播视角中的重要线索。与子图 $G_1$ 中的节点只需考虑其父节点的信息不同的是, 子图 $G_2$ 中的一些节点会被多个节点转发, 因此会关联大量的子节点, 为了使父节点能够关注真正重要的信息, 使用图注意力网络(Graph Attention



Network, GAT)<sup>[26]</sup>实现节点的消息传递。对于观点子图,定义特征矩阵 $\mathbf{X}_2=\{\mathbf{x}_2^0, \mathbf{x}_2^1, \dots, \mathbf{x}_2^n\}$ ,  $\mathbf{X}_2$ 表示通过预训练的BERT提取的所有观点中的文本表征,具体就是将观点文本映射为输入序列中所有词元隐藏状态的平均池化后的向量。邻接矩阵 $\mathbf{A}_1$ 与 $\mathbf{A}_2$ 具有完全相反的方向,故 $\mathbf{A}_2$ 中的每个元素 $a_{st}$ 如公式(14)所示。

$$a_{st} = \begin{cases} 1, e_{s \leftarrow t} \in E_2 \\ 0, \text{其他} \end{cases} \quad (14)$$

在图注意力网络中,每层节点特征的更新过程如公式(15)~公式(17)所示。

$$e_{ij} = \text{LeakyReLU}(\omega^T [\mathbf{W}\mathbf{x}_2^i \parallel \mathbf{W}\mathbf{x}_2^j]) \quad (15)$$

$$\alpha_{ij} = \text{Softmax}(e_{ij}) \quad (16)$$

$$\mathbf{x}_2^{i'} = \text{LeakyReLU}\left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W}\mathbf{x}_2^j\right) \quad (17)$$

其中, $\omega^T$ 、 $\mathbf{W}$ 为相应层的权重参数, $\parallel$ 表示向量的连接操作, $\text{LeakyReLU}$ 为使用的激活函数, $N_i$ 为节点 $i$ 的所有邻接点的集合, $e_{ij}$ 为计算得到的相邻节点 $i$ 和 $j$ 之间的注意力系数,并使用 $\text{Softmax}$ 归一化后得到权重系数 $\alpha_{ij}$ ,该系数表示邻接点 $j$ 对于节点 $i$ 的重要性,通过该系数计算得到聚合了的所有邻接点信息后的节点 $i$ 的表征 $\mathbf{x}_2^{i'}$ ,并在这个过程中,对真正重要的观点施加了更多的注意力。同样经过两层图注意力网络,将所有节点特征 $\mathbf{F}_P$ 表示为 $\{\mathbf{f}_{p_0}, \mathbf{f}_{p_1}, \dots, \mathbf{f}_{p_n}\}$ 。

由于子图 $G_1$ 与 $G_2$ 之间的节点一一对应,将每个节点中的观点信息和用户信息合并经过一个线性层后作为整个新闻的上下文特征矩阵 $\mathbf{F}_S = \{\mathbf{f}_{s_0}, \mathbf{f}_{s_1}, \dots, \mathbf{f}_{s_n}\}$ ,如公式(18)所示。

$$\mathbf{F}_S = \text{LN}(\text{Linear}(\mathbf{F}_P \oplus \mathbf{F}_U)) \quad (18)$$

## 2.4 多视图融合和检测模块

给定一个新闻事件,通过新闻内容模块和新闻上下文模块,分别获取到新闻内容表征向量 $\mathbf{f}_C$ 和新闻上下文表征矩阵 $\mathbf{F}_S$ 。针对多视图特征融合问题,本文参照文献[21]使用双协同注意力网络实现视图间的交互。在这个注意力网络中,查询来自新闻内容特征 $\mathbf{f}_C$ ,键和值来自传播图中的所有节点特征 $\mathbf{F}_S$ , $e_i$ 由第 $i$ 个传播节点特征与新闻内容特征的点积运算得到,表示两者

之间的相关性得分,对于一个具有 $n$ 个节点的传播图,所有节点的相关性分数 $e \in R^n$ , $e_i$ 为通过 $\text{Softmax}$ 函数计算得到第 $i$ 个节点的注意力得分 $\alpha_i$ ,经过注意力计算的节点特征 $\mathbf{F}'_S$ 和内容特征 $\mathbf{f}'_C$ 计算过程如公式(19)~公式(21)所示。

$$\mathbf{f}'_C = \text{LN}\left(\text{ReLU}\left(\mathbf{f}_C + \sum_{i=1}^n \alpha_i \mathbf{v}_{S_i}\right)\right) \quad (19)$$

$$\mathbf{f}'_{S_i} = \text{LN}(\text{ReLU}(\mathbf{f}_{S_i} + \alpha_i \mathbf{q}_C)) \quad (20)$$

$$\mathbf{F}'_S = \{\mathbf{f}'_{S_1}, \mathbf{f}'_{S_2}, \dots, \mathbf{f}'_{S_n}\} \quad (21)$$

其中, $\mathbf{v}_{S_i}$ 、 $\mathbf{q}_C$ 分别为第 $i$ 个节点特征经过线性变换的值和由内容特征得到的查询。在与内容特征融合前,将一个新闻传播图中所有节点的特征进行平均池化,如公式(22)所示。

$$\mathbf{f}_{\text{news}} = \mathbf{f}'_C \oplus \text{AveragePooling}(\mathbf{F}'_S) \quad (22)$$

将新闻的最终表示 $\mathbf{f}_{\text{news}}$ 经过一个线性层计算得到预测结果,如公式(23)所示。

$$\hat{y} = \text{Softmax}(\mathbf{W}_f \mathbf{f}_{\text{news}} + \mathbf{b}_f) \quad (23)$$

$\mathbf{W}_f$ 为线性层的参数, $\mathbf{b}_f$ 为偏置向量。为了训练模型参数,使用交叉损失熵定义需要优化的目标函数,如公式(24)所示。

$$L = -\sum_{i=1}^m [y_i \log p_i + (1-y_i) \log(1-p_i)] + \lambda \|\theta\|_2^2 \quad (24)$$

其中, $m$ 为数据集中新闻的个数, $y_i \in \{0,1\}$ 是新闻的二分类标签, $p_i$ 表示模型预测的第 $i$ 个新闻为假新闻的概率,为防止模型过拟合,加入L2正则化项 $\|\theta\|_2^2$ , $\lambda$ 表示权重系数, $\lambda$ 需要根据多次的实验结果选取最优值。

## 3 实验和结果分析

### 3.1 数据集

为了挖掘新闻的传播特征和用户特征,本文使用了一个包含丰富信息的公开数据集来训练模型并验证模型的性能,即微博(Weibo)数据集<sup>[15]</sup>,其中的虚假信息来自社区管理中心,真实信息通过微博API从普通帖子中抓取。数据集的详细信息如表1所示,数据集中包含4664个新闻事件,其中假新闻与真实新闻数据约各占一半,每个事件都关联若干相关帖子。本

文还根据原始数据中提供的信息抓取了每个新闻附加的图片,并删除了已经损坏和低质量的图片,本次实验将数据集以7:1:2的比例分解为训练集、验证集和测试集。

表1 数据集统计信息

统计数据	数量
帖子	3805656 个
用户	2746818 个
图片	3845 张
新闻事件	4664 个
真实新闻	2351 个
虚假新闻	2313 个
事件的平均传播时间	2460.7 h
每个事件平均帖子数	816 个
单个事件最大帖子数	59318 个
单个事件最小帖子数	10 个
最小转发深度	2
最大转发深度	73
平均转发深度	6.19
转发深度超过5的事件数	2198 个

## 3.2 基线模型

本文将提出的模型与一些具有代表性的深度学习方法在相同的数据集以及实验环境下进行了比较。

1) LSTM<sup>[27]</sup>: 两层隐藏层的长短期记忆网络,使用最后一个时间步的隐藏层作为表征,并通过全连接层进行预测。

2) BERT<sup>[5]</sup>: BERT为本文所使用的预训练模型,但其同样能够进行分类任务。在本次实验中,冻结了BERT中所有层的参数,只更新最后的分类层的参数,并将最后一层的CLS词元嵌入作为新闻的文本表征用来进行二分类任务。

3) BERT+Attention: 通过BERT提出新闻文本中的词元表征,并使用自注意力机制增强其中的关键词元,使用平均池化后的文本表征进行假新闻检测。

4) TextCNN<sup>[28]</sup>: 可以应用到文本分类任务中的卷积神经网络,利用多个不同尺寸的卷积核来提取句子中的关键信息,从而能够更好地捕捉局部相关性。

5) CARMN<sup>[11]</sup>: 基于新闻内容的多模态检测方法,通过交叉注意力和自注意力机制表征新闻内容,并使用多通道卷积网络进一步提取文本特征。

6) Bi-GCN<sup>[8]</sup>: 基于新闻的传播行为以及内容构建

传播图,并使用一个双向的图卷积网络分别自下而上以及自上而下挖掘新闻的扩散特征和传播特征。

7) DA-GCN<sup>[29]</sup>: 基于新闻的文本语义和传播结构构建事件传播图,使用一种结合双重注意力机制和图卷积网络的检测方法,并基于注意力机制关注与假新闻检测相关的交互关系。

8) DAN-Tree<sup>[30]</sup>: 一种基于传播树结构的双注意力网络模型,将传播过程中的深度结构和语义信息有机地融合在一起,并采用路径过采样和结构嵌入的方法来增强深度结构的学习。

## 3.3 实验结果

表2展示了以上深度学习模型在微博数据集上的实验结果,本文提出的MVRFD模型在相同的实验环境和参数下的准确率达到96.7%,在其余各项指标上均优于最佳的基线模型。

表2 微博数据集实验结果

Model	准确率	虚假新闻			真实新闻		
		精确度	召回率	F1	精确度	召回率	F1
LSTM	87.8%	88.5%	87.7%	87.9%	86.5%	88.1%	87.0%
BERT	89.7%	88.7%	91.4%	89.9%	90.5%	87.9%	89.0%
BERT+Attention	91.1%	91.0%	91.8%	91.3%	91.3%	90.1%	90.6%
TextCNN	90.4%	90.3%	91.5%	90.5%	91.0%	89.9%	90.0%
CARMN	92.5%	92.2%	92.7%	92.4%	92.7%	92.3%	92.5%
Bi-GCN	94.4%	92.6%	96.8%	94.6%	96.6%	92.0%	94.1%
DA-GCN	94.4%	94.1%	94.6%	94.4%	94.7%	94.1%	94.4%
DAN-Tree	95.8%	94.6%	97.2%	95.8%	97.2%	94.5%	95.8%
MVRFD	96.7%	96.0%	97.7%	96.8%	97.5%	95.5%	96.4%

由表2可知,文本信息对假新闻检测有很大的贡献,且注意力机制也能够帮助模型捕捉更重要的文本特征以及实现多模态信息之间的信息互补。

随着视角的增加,模型检测假新闻的准确率在不斷上升,这表明通过提取新闻的完整内容信息和传播信息能够提高假新闻检测的效果。但从本文的结果来看,新闻图片对提升模型效果有限,例如,与去除图片信息的DAN-Tree对比,模型的性能提升较小,相反,上下文视角的加入有效地提升了预测的准确率,这一结论也体现在后面的消融实验中。

## 3.4 消融实验

为了进一步验证各个视角对于假新闻检测的贡



献,本文设计了3个用于消融实验的变体模型,分别为 $MVRFD_{mm}$ 、 $MVRFD_{view}$ 、 $MVRFD_{prop}$ 。

1)  $MVRFD_{mm}$ : 去掉新闻的内容视角,仅使用新闻的上下文视角;

2)  $MVRFD_{view}$ : 去掉新闻内容信息中的视觉模态,仅保留文本模态;

3)  $MVRFD_{prop}$ : 去掉上下文视角,仅使用新闻的内容视角。

由图4可知,新闻的上下文视角发挥了较大的作用,去除了上下文特征后的模型检测准确率下降了4.18%,而去除了新闻内容视角的模型性能并没有受到太大影响,尤其是仅去除图片信息的模型的准确率和 $F1$ 值仅分别下降了1.29%和1.36%,造成这一结果的原因可能是数据集中的图片质量较低或是模型未对图片做更加细粒度的处理,从而未充分挖掘文本与图片之间的互补关系。对于上下文视角,由于传播图的根节点中已经包含了新闻的文本内容信息,因此模型在GCN层或GAT层可能就已经获取了较全面的新闻信息。

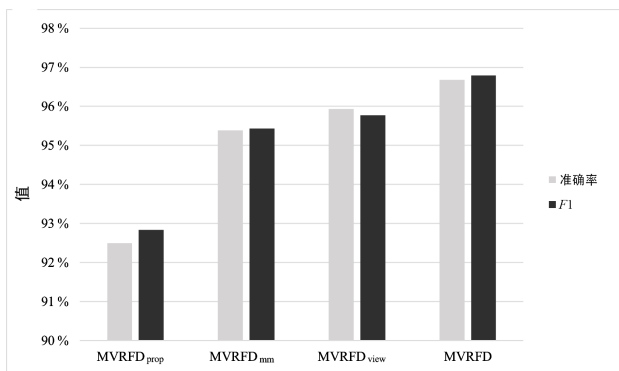
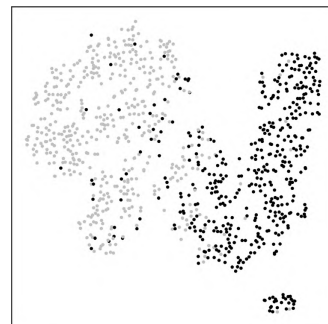


图4 模型去除不同部分后的性能

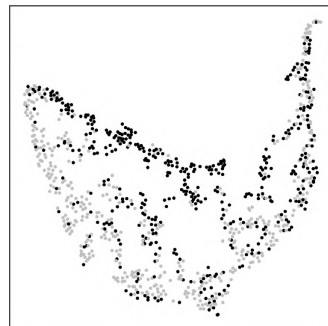
### 3.5 多视图表征可视化

为了进一步证明多视图表征对假新闻检测的帮助,将模型多视图融合模块输出的多视图表征结果运用t-SNE (t-Distributed Stochastic Neighbor Embedding) 算法进行可视化,即对1024维新闻的多视图表征向量进行降维,得到更便于观察的二维特征并在坐标图中展示。与之对比的是Bi-GCN和CARMN两个单视图模型,同样对去除分类模块后的模型输出使用t-SNE算法进

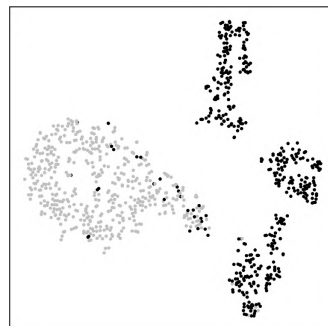
行降维处理。不同模型表征的可视化结果如图5所示,灰色和黑色的点分别表示假新闻和真实新闻。由图5可知,在本文提出的模型表征散点图中,表示假新闻的点与真实新闻的点有较为明显的区分,并且可以观察到MVRFD提取的特征比Bi-GCN具有更强的鉴别性,也稍强于CARMN。



a) CARMN 的表征结果



b) Bi-GCN 的表征结果



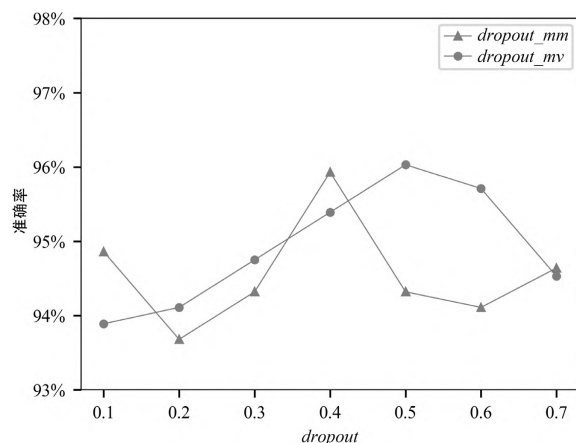
c) MVRFD 的表征结果

图5 不同模型表征的可视化结果

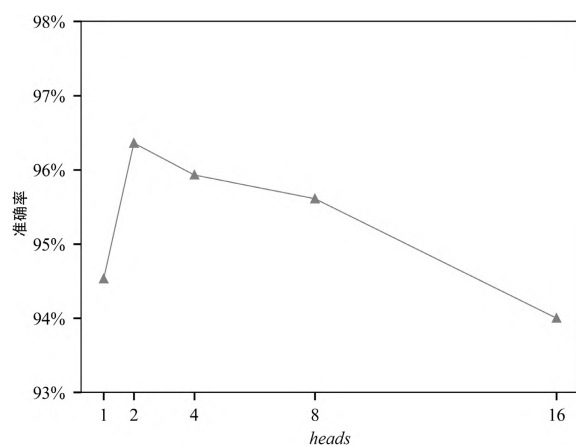
### 3.6 参数分析

本文模型使用Adam优化器更新参数,在模型的训练过程中, $batchsize$ 设置为64,学习率初始化为 $5 \times 10^{-4}$ ,权重衰减参数设置为 $10^{-4}$ ,进行最终分类前的

表征衰减率 *dropout* 设置为0.5, 训练迭代次数 *epoch* 设置为100。为了探究多模态表征时的 *dropout* 大小、多视图表征时的 *dropout* 大小以及注意力头数 *heads* 对模型性能的影响, 进行了参数实验并进行分析, 图6为模型准确率在不同参数下的变化, 其中 *dropout<sub>mm</sub>* 和 *dropout<sub>mv</sub>* 分别指在进行多模态表征和多视图表征注意力计算时被随机丢弃的神经元比例。为控制变量, 在调整一个 *dropout* 值时, 另一个 *dropout* 固定为  $10^{-6}$ , 可以看到在进行多模态表征时, *dropout* 的值0.4时准确率达到最优, 在进行多视图联合表征时, *dropout* 的值为0.5时达到最优, 注意力头数则在2时准确率达到最优。



a) dropout 对准确率的影响



b) heads 对准确率的影响

图6 部分参数对模型准确率的影响

## 4 结束语

本文提出了一种基于多视图表征下的假新闻检测

方法, 该模型能够在新闻内容、用户信息和新闻传播3个视角下提取新闻的表征, 并使用注意力机制和双向的图神经网络挖掘新闻的特征, 通过新闻多个视图的高层特征之间的交互和融合, 最终实现了假新闻的检测。在一个现实世界数据集上的实验结果表明, 本文模型的检测准确率优于对比的基准模型, 体现了基于新闻内容和上下文的多视图方法的有效性。在未来的工作中, 会进一步考虑新闻的文本和图片之间的关系, 从而改进新闻多模态数据的表征与融合方法。此外, 考虑从用户信息中提取更能表现用户喜好和社交关系的特征。

## 参考文献:

- [1] YANG Fan, YU Xiaohui, LIU Yang, et al. Automatic Detection of Rumor on Sina Weibo[C]//ACM. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2012: 1-7.
- [2] ZHAO Zhe, RESNICK P, MEI Qiaozhu. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts[C]//ACM. The 24th International Conference on World Wide Web. New York: ACM, 2015: 1395-1405.
- [3] HOOI B, SHAH N, BEUTEL A, et al. BIRDNEST: Bayesian Inference for Ratings-Fraud Detection[C]//SIAM. The 2016 SIAM International Conference on Data Mining. Philadelphia: SIAM, 2016: 495-503.
- [4] GIRGIS S, AMER E, GADALLAH M. Deep Learning Algorithms for Detecting Fake News in Online Text[C]//IEEE. The 13th International Conference on Computer Engineering and Systems. New York: IEEE, 2018: 93-97.
- [5] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding[C]//ACL. Proceedings of NAACL-HLT. Stroudsburg: ACL 2019, 1: 4171-4186.
- [6] DU Pengfei, LI Xiaoyong, GAO Yali. Survey on Multimodal Visual Language Representation Learning[J]. Journal of Software, 2021, 32(2): 327-348.
- [7] 杜鹏飞, 李小勇, 高雅丽. 多模态视觉语言表征学习研究综述[J]. 软件学报, 2021, 32(2): 327-348.
- [8] WU Ke, SONG Yang, ZHU K Q. False Rumors Detection on Sina Weibo by Propagation Structures[C]//IEEE. The 31st IEEE International Conference on Data Engineering. New York: IEEE, 2015: 651-662.
- [9] BIAN Tian, XIAO Xi, XU Tingyang, et al. Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks[C]//AAAI. Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2020, 34(1): 549-556.
- [10] SONG Chenguang, TENG Yiyang, ZHU Yangfu, et al. Dynamic Graph Neural Network for Fake News Detection[J]. Neurocomputing, 2021(505): 27-31.

- [10] SONG Chenguang, SHU Kai, WU Bin, et al. Temporally Evolving Graph Neural Network for Fake News Detection[J]. *Information Processing & Management*, 2021, 58(6): 102712.
- [11] SONG Chenguang, NING Nianwen, ZHANG Yunlei, et al. A Multimodal Fake News Detection Model Based on Crossmodal Attention Residual and Multichannel Convolutional Neural Networks[J]. *Information Processing & Management*, 2021, 58(1): 102437.
- [12] XIONG Shufeng, ZHANG Guipeng, BATRA V, et al. TRIMOON: Two-Round Inconsistency-Based Multi-Modal Fusion Network for Fake News Detection[J]. *Information Fusion*, 2023, 93: 150–158.
- [13] QI Peng, CAO Juan, LI Xirong, et al. Improving Fake News Detection by Using an Entity-Enhanced Framework to Fuse Diverse Multimodal Clues[C]//ACM. *Proceedings of the 29th ACM International Conference on Multimedia*. New York: ACM, 2021: 1212–1220.
- [14] CHEN Yiyuan, LI Dongsheng, ZHANG Peng, et al. Cross-Modal Ambiguity Learning for Multimodal Fake News Detection[C]//ACM. *The ACM Web Conference*. New York: ACM, 2022: 2897–2905.
- [15] MA Jing, GAO Wei, MITRA P, et al. Detecting Rumors from Microblogs with Recurrent Neural Networks[C]//ACM. *The 25th International Joint Conference on Artificial Intelligence*. New York: ACM, 2016: 3818–3824.
- [16] WEI Lingwei, HU Dou, ZHOU Wei, et al. Uncertainty-Aware Propagation Structure Reconstruction for Fake News Detection[C]//ACL. *The 29th International Conference on Computational Linguistics*. Stroudsburg: ACL, 2022, 29(1): 2759–2768.
- [17] SUN Ling, RAO Yuan, LAN Yuqian, et al. HG-SL: Jointly Learning of Global and Local User Spreading Behavior for Fake News Early Detection[C]//AAAI. *Proceedings of the 37th AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI, 2023, 37(4): 5248–5256.
- [18] DOU Yingdong, SHU Kai, XIA Congying, et al. User Preference-Aware Fake News Detection[C]//ACM. *The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM, 2021: 2051–2055.
- [19] RAN Hongyan, JIA Caiyan, ZHANG Pengfei, et al. MGAT-ESM: Multi-Channel Graph Attention Neural Network with Event-Sharing Module for Rumor Detection[J]. *Information Sciences*, 2022, 592: 402–416.
- [20] HAN Xiaohong, ZHAO Mengfan, ZHANG Yutao. Joint Heterogeneous Graph Convolutional Network and Attention Mechanism for Fake News[EB/OL]. (2022–11–28)[2023–12–01]. <https://doi.org/10.20009/j.cnki.21-1106/TP.2022-0412>.
- [21] BAZMI P, ASADPOUR M, SHAKERY A. Multi-View Co-Attention Network for Fake News Detection by Modeling Topic-Specific User and News Source Credibility[J]. *Information Processing and Management*, 2023, 60(1): 103146.
- [22] YING Qichao, HU Xiaoxiao, ZHOU Yangming, et al. Bootstrapping Multi-View Representations for Fake News Detection[C]//AAAI. *The 37th AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI, 2023, 37(4): 5384–5392.
- [23] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]//IEEE. *The 3th International Conference on Learning Representations*. New York: IEEE, 2015: 730–734.
- [24] LU Jiasen, BATRA D, PARIKH D, et al. ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks[C]//ACM. *The 33rd International Conference on Neural Information Processing Systems*. New York: ACM, 2019: 13–23.
- [25] KIPF T N, WELING M. Semi-Supervised Classification with Graph Convolutional Networks[C]//IEEE. *The 5th International Conference on Learning Representations*. New York: IEEE, 2017: 11305–11312.
- [26] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[EB/OL]. (2017–10–30)[2023–12–01]. <https://arxiv.org/abs/1710.10903>.
- [27] SINGH J P, KUMAR A, RANA N P, et al. Attention-Based LSTM Network for Rumor Veracity Estimation of Tweets[J]. *Information Systems Frontiers*, 2022, 24(2): 459–474.
- [28] KIM Y. Convolutional Neural Networks for Sentence Classification[C]//ACL. *2014 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: ACL, 2014: 1746–1751.
- [29] LIU Xiaoyang, ZHAO Zhengyang, ZHANG Yihao, et al. Social Network Rumor Detection Method Combining Dual-Attention Mechanism with Graph Convolutional Network[J]. *IEEE Transactions on Computational Social Systems*, 2023, 10(5): 2350–2361.
- [30] BAI Lin, HAN Xueming, JIA Caiyan. A Rumor Detection Model Incorporating Propagation Path Contextual Semantics and User Information[J]. *Neural Processing Letters*, 2023, 55(7): 9831–9850.