

一种基于动态异构图的谣言检测模型

朱文龙^{1,2}, 陈羽中^{1,2*}, 饶孟宇²

¹ (福州大学 计算机与大数据学院, 福州 350116)

² (福建省网络计算与智能信息处理重点实验室, 福州 350116)

E-mail: yzchen@fzu.edu.cn

摘 要: 随着互联网技术和自媒体行业的快速发展, 人们可以方便快捷地从社交媒体中获取最新信息, 但也让更多的谣言在网络中盛行。现有谣言检测模型多从文本内容、用户信息和传播模式中挖掘有效特征。然而, 现有模型未充分学习文本的语义信息和谣言传播过程中的结构信息, 并忽略了谣言传播的动态过程。针对上述问题, 本文提出一种基于动态异构图的谣言检测模型 DHGNN (Dynamic Heterogeneous Graph Neural Network)。首先, 为了增强帖子的文本语义表示, 本文提出一种多级注意力网络, 引导模型关注源帖子和相应评论中关键的词和句子, 充分学习源帖与相应评论之间的语义关联。其次, 引入了基于异构图的图神经网络, 通过对异构传播图中的用户、帖子节点和转发(或评论)关系进行建模, 为不同类型的节点和边生成特定的表示, 充分学习异构传播图中的结构信息。最后, 提出一种基于旋转记忆单元的时序注意力, 分别为每个异构传播图快照建立记忆, 捕获谣言动态传播的演化模式。在 Twitter15、Twitter16 数据集上的实验结果表明, DHGNN 模型的性能优于最新的对比模型。

关键词: 谣言检测, 多级注意力, 异构传播图, 图神经网络, 时序注意力

中图分类号: TP391

文献标识码: A

A Rumor Detection Model based on Dynamic Heterogeneous Graph Neural Network

ZHU Wen-long^{1,2}, CHEN Yu-zhong^{1,2*}, RAO Meng-yu²

¹ (College of Computer and Data Sciences, Fuzhou University, Fuzhou 350116, China)

² (Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou 350116, China)

Abstract: With the rapid development of Internet technology and the self-media industry, people can easily and quickly obtain the latest information from social media, but it also allows more rumors to prevail on the Internet. Most existing rumor detection models focused on mining effective features from textual content, user profiles, and propagation patterns. However, these methods do not fully learn the semantic information of the text and structural information in the process of rumor propagation. They also ignore the dynamic process of rumor propagation. To address these issues, this paper proposes a rumor detection model based on dynamic heterogeneous graph neural network (DHGNN). First, we propose a multi-level attention network to enhance the semantic representation of posts. The multi-level attention network can guide DHGNN to pay more attention to key words and sentences in the source posts and comments. Therefore, DHGNN can fully learn the semantic relationships between the source posts and the corresponding comments. Second, we propose a heterogeneous graph neural network by modeling users, post nodes, and retweet (or comment) relationships in a heterogeneous propagation graph. DHGNN can generate specific representations for different types of nodes and edges and fully learn structural information in heterogeneous propagation graphs. Third, we propose a temporal attention mechanism based on the rotating memory unit, which can establish memory for each heterogeneous propagation graph snapshot and capture the dynamic evolution pattern of rumor propagation. Experimental results on Twitter15, Twitter16 datasets show that DHGNN outperforms other state-of-the-art models.

Key words: Rumor Detection, Multi-level Attention, Heterogeneous Propagation Graph, Graph Neural Network, Temporal Attention

收稿日期: 2020-01-01 基金项目: 国家自然科学基金(61672158, 61972097, U21A20472)资助; 福建省科技重大专项专题项目(科教联合)(2021HZ2022007)资助, 福建省高校产学研合作项目(2021H6022)资助; 福建省自然科学基金(2020J01494)资助。 作者简介: 朱文龙, 男, 1997年生, 硕士研究生, CCF 会员(I0561G), 研究方向为自然语言处理; 陈羽中(通讯作者), 男, 1979年生, 博士, 教授, CCF 会员, 研究方向为数据挖掘、自然语言处理; 饶孟宇, 男, 2000年生, 本科生, CCF 会员, 研究方向为自然语言处理。

1 引言

随着互联网的迅速发展, 微博、Twitter 等社交媒体逐渐成为人们生活中不可缺少的交流工具。同时, 由于社交媒体传播速度快、范围广, 谣言也逐渐成为人们关注的一大焦点。在社交媒体中, 谣言是一种以某些途径广泛传播且未经证实的信息表述^[1]。谣言的出现往往会给社会造成严重的危害和巨大的经济损失。尽管各社交媒体机构陆续开通了辟谣平台来加强对谣言的管控, 但这些辟谣平台仍然需要依靠人工收集的信息来识别谣言真实性, 非常耗时耗力。因此, 迫切需要一种能够有效检测谣言的自动化方法。

传统的谣言检测方法主要利用特征工程从文本内容、用户资料和传播模式中提取易识别的特征来训练有监督学习分类器^[2-7], 例如随机森林、支持向量机、决策树等。然而, 特征工程提取特征往往十分费时费力, 并很难应对复杂的社交媒体数据。近年来, 深度学习在自然语言处理领域取得了优异的性能表现, 能够让模型自动学习谣言的深层特征表示, 例如循环神经网络和卷积神经网络。通过将文本内容建模为时间序列, 循环神经网络可以捕获每个帖子及其评论的语义变化^[8]。卷积神经网络可以捕获文本内容中的局部语义信息^[9]。但是, 上述方法仅从传播序列中学习时间特征, 忽略了传播内部的拓扑关系。因此, 一些研究人员研究了基于文本语义信息及其传播结构的谣言检测方法^[13-22]。通过对帖子的传播结构建模, 构建传播图, 并采用图神经网络聚合邻居节点的特征, 从而捕获谣言的传播结构特征。此外, 一些研究人员结合多种异构信息构建异构传播图, 通过学习不同类型节点和边丰富的结构信息来提高谣言检测的性能^[23-25]。

尽管基于文本语义信息及其传播结构的谣言检测方法取得了显著的进展, 但仍然存在一些亟待解决的问题。首先, 现有基于异构传播图的谣言检测模型通常使用常规的同构图神经网络, 不能充分学习传播图中的结构信息, 忽略了异构传播图中节点特定于自身类型的信息。其次, 现有模型往往忽略了谣言传播的动态过程, 无法捕获谣言动态传播的演化模式。最后, 现有模型无法充分学习帖子与相关评论之间的语义关联, 导致文本语义信息损失。针对上述问题, 本文提出了一种基于动态异构图的谣言检测模型 DHGNN(Dynamic Heterogeneous Graph Network With Multi-level Attention), 主要贡献如下:

1) DHGNN 引入基于异构图的图神经网络。该网络对异构传播图中的用户、帖子节点和转发(或评论)关系进行建模, 通过对不同类型的节点和边设置不同的共享参数, 来表征每个边上的异构性注意力, 为不同类型的节点和边生成特定的表示, 学习谣言传播过程中所有用户之间的潜在交互, 从而充分学习异构传播图中的结构信息。

2) DHGNN 提出一种基于旋转记忆单元(Rotational Unit

of Memory, RUM)的时序注意力。该注意力通过 RUM 为每个异构传播图快照建立记忆, 模拟谣言在社交媒体中的动态传播过程, 从而学习传播过程中的动态信息, 且能进一步捕获谣言动态传播的演化模式。

3) DHGNN 提出一种多级注意力机制。该机制通过在词级和句子级中应用多头注意力, 引导模型关注源帖和相应评论中的关键词和关键句, 充分学习源帖与评论之间的语义关联, 从而增强源帖的语义表示。

4) 本文在两个公共数据集上进行了充分实验验证 DHGNN 的性能。与最先进的谣言检测模型相比, DHGNN 在所有数据集上均取得领先的性能。此外, 本文进一步通过消融实验分析了 DHGNN 的各模块对其整体性能的贡献。

2 相关工作

近年来, 国内外很多学者对谣言检测领域进行了深入的研究, 并取得了出色的成果。谣言检测相关研究主要可以分为两类: 基于传统机器学习的检测方法与基于深度学习的检测方法。

传统的机器学习方法需要从社交媒体中提取有效的特征送入分类器对谣言进行检测。Yang 等人^[3]通过提取文本内容、地理位置、转发次数和评论次数等特征来训练 SVM 分类器, 该分类器使用了径向基核函数(Radial Basis Function, RBF)。Castillo 等人^[4]对社交媒体上的热门话题进行统计分析, 提出基于文本特征、用户特征和传播特征的决策树模型。Kwon 等人^[2]通过考虑一些特殊时间节点对谣言产生的影响, 如节假日、美国大选等, 提出了一种基于时间特征的随机森林模型。Ma 等人^[5]应用时间序列建模技术来整合各种社会背景信息, 通过考虑特定话题事件的整个生命周期来判断是否为谣言。Wu 等人^[6]对传播树进行建模, 使用具有不同核的支持向量机来检测谣言。Zhao 等人^[7]通过发掘谣言中隐藏的关键字特征, 并结合 K-means 聚类方法和 SVM 分类器进行谣言检测。上述方法虽然取得了一定的进展, 但严重依赖手工的特征工程来提取特征集, 模型经常缺乏泛化能力, 费时费力。

随着深度学习方法研究的不断深入, 许多研究人员也将深度学习模型应用到谣言检测领域中。Ma 等人^[8]利用循环神经网络来捕获每个源帖子及其转发评论之间的语义变化, 并根据语义变化进行预测。Yu 等人^[9]通过卷积神经网络捕获文本内容的局部语义, 并获得隐藏单元的深层文本特征。Pan 等人^[10]在经典的文本卷积神经网络(TextCNN)上引用注意力机制, 为卷积层提取的每个特征赋予不同权重来进行谣言事件检测。Liu 等人^[11]将每个新闻的传播路径建模成一个多元时间序列, 其中每个元组是一个数值向量, 表示参与传播新闻的用户特征。Wang 等人^[12]采用自动构造方法构建的情感字典来捕获人们对不同事件的细粒度情感反应, 并通过两步动态时间序列算法保留事件的时间跨度分布信息。上述这些方法都是将帖子及其评论建模为时间序列进行处理, 而忽略了谣言传播内部的拓扑关系。

随着图神经网络的快速发展,一些研究学者研究了基于帖子语义信息和传播结构的谣言检测方法。Ma 等人^[13]利用自顶向下和自低向上的树形递归神经网络来捕获谣言的语义信息和传播特征。Khoo 等人^[14]使用 Transformer 网络中的多头注意力机制对帖子之间的长距离交互进行建模,提出一种帖子级注意力模型。Bian 等人^[15]利用双向图卷积网络学习谣言的传播模式,捕捉谣言扩散的结构。Ma 等人^[16]提出一种基于树 Transformer 的检测模型,利用对话中的用户交互来聚合子树的观点。Wu 等人^[17]为每个谣言构建传播图,并使用门控图神经网络对关系路径上的相邻节点交换信息,更新节点的表示,从而增强传播图的节点表示。Tu 等人^[18]提出联合图的概念,构建了一个集成所有推文传播关系的大图,并使用基于 CNN 的模型捕捉推文的文本内容特征和传播结构信息。Luo 等人^[19]提出一种新的顺序编码方法,将源推文的传播树嵌入向量空间,并使用预训练的

Transformer 提取源推文的上下文语义特征。Wei 等人^[20]首次尝试探究传播过程中的不确定性,利用贝叶斯方法自适应地学习传播结构中潜在关系的可靠性,并设计一个新的边一致性训练框架来优化模型。

此外,一些研究人员则尝试构建异构图来模拟谣言传播过程提供新的视角,并取得了出色的性能表现。Yuan 等人^[23]将推文、转推和用户的全局关系建模成异构图,然后从帖子的内容和用户的传播关系中捕获局部语义信息和全局结构信息。Huang 等人^[24]构建了推文-词-用户异构图,利用基于元路径的异构图注意力网络捕获文本内容的全局语义关系和推文传播的全局结构信息。Ran 等人^[25]提出一种多通道图注意力网络,该网络分别对源推文与其评论的关系、源推文与其词的关系、源推文与其相关用户的关系进行建模构建三个子图,并设计一个事件共享模块学习不同事件谣言中的共性特征。

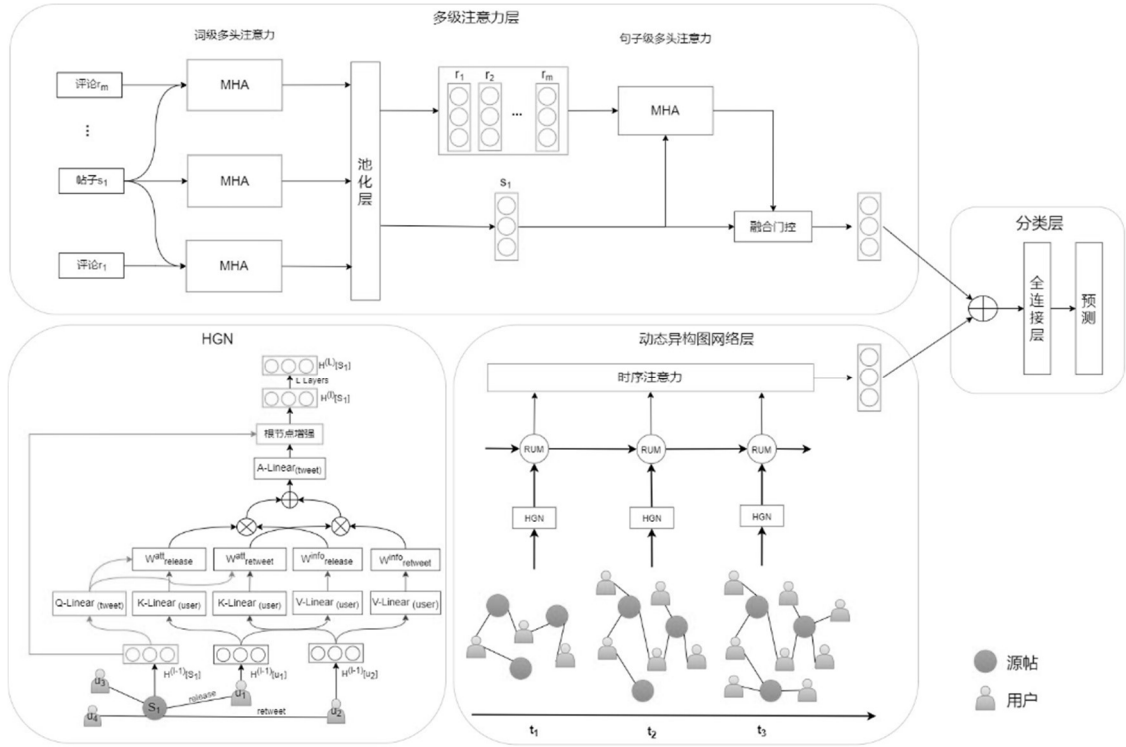


图 1 DHGNN 整体架构

Fig. 1 Overall architecture of DHGNN

3 模型

3.1 任务定义

给定源帖集合 $S = \{s_1, \dots, s_i, \dots, s_N\}$ 和用户集合 $\mathcal{U} = \{u_1, \dots, u_j, \dots, u_{|\mathcal{U}|}\}$, 其中 N 表示帖子的总数量。每篇源帖 s_i 的评论集 R_i 由 m 条评论帖子组成, 可表示为 $R_i = \{r_1, r_2, \dots, r_m\}$ 。谣言检测任务是学习一个函数 $f: (s_i, R_i, \mathcal{U}) \rightarrow y$, 预测源帖是否为谣言, y 表示所属的类别标签。其中, 类别标签包括非谣言、假谣言、真谣言、未经证实的谣言。

3.2 模型架构

本文提出的 DHGNN 模型的整体架构如图 1 所示, 主要包括多级注意力层、动态异构图网络层以及分类层。多级注意力层通过词级多头注意力和句子级多头注意力学习源帖与评论之间潜在的语义关联, 得到源帖基于评论的细粒度语义表示。动态异构图网络层的主要任务是在谣言的动态传播过程中同时捕获结构信息和动态演化模式。动态异构图网络层由两部分组成, 分别为基于异构图的图神经网络和时序注意力。通过图神经网络充分学习异构传播图快照的结构信

息,同时,利用时序注意力来捕获谣言动态传播的演化模式,得到帖子的动态信息表示。最后,分类层将源帖的细粒度语义表示与动态信息表示拼接,通过学习分类函数来预测源帖的标签。

3.3 多级注意力层

谣言的指示特征往往可以从相关用户的评论中获取,评论中表达的语义信息可以进一步增强源帖的语义表示。其中,具有强烈立场的评论在确定谣言的真实性时往往有着至关重要的作用。为了充分学习帖子与相关评论之间的语义关联,DHGNN 采用多级注意力机制选择性地关注相应评论中具有重要谣言指示特征的词和句子,从而得到帖子基于评论的细粒度语义表示。多级注意力层由词级多头注意力和句子级多头注意力组成。以下进行详细说明。

3.3.1 词级多头注意力

给定一条帖子 s_i 和相应评论集合 $R_i = \{r_1, r_2, \dots, r_m\}$,使用预训练的词嵌入模型进行初始化得到对应的词嵌入表示,即 $s_i = (w_1 \dots w_t \dots w_{|s_i|})$, $r_j = (w_1 \dots w_t \dots w_{|r_j|})$,其中每个 $w_t \in \mathbb{R}^d$ 是一个 d 维向量。为了探究评论中的哪些信息可以作为谣言的指示特征,从而引导 DHGNN 模型专注于评论的一些关键单词或短语。DHGNN 模型将帖子 s_i 作为矩阵 Q ,相应的评论 $R_i = \{r_1, r_2, \dots, r_m\}$ 作为 K 和 V ,分别对帖子的每一条评论 r_j 进行多头交叉注意力,将帖子的每个单词作为指导,引导模型关注到每条评论中关键的单词或短语,计算公式如下:

$$c'_j = \text{MultiHead}(s_i, r_j, r_j) \quad (1)$$

$$\begin{aligned} & \text{MultiHead}(Q', K', V') = \\ & \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W_o \end{aligned} \quad (2)$$

$$\text{head}_i = \text{Attention}(Q'W_i^Q, K'W_i^K, V'W_i^V) \quad (3)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (4)$$

其中, h 是并行头数, \sqrt{d} 是比例因子, $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d \times d/h}$ 是线性投影的权重矩阵, r_j 表示帖子 s_i 的第 j 条评论。

为了学习帖子中单词之间的依存关系,并提取句子内部的结构信息。将帖子 s_i 输入多头自注意力,从而关注源帖中的关键词。最后,通过池化层得到帖子 s_i 和相应评论 r_j 的句子级表示:

$$x'_i = \text{MultiHead}(s_i, s_i, s_i) \quad (5)$$

$$x_i = \text{pool}(x'_i) \quad (6)$$

$$c_j = \text{pool}(c'_j) \quad (7)$$

其中, $x_i \in \mathbb{R}^d$, $c_j \in \mathbb{R}^d$, $\text{pool}(\cdot)$ 表示池化函数,常用的池化函数有最大池化和平均池化。

3.3.2 句子级多头注意力

从公式(7)可以得到相应评论的句子表示,将所有评论拼接起来得到评论表示 $C \in \mathbb{R}^{m \times d}$:

$$C = \text{Concat}(c_1, \dots, c_m) \quad (8)$$

在现实生活中,人们判断一个帖子是否是谣言,往往会专注于一些具有关键信息的评论,而忽略其他微不足道或毫无关系的评论。因此,DHGNN 模型需要专注于有意义的评论句子。DHGNN 使用句子级多头交叉注意力产生不同的注意力分布,来表示不同评论对确定帖子真实性的不同贡献。

首先,复制 m 次帖子的句子表示 x_i ,可得到矩阵 $X \in \mathbb{R}^{m \times d}$,通过多头交叉注意力关注重要的评论句,然后通过池化层得到所有评论的语义表示,公式如下:

$$C' = \text{MultiHead}(X, C, C) \quad (9)$$

$$o = \text{pool}(C') \quad (10)$$

其中, $C' \in \mathbb{R}^{m \times d}$ 。之后,通过门控机制将帖子的句子表示 x_i 与相应评论的综合语义表示 o 相融合,得到源帖基于评论的细粒度语义表示 D ,公式如下:

$$\alpha = \sigma(w_1 x_i + w_2 o + b) \quad (11)$$

$$D = x_i \odot \alpha + o \odot (1 - \alpha) \quad (12)$$

其中, $\sigma(\cdot) = \frac{1}{1 + \exp(\cdot)}$ 是 sigmoid 激活函数, $w_1, w_2 \in \mathbb{R}^{d \times 1}$

和 $b \in \mathbb{R}$ 是门控机制的可学习参数。

3.4 动态异构图网络层

3.4.1 异构传播图

具有特定特征用户之间的相关性往往可以帮助揭示一条帖子是否为谣言。本文通过构建异构传播图来模拟用户与源帖之间的交互关系。具体而言,异构传播图 $G = (V, E)$,节点集 V 包含源帖节点、发布源帖的用户节点和转发(或评论)源帖的用户节点,边集合 E 包含发布、转发和评论关系。然后,对所有节点进行统一编号,让邻接矩阵包含所有关系。

在社交媒体中,一个新的热门事件往往会引发人们的广泛关注和热议。因此,随着时间的推移,不断会有新的用户转发(或评论)源帖。为了进一步模拟社交媒体中帖子传播的动态过程,本文设置 T 个快照时间点,根据用户转发(或评论)源帖的时间构建 T 个异构传播图快照 G_1, \dots, G_T ,如图2所示。其中, $G_t = (V_t, E_t)$ 表示为第 t 个异构传播图快照。

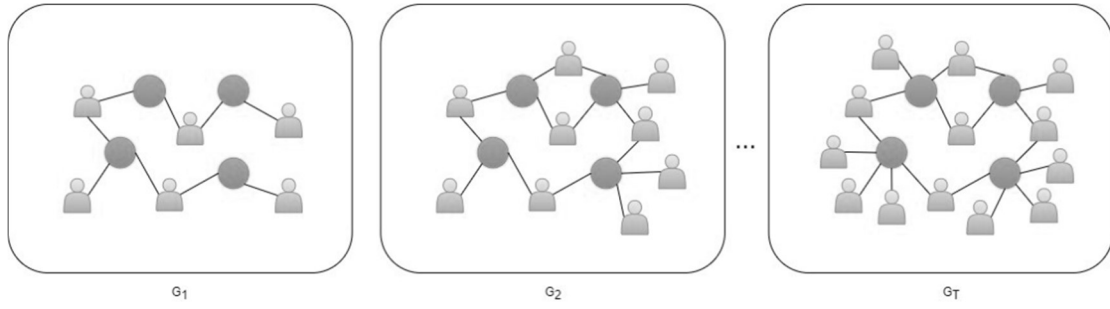


图2 异构传播图快照

Fig. 2 Snapshot of the heterogeneous propagation graph

3.4.2 图神经网络

每个异构传播图快照都可以被视为一个静态异构传播图，为了充分学习快照的传播结构信息，学习所有用户之间的潜在交互。本文引入基于异构图的图神经网络 (Heterogeneous Graph Neural Network, HGN)，通过多层 HGN 对异构传播图的节点和边进行建模，为不同类型的节点和边生成特定的表示。

首先，为了最大程度建模不同类型节点的分布差异性，HGN 为不同类型的节点提供了相应的线性映射函数。同时，由于异构传播图中边类型的多样性，为了捕获不同节点之间的语义关联，HGN 为不同类型的边设置了不同的参数化权重矩阵，进而在保留不同关系特性的同时尽可能的实现参数共享。对于第 t 个异构传播图快照 $G_t = (V_t, E_t)$ ，计算源帖子节点 s_i 与所有邻接用户节点集合 $N(s_i)$ 的注意力分布，公式如下：

$$Q^k(s_i) = Q_Linear_{\tau(s_i)}^k \left(H_t^{(l-1)}[s_i] \right) \quad (13)$$

$$K^k(u) = K_Linear_{\tau(u)}^k \left(H_t^{(l-1)}[u] \right) \quad (14)$$

$$head^k(u, e, s_i) = \left(\frac{K^k(u) W_{\phi(e)}^{att} Q^k(s_i)^T}{\sqrt{d}} \right) \quad (15)$$

$$Attention_{HGN}(u, e, s_i) = \text{Softmax} \left(\bigoplus_{k \in [1, H]} head^k(u, e, s_i) \right), \forall u \in N(s_i) \quad (16)$$

其中， $\tau(s_i)$ 表示节点 s_i 的类型， $\phi(e)$ 表示边 e 的类型。 $Q_Linear_{\tau(s_i)}^k(\cdot)$ ， $K_Linear_{\tau(u)}^k(\cdot)$ 分别表示为 $\tau(s_i)$ 类型和 $\tau(u)$ 类型节点的线性映射函数， $W_{\phi(e)}^{att} \in \mathbb{R}^{d/H}$ 表示 $\phi(e)$ 类型边的权重矩阵， \bigoplus 为拼接操作， H 为并行头数。

同时，源帖节点从相邻用户节点学习信息的过程可以看作是相邻用户节点向源帖节点信息传递的过程，为了缓解不同类型节点和边分布的差异性，HGN 将不同边关系合并到信息传递的过程中。对于节点对 $e = (u, s_i)$ ，信息传递的计算过程如下：

$$Info_head^k(u, e, s_i) = V_Linear_{\tau(u)}^k \left(H_t^{(l-1)}[u] \right) W_{\phi(e)}^{info} \quad (17)$$

$$Info_{HGN}(u, e, s_i) = \text{Concat}(Info_head^1(u, e, s_i), \dots, Info_head^H(u, e, s_i)) \quad (18)$$

其中， $V_Linear_{\tau(u)}^k$ 表示 $\tau(u)$ 类型节点的线性映射函数， $\text{Concat}(\cdot)$ 为拼接函数， $W_{\phi(e)}^{info} \in \mathbb{R}^{d/H}$ 表示 $\phi(e)$ 类型边的权重矩阵。

最后，得到源帖节点与所有邻接用户节点的注意力分布后，源帖节点使用注意力系数作为权重聚合所有邻接用户节点的信息，得到源帖节点 s_i 更新后的向量表示：

$$\tilde{H}_t^{(l)}[s_i] = \sum_{\forall u \in N(s_i)} (Attention_{HGN}(u, e, s_i) \cdot Info_{HGN}(u, e, s_i)) \quad (19)$$

其中， t 表示第 t 个异构传播图快照。

谣言事件的源帖通常包含大量的信息，因此起着至关重要的作用。为了更好地从节点对中学习更精确的节点表示，DHGNN 引入根节点增强来帮助提升模型的谣言检测性能。如下得到源帖节点 s_i 在第 l 层 HGN 的节点表示 $H_t^{(l)}[s_i]$ ：

$$H_t^{(l)}[s_i] = A_Linear_{\tau(s_i)} \left(\sigma \left(\tilde{H}_t^{(l)}[s_i] \right) + H_t^{(l-1)}[s_i] \right) \quad (20)$$

其中， $A_Linear_{\tau(s_i)}(\cdot)$ 表示将源帖子节点映射回对应类别的特征分布空间， $\sigma(\cdot)$ 为非线性激活函数。堆叠 L 层 HGN 后，可以得到源帖 s_i 的传播结构信息表示 $H_t^{(L)}[s_i]$ 。

3.4.3 时序注意力

随着时间的推移，源帖在社交媒体中的演化模式主要表现为新的转发或评论出现，即异构传播图中新的用户节点和边的不断出现。HGN 虽然可以有效地学习静态快照的结构信息，但是并不能模拟随时间演变的模式。最近，用于动态网络嵌入的循环神经网络(RNN)取得了很好的成果。然而，常用的 RNN，例如 Bi-LSTM 和 GRU，在保存与当前节点距离较远的节点的信息和复制过去很久的信息时仍然存在问题，这可能导致模型的长期记忆能力不足。Dangovski 等人^[26]提出一种被称为旋转记忆单元(RUM)的新型 RNN 单元以解决该问题。与通过门机制来操纵隐藏状态的 LSTM 和

GRU 不同, RUM 通过在欧氏空间中执行旋转操作从而实现隐藏状态的操纵。因此, RUM 可以有效地提升模型对长期信息的记忆能力并准确地总结上下文信息, 从而释放模型在长期记忆方面的全部潜力。

DHGNN 扩展了现有的 RUM 模型, 并提出了一种基于 RUM 的时序注意力, 学习谣言传播过程中更深层的演化模式。基于 RUM 的时序注意力主要包含两部分: 旋转记忆单元(RUM)和时序注意力。其中, RUM 计算过程定义如下:

$$\tau_t = W_z^\tau z_t + W_h^\tau h_{t-1} + b_t^\tau \quad (21)$$

$$u_t' = W_z^u z_t + W_h^u h_{t-1} + b_t^u \quad (22)$$

$$u_t = \text{sigmoid}(u_t') \quad (23)$$

$$\tilde{\varepsilon}_t = \tilde{W}_z z_t + b_t \quad (24)$$

$$w_t = \frac{\tilde{\varepsilon}_t}{\|\tilde{\varepsilon}_t\|} \quad (25)$$

$$v_t = \frac{\tau_t - (w_t \cdot \tau_t)w_t}{\|\tau_t - (w_t \cdot \tau_t)w_t\|} \quad (26)$$

$$\theta = \arccos\left(\frac{\tilde{\varepsilon}_t \cdot \tau_t}{\|\tilde{\varepsilon}_t\| \cdot \|\tau_t\|}\right) \quad (27)$$

$$\tilde{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (28)$$

$$\text{Rotation}(\tilde{\varepsilon}_t, \tau_t) = [1 - w_t w_t^\top - v v^\top] + (w_t, v_t) \tilde{R}(\theta) (w_t, v_t)^\top \quad (29)$$

$$R_t = (R_{t-1})^\lambda \text{Rotation}(\tilde{\varepsilon}_t, \tau_t) \quad (30)$$

$$\tilde{h}_t = \sigma(\tilde{\varepsilon}_t + R_t h_{t-1}) \quad (31)$$

$$h_t' = u_t \odot h_{t-1} + (1 - u_t) \odot \tilde{h}_t \quad (32)$$

$$h_t = \frac{\eta h_t'}{\|h_t'\|} \quad (33)$$

其中, $\sigma(\cdot)$ 是非线性激活函数, \odot 是逐元素矩阵乘法运算, $W_z^\tau, W_z^u, \tilde{W}_z \in \mathbb{R}^{l \times d}, W_h^\tau, W_h^u \in \mathbb{R}^{l \times l}, b_t^\tau, b_t^u, b_t \in \mathbb{R}^l$ 是可训练的参数, λ 和 η 是可设置的系数, d 是输入向量的维度, l 是隐藏层神经元数, $\|\cdot\|$ 是 L_2 正则化操作。

通过 HGN, 可以得到每个快照的节点向量集 $H_t^{(L)} = \{H_t^{(L)}[s_1], H_t^{(L)}[s_2], \dots, H_t^{(L)}[s_N], H_t^{(L)}[s_i] \in \mathbb{R}^d\}$ 。其中, t 表示第 t 个快照, i 表示第 i 个源帖节点, d 是每个节点嵌入的维度。在每个时间步长 t , 将给定快照的节点向量集 $H_t^{(L)}$ 作为 z_t 输入到 RUM 中, 得到输出表示为 $\{h_t^1, h_t^2, \dots, h_t^N\} \in \mathbb{R}^{N \times d}$ 。

为了进一步捕获谣言动态传播的演化模式, DHGNN 在 RUM 输出上使用时序注意力来捕获重要的特征向量。通过加入特定的掩码矩阵, 确保模型按照时间顺序学习各异构传播图快照中的动态演化信息。从 RUM 输出中可以得到源帖

子节点 s_i 在不同快照的特征向量, 可表示为 $H_i = \{h_1^i, h_2^i, \dots, h_T^i\} \in \mathbb{R}^{T \times d}$, 将 H_i 输入时序注意力模型中, 得到的输出为 $Z^i = \{z_1^i, z_2^i, \dots, z_T^i\} \in \mathbb{R}^{T \times d}$, 计算公式如下:

$$Z^i = \text{softmax}\left(\frac{(H_i W_q)(H_i W_k)^\top}{\sqrt{d}} + M\right) \cdot (H_i W_v) \quad (34)$$

其中 $W_q, W_k, W_v \in \mathbb{R}^{d \times d}$ 是可训练参数, $M \in \mathbb{R}^{T \times T}$ 表示掩码矩阵。如果 $M_{u,v} = -\infty$, 则意味着时间 u 到 v 不记录注意力系数。当时间 $u \leq v$ 时, 设置 $M_{u,v} = 0$; 否则 $M_{u,v} = -\infty$ 。最后, 将 Z^i 输入池化层中得到源帖 s_i 的动态信息表示 B 。

3.5 分类层

通过上述过程, 得到一个文本细粒度语义表示 D 和一个动态信息表示 B , 这两种表示都对谣言检测非常重要。因此, 将两种表示进行拼接得到最终特征表示 P :

$$P = \text{Concat}(D, B) \quad (35)$$

最后, 使用一个全连接层和 *softmax* 层, 得到帖子的预测标签 y :

$$y = \text{softmax}(FC(P)) \quad (36)$$

DHGNN 使用交叉熵作为训练过程中的损失函数, 并通过 Adam 优化算法进行学习率的更新, 交叉熵损失函数如下:

$$L = \sum_{i=1}^{|C|} \hat{y}_i \log(y_i) \quad (37)$$

其中, $|C|$ 是谣言检测极性类别的数量, \hat{y}_i 是由 *one-hot* 向量表示的谣言检测极性, y_i 是通过分类层计算得到的谣言检测极性概率分布。

4 实验

4.1 数据集

为了评估 DHGNN 模型的性能, 本文在 Twitter15 和 Twitter16 两个公共数据集中进行实验验证。Twitter15 和 Twitter16 数据集由 Ma 等人^[27]公开提供。每个数据集由一组源帖子和相应的响应帖子组成, 并提供自源帖子发布以来经过的时间。在数据集中, 每条源帖子都被标记为四种谣言类别之一: 非谣言、假谣言、真谣言、未经证实的谣言。两个数据集的统计数据如表 1 所示。

表 1 数据集统计信息

Table 1 Statistics of datasets

数据集	Twitter15	Twitter16
source tweets	1490	818
non-rumors	374	205
false rumors	370	205
unverified rumors	374	203
true rumors	372	205
users	276663	173487
posts	331612	204820

4.2 对比模型

为了验证 DHGNN 模型的性能, 本文选择了一些最新的模型进行对比:

DTC^[4]: 该模型手工提取多个全局特征, 并使用基于信息可信度的决策树进行分类。

SVM-TS^[5]: 该模型提取随时间动态变化的手工特征, 并使用线性支持向量机分类器进行谣言检测。

GRU-RNN^[8]: 该模型使用 RNN 对序列结构特征进行建模。

RvNN^[13]: 该模型采用两种递归神经模型, 分别基于自底向上和自顶向下的传播树。

BiGCN^[15]: 该模型利用双向 GCN 来模拟自下而上的传播和自上而下的扩散。

RoBERTa-CNN^[19]: 该模型利用帖子文本的语义特征和帖子传播树的底层模式进行谣言检测。

EBGCN^[20]: 该模型通过采用贝叶斯方法自适应地考虑潜在边关系的可靠性。

MGAT-ESM^[25]: 该模型对源帖子与评论的关系、源帖子与词的关系和、源帖子与用户的关系建模, 并通过事件共享模块学习不同谣言中的事件不变特征。

4.3 评估标准和实验设置

为了公平比较, 本文采用了与先前工作^[20,25]相同的评估指标, 即准确率(Accuracy)和各类别非谣言(NR)、假谣言(FR)、真谣言(TR)、未经证实的谣言(UR)的 F1 分数。

本文随机选择 10% 的样本作为验证数据集, 并将其余的样本以 3:1 的比例分配给数据集的训练数据集和测试数据集。使用 Adam 算法优化模型, 所有权重矩阵使用 Xavier 进行初始化, 并将所有偏置值初始化为零。将初始学习率设置为 0.001, 并在训练过程中逐渐减小。词向量维度设置为 300, 当文本长度小于输入长度时用 0 填充。Dropout 设置为 0.5, 多头注意力头数 h 设置为 8, HGN 层数 L 设置为 4。本文提出的模型基于 PyTorch 开源工具实现, 采用 NVIDIA Tesla P100 的 GPU 训练模型。

表 2 Twitter15 和 Twitter16 数据集上谣言检测的对比结果

Table 2 Performance results of rumor detection on Twitter15 and Twitter16 datasets

模型	Twitter15					Twitter16				
	Acc	NR	FR	TR	UR	Acc	NR	FR	TR	UR
DTC	45.4	73.3	35.5	31.7	41.5	46.5	64.3	39.3	41.9	40.3
SVM-TS	54.4	79.6	47.2	40.4	48.3	57.4	75.5	42.0	57.1	52.6
GRU-RNN	64.1	68.4	63.4	68.8	57.1	63.6	61.7	71.5	57.7	52.7
RVNN	72.3	68.2	75.8	82.1	65.4	73.7	66.2	74.3	83.5	70.8
BiGCN	86.6	89.1	86.0	93.0	86.4	88.0	84.7	86.9	93.7	86.5
RoBERTa-CNN	86.2	85.0	82.4	86.7	81.2	89.6	81.5	83.7	92.9	85.4
EBGCN	89.2	86.9	89.7	93.4	86.7	91.5	87.9	90.6	94.7	91.0
MGAT-ESM	90.8	91.6	92.2	92.2	87.0	91.6	91.0	89.6	94.8	90.8
DHGNN w/o MLAN	86.0	92.2	80.6	90.7	81.0	85.3	82.7	86.7	92.1	80.4
DHGNN w/o DH	82.1	77.5	83.8	91.1	76.9	84.8	74.2	88.6	93.5	83.5
DHGNN w/o HGN	88.4	91.2	91.4	90.1	80.5	89.7	87.0	89.9	94.7	87.0
DHGNN w/o TA	87.5	91.5	85.9	89.9	82.4	88.6	88.1	88.4	90.2	87.5
DHGNN	91.7	92.3	93.6	93.2	87.6	92.9	93.2	91.1	92.8	94.6

4.4 实验结果分析

DHGNN 模型与其他对比模型的实验结果如表 2 所示。从中可以发现基于深度学习的方法明显优于传统方法 DTC、SVM-TS。主要原因是传统基于机器学习的模型使用的是手工制作的特征, 往往会导致特征选择的信息丢失, 从而很难达到性能最优。另一方面, 基于深度学习的模型可以从给定的高维数据中自动学习重要的特征表示。此外, 与基于序列的模型 GRU-RNN 相比, DHGNN 模型表现出了更好的性能, 这可以归因于 DHGNN 模型能够从词级和句子级两方

面充分学习文本的语义信息, 并通过对谣言动态传播过程建模学习其中的结构信息, 而不是仅从转发序列中捕获文本中的时序信息。DHGNN 模型在两个数据集中均取得了良好的性能表现。DHGNN 模型在 Accuracy 指标上分别比 Twitter15 和 Twitter16 数据集上的最佳对比模型高了 0.9% 和 1.3%, 在 F1 值指标上最大提升了 1.4% 和 2.2%, 但是在 TR 的 F1 值中分别比 EBGCN 和 MGAT-ESM 低了 0.2%、2.0%。因此, 可以认为 DHGNN 模型具有最优的综合性能。

4.5 消融分析

为了确定 DHGNN 中每个模块对模型性能的贡献, 本文进行了一系列的消融实验。其中, 消融模型包括 DHGNN w/o MLAN、DHGNN w/o DH、DHGNN w/o HGN 和 DHGNN w/o TA。DHGNN w/o MLAN 表示从 DHGNN 模型中去除多级注意力层, 仅利用动态传播信息; DHGNN w/o DH 表示从 DHGNN 模型中去除动态异构图网络层, 仅利用文本语义信息; DHGNN w/o TA 表示去除时序注意力, 不按照转发时间构建不同的异构传播图快照, 即不使用传播过程中的动态信息; DHGNN w/o HGN 表示不采用本文提出的图神经网络, 而采用传统图神经网络 GAT。

消融实验的结果如表 2 所示, 可以看出 DHGNN w/o DH 的性能与其他消融模型相比存在较大差距, 说明提出的动态异构图网络层可以充分学习异构传播图快照的结构性信息和帖子传播过程中的动态信息。DHGNN w/o HGN 和 DHGNN w/o TA 的性能分别位于第二和第三, 证明了图神经网络 HGN 和时序注意力是有效的。DHGNN w/o MLAN 的性能差于 DHGNN w/o TA 和 DHGNN w/o HGN, 这表明提出的多级注意力层能够引导模型学习源帖与评论之间潜在的语义关联, 从而提高模型性能。

4.6 参数分析

动态异构图网络层中的 HGN 层数 L 和多级注意力层中的注意力头数 h 决定了 DHGNN 模型学习传播结构信息和语义信息的程度, 是影响模型性能的关键因素。本节在 Twitter15、Twitter16 数据集中进行一系列参数实验, 分析上述参数对模型性能的影响。实验结果如表 3 与表 4 所示。从表 3 中可以看出, 当 $L=4$ 时, 模型取得了最好的性能; 当 $1 \leq L \leq 4$ 时, 模型性能随着层数的增加而有所提升; 但当 $L > 4$ 时, 模型性能开始出现下降趋势。这是由于图神经网络层数越多, 各节点状态更新时就可以融合更多跳内的邻居节点信息, 从而增强帖子节点的传播信息表示。但当图神经网络层数过多时, 模型过度学习各邻居节点的细节信息, 可能引入更多的噪声, 影响模型的性能。从表 4 中可以看出, 当 $h=8$ 时, 模型在两个数据集中均取得了最好的综合性能。当 $h=1$ 时, 模型的性能最差。这表明单个注意力函数并不能准确地将注意力权重分配给帖子 (或评论) 文本的上下文, 而多头注意力可以关注到更多不同子空间中的语义表示信息, 增强模型对文本语义的表达能力。但是, 模型性能不会随着头数 h 单调增加。当 $h > 8$ 时, 模型的性能出现下降趋势。这可以归因于过多的注意力函数会引入大量参数, 从而导致模型过拟合。

表 3 HGN 层数对 DHGNN 模型性能的影响

Table 3 Effects of the number of HGN layers on the performance of DHGNN

HGN 层数 L	Twitter15					Twitter16				
	Acc	NR	FR	TR	UR	Acc	NR	FR	TR	UR
1	87.8	89.4	88.2	92.7	80.5	87.0	84.8	85.1	92.8	84.8
2	89.0	90.6	90.0	90.7	84.7	90.2	89.0	91.0	92.6	88.4
3	88.1	91.8	86.2	92.0	82.6	91.8	90.3	91.1	92.3	93.7
4	91.7	92.3	93.6	93.2	87.6	92.9	93.2	91.1	92.8	94.6
5	90.4	92.2	91.3	92.2	86.0	90.2	90.5	90.9	89.6	89.9
6	89.3	89.9	89.3	92.9	85.0	90.2	93.1	87.1	91.8	88.4
7	88.1	88.9	87.4	92.6	83.3	87.5	85.1	88.1	90.0	86.7
8	89.6	90.8	90.2	92.7	84.9	88.0	87.8	84.8	92.5	87.3

表 4 多头注意力头数对 DHGNN 模型性能的影响

Table 4 Effects of the number of multi-head attentions on the performance of DHGNN

多头注意力头数 h	Twitter15					Twitter16				
	Acc	NR	FR	TR	UR	Acc	NR	FR	TR	UR
1	85.1	89.7	86.1	84.9	80.0	85.9	87.1	84.1	92.1	80.0
2	89.9	92.0	93.4	92.2	81.7	89.7	90.9	87.2	90.9	89.1
4	88.1	91.2	90.3	90.1	80.5	89.1	86.0	87.4	92.3	90.9
6	90.5	91.9	92.0	93.1	84.7	91.1	89.4	90.1	91.8	92.9
8	91.7	92.3	93.6	93.2	87.6	92.9	93.2	91.1	92.8	94.6
10	89.9	92.2	91.1	92.1	84.2	90.8	90.9	87.6	92.6	91.6
12	88.4	92.1	89.3	88.6	83.5	89.7	91.1	87.4	92.0	87.9

5 总结

针对现有未充分学习文本的语义信息和谣言传播过程中的结构信息,忽略谣言动态传播过程等问题。本文提出一种基于动态异构图的谣言检测模型 DHGNN。首先,通过多级注意力引导模型关注到源帖和相应评论中关键的词和句子,学习源帖与评论之间潜在的语义关联,增强帖子的文本语义表示。然后,利用基于异构图的图神经网络 HGN 对异构传播图中的用户、帖子节点和转发(或评论)关系进行建模,通过对不同类型的节点和边设置不同的共享参数,来表征每个边上的异构性注意力,从而为不同类型的节点和边生成特定的表示,充分学习异构传播图中的结构信息。同时,利用基于旋转记忆单元的时序注意力为每个异构传播图快照建立记忆,模拟谣言的动态传播过程,捕获谣言动态传播的演化模式。此外,本文实验验证了 DHGNN 模型在谣言检测的性能,通过消融实验分析了多级注意力、图神经网络 HGN 及时序注意力各模块的有效性,通过参数实验分析了动态异构图网络层中的 HGN 层数和多级注意力层中的注意力头数对模型性能的影响。

References:

- [1] DiFonzo N, Bordia P. Rumor Psychology: Social and Organizational Approaches[M]. American Psychological Association, 2007.
- [2] 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.
- [3] 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.
- [4] Castillo C, Mendoza M, Poblete B. Information Credibility on Twitter[C]//Proceedings of the 20th international conference on World wide web. 2011: 675-684.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] MA J, GAO W, MITRA P, et al. Detecting Rumors from Microblogs with Recurrent Neural Networks. (2016)[C] //Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016). 3818-3824.
- [9] Yu F, Liu Q, Wu S, et al. A Convolutional Approach for Misinformation Identification[C]//IJCAI. 2017: 3901-3907.
- [10] D Y, SONG Y R, SONG B. New Microblog Rumor Detection Model based on Attention Mechanism[J]. Journal of Chinese Computer Systems, 2021, 42(2): 348-353.
- [11] 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).
- [12] Wang Z, Guo Y. Rumor Events Detection Enhanced by Encoding Sentimental Information into Time Series Division and Word Representations[J]. Neurocomputing, 2020, 397: 224-243.
- [13] Ma J, Gao W, Wong K F. Rumor Detection on Twitter with Tree-Structured Recursive Neural Networks[C]. Association for Computational Linguistics, 2018.
- [14] Khoo L M S, Chieu H L, Qian Z, et al. Interpretable Rumor Detection in Microblogs by Attending to User Interactions[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(05): 8783-8790.
- [15] 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.
- [16] Ma J, Gao W. Debunking Rumors on Twitter with Tree Transformer[C]//Proceedings of the 28th International Conference on Computational Linguistics. 2020: 5455-5466.
- [17] Wu Z, Pi D, Chen J, et al. Rumor Detection based on Propagation Graph Neural Network with Attention Mechanism[J]. Expert Systems with Applications, 2020, 158: 113595.
- [18] Tu K, Chen C, Hou C, et al. Rumor2vec: A rumor Detection Framework with Joint Text and Propagation Structure Representation Learning[J]. Information Sciences, 2021, 560: 137-151.
- [19] Luo Z, Li Q, Zheng J. Deep Feature Fusion for Rumor Detection on Twitter[J]. IEEE Access, 2021, 9: 126065-126074.
- [20] 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 (Volume 1: Long Papers). 2021: 3845-3854.

[21] HU D, WEI L W, ZHOU W, HUAI X Y, HAN J Z, HU S L. A Rumor Detection Approach Based on Multi-Relation Propagation Tree[J]. Journal of Computer Research and Development, 2021, 58(07): 1395-1411.

[22] XUE H, WANG L, YANG Y, et al. Rumor Detection Model Based on User Propagation Network and Message Content[J]. Journal of Computer Applications, 2021, 41(12): 3540-3545.

[23] 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.

[24] Huang Q, Yu J, Wu J, et al. Heterogeneous Graph Attention Networks for Early Detection of Rumors on Twitter[C]//2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-8.

[25] 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.

[26] Dangovski R, Jing L, Nakov P, et al. Rotational Unit of Memory: A Novel Representation Unit for RNNs with Scalable Applications[J]. Transactions of the Association for Computational Linguistics, 2019, 7: 121-138.

[27] Ma J, Gao W, Wong K F. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning [C]. Association for Computational Linguistics, 2017.

附中文参考文献:

[10]潘德宇,宋玉蓉,宋波.一种新的考虑注意力机制的微博谣言检测模型[J].小型微型计算机系统,2021,42(02):348-353.

[21]胡斗,卫玲蔚,周薇,淮晓永,韩冀中,虎嵩林.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(07):1395-1411.

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