小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems
DOI:10.20009/j.cnki.21-1106/TP.2022-0382
2024 年 1 月 第 1 期
Vol.45 No.1 2024

时序感知的异质图神经谣言检测

陈林威[1],宋玉蓉[1],宋 波[2]

[1](南京邮电大学 自动化学院、人工智能学院,南京 210023)
[2](南京邮电大学 现代邮政学院,南京 210003)
E-mail:songyr@njupt.edu.cn

摘 要: 近年来,在线社交媒体的发展大大加速了谣言的滋生和传播,谣言的危害性使得谣言的自动检测技术受到研究学者的广泛关注.本文同时考虑事件与事件之间的全局结构关系以及事件内部消息传播的时序关系,以异质图为载体共同显式建模两种关系,提出一种新的时序感知的异质图神经谣言检测模型.该模型利用时序感知的自注意力机制捕获事件内部转发(或评论)贴之间的时序关系,并将具有时序信息的转发(或评论)贴与源贴融合,得到事件的局部时序表征;接着利用元素级注意力机制捕捉事件与事件之间的全局结构关系,学习事件的全局结构表征;最后将二者融合用于检测谣言.实验结果表明,该模型优于大多数现有模型,可以提高谣言检测性能,并且同样具有优秀的早期检测性能.

关 键 词: 时序感知;异质图;谣言检测
中图分类号: TP391 **文献标识码:** A **文 章 编 号:** 1000-1220(2024)01-0045-07

Sequence-aware Heterogeneous Graph Neural Rumor Detection

CHEN Linwei[1],SONG Yurong[1],SONG Bo[2]

[1](College of Automation & College of Artificial Intelligence,Nanjing University of Posts and Telecommunications,Nanjing 210023,China)
[2](School of Modern Posts & Institute of Modern Posts,Nanjing University of Posts and Telecommunications,Nanjing 210003,China)

Abstract: In recent years,the development of online social media has greatly accelerated the breeding and spreading of rumors,and the harmful effects of rumors have led to the automatic detection of rumors receiving extensive attention from research scholars. In this paper,we propose a new sequence-aware heterogeneous graph neural(SHGN) rumor detection model by considering both the global structural relation between events and events and the temporal relation of message propagation within events,and modeling the two relations together explicitly with a heterogeneous graph as a carrier. The model captures the temporal relations between retweets(or comment) posts within an event using a temporal-aware self-attention mechanism,and fuses the retweets(or comment) posts with temporal information with the source posts to obtain the local temporal representation of the event;then captures the global structural relationships between the event and the event using an element-level attention mechanism to learn the global structural representation of the event;finally,the two are fused for detecting rumors. Experimental results show that the model outperforms most existing models to improve rumor detection performance and also has excellent early detection performance.

Keywords: sequence-aware;heterogeneous graph;rumor detection

0 引 言

随着互联网技术的快速发展和移动设备的广泛使用,微博、Facebook 等一批在线社交媒体,逐渐成为大众信息消费的主要载体.社交媒体平台的开放性和便捷性为人们提供了自由表达的空间,但这也间接为虚假信息的传播提供了理想场所.在这里,谣言被定义为一种未经证实或故意发布的虚假消息[1].谣言的传播会引起公众的恐慌,对社会稳定造成巨大威胁.尤其在疫情期间,大量新的谣言不断涌现,例如“粮食短缺,赶紧屯粮”、“新冠抗体可使人免受‘二次感染’”等谣言信息广为散布,这无疑会误导群众,一定程度上影响社会秩序.因此,如何准确、快速地识别谣言是亟待解决的问题.

早期的谣言检测方法大多利用特征工程从文本内容[2]、用户配置信息[3]和传播结构[4-6]等方面挖掘有效特征.这类方法依赖繁琐的特征工程,耗时耗力,并且人工设计的特征无法表达更深层的特征,大大限制了模型的检测性能.随着深度学习算法的发展,神经网络在自然语言处理领域中取得了十分显著的效果.受此启发,研究学者们开始利用深度学习模型对文本内容、用户、传播结构等进行建模,自动学习这些信息的高阶特征表示[7-11].最近,基于图模型的方法[12-15]将事件的传播结构抽象为图结构,并利用图神经网络学习事件的特征表示,有效提高了谣言检测的准确率.但这些方法只考虑到事

收稿日期:2022-06-20 收修改稿日期:2022-08-24 基金项目:国家自然科学基金项目(61672298,61873326)资助;江苏高校哲学社会科学研究重点项目(2018SJZDI142)资助;江苏省高等学校自然科学研究面上项目(20KJB120007)资助;江苏省自然科学基金青年基金项目(BK20200758)资助. 作者简介:**陈林威**,男,1998 年生,硕士,研究方向为复杂网络、谣言检测;**宋玉蓉**,女,1971 年生,博士,教授,博士生导师,CCF 会员,研究方向为复杂网络中信息传播及其控制策略、自适应网络建模、仿真与智能优化;**宋 波**,男,1987 年生,博士,讲师,研究方向为复杂网络传播动力学、供应链网络级联失效、节点影响力辨识.

件内部帖子的局部传播结构,忽略了事件在社交媒体上的全局结构关系. Yuan 等人[16]认为每个事件都不是独立的个体,事件与事件之间可能因为相同用户的参与而产生联系,只考虑每个事件本身的特征而忽略事件之间的联系,势必会限制模型的检测性能. 因此,他们从异构网络的视角研究事件与事件之间的关联,提出了一个联合全局与局部关系的异质图来捕获消息传播的局部语义关系和全局结构信息. 尽管该模型取得了良好的效果,但他们忽略了事件内部消息传播过程中的时序信息. 事实上,在现实的消息传播过程中,每个时间阶段发布的转发(或评论)贴可能有不同的含义. 例如,在源贴刚发布不久,由于其没有被证实,所以表示怀疑的转发(或评论)贴可能更多;在消息传播的后期阶段,由于官方的辟谣以及事件真相的逐渐显露,人们的态度随之发生改变. 对于谣言事件,人们更多的转变为反对的态度,而对于非谣言事件,则越来越多的人开始肯定该言论. 并且,用户发布帖子除了受源贴的影响,还可能受到其他已发帖子的影响,进而干扰其判断. 因此,本文认为有必要挖掘消息传播过程中的时序信息,以捕获谣言与非谣言在不同时间阶段产生的响应贴的差异性以及响应贴之间潜在的时序依赖关系.

基于这些观察,本文考虑同时建模事件与事件之间的全局结构关系以及事件内部消息传播的时序关系,提出一种时序感知的异质图神经谣言检测模型. 本文首先基于谣言检测数据集构建一张事件-用户异质图模拟消息在真实社交媒体上的传播;然后,通过位置编码建模事件内部转发(或评论)帖之间的局部时序关系,结合多头注意力机制关注重要的响应贴,并将具有时序信息的响应贴与源贴融合,得到事件节点的局部时序表征;接着利用元素级注意力机制捕捉事件与事件之间的全局结构关系,学习事件节点的全局结构表征;最后,将两种特征融合,得到最终的事件节点表示用于谣言检测. 本文的主要贡献如下:

1)本文同时考虑事件与事件之间的全局结构关系以及事件内部消息传播的时序关系,以异质图为载体共同显式建模两种关系;

2)本文基于转发(或评论)贴与源贴的交互时间构建响应序列,并通过时序感知的自注意力机制挖掘事件内部的时序信息,然后融合源贴和响应贴的表示,得到每个事件节点的局部时序表征;

3)本文基于用户与事件的交互关系,利用元素级注意力机制学习每个事件节点的全局结构表征,以捕获复杂多样的传播结构特征;

4)本文在 3 个公开的真实数据集上验证方法的有效性. 实验结果表明,本文模型优于现有最先进的谣言检测模型,并且具有更高的早期检测性能.

1 相关工作

谣言检测的研究最早是基于帖子的内容信息展开,致力于挖掘文本特征,捕获谣言与非谣言在内容上的差异. 但考虑到谣言的发布者常常刻意效仿真实消息的语言风格,导致单纯基于内容的方法无法捕捉有效的特征进行谣言识别. 随后的一些研究工作开始从消息的传播结构特征着手,捕获谣言事件与非谣言事件在传播模式上的差异,取得了良好的效果. 本文将现有的谣言检测方法大致分为:1)基于内容的谣言检测方法;2)基于传播结构的谣言检测方法.

1.1 基于内容的谣言检测方法

早期基于内容的谣言检测的研究工作主要集中在人工设计特征上,再结合如逻辑回归、决策树等分类算法对消息的真实性进行判别. Wu 等人[17]利用狄利克雷分布(Latent Dirichlet Allocation,LDA)生成文本的主题,并基于随机游走图核的 SVM 算法在微博数据上进行谣言检测;Sun 等人[18]提取了描述事件的动词数、包含事件动词的消息比例、是否包含强消极词以及包含强消极词的消息比例 4 种新的文本特征,并利用决策树分类算法用于谣言检测. 这类方法均基于繁重的手工特征提取,只能挖掘谣言的浅层特征,无法进一步提升谣言检测的精度. 伴随着深度学习的快速发展,各类模型结构层出不穷,神经网络不断加深使得网络能够对特征进行充分挖掘,降低了对人工设计特征的要求,谣言检测方向也逐渐从特征挖掘迁移到模型结构设计上. Yu 等人[19]、Liu 等人[20]将卷积神经网络(Convolutional Neural Network,CNN)应用于谣言检测问题,通过 CNN 自动生成高阶抽象特征,对事件的表示进行学习,提升了谣言检测精度;Zhang 等人[21]利用注意力模型分别学习文本内容和情感符号的特征表示,并将二者融合的结果作为微博文本的语义表示,一定程度上丰富了微博文本的情感语义信息;Pan 等人[22]结合文本卷积神经网络(Text CNN)与引入注意力机制,通过注意力机制对 Text CNN 学习到的文本表示进行加权输出,提取更为显著的微博文本特征. 虽然这类方法取得了一定效果,但目前的谣言发布者常常会通过刻意模仿真实消息的写作手法、表达风格等途径达到逃避检测的目的. 这就导致仅依赖于文本内容的方法无法捕捉有效特征进行谣言的识别.

1.2 基于传播结构的谣言检测方法

基于传播结构的方法重点关注的是谣言事件与非谣言事件传播特征之间的差异. Ma 等人[7]考虑消息在传播过程的时序特征,首次提出利用循环神经网络(Recurrent Neural Network,RNN)捕获每个源贴及其转发(或评论)贴的语义变化,并根据语义变化进行预测,这也是第一次引入深度学习模型进行谣言检测的研究;而后,Ma 等人[8]又探究了一种基于树的递归神经网络(RvNN),分别建模自顶向下和自底向上的消息传播树,用以捕捉源贴的语义信息和传播结构信息;Liao 等[9]提取文本的潜在特征和局部的用户特征,并利用带有注意力机制的双向 GRU 网络学习微博事件的表示;Chen 等人[10]将 RNN 与注意力机制结合,捕获消息传播过程中的语义变化,一定程度上提高了谣言检测性能. 这些方法大多只关注到事件内部的时间变化,将消息的传播结构构建为时间序列,忽略了帖子之间显式的转发(评论)关系. 基于此,Bian 等人[13]根据帖子之间的转发(或评论)关系将传播结构构建成图结构,通过双向图卷积网络(Bi-Directional Graph Convolution Network,Bi-GCN)学习消息转发的结构特征;Hu 等[14]基于消息传播树中蕴含的层间依赖关系与层内依赖关系,利用多关系图卷积网络共同建模两种关系,以捕获更为丰富的传播结构特征. 这些方法虽然取得了不错的效果,但终究只关注到消息的局部传播,即认为每个事件都是独立的,没有考虑到

事件与事件之间的关联,忽略了全局结构信息. 而 Yuan 等人[16]则考虑到事件的全局结构特征,探究了一个联合全局与局部关系的异构网络来捕获消息传播的局部语义关系和全局结构信息,取得了良好的效果. 但其忽略了事件内部重要的时序关系.

本文以异质图为载体,同时融合局部时序关系和全局结构关系,提出了一种时序感知的异质图神经谣言检测方法,共同显式建模局部时序信息和全局结构信息,为事件节点学习更全面的特征表示,用于提升谣言的识别能力.

2　模型方法

2.1　问题描述

设 $C=\{c_1,c_2,\cdots,c_{|C|}\}$ 为谣言事件集,其中,c_i 为第 i 个事件样本,$|C|$ 为事件集的大小,即事件总数. 对于每个 c_i 事件,都包括 1 条源贴 m_i 和 n 条相关的转发(评论)贴,记为 $c_i=\{m_i,r_{i,1},r_{i,2},\cdots,r_{i,n}\}$. 此外,本文定义 $U=\{u_1,u_2,\cdots,u_{|U|}\}$ 表示社交媒体中参与事件的用户集合,其中 $|U|$ 为数据集中用户的总数量.

谣言检测任务可以描述为机器学习中的分类任务,将每个事件 c_i 标注为对应的真实标签 $y_i\in y$,y 为类别标签集. 本文的目标是学习一个函数 $f:f(c_i)\rightarrow y_i$ 来预测当前事件是否为谣言.

2.2　模型介绍

本文提出的谣言检测模型主要由异质图构建、局部时序信息编码、全局结构信息编码和谣言分类 4 个模块组成. 总体框架如图 1 所示. 首先,在异质图构建模块(模块 1)中,本文基于谣言检测数据集构建事件-用户异质图,并利用嵌入技术对图中各节点进行初始化表示;其次,在局部时序信息编码模块(模块 2)中,本文基于事件内部转发(或评论)贴之间的时序关系,利用时序感知的自注意力机制学习具有时序信息的响应贴表示,再融合源贴本身的内容信息,得到每个事件节点的局部时序表征;在全局结构信息编码模块(模块 3)中,本文基于用户与事件之间的交互关系,利用元素级注意力机制学习每个事件节点的全局结构表征;最后,在谣言分类模块(模块 4)中,本文融合事件的局部时序表征与全局结构表征,以预测当前事件为谣言的概率. 接下来,本文将详细介绍每个模块. 为简化说明,本文在后续的描述中,将“转发(或评论)贴”均替代为“响应贴”.

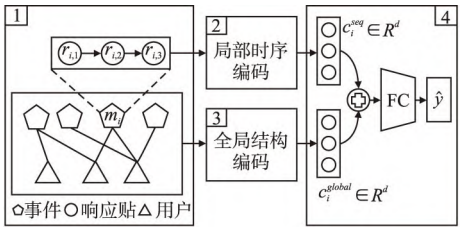

图 1　SHGN 模型总体框架

Fig. 1　Overall architecture of the SHGN model

1)异质图构建

本文将事件以及相关用户抽象为网络中两种不同类型的节点,并根据用户对事件的参与情况(用户对事件中的帖子存在转发或评论的行为),建立用户节点与事件节点的连边关系. 此外,在每个事件内部又包含一条源贴和一系列响应贴. 根据响应贴在源贴发布后的时间延迟,将响应贴构建为时间序列,这样每条源贴就对应一个响应序列. 如图 1 中模块 1 所示,最终构建为具有时序信息的事件-用户异质图.

接下来,对异质图中的各节点进行初始化. 首先对于事件节点,其内部本质是源贴和响应贴的文本内容. 而社交媒体中大多是短文本数据,由几个到几十个单词组成,长度短、内容少;并且在文本语义方面,谣言发布者常常刻意模仿真实消息的写作风格. 这些因素都会导致模型无法挖掘有效的语义信息进行谣言检测. 目前主流的从词向量序列中学习文本语义表示的模型包括卷积神经网络、循环神经网络等. 相比之下,RNN 等序列模型更适用于长文本的内容特征提取,对社交帖子这类短文本数据而言,数据稀疏问题不可避免地影响模型性能;而 CNN 具有关注局部语义信息的特点,在处理短文本的效果上比 RNN 好.

因此,本文采用词向量的方式初始化,并利用 CNN 对其进行编码. 具体而言,固定每个帖子的单词数量为 L 个,当单词数量小于 L 时,用 0 填充;当单词数量超过 L 时,则截断. 接着通过 Word2Vec 算法[23]在特定领域的语料库上进行训练,得到每个单词的向量表示,对于预训练词向量库中没有出现的单词,本文使用均匀分布进行初始化,并且保持词向量在训练过程中可微调. 记每个单词的初始向量为 $x_j\in\mathbb{R}^d$,j 表示帖子中的第 j 个单词,则每条单词数量为 L 的帖子可表示为:

$$x_{1:L}=[x_1;x_2;\cdots;x_L] \tag{1}$$

其中,“;”为拼接操作,$x_{1:L}\in\mathbb{R}^{L\times d}$.

进一步地,本文利用 CNN 对句子序列进行编码. 给定一个由单词向量组成的句子序列 $x_{1:L}$,通过 CNN 的卷积层对每个可能的窗口做一维卷积操作:

$$e_i=\sigma(W*x_{i:i+h-1}) \tag{2}$$

得到特征图 $e=[e_1,e_2,\cdots,e_{L-h+1}]\in\mathbb{R}^{L-h+1}$. 其中,$W\in\mathbb{R}^{h\times d}$ 是大小为 h 的卷积核,$\sigma=\dfrac{1}{1+\exp(\cdot)}$为 *sigmoid* 激活函数. 接着利用最大池化操作 $\hat{e}=\max(e)$ 选择每个特征图的最大值,再通过拼接操作得到每个帖子的初始向量表示. 这样,对于第 i 个事件而言,其源贴表示为 $m_i\in\mathbb{R}^d$,每条响应贴表示为 $r_{i,j}\in\mathbb{R}^d$,将该事件中响应贴组成的矩阵记为 $R_i=[r_{i,1},r_{i,2},\cdots,r_{i,n}]\in\mathbb{R}^{n\times d}$.

另一方面,对于用户节点的初始化,本文对用户的属性信息(包括性别、年龄、粉丝数、关注数等)进行编码,得到用户节点的初始化向量表示. 对于获取不到的用户信息,本文通过正态分布进行初始化.

2)局部时序信息编码

受 Transformer 模型[24]的启发,本文采用时序感知的自注意力机制挖掘事件内部的局部时序信息,捕获谣言事件与非谣言事件在不同时间阶段产生的响应贴的差异以及响应贴之间潜在的时序依赖关系.

首先,为了编码每条响应贴的时延信息,本文使用 Transformer 模型中的位置编码(Positional Encoding,PE)公式为每条响应贴生成一个位置嵌入 $P\in\mathbb{R}^{|P|\times d}$:

$$P_{(pos,2k)} = \sin\left(\frac{pos}{10000^{\frac{2k}{d}}}\right) \tag{3}$$

$$P_{(pos,2k+1)} = \cos\left(\frac{pos}{10000^{\frac{2k}{d}}}\right) \tag{4}$$

其中,pos 表示响应贴在序列中的位置,d 表示位置嵌入的维度,$2k$ 表示偶数的维度,$2k+1$ 表示奇数维度(即 $2k \leqslant d, 2k+1 \leqslant d$).

接着将每条响应贴的嵌入与其对应的位置嵌入相关联,以捕获响应贴之间的时序信息:

$$R_i^{seq} = \{r_{i,t} + p_{i,1}, r_{i,t-1} + p_{i,2}, \cdots, r_{i,1} + p_{i,t}\} \tag{5}$$

其次,利用多头注意力机制对重要的响应贴进行重点关注. 自注意力机制能够自动学习自己与上下文之间的相关性,捕获重要的上下文信息,而多头则可以考虑多方面的影响因素,获得更为全面的节点表示:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{6}$$

$$\tilde{R}_i^{seq} = \|_{k=1}^{K} Attention(R_i^{seq}, R_i^{seq}, R_i^{seq}) \tag{7}$$

最后,将具有时序信息的响应贴融合到源贴的表示中,得到事件的局部时序表征 c_i^{seq}. 具体的融合策略有很多,如平均池化、最大池化、拼接等操作. 而本文将一系列响应贴视为对应源贴的一阶邻居节点,并采用图注意力网络[25]中的聚合函数进行融合,具体计算如下:

$$\alpha_{ii} = softmax(LeakyReLU(a^T[m_i; m_i])) \tag{8}$$

$$\alpha_{ij} = softmax(LeakyReLU(a^T[m_i; \tilde{r}_{i,j}^{seq}])) \tag{9}$$

$$c_i^{seq} = \sigma\left(\alpha_{ii} Wm_i + \sum_{j \in N(m_i)} \alpha_{ij} W \tilde{r}_{i,j}^{seq}\right) \tag{10}$$

其中,α_{ii}、α_{ij}分别表示节点 i 与自身以及节点 i 与节点 j 之间的注意力分数,$N(m_i)$为节点 i 的邻居节点,即当前源贴对应的所有响应贴,$W \in \mathbb{R}^{d^{(l+1)} \times d^l}$为该层节点特征变换的权重参数.

3)全局结构信息编码

基于共同用户而建立的事件与事件之间的全局结构关系,本文考虑如何学习事件节点的全局结构特征 c_i^{global}. 受广泛应用于推荐系统任务的元素级注意力机制[26]的启发,本文提出一种以用户嵌入元素为导向的注意力机制,其假定用户嵌入的每个维度都反映了用户的不同方面信息,并且用户的这些不同属性会对消息的传播产生不同的影响. 具体过程如下.

对于参与到事件 c_i 中的特定用户 $u_j \in \mathbb{R}^d$,计算用户 u_j 在不同方面的注意力向量 γ_j:

$$\gamma_j = \tanh(W_c u_j + b) \tag{11}$$

其中,$W_c \in \mathbb{R}^{d \times d}$为特征变换矩阵,$\gamma_j = \{\gamma_j^1, \gamma_j^2, \cdots, \gamma_j^d\}$是不同方面的注意力向量. γ_j^k 越大,表示用户嵌入 u_j 的第 k 个方面对消息传播的影响越大.

接着,将注意力向量 γ_j 与参与到事件 c_i 中的所有用户以元素乘积方式进行聚合,以捕获事件与事件之间的全局结构关系:

$$c_i^{global} = \sum_{j \in N(c_i)} \gamma_j \odot u_j \tag{12}$$

4)谣言分类

经过上面的过程,可以得到每个事件节点的局部时序表示和全局结构表示,二者对谣言检测都至关重要. 本文将两种特征进行拼接作为事件节点的最终表示,并经过全连接层和 $softmax$ 函数计算该事件的预测结果,即该事件为各个标签的概率值:

$$\hat{y}_i = softmax(Fc([c_i^{seq}; c_i^{global}])), \tag{13}$$

其中,$Fc(\cdot)$表示全连接层,输出的维度与标签的类别一致.

最后,为了训练模型的参数,本文采用交叉熵损失作为模型的目标优化函数:

$$L(\theta) = -\sum_{i=0}^{r-1} y_i \log(\hat{y}_i) \tag{14}$$

其中,θ 为模型的所有参数,r 为样本标签的类别数,$y_i \in \{0,1,2,3\}$(Twitter),$y_i \in \{0,1\}$(Weibo)为真实的标签值.

3 实验结果与分析

3.1 实验数据及设置

1)实验数据

本文在 Twitter15[6]、Twitter16[6]和 Weibo[7]3 个公开的真实数据集上对提出模型的有效性进行评估. 所有标签都是由辟谣网站中文章的真实标签获取. 其中,Twitter15 和 Twitter16 数据集有 4 种标签类别,分别是非谣言(non-rumor, NR)、假谣言(false rumor, FR)、真谣言(true rumor, TR)、未经证实的谣言(unverified rumor, UR);Weibo 数据集包含 2 种标签,分别是真谣言(true rumor, TR)和假谣言(false rumor, FR). 详细的统计情况如表 1 所示.

表 1 数据集统计

Table 1 Dataset statistics

统 计	Twitter15	Twitter16	Weibo
帖子数量	331612	204820	3805656
用户数量	276663	173487	2746818
事件数量	1490	818	4664
真谣言数量	374	205	2351
经过验证的非谣言数量	370	205	2313
未验证谣言数量	374	203	0
非谣言数量	372	205	0
每个事件平均事件跨度/h	1337	848	2460
每个事件平均帖子数	233	251	816
每个事件最大贴子数	1768	2765	59318
每个事件最小贴子数	55	81	10

2)评价指标及参数设置

为了公平验证本文模型的有效性,本文采用与之前研究一致的评估指标[16],分别是准确率、F1 分数、召回率和精准率. 模型中的参数由 Adam 算法更新,其参数 β_1、β_2 分别设置为 0.9 和 0.999,学习率初始化为 1e-3,在训练过程中逐渐降低. 模型中的词嵌入和用户初始化嵌入都是 300 维,dropout 比率设置为 0.2. CNN 卷积核大小设置为(3,4,5),每组 100 个. 多头注意力机制的 K 设置为 8. 实验的 batchsize 设置为 64,epoch 设置为 30. 所有代码均由 Pytorch 实现,所有的实验结果都是在 5 次实验的结果上取平均.

3.2 对比方法

本文选取了一些最先进的基线模型在相同的数据集上进

行实验和比较,所选取的基线模型如下:

1)DTC 模型[3]. 该模型基于特征工程,人工设计文本内容特征和用户特征,并利用决策树分类器进行谣言检测任务.

2)SVM-RBF 模型[2]. 该模型基于帖子的总体统计数据,手工构造特征,并利用基于 RBF 内核的支持向量机进行谣言检测.

3)CAMI 模型[19]. 该模型将段落向量作为 CNN 的输入,自动学习文本的深层特征进行谣言检测.

4)GRU 模型[7]. 该模型基于消息的传播序列,并利用 GRU 学习事件的传播结构特征,进而完成谣言检测任务.

5)RvNN 模型[8]. 该模型利用树结构的递归神经网络,对树状的消息传播模式建模,从而完成谣言检测任务.

6)Bi-GCN 模型[13]. 该模型基于消息的传播方向和扩散方向建立传播树,并利用双向图卷积神经网络学习节点表示,进行谣言检测.

7)GLAN 模型[16]. 该模型基于异质图同时捕获事件的局部语义信息和全局结构信息,进而完成谣言检测任务.

3.3　结果与分析

本文通过实验得到本文模型及所有对比方法在 3 个数据集上的性能,如表 2 ~ 表 4 所示. 其中,Acc 表示分类的总体准确率(accuracy,Acc). 实验结果证明,本文提出的 SHGN 模型要优于其他基线模型,且检测的准确率和精度都得到显著提高. 下面对实验结果进行具体分析:

1)相比基于传播结构的谣言检测方法(GRU、RvNN、Bi-GCN、GLAN 以及 SHGN),基于内容的方法(DTC、SVM-RBF、CAMI)性能整体较差. 这是因为社交帖子具有内容少、噪声多的特点,导致模型无法充分提取有效的语义特征进行谣言识别,而基于传播结构的模型主要侧重于挖掘谣言事件与非谣言事件传播特征之间的差异,通过捕捉复杂多样的传播特征提高检测性能. 可以观察到,相比于基于内容的模型中表现最佳的 CAMI 模型,SHGN 结果更好,在 Twitter15 数据集上其准确率要比 CAMI 模型高 21.1%,在 Twitter16 数据集上的准确率要比 CAMI 模型高 17.9%,在 Weibo 数据集上的准确率要比 CAMI 模型高 2.5%. 实验结果证明了本文模型的优势以及传播结构特征对于谣言检测任务的重要性和必要性.

2)在所有基于传播结构的基线模型中,GLAN 模型表现最佳. 这是因为 GLAN 考虑了消息在真实社交媒体中复杂多样的传播过程,以异质图为载体同时建模局部语义信息和全局结构信息进行谣言检测,而 GRU 模型将消息传播建模为传播序列以及 Bi-GCN 模型将消息传播建模为传播图,都只考虑局部的结构特征,忽略了事件与事件之间的关联,无法捕获全局结构信息. 这也表明,将谣言检测任务建模为异质图以捕获全局结构信息的方法要优于通过建模为传播序列或传播树捕获局部结构信息的方法. 而相比于 GLAN 模型,本文提出的 SHGN 模型结果更好,在 Twitter15 数据集上其准确率要比 GLAN 模型高 2.2%,在 Twitter16 数据集上的准确率要比 GLAN 模型高 2.1%,在 Weibo 数据集上的准确率要比 GLAN 模型高 1.2%. 这是因为 SHGN 模型通过引入位置编码弥补了 GLAN 模型忽略的局部时序信息. 由此证明局部的时序关系在谣言检测中至关重要,捕获局部结构中的时序信息有助于提高谣言检测的性能.

表 2　Twitter15 数据集上的实验结果

Table 2　Experimental results on Twitter15 dataset

Method	Twitter15				
	Acc	NR F_1	FR F_1	TR F_1	UR F_1
DTC	0.454	0.733	0.355	0.317	0.415
SVM-RBF	0.318	0.455	0.037	0.218	0.225
CAMI	0.701	0.664	0.739	0.806	0.635
GRU	0.646	0.792	0.574	0.608	0.592
RvNN	0.723	0.682	0.758	0.821	0.654
Bi-GCN	0.835	0.767	0.843	0.887	0.808
GLAN	0.890	0.936	0.908	0.897	0.817
SHGN	**0.912**	**0.942**	**0.929**	**0.904**	**0.854**

表 3　Twitter16 数据集上的实验结果

Table 3　Experimental results on Twitter16 dataset

Method	Twitter16				
	Acc	NR F_1	FR F_1	TR F_1	UR F_1
DTC	0.465	0.643	0.393	0.419	0.403
SVM-RBF	0.321	0.423	0.085	0.419	0.037
CAMI	0.744	0.676	0.728	0.813	0.683
GRU	0.633	0.772	0.489	0.686	0.593
RvNN	0.737	0.662	0.743	0.835	0.708
Bi-GCN	0.860	0.779	0.859	0.925	0.855
GLAN	0.902	0.921	0.869	0.847	**0.968**
SHGN	**0.923**	**0.937**	**0.911**	**0.935**	0.910

表 4　Weibo 数据集上的实验结果

Table 4　Experimental results on Weibo dataset

Method	Class	Acc	Prec	Recall	F_1
DTC	FR	0.831	0.847	0.815	0.831
	TR		0.815	0.847	0.830
SVM-RBF	FR	0.818	0.822	0.812	0.817
	TR		0.815	0.824	0.819
CAMI	FR	0.933	0.921	0.945	0.933
	TR		0.945	0.921	0.932
GRU	FR	0.910	0.876	0.956	0.914
	TR		0.952	0.864	0.906
RvNN	FR	0.901	0.918	0.896	0.921
	TR		0.914	0.916	0.911
Bi-GCN	FR	0.934	0.940	0.930	0.931
	TR		0.928	0.939	0.929
GLAN	FR	0.946	0.943	0.948	0.945
	TR		0.949	0.943	0.946
SHGN	FR	**0.958**	**0.957**	**0.960**	**0.961**
	TR		**0.961**	**0.958**	**0.961**

3.4　消融实验

为了验证局部时序信息和全局结构信息对提高谣言检测性能均有贡献,本文进行了一系列消融实验,主要包括 3 个部分:

1)w/o LSE(Local Sequence Embedding):移除局部时序信息编码模块,仅考虑全局结构信息.

2)w/o GSE(Global Structure Embedding):移除全局结构

信息编码模块,仅考虑局部时序信息.

3)w/o PE:移除位置编码,即没有引入局部结构中的时序边,仅使用多头注意力机制捕获局部语义信息,主要用于验证局部传播结构中时序边对谣言检测的有效性.

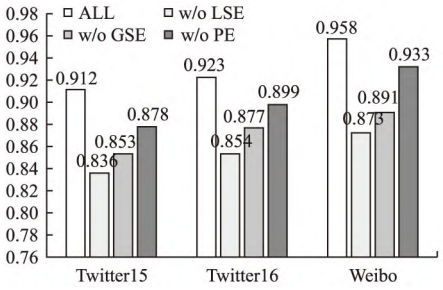

图 2 SHGN 在 3 个数据集上的消融实验结果

Fig 2 Ablation experiment result of the SHGN on three datasets

消融实验的结果如图 2 所示,其中 ALL 为原始的 SHGN 模型. 根据表中的实验结果,可以得到以下结论:

首先,本文研究了局部时序编码模块对谣言检测的影响. 可以观察到,移除 LSE 模块显著影响 SHGN 模型在 3 个数据集上的性能,准确率在 Twitter15 和 Twitter16 数据集上分别下降了 7.6% 和 6.9%,在 Weibo 数据集上下降了 8.5%. 实验结果证明了建模局部时序关系的有效性. 这也说明,随着时间推移,人们态度的转变对于不同事件确实存在差异,并且较早发表的言论对某些用户也确实存在一定影响. 因此,局部时序编码模块能够捕获到谣言事件与非谣言事件在不同时间阶段产生的响应贴的差异性以及响应贴之间潜在的时序依赖关系,对谣言检测至关重要.

其次,本文研究了全局结构编码模块对谣言检测的影响. 社交媒体事件不是独立存在的,事件与事件之间因相同的用户建立起全局关联. 直观地说,建模全局结构关系能够使谣言事件与非谣言事件具有高凝聚性和低耦合性,从而提高谣言检测性能. 可以观察到,GSE 模块的移除导致模型在 3 个数据集上的性能下降,准确率在 Twitter15 和 Twitter16 数据集上分别下降了 6.7% 和 4.6%,在 Weibo 数据集上下降了 7.7%. 实验结果同样可以证明建模全局结构关系的有效性,事件的全局结构关系对于谣言检测亦是不可或缺.

此外,本文还通过评估位置编码的有效性,证明了事件内部的消息不仅仅只有显式的关联关系,还具有潜在的时序依赖关系. 仅使用多头注意力机制虽然可以捕获任意响应贴之间的影响,但是其本身不能表达位置信息,即无法捕获响应贴在传播过程中蕴含的时序依赖关系,而加入位置编码可以显式地编码各响应贴在消息传播序列中的相对位置,从而捕获到消息传播过程中的上下文信息. 可以观察到,SHGN 在没有加入位置编码的模型中,准确率在 Twitter15 和 Twitter16 数据集上分别下降了 3.4% 和 2.4%,在 Weibo 数据集上下降了 2.5%. 因此,事件内部的消息确实具有潜在的时序关系,引入位置编码捕获事件局部的时序信息能够有效提高该场景下的检测能力.

3.5 早期检测研究

早期检测是指在消息传播的早期阶段进行谣言识别,以便及时采取措施,减少谣言传播,这是评价谣言检测方法质量的另一个重要指标. 为了验证本模型具有优秀的早期检测性能,本文在 Twitter15 和 Twitter16 这 2 个数据集上进行了早期谣言检测实验. 具体而言,本文分别选取了源贴发布后的 4h、8h、12h、24h 作为时间节点,代表不同的时间阶段,截取这些时间段内的响应帖以及相关用户作为早期数据. 同样,本文选取了 DTC、GRU、RvNN、Bi-GCN、GLAN 5 个基线模型进行对比,来评估本文模型的性能. 实验结果如图3和图4所示. 可

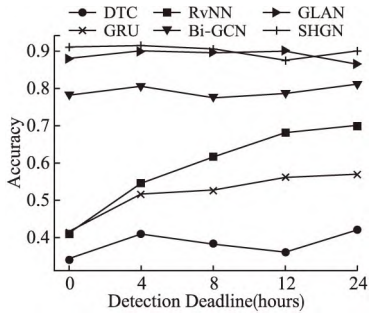

图 3 Twitter15 数据集上早期检测实验结果

Fig. 3 Experimental results of early detection on Twitter15 dataset

以看出,本文的 SHGN 模型使用不到 4 个小时的数据,在 2 个数据集上的准确率就达到 90% 以上,超过了其他基线模型,短时间内的高准确率表明本文模型具有优越的早期检测性能. 随着时间的推移,本文的 SHGN 模型和 GLAN 模型在 4h 到 12h 之间准确率有轻微波动,这是因为随着消息的传播,会有更多复杂的信息加入,不可避免地给模型带来噪声信息. 而对于早期检测任务,一个优秀的模型应该尽可能早的达到高指标的检测效果. 因此,可以判定本文提出的模型在早期检测

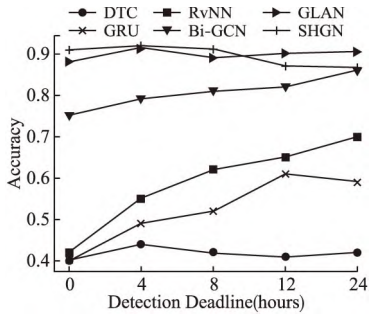

图 4 Twitter16 数据集上早期检测实验结果

Fig. 4 Experimental results of early detection on Twitter16 dataset

任务上比 GLAN 模型更优,也进一步表明本文模型对数据具有更强的鲁棒性和更稳定的性能. 综上所述,在 2 个真实世界的 Twitter 数据集上的实验结果表明,本文提出的方法在早期谣言检测方面比最先进的基线模型具有更好的性能.

4 总结与展望

本文提出了一个时序感知的异质图神经谣言检测模型. 该模型通过位置编码建模事件内部响应贴之间的时序关系,

并利用多头注意力机制关注重要的响应贴,然后利用图注意力机制的聚合函数融合源贴和响应贴,得到事件的局部时序表征;接着,基于用户与事件之间的交互关系,利用元素级注意力机制学习事件的全局结构表征;最后,将两种特征表示拼接用于谣言分类. 实验结果表明,本文提出的 SHGN 模型优于最先进的谣言检测方法,并且在早期检测任务上,也具有更高的检测性能.

在未来的研究中,本文将主要从 3 个方面继续深入工作:1)在局部结构的构建方面,寻找更加合适的建模方法(如根据转发、评论关系构建具有时序边的消息传播树),用以捕获更为丰富的消息传播结构;2)在算法层面,寻找合适的异质图嵌入算法,充分利用异质图带来的丰富异质信息以及不同类型节点之间的交互关系;3)由于社交媒体资源的多样性,本文还将考虑利用音频、图像等多模态信息以提高谣言检测性能.

References:

[1] DiFonzo Nicholas, Bordia Prashant. Rumor psychology: social and organizational approaches[M]. Washington DC: American Psychological Association, 2007.

[2] Yang Fan, Liu Yang, Yu Xiao-hui, et al. Automatic detection of rumor on sina weibo[C]//Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, 2012: 13-20.

[3] Castillo Carlos, Mendoza Marcelo, Poblete Barbara. Information credibility on twitter[C]//Proceedings of the 20th International Conference on World Wide Web, 2011: 675-684.

[4] Jin Fang, Dougherty Edward, Saraf Parang, et al. Epidemiological modeling of news and rumors on twitter[C]//Proceedings of the 7th Workshop on Social Network Mining and Analysis, 2013: 1-9.

[5] Sampson Justin, Morstatter Fred, Wu Liang, et al. Leveraging the implicit structure within social media for emergent rumor detection[C]//Proceedings of the 25th ACM International Conference on Information and Knowledge Management, 2016: 2377-2382.

[6] Ma Jing, Gao Wei, Wong Kam-fai. Detect rumors in microblog posts using propagation structure via kernel learning[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017: 708-717.

[7] Ma Jing, Gao Wei, Mitra Prasenjit, et al. Detecting rumors from microblogs with recurrent neural networks[C]//Proceedings of the 25th International Conference on Artificial Intelligence, New York, USA: AAAI Press, 2016: 3818-3824.

[8] Ma Jing, Gao Wei, Wong Kam-fai. Rumor detection on twitter with tree-structured recursive neural networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 1980-1989.

[9] LIAO X W, HUANG Z, YANG D D, et al. Rumor detection in social media based on a hierarchical attention network[J]. Scientia Sinica Informationis, 2018, 48(11): 1558-1574.

[10] Chen Tong, Wu Lin, Li Xue, et al. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection[C]//Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Berlin, Springer, 2018: 40-52.

[11] Khoo Ling Min Serena, Chieu Hai Leong, Qian Zhong, et al. Interpretable rumor detection in microblogs by attending to user interactions[C]//Proceedings of the Association for the Advancement of Articial Intelligence, 2020, 34(5): 8783-8790.

[12] Wei Peng-hui, Xu Nan, Mao Wen-ji. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019: 4786-4797.

[13] Bian Tian, Xiao Xi, Xu Ting-yang, et al. Rumor detection on social media with bi-directional graph convolutional networks[C]//Proceedings of the Association for the Advancement of Articial Intelligence, 2020, 34(1): 549-556.

[14] HU D, WEI L W, ZHOU W, et al. A rumor detection approach based on multi-relational propagation tree[J]. Journal of Computer Research and Development, 2021, 58(7): 1395-1411.

[15] Yang Xiao-yu, Lyu Yue-fei, Tian Tian, et al. Rumor detection on social media with graph structured adversarial learning[C]//Proceedings of the 29th International Joint Conference on Artificial Intelligence, 2021: 1417-1423.

[16] Yuan Chun-yuan, Ma Qian-wen, Zhou Wei, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection[C]//Proceedings of the IEEE International Conference on Data Mining, 2019: 796-805.

[17] Wu Ke, Yang Song, Zhu Kenny Q. False rumors detection on sina weibo by propagation structures[C]//Proceedings of the IEEE 31st International Conference on Data Engineering, 2015: 651-662.

[18] Sun Sheng-yun, Liu Hong-yan, He Jun, et al. Detecting event rumors on sina weibo automatically[C]//Proceedings of the Web Technologies and Applications 15th Asia-Pacific Web Conference, 2013: 120-131.

[19] Yu Feng, Liu Qiang, Wu Shu, et al. A convolutional approach for misinformation identification[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017: 3901-3907.

[20] LIU Z, WEI Z H, ZHANG R X. Rumor detection based on convolutional neural network[J]. Computer Application, 2017, 37(11): 3053-3056.

[21] ZHANG Y S, ZHENG J, HUANG G J, et al. Sentiment analysis method of Weibo based on dual attention model[J]. Journal of Tsinghua University (Natural Science Edition), 2018, 58(2): 122-130.

[22] PAN D Y, SONG Y R, SONG B. New microblog rumor detection model based on attention mechanism[J]. Journal of Chinese Computer Systems, 2021, 42(8): 1780-1786.

[23] Mikolov Tomas, Chen Kai, Corrado Greg, et al. Efficient estimation of word representations in vector space[C]//arXiv preprint arXiv: 1301.3781, 2013.

[24] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention is all you need[C]//Proceedings of the Advances in Neural Information Processing Systems, 2017: 6000-6010.

[25] Veličković Petar, Cucurull Guillem, Casanova Arantxa, et al. Graph attention networks[C]//arXiv preprint arXiv: 1710.10903, 2017.

[26] Li Chen, Hu Lin-mei, Shi Chuan, et al. Sequence-aware heterogeneous graph neural collaborative filtering[C]//Proceedings of the SIAM International Conference on Data Mining, 2021: 64-72.

附中文参考文献:

[9] 廖祥文,黄　知,杨定达. 基于分层注意力网络的社交媒体谣言检测[J]. 中国科学:信息科学,2018,48(11):1558-1574.

[14] 胡　斗,卫玲蔚,周　薇,等. 一种基于多关系传播树的谣言检测方法[J]. 计算机研究与发展,2021,58(7):1395-1411.

[20] 刘　政,卫志华,张韧弦. 基于卷积神经网络的谣言检测[J]. 计算机应用,2017,37(11):3053-3056.

[21] 张仰森,郑　佳,黄改娟,等. 基于双重注意力模型的微博情感分析方法[J]. 清华大学学报(自然科学版),2018,58(2):122-130.

[22] 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(8):1780-1786.