基于Bert-GNNs异质图注意力网络的早期谣言检测

欧阳祺,陈鸿昶,刘树新,王 凯,李 星
(中国人民解放军战略支援部队信息工程大学,河南郑州 450001)

摘 要: 网络谣言的广泛传播已经造成了很大的社会危害,因此早期谣言检测任务已成为重要的研究热点.现有谣言检测方法主要从文本内容、用户配置和传播结构中挖掘相关特征,但没有同时利用到文本全局语义关系和局部上下文语义关系.为了克服以上局限性,充分利用到谣言数据中的文本全局-局部上下文语义关系、文本语义内容特征和推文传播的结构特征,本文提出了一种基于Bert-GNNs异质图注意力网络的早期谣言检测算法(Bert-GNNs Heterogeneous Graph Attention Network,BGHGAN).该方法根据历史谣言集和用户特征构建一个推文-词-用户异质图,通过采用预训练语言模型Bert和图卷积神经网络(Graph Convolutional Network,GCN)结合的方法进行特征学习,以挖掘谣言的文本语义特征和文本之间的关系,并将异质图分解为推文-词子图和推文-用户子图,采用图注意力网络(Graph Attention network,GAT)的方式分别进行特征学习,从而更充分利用文本全局-局部上下文语义关系和传播图的全局结构关系以加强特征表达;最后,通过子图级注意力机制将不同模块的学习集成进行最终的谣言检测.所提算法在真实的Twitter15和Twitter16数据上进行实验,验证了该算法在检测准确率上分别为91.4%和91.9%,较现有最佳模型分别提高了1%和1.4%,也具备在早期阶段对谣言的检测能力;同时,本文通过实验探讨了不同特征对谣言检测的重要性、对异质图构建质量的重要性.

关键词: 虚假谣言;Bert-GCN模块;子图注意力网络模块;全局语义关系;全局结构关系;局部上下文语义关系
基金项目: 中原英才计划(No.212101510002)
中图分类号: TP393.1 **文献标识码:** A **文章编号:** 0372-2112(2024)01-0311-13
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20220882

Early Rumor Detection Based on Bert-GNNs Heterogeneous Graph Attention Network

OUYANG Qi, CHEN Hong-chang, LIU Shu-xin, WANG Kai, LI Xing
(*Strategic Support Force Information Engineering University*, *Zhengzhou*, *Henan* 450001, *China*)

Abstract: The widespread spread of network rumors has caused great harm to the society, so the task of early rumor detection has become an important research focus. The majority of existing methods for rumor detection focus on mining effective features from text contents, user profiles, and patterns of propagation, but these methods do not take full advantage of both global semantic relationship of text and local context semantic relationship. In order to overcome the above limitations and make full use of the text global-local context semantic relationship, text semantic content feature and the structural feature of tweet propagation in the rumor data, this paper puts forward a kind of early rumors detection algorithm based on Bert-GNNs heterogeneous graph attention network (BGHGAN). This method constructs a tweet-word-user heterogeneous graph according to historical rumor sets and user characteristics, using the method of combining Bert and GCN (Graph Convolutional Network) for feature learning to mine the relationship between the text semantic features and the text of rumors. And by decomposing the heterogeneous graph into tweet-word subgraph and tweet-user subgraph, the method uses GAT (Graph Attention network) to perform feature learning respectively, so as to make full use of the global-local context semantic relationship of the text and the global structure relationship of the propagation graph to strengthen the feature expression. Finally, the learning integration of different modules is carried out through the subgraph-level attention mechanism for final rumor detection. The proposed algorithm is experimented on real Twitter15 and Twitter16 data, and verifies that

收稿日期:2022-07-25;修回日期:2023-03-29;责任编辑:朱梅玉

the detection accuracy of the algorithm is 91.4% and 91.9%, respectively, which is 1% and 1.4% higher than the existing best model, and also has the ability to detect rumors in the early stage. And this paper discusses the importance of different features to rumor detection and the importance of the quality of heterogeneous graph construction.

Key words: fake news; Bert-GCN module; sub-graph attention network module; global semantic relationship; global structure relationship between the text; local contextual semantic relation

Foundation Item(s): Zhongyuan Program of Excellence Project (No.212101510002)

1 引言

随着移动互联网的普及以及各种网络设备的广泛使用,人们开始在网络上表达自己的观点,并会选择从网络社交媒体上获取各种信息.社交媒体带来极大便利性的同时,也滋生出各种各样的社会问题.由于社交媒体用户数量多,网络上的虚假谣言能迅速在社交媒体上传播[1],并且极易引起民众恐慌,甚至可能造成巨大的经济损失,给社会带来巨大的危害.由于人们的局限性,普通人很难从网络信息中区分出谣言[2].正因如此,它们可以在短时间内在人与人之间传播,严重阻碍人们获得真实的信息.比如,2016年美国总统大选期间,社交媒体上的假新闻激增,谣言泛滥对选举过程和选民心态产生了严重危害.这些谣言情绪鼓动性极强,降低了候选人的形象,加剧了政治氛围的紧张,对选举结果产生了巨大的影响[3].因此,开发一种自动辅助的方法来早期检测谣言很有必要.

早期的谣言检测方法大多利用特征工程从文本内容[4]、用户配置[5,6]和传播结构[7~9]中提取识别特征来训练分类器,例如决策树(tree decision)[10]、随机森林(random forest)[11]和支持向量机(Support Vector Machine,SVM)[12].同时,一些研究应用了更有效的特征,如用户评论[13]、文本情感[14]、时间结构特征[15]和主题[16],这些方法主要依赖特征工程,非常耗时[17].

因特征工程方法的缺陷,近期的研究逐步转向运用深度学习的方法从文本内容和传播路径中挖掘特征来检测谣言[18].例如,循环神经网络[19],包括长短时记忆网络(Long Short-Term Memory,LSTM)[20]、门控递归单元(Gate Recurrent Unit,GRU)[21]和递归神经网络(Recursive Neural Network,RvNN)[22].Ma等人[23]利用循环神经网络捕获每个源推文及其转发的语义变异,并根据语义变异进行检测,他们能从长期的谣言传播来学习序列特征.然而,由于时间结构特征只关注谣言的连续传播,而忽略了谣言扩散的影响,这些方法在效率上存在明显的局限性.

谣言的传播结构也反映了谣言的一些传播行为.因此,一些研究试图通过调用基于卷积神经网络(Convolutional Neural Networks,CNN)的方法来挖掘谣言扩散结构中的信息[24].基于CNN的方法可以获得局部邻居内部的相关特征,但不能处理图或树中的全局结构关系[25].这些方法忽略了谣言传播的全局结构特征.Bian等人[26]提出了一个新的双向图模型,称为双向图卷积网络(Binary Graph Convolutional Network,Bi-GCN),通过操作自上而下和自下而上的谣言传播来探索这两个特性,利用GCN和自上而下的定向谣言传播图来了解谣言传播的模式,同时利用一个具有相反方向的谣言扩散图的GCN来捕捉谣言的扩散结构.Yuan等人[27]探索了一种全局-局部注意网络,捕捉源推文传播的局部上下文语义关系和传播图的全局结构关系,用于谣言检测,验证了局部上下文语义关系在对谣言特征学习时起到重要作用.Huang等人[28]构建了一种基于元路径的图注意力框架,用于获取文本之间的全局语义关系和源推文传播的结构信息,这种全局语义关系表示了谣言的共性,是谣言检测的关键因素.

虽然上述方法可对早期谣言进行有效的检测,但没有同时利用文本全局语义关系和局部上下文语义关系,未能有效挖掘数据的潜在特征.为了同时提取文本全局语义关系、文本内容特征和文本的推文传播的全局结构特征,本文提出了一种基于Bert-GNNs异质图注意力网络的早期谣言检测算法.该方法在构建文本-词-用户异质图的基础上,分别采用Bert-GCN模块和子图注意力网络模块,并通过集成模块以融合推文中的各种信息,实现谣言检测.本方法首先基于谣言的文本内容和推文传播情况构建异构推文-词-用户图,如图1所示.两个词节点之间的边(即红线)根据词共现信息构建;单词节点和推文节点之间的边(即黑线)表示单词在推文中;推文节点和用户节点之间的边(即绿线)表示用户转发或回复源推相关推文.同时将异质图分解为推文-词子图和推文-用户子图,对于推文-词子图采用Bert模型初始化文档结点,并通过图卷积来学习文本之间的语义内容,得到节点表示.然后继续利用注意力网络学习两个子图的节点表示,随后引入一种注意机制来融合子图中节点的表示,并将两者的表示进行集成,得到最终的节点表示,用于谣言检测,并具备早期谣言检测能力.

本文的贡献如下:

(1)本文提出了一种基于Bert-GNNs异质图注意力

网络的早期谣言检测算法,通过构建推文-词-用户异质图对两个子图以及推文内容分别进行特征提取学习,再集成节点特征,从而更加充分利用文本语义关系、文本语义特征和文本的全局结构关系;

(2)本文设计了一种基于子图级注意力机制,以自动学习不同模块之间的重要性,从而使不同模块提取特征后的集成更加方便;

(3)所提方法在真实的Twitter15和Twitter16数据集上进行了实验,验证了该方法在准确率上优于现有模型,也具备在早期阶段对谣言的检测能力,并且探讨了不同特征对谣言检测影响的重要性以及异质图构建质量的重要性.

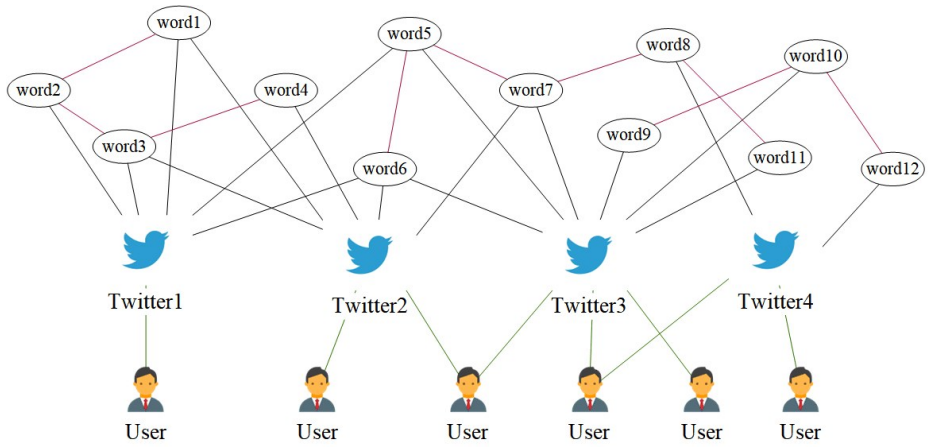


图1　推文-词-用户异质图

2　相关工作

谣言检测的目标是通过各个社交媒体平台上发布的文本信息,如文本内容、传播结构、情感、用户配置等,来区分是否为谣言. 现有谣言检测方法主要分为传统机器学习方法和深度学习方法.

2.1　传统机器学习方法

目前,统计机器学习仍被广泛用于谣言检测,通过特征工程从数据中心提取文本内容、用户配置和传播结构等特征来识别谣言. 早在2011年,Castillo等人[29]通过对热门话题相关微博帖子的分析,并用发布的文本特征和引用的外部源对帖子进行可信度评估,采用支持向量机、决策树、决策规则以及贝叶斯网络4种算法对其进行验证. Qazvinian等人[30]基于内容的特征、网络的特征以及Twitter特定模式特征,分析推文中词汇模式、词性模式以及特定于Twitter模式中提取出的标签和URL. Yang等人[31]在之前的基础上,基于客户和位置的特征,分析用户使用的客户端程序,同时根据微博事件发生的实际地点对谣言进行判断. Ma等人[32]基于谣言生命周期的时间序列,探索了这些时间特征,同时整合了各种社会语境信息. Kwon等人[33]基于谣言传播的时间、结构和语言特征来进行谣言分析检测. 这些方法严重依赖特征工程,非常费时耗力,且由于手工提取特征,不一定适用于所有数据集.

2.2　深度学习方法

近年来,为了解决传统的基于特征的机器学习方法存在的上述问题,研究者们开始利用深度学习模型来自动学习有效特征用于谣言检测. Ma等人[23]首次将递归神经网络用于检测谣言. Chen等人[34]在Ma等人研究的基础上引入了注意力机制,通过对特征加上权重衡量,使谣言的分类模型给某一个或几个分配更大的权重,并通过学习推文时间序列的表示形式以进行谣言识别. Vaibhav等人[35]提出了一种基于图神经网络(Graph Neural Network,GNN)的虚假新闻检测模型,该模型针对推文中语句之间的语义关系进行建模. Bian等人[26]提出了一种双向GCN用于谣言检测的方法,该方法考虑了谣言的序列传播和横向扩散,以此来捕捉谣言传播结构的全局特征. Yuan等人[27]提出了一种全局-局部注意网络(Global-Local Attention Network,GLAN),捕捉源推文传播拓扑树的全局结构特征和谣言局部上下文的关系,用于谣言检测. Sharma等人[36]提出了一种弱监督学习方法,利用用户回复信息并使用推文的情感分析来识别谣言. Huang等人[28]则利用一种基于元路径的异质图注意网络框架来捕获文本内容的全局语义关系. 这种语义关系是谣言的共性特征,也是检测谣言的关键因素. 这些方法虽能进行有效的检测,但没有同时利用到文本全局语义关系和局部上下文语义关系.

3 问题定义

3.1 异质推文-词-用户图构建

本文基于推文内容、转发推文和用户构建了一个推文-词-用户异质图，如图1所示，其中包含了谣言的文本内容和推文传播中涉及的信息．将构造的推文-词-用户异质图定义为$G=(V,E)$，其中，V和E表示图中的节点和边．节点V由推文集合T、推文所含的词集合W和用户集合U组成，每篇推文、每个词和每个用户均作为图的节点．边E由推文-单词边集合E_{tw}，单词-单词边集合E_{ww}和推文-用户边集合E_{tu}组成．E_{tw}反应了推文所含单词的关系，E_{ww}反应了单词于单词之间的上下文语义关系，E_{tu}反应了用户对推文的行为，包括转发和评论等．而边权重则更具体地反应了各种信息特征：E_{tw}边权重由该单词在源推文中的词频逆文档频率(TF-IDF)计算，其中词频(Term Frequency，TF)是某一个给定的词语在该文件中出现的频率，逆文档频率(Inverse Document Frequency，IDF)是该词在源推文中的出现次数与包含该词的推文总数之比．所以如果某个单词在一篇推文中出现的频率(TF)高，并且在其他推文中很少出现，则认为此词或者短语具有很好的类别区分能力．此时，边权值较大．E_{ww}边权重由逐点互信息(Point-wise Mutual Information，PMI)计算词语之间的相关性，两个词相关性越强，边权值越大．E_{tu}边权重由用户转发或回复源推文的时间的倒数计算，用户转发或回复源推文时间越短，边权值越大．

具体的计算如下：

$$A_{ij}=\begin{cases}\mathrm{PMI}(i,j), & i,j\text{都是词},\mathrm{PMI}(i,j)>0\\ \mathrm{TF-IDF}_{ij}, & i\text{是推文},j\text{是词}\\ \dfrac{1}{t+1}, & i\text{是推文},j\text{是用户}\\ 1, & i=j\\ 0, & \text{其他}\end{cases} \tag{1}$$

其中，t表示距离源推文i发布到用户j转发或回复所用的时间，PMI计算公式如下：

$$\mathrm{PMI}(i,j)=\log\frac{p(i,j)}{p(i)p(j)} \tag{2}$$

$$p(i,j)=\frac{W(i,j)}{W} \tag{3}$$

$$p(i)=\frac{W(i)}{W} \tag{4}$$

其中，$W(i,j)$表示词i和词j出现在同一推文的数量，W表示推文总数量，$W(i)$表示包含词i的推文数．TF-IDF计算公式如下：

$$\mathrm{TF\text{-}IDF}_{ij}=\mathrm{TF}_{ij}\times\mathrm{IDF}_j \tag{5}$$

$$\mathrm{TF}_{ij}=\frac{n_{ij}}{\sum_k n_{ik}} \tag{6}$$

$$\mathrm{IDF}_j=\log\frac{W}{W(i)} \tag{7}$$

其中，n_{ij}表示词j在推文i中出现的数量．

3.2 问题定义

给定一个推文-词-用户异质图$G=(V,E)$，其中$V=\{T,W,U\}$表示图中节点集，$E=\{E_{tw},E_{ww},E_{tu}\}$表示图中边集．$T$为推文集，$T=\left\{t_1,t_2,\cdots,t_{|T|}\right\}$，$|T|$表示推文数量；$W$为谣言中的词集，$W=\left\{w_1,w_2,\cdots,w_{|W|}\right\}$，$|W|$表示单词数量；$U$表示用户集，$U=\left\{u_1,u_2,\cdots,u_{|U|}\right\}$，$|U|$表示用户数量．

本文通过学习函数$p(c|t_i,G,\theta)$确定推文t_i的各个标签的概率．c表示种类标签；θ表示待学习的模型参数．

4 模型介绍

本文的目的是对早期谣言进行检测，即判断该谣言类型，具体流程如图2所示．首先，根据历史谣言集构建一个推文-用户-词异质图；然后，考虑到谣言本身的文本语义特征和文本之间的关系，采用Bert和GCN结合的方法进行特征学习；其次，为充分利用到文本全局-局部上下文语义关系和全局结构关系，将异质图分解为推文-词子图和推文-用户子图，采用图注意力网络(Graph Attention network，GAT)的方式分别进行特征学习；最后，采用子图级注意力机制将不同模块的学习集成进行最终的谣言检测．

4.1 Bert-GCN模块

考虑到Bert预训练模型强大的语言表征能力和特征提取能力，为了充分获取每个推文中的文本语义内容，本文对于最初的文档结点用预训练到Bert模型初始化得出节点特征，将节点和节点特征输入至GCN网络中，再联合训练Bert与GCN，充分融合二者处理数据语义结构、提取特征的能力．文档嵌入用$\boldsymbol{X}_{\mathrm{Bert}}\in\mathbb{R}^{|T|\times d}$表示，由Bert预训练模型初始化，$d_{\mathrm{input}}$为嵌入维数．综上所述，初始节点特征矩阵为

$$\boldsymbol{X}_{\mathrm{init}}=\begin{pmatrix}\boldsymbol{X}_{\mathrm{Bert}}\\ 0\end{pmatrix}\in\mathbb{R}^{(|T|+|W|)\times d_{\mathrm{input}}} \tag{8}$$

随后，为了获取文本之间的关联性，将$\boldsymbol{X}$输入到GCN模块[37]中，其中，GCN的输出为

$$\begin{aligned}\boldsymbol{X}_{\mathrm{Bert\text{-}GCN}}&=f\left(\boldsymbol{X}_{\mathrm{init}},\boldsymbol{A}\right)\\&=\mathrm{softmax}\left(\hat{\boldsymbol{A}}\,\mathrm{ReLU}\left(\hat{\boldsymbol{A}}\,\boldsymbol{X}_{\mathrm{init}}\boldsymbol{W}^{(0)}\right)\boldsymbol{W}^{(1)}\right)\end{aligned} \tag{9}$$

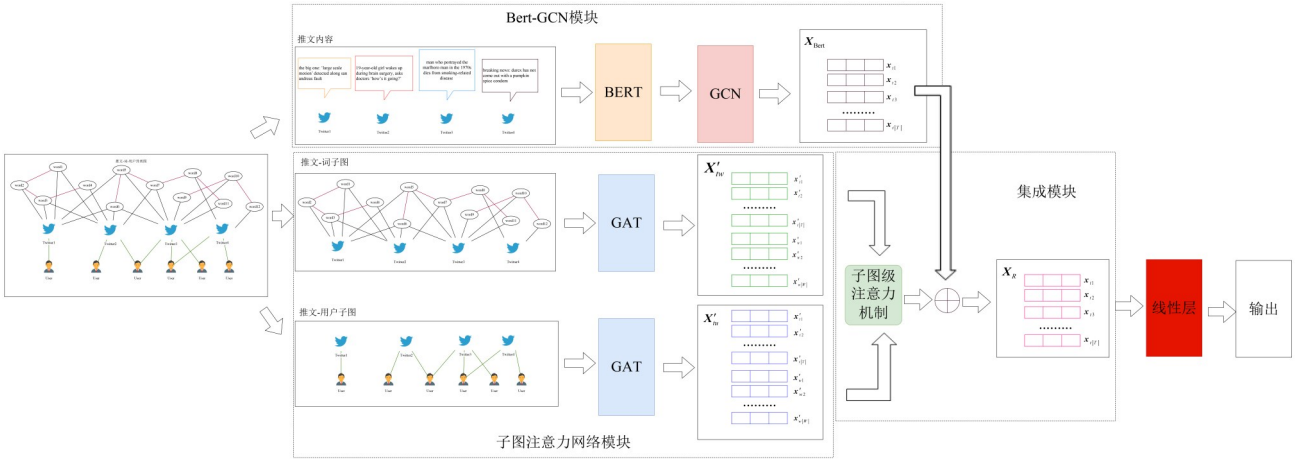


图2　算法框架

其中，$\hat{\boldsymbol{A}}$ 是归一化邻接矩阵，$\boldsymbol{W}^{(0)} \in \mathbb{R}^{d_{\text{input}} \times H}$ 是将模型输入映射到 H 个特征隐藏层的权重矩阵，$\boldsymbol{W}^{(1)} \in \mathbb{R}^{H \times d_{\text{output}}}$ 是将隐藏层的输出作为输入映射到最终模块输出的权重矩阵，$\boldsymbol{X}_{\text{Bert-GCN}}$ 为本模块最终学习到的特征向量.

4.2　子图注意力网络模块

为了充分获取推文中文本之间的语义关系和推文传播的全局结构信息，同时考虑到两个子图中节点的邻居对于学习谣言检测的节点嵌入具有不同的重要性. 受图注意网络[38]的启发，本模块采用一种子图注意力机制，利用注意机制学习每个节点邻居的重要性，并将这些邻居的表示与重要性合并，形成每个节点的表示.

在本模块中，将词集特征表示为 $\boldsymbol{X}_W=\left[\boldsymbol{x}_{w_1},\boldsymbol{x}_{w_2},\cdots,\boldsymbol{x}_{w_{|W|}}\right]$，$\boldsymbol{X}_W \in \mathbb{R}^{|W| \times N}$，$\boldsymbol{x}_{w_i}$ 为词 w_i 的词嵌入表示，N 为特征向量维度. 推文集 T 特征表示为 $\boldsymbol{X}_T=\left[\boldsymbol{x}_{t_1},\boldsymbol{x}_{t_2},\cdots,\boldsymbol{x}_{t_{|T|}}\right]$，$\boldsymbol{X}_T \in \mathbb{R}^{|T| \times N}$. 用户集 U 表示为 $\boldsymbol{X}_U=\left[\boldsymbol{x}_{u_1},\boldsymbol{x}_{u_2},\cdots,\boldsymbol{x}_{u_{|U|}}\right]$，$\boldsymbol{X}_U \in \mathbb{R}^{|U| \times N}$.

在推文-词子图中，$\boldsymbol{X}_{tw}=\left[\boldsymbol{x}_{t_1},\boldsymbol{x}_{t_2},\cdots,\boldsymbol{x}_{t_{|T|}},\boldsymbol{x}_{w_1},\boldsymbol{x}_{w_2},\cdots,\boldsymbol{x}_{w_{|W|}}\right]$，此处的 $\boldsymbol{x}_{t_i}$ 计算公式为 $\boldsymbol{x}_{t_i}=\dfrac{1}{|t_i|}\sum_{w_j \in t_i}\boldsymbol{x}_{w_j}$. 而在推文-用户子图中，$\boldsymbol{X}_{tu}=\left[\boldsymbol{x}_{t_1},\boldsymbol{x}_{t_2},\cdots,\boldsymbol{x}_{t_{|T|}},\boldsymbol{x}_{u_1},\boldsymbol{x}_{u_2},\cdots,\boldsymbol{x}_{u_{|U|}}\right]$，此处的特征表示均为基于词向量的随机初始化，向量值在正态分布 N(0,1) 中随机取值.

然后利用自注意力[39]来学习子图中节点之间的权值. 给定子图中的节点对 (i,j)，自我注意机制 f 可以学习到节点 j 对节点 i 表示的注意系数 $e_{i,j}$，它表示节点 j 对节点 i 表示的重要性. 节点对 (i,j) 的注意系数 $e_{i,j}$ 可以计算如下：

$$e_{i,j}=f\left(\boldsymbol{W}\boldsymbol{x}_i,\boldsymbol{W}\boldsymbol{x}_j\right),\boldsymbol{x}_i,\boldsymbol{x}_j \in \boldsymbol{X}_{tw(tu)} \tag{10}$$

其中，f 可以由一个单层前馈神经网络实现，该神经网络由一个权向量 $\boldsymbol{a}$ 参数化，并应用 LeakyReLU 作为激活函数[40]；$\boldsymbol{W}$ 表示共享的变化矩阵，$\boldsymbol{W} \in \mathbb{R}^{d \times N}$，$d$ 为最终词嵌入向量.

随后，本模块通过用注意机制将子图结构信息注入到模型中，具体来讲就是只计算节点 $j \in \mathcal{N}_i$ 的注意力系数 $e_{i,j}$，其中，$\mathcal{N}_i$ 是图中节点 $\mathcal{N}_i$ 的邻域（包括 i 节点本身）. 在所有的实验中，这些将恰好是 i（包括 i 节点本身）的一阶邻居. 为了使不同节点之间的系数易于比较，使用 softmax 函数对所有 j 进行规范化：

$$\begin{aligned}\alpha_{i,j}&=\text{softmax}\left(e_{i,j}\right)\\&=\frac{\exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^{\text{T}}\cdot\left[\boldsymbol{W}\boldsymbol{x}_i /\!/ \boldsymbol{W}\boldsymbol{x}_j\right]\right)\right)}{\sum_{k \in \mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^{\text{T}}\cdot\left[\boldsymbol{W}\boldsymbol{x}_i /\!/ \boldsymbol{W}\boldsymbol{x}_k\right]\right)\right)}\end{aligned} \tag{11}$$

其中，$\boldsymbol{a}$ 是神经网络的权值向量；$(\cdot)^{\text{T}}$ 表示转置矩阵；$/\!/$ 表示拼接操作. 之后聚合子图节点 i 的邻居表示和其对应的系数来更新节点 i 的嵌入表示，具体计算如下：

$$\boldsymbol{x}_i^{(1)}=\sigma\left(\sum_{j \in \mathcal{N}_i}\alpha_{i,j}^k\boldsymbol{W}^k\boldsymbol{x}_j\right) \tag{12}$$

其中，$\boldsymbol{x}_i^{(1)}$ 表示节点 i 的更新嵌入；σ 表示一个非线性函数；$\mathcal{N}_i$ 表示节点 i 和其邻居的集合.

最后，为了稳定自注意力学习的过程，将自注意扩展到类似于图注意网络[38]的多头注意机制入. 具体来说，对式(12)进行 K 次变换，得到 K 个独立的注意力特征，并将它们学习到的特征进行拼接，得到最终的输出表示为

$$\boldsymbol{x}_i'=/\!/_{k=1}^{K}\sigma\left(\sum_{j \in \mathcal{N}_i}\alpha_{i,j}^k\boldsymbol{W}^k\boldsymbol{x}_j\right) \tag{13}$$

由上述操作，给定上述推文-词子图中特征表示 $\boldsymbol{X}_{tw}$ 和推文-用户中的特征表示 $\boldsymbol{X}_{tu}$，可以最终得到包含推文

中文本内的全局-局部上下文语义关系的推文-词的嵌入表示$\boldsymbol{X}'_{tw}=\left[\boldsymbol{x}'_{t_1},\boldsymbol{x}'_{t_2},\cdots,\boldsymbol{x}'_{t_{|T|}},\boldsymbol{x}'_{w_1},\boldsymbol{x}'_{w_2},\cdots,\boldsymbol{x}'_{w_{|W|}}\right]$和包含推文传播的全局结构信息推文-用户的词嵌入表示$\boldsymbol{X}'_{tu}=\left[\boldsymbol{x}'_{t_1},\boldsymbol{x}'_{t_2},\cdots,\boldsymbol{x}'_{t_{|T|}},\boldsymbol{x}'_{u_1},\boldsymbol{x}'_{u_2},\cdots,\boldsymbol{x}'_{u_{|U|}}\right]$,其中,$\boldsymbol{X}'_{tw}\in\mathbb{R}^{(|T|+|W|)\times d_{\text{output}}}$,$\boldsymbol{X}'_{tu}\in\mathbb{R}^{(|T|+|U|)\times d_{\text{output}}}$.

4.3 集成模块

4.3.1 子图级注意力机制

异质图中按元路径分解的子图包含不同的信息.推文内容子图中包含文本内容的特征信息,经过GCN模块后进一步学习到其推文与推文之间的全局语义关系特征,推文-词子图包含文本内容的局部语义上下文关系信息,推文-用户子图包含源推文传播过程中涉及的结构信息.为了准确识别谣言,需要融合3个子图中包含的信息.

首先,对于推文-词子图和推文-用户子图,本模块采用子图级注意机制来学习子图权值.计算权值如下:

$$\left(\beta_{tw},\beta_{tu}\right)=\text{att}_{\text{FNN}}\left(\boldsymbol{X}'_{tw},\boldsymbol{X}'_{tu}\right) \tag{14}$$

其中,att_{FNN}代表执行子图级注意的前馈神经网络.具体来说,就是先通过非线性变换,而后将转换后的节点和子图注意力向量$\boldsymbol{a}$的相似度作为节点的重要性,子图的重要性则有节点重要性的加权平均得到.具体的计算公式如下:

$$\boldsymbol{w}_{tw}=\frac{1}{\left|\boldsymbol{X}'_{tw}\right|}\sum_{\boldsymbol{x}_i\in\boldsymbol{X}'_{tw}}\boldsymbol{a}^{\mathrm{T}}\cdot\tanh\left(\boldsymbol{W}_{\text{sub}}\boldsymbol{x}_i\right) \tag{15}$$

$$\boldsymbol{w}_{tu}=\frac{1}{\left|\boldsymbol{X}'_{tu}\right|}\sum_{\boldsymbol{x}_i\in\boldsymbol{X}'_{tu}}\boldsymbol{a}^{\mathrm{T}}\cdot\tanh\left(\boldsymbol{W}_{\text{sub}}\boldsymbol{x}_i\right) \tag{16}$$

其中,$\boldsymbol{W}_{\text{sub}}$为权重矩阵,$\boldsymbol{a}$为子图注意力向量,且两者共用该向量.随后用softmax函数对其进行归一化,得到两者权值:

$$\beta_{tw}=\text{softmax}\left(\boldsymbol{w}_{tw}\right)=\frac{\exp\left(\boldsymbol{w}_{tw}\right)}{\sum_{\varphi\in\{tw,tu\}}\exp\left(\boldsymbol{w}_{\varphi}\right)} \tag{17}$$

因此,这两个模块融合后的特征向量为$\boldsymbol{X}_{\text{sub}}=\left\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{|T|}\right\}$.其中,

$$\boldsymbol{x}_i=\sum_{\varphi\in tw,tu}\beta_{\varphi}\cdot\boldsymbol{x}_{t_i},\boldsymbol{x}_{t_i}\in\boldsymbol{X}'_{\varphi} \tag{18}$$

4.3.2 集成

对于Bert-GCN模块得到的特征向量$\boldsymbol{X}_{\text{Bert-GCN}}$和子图注意力网络模块融合后得到的特征向量$\boldsymbol{X}_{\text{sub}}$,最终训练目标是将两者特征进行线性加权,得到

$$\boldsymbol{X}_T=\lambda\boldsymbol{X}_{\text{sub}}+(1-\lambda)\boldsymbol{X}_{\text{Bert-GCN}} \tag{19}$$

本文用λ控制两个训练目标之间的权衡.通过调整λ就能平衡模块之间的特征,使最终结果能得到更好的优化.

4.4 损失函数和谣言分类

在本工作中,将上述得到的源推文的特征表示$\boldsymbol{X}_T$输入到一层前反馈神经网络(Feedforward Neural Network,FNN)中,并通过softmax函数归一化的方式得到预测源推文类别的类概率分布:

$$p\left(c\left|t_i,G;\theta\right.\right)=\text{softmax}\left(\text{FNN}\left(\boldsymbol{x}_i\right)\right),\boldsymbol{x}_i\in\boldsymbol{X}_R \tag{20}$$

同时,为了训练模型的参数,通过交叉熵损失和正则化项作为目标来优化函数:

$$\mathcal{L}=-\sum_{i\in|T|}y_i p\left(c\left|t_i,G;\theta\right.\right)+\lambda\left\|\theta\right\|_2^2 \tag{21}$$

其中,y_i表示真实推文根据其所属种类的独热编码;λ表示权衡系数;$\|\cdot\|_2^2$表示L2的正则化项,以防止过拟合.

综上,对谣言数据集,为充分利用到谣言数据中的文本全局-局部上下文语义关系、文本语义内容特征和推文传播的全局结构关系,本文提出了一种基于Bert-GNNs异质图注意力网络的早期谣言检测算法,具体流程如算法1所示.

算法1 HGBGNN算法

输入: 异质推文-词-用户图$G(V,E)$;推文集$T=\left\{t_1,t_2,\cdots,t_{|T|}\right\}$

输出: 推文检测结果

步骤1: 通过式(1)~(3)得到邻接矩阵$\boldsymbol{A}$

步骤2: 分解异质推文-词-用户图$G(V,E)$为推文-词子图$G(V,E)_{TW}$和推文-用户子图$G(V,E)_{TU}$

步骤3: 对于推文集$T=\left\{t_1,t_2,\cdots,t_{|T|}\right\}$,采用Bert-GCN模块,将推文内容输入Bert预训练模型,得到文本特征后将文本视为节点再输入GCN模型得到该模块表示

步骤4: 对于推文-词子图$G(V,E)_{TW}$,将文本、词节点特征和邻接矩阵$\boldsymbol{A}$输入到GAT模型中

步骤5: 对于推文-用户子图$G(V,E)_{TU}$,将文本、用户节点特征和邻接矩阵$\boldsymbol{A}$输入到GAT模型中

步骤6: 共同训练两个子图分别得到两个子图的图嵌入表示

步骤7: 通过子图级注意力机制来学习两个子图的重要性,并通过集成模块将两个模块学习到的图嵌入表示集成

步骤8: 通过单层MLP后得到最终结果

5 实验

5.1 数据集

本文实验基于两个公开的Twitter数据分别记录为Twitter15和Twitter16,这些数据来源于Ma等人[9].其中Twitter15包含1 490条源推文,Twitter16包含818条谣言源推文,具体数据如表1所示.每一种推文都有4种标签:非谣言(non-rumors),假谣言(false-rumors),真谣言(true-rumors),未经证实的谣言(unverified rumors).由于原始数据集不包括用户信息,本文采取的是Huang

等人[28]与源推文有关的所有用户相关信息.

表1　数据集

统计种类	Twitter15	Twitter16
源推文数量	1 490	818
用户数量	276 663	173 487
推文数量	331 612	204 820
非谣言	374	205
假谣言	370	205
真谣言	372	207
未经证实的谣言	374	201

5.2　实验设置

本实验是通过Intel(R)Core(TM)i5-7400 CPU @ 2* 3.00 GHz, RAM 8.0 GB, Windows7并连接服务器NVIDIA RTX3090的单卡GPU来完成的.软件方面采用Python3.7.10实现,同时使用深度图库(Deep Graph Library,DGL),后端采用PyTorch框架,并采用Adam作为模型的优化器.

实验设置方面,本实验设置Bert-GCN模块中Bert模型学习率(learning rate)为0.000 01,其他学习率均设置为0.001,学习率为0.000 1,权重衰减(weight decay)为0.000 5.为了防止过拟合,设置丢弃率(dropout)为0.5. 为了保持较高的计算效率,设置多头注意力头数K为8,为了充分训练模型的参数,训练批次(epochs)设置为100,一次训练所选样本数(batch size)为64,每个Batch包括推文集、图结构和特征向量.

在实验数据方面,本文随机选取10%的数据集作为验证集,剩余数据按照3∶1的比例分别作为训练集和测试集.

5.3　评价指标

本文采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)以及F_1值(F_1-score)作为模型的评价指标,具体计算公式如下:

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{22}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{23}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{24}$$

$$F_1\text{-score}=\frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}} \tag{25}$$

其中,TP,TN,FP,FN分别表示真正例数、真负例数、假正例数和假负例数. 由于这是一个多分类问题,所以指标比较时,分别选择对非谣言(NR)、假谣言(FR)、真谣言(TR)、未经证实的谣言(UR)4类进行F_1值的计算.

5.4　基准模型

将本文的模型和一些基线谣言检测模型进行对比:

(1)DTR[10]:基于决策树的模型,通过正则表达式对从推文中提取的集群进行排序来识别谣言.

(2)SVM-TS[32]:利用时间序列中人工社会背景特征变化的线性支持向量机分类器.

(3)BU-RvNN和TD-RvNN[12]:基于传播树遍历的递归神经网络对谣言的传播结构进行建模和学习,分为Bottom-Up(BN)模型和Top-Down(TD)模型,两者以不同方式对传播树结构进行表示.

(4)Bi-GCN[26]:一个同时处理谣言传播深度和广泛散布结构的方法,从谣言自顶向下(top-down)和自底向上(bottom-up)的传播方向上发掘这两个特征.

(5)PPC_RNN+CNN[41]:一个结合RNN和CNN来构建时间序列分类器分别捕捉用户特征在传播路径上的全局和局部变化,以检测假新闻.

(6)GLAN[27]:提出了一种新型的全局-局部注意网络,将局部语义信息和全局结构信息联合编码,用于谣言检测.

(7)EBGCN[42]:提出了一种新的边缘增强贝叶斯图卷积网络来捕获更具鲁棒的结构特征,采用贝叶斯方法自适应地重新考虑潜在关系的可靠性,以检测虚假新闻.

(8)RDGCN[43]:对谣言的区域化传播模式进行了研究,提出了一种新的区域增强深度图卷积网络,通过学习区域化传播模式来增强谣言的传播特性,并通过无监督学习来训练谣言传播模式的学习.

(9)FedGCN[44]:将联邦学习范式与双向图注意网络谣言检测模型相结合,利用横向联邦学习的优势进行跨平台谣言检测,保证各社交平台数据信息的安全性,并用于检测假新闻.

5.5　模型检测能力评估

如表2和表3所示(加粗数据表示最优结果),本文的方法在Twitter15和Twitter16两个数据集训练后的测试性能优于所有其他基准模型. 本文提出的方法在谣言检测实现方面准确率分别达到了91.4%和91.9%,较现有最佳模型分别提高了1%和1.4%. 这表明了本文算法性能的优越性,因为其充分利用文本全局-局部上下文语义关系、文本语义内容特征和推文传播的全局结构关系,对谣言检测有着更显著的改善.

深度学习方法(如BU-RvNN, TD-RvNN, PPC, GLAN)都比基于传统机器学习的方法有更好的性能,这说明深度学习方法更容易捕获对谣言识别有用的特征用于谣言检测.

经对比还发现,基于传统机器学习的方法(如DTR,SVM-TS)的对比基线表现不佳,因为基于传统机器学习的方法难以挖掘数据集的深度特征. 在这些基线方法中,SVM-TS比DTR好,这主要是因为SVM-TS相

较于其他方法利用了额外的时间特性.

深度学习方法(如 BU-RvNN, TD-RvNN, PPC_RNN+CNN, Bi-GCN, GLAN, FedGCN, RDGCN(6-layers),EBGCN)都比基于传统机器学习的方法有更好的性能,这表明深度学习方法更容易捕获有效的特征用于谣言检测,并且运用到图神经网络的方法(如 Bi-GCN,GLAN,FedGCN,RDGCN(6-layers),EBGCN)要比未用到图神经网络的方法(如 BU-RvNN,TD-RvNN,PPC_RNN+CNN)更有效,因为图神经网络更能捕获谣言之间的关系以及谣言传播结构的特征.

表 2 Twitter15 数据集谣言检测结果

方法	准确率	NR F_1值	FR F_1值	TR F_1值	UR F_1值
DTR	0.409	0.501	0.311	0.364	0.473
SVM-TS	0.544	0.796	0.472	0.404	0.483
BU-RvNN	0.708	0.695	0.728	0.759	0.653
TD-RvNN	0.723	0.682	0.758	0.821	0.654
Bi-GCN	0.866	0.892	0.896	0.921	0.861
PPC_RNN+CNN	0.842	0.811	0.875	0.790	0.818
GLAN	0.904	0.924	**0.917**	0.852	**0.927**
FedGCN	0.841	0.809	0.850	0.815	0.809
RDGCN(6-layers)	0.856	0.807	0.876	0.910	0.816
EBGCN	0.892	0.869	0.897	0.934	0.867
HGBGAN	**0.914**	**0.928**	0.907	**0.935**	0.885

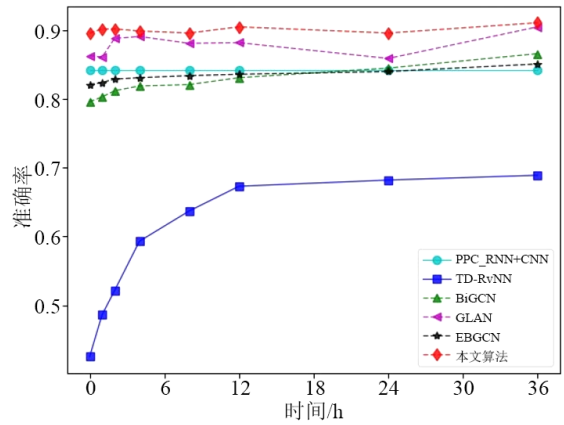
表 3 Twitter16 数据集谣言检测结果

方法	准确率	NR F_1值	FR F_1值	TR F_1值	UR F_1值
DTR	0.414	0.394	0.630	0.344	0.473
SVM-TS	0.544	0.796	0.472	0.404	0.483
BU-RvNN	0.718	0.723	0.712	0.779	0.659
TD-RvNN	0.737	0.662	0.743	0.835	0.708
Bi-GCN	0.876	0.847	0.862	0.931	0.862
PPC_RNN+CNN	0.863	0.820	0.898	0.837	0.843
GLAN	0.902	**0.921**	0.869	0.847	**0.968**
FedGCN	0.891	0.799	0.876	0.959	0.885
RDGCN(6-layers)	0.878	0.810	0.881	0.945	0.879
EBGCN	0.905	0.879	0.906	**0.967**	0.910
HGBGAN	**0.919**	0.868	**0.936**	0.948	0.915

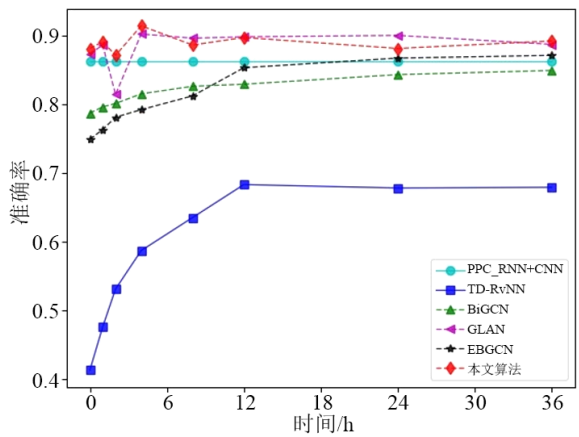
5.6 早期谣言检测能力评估

众所周知,越早对谣言进行检测,最终所造成的恶劣影响就越小,所以及早发现谣言就显得尤为重要. 为了模拟谣言早期检测的过程,本文通过控制谣言源推文发布后所经过的时间或用户转发数来模拟谣言传播的不同时期,然后计算不同时期的谣言检测准确率来评估性能.

为了验证本文所提方法谣言早期检测能力,本节将该算法与 PPC_RNN+CNN,TD-RvNN,BiGCN,GLAN,EBGCN 这 5 个最新的基线进行了比较,经测试后实验结果如图 3 和图 4 所示.



(*a*) Twitter15 数据集经过不同时间准确性对比

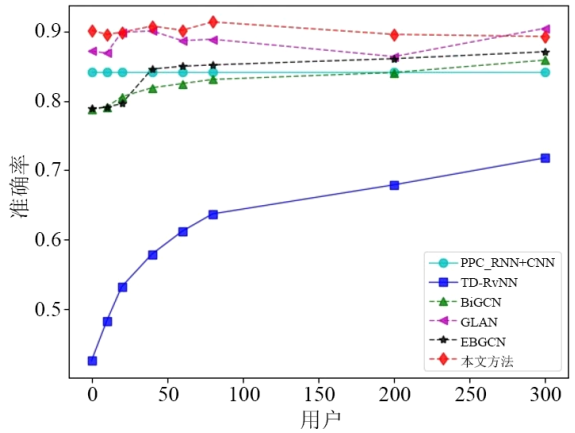


(*b*) Twitter16 数据集经过不同时间准确性对比

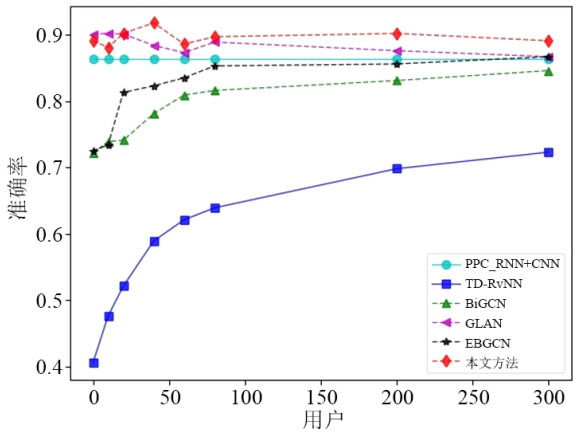
图 3 源推文发布经过不同时间准确性对比

图 3 显示了源推文发布经过不同时间准确性对比,图 4 显示了源推文发布由不同用户转发数的准确性对比. 可以看出,本文算法在任意时间和转发次数下性能均优于 PPC_RNN+CNN,TD-RvNN,BiGCN,EBGCN 这 4 种算法,大部分情况优于 GLAN 算法,而经过时间或用户转发次数的增加,本文方法的谣言检测准确率会有轻微的波动,在少部分情况下略差于 GLAN 算法. 这是因为次数的变化或者时间的变化会导致信息量的增加,这也可能给模型的学习带来噪声信息或者冗余信息,影响最终精度.

但是对于早期检测任务,一个好的解决方案应该尽早达到合适的性能. 基于此,本文方法在早期谣言检测的任务上优于现有算法. 结果表明,该方法具有更强的鲁棒性和更稳定的性能.

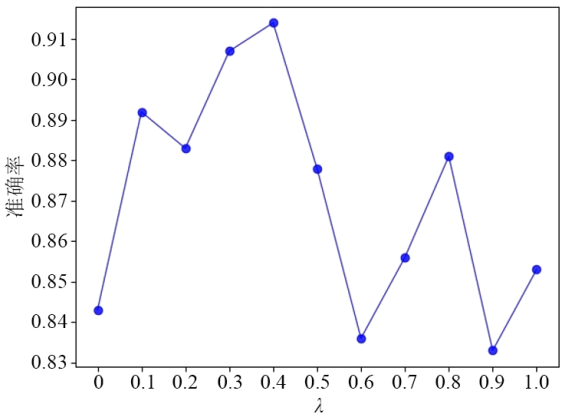


(*a*) Twitter15数据集由不同用户转发数的准确性对比

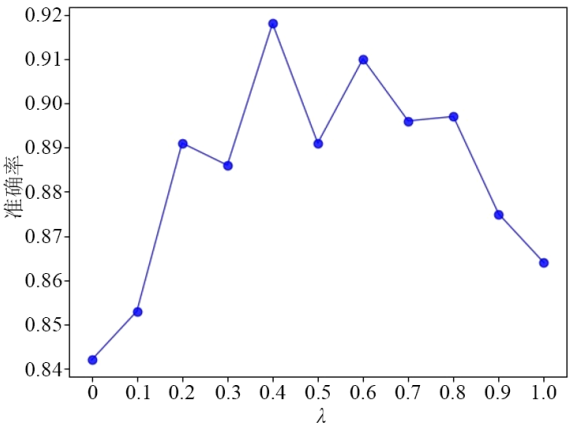


(*b*) Twitter16数据集由不同用户转发数的准确性对比

图4　源推文发布由不同用户转发数的准确性对比



(*a*) Twitter15数据集的准确率变化



(*b*) Twitter16数据集的准确率变化

图5　超参数λ对最终结果的影响

5.7　超参数λ的影响

超参数λ控制着各个模块之间的权衡,λ的最优值对于不同的训练任务也是不同的. 图5显示了在Twitter15和Twitter16数据集下测试,不同的λ对最终结果的影响,选用的延迟时间是36 h,转发用户数量500.

由图5可知,模型精度随λ的增加呈现先增加后减少的趋势,结合式(19)可知,图结构的应用能在模型的特征学习方面起到一定的作用. 在两个数据集上均显示λ=0.4模型达到最佳效果,说明对于原输入的内容信息和图结构在模型训练时均起到了一定的效果. 其中,图结构所起的效果更大,由此也能说明文本全局-局部上下文语义关系和推文传播结构特征对最终结果影响更大.

5.8　异质图构建质量的影响

本文所提方法均基于所构建的推文-词-用户异质图,所以对异质图的构建至关重要. 为了验证构建质量对最终性能的影响,在本节中,分别改变边权重,进行4组对比实验:

(1)将词-词边权值设为1;

(2)将推文-词边权值设为1;

(3)将推文-用户边权值设为1;

(4)将所有边权值设为1,即将带权图变为无权图.

从表4和表5(实验测试结果中加粗数据表示最优结果)可以看出,本文方法所构建的推文-词-用户异质图较4种情况而言,在Twitter15数据集上分别高出7.2%,12.1%,10.2%,36.4%,在Twitter16数据集上分别高出4.7%,11.5%,9.3%,39.2%. 实验结果说明了在边权值设定上,3种边权值设定的合理性. 其中,词-词边权值由逐点互信息(Pointwise Mutual Information,PMI)计算,反应了单词之间的上下文语义关系;推文-词边权值由词频逆文档频率(TF-IDF)计算,反应了推文与推文之间的全局语义关系;推文-用户边权值由用户转发或回复源推文的时间的倒数计算,反应了推文的传播关系. 每一种边权值均带有一定特征,而主模型每一部

分均优于变体,说明所提模型中充分利用到全局-局部上下文语义特征和传播结构特征,从而使性能提升.

表4 Twitter15数据集构图方式对比分析

方法	准确率	NR F_1值	FR F_1值	TR F_1值	UR F_1值
词-词边权值为1	0.842	0.755	0.809	0.938	0.857
推文-词边权值为1	0.793	0.813	0.751	0.790	0.818
推文-用户边权值为1	0.812	0.721	0.801	0.912	0.814
所有边权值为1	0.550	0.571	0.302	0.680	0.561
HGBGAN	**0.914**	**0.928**	**0.907**	**0.935**	**0.885**

表5 Twitter16数据集构图对比分析

方法	准确率	NR F_1值	FR F_1值	TR F_1值	UR F_1值
词-词边权值为1	0.872	0.853	0.912	0.902	0.815
推文-词边权值为1	0.804	0.791	0.836	0.854	0.735
推文-用户边权值为1	0.826	0.721	0.891	0.817	0.875
所有边权值为1	0.527	0.550	0.297	0.711	0.550
HGBGAN	**0.919**	**0.868**	**0.936**	**0.948**	**0.915**

5.9 消融实验

5.9.1 子图模块的重要性

为了验证两个子图模块对于最终模型性能的重要性和影响,在本节中,分别单独利用Bert-GCN模块和子图注意力网络模块,学习子图节点的表示用于谣言检测,并分别对两个数据集进行验证. 两个对比实验分别如下.

(1)单Bert-GCN模块:仅对推文内容进行建模,学习推文表示用于谣言检测.

(2)单子图注意力网络模块:利用子图注意网络对推文-词和推文-用户子图建模,通过子图级注意力机制后学习推文表示. 实验结果如表6和表7所示.

从表6和表7(实验结果中,加粗数据表示最优结果)可以看出:总体上,整个模型框架而言,包含文本全局-局部上下文语义关系和推文传播结构特征的子图注意力网络模块较比包含推文语义内容特征的Bert-GCN模块在对谣言检测效果要更好. 具体而言,单子图注意力网络模块在Twtter15和Twitter16数据集上测试精度分别较单Bert-GCN模块高出了4%和2.2%. 这一结果表明,谣言的文本全局-局部上下文语义关系和推文传播结构特征对谣言检测更为重要. 这是因为现在社交网络造谣者都会通过各种手段使谣言本身具有伪装性和隐蔽性,所以单从文本内容一种特征来分辨,难以判断该推文是否为谣言. 必须结合谣言的文本全局-局部上下文语义关系和推文传播结构特征判断才能更有效.

而对于非谣言类而言,在Twitter15数据集上单子图注意力网络模块精度高于主模型. 这是因为很多情况下,非谣言更容易得到可信度较高的用户的回应,其传播路径也相对固定. 谣言散布者可以改变推文内容和自身的转发行为,但无法改变网络中正常用户的连接关系,而正常用户是占大多数的. 所以在非谣言类,更容易通过只学习到传播结构特征的单子图注意力网络模块来区分非谣言类.

表6 Twitter15数据集模块对比分析

方法	准确率	NR F_1值	FR F_1值	TR F_1值	UR F_1值
单Bert-GCN模块	0.843	0.760	0.804	0.936	0.867
单子图注意力网络模块	0.883	**0.951**	0.886	0.876	0.820
HGBGAN	**0.914**	**0.928**	**0.907**	**0.935**	**0.885**

表7 Twitter16数据集模块对比分析

方法	准确率	NR F_1值	FR F_1值	TR F_1值	UR F_1值
单Bert-GCN模块	0.842	0.800	0.860	0.927	0.779
单子图注意力网络模块	0.864	0.826	0.818	0.937	0.869
HGBGAN	**0.919**	**0.868**	**0.936**	**0.948**	**0.915**

5.9.2 异质图分解的重要性

为了验证异质图分解对最终模型性能的重要性和具体影响,本节将设置异质图不分解,直接输入到GAT网络中以学习到节点特征的对比实验,与原实验进行性能对比,同时分别对两个数据集进行验证. 两个对比实验分别如表8和表9所示(加粗数据表示最优结果).

表8 Twitter15数据集异质图处理对比分析

方法	准确率	NR F_1值	FR F_1值	TR F_1值	UR F_1值
异质图不分解	0.845	0.787	0.849	0.915	0.819
异质图分解	**0.914**	**0.928**	**0.907**	**0.935**	**0.885**

表9 Twitter16数据集异质图处理对比分析

方法	准确率	NR F_1值	FR F_1值	TR F_1值	UR F_1值
异质图不分解	0.821	0.735	0.867	0.917	0.764
异质图分解	**0.919**	**0.868**	**0.936**	**0.948**	**0.915**

从表8和表9的实验结果看出,异质图分解的精度在Twitter15和Twitter16数据集上测试结果分别高出异质图不分解时7.4%和9.8%,说明异质图分解的效果要优于异质图不分解的效果,这也充分证明了异质图分解的必要性. 同时,证明了子图注意力机制的必要性. 在分解中推文-词子图学习到的是文本全局-局部上下文特征,而推文-用户子图学习到的是文本传播结构特征,将异质图分解后再通过子图注意力机制模块结合,可以学习到两个特征的权重,让模型更加聚焦重要特征. 同时,将异质图分解后输入到GAT网络中. 由于图

节点和边的减少,GAT处理分解后异质图的复杂度要远远低于分解前,从而大幅节约了训练成本.

6 结论

本文基于推文内容、转发推文和用户配置提出了一种基于Bert-GNNs异质图注意力网络的早期谣言检测算法. 该算法学习文本内容的全局-局部上下文语义关系和文本语义内容特征,并将其与源推文传播相关信息有效整合,用于谣言检测. 该方法在构建的文本-词-用户异质图基础上,融合Bert-GCN模块和子图注意力网络模块,利用大量原始数据的大规模预训练学习文本的语义内容信息,并通过图神经网络来学习文本之间的全局-局部上下文语义关系和传播图的全局结构关系,两者共同训练得出图嵌入表示,以尽可能利用现有信息挖掘特征. 在两个真实的Twitter15和Twitter16数据集上的实验表明,本文所提算法在准确性上比现有方法有更好的性能,也具备在早期阶段对谣言进行检测的能力,并验证了谣言的文本全局-局部上下文语义关系和推文传播结构特征较文本语义内容特征对谣言检测更为重要.

在未来的工作中,将尝试将用户的社会关系整合到异构的推文-词-用户异质图中,以期进一步提高早期谣言检测的性能.

参考文献

[1] 刘树新, 季新生, 刘彩霞, 等. 一种信息传播促进网络增长的网络演化模型[J]. 物理学报, 2014, 63(15): 429-439.
LIU S X, JI X S, LIU C X, et al. A complex network evolution model for network growth promoted by information transmission[J]. Acta Physica Sinica, 2014, 63(15): 429-439. (in Chinese)

[2] LIANG G, HE W B, XU C, et al. Rumor identification in microblogging systems based on users' behavior[J]. IEEE Transactions on Computational Social Systems, 2015, 2 (3): 99-108.

[3] GRINBERG N, JOSEPH K, FRIEDLAND L, et al. Fake news on Twitter during the 2016 US presidential election [J]. Science, 2019, 363(6425): 374-378.

[4] POPAT K. Assessing the credibility of claims on the web [C]//Proceedings of the 26th International Conference on World Wide Web Companion. New York: ACM, 2017: 735-739.

[5] 程晓涛, 刘彩霞, 刘树新. 基于关系图特征的微博水军发现方法[J]. 自动化学报, 2015, 41(9): 1533-1541.
CHENG X T, LIU C X, LIU S X. Graph-based features for identifying spammers in microblog networks[J]. Acta Automatica Sinica, 2015, 41(9): 1533-1541. (in Chinese)

[6] 袁得嵛, 章逸钒, 高见, 等. 基于用户特征提取的新浪微博异常用户检测方法[J]. 计算机科学, 2020, 47(S1): 364-368, 385.
YUAN D Y, ZHANG Y F, GAO J, et al. Abnormal user detection method in sina weibo based on user feature extraction[J]. Computer Science, 2020, 47(S1): 364-368, 385. (in Chinese)

[7] JIN F, DOUGHERTY E, SARAF P, et al. Epidemiological modeling of news and rumors on Twitter[C]//Proceedings of the 7th Workshop on Social Network Mining and Analysis. New York: ACM, 2013: 1-9.

[8] SAMPSON J, MORSTATTER F, WU L, et al. Leveraging the implicit structure within social media for emergent rumor detection[C]//Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. New York: ACM, 2016: 2377-2382.

[9] 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 (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2017: 708-717.

[10] ZHAO Z, RESNICK P, MEI Q Z. Enquiring minds: Early detection of rumors in social media from enquiry posts [C]//Proceedings of the 24th International Conference on World Wide Web. New York: ACM, 2015: 1395-1405.

[11] FARNAAZ N, JABBAR M A. Random forest modeling for network intrusion detection system[J]. Procedia Computer Science, 2016, 89: 213-217.

[12] 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 (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2018: 1980-1989.

[13] LI Q Z, ZHANG Q, SI L. Rumor detection by exploiting user credibility information, attention and multi-task learning[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 1173-1179.

[14] 蔡国永, 林强, 任凯琪. 基于域对抗网络和BERT的跨领域文本情感分析[J]. 山东大学学报(工学版), 2020, 50 (1): 1-7, 20.
CAI G Y, LIN Q, REN K Q. Cross-domain text sentiment

classification based on domain-adversarial network and BERT[J]. Journal of Shandong University (Engineering Science), 2020, 50(1): 1-7, 20. (in Chinese)

[15] MADHAV KOTTETI C M, DONG X S, QIAN L J. Multiple time-series data analysis for rumor detection on social media[C]//2018 IEEE International Conference on Big Data (Big Data). Piscataway: IEEE, 2019: 4413-4419.

[16] 马鸣, 刘云, 刘地军, 等. 基于主题和预防模型的微博谣言检测[J]. 北京理工大学学报, 2020, 40(3): 310-315.
MA M, LIU Y, LIU D J, et al. Rumor detection in microblogs based on topic and prevention model[J]. Transactions of Beijing Institute of Technology, 2020, 40(3): 310-315. (in Chinese)

[17] 贾硕, 张宁, 沈洪洲. 网络谣言传播与消解的研究进展[J]. 信息资源管理学报, 2019, 9(3): 62-72.
JIA S, ZHANG N, SHEN H Z. Overviews on information dissemination and disappearance for online rumor[J]. Journal of Information Resources Management, 2019, 9(3): 62-72. (in Chinese)

[18] 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.
GAO Y J, LIANG G, JIANG F T, et al. Social network rumor detection: A survey[J]. Acta Electronica Sinica, 2020, 48(7): 1421-1435. (in Chinese)

[19] 任文静, 秦兵, 刘挺. 基于时间序列网络的谣言检测研究[J]. 智能计算机与应用, 2019, 9(3): 300-303.
REN W J, QIN B, LIU T. Rumor detection based on time series model[J]. Intelligent Computer and Applications, 2019, 9(3): 300-303. (in Chinese)

[20] TORSHIZI A S, GHAZIKHANI A. Automatic Twitter rumor detection based on LSTM classifier[C]//Communications in Computer and Information Science. Cham: Springer, 2019: 291-300.

[21] ZHOU Z Y, QI Y, LIU Z, et al. A C-GRU neural network for rumors detection[C]//2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS). Piscataway: IEEE, 2019: 704-708.

[22] JIN Z W, CAO J, GUO H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]//Proceedings of the 25th ACM international conference on Multimedia. New York: ACM, 2017: 795-816.

[23] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: ACM, 2016: 3818-3824.

[24] YU F, LIU Q, WU S, et al. A convolutional approach for misinformation identification[C]//5th International Conference on Learning Representations (ICLR). Toulon: OpenReview, 2017: 3901-3907.

[25] IANCU B, RUIZ L, RIBEIRO A, et al. Graph-adaptive activation functions for graph neural networks[C]//2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP). Piscataway: IEEE, 2020: 1-6.

[26] BIAN T A, XIAO X, XU T Y, et al. Rumor detection on social media with bi-directional graph convolutional networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 549-556.

[27] YUAN C Y, MA Q W, 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). Piscataway: IEEE, 2020: 796-805.

[28] HUANG Q, YU J S, WU J, et al. Heterogeneous graph attention networks for early detection of rumors on twitter [C]//2020 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2020: 1-8.

[29] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C]//Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 675-684.

[30] QAZVINIAN V, ROSENGREN E, RADEV D R, et al. Rumor has it: Identifying misinformation in microblogs [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. New York: ACM, 2011: 1589-1599.

[31] YANG F, LIU Y, YU X H, et al. Automatic detection of rumor on Sina Weibo[C]//Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. New York: ACM, 2012: 1-7.

[32] MA J, GAO W, WEI Z Y, 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. New York: ACM, 2015: 1751-1754.

[33] 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. Piscataway: IEEE, 2014: 1103-1108.

[34] CHEN T, LI X, YIN H Z, et al. Call attention to rumors: Deep attention based recurrent neural networks for early

rumor detection[C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer, 2018: 40-52.

[35] VAIBHAV V, MANDYAM R, HOVY E. Do sentence interactions matter? Leveraging sentence level representations for fake news classification[C]//Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13). Stroudsburg: Association for Computational Linguistics, 2019: 134-139.

[36] SHARMA S, SHARMA R. Identifying possible rumor spreaders on twitter: A weak supervised learning approach [C]//2021 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2021: 1-8.

[37] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[C]//5th International Conference on Learning Representations (ICLR). Toulon: OpenReview, 2017: 3294-3302.

[38] SONG W P, XIAO Z P, WANG Y F, et al. Session-based social recommendation via dynamic graph attention networks[C]//Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. New York: ACM, 2019: 555-563.

[39] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[40] MASETTI G, DI GIANDOMENICO F. Analyzing forward robustness of feedforward deep neural networks with LeakyReLU activation function through symbolic propagation[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer, 2020: 460-474.

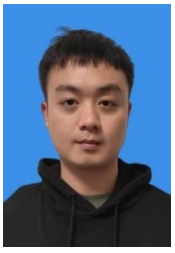
[41] LIU Y, WU Y F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 354-361.

[42] WEI L W, 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). Stroudsburg: Association for Computational Linguistics, 2021: 3845-3854.

[43] WANG G, TAN L, SONG T B, et al. Region-enhanced deep graph convolutional networks for rumor detection [EB/OL]. (2022-06-15)[2022-07-25]. https://arxiv.org/abs/2206.07665.

[44] WANG H D, BAI C Z, YAO J L. Federated graph attention network for rumor detection[EB/OL]. (2022-06-12)[2022-07-25]. https://arxiv.org/abs/2206.05713.

作者简介



欧阳祺　男,1999年生,江西吉安人. 中国人民解放军战略支援部队信息工程大学研究生. 主要研究方向为复杂网络、数据挖掘.
E-mail: oyq126126@126.com

陈鸿昶　男,1964生,河南新密人. 中国人民解放军战略支援部队信息工程大学博士研究生教授、博士生导师. 主要研究方向为未来网络体系结构、人工智能等.
E-mail: chenhongchang@ndsc.com.cn

刘树新　男,1987年生,山东潍坊人. 中国人民解放军战略支援部队信息工程大学助理研究员. 研究方向为复杂网络、移动通信网安全.
E-mail: liushuxin11@gmail.com

王　凯　男,1980年生,河南许昌人. 国家数字交换系统工程技术研究中心副研究员. 主要研究方向为电信网信息关防.
E-mail: wangkai@ndsc.com.cn

李　星　男,1981年生,河南新乡人. 博士. 中国人民解放军战略支援部队信息工程大学助理研究员. 主要研究方向为链路预测、社团挖掘.
E-mail: lixing@ndsc.com.cn