计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2021.20200801
58(7): 1412-1424, 2021

融合源信息和门控图神经网络的谣言检测研究

杨延杰　王　莉　王宇航
(太原理工大学大数据学院　山西晋中　030600)
(yangyanjie1073@link.tyut.edu.cn)

Rumor Detection Based on Source Information and Gating Graph Neural Network

Yang Yanjie, Wang Li, and Wang Yuhang
(*College of Big Data, Taiyuan University of Technology, Jinzhong, Shanxi* 030600)

Abstract　Social media not only brings convenience to people, but also provides a platform for spreading rumors. Currently, most rumor detection methods are based on text content information. However, in social media scenarios, text content is mostly short text, which often leads to poor performance due to data sparsity. Message propagation on social networks can be modeled as a graph structure. Previous studies have taken into account the characteristics of message propagation structure, and detected rumors through GCN. GCN aggregates neighbors based on structural information to enhance node representation, but some neighbor aggregation is useless and may even cause noise, which making the representation obtained from GCN unreliable. Meanwhile, these methods can not effectively highlight the importance of the source post information. In this paper, we propose a propagation graph convolution network model GUCNH. In GUCNH model, information forwarding graph is constructed first, and the representation of neighbor nodes is aggregated by two fusion gated convolution network modules. Fusion gating can select and combine the feature representation before and after the graph convolution to get a more reliable representation. Considering that in forwarding graph, any post may interact with each other rather than just with its neighbors, a multi-headed self-attention module is introduced between two integrated gated convolution network modules to model the multi-angle influence between posts. In addition, in forwarding graph, the source posts often contain the richest information than reposts. After generating each node representation, we selectively enhance the source node's information to enhance the influence of the source posts. Experiments on three real datasets show that our proposed model outperforms the existing methods.

Key words　rumor detection; communication structure; graph convolution network with gating; multi-head attention; source information enhancement

摘　要　社交媒体在带给人们便利同时,也为谣言的发布和传播提供了平台.目前,大多数的谣言检测方法都是基于文本内容信息,但在社交媒体场景下,文本内容大多是短文本,这类方法往往会因为数据稀疏性的问题导致性能下降.社交网络上的消息传播可建模为图结构,已有研究考虑消息传播结构特点,通过GCN等模型进行谣言检测.GCN依据结构信息聚合邻居来提升节点表示,但有些邻居聚合是无用的,甚至可能带来噪声,使得通过GCN得到的表示并不可靠.此外,这些研究不能有效的突出源帖信息

收稿日期:2019-10-08;**修回日期**:2020-12-11
基金项目:国家自然科学基金项目(61872260)
This work was supported by the National Natural Science Foundation of China (61872260).
通信作者:王莉(wangli@tyut.edu.cn)

的重要性.针对这些问题提出了一种融合门控的传播图卷积网络模型 GUCNH,在 GUCNH 模型中,首先利用消息转发关系构建信息转发图,通过 2 个融合门控的图卷积网络模块来聚合邻居节点信息生成节点的表示,融合门控能够对图卷积之前的特征表示和之后的特征表示进行选择与组合,以得到更加可靠的表示.考虑到在转发图中,任意的帖子之间都可能存在相互影响,而不仅仅是基于邻接关系,因此在 2 个融合门控的图卷积网络模块之间引入多头自注意力模块来建模任意帖子之间的多角度影响.此外,在转发图中,源帖包含的信息往往是最原始、最丰富的,在生成各节点表示之后,选择性的增强了源节点的信息以增强根源信息的影响力.在 3 个真实数据集上进行的实验表明,提出的模型优于现有的方法.

关键词 谣言检测;传播结构;融合门控的图卷积网络;多头注意力;源信息增强

中图法分类号 TP18

随着互联网的飞速发展,社交媒体已经成为用户获取信息、交流意见的主要平台,根据 Kantar Media 在 2019 年发布的一份报告,全球 40%的人使用社交媒体[1],而且这一数字还在不断地增加,这就极大地促进了谣言的快速滋生和广泛传播,对社会稳定造成巨大的威胁.例如据 BuzzFeed News 报道[2],在 2016 年美国总统大选期间,谣言的传播在网络上造成了不小的负面影响.2020 年 COVID-19 疫情爆发期间,有些人在社交平台上散布一些有关疫情传播的谣言,引发了人们的不安.谣言的迅速传播,已经开始从各个方面影响人们的正常生活,因此,谣言检测是一个亟待解决的关键问题.

然而,谣言检测是一项非常有挑战性的任务,主要体现为 3 个方面:1)谣言具有强迷惑性和误导性,使得单独从谣言文本内容本身检测谣言存在困难.因此除了从谣言本身的内容信息出发,我们还应该探索和利用其他信息,如社交媒体上的用户信息以及社会上下文信息.2)早期检测的需求.社交媒体上的用户较为活跃,使得谣言能够在短时间内广泛传播,谣言造成的负面影响随之剧增,使得早期检测尤为重要.3)谣言的传播过程复杂多样[3],数据流动没有固定的规律,谣言内容涵盖的方面非常大,使得数据的处理和使用成为一大困难.

为了有效检测谣言,人们已经做了大量的研究,常见的方法利用文本内容进行谣言检测,研究人员从文本内容中提取一些低级特征如 n-gram,TF-IDF,bag-of-word[4-6]和一些高级的特征如文体特征、事实主观性、写作风格一致性[6-8]等,然后将这些特征应用于机器学习算法进行谣言检测.这些方法基于手工构建的特征,特征提取类别较为单一,无法很好的应对复杂多变的真实环境.深度学习不依赖于手工特征的构建,而且还能提取得到高层次的特征表示.近年来,研究者开始利用深度学习方法建模文本语言[9-11]、文本结构[12-14]等,取得了非常好的效果.这一类方法需要较长的文本才能够训练得到好的特征表示以提高检测效果.但是社交媒体上,人们发表见解的帖子通常是较短的文本[15],这就可能影响基于内容的方法的检测性能.此外,还有方法利用参与社交媒体的用户信息来检测谣言[16-17],这些方法受到现实场景的限制,出于隐私考虑,用户的真实信息往往难以获得.研究者们开始关注于利用社交网络上的传播信息进行谣言检测,一些研究利用传播路径构建传播树,然后利用长短期记忆(long short-term memory, LSTM)网络、门控递归单元(gated recurrent unit, GRU)来学习传播过程中的序列特征[18-19],但是传播的序列特征无法反映传播内部的结构信息,此类方法有一定的局限.图卷积网络(graph convolutional network, GCN)[12]的诞生,为我们提供了很好的思路,最近的一些研究使用 GCN 解决谣言检测问题[20-21]并取得了较好的效果.

受上述研究启发,社交媒体上的消息转发可以建模为图结构,图 1(a)展示了来自公共数据集 FakeNewsNet 的一条"凯瑟琳生下第 3 个孩子后 5 个小时就出现在伦敦一家医院外"① 的谣言以及它的转发路径,根据图 1(a)的转发关系可以得到如图 1(b)所示的转发图.消息转发图中某一帖子的上游信息和下游信息对于研究当前帖子都非常重要,我们认为这样的转发图中蕴含着丰富的结构关系可以为谣言检测提供帮助.另外,转发过程是一种信息逐步扩展的过程,源帖表达出最原始且最重要的信息,更好地利用源帖的信息对于谣言检测至关重要.

本文主要研究:1)如何有效地利用转发图来整合复杂的转发结构信息用于分类;2)如何更好地利用源帖的信息以提高谣言检测的性能.为了解决这 2 个

① https://twitter.com/CNN/status/988463960159608833

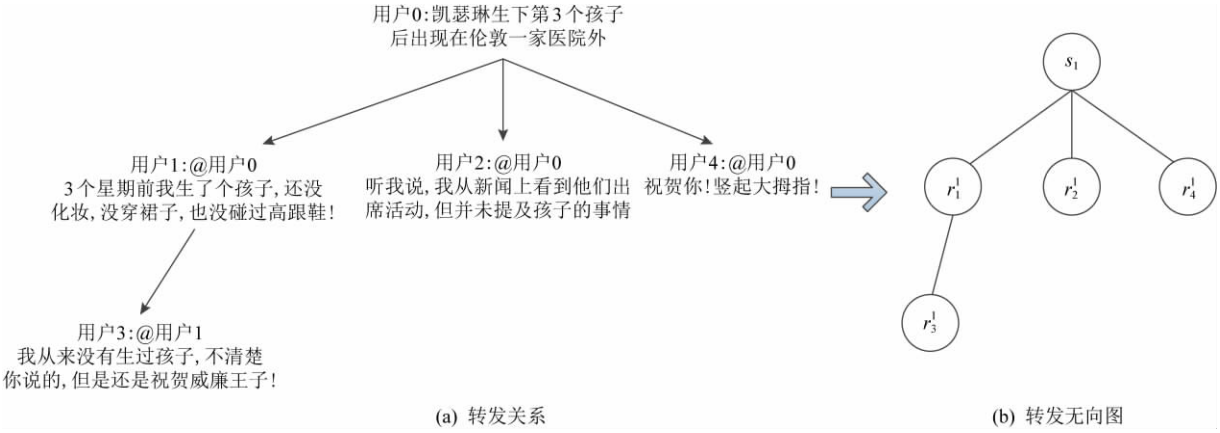

Fig. 1 Construction of forwarding graph in social media environment

图 1 社交媒体场景下的转发图的构造方法

问题,提出了一种谣言检测模型 GUCNH.首先,我们利用社交网络中帖子的转发关系构造转发图,然后提出了一种融合门控的图卷积网络模块用于捕获转发图中的各节点之间的结构信息,融合门控的目的是对图卷积之前的特征表示和之后的特征表示进行选择与组合,以得到更加可靠的表示.为了更好地利用源帖信息,我们在源帖对应节点的原始表征和通过融合门控的图卷积网络模块之后得到的表示之间进行选择与组合,将选择后的结果与每个节点的表征拼接.最后将所有节点表征取平均用于分类.本文工作的主要贡献可以概括为 3 个方面:

1) 提出了一种融合门控的图卷积网络模块 GUCN,该模块通过门控单元来对图卷积之前的特征表示和之后的特征表示进行选择与组合,以得到更加可靠的表示.通过该模块来捕获转发图节点之间的结构关系,并结合多头自注意力模块来考虑任意节点之间可能存在的影响,最终生成节点表示.

2) 源贴信息往往最为重要,为了充分利用源贴信息,在生成节点表示之后,模型将经过选择的源贴特征表示与转发图中生成的所有节点表示拼接起来,以加强源帖的重要性.

3) 在 3 个真实的数据集进行了一系列的实验.实验结果表明:本模型在谣言分类和早期检测任务方面都取得了优于现有模型的结果.

1 相关工作

谣言检测的目标是根据用户发布在社交媒体平台上的相关信息(如文本内容、用户配置文件、评论、传播模式等)来检测谣言的真假.根据研究对象的不同,相关工作可以大致的分为 3 类:1)基于内容的方法;2)基于用户的方法;3)基于传播的方法.

1) 基于内容的方法.基于内容的方法主要依赖于文本的内容信息来检测谣言,这些研究通常面向于长文本数据.一部分研究者从机器学习的角度进行谣言检测,Pérez-Rosas 等人[22]从新闻中提取手工特征建立组合特征集训练线性支持向量机 SVM 模型用于谣言检测;Popat 等人[7]通过研究文本内容的语言风格来进行谣言检测;Takahashi 等人[23]通过应用命名实体和线索关键字来训练分类器进行谣言检测,这类方法均基于机器学习,需要人工设计特征并进行提取,在通用性和扩展性上存在一定的缺陷.近年来,深度学习的发展为谣言检测提供了很多新的方法,Ma 等人[18]利用递归神经网络(recurrent neural network, RNN)从文本内容中提取隐藏的向量表示用于分类;Ahn 等人[10]将预训练的 BERT 模型用于谣言检测任务,取得了非常好的效果;Vaibhav 等人[13]提出了一种用于虚假新闻检测的图神经网络模型,该模型对新闻中所有句子对之间的语义关系进行建模,从而进行谣言检测;Wang 等人[14]依赖文本内容,提出了 SemSeq4FD 模型来检测虚假信息,该模型同时考虑了新闻中句子之间的全局语义关系和局部上下文顺序特征,取得了很好的效果.本节介绍的基于文本内容的方法局限性是它们更适用于长文本,基于机器学习的方法需要长文本才能提取到所需要的特征进行分类,基于深度学习的方法也需要较长的文本才能够训练得到好的特征表示以提高检测效果,而社交媒体上的帖子大多是短文本,造成数据稀疏问题从而影响该类方法的检测性能.

2) 基于用户的方法.基于用户的方法主要针对参与社交媒体的用户进行建模.其中用户的特征信息是从用户配置文件中收集的,如描述、性别、关注者、朋友、位置和验证类型等.Yang 等人[16]提取用户特征进行分类,如性别、地理位置和追随者数量;Castillo 等人[4]利用 Twitter 上的用户特征来检测假新闻,这些特征包括关注者数量、好友数量、注册年龄等;Shu 等人[24]充分研究了用户配置文件在虚假信息检测中的作用,他们的工作为深入探索社交媒体的用户特征提供了基础;Liu 等人[17]结合 RNN 和卷积神经网络(convolutional neural network, CNN)来捕获基于时间序列的用户特征;Lu 等人[20]将参与社交的所有用户构建为一个完全连通的图以辅助检测谣言.这类方法的局限性主要表现在由于隐私问题,许多用户会隐藏自己的信息或使用虚假的个人信息,这使得获取真实的用户信息变得非常困难.

3) 基于传播的方法.与基于内容和基于用户的 2 种方法不同,基于传播的方法主要侧重于真假信息传播特征的差异,现有的研究根据建模类型的不同主要可以分为 3 种:基于传播链的方法、基于传播树的方法、基于传播图的方法.①基于传播链的方法主要将信息传播按照时间顺序看为一个时间链来检测谣言.Kwon 等人[25]确定了真假新闻在传播中存在语言差异,从时间、内容等方面分析了谣言的传播特征,并根据这些特征,利用决策树、随机森林和支持向量机来检测谣言;Ma 等人[26]提出了一系列基于谣言生命周期的时间序列特征,将这些特征用于分类,一定程度上提高了谣言的检测效果.②基于传播树的方法主要将信息的传播建模为一棵消息传播树,通过对消息传播树中的传播链进行一系列操作以检测谣言.Wu 等人[27]提出了一种随机游走的核来建模消息的传播树,以提高谣言的检测能力;Ma 等人[19]建立了树结构递归神经网络(RvNN),从传播结构和文本内容中捕捉各节点的隐藏表示,取得了不错的效果.然而,这些方法通常只关注于从传播树上学习序列化特征,忽略了社交网络上帖子之间的全局转发关系.③最近的一些研究将信息的传播建模为一个传播图,利用图神经网络技术解决谣言检测问题,Wei 等人[28]针对谣言检测问题,提出了一种多深度 M-GCN 模型,该模型能够捕获多尺度的邻居信息;Wu 等人[29]对于传播图迭代的使用图神经网络直到收敛,将收敛之后的节点表示用于分类;最近,Bian 等人[21]提出了一种用于谣言检测的双向 BiGCN 模型.通过双向图卷积网络学习消息转发的结构特征,取得了良好的效果.这些现有的基于传播图的方法虽然已经开始注意使用消息传播结构信息,但是他们过分依赖于 GNN,GCN 等单一模型的处理结果,同时源帖子的重要性并没有得到充分利用.

本文的研究主要是根据文本内容和转发结构进行谣言检测,与本研究最相关的是基于文本内容的方法和基于传播的方法.本文工作的贡献在于:考虑到帖子之间的转发结构信息、融合门控单元和图卷积网络进行建模、充分利用源帖的信息.

2 问题定义

设 $P=\{p_1,p_2,\cdots,p_m\}$ 是一个谣言检测数据集,其中 p_i 为第 i 个事件的消息集,m 为数据集中事件的个数,$p_i=\{s_i,r_1^i,r_2^i,\cdots,r_{k_i-1}^i,G_i\}$,其中 k_i 为事件 i 中的帖子个数,s_i 是事件 i 的源帖,r_j^i 是源帖 s_i 的第 j 个转发,$G_i=\langle V_i,E_i\rangle$ 是根据 s_i 的转发关系构建的转发图.当前谣言检测任务可分为 2 种类别.根据文献[30],可以分为非谣言(N)、真谣言(T)、经证实的非谣言(F)、未经证实的谣言(U).根据文献[18],可以分为谣言(T)和非谣言(F).本文根据数据集的不同使用了不同的问题定义.谣言检测任务可以描述为学习一个函数 $f:f(p_i)\rightarrow y_i$,其中标签值 $y_i\in\{0,1,2,3\}$(四分类)或者 $y_i\in\{0,1\}$(二分类).

3 模 型

本文提出一种谣言检测模型——GUCNH,如图 2 所示,主要分为 4 个模块:转发图构建、节点表示、选择性增强根节点表示、谣言分类.

3.1 构建转发图

根据事件 i 的转发关系构建转发图 $G_i=\langle V_i,E_i\rangle$,$G_i$ 是无向图.原因是图神经网络通过聚合邻居节点进行表征,消息转发中上游信息和下游信息对于节点的表征作用是相同的,同时,建模为无向图可降低数据稀疏程度.$V_i=\{s_i,r_1^i,r_2^i,\cdots,r_{k_i-1}^i\}$ 表示源帖 s_i 的节点和它对应的 k_i-1 个转发节点.$E_i=\{e_{uv}\mid u=0,\cdots,k_i-1;\ v=0,\cdots,k_i-1\}$ 表示转发图中的所有边集.例如,图 1(b)所示为事件消息集 p_1 对应的传播图,在图中,r_3^1 转发了 r_1^1,边集合 E_i 中则包含 e_{13} 和 e_{31};r_1^1 转发了源帖 s_1,边集合 E_i 中就

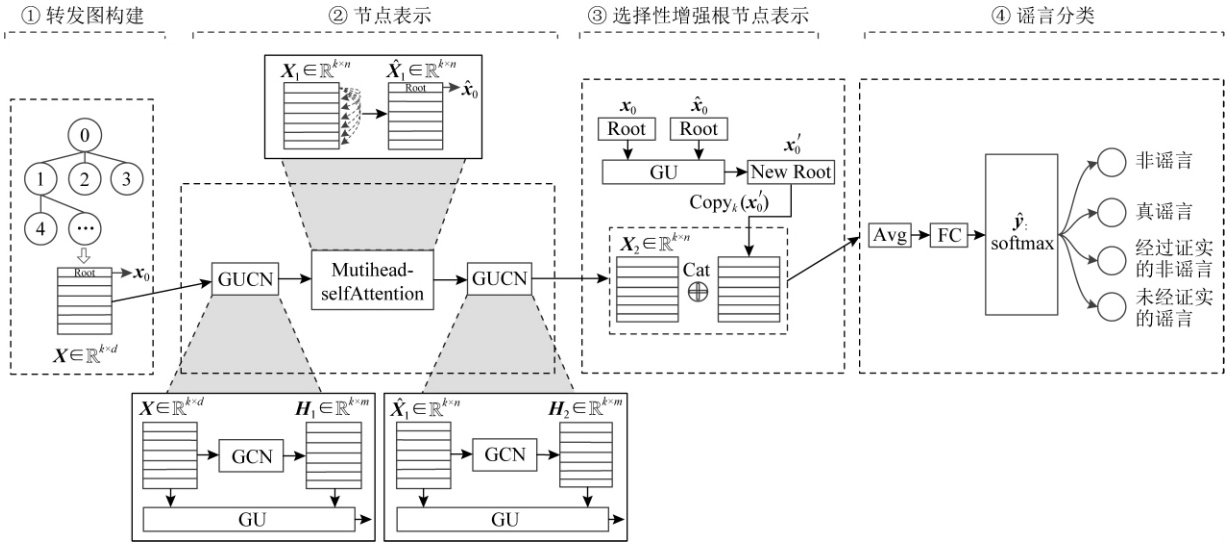

Fig. 2 Four modules in GUCNH model

图 2 GUCNH 模型的 4 个模块

包含 e_{01} 和 e_{10}.设 $\boldsymbol{A}_i \in \{0,1\}^{k_i \times k_i}$ 是邻接矩阵,其元素为

$$a_{uv}^i = \begin{cases} 1, & \text{若 } e_{uv}^i \in E_i, \\ 0, & \text{其他.} \end{cases} \tag{1}$$

借鉴 Bian 等人[21]的方法,本文引入了一种 DropEdge[31]的方法以减少 GCN 过拟合,在训练的每个阶段,随机的将输入图中的一部分边去掉,增加了输入数据的随机性和多样性,能够有效地防止过拟合.本文模型中,随机删除边的比率设定为 q,通过 DropEdge 之后,邻接矩阵变为

$$\boldsymbol{A}_i = \boldsymbol{A}_i - \boldsymbol{A}_i^{\text{drop}}, \tag{2}$$

其中,$\boldsymbol{A}_i^{\text{drop}}$ 为根据原邻接矩阵和比率 q 生成的删除边后的邻接矩阵.设 $\boldsymbol{X}_i = [\boldsymbol{x}_0^i, \boldsymbol{x}_1^i, \cdots, \boldsymbol{x}_{k_i-1}^i]^{\mathrm{T}} \in \mathbb{R}^{k_i \times d}$ 是事件 i 对应的特征矩阵,其中 $\boldsymbol{x}_0^i$ 表示源帖 s_i 的特征向量,$\boldsymbol{x}_j^i$ 表示 s_i 的第 j 个转发 r_j^i 的特征向量,我们利用词袋模型为事件 i 中的所有帖子构建 d 维的特征向量.

3.2 节点表示

构建好转发图之后,通过融合门控的图卷积网络模块 GUCN 和多头自注意力模块来得到包含转发结构信息的节点表示,前者利用图卷积网络聚合一定的邻居信息,融合门控机制来获取更好的中间表示,后者主要通过注意力机制来捕获任意节点之间的多方面影响,具体介绍如下:

1) 融合门控的图卷积网络模块 GUCN

为了充分利用转发图中的转发结构信息,使转发图中的各个节点能很好地融合邻居信息以获得更好的特征表示,引入了融合门控的图卷积网络模块 GUCN,图卷积网络[12]能够依据结构信息对图中的节点进行融合,得到聚合邻居信息后的特征表示.但是 GCN 依靠聚合邻居信息来提升自己的表示,有些聚合可能带来噪声.受文献[32]的启发,本文提出了一种名为 GU 的门控单元,实现从不同的数据组合中找到合适的中间表示.门控单元 GU 的结构如图 3 所示:

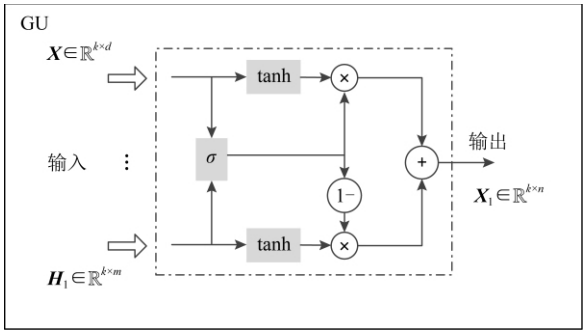

Fig. 3 GU network structure diagram

图 3 GU 网络结构图

为了提高表示的质量,门控单元对图卷积之前的特征表示和之后的特征表示进行选择与组合,最终通过堆叠 GUCN 模块得到融合邻居信息的节点高级特征表示:

$$\boldsymbol{X}_1 = \text{GUCN}(\boldsymbol{X}), \tag{3}$$

$$\boldsymbol{X}_2 = \text{GUCN}(\boldsymbol{X}_1), \tag{4}$$

其中,GUCN(·)为 GUCN 模块运算,$\boldsymbol{X} \in \mathbb{R}^{k \times d}$ 为最初输入的特征矩阵,$\boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathbb{R}^{k \times n}$ 分别为经过 2 次

GUCN 模块运算的特征矩阵.GUCN 模块通过计算得到输出的特征矩阵:

$$\boldsymbol{H}_1=\mathrm{ReLU}(\widetilde{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W}_0),\tag{5}$$

$$\boldsymbol{H}^1=\tanh(\boldsymbol{W}_1\boldsymbol{X}^{\mathrm{T}}),\tag{6}$$

$$\boldsymbol{S}^1=\tanh(\boldsymbol{W}_2\boldsymbol{H}_1^{\mathrm{T}}),\tag{7}$$

$$\boldsymbol{Z}=\sigma(\boldsymbol{W}_3[\boldsymbol{H}^1,\boldsymbol{S}^1]^{\mathrm{T}}),\tag{8}$$

$$\boldsymbol{X}_1=\boldsymbol{Z}\boldsymbol{H}^1+(1-\boldsymbol{Z})\boldsymbol{S}^1,\tag{9}$$

其中,$\widetilde{\boldsymbol{A}}=\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}(\boldsymbol{A}+\boldsymbol{I})\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}$,表示标准化之后的邻接矩阵,$\boldsymbol{A}+\boldsymbol{I}$ 为加上自连接之后的邻接矩阵,$\widetilde{\boldsymbol{D}}$ 为转发图对应的度矩阵,其中 $\widetilde{\boldsymbol{D}}_{ii}=\sum_j\widetilde{\boldsymbol{A}}_{ij}$,$\boldsymbol{X}\in\mathbb{R}^{k\times d}$ 为输入特征矩阵,$\boldsymbol{H}_1\in\mathbb{R}^{k\times m}$ 为经过图卷积操作之后的特征矩阵,ReLU(·)为激活函数,$\sigma(\cdot)$为 sigmoid 激活函数.$\boldsymbol{W}_0\in\mathbb{R}^{d\times m}$,$\boldsymbol{W}_1\in\mathbb{R}^{n\times d}$,$\boldsymbol{W}_2\in\mathbb{R}^{n\times m}$,$\boldsymbol{W}_3\in\mathbb{R}^{n\times 2n}$ 为可学习的参数,$\boldsymbol{H}^1,\boldsymbol{S}^1\in\mathbb{R}^{k\times n}$ 为生成的中间隐藏表示,k 为传播图中节点个数,m 为图卷积网络输出表征的维度,n 为门控网络输出表征的维度.

2) 多头自注意力模块

虽然使用图卷积网络能够考虑到一定的结构信息,让每个节点较好地聚合到邻居信息,但是仅仅通过一次图卷积网络,无法较好地利用远距离非邻居节点的信息,如图 1(b)所示,在实际的社交场景中,评论者 r_3^1 在转发 r_1^1 信息并发表一定的见解时,它很可能已经参考了 r_4^1 的见解,所以 r_3^1 同时受到了 r_1^1 和 r_4^1 的影响,单纯一次 GCN 并不能很好的让 r_3^1 聚合到 r_4^1的信息.自注意力机制可以显式的将对自己影响较大的信息赋予较大的权重且加权到自己的信息中,这样可以极大地丰富节点的表示,而多头则可以尽可能多方面地考虑外部信息的影响.为了将转发图中任意信息之间的影响考虑在内,所以我们显式的引入了多头自注意力模块[33]来捕获任意节点之间的影响,从而使得在进行下一次节点信息融合之前所有节点的信息尽可能全面.多头自注意力模块运算结果通过:

$$\hat{\boldsymbol{A}}_1=\mathrm{MutiHeadAttention}(\boldsymbol{X}_1,\boldsymbol{X}_1,\boldsymbol{X}_1),\tag{10}$$

其中,$\boldsymbol{X}_1\in\mathbb{R}^{k\times n}$ 为输入特征矩阵,$\hat{\boldsymbol{A}}_1=[\hat{\boldsymbol{x}}_0,\hat{\boldsymbol{x}}_1,\cdots,\hat{\boldsymbol{x}}_{k-1}]\in\mathbb{R}^{k\times n}$ 为经过多头自注意力模块后的输出特征矩阵,$\hat{\boldsymbol{x}}_0\in\mathbb{R}^n$ 为源帖对应的特征向量,MutiHeadAttention(·)为多头自注意力模块运算函数,它通过式(11)~(13)得到运算结果:

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V},\tag{11}$$

$$\boldsymbol{Head}_i=\mathrm{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^{\mathrm{Q}},\boldsymbol{K}\boldsymbol{W}_i^{\mathrm{K}},\boldsymbol{V}\boldsymbol{W}_i^{\mathrm{V}}),\tag{12}$$

$$\mathrm{MutiHeadAttention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Concat}(\boldsymbol{Head}_1,\boldsymbol{Head}_2,\cdots,\boldsymbol{Head}_h)\boldsymbol{W}^{\mathrm{O}},\tag{13}$$

其中,$\boldsymbol{Q}\in\mathbb{R}^{k\times n}$,$\boldsymbol{K}\in\mathbb{R}^{k\times n}$,$\boldsymbol{V}\in\mathbb{R}^{k\times n}$分别为 query 矩阵、key 矩阵和 value 矩阵,$\boldsymbol{W}_i^{\mathrm{Q}},\boldsymbol{W}_i^{\mathrm{K}},\boldsymbol{W}_i^{\mathrm{V}}\in\mathbb{R}^{n\times n/h}$,其中 d 是防止分子点积值过大的比例因子,其值为输入表示的维度.h 为多头注意力中头的个数.

将计算得到的特征矩阵 $\hat{\boldsymbol{A}}_1$ 用于下一次的 GUCN 模块,用 $\hat{\boldsymbol{A}}_1$ 替换式(4)中的 $\boldsymbol{X}_1$ 以得到输出矩阵 $\boldsymbol{X}_2$:

$$\boldsymbol{X}_2=\mathrm{GUCN}(\hat{\boldsymbol{A}}_1).\tag{14}$$

3.3 选择性增强根节点表示

谣言事件的源帖总是有着最丰富且重要的原始信息,更好地利用源帖信息能够提高谣言检测的性能.具体的,我们通过 GU 门控单元对源贴的初始向量表示 $\boldsymbol{x}_0$ 和经过多头自注意力模块之后高级向量表示的 $\hat{\boldsymbol{x}}_0$ 进行选择与组合,然后将选择与组合的结果复制后同最后一次 GUCN 模块运算的结果拼接作为最终的输出特征矩阵 $\boldsymbol{X}_{\mathrm{last}}$:

$$\boldsymbol{h}^1=\tanh(\boldsymbol{W}_4\boldsymbol{x}_0^{\mathrm{T}}),\tag{15}$$

$$\boldsymbol{s}^1=\tanh(\boldsymbol{W}_5\hat{\boldsymbol{x}}_0^{\mathrm{T}}),\tag{16}$$

$$\boldsymbol{z}=\sigma(\boldsymbol{W}_6[\boldsymbol{h}^1,\boldsymbol{s}^1]^{\mathrm{T}}),\tag{17}$$

$$\boldsymbol{x}_0'=\boldsymbol{z}\boldsymbol{h}^1+(1-\boldsymbol{z})\boldsymbol{s}^1,\tag{18}$$

$$\boldsymbol{X}_0'=\mathrm{copy}_k(\boldsymbol{x}_0'),\tag{19}$$

$$\boldsymbol{X}_{\mathrm{last}}=\mathrm{concat}(\boldsymbol{X}_2,\boldsymbol{X}_0'),\tag{20}$$

其中,$\boldsymbol{W}_4\in\mathbb{R}^{n\times d}$,$\boldsymbol{W}_5\in\mathbb{R}^{n\times n}$,$\boldsymbol{W}_6\in\mathbb{R}^{n\times 2n}$ 为可学习的参数,$\boldsymbol{h}^1,\boldsymbol{s}^1\in\mathbb{R}^n$ 为中间隐藏向量表示.$\boldsymbol{x}_0'\in\mathbb{R}^n$ 为通过 GU 单元组合之后的向量表示.$\mathrm{copy}_k(\cdot)$为复制函数,作用是将某向量表示复制 k 次,concat(·)为拼接函数,最终可以得到通过选择性增强头节点表示模块的特征表示 $\boldsymbol{X}_{\mathrm{last}}\in\mathbb{R}^{k\times 2n}$.

3.4 谣言分类

本节主要讨论如何使用得到的节点表示 $\boldsymbol{X}_{\mathrm{last}}$进行分类,我们认为基于转发图的谣言检测可以看作是一个图分类任务,所以需要一个单独的向量作为整图的特征表示用于分类.具体的,首先通过选择性增强根节点表示模块得到了转发图中每个节点的表示,然后通过平均这些节点表示得到整个转发图的向量表示,将该向量表示作为全连接神经网络的输入,得到预测结果,计算过程为

$$\hat{\boldsymbol{y}}=\mathrm{softmax}\left(\mathrm{Fc}\left(\frac{1}{k}\sum_{v\in\boldsymbol{V}}\boldsymbol{x}_v\right)\right),\tag{21}$$

其中,Fc(·)为全连接网络,输出维度为分类的类别数 r,$\boldsymbol{x}_v$ 表示 $\boldsymbol{X}_{\mathrm{last}}$中第 v 个节点的向量表示,k 为节点

个数.最后生成的二进制预测向量 $\hat{\boldsymbol{y}}=\{\hat{\boldsymbol{y}}_0,\hat{\boldsymbol{y}}_1,\hat{\boldsymbol{y}}_2,\hat{\boldsymbol{y}}_3\}$(Twitter)或 $\hat{\boldsymbol{y}}=\{\hat{\boldsymbol{y}}_0,\hat{\boldsymbol{y}}_1\}$(Weibo),其中 $\hat{\boldsymbol{y}}_0,\hat{\boldsymbol{y}}_1,\hat{\boldsymbol{y}}_2,\hat{\boldsymbol{y}}_3$ 分别表示标签 0,1,2,3 的预测概率值.

最后,将模型的损失函数定义为预测结果与真实标签之间的交叉熵:

$$L(\theta)=-\sum_{i=0}^{r-1}\boldsymbol{y}_i\log(\hat{\boldsymbol{y}}_i),\tag{22}$$

其中,r 为分类的类别数,θ 为整个模型的参数,$\boldsymbol{y}_i\in\{0,1,2,3\}$(Twitter),$\boldsymbol{y}_i\in\{0,1\}$(Weibo)为真实标签值.

3.5 时空复杂度分析

对所提 GUCNH 模型的时间复杂度和空间复杂度进行分析.对于端到端的深度学习算法而言,相比训练的时间复杂度,实际应用中,更关注其预测时间复杂度,因此,在进行时间复杂度分析的时候,我们只分析所提模型预测一个谣言需要的时间.在进行空间复杂度分析的时候,我们则更关注于训练参数的个数.分 2 个方面进行分析:

1) 时间复杂度分析.对于本文提出的方法,当来自邻居的信息根据式(3)进行 GCN 运算的时候,时间复杂度与转发图中节点的个数 k 以及平均入度 β 有关,所以式(3)的时间复杂度为 $O(\beta kd^2)$,其中 d 为节点表示维度.式(6)~(9)的时间复杂度为 $O(kd^2)$,所以 GUCN 模块的总体时间复杂度为 $O((\beta+1)kd^2)$.多头自注意力模块的时间复杂度除了与节点个数 k 相关,还与头的个数相关,文章中使用了 4 个,所以该模块的时间复杂度为 $O(4k^2d^2)$,综合可得在节点表示模块,时间复杂度为 $O(4k^2d^2+2(\beta+1)kd^2)$.根节点选择性增强模块的时间复杂度为 $O(d^2)$.谣言分类阶段的时间复杂度则为 $O(rkd^2)$,其中 r 为最终分类的类别数.

2) 空间复杂度分析.只关注模型训练时候的可学参数,为了方便叙述,我们统一的将所有参数的维度设为 u,而忽略那些不同模块参数维度之间的差异.在本文的模型中,GUCN 模块单元中包含 $\boldsymbol{W}_0$,$\boldsymbol{W}_1$,$\boldsymbol{W}_2$,$\boldsymbol{W}_3$,所以该模块的可学习参数量可近似为 $4u^2$.在多头注意力模块,可学习的权重参数为 $\boldsymbol{W}^{\mathrm{O}}$,$\boldsymbol{W}_i^{\mathrm{Q}}$,$\boldsymbol{W}_i^{\mathrm{K}}$,$\boldsymbol{W}_i^{\mathrm{V}}$,其中 $i=4$,所以这部分参数量约为 $13u^2$.在选择性增强根节点表示模块,$\boldsymbol{W}_4$,$\boldsymbol{W}_5$,$\boldsymbol{W}_6$.为可学习参数,所以参数量约为 $3u^2$.在谣言分类模块,如式(12)我们引入了 Fc(·)全连接网络,参数量约为 u^2+u.综上,我们所提出算法的可训练参数总数约为 $25u^2+u$.

4 实 验

将通过实验回答 3 个问题:

1) 问题 1.与现有的谣言检测方法相比,本模型 GUCNH 是否能够获得较好的谣言检测性能?

2) 问题 2.GUCNH 的每个模块对于谣言检测的性能是否有贡献?

3) 问题 3.与现有的谣言检测方法相比,GUCNH 是否具有优秀的早期检测性能?

4.1 实验数据和设置

1) 实验数据

我们在 3 个真实数据集上评估了我们提出方法的有效性:Twitter15[30], Twitter16[30] 和 Weibo[18].Twitter15,Twitter16 数据集均包含 4 个标签类别,分别是非谣言(N)、经过验证的非谣言(F)、真谣言(T)、未经证实的谣言(U).而 Weibo 数据集包含 2 个标签类别,分别是谣言(T)和非谣言(F).数据集中的每个事件标签都是根据辟谣网站上文章的真实性标签来标注的,这 3 个数据集的详细统计情况如表 1 所示:

Table 1 Dataset Statistics

表 1 数据集统计

统计	Twitter15	Twitter16	Weibo
帖子数量	331 612	204 820	3 805 656
用户数量	276 663	173 487	2 746 818
事件数量	1 490	818	4 664
真谣言数量	374	205	2 351
经过验证的非谣言数量	370	205	2 313
未验证谣言数量	374	203	0
非谣言数量	372	205	0
每个事件平均时间跨度/h	1 337	848	2 460
每个事件平均贴子数	223	251	816
每个事件最大贴子数	1 768	2 765	59 318
每个事件最小贴子数	55	81	10

2) 对比方法

为了验证我们的模型,我们将提出的方法和一些最先进的基线方法进行了比较,这些方法大致可以分为基于机器学习的方法、基于传播链和传播树的方法、基于传播图的方法:

① 基于机器学习的方法

Ⅰ DTC[4]:使用基于人工设计的各种统计特征进行分类的决策树分类模型.

Ⅱ SVM-RBF[16]:一种基于支持向量机的 RBF 核模型,利用手工制作的特征对帖子进行总体统计.

② 基于传播链和传播树的方法

Ⅰ BU-RvNN[19]:基于递归网络的自底向上树状结构的谣言检测模型.

Ⅱ TD-RvNN[19]:基于递归神经网络的自顶向下树状结构的谣言检测模型.

Ⅲ PPC_RNN+CNN[17]:一种结合递归神经网络和卷积神经网络的模型,通过谣言传播链中的用户特征来进行谣言检测.

Ⅳ CED(0.975)[34]:一种基于谣言转发序列的可信度检测模型,该模型通过寻找一个时间点来做出可信的预测,其中 0.975 为预测阈值.

③ 基于传播图的方法

BiGCN[21]:利用信息传播时的双向传播结构使用图卷积网络进行谣言检测的模型.

3) 实现细节和评价指标

首先,本文所有实验的机器配置以及环境为:Intel i7 2.20 GHz(处理器),8.0 GB(内存),GTX-1050 ti(GPU),所有代码都是用 Python(3.7.6)实现,scikit-learn(0.22.1),Theano(1.0.4),Pytorch(1.4.0).

① 基于机器学习的方法:

使用 scikit-learn 实现基于机器学习的对比方法 DTC 和 SVM-RBF,对于特征的选择与提取,完全按照原文描述基于我们的数据集提取了有效特征(主要包括:转发数、粉丝数、发布设备类型、好友数量、用户所在地、是否认证、发帖数、性别、评论数等).

② 基于传播链和传播树的方法:

使用 Theano 实现了基于传播链的方法 BU-RvNN 和 TD-RvNN①,使用 pytorh 实现了基于用户传播链的方法 PPC_RNN+CNN②.在 BU-RvNN 和 TD-RvNN 中,所有模型的参数通过 Adam[35]算法更新,模型参数的初始化使用均匀分布,词汇大小设置为 5000,隐层单元大小设置为 100.在 PPC_RNN+CNN 中,我们设置 $epoch=200$,早停机制轮数设置为 10,GRU 输出维度设置为 32,CNN 窗口大小设置为 3,$dropout$ 率设置为 0.5.对于 CED 方法,由于可复现性问题,我们仅在 Weibo 数据集上得到了结果(结果来自原文).

③ 基于传播图的方法:

使用 Pytorch 实现了基于传播图的方法 BiGCN③以我们提出的模型 GUCNH.其中 BiGCN 的复现代码由原作者提供,每个节点的隐层特征向量维度设置为 64,随机删除边的比率 q 设置为 0.2,$dropout$ 率设置为 0.5,$epoch$ 设置为 200,其余参数设置严格按照原文设定.

我们所提模型中的参数由 Adam[35]算法更新,学习率初始化为 1E-4,在训练过程中逐渐降低.我们利用 TF-IDF 值提取前 d 个单词构建词袋模型作为文本的初始表征,设置 $d=5000$,模型中图卷积网络输出表征的维度 m 和门控单元输出表征的维度 n 均设置为 64,多头自注意力模块头的个数 $h=4$.对于原始的转发图,我们设置随机的删除边的比率 $q=0.2$,即随机删除 20%的边.实验的 $batchsize=128$,$epoch=100$,为了防止过拟合,模型中用到了 $dorpout$ 机制,其比率为 0.3,我们将数据集随机分成 5 部分进行 5 折交叉验证以获得结果,除此之外还应用了早停机制[36].

我们采用了与先前工作中相同评估指标[20,37],即准确度、$F1$ 分数、召回率和精准率进行评估.为了公平比较,我们的方法和对比方法在所有数据集上的结果都是在 5 次实验的结果上取平均.

4.2 实验结果分析

为了回答问题 1,通过实验得到分类的总体准确率 Acc 和各类别的 $F1$ 值来验证本文模型的谣言检测性能.表 2~4 分别展示了本文模型以及所有比较方法在 3 个数据集上的性能.显然,我们提出的模型优于选定的对比模型.对实验结果进行分析:

1) 可以观察到深度学习方法的性能要明显地优于机器学习方法,理由是因为深度学习方法可以捕捉到更有价值的高层特征,而机器学习的方法需要手工提取特征,检测能力较为局限.这进一步说明了研究深度学习方法在谣言检测中的重要性和必要性.

2) 可以观察到我们提出的 GUCNH 模型在 Twitter15 数据集上的结果要比 BU-RvNN 和 TD-RvNN 模型分别高 17.6 个百分点和 16.1 个百分点,在 Twitter16 数据集上的结果比 BU-RvNN 和 TD-RvNN 模型分别高 16.8 个百分点和 14.9 个百分点,在 Weibo 数据集上的结果比 BU-RvNN 和 TD-RvNN 模型分别高 7.1 个百分点和 6.3 个百分点.实验结果表明传播结构中包含很多重要信息,捕获这部分结构信息有助于谣言检测任务,将任务建模为传播图以捕获全局结构信息的方法要优于通过建模为传播树捕获局部序列特征的方法.

① https://github.com/majingCUHK/Rumor_RvNN

② https://github.com/yumere/early-fakenews-detection

③ https://github.com/TianBian95/BiGCN

Table 2 Experimental Results on Twitter15 Dataset

表 2 Twitter15 数据集上的实验结果

模型	*Acc*	*F*1			
		非谣言(N)	经过证实的非谣言(F)	真谣言(T)	未经证实的谣言(U)
DTC(①)	0.454	0.733	0.355	0.317	0.415
SVM-RBF(①)	0.318	0.225	0.282	0.455	0.218
BU-RvNN(②)	0.708	0.695	0.728	0.759	0.653
TD-RvNN(②)	0.723	0.682	0.758	0.759	0.654
PPC_RNN+CNN(②)	0.812	0.810	0.813	0.790	0.785
Bi-GCN(③)	0.835	0.767	0.843	0.887	0.808
GUCNH(③)	**0.884**	**0.816**	**0.876**	**0.954**	**0.858**

注:下划线为对比模型的最优结果;黑体为本文模型且为最优结果.

Table 3 Experimental Results on Twitter16 Dataset

表 3 Twitter16 数据集上的实验结果

模型	*Acc*	*F*1			
		非谣言(N)	经过证实的非谣言(F)	真谣言(T)	未经证实的谣言(U)
DTC(①)	0.465	0.643	0.393	0.419	0.403
SVM-RBF(①)	0.553	0.670	0.485	0.317	0.361
BU-RvNN(②)	0.718	0.723	0.712	0.779	0.659
TD-RvNN(②)	0.737	0.662	0.743	0.835	0.708
PPC_RNN+CNN(②)	0.855	0.811	0.871	0.837	0.842
Bi-GCN(③)	0.860	0.779	0.859	0.925	0.855
GUCNH(③)	**0.886**	**0.837**	**0.897**	**0.934**	**0.897**

注:下划线为对比模型的最优结果;黑体为本文模型且为最优结果.

Table 4 Experimental Results on Weibo Dataset

表 4 Weibo 数据集上的实验结果

模型	*Acc*	*Prec*		*Rec*		*F*1	
		T	F	T	F	T	F
DTC(①)	0.831	0.815	0.847	0.824	0.815	0.819	0.831
SVM-RBF(①)	0.878	0.629	0.776	0.695	0.646	0.615	0.711
BU-RvNN(②)	0.894	0.910	0.912	0.914	0.907	0.905	0.918
TD-RvNN(②)	0.901	0.918	0.914	0.916	0.896	0.911	0.921
PPC_RNN+CNN(②)	0.913	0.927	0.884	0.901	0.932	0.907	0.922
CED(0.975)(②)	0.921		0.934		0.899		0.916
Bi-GCN(③)	0.934	0.928	0.940	0.939	0.930	0.929	0.931
GUCNH(③)	**0.958**	**0.956**	**0.960**	**0.959**	**0.957**	**0.958**	**0.958**

注:下划线为对比模型的最优结果;黑体为本文模型且为最优结果;"T"为谣言;"F"为非谣言.

3) 相比于 PPC_RNN+CNN,我们提出的模型结果更好.一方面,PPC_RNN+CNN 仅仅使用传播链上的用户信息进行建模,单一使用用户的一些特征来检测谣言有一定的片面性;另一方面,PPC_RNN+CNN 并没有考虑到实际的转发结构.我们提出的模型根据实际的转发结构充分了利用了每个帖子的内容信息,从而取得了更好的结果,由此可见实际的转发结构在检测谣言中的重要性.相较于 CED(0.975),我们的模型在 Weibo 数据集上的准确率要高 4 个百分点,这进一步说明了利用全局传播结构的优势.

4) 本文模型的实验结果要优于 BiGCN,BiGCN

虽然使用了双向的 GCN 对于转发图结构进行了建模,同时还在 2 次 GCN 之间融入了一定的源节点信息,但是仅仅使用 GCN 聚合得到节点表示的方法太过于依赖 GCN 的表现,这一点本文模型通过引入门控单元来弥补.此外,本文模型引入了多头自注意力模块来考虑任意节点之间的多方面影响,可以有效弥补有限次 GCN 不能很好地捕获任意节点信息的缺陷.

4.3 消融实验

为了回答问题 2,证明我们提出模型各模块的有效性,进行了一系列的消融实验.主要包括 4 部分:

1) w/o Matt.移除多头自注意力模块,在节点表示模块,只使用 2 次 GUCN 的堆叠,其余部分不变.

2) w/o 1GUCN.移除一个 GUCN 模块,主要用于验证 GUCN 模块堆叠的有效性,将多头自注意力模块输出的结果作为节点表示模块的输出,然后拼接源帖表示进行分类.

3) w/o Head.移除选择性增强根节点表示模块,主要用于验证增强源帖信息对于该场景分类的有效性.

4) w/o GU.移除每个 GUCN 模块中的 GU 门控单元,只保留图卷积操作,用于验证我们引入的门控网络与图卷积网络融合的有效性.

如图 4 为消融实验的结果,其中 ALL 为不做任何消融的原始模型 GUCNH,根据表中的实验结果,可以得到结论为:

首先研究多头自注意力模块带来的影响,根据实验结果可以看到,删除多头自注意力模块会影响我们的模型在 3 个数据集上的结果,其中 GUCNH 在消融多头自注意力模块后,Twitter15 和 Twitter16 数据集上的结果分别下降了 3.3 个百分点和 2.0 个百分点,Weibo 数据集上的结果下降了 1.5 个百分点.多头自注意力模块可以捕获任意节点之间的影响,而不仅仅限于具有邻接关系的节点之间,使得在进行下一次节点信息融合之前所有节点的信息尽可能的全面,对于结果的提升有很大的帮助.结果同样可以证明我们引入该模块的动机,并非具有直接转发关系的帖子之间会相互影响,任意的帖子之间也会存在相互影响,而使用多头注意力模块能够很好地考虑到这些影响,取得较好的结果.

随后我们评估了 GUCN 模块堆叠的有效性.GCN 的适当堆叠有助于节点聚合高阶邻居的信息,所以我们的模型采用了融合门控的图卷积网络模块堆叠的方式.一方面使得节点能够聚合到更远节点上的信息;另一方面为了在多头注意力机制之后重新让节点数据考虑到结构信息.为了验证 GUCN 模块堆叠的有效性,我们进行了 w/o 1GUCN 消融实验,根据实验结果可以看到,不进行 GUCN 模块堆叠会影响我们所提模型在 3 个数据集上的结果,GUCNH 在不堆叠 GUCN 模块的实验中,Twitter15 和 Twitter16 数据集上的结果分别下降了 3.4 个百分点和 0.8 个百分点,Weibo 数据集上的结果下降了 2.6 个百分点.结果表明,对融合门控的图卷积网络模块 GUCN 进行堆叠使用可以使得节点更好地融合邻居节点甚至更远节点的信息,同时对于多头自注意力模块有可能造成的结构信息破坏问题有一定的解决,所以取得比单一使用该模块更好的结果.

谣言事件的源帖总是有着最丰富且重要的信息,

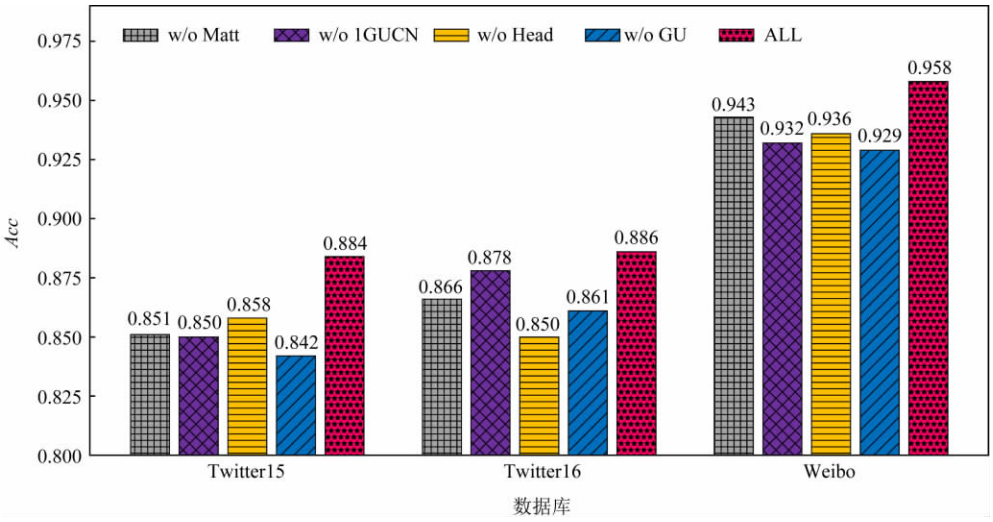

Fig. 4 The ablation experiment result of the GUCNH on three datasets

图 4 GUCNH 在 3 个数据集上的消融实验结果

所以我们的模型包含选择性增强根节点模块,作用就是额外的为每个节点增加源帖的信息.为了证明设计的有效性,进行了该模块的消融实验.根据实验结果可以看到,不增强头节点的信息会影响我们所提模型在3个数据集上的结果. GUCNH在没有选择性增强头节点模块的实验中,Twitter15和Twitter16数据集上的结果分别下降了2.6个百分点和3.6个百分点,Weibo数据集上的结果下降了2.2个百分点.结果表明,源帖有着非常重要且原始的信息,为每个节点额外的增加源节点对应的信息,能够有效地提高该场景下的检测能力.

最后研究了引入融合门控的图卷积网络的有效性,实验过程是将原模型中所有融合门控的图卷积网络模块GUCN换为单一的图卷积网络模块GCN进行实验,根据实验结果可以看到,使用单一的GCN会影响我们所提模型在3个数据集上的结果,GUCNH在使用单一GCN的实验中,Twitter15和Twitter16数据集上的结果分别下降了4.2个百分点和2.5个百分点,Weibo数据集上的结果下降了2.9个百分点.结果表明,引入门控单元GU能够对进行图卷积之前的特征表示和之后的特征表示进行选择与组合,从而得到更好的表示使得分类结果有了一定的提升.

4.4 早期检测研究

在谣言检测任务中,最关键的目标之一是尽早发现谣言,以便及时进行干预[38].为了回答问题3,验证我们提出的模型具有优秀的早期检测性能,我们在Twitter15和Twitter16这2个数据集上设计了早期检测实验,具体的方法是设置检测截止时间节点,即仅使用在发布时间到检测截止时间节点之间的帖子内容来评估模型检测的性能.通过改变检测截止时间节点(我们设置节点分别是源帖发布后4 h,8 h,12 h,24 h,36 h),分别得出了2个数据集上的早期检测结果,如图5和图6分别为2个数据集上进行早期检测的结果.可以看到,在源帖发布的最早期,也就是图5、图6中4 h时,我们提出模型的在Twitter15数据集和Twitter16数据上分别取得了82.1%和84.1%的结果,可以看出这些结果比其余对比方法的结果好,这表明我们提出的模型具有良好的早期检测性能.当检测截止时间节点逐渐增大时,我们模型的性能仍然呈上升趋势,这一点与BiGCN等模型不同,随着时间节点的变大,转发结构更加复杂,言论种类也逐渐增多,我们的模型仍然可以保持很好的结果,说明我们的模型对复杂的数据不敏感,具有较好的稳定性和鲁棒性.

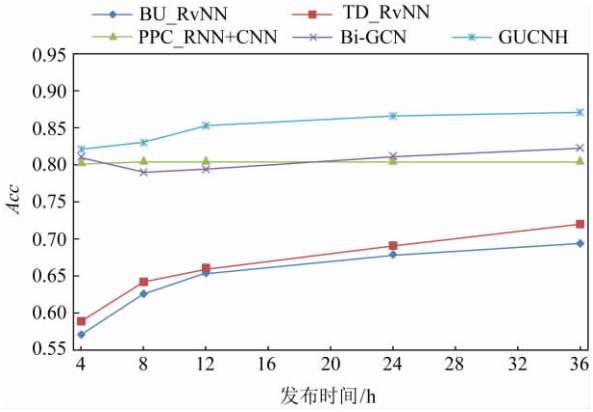

Fig. 5 Experimental results of early detection on Twitter15 dataset

图5 Twitter15数据集上早期检测实验结果

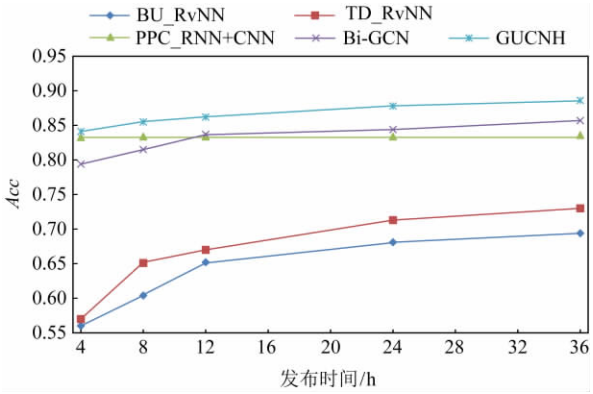

Fig. 6 Experimental results of early detection on Twitter16 dataset

图6 Twitter16数据集上早期检测实验结果

5 总结与展望

本文提出了一个融合门控的传播图卷积网络模型GUCNH,该模型首先通过融合门控的图卷积网络模块GUCN来根据实际转发结构聚合邻居信息以生成节点的表示,即门控机制用来对进入图卷积网络之前的特征表示和经过图卷积网络之后的特征表示进行选择与组合得到质量更高的特征表示,同时在2个融合门控的图卷积模块之间引入了多头自注意力模块来考虑任意节点之间的影响,使得节点信息在进入下一次融合之前包含尽可能全面的信息.在生成节点的高级特征表示之后,我们选择性的增强了源节点的信息,理由是往往转发源的信息最为丰富.为了确保增强的源节点信息的质量,同样加入门控单元对于源节点的信息进行了选择与组合,

最终将选择后的源节点特征表示与所有节点的特征表示拼接用于分类.在3个真实数据集上的实验结果表明,我们提出的方法优于最先进的方法.

在未来的研究中,我们将主要从2个方面继续深入工作:1)在转发图的构建方面,寻找更加合适的建模方法(如加入用户构建异构图),以提高检测性能.2)一般来说,完整的社交帖子不仅只有文本内容,同样还会包含图像或视频等信息,在接下来的研究中,我们还将考虑利用多模态信息来解决谣言检测问题.

参 考 文 献

[1] Kantar M. Social Media Trends 2019 [EB/OL]. 2019 [2020-10-05]. https://www.kantarmedia.com/global/thinking-and-resources/latest-thinking/socialmediatrends2019

[2] Silverman, C. This analysis shows how viral fake election news stories outperformed real news on facebook [EB/OL]. BuzzFeed News, 2016 [2020-10-05]. https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

[3] Liu Bo, Li Yang, Meng Qing, et al. Evaluation of content credibility in social media [J]. Journal of Computer Research and Development, 2019, 56(9): 1939-1952 (in Chinese)
(刘波, 李洋, 孟青, 等. 社交媒体内容可信性分析与评价[J]. 计算机研究与发展, 2019, 56(9): 1939-1952)

[4] Castillo C, Mendoza M, Poblete B. Information credibility on Twitter [C] //Proc of the 20th Int Conf on World Wide Web. New York: ACM, 2011: 675-684

[5] Shu Kai, Sliva A, Wang Suhang, et al. Fake news detection on social media: A data mining perspective [J]. SIGKDD Explor, 2017, 19(1): 22-36

[6] Horne B D, Adali S. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news [J]. arXiv preprint arXiv: 1703.09398, 2017

[7] Popat K. Assessing the credibility of claims on the Web [C] //Proc of the 26th Int Conf on World Wide Web Companion. New York: ACM, 2017: 735-739

[8] Potthast M, Kiesel J, Reinartz K, et al. A stylometric inquiry into hyperpartisan and fake news [C] //Proc of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 231-240

[9] Wang W Y. "Liar, liar pants on fire": A new benchmark dataset for fake news detection [C] //Proc of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2017: 422-426

[10] Ahn Y, Jeong C, et al. Natural language contents evaluation system for detecting fake news using deep learning [C] //Proc of 16th Int Joint Conf on Computer Science and Software Engineering. Piscataway, NJ: IEEE, 2019: 289-292

[11] Yu Feng, Liu Qiang, Wu Shu, et al. A convolutional approach for misinformation identification [C] //Proc of the 26th Int Joint Conf on Artificial Intelligence. Melbourne: IJCAI, 2017: 3901-3907

[12] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [J]. arXiv preprint arXiv:1609.02907, 2016

[13] Vaibhav V, Annasamy R, Hovy E. Do sentence interactions matter? leveraging sentence level representations for fake news classification [C] //Proc of the 13th Workshop on Graph-Based Methods for Natural Language Processing. Stroudsburg, PA: ACL, 2019: 134-139

[14] Wang Yuhang, Wang Li, Yang Yanjie, et al. SemSeq4FD: Integrating global semantic relationship and local sequential order to enhance text representation for fake news detection [J]. Expert Systems with Applications, 2021, 166: No. 114090

[15] Yan Rui, Yen Ian E. H, Li Chengte, et al. Tackling the achilles heel of social networks: Influence propagation based language model smoothing [C] //Proc of the 24th Int Conf on World Wide Web, New York: ACM, 2015: 1318-1328

[16] Yang Fan, Yu Xiaohui, Liu Yang, et al. Automatic detection of rumor on Sina Weibo [C] //Proc of the ACM SIGKDD Workshop on Mining Data Semantics. New York: ACM, 2012: 1-7

[17] Liu Yang, Wu Yifang Brook. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks [C] //Proc of the 32nd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2018: 354-361

[18] Ma Jing, Gao Wei, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks [C] //Proc of the 25th Int Joint Conf on Artificial Intelligence. PaloAlto, CA: AAAI, 2016: 3818-3824

[19] Ma Jing, Gao Wei, Wong K, et al. Rumor detection on Twitter with tree-structured recursive neural networks [C] //Proc of the 56th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2018: 1980-1989

[20] Lu Yiju, Li Chengte. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media [C] //Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 505-514

[21] Bian Tian, Xiao Xi, Xu Tingyang, et al. Rumor detection on social media with bi-directional graph convolutional networks [C] //Proc of the AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2020: 549-556

[22] Pérez-Rosas V, Kleinberg B, Lefevre A, et al. Automatic detection of fake news [J]. arXiv preprint arXiv:1708.07104, 2017

[23] Takahashi T, Igata N. Rumor detection on Twitter [C] // Proc of the 6th Int Conf on Soft Computing and Intelligent Systems, and the 13th Int Symp on Advanced Intelligence Systems. Kobe. Piscataway, NJ: IEEE, 2012: 452-457

[24] Shu Kai, Zhou Xinyi, Wang Suhang, et al. The role of user profiles for fake news detection [C] //Proc of the 2019 IEEE/ACM Int Conf on Advances in Social Networks Analysis and Mining. New York: ACM, 2019: 436-439

[25] Kwon S, Cha M, Jung K, et al. Prominent features of rumor propagation in online social media [C] //Proc of 2013 IEEE 13th Int Conf on Data Mining. Piscataway, NJ: IEEE, 2013: 1103-1108

[26] Ma Jing, Gao Wei, Wei Zhongyu, et al. Detect rumors using time series of social context information on microblogging websites [C] //Proc of the 24th ACM Int Conf on Information and Knowledge Management (CIKM'15). New York: ACM, 2015: 1751-1754

[27] Wu Ke, Yang Song, et al, False rumors detection on sina Weibo by propagation structures [C] //Proc of 31st 2015 IEEE Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2015: 651-662

[28] Wei Penghui, Xu Nan, Mao Wenji. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity [J]. arXiv preprint arXiv: 1909.08211, 2019

[29] Wu Zhiyuan, Pi Dechang, Chen Junfu, et al. Rumor detection based on propagation graph neural network with attention mechanism [J]. Expert Systems with Applications, 2020, 158: 113595

[30] Ma Jing, Gao Wei, et al. Detect rumors in microblog posts using propagation structure via kernel learning [C] //Proc of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2017: 708-717

[31] Rong Yu, Huang Wenbing, et al. DropEdge: Towards deep graph convolutional networks on node classification [J]. arXiv preprint arXiv: 1907.10903, 2019

[32] Cho K, Merrienboer B, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C] //Proc of EMNLP'2014. Stroudsburg, PA: ACL, 2014: 1724-1734

[33] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need [C] //Proc of Conf on NIPS'2017. Cambridge, MA: NIPS, 2017: 5998-6008

[34] Song Changhe, Tu Cunchao, Yang Cheng, et al. CED: Credible early detection of social media rumors [J/OL]. IEEE Transactions on Knowledge and Data Engineering, 2019 [2019-09-03]. https://doi.org/10.1109/tkde.2019.2961675

[35] Kingma D, Ba J. Adam: A method for stochastic optimization [C/OL] //Proc of the 3rd Int Conf on Learning Representations. San Diego, CA: ICLR, 2015 [2019-09-03]. http://arxiv.org/abs/1412.6980

[36] Yao Y, Rosasco L, Caponnetto A. On early stopping in gradient descent learning [J]. Constructive Approximation, 2007, 26(2): 289-315

[37] Yuan Chunyuan, Ma Qianwen, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection [C] //Proc of 2019 IEEE Int Conf on Data Mining. Piscataway, NJ: IEEE, 2019: 796-805

[38] Zhao Z, Pesnick P, Mei Q. Enquiring minds: Early detection of rumors in social media from enquiry posts [C] //Proc of the 24th Int Conf on World Wide Web. New York: ACM, 2015: 1395-1405

Yang Yanjie, born in 1995. Master candidate. His main research interests include Natural Language Processing, data mining.

杨延杰,1995 年生.硕士研究生.主要研究方向为自然语言处理、数据挖掘.

Wang Li, born in 1971. PhD, professor. Senior member of CCF. Her main research interests include big data computation and analysis, knowledge graph, data mining.

王　莉,1971 年生.博士,教授,CCF 高级会员.主要研究方向为大数据计算与分析、知识图谱、数据挖掘.

Wang Yuhang, born in 1997. Master candidate. Student member of CCF. Her main research interests include natural language processing, data mining. (wangyuhang0983@link.tyut.edu.cn)

王宇航,1997 年生.硕士研究生,CCF 学生会员,主要研究方向为自然语言处理、数据挖掘.