计算机应用
Journal of Computer Applications
ISSN 1001-9081,CN 51-1307/TP

《计算机应用》网络首发论文

题目：融合评论序列二义性与生成用户隐私特征的谣言检测
作者：孟文凡，周丽华，王晓旭
收稿日期：2023-08-31
网络首发日期：2023-11-02
引用格式：孟文凡，周丽华，王晓旭．融合评论序列二义性与生成用户隐私特征的谣言检测[J/OL]．计算机应用．
https://link.cnki.net/urlid/51.1307.TP.20231102.1326.016

Journal of Computer Applications
计算机应用
ISSN 1001-9081
CODEN JYIIDU
http://www.joca.cn

DOI:10.11772/j.issn.1001-9081.2023081176

融合评论序列二义性与生成用户隐私特征的谣言检测

孟文凡[1],周丽华[1*],王晓旭[2]
(1.云南大学 信息学院,昆明 650504; 2.云南大学滇池学院,昆明 650228)
(*通信作者电子邮箱lhzhou@ynu.edu.cn)

摘　要:针对现有谣言检测工作存在的以下问题:1)没有同时捕获评论序列的文本语义特征和时间周期特征;2)在隐私保护环境下无法获取用户个人资料,导致传播结构中的信息难以充分融合。为此,提出了融合评论序列二义性与生成用户隐私特征的谣言检测模型(RD-CSGU)。该模型综合考虑了评论序列不同视角下的文本语义特征和时间周期特征,同时构建了反映传播过程中用户之间社交互动关系的谣言传播异质网络,并基于该网络中的语义关系通过生成对抗网络生成用户的隐私特征,解决了用户个人资料访问受限的问题。在Twitter15、Twitter16、Weibo数据集上展开有效性验证,与次优基线模型GLAN(Global-Local Attention Network)相比,RD-CSGU的Acc指标分别提升了0.9个百分点、2.2个百分点和1.8个百分点,TR-F1值分别提升了2.6个百分点、6.8个百分点和1.9个百分点,结合消融实验、GAN生成嵌入分析的实验结果表明,RD-CSGU能够有效检测出社交媒体平台上发布的谣言帖子。

关键词:谣言检测;评论序列;传播异质网络;生成特征;传播结构
中图分类号:TP391.1　　**文献标志码**:A

Rumor detection by fusing ambiguity in comment sequences and generating user privacy features

MENG Wenfan[1], ZHOU Lihua[1*], WANG Xiaoxu[2]
(1. *School of Information Science and Engineering, Yunnan University, Kunming* 650504, *China*;
2. *Dianchi College of Yunnan University, Kunming* 650228, *China*)

Abstract: Aiming to the difficult of fully integrating information within the propagation structure resulting from the deficiency of simultaneously capturing text semantic features and temporal periodicity features in comment sequences and the inability to access the users′ personal profiles in a privacy-protected environment, a rumor detection model integrated the Ambiguity of comment sequences and generated user privacy characteristics(RD-CSGU) was proposed, where text semantic features and the temporal periodicity features from different perspectives of comment sequences was comprehensively considered. Meanwhile, a heterogeneous network of rumor propagation for describing the social interaction among users during the propagation process was constructed. Furthermore, users′ privacy features were generated through generative adversarial network) based on the semantic relationships, overcoming the limitation of insufficient users′ personal profiles. The effectiveness of the proposed model was validated on Twitter15、Twitter16、Weibo datasets, Compared with the suboptimal baseline model GLAN (Global-Local Attention Network), RD-CSGU achieved improvements of 0.9 percentage points, 2.2 percentage points, and 1.8 percentage points in terms of accuracy, as well as improvements of 2.6 percentage points, 6.8 percentage points, and 1.9 percentage points in terms of TR-F1 score. The results combined with those from ablation experiments and analysis of GAN-generated embeddings show that RD-CSGU could effectively detect rumor posts on social media platforms.

Key words: rumor detection; comment sequence; propagation heterogeneous network; generated feature; propagation structure

0　引言

谣言是在人与人之间传播,含有人们广泛关注的信息,且真实性未经证实的一种信息表述[1]。近年来,社交媒体的发展促进了信息的传播和交换,每个用户都可以非常便捷地在社交平台上表达意见和观点,但恶意用户也借助着这种便捷性大量散布谣言。这些谣言误导社会公众、扰乱社会秩序、给社会造成严重的影响。因此,有必要在社交媒体上构建有效的谣言检测方法。

早期的谣言检测工作主要采用基于特征工程的检测方法。这些方法通过手工提取消息帖子中的各种特征,如消息文本特征、转发用

收稿日期:2023-08-31;**修回日期**:2023-09-14;**录用日期**:2023-10-09。
基金项目:国家自然科学基金资助项目(62062066);云南省基础研究计划项目(202201AS070015)。
作者简介:孟文凡(1996—),男,贵州独山人,硕士研究生,CCF学生会员,主要研究方向:数据挖掘、信息扩散、谣言检测; 周丽华(1968—),女,云南华坪人,教授,博士,CCF会员,主要研究方向:数据挖掘、多视角学习、社会网络分析; 王晓旭(1995—),女,安徽阜阳人,讲师,硕士,主要研究方向:数据挖掘、大数据技术。

户特征、传播结构特征等，并在一个有监督的机器学习框架下进行谣言检测工作[2-3]。但随着社交媒体的发展，用户与帖子的数据量呈几何式增长，使得这种依赖手工处理特征的方法不再适合，并且不同社交平台、转发时间、文本内容对模型的可移植性与可伸缩性提出了更高的要求。因此，近期的研究工作普遍采用深度学习的方法，这些方法从谣言帖子、回复评论（回复该帖子的评论集合）、转发帖子的用户文件等可用资料中建立模型自动学习特征。其中帖子的文本内容一般长度较短，导致无法获取充足的语义信息，所以往往将帖子和评论以时间尺度上先后出现的顺序连接起来形成序列，序列以帖子的回复评论集合为主，因此在本文中统一称其为评论序列。

评论序列具有二义性，即文本语义性和时间周期性。文本语义性是指在文本内容中词汇、句法、段落层次不同语言风格的性质。Chen等[4]通过循环神经网络（Recurrent Neural Network，RNN）捕获帖子和评论文本序列的长距离依赖关系来学习文本内容隐藏的特征，依据评论序列的文本语义性，得到包含煽动焦虑、影响判断以及迎合人性跟风心态等文本语义特点从而检测谣言。时间周期性是指在时间上事件展现循环行为的性质。Sejeong等[5]联合周期性外部冲击（Periodic External Shock，PES）模型和流行病学中的易感-感染（Susceptibility-Infection，SI）模型捕捉这种周期性循环行为。谣言生命周期过程中爆发的评论分为潜伏期、繁殖期、传播期和衰落期[6]，依据评论序列在这四个时间段之间的相似特征和变化趋势得到周期性特征，从而区分谣言与非谣言。因此，在进行谣言检测时评论序列数据不仅携带着文本信息，而且隐含着时间周期规律信息，二义性阐述了评论序列不同方面的可用信息，融合评论序列的二义性有助于发掘可用数据的隐含特征，提升模型的检测能力。

除了评论序列特征能区分帖子是否为谣言以外，以用户转发帖子形成特定的传播结构在判别谣言帖子中同样发挥着重要作用，Huang等[7]指出谣言和非谣言在传播结构上存在显著差异。要想捕捉谣言帖子特定的传播结构特征，需要使用两方面的信息，一是传播过程中形成的传播图拓扑结构信息，二是传播图上的帖子以及转发帖子的用户节点属性信息。传播图的拓扑结构信息能直观地从全局关系上了解各个节点之间的连接，反映了谣言传播过程中相似帖子、用户间的结构关系，拓扑结构信息一般可以通过共现用户转发帖子的路径得到。传播图上的节点属性信息提供了拓扑结构以外的补充，即使两个节点之间没有直接联系，但是节点之间的相关性可能隐藏在节点属性内容中，节点属性信息需要访问社交网络上的账号资料获取。近年来，由于保护用户个人隐私信息相关法律法规的完善，很多社交媒体机构出台了保护用户资料的政策，例如Twitter、Facebook等一些社交网站已经限制访问用户个人资料，使得获取用户配置文件成为一个挑战。为了应对这个挑战，Yuan等[8]在将传播结构特征纳入谣言检测模型中时，放弃了通过访问网站获取用户资料的途径，转而假设传播过程中节点之间的关系是独立的，让参与转发的用户节点特征整体上服从正态分布。但是在现实中用户与帖子以及用户与用户之间的节点关系并非独立同分布的[9]，采用正态分布的方法丢弃了谣言传播网络中用户节点之间的社交互动关系，因此检测效果不尽如人意。

为了解决上述没有同时使用评论序列二义性信息以及在保护用户隐私环境下缺失传播节点特征的问题，本文提出了一种融合评论序列二义性与生成用户隐私特征的谣言检测模型（Rumor Detection by Fusing Ambiguity in Comment Sequences and Generating User Privacy Features，RD-CSGU），具体而言，RD-CSGU模型学习了两种表征。第一种表征的学习旨在融合评论序列的文本内容特征和时间周期特征，为此，RD-CSGU模型首先使用具有局部感受野的卷积神经网络（Conventional Neural Network，CNN）捕捉评论文本中多个连续词之间的关联形成文本表示，紧接着使用注意力机制从不同的角度获取评论序列内容深层次的文本语义表示，同时利用时间序列预测敏感的长短期记忆（Long Short-Term Memory，LSTM）网络捕获评论序列的时间周期性行为，然后以门控机制融合评论序列的文本特征和周期特征的权重。第二种表征的学习旨在隐私保护环境中生成用户节点特征分布并将其与传播结构中传播图的拓扑结构进行融合，为此，RD-CSGU模型首先构建包含“用户-帖子-词”三种节点类型的传播异质网络，接着依据该传播异质网络中不同的连接语义关系利用生成对抗网络（Generative Adversarial Network，GAN）生成传播过程中缺失的用户节点特征分布，然后利用图注意力网络（Graph Attention Network，GAT）聚合帖子传播过程中的拓扑结构信息和生成的节点属性信息，得到谣言与非谣言特定的传播结构。其中在生成方法的选择时，由于GAN采用的是一种无监督的训练方式，并且GAN在数据稀疏的异质网络上能捕获不同节点类型或相同节点类型之间远距离的联系[10-11]，因此在没有反对、支持等用户节点标签的传播过程中，基于GAN生成的节点特征分布能够表达用户在没有拓扑结构直连下的社交关系互动情况。谣言检测将学到的两种表征组合起来输入到一个全连接网络，通过分类完成检测任务。

综上所述，本文主要贡献如下：

1）使用多种神经网络融合评论序列的文本内容特征和时间周期特征学习节点的表征表示，既提取了评论序列文本语义特征，又捕获了评论序列的时间周期规律特征。

2）构建了包含“用户-帖子-词”三种节点类型的传播异质网络，并依据不同语义连接关系利用GAN生成用户与帖子的节点特征分布，弥补隐私保护环境中用户节点特征缺失的不足。

3）在三个真实的公共数据集上进行了广泛的实验，实验结果表明，RD-CSGU模型在英文或中文数据集上均具有卓越的检测能力，在三个数据集上比最佳对比模型提升了0.9个百分点、2.2个百分点和1.8个百分点。

1 相关工作

近年来，在社交媒体上进行谣言检测已经成为广泛的研究热点。根据目前研究对象的不同，相关工作可以分为以下：

1）基于文本内容的方法主要通过分析谣言和非谣言在文本中的写作风格和语义信息来进行区分。通常采用Word2Vec和BERT等模型将文本转化为特征向量，然后利用神经网络模型从中提取丰富的信息。例如，Feng等[12]利用卷积神经网络CNN从文本中提取信息，学习谣言的特定写作风格和局部语义特征；Serveh等[13]提出了一种利用图卷积网络（Graph Convolutional neural Network，GCN）检测谣言会话的模型，通过对图进行建模，获取了图的结构信息和会话Tweets的内容；Silva等[14]提出了语义相似性和语义对立性来学习帖子及其回复评论之间的关系。但是神经网络学习到的文本特征只能作为判别谣言的依据之一，完全依靠会导致检测效果不理想。

2）基于用户属性的方法主要通过分析用户的属性信息来推断用户在社交网络上的可信度。通过手动获取用户资料，如性别、年龄、国籍、粉丝数量等，然后利用这些属性信息进行谣言检测。例如，Shu等[15]考虑了用户的政治肤色、个人头像等信息进行谣言检测，并验证了谣言散布者与非谣言散布者的用户特征存在明显差异；Hamdi等[16]从用户的关注者和关注者社交网络中提取特征进行谣言检测；Jiang等[17]通过用户的社交网络发现潜在的用户连接，有效识别出倾向于传播假消息的用户；Lu等[18]利用双重协作注意机制来捕捉用户与邻居节点之间的相关性来检测谣言。然而，在保护用户隐私的环境下，难以获取社交网络上的用户资料，导致这些方法的可移植能力较低。

3）基于传播结构的方法主要通过分析消息传播的结构特征和节点属性来识别谣言和非谣言。消息的传播结构反映了消息的传播行为模式，通过计算不同传播结构的特征来区分谣言和非谣言。例如，Bai等[19]提出了源-回复对话树卷积神经网络（Tree Convolutional neural Network，TCN），用于提取内容特征和结构特征；Jing等[20]利用GCN来聚合树状节点的结构特征，区分谣言和非谣言的传播模式；Bian等[21]引入了GCN来探讨自上而下和自下而上的谣言传播机制，分别通过自顶向下的图卷积网络和自底向上的图卷积网络获得传播和扩散特性；Huang等[22]关注文本内容的全局语义关系，结合用户、帖子将谣言传播过程整理为传播异质图，并利用注意力机制融合帖子-用户和帖子-词子图特征。但是这些方法建立的传播结构没有关

注传播过程中非拓扑直连节点的社交关联。

上述所提的方法往往没有考虑到帖子及评论组成的评论序列具有时间周期特性,同时在保护用户隐私的背景下,无法访问用户资料,使得建立的模型对于缺失的数据具有较低的泛化能力。因此,本文提出了一种融合评论序列二义性与GAN生成用户隐私特征的谣言检测模型。该模型同时考虑了评论序列的文本语义特征和时间周期特征,并构建了包含"用户-帖子-词"节点的传播异质网络,利用该网络上不同的社交关系语义生成用户隐私特征,弥补在隐私保护背景下缺失用户节点特征的不足。

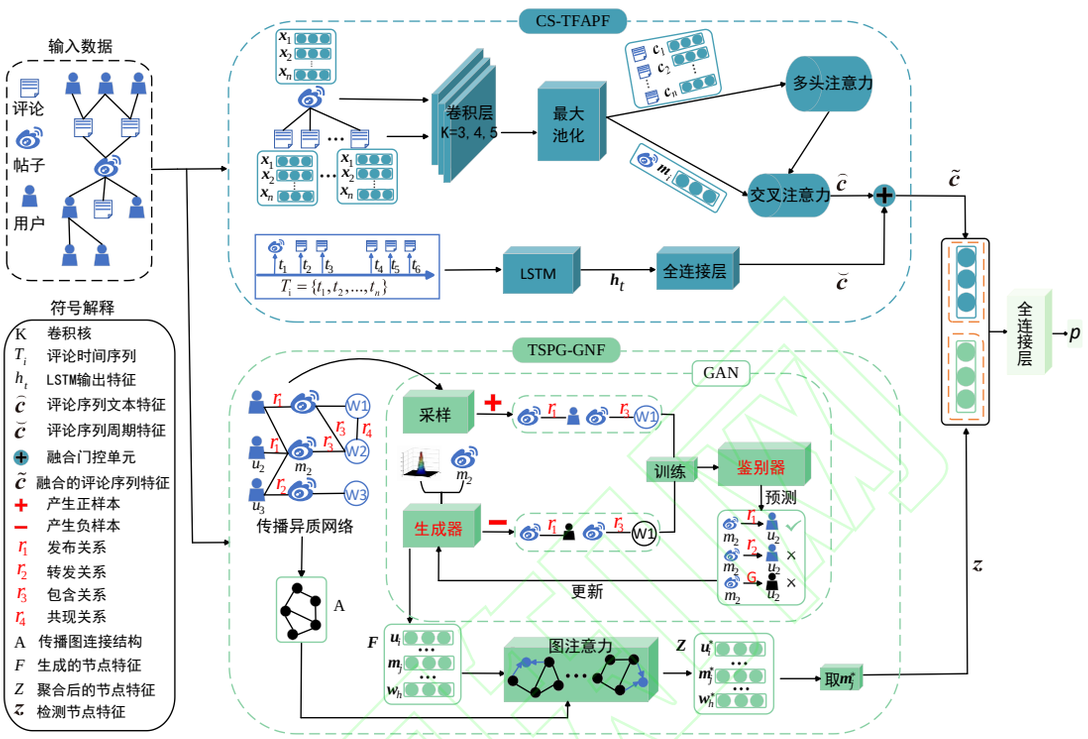


图 1 RD-CSGU模型框架图

Fig 1 RD- CSAGU Model Framework Diagram

2 RD-CSGU模型

2.1 问题描述

假定发布的帖子集合表示为 $\mathcal{M}=\{m_1,m_2,...,m_{|\mathcal{M}|}\}$,在帖子集合上设置一个固定大小的滑动窗口来收集"共现词"信息,"共现词"集合表示为 $\mathcal{W}=\{w_1,w_2,...,w_{|\mathcal{W}|}\}$,参与转发的用户集合表示为 $\mathcal{U}=\{u_1,u_2,...,u_{|\mathcal{U}|}\}$。每条帖子 m_i 的回复评论集合表示为 $C=\{c_1,c_2,...,c_{|C|}\}$,用 t_i 表示每条评论 c_i 发布的时间,将 t_i 按照升序排序得到评论时间序列为 $\mathcal{T}=\{t_1,t_2,...,t_{|\mathcal{T}|}\}$。

谣言检测任务的目标是学习一个函数 $p(c|m_i,C,\mathcal{T},\mathcal{U},\mathcal{W};\theta)$ 来预测社交媒体上发布的帖子是否为谣言,c 为标签类别,θ 表示模型的参数。

2.2 模型介绍

本文提出的RD-CSGU模型主要包含两个模块:1)评论序列文本特征和时间周期特征(Comment Sequence Text Features and Time Periodic Features, CS-TFAPF);2)聚合传播图的拓扑结构与生成的节点特征(Topological Structure of Propagation Graphs and Generated Node Features, TSPG-GNF)。模型总体结构如图1所示,CS-TFAPF模块使用门控机制融合评论序列的文本语义特征 $\hat{c}$ 和时间周期特征 $\tilde{c}$ 来学习表征 $\tilde{c}$,TSPG-GNF模块使用GAT聚合传播拓扑结构和隐私保护环境下生成的节点特征来学习表征 z,最后将表征 $\tilde{c}$ 和表征 z 连接输入全连接层进行谣言检测。CS-TFAPF以独立的帖子单点扩散为出发点,充分发掘评论序列的二义性,TSPG-GNF以全局网络中帖子之间相互影响为出发点,使用GAN生成方式解决访问用户资料受限问题,最后不同角度下的表征共同对分类任务发挥作用。

2.3 CS-TFAPF模块

在CS-TFAPF模块中,RD-CSGU模型一方面利用CNN和注意力机制提取评论序列文本语义特征 $\hat{c}$,另一方面利用LSTM捕获评论序列时间周期特征 $\tilde{c}$,然后将文本语义特征和时间周期特征融合获取表征 $\tilde{c}$。

2.3.1 提取文本语义特征

假设每条评论文本都有 L 个词,长度小于 L 时,在文本开始处填充零直至长度达 L,长度大于 L,则在位置结束处截断评论。在预处理组件word2vec生成的嵌入词库中查询词索引,得到对应的词嵌入 $(\boldsymbol{x}_1;\boldsymbol{x}_2;\cdots;\boldsymbol{x}_L)\in\mathbb{R}^{L\times d}$,$\boldsymbol{x}_i\in\mathbb{R}^d$ 定义为评论 c_i 中第 i 个词对应的 d 维词嵌入,将单词嵌入输入到CNN组件中,首先通过式(1)得到捕获域内具有上下词关系的特征 $\boldsymbol{e}=[e_1,e_2,...,e_{L-h+1}]\in\mathbb{R}^{L-h+1}$,然后将特征 $\boldsymbol{e}$ 输入到最大值池化层计算输出 $\hat{e}=\max(\boldsymbol{e})$。

$$e_i=\sigma(\boldsymbol{W}\cdot\boldsymbol{x}_{i:i+h-1}) \tag{1}$$

其中:$\boldsymbol{W}\in\mathbb{R}^{h\times d}$ 是大小为 h 的卷积核,$\sigma(\cdot)$ 是非线性变化函数。在CNN组件的卷积层中分别利用 $h=3,4,5$ 的卷积核来构造三个捕获域,从每个捕获域中得到 $d/3$ 维度的特征,将三个捕获域的输出连接成 $\boldsymbol{e}\in\mathbb{R}^d$。分别将帖子和回复该帖子的每条评论输入CNN组件执行该过程,得到帖子的文本表示 $\boldsymbol{m}_i\in\mathbb{R}^d$ 和每条评论的文本表示 $\boldsymbol{c}_i\in\mathbb{R}^d i\in[1,n]\in\mathbb{R}^{n\times d}$。将每条评论的文本表示堆叠到一起形成评论矩阵 $\boldsymbol{C}=[\boldsymbol{c}_1;\boldsymbol{c}_2;\cdots;\boldsymbol{c}_n]$,接着通过多头注意力模块得到 $\hat{\boldsymbol{C}}=\text{MultiHeadAttention}(\boldsymbol{C},\boldsymbol{C},\boldsymbol{C})$,然后使用交叉注意模块来评估每一条评论与帖子之间的权重关系,使用式(2)计算每条评论与帖子 $\boldsymbol{m}_i$ 的注意力分数 $\boldsymbol{s}\in\mathbb{R}^{n\times 1}$,$\boldsymbol{A}$ 表示注意力运算矩阵,根据注意力分数结合回复

评论矩阵获得文本语义特征 $\hat{\boldsymbol{c}}$。

$$\begin{cases} \boldsymbol{s} = \mathrm{softmax}\left(\hat{\boldsymbol{C}}\boldsymbol{A}\boldsymbol{m}_i^{\mathrm{T}}\right), \\ \hat{\boldsymbol{c}} = \boldsymbol{s}^{\mathrm{T}}\hat{\boldsymbol{C}} \end{cases} \tag{2}$$

2.3.2 提取时间周期特征

对评论序列的时间 $\mathcal{T}=\left\{t_1, t_2, ..., t_{|\mathcal{T}|}\right\}$ 进行reshape操作，使得每个输入LSTM中的 $\boldsymbol{x}_t$ 对应每条评论的时间 t_t，LSTM神经网络表达式为：

$$\begin{cases} \boldsymbol{i}_t = \sigma\left(\boldsymbol{W}_{\mathrm{i}} \cdot \left[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t\right] + \boldsymbol{b}_{\mathrm{i}}\right), \\ \tilde{\boldsymbol{S}}_t = \tanh\left(\boldsymbol{W}_{\mathrm{c}} \cdot \left[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t\right] + \boldsymbol{b}_{\mathrm{c}}\right), \\ \boldsymbol{S}_t = \sigma\left(\boldsymbol{W}_{\mathrm{f}} \cdot \left[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t\right] + \boldsymbol{b}_{\mathrm{f}}\right) * \boldsymbol{S}_{t-1} + \boldsymbol{i}_t * \tilde{\boldsymbol{S}}_t, \\ \boldsymbol{h}_t = \sigma\left(\boldsymbol{W}_{\mathrm{o}}\left[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t\right] + \boldsymbol{b}_{\mathrm{o}}\right) * \tanh\left(\boldsymbol{S}_t\right) \end{cases} \tag{3}$$

其中：$\boldsymbol{W}_{\mathrm{i}}$、$\boldsymbol{W}_{\mathrm{c}}$、$\boldsymbol{b}_{\mathrm{i}}$、$\boldsymbol{b}_{\mathrm{c}}$ 为更新门的参数权重和偏置，$\boldsymbol{W}_{\mathrm{f}}$、$\boldsymbol{b}_{\mathrm{f}}$ 为遗忘门的参数权重和偏置，$\boldsymbol{W}_{\mathrm{o}}$、$\boldsymbol{b}_{\mathrm{o}}$ 为结果输出的参数权重和偏置，$\sigma(\cdot)$ 为激活函数。更新门选择有多少信息需要保存并更新当前时刻，遗忘门选择忘记旧状态 $\boldsymbol{S}_{t-1}$ 的一部分，输入控制选择添加候选状态 $\tilde{\boldsymbol{S}}_t$ 的一部分得到新状态 $\boldsymbol{S}_t$，$\boldsymbol{h}_{t-1}$ 为前一时刻的输出值，式(3)最后输出当前时刻特征 $\boldsymbol{h}_t$。将LSTM输出的特征 $\boldsymbol{h}_t$ 输入式(4)所表示全连接层，获取评论序列的时间周期特征 $\breve{\boldsymbol{c}}$。

$$\breve{\boldsymbol{c}} = \mathrm{softmax}\left(\boldsymbol{W}\boldsymbol{h}_t + \boldsymbol{b}\right) \tag{4}$$

2.3.3 融合评论序列特征

利用式(5)的融合门机制，将评论序列的文本语义特征 $\hat{\boldsymbol{c}}$ 和时间周期规律特征 $\breve{\boldsymbol{c}}$ 进行融合，从而得到表征 $\tilde{\boldsymbol{c}}$。

$$\tilde{\boldsymbol{c}} = \hat{\boldsymbol{c}} \odot \alpha + \breve{\boldsymbol{c}} \odot (1-\alpha);\ \alpha = \sigma\left(\boldsymbol{w}_1\hat{\boldsymbol{c}} + \boldsymbol{w}_2\breve{\boldsymbol{c}} + \boldsymbol{b}\right) \tag{5}$$

其中：$\sigma(\cdot) = \dfrac{1}{1+\exp(\cdot)}$ 是sigmoid激活函数，$\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathbb{R}^{d\times 1}$ 和 $\boldsymbol{b}$ 表示融合门的可学习参数。

2.4 TSPG-GNF模块

在TSPG-GNF模块中，RD-CSGU模型借助于GAN在构建的传播异质网络上生成缺失的节点特征 $\boldsymbol{F}$，然后使用GAT聚合传播拓扑结构 $\boldsymbol{A}$ 与生成的节点特征 $\boldsymbol{F}$ 得到特征矩阵 $\boldsymbol{Z}$，从 $\boldsymbol{Z}$ 中获取帖子 m_i 节点的表征 $\boldsymbol{z}$ 作为输出。

2.4.1 构建传播异质网络

本文将谣言传播过程中的参与对象和相关转发关系构建为具有3种类型节点 $\mathcal{V}=\{\mathcal{M}, \mathcal{U}, \mathcal{W}\}$ 和4种连接关系 $\Re=\{r_1, r_2, r_3, r_4\}$ 的带权传播异质网络(Propagation Heterogeneous Network，PHN)，其中，$\mathcal{M}$ 表示帖子节点，$\mathcal{U}$ 表示用户节点，$\mathcal{W}$ 表示全局共现的词节点。连接关系 r_1 表示用户发布帖子行为，权重设置为1。连接关系 r_2 表示用户转发帖子行为，权重是用户转发帖子时刻相距帖子发布时刻的倒数，数值越大表明转发时间越早、关联越深。连接关系 r_3 表示帖子 m_i 文本中包含的全局共现词，权重是词频率和逆文档频率(Term Frequency and Inverse Document Frequency，TF-IDF)的比率，词频率是词出现的次数，逆文档频率是帖子总数和包含这个词的帖子数量的比率，用式(6)计算一条帖子 i 和一个词 j 的TF-IDF值。

$$\begin{aligned} &TF\text{-}IDF_{ij} = TF_{ij} \times IDF_j \\ &TF_{ij} = \frac{n_{ij}}{\sum_k n_{ik}} \\ &IDF_j = \log\frac{|\Gamma|}{\left|\left\{k: w_j \in t_k\right\}\right|} \end{aligned} \tag{6}$$

其中：n_{ij} 表示词 j 在一条帖子 i 中出现的次数，$|\Gamma|$ 为帖子总数，$\left|\left\{k: w_j \in t_k\right\}\right|$ 为包含词 j 的帖子数。

连接关系 r_4 表示在不同帖子中出现的全局共现词，在数据集中的所有帖子上设置一个固定大小的滑动窗口来收集词的共现统计，使用词关联度量点向互信息(Point Mutual Information，PMI)[23]来计算权重，以式(7)计算词对 i,j 之间的PMI值。

$$\begin{aligned} PMI(i,j) &= \log\frac{p(i,j)}{p(i)p(j)} \\ p(i,j) &= \frac{\#W(i,j)}{\#W} \\ p(i) &= \frac{\#W(i)}{\#W} \end{aligned} \tag{7}$$

其中：$\#W(i,j)$ 表示滑动窗口同时包含词 i 和词 j 的数量，$\#W$ 表示滑动窗口的数量，$\#W(i)$ 表示包含词 i 的滑动窗口的数量。

2.4.2 GAN生成节点特征

GAN利用PHN上不同类型节点之间的不同连接语义关系 $\{r_1, r_2, r_3, r_4\}$ 区分鉴别器和生成器样本的真实性，如图1中TSPG-GNF模块的传播异质网络所示，给定一个帖子 m_2 以及一个关系 r_1，形成三元组样本 (m_2, u_2, r_1)，鉴别器能够根据 (m_2, u_i, r_1) 区分转发的用户 u_3 和发布的用户 u_2 这2个不同的节点特征，生成器将尝试产生 (m_2, u_2, r_1) 和 (m_2, u_3, r_2) 约束下 u_2 和 u_3 节点特征，使得 u_2 和 u_3 节点的特征具有在没有拓扑结构直连下的社交联系以及与帖子之间的关系比重，而非简单认为它们的节点特征整体上服从正态分布。在迭代训练后，生成器的节点特征由服从正态分布转化为带有PHN上丰富和复杂语义关系的节点特征分布，生成的用户节点特征表达倾向于传播谣言的用户属性，帖子节点处于这些相似用户节点的社交圈中，形成不同类别的帖子节点特征。

鉴别器(Discriminator，D)评估节点对 u 和 v 之间的连通关系的正确性，$u, v \in \mathcal{V}$ 为给定节点，$r \in \Re$ 为来自PHN的给定关系 $\Re=\{r_1, r_2, r_3, r_4\}$，$\boldsymbol{e}_v$ 是样本节点的嵌入，鉴别器输出样本 v 在关系 r 下连接到 u 的概率，即 $D\left(\boldsymbol{e}_v \middle| u, r;\ \theta^{\mathrm{D}}\right)$，使用式(8)得到量化的概率数值。

$$D\left(\boldsymbol{e}_v \middle| u, r;\ \theta^{\mathrm{D}}\right) = \frac{1}{1+\exp\left(-\boldsymbol{e}_u^{\mathrm{D}}\boldsymbol{M}_r^{\mathrm{D}}\boldsymbol{e}_v\right)} \tag{8}$$

其中：$\boldsymbol{e}_v \in \mathbb{R}^{d\times 1}$ 是样本 v 的输入嵌入，$\boldsymbol{e}_u^{\mathrm{D}} \in \mathbb{R}^{1\times d}$ 是节点 u 的可学习嵌入，$\boldsymbol{M}_r^{\mathrm{D}} \in \mathbb{R}^{d\times d}$ 是关系 r 的可学习关系矩阵，$\theta^{\mathrm{D}} = \left\{\boldsymbol{e}_u^{\mathrm{D}};\ u \in \mathcal{V}, \boldsymbol{M}_r^{\mathrm{D}};\ r \in \Re\right\}$ 是鉴别器的模型参数。当 v 是通过 r 与 u 相关的正样本时，应该是高概率，当它是负样本时，应该是低概率。鉴别器考虑关系类型时，通过采样单元或生成器以三元组 (u, v, r) 形式得到以下三种样本：

1) 在给定的关系下进行连接，节点 u 和 v 确实通过PHN上的正确关系 r 连接，如 (u_2, m_2, r_1)，这样的三元组被认为是正样本的，用损失函数 ς_1^{D} 来更新训练。

$$\varsigma_1^{\mathrm{D}} = \mathbb{E}_{(u,v,r)\sim \mathrm{P}_{\mathcal{G}}} - \log D\left(\boldsymbol{e}_v^{\mathrm{D}} \middle| u, r\right) \tag{9}$$

2) 在不正确的关系下进行了连接，u 和 v 以一个错误的关系 $r' \neq r$ 连接，如 (u_2, m_2, r_2)，期望鉴别器将这种样本标记为负，因为它们的连通性与给定关系所携带的期望语义不匹配，用损失函数 ς_2^{D} 来更新训练。

$$\varsigma_2^{\mathrm{D}} = \mathbb{E}_{(u,v)\sim \mathrm{P}_{\mathcal{G}}, r'\sim \mathrm{P}_{\Re'}} - \log D\left(1 - D\left(e_v^{\mathrm{D}} \middle| u, r'\right)\right) \tag{10}$$

3) 来自生成器的假节点，给定一个节点 $u \in \mathcal{V}$ 和正确连接关系 $r \in \Re$，它可以与生成器提供的节点 v 形成一个假节点对 (u, v, r)，生成器试图生成一个假节点 v 的嵌入，模拟在正确的关系 r 下连接到 u 的真实节点，用损失函数 ς_3^{D} 来更新训练。

$$\varsigma_3^{\mathrm{D}} = \mathbb{E}_{(u,v)\sim \mathrm{P}_{\mathcal{G}}, \boldsymbol{e}_v'\sim G\left(u,r;\ \theta^{\mathrm{G}}\right)} - \log D\left(1 - D\left(\boldsymbol{e}_v^{\mathrm{D}} \middle| u, r\right)\right) \tag{11}$$

综合上述 $\varsigma_1^{\mathrm{D}}, \varsigma_2^{\mathrm{D}}, \varsigma_3^{\mathrm{D}}$ 训练鉴别器：

$$\varsigma^{\mathrm{D}} = \varsigma_1^{\mathrm{D}} + \varsigma_2^{\mathrm{D}} + \varsigma_3^{\mathrm{D}} + \lambda^{\mathrm{D}}\left\|\theta^{\mathrm{D}}\right\|_2^2 \tag{12}$$

其中：$\boldsymbol{e}_v^{\mathrm{D}}$ 表示输入到鉴别器中的节点向量，$\lambda^{\mathrm{D}} > 0$ 为控制正则化项 ς^{D}，以避免过拟合，通过最小化 ς^{D} 优化鉴别器的参数 θ^{D}。

生成器(Generator, G)的目标是生成一个在关系r上下文中可能连接到u的假节点v,v应该尽可能地接近真实的节点。比如给定节点m_2和关系r_3,生成的节点嵌入w_i尽可能使得$(m_2, w_i, r_3) \sim \mathrm{P}_{\mathcal{G}}$,即$G(u, r;\ \theta^{\mathrm{G}})$。生成器使用式(13)从正态分布中生成样本,以实现关系感知和泛化能力。

$$\mathrm{N}\left(\boldsymbol{e}_u^{\mathrm{G}}\boldsymbol{M}_r^{\mathrm{G}}, \sigma^2\boldsymbol{I}\right) \tag{13}$$

其中:$\boldsymbol{e}_u^{\mathrm{G}} \in \mathbb{R}^{1\times d}$和$\boldsymbol{M}_r^{\mathrm{G}} \in \mathbb{R}^{d\times d}$表示生成器节点$u \in \mathcal{V}$的嵌入和关系$r \in \Re$的矩阵,对于某些$\sigma$的选择,使得正态分布中的均值为$\boldsymbol{e}_u^{\mathrm{G}}\boldsymbol{M}_r^{\mathrm{G}}$,协方差为$\sigma^2\boldsymbol{I} \in \mathbb{R}^{d\times d}$,直观地说,均值表示一个通过关系$r$可能连接到$u$的伪节点,协方差表示潜在的偏差。将多层感知器(Multi-Layer Perceptron, MLP)集成到生成器中,得到的生成器形式化表达为式(14),增强了产生的假样本表达能力。

$$G(u, r;\ \theta^{\mathrm{G}}) = f\left(\boldsymbol{W}_n \ldots f\left(\boldsymbol{W}_2 f\left(\boldsymbol{W}_1\boldsymbol{e}_1 + \boldsymbol{b}_1\right) + \boldsymbol{b}_2\right) + \boldsymbol{b}_n\right) \tag{14}$$

其中:从分布$\mathrm{N}\left(\boldsymbol{e}_u^{\mathrm{G}}\boldsymbol{M}_r^{\mathrm{G}}, \sigma^2\boldsymbol{I}\right)$得到$\boldsymbol{e}$。这里$\boldsymbol{W}_*$和$\boldsymbol{b}_*$分别表示每一层的权值矩阵和偏置向量,$f$为激活函数。因此,生成器的参数集为$\theta^{\mathrm{G}} = \left\{\boldsymbol{e}_u^{\mathrm{G}}: u \in \mathcal{V}, \boldsymbol{M}_r^{\mathrm{G}}: r \in \Re, \boldsymbol{W}_*, \boldsymbol{b}_*\right\}$,即所有节点嵌入和关系矩阵的集合,以及MLP的参数。

生成器的目标是通过生成接近真实的伪样本来欺骗鉴别器,从而获得高分。为此,使用式(15)的损失函数来训练和更新生成器。

$$\varsigma^{\mathrm{G}} = \mathbb{E}_{(u,r)\sim \mathrm{P}_{\mathcal{G}}, e_v' \sim G(u, r;\ \theta^{\mathrm{G}})} - \log D\left(\boldsymbol{e}_v' | u, r\right) + \lambda \left\| \theta^{\mathrm{G}} \right\|_2^2 \tag{15}$$

其中:$\lambda^{\mathrm{G}} > 0$控制正则化项,通过最小化ς^{G}来对生成器的参数θ^{G}进行优化。

2.4.3 聚合传播结构特征

结合生成的节点特征和传播过程中以用户转发帖子路径形成的拓扑结构作为GAT的输入,GAT通过收集本地邻居节点的嵌入来更新每个节点的表示,以待检测的帖子为中心,在每次迭代中,GAT汇聚生成的用户节点特征和不同帖子间关联的全局'共现词'节点特征,GAT使用式(16)计算邻居节点的注意力分数。

$$\alpha_{jih}' = \mathrm{softmax}\left(\mathrm{LeakyReLU}\left(\boldsymbol{a}^{\mathrm{T}}\left[\boldsymbol{m}_j;\ \boldsymbol{u}_i;\ \boldsymbol{w}_h\right]\right)\right) \tag{16}$$

其中:$\boldsymbol{a} \in \mathbb{R}^{3d\times 1}$为可学习参数,非线性函数LeakyReLU参数为0.2。

在迭代过程中,当前节点逐渐从图的多跳中获得越来越多的信息,为了从不同角度的关系中捕获多种表示,捕获谣言传播过程方式也扩展了图注意力机制采用多头范式,使用K个独立的图注意力机制执行式(17)的变换,然后将它们的输出特征连接得到帖子节点传播结构表征$\boldsymbol{z}$。

$$\boldsymbol{z} = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{i\in|\mathcal{M}|, h\in|\mathcal{W}|} \alpha_{jih}^k \boldsymbol{W}_u^k\left[\boldsymbol{u}_i;\ \boldsymbol{w}_h\right]\right) \tag{17}$$

其中:$\Vert$表示连接操作,$\sigma(\cdot)$表示激活函数ELU。

2.5 谣言分类

获得CS-TFAPF模块和TSPG-GNF模块的输出表示$\tilde{\boldsymbol{c}}$、$\boldsymbol{z}$之后,使用式(18)将最终的表示形式投影到概率分类的目标空间中。

$$p\left(c | m_i, C, \mathcal{T}, \mathcal{U}, \mathcal{W};\ \theta\right) = \mathrm{softmax}\left(\boldsymbol{W}\left[\tilde{\boldsymbol{c}};\ \boldsymbol{z}\right] + \boldsymbol{b}\right) \tag{18}$$

在谣言分类模型的训练过程中,使用交叉熵损失作为优化目标函数来优化模型参数。

$$\varsigma = -\sum_i y_i \log p_i\left(c | m_i, C, \mathcal{T}, \mathcal{U}, \mathcal{W};\ \theta\right) \tag{19}$$

其中:y_i为谣言类的预测概率,θ为模型的所有参数。

3 实验与结果分析

为了全面地评估提出的RD-CSGU模型,本文将通过对比实验、消融实验、GAN生成嵌入分析、训练时间和空间分析、超参数分析回答5个问题:

问题1 与RD-CSGU模型所用特征相关的其他谣言检测模型相比,其是否能够获得较好的谣言检测性能?

问题2 RD-CSGU模型中各模块对于谣言检测的性能是否有贡献、对数据有什么依赖?

问题3 GAN生成的用户节点特征能否表达出谣言传播网络中具有相似社交选择行为的用户隐私属性?

问题4 RD-CSGU模型在时间和空间方面的优劣,适用于什么场景?

问题5 RD-CSGU模型中不同的超参数取值如何影响检测性能?

3.1 实验数据和设置

3.1.1 数据集

本文实验使用了英文Twitter15、Twitter16和中文Weibo数据集。其中Twitter15和Twitter16数据集由Ma等[20]从Twitter平台上抓取整理,分别包含1490和818条谣言数据,并在辟谣网站(snopes.com和Emergent.info)上获取数据的标签信息,包括4个标签:非谣言(Non-Rumor, NR)、验证为假的谣言(False Rumor, FR)、验证为真的谣言(True Rumor, TR)和未经验证的谣言(Unverified Rumors, UR)。同样由Ma等[24]收集整理的Weibo数据集来自Weibo平台,是当前使用最多的中文谣言数据集,一共有4664条谣言数据,并通过新浪微博举报处理大厅抓取数据的标签信息,包括2个标签:FR和TR。表1显示了三个数据集的统计信息。

表1 数据集统计信息

Tab. 1 Data Set Statistics

数据集	帖子数量	用户数量	评论数量	NR	FR	UR	TR
Twitter15	1490	276,663	331,612	374	370	374	372
Twitter16	818	173,487	204,820	205	205	203	205
Weibo	4664	2,746,818	3,805,656	0	2,313	0	2351

3.1.2 评价指标和实验设置

本文在Twitter15、Twitter16数据集上使用准确率(Accuracy)和非谣言(NR)、假谣言(FR)、真谣言(TR)、未经证实的谣言(UR)的F1分数作为评价指标。在Weibo数据集上使用准确率(Accuracy)和假谣言(FR)、真谣言(TR)的精度(Precision)、召回率(Recall)和F1分数作为评价指标。

随机选择10%的样本作为验证数据集,并将其余的样本以3:1的比例分配给数据集的训练集和测试集。使用预训练的词向量库Tencent AI Lab Chinese and English Term Embedding Corpora分别提取中文和英文的词向量表示,词向量维度都是200。Dropout设置为0.5,多头注意力头数设置为8,使用Adam优化器进行训练,学习率设置为0.001,在Twitter15、Twitter16数据集上batch_size=16,在Weibo数据集上batch_size=64。在Twitter15,Twitter16和Weibo数据集上模型采用相同的实验参数。

GAN在总体检测模型训练之前单独进行博弈训练,生成器提供传播结构中的帖子特征向量表示和用户特征向量表示,特征向量维度都是300。生成器初始正态分布的方差$\sigma^2 = 1.0$,鉴别器考虑正确连接关系、错误连接关系、生成器生成的假连接关系,每次迭代中采样单元以三元组(u, v, r)形式在PHN上随机采集每个节点包含上述连接关系的10个正样本用于训练,生成器和鉴别器的学习率设置为0.0001,batch_size=128。

实验训练基于PyTorch开源工具实现,采用单卡NVIDIA[RTX 3090]的GPU,32核CPU、128GB内存。

3.2 对比实验

3.2.1 基线模型

为了验证RD-CSGU模型的性能,本文选择了7个与RD-CSGU模型所用特征相关的其他谣言检测模型进行对比,大致可分为以下3种类别:

1) 基于时间序列建模的方法,探究这些方法与本文利用的时序信息在检测性能上的差异。

GRU(Gate Recurrent Unit)[24]:一种基于RNN的深度学习模型,从用户评论中学习时间语言模式,对序列结构特征进行建模。

PPC(Propagation Path Classification)[25]：一种在时间序列上联合使用RNN和CNN组成的传播路径分类器，通过谣言传播路径上的用户特征得到谣言表示。

PLAN(Post-Level Attention Network)[26]：一种基于Transformer结构的谣言表示学习模型，它将谣言传播过程建模为时间序列，并利用注意力机制学习到序列化的帖子间较长距离的依赖关系。

2）基于传播树结构的检测方法，考察本文构建的传播异质网络与传播树状结构在聚合传播结构信息时不同的表现。

PTK(Propagation Tree Kernel)[20]：一种带有传播树核的支持向量机(Support Vector Machines，SVM)分类器，它通过核方法捕获传播树结构之间的相似性，从而进行谣言分类。

Bi-GCN(Bi-directional GCN)[21]：一种基于GCN的谣言传播树表示学习模型，从自顶向下和自底向上两个方向对谣言的传播结构进行联合建模，模拟谣言的传播和扩散过程。

3）基于联合多特征的检测方法，验证在多特征情况下，GAN生成节点特征对比正态分布初始化节点特征的优势。

GLAN(Global-Local Attention Network)[8]：一种基于局部-全局的注意力网络模型，联合编码帖子与评论的局部语义和全局传播结构信息。

MiSTR (Multiview Structural-Temporal Learning Framework for Rumor Detection)[27]：一种组合的谣言检测多视图结构-时间学习框架，共同学习转发动态的时间特征、传播图的结构特征和帖子的文本特征。

表2 Twitter15和Twitter16数据集上对比结果

Tab. 2 Comparison results between Twitter15 and Twitter16 Datasets

模型	Twitter15					Twitter16				
	Acc	NR-F1	FR-F1	TR-F1	UR-F1	Acc	NR-F1	FR-F1	TR-F1	UR-F1
GRU	0.646	0.792	0.574	0.608	0.592	0.633	0.772	0.489	0.686	0.593
PPC	0.842	0.811	0.875	0.818	0.790	0.863	0.820	0.898	0.843	0.837
PLAN	0.799	0.754	0.521	0.836	0.799	0.816	0.761	0.853	0.870	0.774
PTK	0.667	0.619	0.669	0.772	0.645	0.863	0.820	0.898	0.843	0.837
Bi-GCN	0.836	0.816	0.870	0.866	0.786	0.839	0.838	0.855	0.844	0.819
GLAN	0.905	0.924	**0.917**	0.852	0.927	0.902	0.921	0.869	0.847	**0.968**
MiSTR	0.862	0.848	0.885	0.861	0.854	0.865	0.849	0.882	0.878	0.852
RD-CSGU	**0.914**	**0.939**	0.905	**0.878**	**0.933**	**0.924**	**0.936**	**0.907**	**0.915**	0.936

表中加粗的字体表示评价指标的最优值

表3 Weibo数据集上对比结果

Tab. 3 Comparison results on Weibo dataset

模型	Acc	FR-Precision	FR-Recall	FR-F1	TR-Precision	TR-Recall	TR-F1
GRU	0.910	0.952	0.864	0.906	0.876	0.956	0.914
PPC	0.921	0.949	0.889	0.918	0.896	0.962	0.923
PLAN	0.922	0.920	0.930	0.911	0.923	0.914	0.932
PTK	0.891	0.907	0.868	0.887	0.876	0.913	0.894
Bi-GCN	0.924	0.923	0.927	0.919	0.925	0.921	0.929
GLAN	0.946	0.949	0.943	0.946	0.943	0.948	0.945
MiSTR	0.947	0.921	**0.977**	0.948	**0.976**	0.917	0.946
RD-CSGU	**0.964**	**0.957**	0.971	**0.964**	0.971	**0.957**	**0.964**

表中加粗的字体表示评价指标的最优值

3.2.2 结果分析

为了回答问题1，即与RD-CSGU模型所用特征相关的其他谣言检测模型相比，其是否能够获得较好的谣言检测性能，本文通过对比实验得到了表2和表3所示的对比模型和RD-CSGU模型在3个数据集上分类的总体准确率(Acc%)和各类别的F1值，对表2-3中的实验结果进行分析：

1）在基于时间序列的方法(GRU、PPC、PLAN)中，PPC模型由于同时使用了RNN和CNN结构，Acc指标在Twitter15、Twitter16和Weibo数据集上比GRU模型分别提升了19.6个百分点、23个百分点、1.1个百分点，而PLAN模型关注到了帖子之间的隐式关系，在检测效果上得到了进一步提升。RD-CSGU模型Acc指标在Twitter15、Twitter16和Weibo数据集上比基于时间序列的模型分别提升了7.2个百分点、6.1个百分点、4.2个百分点，这归因于模型在学习评论基于时序特征的同时捕捉了评论序列的文本语义特征。

2）与基于传播树结构的方法(PTK、Bi-GCN)相比，RD-CSGU模型将谣言传播过程构建为异质网络，利用了节点与边的异质性，在Twitter15、Twitter16和Weibo数据集上Acc指标提升了7.8个百分点、6.1个百分点、4个百分点。

3）与利用多特征的方法(GLAN、MiSTR)相比，RD-CSGU模型通过GAN生成用户隐私特征反映了传播过程中相似用户的社交关系，Acc指标在Twitter15、Twitter16和Weibo数据集上相比于使用正态分布初始化传播节点特征的GLAN模型分别提升了0.9个百分点、2.2个百分点、1.8个百分点。

本文提出的RD-CSGU模型在3个数据集上均取得了良好的性能表现，RD-CSGU模型的Acc指标在Twitter15、Twitter16和Weibo数据集上分别比最佳对比模型高了0.9个百分点、2.2个百分点和1.7个百分点。结合上述分析，RD-CSGU模型在使用评论序列的文本语义信息之外加入了时间周期信息，并且使用GAN生成用户隐私特征，有效提升谣言检测性能。

3.3 消融实验

为了回答问题2，即RD-CSGU模型中各模块对于谣言检测的性能是否有贡献、对数据有什么依赖，本文设计了3种变体与RD-CSGU模型进行对比，3种变体描述如下：

1）变体1：从RD-CSGU模型中去除评论序列的文本语义模块，利用评论序列的时间周期模块和GAN生成初始节点特征的传播结构模块进行谣言检测，验证评论序列的文本语义模块有效性。

2）变体2：从RD-CSGU模型中去除评论序列的时间周期模块，利用评论序列的文本语义模块和GAN生成初始节点特征的传播结构模块进行谣言检测，验证评论序列的时间周期模块有效性。

3）变体3：从RD-CSGU模型中去除GAN生成节点特征，利用评论序列的文本语义模块、时间周期模块和正态分布初始化节点特征的传播结构模块进行谣言检测，验证GAN生成节点特征模块有效性。

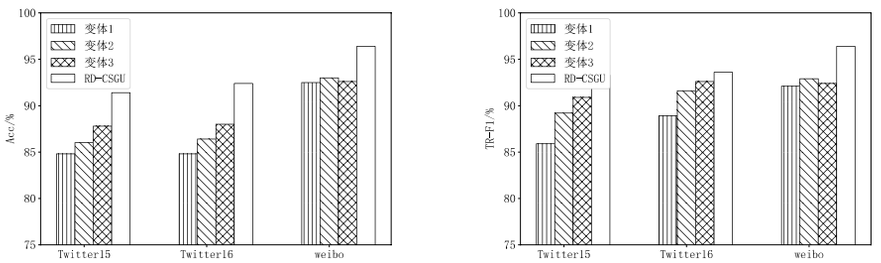
3种变体与RD-CSGU模型在Twitter15、Twitter16和Weibo数据集上的消融实验结果如图2所示，可以观察到：

1）变体1的谣言检测性能在3个数据集上的Acc指标和TR-F1值明显下降，表明评论序列的文本语义模块在模型进行检测时所作的贡献较大，其更容易将带有煽动、蛊惑性词语的谣言帖子及其评论进行划分和判定。但Acc指标和F1值在Twitter15和Twitter16数据集上仍保持在80%以上，在Weibo数据集上仍保持在90%以上，表明剩余的特征共同组合起来仍可以有效地检测谣言。

2）变体2的Acc指标在评论规模较小的Twitter15和Twitter16数据集上检测性能下降较大，但在评论规模较大的Weibo数据集上波动较小，这可能是评论序列的时间周期模块使用的LSTM在提取谣言事

件随时间发展过程中面对不断周期性涌现的评论时间周期特征产生了过拟合,导致评论序列的时间周期模块面对规模较大的数据集时检测性能下降。

3) 变体3的Acc指标和TR-F1值下降较小,但明显低于RD-CSGU模型的性能,这说明GAN生成节点特征模块对模型整体性能起增强效果。因此,RD-CSGU模型中评论序列文本模块、评论序列周期模块、GAN生成特征模块在进行谣言检测时都起到了不同程度的贡献。



(a) 各变体与RD-CSGU模型的Acc%　　(b) 各变体与RD-CSGU模型的TR-F1%

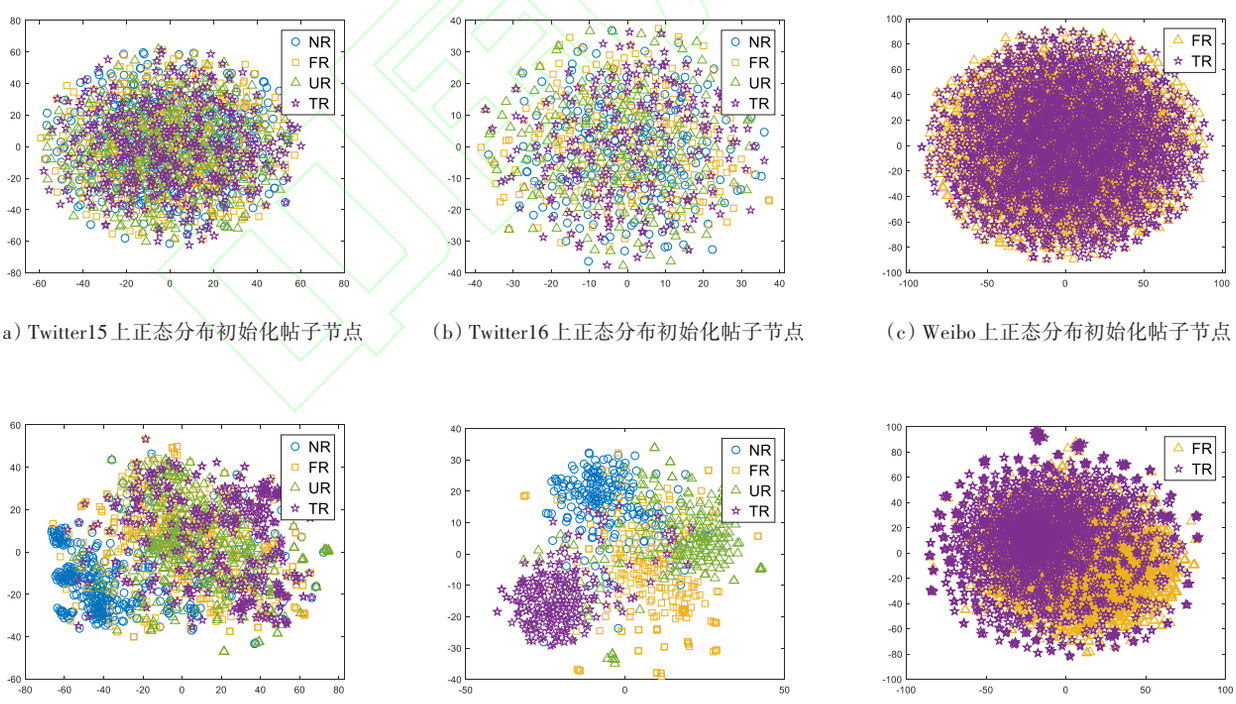
图2　RD-CSGU在3个数据集上的消融实验结果

Fig. 2　Ablation results of RD-CSGU on three datasets

3.4　GAN生成嵌入分析

为了回答问题3,即GAN生成的用户节点特征能否表达出谣言传播网络中具有相似社交选择行为的用户隐私属性,本文在用户节点数量规模较大且无先验标签的情况下,分别利用正态分布初始化的帖子节点特征和GAN生成的帖子节点特征进行t-SNE投影对比,这是因为具有相似社交选择行为的用户形成不同范围的社交圈,在同一个相似用户社交圈中GAN生成的帖子节点特征会趋向于形成同一个簇。在Twitter15、Twitter16数据集上以帖子自带的标签NR、FR、TR、UR作为每一条帖子节点特征的投影标识,在图3中使用4种形状分别表示4个标识;在Weibo数据集上以帖子自带的标签FR、TR作为每一条帖子节点特征的投影标识,在图3中使用2种形状分别表示。

如图3(a)~(c)所示,当用户和帖子的初始隐私数据不可访问时,在3个数据集上使用正态分布初始化的帖子节点特征投影显示无法相互区分。相比之下,如图3(d)~(f)所示,在Twitter15、Twitter16数据集上通过GAN生成的帖子节点特征投影显示同一个簇的节点在一定程度上形成4个紧密的区域,在Weibo数据集上通过GAN生成的帖子节点特征投影显示同一个簇的节点一定程度上形成2个紧密的区域。这验证了利用GAN生成的节点特征具有一定程度的相似d维生成嵌入,回答了GAN生成的用户节点特征能够表达出谣言传播网络中具有相似社交选择行为的用户隐私属性。



(a) Twitter15上正态分布初始化帖子节点　(b) Twitter16上正态分布初始化帖子节点　(c) Weibo上正态分布初始化帖子节点

(d) Twitter15上GAN生成帖子节点　(e) Twitter16上GAN生成帖子节点　(f) Weibo上GAN生成帖子节点

图3　传播异质网络中帖子节点特征的t-SNE投影结果

Fig. 3　t-SNE Projection Results of Node Features in Heterogeneous Network Propagation

3.5　训练时间和空间分析

为了回答问题4,即RD-CSGU模型在时间和空间方面的优劣,适用于什么场景,本文将RD-CSGU与Bi-GCN和GLAN进行对比观察,分别比较它们的参数量、一个epoch下训练消耗的时间和内存。

从表4中可以看到:1) 3个模型中,RD-CSGU的参数量最大,GLAN次之,Bi-GCN的参数量最小,这是因为RD-CSGU包含两个阶

段的参数量，而Bi-GCN仅在GCN基础上进行改进；2）Bi-GCN在3个数据集上训练时间消耗最高，这是因为GLAN和RD-CSGU将所有帖子和转发用户构建为全局网络，而Bi-GCN将用户转发帖子视为每棵独立的转发树，大量重复计算用户转发节点；3）GLAN在3个数据集上的内存消耗最低，而RD-CSGU在3个数据集上的内存消耗较高。由此可见，RD-CSGU具有较大的参数量，适中的训练时间，适用于无法获取转发、评论社交帖子用户账号资料时，机器运行内存空间较大的场景。

表4 时间和空间消耗对比结果

Tab. 4 Comparison of Time and Space Consumption Results

模型	参数量	Twitter15		Twitter16		Weibo	
		时间	内存	时间	内存	时间	内存
RD-CSGU	10958798	24.0s	31.7G	20.4s	30.0G	78.9s	42.7G
Bi-GCN	1289476	73.2s	19.9G	161.5s	17.5G	1127.3s	35.4G
GLAN	6393798	5.6s	3.4G	3.6s	2.3G	7.8s	4.7G

3.6 超参数分析

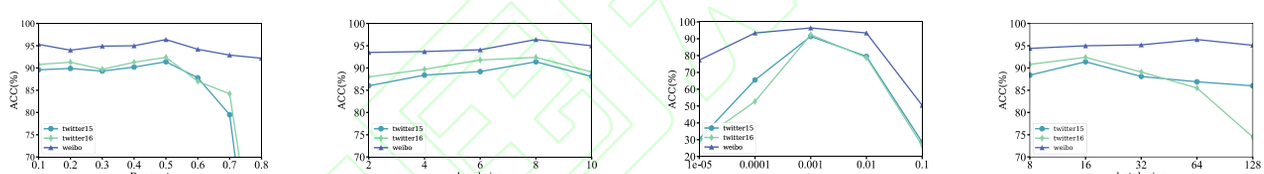
为了回答问题5，即RD-CSGU模型中不同的超参数取值如何影响检测性能，本文验证RD-CSGU模型中的4个超参数，在Twitter15、Twitter16和Weibo数据集上研究某个超参数的变化时，固定其他超参数不变，分析检测性能变化：

1）丢失率D_{Dropout}：一种在训练过程中随机删除神经元的概率。本文让D_{Dropout}在0.1到0.8之间的范围取值，从图4(a)中Acc指标可以看出，当$D_{\text{Dropout}}=0.5$时，模型取得了最好的性能；当$0.1<D_{\text{Dropout}}<0.5$时，模型性能随着学习率的增加而有所提升；但当$D_{\text{Dropout}}>0.5$时，模型性能开始出现下降趋势，$D_{\text{Dropout}}>0.8$之后模型性能呈现断崖式下降。

2）注意力头数(head-size)：利用GAT提取传播结构时使用几个方面的注意力。本文让head-size在2到10之间的范围取值，从图4(b)中Acc指标可以看出，head-size=8时，模型取得了最好的性能，这是由于注意力头数越多，各节点状态更新时融合各个视角信息，从而增强节点的传播信息表示。但当注意力头数过多时，模型过度学习不同视角的细节信息，可能引入噪声，影响模型的性能。

3）优化器的学习率γ：影响神经网络优化的速度，用来减小神经网络的损失。本文让γ在1e-5到0.1之间的范围取值，从图4(c)中Acc指标可以看出，当优化器的学习率$\gamma=10^{-3}$时，模型在3个数据集中均取得了最好的性能，优化器学习率大小设置的不同对模型的检测性能具有较大影响。

4）训练批次大小(batch_size)：表示单次传递给神经网络用以训练的数据(样本)个数。本文让batch_size在8到128之间的范围取值，从图4(d)中Acc指标可以看出，在Twitter15、Twitter16数据集上batch_size=16时效果达到最优，这是因为Twitter15、Twitter16数据集的规模较小，batch_size设置过大导致每个epoch中传递训练样本次数过少，使得模型的泛化能力下降，在Weibo数据集上batch_size=64时效果达到最优，batch_size虽然没有学习率那么敏感，但在进一步提升模型性能时，batch_size就会成为一个非常关键的参数。



(a) 丢失率不同的性能 (b) 注意力头数不同的性能 (c) 学习率不同的性能 (d) 训练批次不同的性能

图4 超参数在3个数据集上的实验结果

Fig. 4 Experimental results of hyperparameters on 3 datasets

4 结语

本文提出了一种融合评论序列二义性与生成用户隐私特征的谣言检测模型(RD-CSGU)。该模型融合了评论序列不同视角下的文本语义信息和时间周期信息来提升检测的准确性，同时在如今注重用户隐私保护的环境下，通过社交网站账号获取用户资料的方式已受到限制，为了得到转发谣言帖子的用户属性特征，RD-CSGU模型首先构建了谣言传播异质网络，然后依据该异质网络上不同的连接关系语义利用生成对抗网络生成用户隐私特征，替代了传播过程中缺失的用户节点属性特征。在三个真实数据集上进行广泛的实验，对比实验结果表明提出的RD-CSGU模型性能均优于对比基线模型，消融实验说明了融合评论序列文本语义特征和时间周期特征的重要性，并证明GAN方法生成缺失用户节点特征的有效性，解决了谣言传播过程中转发用户信息缺失问题。在未来的工作中，将在模型中尝试更多半监督或无监督学习方法以便处理社交媒体上大量未标记的谣言帖子。

参考文献(References)

[1] 徐铭达，张子柯，许小可. 基于模体度的社交网络虚假信息传播机制研究[J]. 计算机研究与发展，2021，58(07)：1425-1435.(XU M D，ZHANG Z K，XU X K. Research on the Mechanism of False Information Propagation in Social Networks based on Modularity Degree [J]. Journal of Computer Research and Development，2021，58(07)：1425-1435.)

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

[3] KWON S，CHA M，JUNG K，et al. Prominent features of rumor propagation in online social media [C]//Proceedings of the 13th IEEE International Conference on Data Mining，Piscataway：IEEE，2013：1103-1108.

[4] CHEN T，LI X，YIN H，et al. Call attention to rumors：Deep attention based recurrent neural networks for early rumor detection [C]// Proceedings of the 20th of Pacific-Asia Conference on Knowledge Discovery and Data Mining，Cham：Springer，2018：40-52.

[5] KWON S，CHA M，JUNG K. Rumor detection over varying time windows[J]，PloS one，2017，12(1)：536－548.

[6] PENG Y，WANG J. Rumor detection based on attention cnn and time series of context information [J]. Future Internet，2021，13(11)：267.

[7] HUANG Q，ZHOU C，WU J，et al. Deep structure learning for rumor detection on twitter[C]//Proceedings of the 19th International

Joint Conference on Neural Networks, Piscataway: IEEE, 2019: 1-8.

[8] YUAN C, MA Q, ZHOU W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection [C]// Proceedings of the 19th IEEE International Conference on Data Mining, Piscataway: IEEE, 2019: 796-805.

[9] NGUYEN V-H, SUGIYAMA K, NAKOV P, et al. Fang: Leveraging social context for fake news detection using graph representation [C]//Proceedings of the 29th ACM international conference on information & knowledge management, New York: ACM, 2020: 1165-1174.

[10] 周丽华, 王家龙, 王丽珍, et al. 异质信息网络表征学习综述[J]. 计算机学报, 2022, 45(01): 160-189. (ZHOU L H, WANG J L, WANG L Z, et al. A Survey of Representation Learning in Heterogeneous Information Networks [J]. Chinese Journal of Computers, 2022, 45(01): 160-189.)

[11] 蒋宗礼, 樊珂, 张津丽. 基于生成对抗网络和元路径的异质网络表示学习[J]. 计算机科学, 2022, 49(01): 133-139. (JIANG Z L, FAN K, ZHANG J L. Heterogeneous Network Representation Learning based on Generative Adversarial Networks and Meta-paths [J]. Computer Science, 2022, 49(01): 133-139.)

[12] YU F, LIU Q, WU S, et al. A Convolutional Approach for Misinformation Identification [C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence, New York: ACM, 2017: 3901-3907.

[13] LOTFI S, MIRZAREZAEE M, HOSSEINZADEH M, et al. Detection of rumor conversations in Twitter using graph convolutional networks[J]. Applied Intelligence, 2021, 51(1): 4774-4787.

[14] DE SILVA N, DOU D. Semantic oppositeness assisted deep contextual modeling for automatic rumor detection in social networks[C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, New York: ACM, 2021: 405-415.

[15] SHU K, ZHOU X, WANG S, et al. The role of user profiles for fake news detection[C]//Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, New York: ACM, 2019: 436-439.

[16] HAMDI T, SLIMI H, BOUNHAS I, et al. A hybrid approach for fake news detection in twitter based on user features and graph embedding [C]//The 16th International Conference of Distributed Computing and Internet Technology, New York: ACM, 2020: 266-280.

[17] JIANG S, CHEN X, ZHANG L, et al. User-characteristic enhanced model for fake news detection in social media [C]// Proceedings of the 8th CCF International Conference of Natural Language Processing and Chinese Computing, New York: ACM, 2019: 634-646.

[18] LU Y-J, LI C-T. GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Stroudsburg, PA: Association for Computational Linguistics, 2020: 505-514.

[19] BAI N, MENG F, RUI X, et al. Rumor detection based on a source-replies conversation tree convolutional neural net [J]. Computing, 2022, 104(1): 1155-1163.

[20] 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, Stroudsburg, PA: Association for Computational Linguistics, 2017: 708-717.

[21] BIAN T, XIAO X, XU T, et al. Rumor detection on social media with bi-directional graph convolutional networks[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence, Menlo Park, CA:AAAI, 2020: 549-556.

[22] HUANG Q, YU J, WU J, et al. Heterogeneous graph attention networks for early detection of rumors on twitter[C]//Proceedings of the 20th International Joint Conference on Neural Networks, Piscataway: IEEE, 2020: 1-8.

[23] ZHANG X, ZHANG T, ZHAO W, et al. Dual-attention graph convolutional network [C]//Proceedings of the 5th Asian Conference of Pattern Recognition, Cham: Springer, 2020: 238-251.

[24] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York: ACM, 2016: 3818-3824.

[25] LIU Y, WU Y-F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks [C]//Proceedings of the 32th AAAI Conference on Artificial Intelligence, Menlo Park, CA: AAAI, 2018: 354-361.

[26] KHOO L M S, CHIEU H L, QIAN Z, et al. Interpretable rumor detection in microblogs by attending to user interactions [C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence, Menlo Park, CA:AAAI, 2020: 8783-8790.

[27] LI J, BAO P, SHEN H, et al. MiSTR: A multiview structural-temporal learning framework for rumor detection [J]. IEEE Transactions on Big Data, 2021, 8(4): 1007-1019

This work is partially supported by National Natural Science Foundation of China Funded Projects (62062066) And Yunnan Basic Research Project (202201AS070015).

MENG Wenfan, born in 1996, M. S. candidate. His research interests include data mining and rumor detection.

ZHOU Lihua, born in 1968, Ph. D., professor. Her research interests include data mining, multi-view learning, and social network analysis.

WANG Xiaoxu, born in 1995, M. S. lecturer. Her research interests include data mining, big data technology.