

# 基于时序特征和结构特征的社交网络谣言检测方法

卫玲蔚<sup>1,3</sup>, 胡斗<sup>1,2,3</sup>, 鲍祎楠<sup>1,3</sup>, 周薇<sup>1</sup>, 杨近朱<sup>1,3</sup>, 虎嵩林<sup>1,3</sup>

<sup>1</sup>中国科学院信息工程研究所 北京 中国 100085

<sup>2</sup>华北计算机系统工程研究所 北京 中国 100083

<sup>3</sup>中国科学院大学网络空间安全学院 北京 中国 100049

**摘要** 随着社交网络的快速发展,越来越多的人在社交网络平台获取或分享信息。但是,在收获便利的同时,也为谣言提供了新的传播媒介。谣言的传播严重影响网络空间清朗环境的建设,开展自动化谣言检测至关重要。现有基于深度学习的谣言检测模型往往基于内容特征或传播特征展开,而这些基于传播特征的模型要么只关注传播过程中的时序关系,或是仅挖掘谣言传播网络的结构特征来识别谣言,不能很好地学习一个全面的特征表示描述谣言传播过程中的时间和空间变化,限制了谣言检测的性能。针对此问题,本文提出一种通用的基于时序特征和结构特征的谣言检测方法,共同探索谣言传播过程中的时间模式与传播树结构特征,学习一个全面的谣言特征表示,提高谣言检测的性能。为评估模型的有效性,本文在 3 个谣言真实数据集进行实验。实验结果表明,本文方法平均获得了 4.8% 准确率绝对提升。大量实验验证了所提方法在谣言检测的有效性。

**关键词** 谣言检测; 时序特征; 传播结构学习; 门控循环单元; 图卷积神经网络; 社交网络

中图分类号 TP391 DOI号 10.19363/J.cnki.cn10-1380/tn.2026.03.18

## Jointly Exploiting Temporal and Structural Features for Rumor Detection on Social Media

WEI Lingwei<sup>1,3</sup>, HU Dou<sup>1,2,3</sup>, BAO Yinan<sup>1,3</sup>, ZHOU Wei<sup>1</sup>, YANG Jinzhu<sup>1,3</sup>, HU Songlin<sup>1,3</sup>

<sup>1</sup> Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China

<sup>2</sup> National Computer System Engineering Research Institute of China, Beijing 100083, China

<sup>3</sup> School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** With the rapid development of the social network, more and more people obtain or share information on social network platforms. Unfortunately, the convenient environment of social network platforms has also provided a new medium for the spread of rumors. The spread of rumor has become a significant challenge that seriously undermines the credibility of information in social network and posts a threat to building a clear cyberspace environment. Automatic rumor detection is essential for timely prevention of rumor spread and maintaining social stability. The existing deep learning-based rumor detection models have been developed based on content characteristics or propagation characteristics including temporal features and structural features. However, most of these detection models either only model the temporal information in rumor propagation or only focus on the structure features of rumor propagation to identify rumors. This limitation cannot learn a comprehensive eigenvector representation well and hinders the performance of rumor detection. To alleviate the above problem, in this paper, we propose a novel graph-based rumor detection model. It combines the power of graph networks and sequence models to jointly model both structural features and temporal patterns in rumor propagation. Specifically, based on the textual features extracted by embedding layer and propagation information, we utilize a time-aware bidirectional gated recurrent unit to explore temporal features and a graph convolutional network to learn structural features. Then, we combine them to make prediction. By doing so, the model can learn a comprehensive representation of rumor characteristics, enabling it to detect rumors with greater accuracy. In addition, the model can effectively alleviate the time mode distortion caused by pruning. To evaluate the performance of the proposed model, we conduct experiments on three real-world rumor detection benchmark datasets. The experimental results show that the proposed method achieves 4.8% average absolute improvements in terms of the accuracy score across all three datasets. Extensive experiments demonstrate the effectiveness of the proposed model for rumor detection.

**Key words** rumor detection; temporal information; propagation structure learning; gated recurrent unit; graph convolutional network; social network

通讯作者: 周薇, 博士, 高级工程师, Email: zhouwei@iie.ac.cn。

本课题得到国家自然科学基金(No.62102412)和国家重点研发计划(No. 2022YFC3302102)资助。

收稿日期: 2020-12-25; 修改日期: 2021-03-04; 定稿日期: 2023-08-09

## 1 引言

社交网络的快速发展,为信息传播带来便利的同时也滋生了谣言等虚假信息的快速传播。大量的虚假信息传播严重降低了社交网络中信息的可信度,为建设清朗的网络空间环境带来巨大的危害与挑战。因此,识别网络中的谣言成为工业界和学术界研究的热点问题之一。为降低谣言的不良影响,国内外多家媒体机构或网络服务供应商都相继构建了谣言举报核查平台,例如,中国互联网联合辟谣平台<sup>①</sup>、美国的 Snopes<sup>②</sup>等。但是,这些人工核查的方式需要耗费大量时间、人力和物力来验证信息的真实性,效率低下。因此,如何自动识别谣言,对于及时阻止谣言传播,降低谣言危害,维护社会的长治久安有着重要意义。

谣言<sup>[1]</sup>是指一种从一个人传到另一个人的故事,其中的真相未经证实或值得怀疑。随着时间的推移,基于用户的转发或分享等传播行为,形成了特有的传播路径。研究表明<sup>[2-3]</sup>,真假消息的传播方式存在一定的差异。与真实消息相比,谣言的传播更快、更广、更深入,其传播路径具有独特的结构特征。原因可能是谣言比真实信息更加新奇,而人们更加倾向于分享新奇的消息。上述研究为研究者利用传播结构检测谣言提供了可能。早期的谣言检测模型<sup>[1-14]</sup>通过人工构建内容特征、传播过程中的时序特征和结构特征等,应用机器学习分类器完成谣言分类。此类方法受限于繁重的人工特征工程,无法学习到谣言的潜在特征表示。随着深度学习模型的日趋成熟,研究者提出了许多有效的基于深度学习的谣言检测模型,从谣言传播内容<sup>[15-23]</sup>、传播时序<sup>[24-30]</sup>或者传播结构<sup>[31-37]</sup>中自动学习高阶特征向量表示,用于谣言检测。

但是,这些研究要么仅关注谣言的时序特征,无法利用传播树特有的结构特征,或仅考虑了传播的结构特征,而忽略了谣言传播的时间信息,二者虽然都取得了一定的效果,但是无法学习一个全面的谣言特征向量表示,共同描述谣言在时间和空间上的传播特点<sup>[38]</sup>,从而限制了谣言检测的性能。为了填补这一空白,本文设计模型同时探究谣言传播过程中的时序特征和结构特征,旨在深入地挖掘潜在的谣言传播特征,有效地优化了谣言特征向量,从而提升谣言检测的性能。

本文提出了一个通用的谣言检测框架来综合建模谣言传播的结构信息和时序信息。首先,基于谣言

的传播序列,提出一种基于时间感知的谣言检测模型,融合传播过程中的序列上下文信息,探索谣言传播的时序特征,获得谣言的时序特征表示。其次,本文将每个谣言的传播序列转化为一个自顶向下的传播图(top-down propagation graph)和一个自底向上的扩散图(bottom-up diffusion graph),利用图卷积神经网络模型分别建模谣言传播和扩散方向上的结构特征,学习得到谣言传播的结构特征向量表示。最后,融合谣言传播的时序特征向量表示和结构特征向量表示,完成谣言分类。

此外,为了缓解图模型存在的过拟合(over-fitting)问题<sup>[39]</sup>,往往需要进行剪枝或采样操作,以提高模型的泛化能力。但是现有方法剪枝后常常会导致时间模式失真问题<sup>[40]</sup>,而本文提出的谣言检测框架采用并行方案共同挖掘传播过程中的时序特征和结构特征,可有效缓解此类模型失真问题。

本文在 3 个公开的谣言检测数据集上进行实验,评估本文提出的模型,并将其结果与谣言检测现有的经典模型进行比较。大量的实验结果证明了本文模型的有效性。

综上,本文的主要贡献包含 3 个方面:

1) 提出一种基于时序特征和结构特征的谣言检测模型,不仅保留谣言传播中的时序信息,同时捕获传播和扩散方向的结构特征,学习一个更全面的谣言特征表示。

2) 提出一个通用的融合传播结构信息和时序信息的框架方法,该方法可以有效缓解剪枝带来的时间模式失真问题。

3) 在 3 个真实数据集上进行实验,实验结果表明本文提出的模型可以有效发现社交网络中潜在的谣言,尤其对于识别早期阶段传播的谣言具有重要意义。

接下来,本文的章节安排如下:第 2 节将阐述与谣言检测相关研究工作;第 3 节对本文研究的问题进行定义;第 4 节详细描述本文提出的基于时序特征和结构特征的谣言检测模型;第 5 节阐述实验设置并分析实验结果;最后在第 6 节中,对全文进行总结。

## 2 相关研究

近年来,如何在社交媒体上实现自动的谣言检测受到了人们的广泛关注。本文主要关注基于文本的和基于传播的谣言自动检测方法,将从传统方法

① <http://www.piyao.org.cn/>

② <http://snopes.com/>

和基于深度学习的方法介绍相关研究工作。

## 2.1 传统的谣言检测方法

早期的谣言检测方法主要通过人工设计的特征来实现谣言检测。

### 1) 基于内容特征的机器学习方法

一些研究者提出,利用从文本中抽取出来的人工特征,例如语言学特征<sup>[4-7]</sup>、语言风格特征<sup>[8]</sup>等,利用机器学习方法实现谣言的自动检测。此外,Dungs等人<sup>[9]</sup>还利用立场特征,使用隐式马尔可夫模型(hidden Markov models, HMM)实现对谣言的验证。祖坤琳等人<sup>[10]</sup>提出将评论内容的情感倾向作为 SVM 分类器的特征之一,用于检测谣言。

此外,也有一些研究者利用谣言传播模式中的特征,实现谣言的检测。

### 2) 基于传播时序特征的机器学习方法

Kwon 等人<sup>[11]</sup>利用基于博文数量随时间变化的时间序列拟合模型,实现谣言检测。Ma 等人<sup>[12]</sup>利用按照时间先后顺序的社交上下文特征,以实现谣言的自动检测。毛二松等人<sup>[13]</sup>从内容、传播以及用户信息三方面进行特征提取得到深层分类特征,并利用分类特征训练集成分类器,用于微博谣言的检测。

### 3) 基于传播结构特征的机器学习方法

Wu 等人<sup>[14]</sup>使用基于核(kernel)的方法,提出了基于图-核(graph-kernel)的混合支持向量机(support vector machines, SVM)分类器。Ma 等人<sup>[41]</sup>定义了树核(tree kernel)来计算相似性,并使用 SVM 用于谣言分类。

但是,上述机器学习方法大多依赖复杂的预处理过程和特征工程,不仅耗时耗力,并且难以自动捕获到高层次的表示。

## 2.2 基于深度学习的谣言检测方法

基于深度学习的谣言检测方法可以分为基于内容特征的方法以及基于传播特征的方法两大类。

### 1) 基于内容特征的深度学习方法

Ma 等人<sup>[20]</sup>利用循环神经网络(recurrent neural network, RNN)和生成对抗网络(generative adversarial networks, GAN),从增强的文本数据中学习更强大的谣言特征。Przybyla<sup>[21]</sup>使用双向的长短时记忆网络(long short-term memory, LSTM)建模文本,实现对假新闻的识别。Shu 等人<sup>[22]</sup>同时利用了新闻内容和用户评论内容,基于门控循环单元(gated recurrent unit, GRU)和注意力机制提出了具有可解释性的假新闻检测模型。Yu 等人<sup>[15]</sup>和刘政等人<sup>[16]</sup>均提出使用卷积神经网络(convolutional neural networks, CNN)从文本中得到关键特征。李力钊等人<sup>[17]</sup>结合了 CNN 和 GRU

的优点,对谣言文本进行处理。Wang 等人<sup>[23]</sup>将 CNN 和强化学习相结合,利用用户评论信息作为弱监督信息,对数据进行标注,以增加用于假新闻检测的训练数据。

### 2) 基于传播时序特征的深度学习方法

基于传播时序特征的深度学习方法旨在学习谣言传播随时间的变化特征。针对谣言传播的时序特征, Ma 等人<sup>[24]</sup>使用 RNN 建模用户评论的时序关系。随后, Chen 等人<sup>[25]</sup>在其基础上引入了注意力机制,对不同的文本特征分配不同的注意力,提高谣言检测性能。此外, Ma 等人<sup>[29]</sup>还使用基于树结构的 RvNN, 从不同方向的谣言传播结构中捕获结构特征。Wu 等人<sup>[26]</sup>利用了用户间的社交网络结构,并使用 LSTM-RNN 建模消息传播时序关系,实现对假新闻的检测。一些研究引入传播过程中评论或转发中的立场信息,提出多任务学习方法,例如, Ma 等人<sup>[27]</sup>考虑立场检测任务和谣言检测任务的密切联系,利用 GRU 模块通过提取共享特征,共同建模谣言检测和立场检测任务。但是基于 RNN 的方法难以捕获到谣言传播过程中的长依赖关系,并且难以用于谣言的早期检测。为了缓解这些问题, Carley 和 Kathleen<sup>[28]</sup>使用带有卷积单元的树形结构 LSTM, 建模文本内容和时序结构,用于社交媒体对话中的谣言验证和立场检测任务。Liu 等人<sup>[30]</sup>使用 RNN 和 CNN 分别学习到全局的和局部的用户特征变化,对谣言传播路径进行分类,从而实现对谣言的识别。

### 3) 基于传播结构特征的深度学习方法

谣言通过转发、评论等社交行为,形成了多级传播树。为了挖掘谣言传播树的结构特征,现有研究往往将其建模为序列或图模型,应用深度学习学习方法学习特征向量表示,用于谣言的自动检测。例如, Wei 等人<sup>[33]</sup>利用图卷积网络(graph convolutional network, GCN)编码谣言传播结构以学习立场特征实现立场分类,然后使用 RNN 建模动态变化的立场信息,从而实现对立场的预测以及谣言的验证。Yuan 等人<sup>[34]</sup>考虑到以往的工作没有充分利用到消息传播图中局部的语义关联以及全局的结构信息,提出使用 CNN 和注意力机制以同时结合这两类信息,模型最终在谣言检测和早期检测任务中都取得了较好的效果。Lu 等人<sup>[35]</sup>考虑到时序特征难以用于谣言的早期检测并且易于操控,利用源贴中的短文本、用户转发序列和用户信息,基于 GCN、GRU 和 CNN 提出图感知的共注意力网络,以实现对假新闻的预测。Bian 等人<sup>[36]</sup>提出双向 GCN 模型用于建模谣言传播的不同方向,使用两个 GCN 分别从自顶向下的谣言传播结构和自

底向上的谣言扩散结构中学习谣言传播的结构特征。胡斗等人<sup>[37]</sup>利用多关系图神经网络探究信息传播中多种潜在依赖关系, 捕获了更丰富的结构特征用于检测。Yang 等人<sup>[42]</sup>从对抗学习的角度出发, 提出具有鲁棒性的基于图的谣言检测器, 将用户、帖子和评论等丰富信息建模成异质信息网络, 使用 GCN、RNN、注意力机制以及对抗学习等方法, 提出图对抗学习框架, 生成器在图结构上动态添加干扰信息以欺骗检测器, 同时检测器学习到更多的具有区分性的结构特征以抵抗干扰。

尽管上述深度学习方法在谣言检测性能上均取得了一定的提升, 但是他们要么只建模了谣言的时序特征, 或者仅考虑了传播的结构特征, 不能很好地学习一个全面的特征向量表示用于描述谣言在时间和空间的传播特征, 从而限制了谣言检测的性能。为了填补这一空白, 本文同时建模了谣言传播的时序特征和结构特征, 能够更加深入地挖掘潜在的谣言传播特征, 有效地优化了谣言特征向量, 从而提升谣言检测的性能。

### 3 问题定义

定义  $\mathcal{C} = \{C_1, C_2, \dots, C_{|C|}\}$  表示谣言数据集, 其中,  $C_i$  是第  $i$  个样本,  $|C|$  表示数据集中样本的总个数。对于每个样本  $C_i = \{r_i, x_{i1}, x_{i2}, \dots, x_{i(n_i-1)}\}$ , 其中,  $r_i$  表示该谣言的源博文,  $x_{ij}$  表示该谣言的第  $j$  个转发/评论博文。  $n_i$  表示该谣言下的博文数。每个谣言被标

记一个类别标签  $y_i$ ,  $y_i \in \mathcal{Y}$ 。

给定谣言的文本内容以及它的传播树信息, 谣言检测任务是学习一个分类器, 预测谣言的类别概率分布。谣言检测的分类器可定义为

$$f: \mathcal{C} \rightarrow \mathcal{Y}.$$

## 4 基于时序关系和传播结构的谣言检测模型

本节提出基于传播时序特征和结构特征的谣言检测模型, 共同建模谣言传播过程中的时序关系和传播树固有的结构特征, 学习一个谣言的高阶特征表示, 用于谣言检测任务。

### 4.1 模型整体架构

本文提出的谣言检测模型整体架构如图 1 所示, 主要包含 4 个模块, 分别是嵌入层、时序特征编码层、结构特征编码层以及谣言分类层。

(1) 嵌入层: 提取谣言文本内容的 TF-IDF 特征作为节点的初始向量表示。

(2) 时序特征编码层: 基于谣言的传播序列, 通过基于时间感知的双向门控循环单元提取谣言的时序特征, 得到谣言的时序特征向量表示。

(3) 结构特征编码层: 基于谣言传播树构建图结构, 采用图卷积网络模型学习传播结构特征, 得到谣言的结构特征向量表示。

(4) 谣言分类层: 基于融合后谣言的时序特征向量表示和结构特征向量表示, 判断谣言的类型。

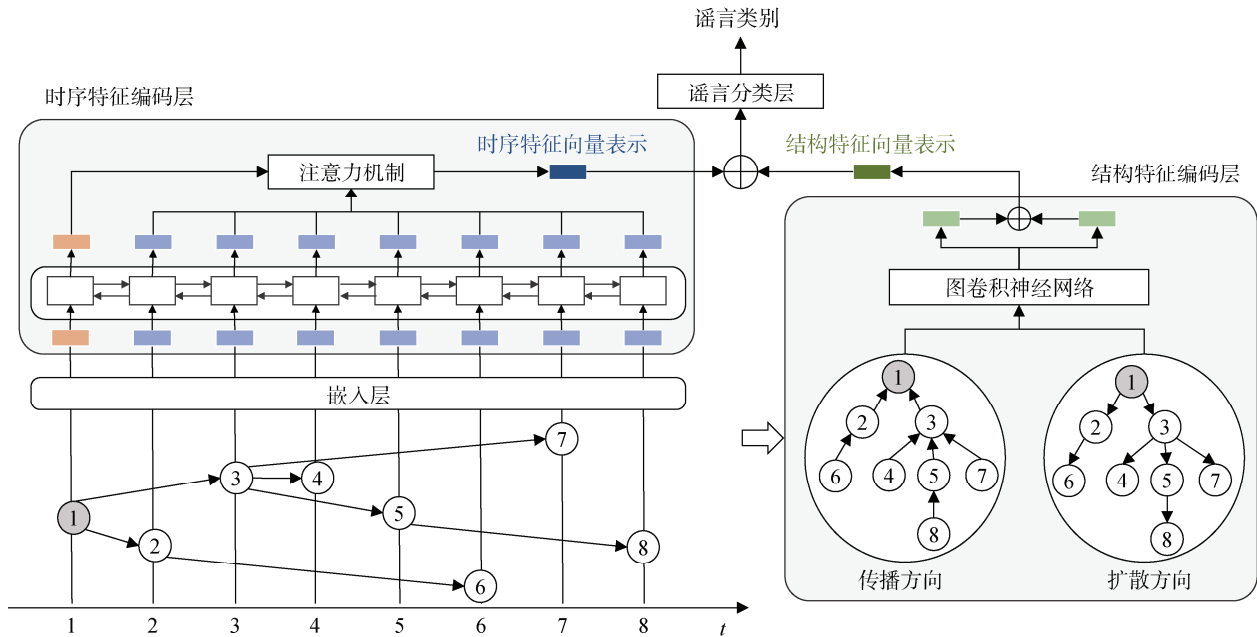


图 1 基于时序特征和结构特征的谣言检测模型总体框架  
Figure 1 The architecture of the proposed rumor detection model

## 4.2 嵌入层

本文采用 TF-IDF 提取文本特征, 作为节点的初始化向量表示。具体地, 对于谣言发布博文和其对应的后续一系列转发/评论博文, 提取频率最高的 5000 个单词, 计算每个博文的 TF-IDF 特征, 得到每个博文的初始向量表示  $\{\mathbf{x}_{ir}, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{i(n-1)}\}$ , 其中,  $\mathbf{x}_{ir}$  表示源博文的初始向量表示,  $\{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{i(n-1)}\}$  表示后续转发/评论的博文的初始向量表示。

## 4.3 时序特征编码层

门控循环单元(gated recurrent unit, GRU)<sup>[43]</sup>可以很好地整合序列的上下文特征, 对时序类特征有着良好的捕获能力。本文提出基于时间感知的谣言检测模型, 利用 Bi-GRU 捕获传播过程中的时序特征, 并利用注意力机制进行特征融合, 得到谣言的时序特征向量表示。

具体地, 给定谣言发布博文及后续转发/评论博文的初始向量表示  $\{\mathbf{x}_{ir}, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{i(n-1)}\}$ , 隐状态向量表示计算如式(1)和式(2)所示:

$$\bar{\mathbf{h}}_{ij} = \overline{\text{GRU}}(\mathbf{x}_{ij}), j \in \{r\} \cup [1, L-1], \quad (1)$$

$$\bar{\mathbf{h}}_{ij} = \overline{\text{GRU}}(\mathbf{x}_{ij}), j \in [L-1, 1] \cup \{r\}, \quad (2)$$

其中,  $\overline{\text{GRU}}$  表示一个前向的门控循环单元,  $\overline{\text{GRU}}$  表示一个反向的门控循环单元,  $L$  表示序列长度。通过拼接前向隐状态向量表示  $\bar{\mathbf{h}}_{ij}$  和后向隐状态向量表示  $\bar{\mathbf{h}}_{ij}$ , 得到隐状态向量表示, 即

$$\mathbf{h}_{ij} = [\bar{\mathbf{h}}_{ij}; \bar{\mathbf{h}}_{ij}]. \quad (3)$$

在谣言传播的过程中, 随着时间的推移, 谣言在各个时刻的影响力度是不同的, 为了建模这一特征, 本文应用注意力机制, 得到最终谣言的时序特征向量表示  $\mathbf{v}$ , 计算公式如式(4)~式(6)所示。

$$\mathbf{u}_{ij} = \tanh(\mathbf{W}_s \mathbf{h}_{ij} + \mathbf{b}_s), \quad (4)$$

$$\alpha_{ij} = \frac{\exp(\mathbf{u}_{ij}^T \mathbf{u}_{i0})}{\sum_j \exp(\mathbf{u}_{ij}^T \mathbf{u}_{i0})}, \quad (5)$$

$$\mathbf{v}_i = \sum_j \alpha_{ij} \mathbf{h}_{ij}, \quad (6)$$

其中,  $\mathbf{W}_s$ 、 $\mathbf{b}_s$  是需要学习的参数。

## 4.4 结构特征编码层

为了捕获传播过程中的结构特征, 本文将每个谣言的传播树建模为图结构, 并通过图卷积网络模型聚合邻居节点信息, 生成图的特征表示。

### 4.4.1 构建图结构

本文将谣言发布博文和后续转发/评论博文建模

为节点, 并依据转发/评论关系, 构建一条有向边。节点的初始化特征向量表示为  $\mathbf{X} = \{\mathbf{x}_r, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n-1}\}$ , 该图的邻接矩阵  $\mathbf{A}$  定义为

$$a_{ij} = \begin{cases} 1 & \text{节点}i\text{转发/评论节点}j \\ 0 & \text{其他.} \end{cases} \quad (7)$$

对于每个谣言传播过程, 依据传播方向不同, 构建一个自顶向下的传播图结构和一个自底向上的扩散图结构, 记为  $G^{\text{TD}}$  和  $G^{\text{BU}}$ , 其对应的邻接矩阵为  $\mathbf{A}^{\text{TD}}$  和  $\mathbf{A}^{\text{BU}}$ 。

### 4.4.2 学习节点表示

图卷积网络模型(graph convolutional network, GCN)定义了一个卷积层操作, 描述了信息传递过程, 通过迭代地聚合边的信息和节点的信息, 不断生成新的节点表示。第  $l+1$  层的信息传递可定义如式(8)所示:

$$\mathbf{H}^{(l+1)} = \sigma(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)} + \mathbf{b}^{(l)}), \quad (8)$$

其中,  $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2}$  是归一化后的邻接矩阵, 其中,  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ ,  $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$ ,  $\mathbf{W}^{(l)}$  是训练参数,  $\sigma(\cdot)$  是非线性激活函数,  $\mathbf{H}^{(0)} = \mathbf{X}$  表示输入节点的特征向量表示, 其维度为  $d$ 。

现有研究证明<sup>[29,36]</sup>, 在谣言传播的过程中, 根节点起着重要作用。本文将上述根节点的特征向量表示与节点的原特征向量表示进行融合, 如式(9)所示:

$$\tilde{\mathbf{H}}^{(l)} = [\mathbf{H}^{(l)}; \mathbf{H}_{\text{root}}^{(l-1)}]. \quad (9)$$

在实验中, 本文设置该卷积神经网络的层数为 2, 即谣言传播树和谣言扩散树的最终节点特征表示分别为:  $\mathbf{H}^{\text{TD}} = (\mathbf{H}^{\text{TD}})^{(2)}$ ,  $\mathbf{H}^{\text{BU}} = (\mathbf{H}^{\text{BU}})^{(2)}$ , 其中,  $(\mathbf{H}^{\text{TD}})^{(0)} = (\mathbf{H}^{\text{BU}})^{(0)} = \mathbf{X}$ 。

为缓解图模型的过拟合问题, 往往会进行剪枝操作。参考文献[36], 本文选用 DropEdge<sup>[39]</sup>, 通过一定概率随机失活给定输入图结构中的边, 以提升模型泛化能力。

### 4.4.3 学习图表示

本文利用平均池化操作来聚合各个图中的节点的特征向量表示, 生成不同传播方向和扩散方向的传播结构特征向量表示, 如式(10)所示:

$$\begin{aligned} \mathbf{C}^{\text{TD}} &= \text{Aggergate}(\tilde{\mathbf{H}}^{\text{TD}}), \\ \mathbf{C}^{\text{BU}} &= \text{Aggergate}(\tilde{\mathbf{H}}^{\text{BU}}), \end{aligned} \quad (10)$$

其中, Aggergate 表示平均池化操作。最后, 本文融合不同方向上的向量表示, 得到最终结构特征表示:

$$\mathbf{C} = \mathbf{W}_c [\mathbf{C}^{\text{TD}}; \mathbf{C}^{\text{BU}}] + \mathbf{b}_c, \quad (11)$$

其中,  $W_c$  和  $b_c$  是需要学习的参数。

## 4.5 谣言分类层

融合上述学习到的每个谣言的时序特征向量表示  $V = [v_r, v_1, \dots, v_{n-1}]$  和结构特征表示  $C$ , 并计算该谣言属于每个类别的概率为

$$\hat{Y} = \text{softmax}(W_o[V; C] + b_o), \quad (12)$$

其中,  $W_o$  和  $b_o$  是需要学习的参数。

训练过程中, 本文利用交叉熵损失函数训练整个模型, 损失函数定义如下:

$$\mathcal{L} = -\sum_i y^i \log \hat{y}^i + \frac{\beta}{2} \|\Theta\|_2^2, \quad (13)$$

其中,  $\beta$  是一个超参数, 表示 L2 正则化的大小。

## 5 实验与分析

本节先对实验设置进行介绍, 包括数据集、实验对比模型、实验参数以及实验结果评估方法; 之后对实验结果进行对比分析。

### 5.1 实验设置

#### 5.1.1 数据集

本文在 Twitter15<sup>[8]</sup>、Twitter16<sup>[8]</sup>、PHEME<sup>[44]</sup> 三个公开的数据集上评估了本文提出的模型。这三个数据集的数据特性如表 1 所示。

表 1 谣言检测数据集统计信息  
Table 1 Statistics of the Datasets

数据集	Twitter15 数据集	Twitter16 数据集	PHEME 数据集
博文总数	331,612	204,820	105,354
假谣言的数量	370	205	638
真谣言的数量	374	205	1,067
未证实谣言的数量	374	203	697
非谣言的数量	372	205	-
总计	1,490	818	2,402

**Twitter15** 和 **Twitter16** 数据集分别包含 1490 和 818 个谣言源博文。源博文的标签一共有非谣言 (Non-rumor, NR)、假谣言(False Rumor, F)、真谣言 (True Rumor, T) 和未验证的谣言(Unverified Rumor, U) 四种。受已有工作<sup>[29,36]</sup>的启发, 本文将数据集随机分为 5 份, 并采用 5 折交叉验证以获得更鲁棒的结果。**PHEME** 数据集包含 9 个事件相关的共 2402 个谣言源博文。标签一共有假谣言、真谣言和未验证的谣言 3 种。参考前人工作<sup>[33]</sup>, 本文采用留一法(leave-one-event-out)交叉验证, 即在每折中, 使用和一个事件相关的样本用作测试, 剩余样本用于训练。

#### 5.1.2 对比方法

对于 Twitter15 和 Twitter16 数据集, 本文将本文提出的模型和以下方法进行对比。

**DTR**<sup>[7]</sup>: 是基于决策树的排序模型, 通过搜索查询短语来识别谣言。

**SVM-TS**<sup>[12]</sup>: Ma 等人利用时间序列, 通过线性的支持向量机(SVM)分类器来建模社交上下文特征的时间变化。

**SVM-TK**<sup>[41]</sup>: Ma 等人应用 SVM 分类器和传播树核来建模谣言的传播结构。

**RvNN**<sup>[29]</sup>: Ma 等人采用了两种基于自底向上和自顶向下的树型结构递归神经网络, 挖掘时序特征。

**StA-PLAN**<sup>[31]</sup>: Khoo 等人利用转换网络(transformer networks)整合博文间的长距离交互以及博文传播树结构中的信息。

**Bi-GCN**<sup>[36]</sup>: Bian 等人利用双向图卷积网络建模自底向上的谣言传播结构和自顶向下的谣言扩散结构, 挖掘传播过程的结构特征。

对于 PHEME 数据集, 本文选取上述基于 RNN 的典型模型, **RvNN** 和基于图的代表模型 **Bi-GCN**。此外, 还对比了 **Hierarchical GCN-RNN**<sup>[33]</sup>: 基于 GCN 建模了结构属性并联合预测了谣言的可信度和立场。为保证对比实验的公平性, 本文与该模型在单一任务中的检测性能作比较。

#### 5.1.3 参数设置

参考基线模型<sup>[45]</sup>, 每个节点的隐层特征向量维度设为 64, DropEdge 的失活概率设为 0.2, Dropout 的概率设为 0.5。本文使用 Adam 算法<sup>[46]</sup>对模型进行优化。训练过程重复 200 轮, 当验证集的损失在连续的 10 轮停止下降时, 则训练过程将提前终止<sup>[47]</sup>, 最终取实验多次结果的平均值作为最终结果。

#### 5.1.4 评估方法

对于 Twitter15 和 Twitter16 数据集, 参考文献 [29,36], 本文评估模型在 4 个标签类别上的总准确率 (Acc.), 记录在每个类别上的 F1 值。计算公式如下所示:

$$\text{Acc.} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}},$$

$$\text{F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}},$$

其中, TP 表示预测值和目标值都为 1 的个数, TN 表示预测值和目标值都为 0 的个数, FP 表示预测值为 1 但目标值为 0 的个数, FN 表示预测值为 0 但目标值为 1 的个数。

对于 PHEME 数据集, 参考文献 [33], 本文评估准确率和宏平均 F1 值(macro-averaging F1, Macro-F1),

即各个类别在 F1 指标下的算术平均值。由于类别的不均衡问题, 本文还比较了加权 F1 值(weighted-averaging F1, Weighted-F1), 即各个类别在 F1 指标下的加权平均值。

## 5.2 谣言检测及其实验结果

### 5.2.1 Twitter15 和 Twitter16 数据集

表 2、表 3 分别展示了在 Twitter15、Twitter16 数据集上的实验结果, 其中, 笔者基于文献[36]的开源项目得到 Bi-GCN 的实验结果, 其他对比模型结果参考文献[31]。

表 2 谣言检测实验结果(Twitter15)

Table 2 Rumor detection results on Twitter15 dataset. (单位: %)

模型	Acc	NR	F	T	U
		F1	F1	F1	F1
DTR	40.9	50.1	31.1	36.4	47.3
SVM-TS	54.4	79.6	47.2	40.4	48.3
SVM-TK	66.7	61.9	66.9	77.2	64.5
RvNN	72.3	68.2	75.8	82.1	65.4
StA-PLAN	85.2	84.0	84.6	88.4	83.7
Bi-GCN	87.1	86.0	86.7	<b>91.3</b>	83.6
<b>本文模型</b>	<b>88.0</b>	<b>87.1</b>	<b>87.7</b>	<b>91.3</b>	<b>85.7</b>

(注: 黑体表示当前指标下的最优结果)

表 3 谣言检测实验结果(Twitter16)

Table 3 Rumor detection results on Twitter16 dataset (单位: %)

模型	Acc	NR	F	T	U
		F1	F1	F1	F1
DTR	41.4	39.4	27.3	63.0	34.4
SVM-TS	57.4	75.5	42.0	57.1	52.6
SVM-TK	66.2	64.3	62.3	78.3	65.5
RvNN	73.7	66.2	74.3	83.5	70.8
StA-PLAN	86.8	82.6	83.3	92.7	88.8
Bi-GCN	88.5	82.9	89.9	93.2	88.2
<b>本文模型</b>	<b>89.2</b>	<b>85.6</b>	<b>90.1</b>	<b>93.8</b>	<b>90.1</b>

(注: 黑体表示当前指标下的最优结果)

在 2 个 Twitter 数据集上的实验结果表明, 本文提出模型的性能要优于基线模型, 证明了本文模型共同挖掘时序特征和传播结构特征在谣言检测任务中的有效性。

DTR 的识别准确率较低, 说明仅仅通过文本内容, 很难识别谣言, 因为谣言产生者往往会故意博人眼球, 模仿真实消息的内容特征以伪装自己<sup>[48-49]</sup>。SVM-TS 通过分析谣言的时序特征, 可在一定程度上提升谣言的检测性能。这也说明时序特征在谣言

检测任务中的有效性。但是, DTR、SVM-TS 均通过人工构建特征, 利用传统机器学习模型完成谣言检测任务, 该类模型的检测性能普遍较差, 这也说明了人工提取特征的局限性, 并且很难学习到高阶特征表示。

SVM-TK 通过核函数分析了谣言和真相形成的不同传播树的特点, 该模型的检测性能优于基于人工构建特征的方法, 说明基于传播结构识别谣言的可行性及先进性。

相比之下, 基于深度学习模型的方法, 例如 RvNN、StA-PLAN、Bi-GCN 等的模型性能更好, 证实了深度学习模型在谣言检测任务中的优势。与传统的机器学习方法相比, 此类模型可以更好地挖掘文本信息和传播信息中隐藏的特征向量表示, 从而完成更准确的分类效果。

Bi-GCN 模型利用双向图卷积网络学习了谣言传播的结构特征, 但是忽略了谣言传播过程本身的时间序列特征, 而本文提出的模型在充分挖掘传播结构特征的同时, 还可融入传播过程中的时序特征, 可学习到更丰富的特征向量表示, 因此, 检测性能有所提升。

### 5.2.2 PHEME 数据集

表 4 是 PHEME 数据集中的实验结果, 其中, 笔者基于文献[36]的开源项目得到 BiGCN 的实验结果, 其他对比模型的结果参考文献[33]。

表 4 数据集谣言检测实验结果(PHEME)

Table 4 Rumor detection results on PHEME dataset (单位: %)

模型	Acc	Macro-F1	Weighted-F1
RvNN	34.1	26.4	-
Hierarchical GCN-RNN	35.6	31.7	-
Bi-GCN	49.2	46.7	63.2
<b>本文模型</b>	<b>62.0</b>	<b>52.1</b>	<b>67.9</b>

(注: 黑体表示当前指标下的最优结果)

从实验结果中, 本文得出: 基于图模型的谣言检测模型 Bi-GCN、Hierarchical GCN-RNN、本文模型均优于基于 RNN 的模型 RvNN, 该结果证实了图模型强大的特征表示能力。

此外, 本文模型在 PHEME 数据集与 2 个 Twitter 数据集的提升程度不同。由表 1 可知, PHEME 数据集包含的博文总数较少, 所形成的传播图结构往往更稀疏, 所有模型在该数据集上的表现较差。而本文模型同时挖掘了传播的时序特征和结构特征, 可提取更加丰富的特征, 有效提升了谣言检测的性能。

### 5.3 消融分析

本文通过消融分析,进一步探究了各个模块对谣言检测性能的影响。对比模型设置如下所示。

(1) w/o Att: 在时序特征编码层,移除注意力机制,通过简单的线性变化提取最终时序特征表示。

(2) w/o AGRU: 移除时序特征编码层,仅考虑建模传播的结构特征。

(3) w/o GCN: 移除结构特征编码层,仅考虑建模传播的时序特征。

实验结果如表 5 所示, w/o Att 模型的效果略低于本文的模型,这表明在传播过程中,谣言的影响会随着时间的推移而不断变化。本文的模型通过注意力机制,可以很好地建模发布博文与其他转发博文的交互关系,有助于更好地挖掘谣言的潜在向量表示。当移除时序特征编码层或移除结构特征编码层时,检测性能均呈现明显下降,尤其是移除结构特征编码层。而本文的模型在 3 个数据集上均表现最佳检测性能。这证实了共同建模时序特征和传播结构特征的必要性和重要性。

表 5 消融分析结果(准确率)

Table 5 Rumor detection results (Accuracy)

模型	Twitter15	Twitter16	PHEME
本文模型	88.0	89.2	62.0
- w/o Att	87.5	88.7	59.6
- w/o AGRU	87.1	88.0	56.3
- w/o GCN	78.1	79.5	40.9

### 5.4 模型分析

本文进一步考虑模型的两种不同变体,如图 2(2)和图 2(3)所示。在变体 1 中,基于传播结构特征编码的输出,建模时序关系特征,即用式(9)计算得到的隐向量表示替换式(1)和式(2)中的输入特征向量;变体 2 中,基于时序关系特征输出,建模传播结构特征,即用式(6)中计算得到的隐向量表示作为图中节点的初始化向量表示。

结果如表 6 所示,从表中观察出,本文的模型取得了最佳性能,在三个数据集上均优于变体 1 和变体 2。合理的解释如下:在图卷积网络模型中,为防止模型过拟合,往往会进行剪枝操作,虽然提升了模型的泛化能力,但同时也导致了时间模式失真,不能很好地建模传播的时序特征,因此,限制了谣言检测的性能。

### 5.5 早期谣言检测及其实验结果

早期谣言检测是评估谣言检测性能的另一个重

要指标,即在谣言传播的早期阶段,利用较少的传播信息识别谣言。为了构建早期检测任务,本文分别设置了一系列截止日期和转发数/评论数阈值,仅基于谣言刚产生的几小时内或者转发/评论数较少的情况,评估本文模型和基线模型的准确性。

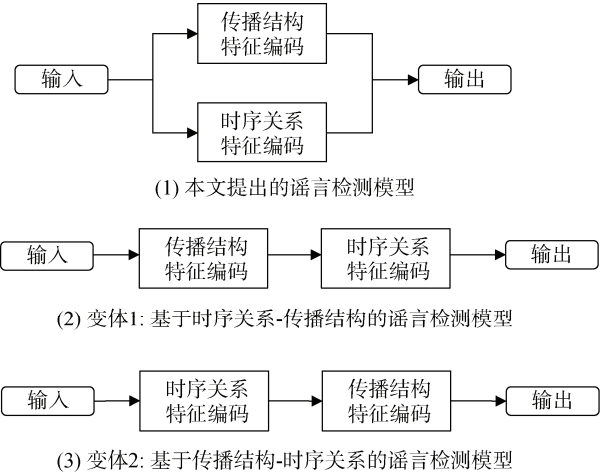


图 2 本文模型变体

Figure 2 The variants of our model

表 6 本文模型变体的实验结果(准确率)

Table 6 Results of the variants (Accuracy)

模型	Twitter15	Twitter16	PHEME
本文模型	88.0	89.2	62.0
变体 1	86.9	87.9	58.1
变体 2	86.6	88.3	58.3

实验结果如图 3 和图 4 所示。从图中可以看出,本文提出的模型由于充分建模传播过程中的时序特征和结构特征,在谣言开始传播的很早期就达到了较高的准确率。此外,在每个截止日期上或每个转发数阈值下的性能均优于其他基线模型,这证实了本文模型不仅有利于谣言的长期检测,而且有利于谣言的早期检测。

## 6 总结

本文提出一种基于时序特征和结构特征的谣言检测模型,不仅可以保留谣言传播中的时序信息,同时还能捕获传播和扩散方向的结构特征,学习一个更全面的谣言特征表示。在 3 个真实数据集上进行实验,实验结果表明本文提出的模型可以有效发现社交网络中潜在的谣言,尤其对于识别早期阶段传播的谣言具有重要意义。此外,本文提出的谣言检测模型还可以有效缓解剪枝带来的时间模式失真问题,具有很好的鲁棒性。

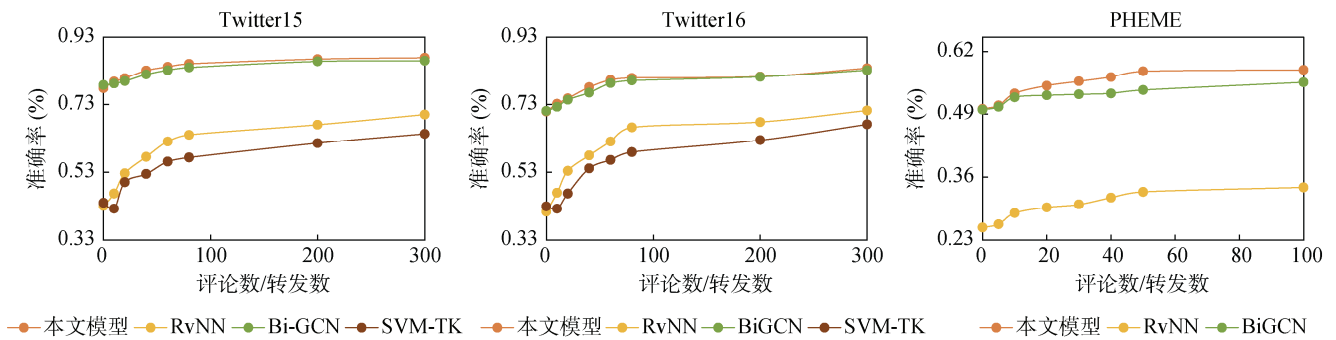


图3 评论转发维度的早期谣言检测

Figure 3 Results of early rumor detection against the comment forwarding dimension

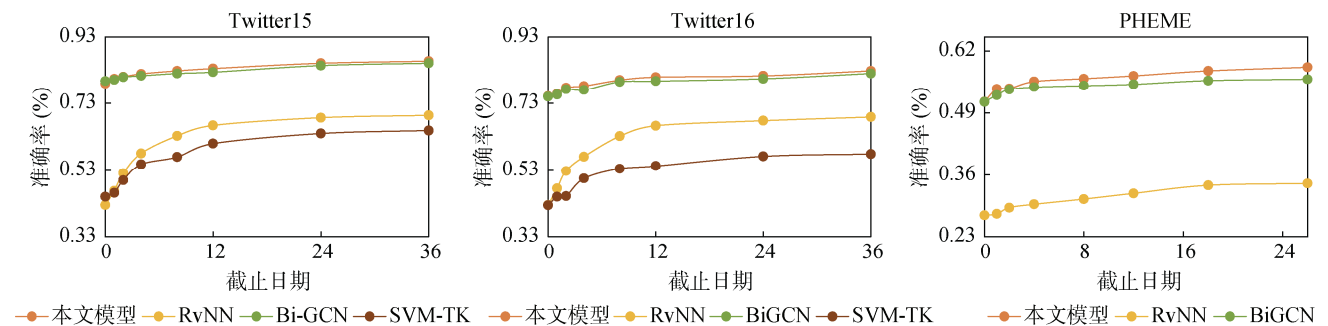


图4 时间维度的早期谣言检测

Figure 4 Results of early rumor detection against time dimension

## 参考文献

- [1] DiFonzo N, Bordia P. Rumor, Gossip and Urban Legends[J]. *Dio-genes*, 2007, 54(1): 19-35.
- [2] Vosoughi S. Automatic detection and verification of rumors on Twitter[D]. Massachusetts Institute of Technology, 2015.
- [3] Vosoughi S, Roy D, Aral S. The Spread of True and False News Online[J]. *Science*, 2018, 359(6380): 1146-1151.
- [4] Ma J, Gao W, Wong K F. Rumor Detection on Twitter with Tree-Structured Recursive Neural Networks[C]. *The 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018: 1980-1989.
- [5] Yu F, Liu Q, Wu S, et al. A Convolutional Approach for Misinformation Identification[C]. *The Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017: 3901-3907.
- [6] Liu Z, Wei Z H, Zhang R X. Rumor Detection Based on Convolutional Neural Network[J]. *Journal of Computer Applications*, 2017, 37(11): 3053-3056, 3100.  
(刘政, 卫志华, 张初弦. 基于卷积神经网络的谣言检测[J]. *计算机应用*, 2017, 37(11): 3053-3056, 3100.)
- [7] Liu Y, Wu Y B. Early Detection of Fake News on Social Media through Propagation Path Classification with Recurrent and Convolutional Networks[C]. *The Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, 2018: 354-361.
- [8] Li L Z, Cai G Y, Pan J. A Microblog Rumor Events Detection Method Based on C-GRU[J]. *Journal of Shandong University (Engineering Science)*, 2019, 49(2): 102-106, 115.  
(李力钊, 蔡国永, 潘角. 基于 C-GRU 的微博谣言事件检测方法[J]. *山东大学学报(工学版)*, 2019, 49(2): 102-106, 115.)
- [9] Khoo L M S, Chieu H L, Qian Z, et al. Interpretable Rumor Detection in Microblogs by Attending to User Interactions[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(5): 8783-8790.
- [10] 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*, 2020: 1-8.
- [11] Cheng M X, Nazarian S, Bogdan P. VRoC: Variational Autoencoder-Aided Multi-Task Rumor Classifier Based on Text[C]. *Proceedings of The Web Conference 2020*, 2020: 2892-2898.
- [12] Li Q Z, Zhang Q, Si L. Rumor Detection by Exploiting User Credibility Information, Attention and Multi-Task Learning[C]. *The 57th Annual Meeting of the Association for Computational Linguistics*, 2019: 1173-1179.
- [13] Meel P, Vishwakarma D K. Fake News, Rumor, Information Pollution in Social Media and Web: A Contemporary Survey of State-of-the-Arts, Challenges and Opportunities[J]. *Expert Systems with Applications*, 2020, 153: 112986.
- [14] Rong Yu, Huang Wenbing, Xu Tingyang, et al. DropEdge: Towards deep graph convolutional networks on node classification [C]. *The 8th International Conference on Learning Representations*, 2020: 1-16.
- [15] Sharma K, Qian F, Jiang H, et al. Combating Fake News: A Survey on Identification and Mitigation Techniques[J]. *ACM Transactions*

- on *Intelligent Systems and Technology*, 2019, 10(3): 1-42.
- [16] Castillo C, Mendoza M, Poblete B. Information Credibility on Twitter[C]. *The 20th International Conference on World Wide Web*, 2011: 675-684.
- [17] Takahashi T, Igata N. Rumor Detection on Twitter[C]. *The 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium on Advanced Intelligence Systems*, 2012: 452-457.
- [18] Gupta A, Kumaraguru P, Castillo C, et al. TweetCred: Real-Time Credibility Assessment of Content on Twitter[C]. *Social Informatics*, 2014: 228-243.
- [19] Zhao Z, Resnick P, Mei Q Z. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts[C]. *The 24th International Conference on World Wide Web*, 2015: 1395-1405.
- [20] Popat K. Assessing the Credibility of Claims on the Web[C]. *The 26th International Conference on World Wide Web Companion*, 2017: 735-739.
- [21] Dungs S, Aker A, Fuhr N, et al. Can Rumour Stance Alone Predict Veracity? [C]. *The 27th International Conference on Computational Linguistics*, 2018: 3360-3370.
- [22] Zu K L, Zhao M W, Guo K, et al. Research on the Detection of Rumor on Sina Weibo[J]. *Journal of Chinese Information Processing*, 2017, 31(3): 198-204.  
(祖坤琳, 赵铭伟, 郭凯, 等. 新浪微博谣言检测研究[J]. *中文信息学报*, 2017, 31(3): 198-204.)
- [23] 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*, 2013: 1103-1108.
- [24] Ma J, Gao W, Wei Z Y, et al. Detect Rumors Using Time Series of Social Context Information on Microblogging Websites[C]. *The 24th ACM International on Conference on Information and Knowledge Management*, 2015: 1751-1754.
- [25] Mao E S, Chen G, Liu X, et al. Research on Detecting Micro-Blog Rumors Based on Deep Features and Ensemble Classifier[J]. *Application Research of Computers*, 2016, 33(11): 3369-3373.  
(毛二松, 陈刚, 刘欣, 等. 基于深层特征和集成分类器的微博谣言检测研究[J]. *计算机应用研究*, 2016, 33(11): 3369-3373.)
- [26] Wu K, Yang S, Zhu K Q. False Rumors Detection on Sina Weibo by Propagation Structures[C]. *2015 IEEE 31st International Conference on Data Engineering*, 2015: 651-662.
- [27] Ma J, Gao W, Wong K F. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning[C]. *The 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017: 708-717.
- [28] Ma J, Gao W, Wong K F. Detect Rumors on Twitter by Promoting Information Campaigns with Generative Adversarial Learning[C]. *The World Wide Web Conference*, 2019: 3049-3055.
- [29] Przybyla P. Capturing the style of fake news[C]. *The 34th Conference on Artificial Intelligence, The 32nd Innovative Applications of Artificial Intelligence Conference, The 10th Symposium on Educational Advances in Artificial Intelligence*, 2020: 490-497.
- [30] Shu K, Cui L M, Wang S H, et al. DEFEND: Explainable Fake News Detection[C]. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 395-405.
- [31] Wang Yaqing, Yang Weifeng, Ma Fenglong, et al. Weak supervision for fake news detection via reinforcement learning[C]. *The 34th Conference on Artificial Intelligence, The 32nd Innovative Applications of Artificial Intelligence Conference, The 10th Symposium on Educational Advances in Artificial Intelligence*, 2020: 516-523.
- [32] Ma J, Gao W, Mitra P, et al. Detecting Rumors from Microblogs with Recurrent Neural Networks[C]. *The Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016: 3818-3824.
- [33] Chen T, Li X, Yin H Z, et al. Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection[M]. *Trends and Applications in Knowledge Discovery and Data Mining*. Cham: Springer International Publishing, 2018: 40-52.
- [34] Wu L, Liu H. Tracing Fake-News Footprints: Characterizing Social Media Messages by how they Propagate[C]. *The Eleventh ACM International Conference on Web Search and Data Mining*, 2018: 637-645.
- [35] Kumar S, Carley K. Tree LSTM with Convolution Units to Predict Stance and Rumor Veracity in Social Media Conversations[C]. *The 57th Annual Meeting of the Association for Computational Linguistics*, 2019: 5047-5058.
- [36] Ma J, Gao W, Wong K F. Detect Rumor and Stance Jointly by Neural Multi-Task Learning[C]. *The Web Conference 2018*, 2018: 585-593.
- [37] Wei P H, Xu N, Mao W J. Modeling Conversation Structure and Temporal Dynamics for Jointly Predicting Rumor Stance and Veracity[C]. *The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, 2019: 4786-4797.
- [38] 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*, 2019: 796-805.
- [39] Lu Y J, Li C T. GCAN: Graph-Aware Co-Attention Networks for Explainable Fake News Detection on Social Media[C]. *The 58th Annual Meeting of the Association for Computational Linguistics*, 2020: 505-514.
- [40] Bian Tian, Xiao Xi, Xu Tingyang, et al. Rumor detection on social media with bidirectional graph convolutional networks[C]. *The 34th Conference on Artificial Intelligence, The 32nd Innovative Applications of Artificial Intelligence Conference, The 10th Symposium on Educational Advances in Artificial Intelligence*, 2020: 549-556.
- [41] Yang X Y, Lyu Y F, Tian T, et al. Rumor Detection on Social Media with Graph Structured Adversarial Learning[C]. *The Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020: 1417-1423.
- [42] Cho K, van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation[C]. *The 2014 Conference on Empirical Methods in Natural Language Processing*, 2014: 1724-1734.
- [43] Zubiaga A, Liakata M, Procter R, et al. Analysing how People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads[EB/OL]. 2015: arXiv: 1511.07487.

<https://arxiv.org/abs/1511.07487>.

- [44] Chung J, Gulcehre C, Cho K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[EB/OL]. 2014: arXiv: 1412.3555. <https://arxiv.org/abs/1412.3555>.
- [45] Kingma D, Ba J. Adam: A method for stochastic optimization[C]. *The 3rd International Conference on Learning Representations*, 2015.
- [46] Yao Y, Rosasco L, Caponnetto A. On Early Stopping in Gradient Descent Learning[J]. *Constructive Approximation*, 2007, 26(2): 289-315.
- [47] Hamdi T, Slimi H, Bounhas I, et al. A Hybrid Approach for Fake News Detection in Twitter Based on User Features and Graph Embedding[M]. *Distributed Computing and Internet Technology*. ChamSpringer International Publishing 2019: 266-280.
- [48] Zhang Z A, Shu K R, He L N. The Theme and Characteristics of WeChat Rumors[J]. *News and Writing*, 2016(1): 60-64. (张志安, 束开荣, 何凌南. 微信谣言的主题与特征[J]. *新闻与写作*, 2016(1): 60-64.)



卫玲蔚 于 2023 年在中国科学院大学网络空间安全专业获得博士学位。现任中国科学院信息工程研究所助理研究员。研究领域为数据挖掘、自然语言处理。研究兴趣包括: 虚假信息检测、情感分析、社交网络分析。Email: weilingwei@iie.ac.cn



胡斗 于 2021 年在华北计算机系统工程研究所软件工程专业获得硕士学位。现在中国科学院信息工程研究所第三研究室攻读博士学位。研究领域为自然语言处理、数据挖掘。研究兴趣包括: 表示学习、情感分析、社交网络分析等。Email: hudou@iie.ac.cn



鲍伟楠 于 2019 年在北京科技大学物联网工程专业获得学士学位。现在中国科学院信息工程研究所第三研究室攻读博士学位。研究领域为自然语言处理。研究兴趣包括社交网络、虚假信息检测、情感分析等。Email: baoyinan@iie.ac.cn



周薇 于 2019 年在中国科学院大学计算机系统结构专业获得博士学位。现任中国科学院信息工程研究所高级工程师、硕士生导师。研究领域为数据挖掘、自然语言处理等。研究兴趣包括: 虚假信息检测、知识图谱等。Email: zhouwei@iie.ac.cn



杨近朱 于 2021 年在中国科学院大学网络空间安全专业获得博士学位。现任职于国家计算机网络应急技术处理协调中心。研究领域为社交网络、知识图谱等。研究兴趣包括: 账号对齐、知识图谱对齐等。Email: yangjinzhu@cert.org.cn



虎嵩林 于 2002 年在北京航空航天大学获得博士学位。现任中国科学院信息工程研究所研究员、博士生导师。主要研究领域为大数据存储与智能处理、知识图谱、语义服务等。Email: husonglin@iie.ac.cn