

基于相似性增强传播结构的谣言检测

林熠笛, 李弼程, 杨海君

引用本文

林熠笛, 李弼程, 杨海君. 基于相似性增强传播结构的谣言检测[J]. 计算机科学, 2024, 51(11A): 240200116-8.

LIN Yidi, LI Bicheng, YANG Haijun. Rumor Detection Based on Similarity-enhanced Propagation Structure [J]. Computer Science, 2024, 51(11A): 240200116-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于BERT模型和图注意力网络的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on BERT Model and Graph Attention Network
计算机科学, 2024, 51(11A): 240400018-7. <https://doi.org/10.11896/jsjcx.240400018>

[CCSD:面向话题的讽刺识别方法](#)

CCSD:Topic-oriented Sarcasm Detection
计算机科学, 2024, 51(9): 310-318. <https://doi.org/10.11896/jsjcx.230600217>

[基于文本和图像门控融合机制的多模态方面级情感分析](#)

Text-Image Gated Fusion Mechanism for Multimodal Aspect-based Sentiment Analysis
计算机科学, 2024, 51(9): 242-249. <https://doi.org/10.11896/jsjcx.230600117>

[基于外部先验和自先验注意力的图像描述生成方法](#)

Image Captioning Generation Method Based on External Prior and Self-prior Attention
计算机科学, 2024, 51(7): 214-220. <https://doi.org/10.11896/jsjcx.230600167>

[融入类别标签和主题信息的用户兴趣识别方法](#)

User Interest Recognition Method Incorporating Category Labels and Topic Information
计算机科学, 2024, 51(6A): 230500169-8. <https://doi.org/10.11896/jsjcx.230500169>

基于相似性增强传播结构的谣言检测

林熠笛 李弼程 杨海君

华侨大学计算机科学与技术学院 福建 厦门 361021

(22014083038@stu.hqu.edu.cn)

摘要 社交媒体的快速崛起引发了谣言传播问题,对社会造成负面影响。现有的谣言检测算法主要关注新闻内容和传播结构,却往往忽视了用户偏好的相似性可能带来的潜在影响。在浏览帖文时,用户更容易接触到与自己具有相似偏好的其他用户所传播的信息,从而助长谣言的传播。此外,现有研究常常忽视了传播结构的多样性,也忽视了新闻内容与其传播结构之间的关联。不同类型的新闻应具备不同的传播模式。因此,本文提出了一种名为“SEPS”的模型,旨在在偏好相似用户之间建立联系,再将传播结构划分为多种形式,以提取不同传播模式的特征。最后,引入对比学习和共注意力模块,增强了新闻内容和传播结构之间的相关性。实验证明,“SEPS”模型能有效检测谣言,其性能超越了最佳基线模型。

关键词: 谣言传播;传播结构;用户偏好;传播模式;相关性

中图分类号 TP391

Rumor Detection Based on Similarity-enhanced Propagation Structure

LIN Yidi, LI Bicheng and YANG Haijun

College of Computer Science and Technology, Huaqiao University, Xiamen, Fujian 361021, China

Abstract The rapid rise of social media has led to the issue of rumor dissemination, causing negative impacts on society. Existing rumor detection algorithms mainly focus on the contents and propagation structures of news, but often overlook the potential influence of user preference similarity. When browsing posts, users are more likely to encounter information spreaded by other users with similar preferences, which can further fuel the spread of rumors. Moreover, existing research frequently neglects the diversity of propagation structures and the relationship between news content and its propagation structure. Different types of news should exhibit different propagation patterns. Therefore, this paper proposes a model named “SEPS”, which aims to establish connections between users with similar preferences and then categorize propagation structures into various forms to extract features of different propagation patterns. Finally, by introducing contrastive learning and co-attention modules, the model enhances the correlation between news content and propagation structure. Experiments demonstrate that the “SEPS” model can effectively detect rumors, and its performance outperforms that of the best baseline models.

Keywords Rumor propagation, Propagation structure, User preferences, Propagation patterns, Relevance

1 引言

先前的研究明确指出,用户偏好的相似性在社交媒体平台上的信息选择和传播行为中起着重要作用^[1]。用户的社交网络结构和个人动机影响着他们在社交媒体上的信息行为。在社交媒体平台上,用户通常与拥有相似观点和兴趣的个体建立联系,形成同质化的社交网络。这种同质性使用户更容易接触到与其观点一致的信息,而难以接触到不同或相反的信息,导致回声室效应^[2]。回声室效应使用户选择性地接受和传播信息,加剧了虚假信息或假新闻在社交媒体上的传播,从而影响了谣言的传播。另一方面,用户偏好的相似性可能削弱用户对谣言信息的批判性思维和验证能力,因为用户容易受到具有相似偏好的用户的影响,导致他们忽视或拒绝与其观点或信仰不同的信息。

以过去的疫情为例,一些反疫苗的用户可能会相信并

传播一些关于疫苗的负面或错误的信息,如疫苗会导致自闭症、不孕、死亡等,而不愿意听取科学的证据和权威的建议。这样的用户往往会在社交媒体上与其他反疫苗的用户建立联系,形成一个回声室,增强了他们的信念,也扩大了谣言的传播范围和影响力。这种用户偏好的相似性不仅妨碍了用户对谣言的辨别和验证,产生从众效应,也危害了公共卫生和社会稳定。因此,本文认为,通过考虑用户偏好的相似性,可以提取更符合实际场景的传播特征。

此外,由于各新闻的性质不同,公众的反应也会不同,谣言的传播方式也会有所不同。因此需要注重新闻传播结构和新闻内容本身的相关性。过去的方法在整合新闻内容和谣言传播结构这两种不同的模态时,往往只进行简单的串联,缺乏进一步相关分析。

为此,本文提出了基于相似性增强传播结构的谣言检测(SEPS)模型。首先,该模型通过分析用户偏好的相似性并揭

基金项目:装备预研教育部联合基金项目(8091B022150)

This work was supported by the Joint Fund Project of Ministry of Education for Equipment Pre-research(8091B022150).

通信作者:李弼程(lbclm@163.com)

示相似用户之间的潜在关系,补充社交图的拓扑信息。其次, Ma 等^[3]的研究表明,考虑不同的传播结构对谣言检测有益。因此,本文将传播结构划分为两种(如图 1 所示):首先根据社交网络中的影响者角色,是“由内向外(IN2O)”的传播结构,其中影响者、意见领袖或活跃个体可能具有更大的影响力。这些个体在社交网络上拥有大量的关注者和粉丝,他们的观点和行为可以迅速传播到更广泛的受众中。另一种形式是“由外向内(O2IN)”的传播结构。在“由外向内”的传播结构中,个体可能在与其他人互动后改变他们的观点,或者在群体意见的压倒性影响下接受虚假信息。使用图卷积网络(GCN)从这两种结构中提取特征,然后进行串联和汇聚,获得统一的传播结构特征。为了更有效地整合传播结构特征和新闻内容之间的关联,采用对比学习来使传播结构特征与新闻内容更具相关性,增强模型内的数据一致性。其次引入一个共注意力模块进一步加强融合效果,增强新闻内容特征和传播结构特征之间的交互影响。最终,使用分类器来判断信息是否是谣言。

本文的主要贡献如下:

- 1) 利用用户偏好相似性揭示谣言检测中潜在的用户关系,丰富了图拓扑信息。
- 2) 考虑了两种不同形式的谣言传播结构,全面捕捉谣言传播的多样性。
- 3) 通过对比学习和共注意力机制,有效地融合了传播结构特征和新闻内容特征,增强了传播结构特征与新闻内容之间的相关性。

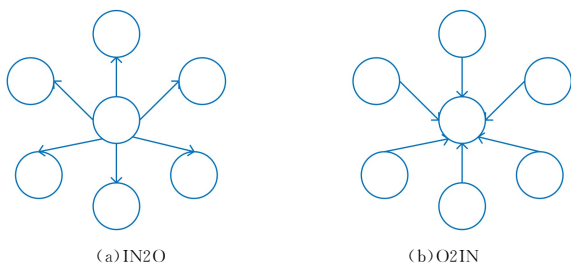


图 1 不同的谣言传播结构

Fig. 1 Different rumor propagation structures

2 相关工作

2.1 基于内容的谣言检测

基于内容的谣言检测方法主要关注传播的信息本身,而不依赖于其他因素。Wu 等^[4]引入了一种创新的分类方法,即基于图核函数的混合支持向量机分类器。SAFE^[5]利用了 TextCNN^[6],而 FakeBert^[7]则使用 BERT^[8]来编码新闻文本内容。此外,Chen 等^[9]将注意机制与循环神经网络(RNNs)相结合,更专注于不同的文本特征。Ma 等^[10]使用鉴别器作为分类器,通过生成产生冲突噪声来增强鉴别器的性能,并引入了对抗性学习方法。由于谣言在社交媒体上的传播模式与真实信息存在显著差异,随着图深度学习的不断进步,越来越多的学者将传播结构视为谣言检测的重要特征。

2.2 基于传播结构的谣言检测

基于传播结构的谣言检测方法大多关注信息传播的结构和路径。Ma 等^[11]通过对传播树进行建模来检测谣言。GC-

NFN^[12]和 GNN-CL^[13]采用图卷积网络(GCN)^[14]来编码社交网络中的谣言传播特征。GTN^[15]和 TGNF^[16]考虑了传播树中不同帖文或转推之间的时间间隔。这些方法使用了时间感知的消息聚合策略和注意力网络机制来处理时间信息。此外,Silva 等^[17]采用将图神经网络与自编码器相结合的方法,将传播网络的子集嵌入其中。这种方法能够早期检测假新闻,同时还提供了可解释性。这些研究表明,谣言的传播结构是检测谣言的有益监督信号,能够结合内容来进行谣言检测。尽管这些研究取得了不错的效果,但也有学者认为这些结构难以很好地模拟真实场景。因此,随后的许多研究将现实因素纳入考虑,以增强对传播结构的理解。

UPFD^[18]全面考虑了用户的内生偏好、新闻文本内容以及谣言的传播结构。通过分析用户的社交互动历史(如发帖行为)来获取用户对新闻的偏好信息,并利用这些信息来增强传播结构中用户节点的表示,以检测假新闻。Wei 等^[19]的研究解决了节点传播的不确定性问题,但未能充分挖掘社交网络中用户间的潜在关系。相反,MFAN^[20]强调了用户相似性对交互的影响,通过挖掘潜在关系并采用多模态融合网络来结合图像、社交网络和新闻文本内容,提升了谣言检测的性能。BiGCN^[21]通过双向图网络学习谣言的传播和扩散特征,考虑了不同的传播模式。通过不同传播方式的图卷积操作,能够更全面地捕获谣言特征。Gong 等^[22]从新闻传播的时序图中提取一系列子图,每个子图表示一次用户交互事件,并学习每个子图的表示。Yuan 等^[23]通过挖掘消息传播图中的局部语义和全局结构信息,有效地增强了传播结构与内容之间的联系;另一方面,Wu 等^[24]利用协同注意力网络进行多模态融合,着重于不同模态之间的相互作用,同样增强了传播结构与内容的联系。

然而,目前的研究在强化传播结构方面存在一些不足之处。在探索用户潜在关系时,未充分考虑用户偏好可能是一个关键因素。对于提高内容与传播结构相关性的尝试,过于简单的线性变化可能不能被视为真正提升了相关性。此外,过去的研究也未全面考虑多种因素的影响。因此,本文的目标在于充分考虑用户偏好相似性、内容与传播结构的关联性,以及多样的传播结构,以进一步增强图的传播结构,使其更贴近真实而复杂的传播场景。

3 方法

本文的研究旨在通过整合用户偏好相似性、谣言传播形式的多样性以及新闻内容与传播图结构之间的相关性,提高谣言检测的效能。

首先,在传播图特征方面,利用用户偏好特征增强图的拓扑结构。通过将用户偏好特征纳入传播图的构建中,能够更准确地刻画用户之间的关系,从而提高图的表征能力。

其次,利用图卷积网络(GCN)从图中捕捉不同形式的谣言演变的特征。通过串联和汇聚这些来自不同传播形式的特征,获得了更全面和丰富的特征表示。

然后,为增强新闻内容与传播结构之间的关联性,引入了对比学习策略。这种方法更好地促进了新闻内容和传播结构之间相互影响的理解,从而提高了特征的语义表达能力。此外,采用了一个共注意力模块,确保在融合过程中不同特征之间

的一致性和相关性,进一步提高了特征的代表效果。

最后,将上述步骤获取的特征用于谣言检测任务。整体

架构如图 2 所示。通过这个过程,实现了更准确可靠的分类结果。

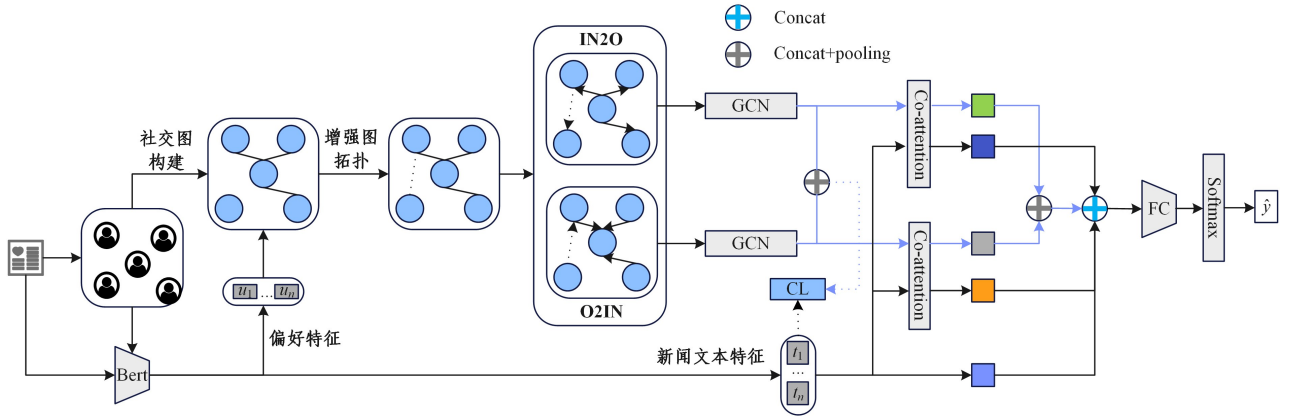


图 2 基于相似性增强传播结构的谣言检测模型

Fig. 2 Rumor detection model based on similarity-enhanced propagation structure

3.1 传播图构建

预设 $c = \{c_1, c_2, \dots, c_m\}$ 为谣言检测数据集,其中 c_i 为第 i 个事件, m 为事件个数。对于每个事件 $c_i = \{r_i, u_1^i, u_2^i, \dots, u_{n_i-1}^i, G_i\}$ 其中源发布者 r_i 是根节点, n_i-1 为事件 c_i 中的参与人数, u_j^i 表示第 i 个事件中的第 j 个参与者, G_i 为传播结构。具体来说, G_i 被定义为一个图 $G_i = \{V_i, X_i, E_i\}$, 其中 X_i 为用户特征矩阵, $E_i = \{e_{st}^i \mid s, t = 0, \dots, n_i-1\}$ 代表用户间的转发关系。传播图的构建规则为: 1) 定义一个源新闻发布者 $r_i, \{u_1^i, u_2^i, \dots, u_{n_i}^i\}$ 作为转发的用户列表, 按时间排序, 若后续转发用户关注列表中的任何其他用户, 那么将该用户与其所关注的、在用户列表中最晚转发新闻的用户之间连上一条边。 2) 若没有关注用户列表中的任何一位, 则与用户列表中粉丝数最多的用户构成一条边。 $A_i \in \{0, 1\}^{n_i \times n_i}$ 记为第 i 个事件传播图的邻接矩阵, 其中

$$a_{st}^i = \begin{cases} 1, & \text{if } e_{st}^i \in E_i \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

记 $X_i = [x_0^{i,T}, x_1^{i,T}, \dots, x_{n_i-1}^{i,T}]$ 作为图中各个节点的特征矩阵, 其中 x_0^i 代表发布者 r_i 的特征向量, x_j^i 代表用户 j 的特征向量。本文将用户偏好特征作为每个节点的特征表示。

3.2 用户偏好与新闻内容特征提取

为了方便对用户偏好配置文件的利用, 基于文献[18]中提出的方法, 利用用户的历史发布内容隐式地捕捉他们的偏好。利用包含 Twitter 上新闻内容及其社交互动的 FakeNewsNet 数据集[25], 包含新闻内容及其在 Twitter 上的社会参与信息。根据 Duo 等[18]的方法, 若用户 u 的帖文不足 200 条, 则反复利用该用户的历史帖文直到补齐 200 条; 若用户 u 的帖文超过 200 条, 则将多余的帖文裁剪。利用 Bert 取最大输入序列长度 (即 512 个令牌), 将 200 条帖文 $\{T_1, T_2, \dots, T_{200}\}$ 进行编码得到 $[t_1^T, t_2^T, \dots, t_{200}^T]^T$, 然后平均得到用户偏好表示 x :

$$x = \frac{\sum_{j=1}^{200} t_j}{200} \quad (2)$$

其中, t_j 为用户 u 的第 j 条帖文经过 Bert 编码得到的向量表示。新闻的内容也是利用 Bert 进行编码, 得到内容特征 $N = [N_1, N_2, \dots, N_n]$ 。

3.3 用户偏好相似性增强图拓扑

在社交媒体平台上, 用户往往与拥有相似观点和兴趣的人建立联系, 形成同质性的社交网络。因此, 计算不同节点之间的偏好特征相似度, 并基于相似度高的节点之间建立潜在关系。具体而言, 定义了节点嵌入矩阵 $X \in \mathbb{R}^{n_i \times d}$, 其中 d 表示特征维数, 矩阵 X 由事件中的用户偏好构成。

在此基础上, 利用用户 u_i 和 u_j 之间的余弦相似度来衡量其偏好相似性, 即:

$$\beta_{ij} = \frac{x_i \cdot x_j}{\|x_i\| \cdot \|x_j\|} \quad (3)$$

其中, x_i 和 x_j 为事件中用户 u_i 和 u_j 的偏好特征。接着, 本文推断, 当相似度大于 0.7 时, 表明两个用户之间存在潜在的关系, 即:

$$e_{ij} = \begin{cases} 0, & \text{if } \beta_{ij} < 0.7 \\ 1, & \text{otherwise} \end{cases} \quad (4)$$

其中, e_{ij} 表示用户 u_i 和 u_j 之间是否存在潜在关系。

随后, 利用推断得到的潜在关系来增强邻接矩阵 $A \in \{0, 1\}^{n_i \times n_i}$, 其中 a_{ij} 表示 A 的元素。若 $a_{ij} = 1$ 表示用户 u_i 和 u_j 之间存在一条边, 否则 $a_{ij} = 0$ 。增强后的邻接矩阵 A' 的元素 a'_{ij} 定义为:

$$a'_{ij} = \begin{cases} 0, & \text{if } e_{ij} = 0 \text{ and } a_{ij} = 0 \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

3.4 不同传播模式的特征提取

本文旨在研究两种不同形态的传播结构, 将由内向外的传播结构称为 IN2O-Graph, 将由外向内的传播结构称为 O2IN-Graph。借助 GCN 对这两种传播结构进行特征抽取, 以便分别捕获谣言不同的传播过程。

对于 IN2O-Graph, 通过如下方式进行特征抽取:

$$H_1^{\text{IN2O}} = \sigma(A'^{\text{IN2O}} X W_0^{\text{IN2O}}) \quad (6)$$

$$H_2^{\text{IN2O}} = \sigma(A'^{\text{IN2O}} H_1^{\text{IN2O}} W_1^{\text{IN2O}}) \quad (7)$$

其中, $H_1^{\text{IN2O}} \in \mathbb{R}^{n_i \times v_1}$ 和 $H_2^{\text{IN2O}} \in \mathbb{R}^{n_i \times v_2}$ 代表经过两层 GCN 抽取的 IN2O-Graph 的隐层特征, $A'^{\text{IN2O}} \in \mathbb{R}^{n_i \times n_i}$ 为 IN2O-Graph 增强后的邻接矩阵, $W_0^{\text{IN2O}} \in \mathbb{R}^{d \times v_1}$ 和 $W_1^{\text{IN2O}} \in \mathbb{R}^{v_1 \times v_2}$ 是 IN2O-Graph 的滤波参数矩阵。在此过程中, 采用了 ReLU 函数作为激活函数, 表示为 $\sigma(\cdot)$ 。为防止过拟合, 在 GCN 层中应用了 Dropout。同样地, 可以根据式 (6) 和式 (7) 计算 O2IN-

Graph 的隐层特征 H_1^{O2IN} 和 H_2^{O2IN} 。为了融合 IN2O-Graph 和 O2IN-Graph 的两种不同形式的传播特征,将 H_2^{IN2O} 和 H_2^{O2IN} 两个隐层特征连接起来,并通过一个全连接层处理,以实现隐层特征的统一。为了获得传播图特征的综合表示,采用均值池操作来聚合这两个节点表示集合中的信息,具体表达式如下:

$$H = \sigma(W_n * (H_2^{IN2O} \parallel H_2^{O2IN})) \quad (8)$$

$$s = \text{MEAN}(H) \quad (9)$$

其中, W_n 代表全连接层的权重矩阵, H 为隐层特征的总体表达, s 为统一的传播结构特征。

通过这种方法,得到了统一的传播结构特征,进一步深化了对传播图结构中信息传递机制的理解。这种对 IN2O-Graph 和 O2IN-Graph 的分析和特征抽取,有助于揭示谣言在不同传播模式下的行为和影响。

3.5 增强内容与传播结构的关联性

在融合新闻内容和谣言传播结构的特征时,传统方法通常仅采用简单的特征拼接,缺乏对模态之间深层关联的深入分析。然而,这种简单的融合方式往往难以充分挖掘不同模态数据之间的内在联系,从而限制了模型性能。为了充分发挥不同模态数据的优势,本节中引入了两种关键技术:对比学习和共注意力机制。这两种技术的综合应用不仅有助于增强模态之间的关联性,还能提升模型的特征表达能力。

3.5.1 对比学习

正负样本生成。首先,从数据集中生成正负样本。正样本包括新闻内容特征 N 和谣言传播结构特征 s 构成的对,而负样本则通过随机打乱 s 特征而产生。

相似性分数的计算。接下来,通过计算相似性分数来评估嵌入空间中正负样本之间的距离。具体而言,对于正样本,在前 $n-1$ 个维度和最后一个维度之间执行点乘运算,然后将结果相加以获得正样本的相似性分数。同样,对负样本进行类似的计算。

对比损失计算。采用对比学习损失进行模型优化。这推动正样本的相似性接近 1,而负样本的相似性接近 0。这一过程的数学表达式如下:

$$\mathcal{L}_{\text{contra}} = -\frac{1}{k} \sum_{i=1}^k \log \left(\frac{\text{pos}_i}{\text{pos}_i + \text{neg}_i + \eta} \right) \quad (10)$$

其中, k 代表样本数量, pos_i 和 neg_i 分别代表第 i 个正样本和负样本的相似性分数, η 表示边距。

3.5.2 共注意力块

先前的研究通常忽视了不同模态之间的重要关系,而这些关系对于上下文和情境的完整性至关重要。透过了解社交网络中信息的传播路径,可以洞察信息是如何从源头传播至各个用户的。结合新闻内容特征,能够揭示信息在不同用户之间的解读和互动方式。这不仅能更好地理解信息在社交媒体中的影响力与共鸣,还能为分析信息传播的深层动态提供更全面的认识。

为了打破这一限制,引入了共注意力块,一种经过改进的注意力机制。该机制接收来自两个模态的输入,其设计目标在于学习这些模态之间的相互作用和关系。更具体地说,将新闻内容特征与传播结构(IN2O-Graph 和 O2IN-Graph)特征分别输入共注意力块中,下文以某个事件的新闻内容特征 N 和传播结构特征 H_2^{IN2O} 为例进行阐述。

$$Q_N = NW^{Q_N}, K_N = NW^{K_N}, V_N = NW^{V_N} \quad (11)$$

$$Q_{IN} = H_2^{IN2O} W^{Q_{IN}}, K_{IN} = H_2^{IN2O} W^{K_{IN}} \quad (12)$$

$$V_{IN} = H_2^{IN2O} W^{V_{IN}} \quad (13)$$

$$\text{Att}(N, H_2^{IN2O}) = \text{softmax} \left(\frac{Q_N (K_{IN})^T}{\sqrt{d}} \right) V_{IN} \quad (14)$$

$$\text{Att}(H_2^{IN2O}, N) = \text{softmax} \left(\frac{Q_{IN} (K_N)^T}{\sqrt{d}} \right) V_N \quad (15)$$

$$N_{\text{ENH1}} = \text{Norm}(N + \text{Att}(N, H_2^{IN2O})) \quad (16)$$

$$H_{2\text{ENH}}^{IN2O} = \text{Norm}(H_2^{IN2O} + \text{Att}(H_2^{IN2O}, N)) \quad (17)$$

其中, $W^{Q_N}, W^{K_N}, W^{V_N}, W^{Q_{IN}}, W^{K_{IN}}, W^{V_{IN}}$ 是共注意力机制中用于进行查询、键和值的线性变换的权重矩阵。 N_{ENH1} 和 $H_{2\text{ENH}}^{IN2O}$ 分别是增强后的新闻内容特征和从 IN2O-Graph 抽取的增强传播结构特征。类似地,新闻内容特征 N 和 O2IN-Graph 的结构特征也采用了同样的步骤,可以得到 N_{ENH2} 和 $H_{2\text{ENH}}^{O2IN}$ 。

3.6 分类器

在将输入送入分类器之前,类似式(8)和式(9),让增强的传播结构特征 $H_{2\text{ENH}}^{IN2O}$ 和 $H_{2\text{ENH}}^{O2IN}$ 拼接并池化得到统一的传播结构特征 s_1 。

最后,将特征 N 、增强过的新闻内容特征 N_{ENH1} 和 N_{ENH2} 、增强过的统一结构特征 s_1 放入分类器得到最终分类结果:

$$\hat{\mathbf{y}} = \text{softmax}(\text{FC}(N \parallel s \parallel N_{\text{ENH1}} \parallel N_{\text{ENH2}} \parallel s_1)) \quad (18)$$

其中, $\hat{\mathbf{y}} \in \mathbb{R}^{1 \times m}$ 是用于预测事件标签的所有类的概率向量。然后使用交叉熵损失函数。

为了减少模型的过拟合风险,并提高泛化性,使用 R-Drop 的思想训练模型。由于 Dropout 的存在,两次完全相同的输入会产生不同的结果与分布,故把损失函数分为两个部分,即交叉熵部分和 KL 散度部分。

首先是交叉熵部分:

$$\hat{y}_1 = \text{softmax}(\text{FC}(N \parallel s \parallel N_{\text{ENH1}} \parallel N_{\text{ENH2}} \parallel s_1)) \quad (19)$$

$$\hat{y}_2 = \text{softmax}(\text{FC}(N \parallel s \parallel N_{\text{ENH1}} \parallel N_{\text{ENH2}} \parallel s_1)) \quad (20)$$

$$\mathcal{L}_{\text{classify1}} = -y \log(\hat{y}_1) - (1-y) \log(1-\hat{y}_1) \quad (21)$$

$$\mathcal{L}_{\text{classify2}} = -y \log(\hat{y}_2) - (1-y) \log(1-\hat{y}_2) \quad (22)$$

$$\mathcal{L}_{\text{classify}} = 0.5 * (\mathcal{L}_{\text{classify1}} + \mathcal{L}_{\text{classify2}}) \quad (23)$$

KL 散度部分:

$$\mathcal{L}_{\text{KL}} = \frac{1}{2} [KL(p_{\theta}(\hat{y}_1) \parallel p_{\theta}(\hat{y}_2)) + KL(p_{\theta}(\hat{y}_2) \parallel p_{\theta}(\hat{y}_1))] \quad (24)$$

$$\text{最终的损失函数可以表达为:} \quad (25)$$

$$\mathcal{L} = \mathcal{L}_{\text{classify}} + \lambda_1 \mathcal{L}_{\text{KL}} + \lambda_2 \mathcal{L}_{\text{contra}} \quad (26)$$

其中, λ_1 和 λ_2 可以用来控制两个损失的占比。

4 实验

4.1 数据集

本文的模型在两个真实世界的数据集上进行了评估: Politifact 数据集(POL)和 Gossipcop 数据集(GOS)。这两个数据集是 FakeNewsNet 数据集从两个事实检查网站收集的虚假和真实新闻的信息,同时还包括来自 Twitter 的相关社交活动数据。通过 2.1 节中的图构建方法可获得数据集的图数据,其基本统计信息如表 1 所列。表中展示了两个数据集的关键统计指标,包括图的数量、虚假新闻的数量、真实新闻的数量、节点数以及边数。

表1 数据集中传播结构图的统计

Table 1 Statistics of propagation structures in two datasets

Dataset	# Graph	# Fake	# True	# Node	# Edge
Politifact	314	157	167	41 054	40 740
Gossipcop	5 464	2 732	2 732	314 262	308 798

4.2 基线模型

本文在 POL 与 GOS 数据集上进行了对比实验,将 SEPS 模型与以下基线模型进行了比较:

1)CSI^[8]:该模型使用 LSTM 对新闻内容进行编码,用于虚假新闻检测。

2)SAFE^[5]:使用 TextCNN^[6]对新闻文本信息进行编码,用于虚假新闻检测。

3)GTN^[15]:该方法对传播树进行编码,边权重由帖文和转推之间的时间间隔确定。每个帖文/转推的表示是从用户的最近 200 条帖文中生成的,并经过 BERT 编码。

4)TGNF^[16]:该方法捕捉了传播树中的时间演变模式,聚合在时间上相邻的帖文/转推邻居。该模型还利用 BERT 获取每个帖文/转推的词向量表示。

5)GCNFN^[12]:GCNFN 是第一个采用 GCN^[14]对新闻传播图进行编码的虚假新闻检测框架,将个人信息作为用户特征和评论文本进行嵌入。

6)GNN-CL^[13]:该模型使用 DiffPool^[26]对新闻传播图进行编码。它主要用于图分类任务,节点特征从 Twitter 用户配置文件属性中提取。

7)BiGCN^[21]:采用双向 GCN,该模型对新闻传播图进行编码,以学习谣言传播的扩散特征。

8)SEAGEN^[22]:基于 Transformer 的全局演化捕捉模块,它使用自注意力机制来整合子图的表示。

9)UPFD^[18]:UPFD 模型使用 GCN 对新闻传播图进行编码,结合用户固有偏好和外部上下文信息。

10)MFAN^[20]:MFAN 模型使用 GAT^[27]对新闻传播图进行编码,集成视觉、文本和结构传播特征。为了比较,省略了该模型的视觉特征提取部分。

本文对基线模型的复现结果与本文模型进行了对比统计,表 2 列出了不同模型使用的方法。这些方法包括了内容(T)、传播结构(PS)、用户偏好(UP)、用户间相似性(US)、不同传播结构(DPS),以及新闻内容与传播结构的相关性(MMC)。特别值得注意的是,后 4 种方法,即用户偏好、用户间相似性、不同传播结构、新闻内容与传播结构的相关性,被认为是传播结构特征的增强。

表2 实验中模型所使用方法的对比

Table 2 Comparison of methods used by different models in experiment

Model	T	PS	UP	US	DPS	MMC
CSI	✓					
SAFE	✓					
GTN	✓	✓				
TGNF	✓	✓				
GCNFN	✓	✓				
GNN-CL	✓	✓				
BiGCN	✓	✓				✓
SEAGEN	✓	✓			✓	
UPFD	✓	✓	✓			
MFAN	✓	✓		✓		✓
SEPS	✓	✓	✓	✓	✓	✓

从表 2 可知,在上述模型中,CSI 和 SAFE 主要利用新闻内容对谣言进行识别;GTN,TGNF,GCNFN 和 GNN-CL 通过谣言的传播结构与内容来识别谣言;BiGCN,SEAGEN,UPFD 和 MFAN 利用增强的传播结构特征进行谣言检测。

4.3 实验环境与参数设置

在实验设置方面,本文采用 PyTorch 搭建实验环境,具体实验环境如表 3 所列。

表3 实验环境

Table 3 Experimental environment

实验环境	环境配置
操作系统	Windows10 专业版 64 位
CPU	Intel(R)Core(TM)i7-8700k
内存/GB	16
GPU	NVIDIA GeForce RTX 2080Ti
开发工具	PyCharm

SEPS 模型的参数设置如表 4 所列。

表4 参数设置

Table 4 Parameter settings

参数	值
优化器	Adam
图嵌入大小	128
批量大小	128
L2 正则化权重	0.001
迭代次数	200
η	0.2
λ_1/λ_2	0.3

4.4 实验评价指标

本实验的模型均是二分类模型,即判断新闻内容是谣言或非谣言,实验的评价指标使用准确率(ACC)和 F1 值(F1 score)。

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (25)$$

$$\begin{cases} p = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \\ F1 = \frac{2 \times p \times R}{p + R} \end{cases} \quad (26)$$

其中,TP 表示正确分类的正样本数,TN 表示正确分类的负样本数,FP 表示错误分类的正样本数,FN 表示错误分类的负样本数,p 代表精确率,R 代表召回率。

4.5 结果与讨论

4.5.1 对比实验结果与分析

首先,将基线模型与本文提出的模型在 POL 与 GOS 这两个存在真实与虚假新闻的两个数据集进行了实验,实验结果如表 5 所列。

SEPS 相比基线模型在 POL 与 GOS 这两个数据集上取得了最佳的性能。另外从表中还可以获得以下信息:

1)仅使用单一特征的方法如 SAFE 和 CSI 不如同时考虑新闻内容与传播结构进行谣言检测的方法表现得好。这表明将传播结构纳入考虑对谣言检测具有益处。此外,虽然 BiGCN 仅利用传播结构特征,但通过不同的传播形式增强了传播结构特征,在两个数据集的检测任务上超过了仅使用单一特征的谣言识别方法。同样,对于使用新闻内容与传播结构进行谣言检测的方法,如 GCNFN 和 GNN-CL,在性能上不及

UPFD, SEAGEN, MFAN 这类在内容和传播结构特征的基础上增强传播结构特征的方法。这说明在谣言检测任务中, 增强谣言的传播结构特征是有益的。

2) 比较 MFAN 与 UPFD 可以发现, 在 POL 数据集上, MFAN 的表现优于 UPFD, 而在 GOS 数据集上略逊于 UPFD。在规模较小的数据集 POL 中, MFAN 通过挖掘用户之间的相似性, 更充分地利用了用户的潜在关系, 使得信息源比 UPFD 更为丰富。此外, MFAN 通过引入注意力机制增强了内容与传播结构的关联性, 提高了特征表示的质量, 从而在 POL 数据集上超过了 UPFD。然而, 在规模较大的数据集 GOS 上, 这种优势并不十分明显, 可能是因为数据规模较大, MFAN 的优化更受到限制, 而 UPFD 在该情境下的表现相对稳定。

3) SEPS 相比 MFAN 与 UPFD 的性能差异表明, 有效利用多种增强传播结构特征能够显著提升谣言检测的性能。

表 5 模型在 POL 和 GOS 数据集上的比较结果

Table 5 Comparative results of models on POL and GOS datasets

Model	Politifact		Gossipcop	
	ACC	F1	ACC	F1
SAFE	73.30	72.87	77.19	77.19
CSI	76.02	75.99	75.01	75.01
GTN	81.67	81.53	92.38	92.38
TGNF	74.28	74.09	85.07	85.07
GNN-CL	62.90	62.25	95.09	95.09
GCNFN	83.16	83.56	96.36	96.36
BiGCN	77.73	77.73	87.44	87.44
SEAGEN	84.25	84.16	96.14	96.14
MFAN	87.35	87.69	96.33	96.33
UPFD	84.62	84.65	97.23	97.22
SEPS	88.24	88.21	97.23	97.23

4.5.2 消融实验结果与分析

为了验证 SEPS 模型中各组件的有效性, 对其子模型进行了全面比较, 分别为“-US”“-IN2O”“-O2IN”“-P”和“-MMC”。这些子模型代表 SEPS 的不同变体, 分别排除用户相似性、IN2O 传播结构特征、O2IN 传播结构特征、用户偏好特征和跨模态相关性。比较结果如表 6 所列。

表 6 SEPS 的不同变体

Table 6 Various variants of SEPS

Model	Politifact		Gossipcop	
	ACC	F1	ACC	F1
-US	83.71	83.26	94.85	94.03
-IN2O	85.07	85.28	96.26	96.23
-O2IN	87.33	87.11	96.76	96.07
-P	83.26	83.48	92.08	92.08
-MMC	86.23	86.08	96.17	96.27
SEPS	88.24	88.21	97.23	97.23

值得注意的是, 在两个数据集中, 所有消融变体的性能均不及完整的 SEPS 模型。

通过比较, 得出以下结论:

1) 当除去用户相似性时(-US), 即模型不考虑潜在关联, 其准确率相对于 SEPS 模型在两个数据集上都有所下降, 这表明考虑用户之间的相似性有助于获取更高质量的谣言传播结构特征。

2) 消融变体-IN2O 和-O2IN(不考虑多种传播结构)的性能下降说明, 不同的谣言传播结构对于谣言检测至关重要。模型需要考虑不同的传播形式, 以更全面地理解和

捕捉谣言的传播模式。

3) 从模型-P(不考虑用户偏好)的性能下降来看, 充分利用用户偏好信息有助于获得更高质量的用户关系, 提高谣言检测的准确性。

4) 通过对比-MMC 和 SEPS 的结果可以得知, 考虑新闻内容与传播结构的关系对谣言检测是有益的。不同性质的新闻内容可能导致不同的传播模式, 因此综合考虑这两者可以提升模型性能。

4.5.3 不同偏好相似度阈值的研究

为了研究建立连接的不同偏好相似度阈值对谣言检测的影响, 我们对不同相似度阈值进行了实验, 并统计了实验结果, 如图 3 所示。

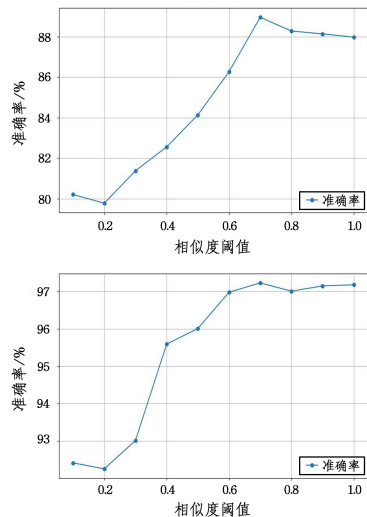


图 3 不同相似度阈值对谣言检测的影响(POL 与 GOS 数据集)

Fig. 3 Impact of different similarity threshold values on rumor detection(POL and GOS datasets)

在对不同偏好相似度进行探究的实验中, 通过取不同相似度阈值(0.1, 0.2, 0.3, ..., 1)作为建立连接的边界, 观察到了不同的检测准确度(ACC)结果。具体来说, 相似度阈值为 0.2 时, 检测准确度最差, 而相似度阈值在 0.2~0.7 之间时, 检测准确度呈现稳定上升的趋势, 之后趋于平缓甚至下降。

实验结果表明, 在建立连接时, 考虑偏好相差较大的用户可能是无益的。当相似度阈值较低时, 可能会将与用户偏好相差很大的用户纳入考虑范围, 从而导致整体性能下降。相反, 当相似度阈值在一定范围内上升时, 代表着只有相对相似的用户才被纳入考虑范围, 进而提高了检测的准确度。

这一观察揭示了在社交网络中, 用户之间的相似性对于谣言检测的重要性。适当提高相似度阈值可以过滤掉那些与用户偏好相差较大的连接, 有助于提升模型的整体性能。然而, 需要注意的是, 过高的相似度阈值也可能导致漏报, 因为某些相对相似但仍可能传播谣言的用户可能被排除在考虑范围之外。因此, 在确定相似度阈值时需要平衡准确性和全面性。

4.5.4 新闻内容与传播结构的关联性分析

消融实验的结果表明提高新闻内容与传播结构的关联性有益于谣言检测任务。为了展示本文使用对比学习提高关联性的有效性, 本文将对比学习替换为 MFAN^[20]中的模态对齐方法来进行对比, 实验结果如表 7 所列。

表 7 不同的关联性增强的方法

Table 7 Different methods of association enhancement

Model USE	Politifact		Gossipcop	
	ACC	F1	ACC	F1
模态对齐	86.97	85.61	95.35	95.77
对比学习	88.24	88.21	97.23	97.23

从表中可以观察到,采用对比学习的方法相较于模态对

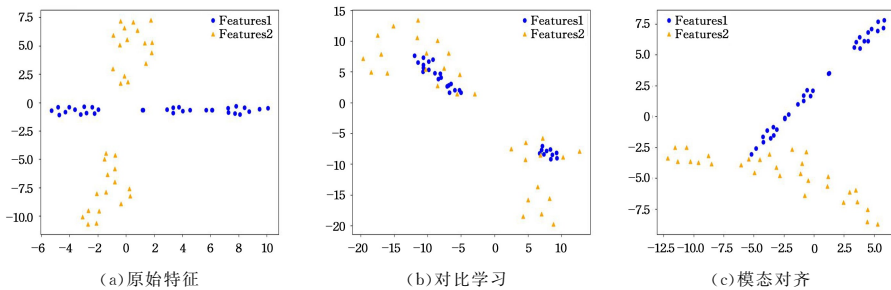


图 4 特征可视化结果

Fig. 4 Feature visualization results

观察图 4 明显可以发现,使用对比学习后的特征与使用模态对齐后的特征在分布上更为接近。这清晰地表明了这两种特征之间更高的相关性。另外,从图中还可以观察到,使用模态对齐的方法的分布情况与原始特征类似,这表明该方法只是做了线性变换,并没有提高特征之间的相关性,且经过实验证明,采用对比学习的方法能够显著提升检测性能。这与本文的观点一致,即不同性质的新闻应该具有独特的传播模式。通过增强新闻与传播结构之间的相关性,可以有效提高谣言检测的性能。

4.5.5 早期谣言检测实验结果与分析

早期检测的目标是在信息传播的初始阶段识别谣言,即在信息发布初期用户参与度较低的情况下进行谣言的识别。这是方法质量评估的另一个关键指标。由于用户列表按照时间顺序排列,通过设置一系列截断比率构建了一个早期谣言检测任务。本文认为,考虑用户偏好相似性的方法可以增强早期谣言检测的性能,即使在参与用户较少的早期阶段,也能产生显著效果。因此,选择了 3 种方法,即 SEPS, SEPS(-UPS) 和 BiGCN, 进行对比实验。其中, SEPS(-UPS) 和 BiGCN 都不考虑捕捉用户之间的潜在关系。实验结果(见图 5 和图 6)清楚地表明,在早期谣言检测任务中, SEPS 相比 SEPS(-UPS) 和 BiGCN 表现出更优越的性能。这一结果强调了即使在只有少数参与用户的早期阶段,考虑用户偏好相似性以捕捉潜在的互动关系,在少样本的情况也能有很好的表现。

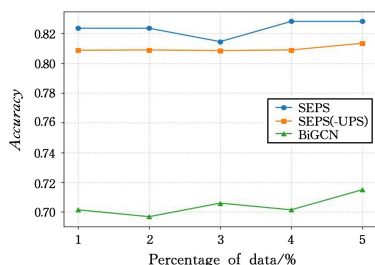


图 5 模型在早期谣言检任务的效果(POL 数据集)

Fig. 5 Performance of different models in early rumor detection tasks(POL Dataset)

齐方法在谣言检测准确率上有显著提升。这种差异可能源于对比学习在学习相似性和差异性方面的优越性,使得模型更能准确捕捉谣言与非谣言之间的细微差别。

为了进一步展示这些方法是否能提高不同模态之间的相关性,本文将原始特征分布、使用模态对齐后的特征分布与使用对比学习后的特征分布进行了可视化,如图 4 所示。

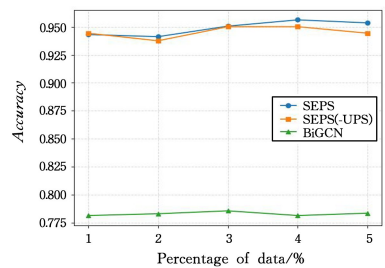


图 6 模型在早期谣言检任务的效果(GOS 数据集)

Fig. 6 Performance of different models in early rumor detection tasks(GOS Dataset)

结束语 本文深入探讨了社交网络中用户偏好相似性对谣言检测性能的影响,并提出了 SEPS 模型来全面利用这一信息。本文的模型不仅考虑了社交网络内的传播结构,还专注于用户之间的潜在关系以及新闻内容与传播结构的关联性,从而更全面地了解谣言传播的特征。通过与其他基线模型比较,实验结果验证了 SEPS 在谣言检测方面的有效性。

然而,本文研究中也存在一些限制。首先,虽然模型在考虑用户偏好相似性方面表现出显著的有效性,但仍可能忽视其他可能影响谣言传播的因素。未来的研究可以探索如何将更多的社交网络属性和用户行为特征整合到模型中,以全面分析谣言传播的机制。其次,本文的分析是建立在传播结构和用户偏好相似性的基础上的。引入更丰富的数据源,如情感分析和文本内容,以增强谣言检测效果,是未来研究的另一个有前途的方向。

本文为更深入理解社交媒体上谣言传播背后的机制提供了新的视角,并为谣言检测的进一步研究提供了有价值的见解。未来的研究可以在本文工作的基础上探索更多的特征和方法,以解决社交媒体中谣言传播不断演变的挑战。

参考文献

- [1] ACEMOGLU D, OZDAGLAR A, SIDERIUS J. Misinformation: Strategic sharing, homophily, and endogenous echo chambers[M]. National Bureau of Economic Research, 2021.
- [2] CINELLI M, DE FRANCISCI M G, GALEAZZI A, et al. The

- echo chamber effect on social media[J]. Proceedings of the National Academy of Sciences, 2021, 118(9): e2023301118.
- [3] MA J, GAO W, WONGK F. Rumor Detection on Twitter with Tree-structured Recursive Neural Networks[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers). 2018: 1980-1989.
- [4] WU K, YANG S, ZHUK Q. False rumors detection on sina weibo by propagation structures[C]// 2015 IEEE 31st International Conference on Data Engineering. IEEE, 2015: 651-662.
- [5] ZHOU X, WU J, ZAFARANI R. Similarity-Aware Multi-modal Fake News Detection [C] // Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer International Publishing, 2020: 354-367.
- [6] ZHANG Y, WALLACE B C. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification[C]// Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2017: 253-263.
- [7] KALIYAR R K, GOSWAMI A, NARANG P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach[J]. Multimedia Tools and Applications, 2021, 80(8): 11765-11788.
- [8] KENTON J D M W C, TOUTANOVAL K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]// Proceedings of NAACL-HLT. 2019: 4171-4186.
- [9] CHEN T, LI X, YIN H, et al. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection[C]// Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2018 Workshops. Springer International Publishing, 2018: 40-52.
- [10] MA J, GAO W, WONGK F. Detect rumors on twitter by promoting information campaigns with generative adversarial learning [C]// The World Wide Web Conference. 2019: 3049-3055.
- [11] MA J, GAO W. Debunking Rumors on Twitter with Tree Transformer[C]// Proceedings of the 28th International Conference on Computational Linguistics. 2020: 5455-5466.
- [12] MONTI F, FRASCA F, EYNARD D, et al. Fake news detection on social media using geometric deep learning[J]. arXiv: 1902. 06673, 2019.
- [13] HAN Y, KARUNASEKERA S, LECKIE C. Graph neural networks with continual learning for fake news detection from social media[J]. arXiv: 2007. 03316, 2020.
- [14] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv: 1609. 02907, 2016.
- [15] MATSUMOTO H, YOSHIDA S, MUNAYASU M. Propagation-based fake news detection using graph neural networks with transformer[C]// 2021 IEEE 10th Global Conference on Consumer Electronics(GCCE). IEEE, 2021: 19-20.
- [16] SONG C, SHU K, WU B. Temporally evolving graph neural network for fake news detection [J]. Information Processing & Management, 2021, 58(6): 102712.
- [17] SILVA A, HAN Y, LUO L, et al. Propagation2Vec: Embedding partial propagation networks for explainable fake news early detection [J]. Information Processing & Management, 2021, 58(5): 102618.
- [18] DOU Y, SHU K, XIA C, et al. User preference-aware fake news detection[C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 2051-2055.
- [19] WEI L, HU D, ZHOU W, et al. Towards Propagation Uncertainty: Edge-enhanced Bayesian Graph Convolutional Networks for Rumor Detection [C] // Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing(Volume 1: Long Papers). 2021: 3845-3854.
- [20] ZHENG J, ZHANG X, GUO S, et al. MFAN: Multi-modal Feature-enhanced Attention Networks for Rumor Detection[C]// IJCAI, 2022: 2413-2419.
- [21] BIAN T, XIAO X, XU T, et al. Rumor detection on social media with bi-directional graph convolutional networks[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 549-556.
- [22] GONG S, SINNOTT R O, QI J, et al. Fake News Detection Through Temporally Evolving User Interactions[C]// Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer Nature Switzerland, 2023: 137-148.
- [23] YUAN C, MA Q, ZHOU W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection [C]// 2019 IEEE International Conference on Data Mining(ICDM). IEEE, 2019: 796-805.
- [24] WU Y, ZHAN P, ZHANG Y, et al. Multimodal fusion with co-attention networks for fake news detection[C]// Findings of the Association for Computational Linguistics(ACL-IJCNLP 2021). 2021: 2560-2569.
- [25] SHU K, MAHUESWARAN D, WANG S, et al. Fakenewsnet: A data repository with news content, social context, and spatio-temporal information for studying fake news on social media [J]. Big Data, 2020, 8(3): 171-188.
- [26] YING R, YOU J, MORRIS C, et al. Hierarchical graph representation learning with differentiable pooling[C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018: 4805-4815.
- [27] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[J]. Stat, 2018, 1050: 4.



LIN Yidi, born in 2000, postgraduate. His main research interests include natural language processing and graph neural network.



LI Bicheng, born in 1970, Ph.D, professor, Ph.D supervisor. His main research interests include intelligent information processing, network ideological security, network public opinion monitoring and guidance, and big data analysis and mining.