计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目： 融合动态传播和社区结构的社交媒体谣言检测模型
作者： 强子珊，顾益军
网络首发日期： 2023-11-29
引用格式： 强子珊，顾益军. 融合动态传播和社区结构的社交媒体谣言检测模型[J/OL]. 计算机工程与应用. https://link.cnki.net/urlid/11.2127.TP.20231129.1051.006

网络首发时间：2023-11-29 14:43:55
网络首发地址：https://link.cnki.net/urlid/11.2127.TP.20231129.1051.006

融合动态传播和社区结构的社交媒体谣言检测模型

强子珊， 顾益军

中国人民公安大学 信息网络安全学院，北京 100032

摘 要：为解决现有谣言检测模型对时间信息利用不充分的问题，同时验证利用谣言传播的社区结构特征可以提高谣言检测模型的性能，提出一种融合动态传播和社区结构的社交媒体谣言检测模型 Dy_PCRD (Rumor Detection Model Based on Dynamic Propagation and Community Structure)，一方面使用图卷积网络提取谣言传播的结构特征，另一方面先根据谣言内容和传播结构划分话题社区，再使用一种新型的注意力计算方法提取谣言的社区结构特征，将二者分别输入时间注意力单元对其动态变化规律进行建模，最后基于所获得的嵌入表示对谣言进行分类。三个公开数据集上的实验结果表明，在相同条件下，相较于基线模型，其准确率及其他各评价指标均有所提升，验证了社区结构特征、动态性特征以及相关注意力计算方法对提升谣言检测模型性能的有效性。

关键词：谣言检测；动态图；社区结构；传播；注意力机制
文献标志码:A 中图分类号:TP391 doi：10.3778/j.issn.1002-8331.2306-0347

Rumor Detection Model Based on Dynamic Propagation and Community Structure

QIANG Zishan, GU Yijun

College of Information and Cyber Security, People's Public Security University of China, Beijing 100032, China

Abstract: To address the insufficient utilization of time information, and to verify that the community structure features of rumor propagation can improve rumor detection model's performance. Dy_PCRD(Rumor Detection Model Based on Dynamic Propagation and Community Structure) model is proposed which integrates dynamic propagation and community structure. On the one hand, GCN can extract structural features of rumor propagation, and on the other hand, topic communities are divided based on rumor content and propagation structure, and then uses a new attention calculation method to extract community structural features. They are inputted into temporal attention units to model their dynamic changes and classify the rumors. The experimental results on three public datasets show that under the same condition, its accuracy and other evaluation indicators have been improved compared to the baseline model, verifying the effectiveness of community structure features, dynamic features, and related attention calculation methods in improving the performance of rumor detection models.

Key words: rumor detection; dynamic graph; community structure; propagation; attention mechanism

学术界通常将谣言定义为一种未经证实的、毫无根据的观点，由于其具有较强的迷惑性、煽动性，且

基金项目：中国人民公安大学网络空间安全执法技术双一流创新研究专项（2023SYL07）。
作者简介：强子珊(1999－),女,硕士研究生,CCF 学生会员,研究方向为谣言检测,E-mail: 201721430042@stu.ppsuc.edu.cn;**顾益军**(1968－),男,通讯作者,博士,教授,研究方向为网络情报技术,E-mail: guyijun@ppsuc.edu.cn。

通常符合当时的时代特征，因而很容易被公众所轻信。近年来，随着互联网的迅速发展和移动设备的广泛使用，各种社交平台的使用人群也逐渐向不同年龄层扩展，其中部分人群由于缺乏相关知识容易相信这些谣言，进而产生恐慌情绪甚至损失财产，严重影响了社会的和谐稳定，同时破坏了政府公信力，因此进行谣言检测是一项至关重要的工作。

早期的检测方法是通过人工选择提取谣言的内容、相关用户、传播结构等特征，然后使用决策树分类器(Decision Tree Classification, DTC)、随机森林分类器(Random Forest Classification, RFC)、支持向量机(Support Vector Machine, SVM)、贝叶斯网络等传统机器学习算法对其进行分类，由于此类方法是基于谣言的浅层特征，模型的性能受到限制。随着神经网络和自然语言处理的不断发展，BERT(Bi-directional Encoder Representation from Transformers)、CNN(Convolutional Neural Networks)、RNN(Recurrent Neural Networks)[1-5]等深度学习算法开始被用于提取谣言内容、情感等方面的深层特征，有效提升了模型效果，但此类算法只能按照序列提取信息，无法有效提取谣言整体的传播结构特征，而图神经网络GNN(Graph Neural Networks)[6-8]不仅解决了上述问题，还可以将谣言整体的传播结构与其内容、相关用户等信息结合起来，因而得到了广泛的应用，同时随着研究的深入，有学者发现谣言的各类特征之间还存在着异质性，因而将异质图的相关模型应用其中。但上述研究通常针对谣言传播所形成的最终状态，包括内容、结构等，忽略了在谣言传播过程中，其相关特征往往会随时间动态变化。

现有的基于时间信息构建的谣言检测模型还存在一定的不足：这些模型大多采用显式的时间特征，即通过单独提取时间信息和其他类型特征，再对二者进行拼接[9][10]，这会导致不同类型特征之间缺乏交互；另外，在谣言传播过程中会产生与之相关的各类话题，将同一话题有关的帖子集定义为一个话题社区，同一社区内的帖子所表达的观点、立场、情感等存在相似性，而不同社区内的帖子则存在较大差异，由此可见，谣言传播还存在一定的社区性，但现有研究忽略了此类特征，同时，随着谣言的传播、知识的增加以及一些意见领袖发表看法，相关用户的观点、兴趣等也会产生变化，使得不同时间段的话题社区也会存在一定的差别。

总之，现有研究中存在时间信息利用不充分、忽略了谣言社区结构特征等问题，如何有效利用上述信息提高谣言检测模型性能是当前研究面临的挑战。因此，本文考虑到谣言传播的动态性和社区性，提出了一种融合动态传播和社区结构的社交媒体谣言检测模型 Dy_PCRD，同时开展谣言动态传播结构和动态社区结构两类特征的提取工作，最后，融合两类特征并基于获得的特征表示对谣言进行分类。本文的创新点和贡献如下：

（1）提出谣言话题社区的概念，将内容和关联关系相近的谣言帖子集合定义为一个话题社区，并将其引入谣言检测工作。

（2）提出一种融合动态传播和社区结构的社交媒体谣言检测模型 Dy_PCRD，通过学习谣言传播结构和社区结构两类特征的动态变化规律对谣言传播过程进行建模。

（3）采用一种新的注意力机制计算方式提取谣言的社区结构特征，通过卷积计算同一话题社区内部和不同话题社区之间的注意力权重更新特征。

（4）利用公开的数据集进行实验，实验结果验证了 Dy_PCRD 模型的有效性以及社区结构特征、动态性特征、相关话题社区融合方法的优越性。

1 相关研究

谣言检测的具体工作流程可总结为先提取谣言的相关特征，然后基于所提取的特征对谣言进行分类，现有研究通常将其定义为一项二分类任务，即判断其是否为谣言，Ma 等人[11]提出谣言还存在未经验证的情况，因而将分类粒度定义得更为细致。本文将现有的谣言检测方法按照其所提取的特征类型分为内容特征和传播结构特征。

1.1 基于内容特征的谣言检测方法

内容特征是通过谣言当中的词汇特征、文本长度、语法特征等反映其所表达的思想、情感等信息，早期的研究[12]使用相关机器学习算法基于内容表示对谣言进行分类，之后也有研究使用 TF-IDF[13]先对谣言内容进行初步分析，再进行谣言检测。此类方法基于谣言的浅层特征，不能得到深层的语义信息。

随着深度学习技术和自然语言处理的不断发展，出现了很多的新型模型，Yang 等人[1]使用 BERT 和 CNN 从用户评论当中提取信息，再利用 LSTM（Long Short Term Memory）进一步提取其中的情感信息并将其与内容特征进行融合作为谣言特征，刘政等人[2]提出使用 CNN 挖掘帖子之间的关联进而获得谣言的特征表示，周丽娜等人[3]提出使用 TextCNN 提取深层的文本特征，但谣言本身是一种时序数据，时间是谣言的一个重要信息，CNN 无法有效提取谣言的时间信息，其性能受到限制。

由于 RNN 能够解决 CNN 不能提取时间信息的问题[4][5]，Ma 等人[14]提出使用 RNN 从谣言帖子当中同时学习文本和时间特征，Wu 等人[15]采用 LSTM 基于谣言的时序特征进行建模，Chen 等人[16]提出使用 RNN 学习帖子序列的时间表示，有效提升了模型的效果。但与此同时，一些不法分子使用深度伪造技术模仿真实新闻的语言风格、情感特征等生成谣言，给检测工作带来一定的困难，因此仅使用谣言的内容特征不能满足当前的现实需要。

1.2 基于传播结构的谣言检测方法

传播结构特征指的是谣言传播呈现出的一种模式特征，具体包括帖子之间的交互情况、所形成的传播路径长度、传播时间长短等，考虑到谣言传播结构会随时间变化，本文将现有的基于传播结构的检测方法分为静态的传播结构和动态的传播结构。

1.2.1 静态的传播结构

Wu 等人[17]将传播特征引入谣言检测当中，并结合随机游走内核和 RBF 内核提取谣言的结构特征，还有学者使用 SIS[18]、SEIR[19]等传染病模型进行谣言源头的检测，其原理类似于限制流行病传播，通过识别网络中的谣言源，及时隔离谣言传播，以限制其造成的损害，但信息传播与传染病在很多方面都有明显的区别[20]，因而会影响检测效果。

Ma 等人[21]提出一种基于树结构的递归神经网络模型 RvNN（Recursive Neural Network），结合谣言的内容和传播方向提取高维特征，但该模型提取到的传播信息依然有限。

图神经网络不仅能提取谣言传播的结构特征，还可以有效地将其与内容、用户等建立关系，Monti 等人[6]使用堆叠图卷积神经网络从用户、用户评论等数据组成的异质图中提取谣言传播的结构特征，解决了相关 NLP 算法缺乏相关知识的不足，Bian 等人[7]使用双向的图卷积网络 Bi-GCN，提取到了谣言自上而下和自下而上两种传播模式，Cui 等人[8]提出一种 KAGN 模型，从知识图谱中融合外部知识来检测谣言。但此类方法通常是对谣言传播形成的最终状态提取特征，忽略了谣言传播结构随时间的动态变化。

1.2.2 动态的传播结构

Wang 等人[9]提出一种根据谣言的动态传播结构进行谣言检测的模型，使用分区方法来建模传播结构的动态演化，然后使用 GRU(Gated Recurrent Unit）学习动态结构的表示，最后对获得的动态结构表示和文本表示进行融合，该方法分别提取谣言的内容和结构特征，这样会使得两类特征之间缺乏交互。

Song 等人[22]提出真假新闻相关的推文数量随时间的变化趋势存在差别，因而构建了一种基于时间演化的图神经网络进行假新闻检测，Wei 等人[23]构建了一种基于 transformer 的动态图表示学习方法用于谣言识别，捕捉到了结构和时间序列的长期依赖性，Choi 等人[24]提出一种基于注意力机制的动态图卷积网络模型 DyGCN(Dynamic Graph Convolution Neural Network)进行谣言检测，有效地将谣言的内容与传播结构结合在一起提取特征，同时也捕捉到了谣言传播的动态信息，但上述方法均忽略了谣言传播的社区性特征。

2 融合动态传播和社区结构的社交媒体谣言检测模型

针对上述问题，本文提出了一种融合动态传播和社区结构的社交媒体谣言检测模型 Dy_PCRD，包含动态传播结构提取和动态社区结构提取两个模块，其整体结构如图 1 所示，首先按照谣言传播的时间线划分各时段的谣言帖子及其间的关联关系，构建各时段的谣言传播图，然后两个特征提取模块分别提取谣言传播结构和社区结构的动态变化特征，最后融合两类特征并输出分类结果。

2.1 谣言动态传播图的构建

考虑到谣言传播的时间信息，结合相关帖子的发表时间 T，按照固定的时间间隔 γ 对相关帖子及连边

进行划分，将帖子作为节点，帖子之间的转发、评论等关系建立边，进而构建不同时段的谣言传播图，获得谣言传播图快照集合 $G=\{G_{t_1},G_{t_2},\cdots,G_{t_n}\}$，$G_{t_i}=\{V_{t_i},E_{t_i}\}$，同时获得各图快照所对应的邻接矩阵 $A=\{A_{t_1},A_{t_2},\cdots,A_{t_n}\}$，其中时间间隔 γ 的计算过程如下：

$$\gamma=\frac{T_{\max}-T_{\min}}{n} \tag{1}$$

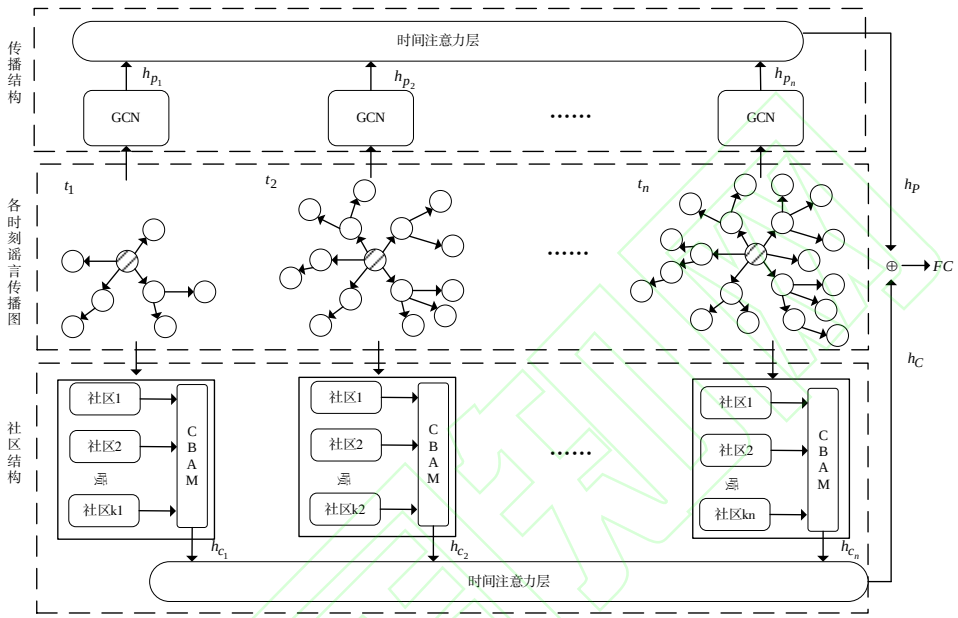

图 1 Dy_PCRD 框架

Fig.1 Dy_PCRD framework

2.2 动态传播结构特征提取

该模块的作用是提取谣言传播结构的动态变化特征：根据谣言帖子之间的关系和帖子内容利用图卷积网络提取各时段谣言的特征表示。在动态图 G 中，令 $A_{t_i}\in R^{N_{t_i}\times N_{t_i}}$ 和 $X\in R^{N_{t_i}\times d_0}$ 分别代表图快照 G_{t_i} 的邻接矩阵和特征矩阵，其中 N_{t_i} 表示该图快照中节点的个数，d_0 表示原始特征的维度。

传播结构提取过程如图 2 所示，共设置两个图卷积层提取各图快照特征，由于原始帖子所包含的信息量最大，计算过程中将各节点特征与其根节点特征进行拼接，从而增强源帖的信息[7]，为便于表示，此后对各图快照特征提取过程的描述中均省去时间下标，传播特征提取过程如下：

$$H_1^{TD}=\sigma(\hat{D}^{-1/2}\tilde{A}^{TD}\hat{D}^{-1/2}XW_0) \tag{2}$$

$$\tilde{H}_1^{TD}=concat(H_1^{TD},X^{root}) \tag{3}$$

$$H_2^{TD}=\sigma(\hat{D}^{-1/2}\tilde{A}^{TD}\hat{D}^{-1/2}\tilde{H}_1^{TD}W_1) \tag{4}$$

$$H^{TD}=concat(H_2^{TD},H_{1TD}^{root}) \tag{5}$$

其中，$\tilde{A}=A+I_N$，W_0 和 W_1 均为可学习的参数，所获得的 H^{TD} 即为该时段内谣言自上而下的传播特征，同理，还可获得自下而上的传播特征 H^{BU}，对二者拼接后进行平均池化即可获得该时段内传播特征的表示向量 $h_p\in R^{1\times d_p}$：

$$h_p=MEAN(concat(H^{TD},H^{BU})) \tag{6}$$

各时段传播特征可表示为 $h_{p_1},h_{p_2},\cdots,h_{p_n}$，对其进行 stack 得到矩阵 $H\in R^{n\times d_p}$，传入时间注意力单元，该单元应用自注意力机制融合不同时段的特征，计算方式使用缩放的点积型注意力：

$$Attention(Q,K,V)=Soft\max(\frac{QK^T}{\sqrt{d_p}})V \tag{7}$$

其中，令 $Q=K=V=H$，d_p 表示输出向量的维度，最终获得谣言动态传播结构特征的表示向量 h_P。

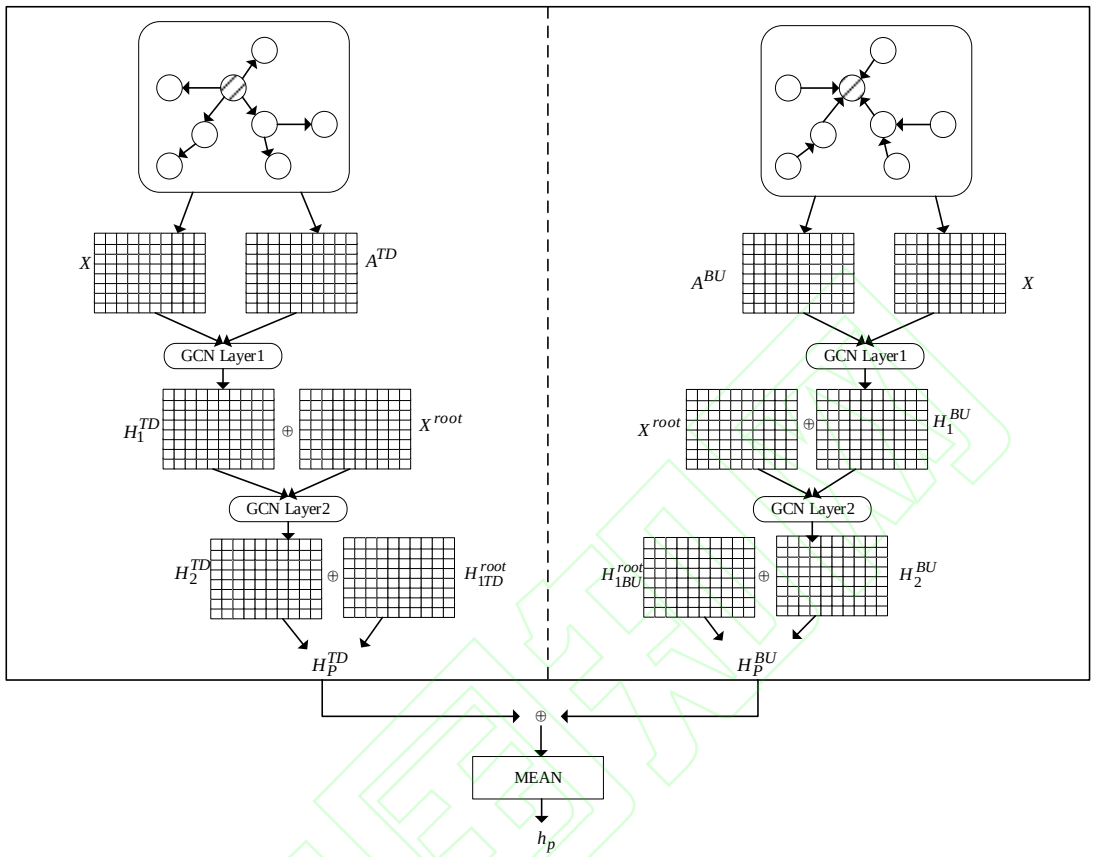

图 2 传播结构特征提取

Fig.2 Propagation structure feature extraction

2.3 动态社区结构特征提取

该模块旨在提取谣言社区结构特征的动态变化：首先基于各图快照中节点的特征划分话题社区，然后提取各话题社区的特征，并对其进行融合获得各时段谣言的社区结构特征，最后将各时段特征输入时间注意力单元获得其动态表示。

2.3.1 话题社区的划分

同一社区内的谣言帖子相似度较高，原因在于，一方面，同一社区内的数条帖子中所包含的词语、标点符号等可能存在重合，另一方面，这些帖子也极可能是对同一条帖子的转发或者评论，因此，同一社区内帖子的相似性不仅体现于文本内容，还体现在传播结构上。

考虑到上述两个方面，社区划分过程如图 3 所示，包括图自编码器模块和聚类模块。

（1）图自编码器模块

将相应图快照的邻接矩阵 A 和原始特征矩阵 X 输入编码器，经过两层 GCN 获得嵌入表示 Z：

$$Z=\sigma(\hat{D}^{-1/2}\tilde{A}\hat{D}^{-1/2}\sigma(\hat{D}^{-1/2}\tilde{A}\hat{D}^{-1/2}XW_0)W_1) \tag{8}$$

Z 同时包含谣言的内容和传播结构信息，然后将其输入到解码器，计算内积获得重构的邻接矩阵 $\hat{A}$：

$$\hat{A}=\delta(Z,Z^T) \tag{9}$$

设置损失函数，通过最小化损失优化图自编码器模型，使得 $\hat{A}$ 与原始邻接矩阵 A 的相似度更高，进而获得各谣言事件最优的特征表示 Z。

（2）聚类模块

聚类模块应用 k-means 算法划分各话题社区，该算法通过计算距离衡量数据之间的相似性，距离越大，表明其相似度越低，本文选取欧氏距离作为距离计算标准，根据帖子嵌入表示的相似性对除源帖外的帖子进行聚类，首先随机指定聚类数目 k 和质心 $c=\{c_1,c_2,\cdots,c_k\}$，再计算各帖子与质心之间的距离，计算过程如下：

$$dis(p,c_i)=\sqrt{\sum_{j=1}^{d_c}(p_j-c_{ij})^2} \tag{10}$$

其中，p 和 c_i 分别表示目标帖子和第 i 个质心帖子的特征表示，d_c 表示二者的特征维度，根据所计算的欧氏距离衡量目标帖子与该质心的相似度，将相似度最高的帖子分配至 c_i 的簇中，再对 k 个簇中的对象求平均值，确定新的聚类中心，降低误差，同时计算聚类结果的轮廓系数，将轮廓系数最高的 k 值确定为聚类数目，由于所有帖子均围绕源帖展开讨论，因而最后将源帖分配至各社区，即可得到相应图快照的各话题社区。

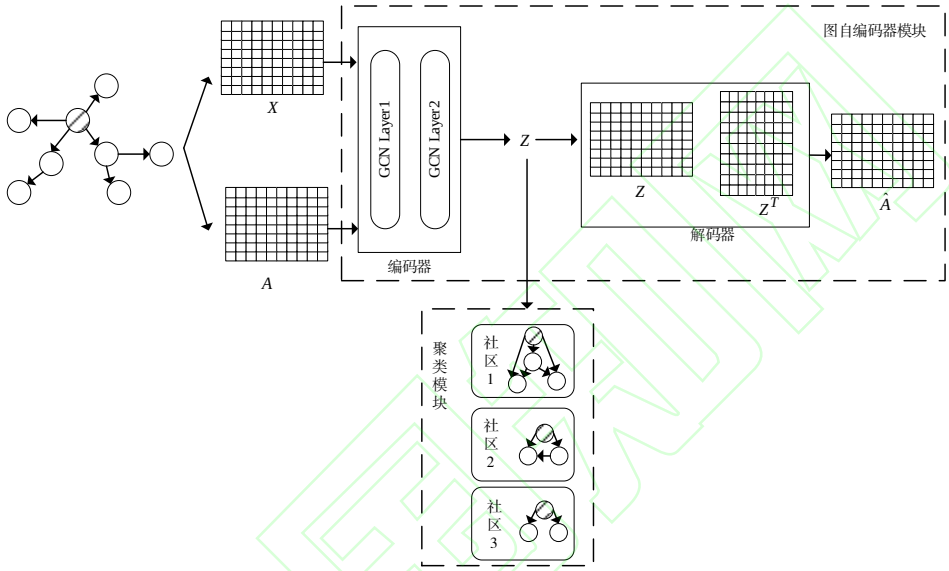

图 3 社区划分流程

Fig.3 Community division process

2.3.2 社区结构特征提取

由于源贴表达了谣言的中心话题，因此规定其与同一社区中的其他帖子之间均有关联，其余帖子则按照原传播图快照中节点之间的关联关系保留信息，这里使用图注意力网络(Graph attention network, GAT)按照社区关联提取各时段的特征。令 com_i 表示图快照中的第 i 个社区，$i\in[1,k]$，$com_i=\{p_0,p_1,\cdots,p_{N-1}\}$，其中 p_0 表示源帖，N 表示该话题社区中帖子的个数，对于帖子 p_j，按照社区节点关联计算其与其他帖子的相似系数：

$$e_{p_jp_z}=a([Wh_{p_j}\,\|\,Wh_{p_z}])\quad z\in[0,N-1] \tag{11}$$

使用 $Softmax$ 对社区中其他帖子进行归一化计算注意力权重：

$$\alpha_{p_jp_z}=\frac{\exp(Leaky\ \mathrm{Re}LU(e_{p_jp_z}))}{\sum_{m\in[0,N-1]}\exp(Leaky\ \mathrm{Re}LU(e_{p_jp_m}))} \tag{12}$$

最后，根据所计算的注意力权重，对相关特征进行加权求和：

$$h'_{p_j}=\sigma(\sum\nolimits_{m\in[0,N-1]}\alpha_{p_jp_m}Wh_{p_m}) \tag{13}$$

获得的 $h'_{p_j}\in R^{1\times d_c}$ 即为 com_i 中帖子 p_j 的表示向量，按照相同的方法获得同社区中其他帖子的特征表示，即可获得该社区的特征表示矩阵 H_{com_i}，平均池化后获得其表示向量 $h_{com_i}\in R^{1\times d_c}$，进而获得各社区的特征表示集合 $h=\{h_{com_1},h_{com_2},\cdots,h_{com_k}\}$，$h\in R^{k\times 1\times d_c}$。

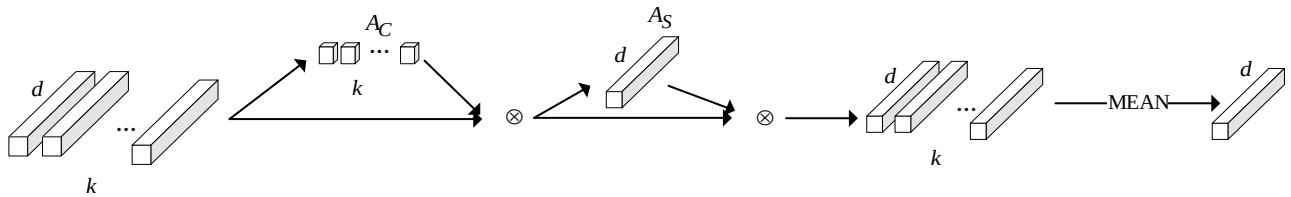

图 4 CBAM 注意力模型

Fig.4 CBAM attention model

融合不同话题社区的特征表示即可获得该时段谣言的社区结构特征。受计算机视觉领域的注意力机制启发，应用 CBAM(Convolutional Block Attention Module)[25]，从社区内部和社区之间两个维度进行卷积计算注意力权重，其结构如图 4 所示，首先将各社区的特征表示集合 h 输入模型，通过卷积计算各社区内部的注意力权重 $A_C \in R^{k\times1\times1}$，并对二者对应相乘，随后对结果再次进行卷积计算社区之间的注意力权重 $A_S \in R^{1\times1\times d_c}$，并将其与先前的结果再次对应相乘，对结果进行平均化和降维获得对应时段的社区结构表示 $h_c \in R^{1\times d_c}$。

同理，可以获得各时段社区结构表示 $h_{c_1}, h_{c_2}, \cdots, h_{c_n}$，将其传入时间注意力单元，该单元同样利用自注意力机制，计算方式与 2.2 节一致，最终获得谣言动态社区结构特征的表示向量 h_C。

2.4 谣言分类

经过上述模块可获得谣言的动态传播结构特征 h_P 与社区结构特征 h_C，将二者进行拼接作为谣言最终的特征表示 h：

$$h = concat(h_P, h_C) \tag{14}$$

将其输入分类层，得到谣言的预测标签 $\hat{y}$：

$$\hat{y} = Soft\max(wh + b) \tag{15}$$

最后使用交叉熵损失函数计算损失：

$$L = \sum_{i=1}^{|class|} - y_i \log \hat{y}_i \tag{16}$$

3 实验

3.1 实验数据集

为有效评估模型效果，本文在 Pheme[26]和 Twitter15[21]、Twitter16[21]三个公开的数据集上进行实验。其中，Pheme 数据集中包含两种标签，分别是真谣言(True Rumor, TR)和假谣言(False Rumor, FR)，Twitter15 和 Twitter16 数据集包含四种标签，分别是真谣言(True Rumor, TR)、假谣言(False Rumor, FR)、未经验证的谣言(Unverified Rumor, UR)和非谣言(Non-rumor, NR)。相关数据的统计情况如表 1 所示：

表 1 数据统计情况

Table 1 Data statistics

	Pheme	Twitter15	Twitter16
事件总数	5748	1490	818
真谣言数(TF)	3654	372	205
假谣言数(FR)	2094	370	205
未经验证(UR)	0	374	203
非谣言数(NR)	0	374	205

3.2 基线模型

本文选取基于内容特征和传播结构的几种经典模型作为基线方法同本文模型进行对比：

DTC[12]：一种基于决策树的谣言检测模型。

RFC[12]：一种基于随机森林的谣言检测模型。

SVM-RBF[12]：一种带有径向基核函数(Radial Basis Function，RBF)的支持向量机谣言检测模型。

RvNN[21]：一种基于递归神经网络的谣言检测模型，基于传播结构和内容采用递归的方法进行建模开展谣言检测工作。

Bi-GCN[7]：一种基于双向图卷积网络的谣言检测模型，使用图卷积网络根据谣言传播的两个方向提取特征。

DyGCN[24]：一种基于动态图和注意力机制的谣言检测模型，分别提取每个图快照的特征，再应用注意力机制对其融合。

Dy_PCRD：本文提出的融合动态传播和社区结构的社交媒体谣言检测模型，在 DyGCN 的基础上增加了动态社区结构特征提取模块。

其中，DTC、RFC 和 SVM-RBF 三种基于机器学习的方法将谣言的内容作为输入，RvNN、Bi-GCN、DyGCN 和 Dy_PCRD 四种基于深度学习的方法将谣言的内容和传播结构作为输入。

3.3 实验设置和评价指标

参考基线模型的设置，由于 Twitter15 和 Twitter16 数据集规模较小，实验中使用五折交叉验证，预处理方法与 RvNN、Bi-GCN、DyGCN 一致，即选用数据集中 TF-IDF 值最高的 5000 个单词的词频作为原始特征，其输入模型的维度为 5000。

由于上述基线模型在实验中未使用 Pheme 数据集，因此预处理阶段使用 BERT 预训练模型提取每条推文的特征表示，其输入模型的维度设置为 128。

为公平地检验模型效果，三个数据集的其他实验设置保持一致。实验环境使用 Pytorch，按照 7：1：2 划分训练集、验证集和测试集，图编码器模块输出特征的初始维度设置为 128，学习率设置为 0.0005，batchsize 设置为 64，dropout 设置为 0.5，使用 Adam 算法对参数进行优化，epoch 设置为 100，同时应用早停法，当验证集的损失值在 10 个 patience 内不再下降时停止训练，使用验证集的最优参数在测试集上检验模型效果。另外由于文献[24]已经探究过图快照的划分方法及图快照数量对实验结果的影响，因此本实验与其最优设置保持一致，即按照时间线划分图快照，数量设置为 3。

为验证模型的有效性，同时也为便于同基线模型进行对比，对 Twitter15 和 Twitter16 数据集使用准确率和各类别的 F1 值进行评价，Pheme 数据集标签有两种，因此使用二分类的评价指标准确率、精确率、召回率和 F1 值进行评价。

3.4 实验结果与分析

3.4.1 对比实验

模型在三个数据集上的结果如表 2、表 3、表 4 所示，其中 Twitter15 和 Twitter16 数据集上 RvNN 的结果引用自文献[21]。

表 2 Pheme 数据集实验结果

Table 2 Results of Pheme dataset

模型名称	Acc	Prec	Rec	F1
DTC	0.537	0.536	0.537	0.536
RFC	0.609	0.554	0.609	0.555
SVM-RBF	0.627	0.393	0.627	0.483
RvNN	0.633	0.558	0.633	0.505
Bi-GCN	0.666	0.414	0.498	0.403
DyGCN	0.831	0.817	0.822	0.820
Dy_PCRD	**0.858**	**0.847**	**0.848**	**0.847**

表 3 Twitter15 数据集实验结果

Table 3 Results of Twitter15 dataset

模型名称	Acc	F1			
		TR	FR	UR	NR
DTC	0.660	0.726	0.622	0.675	0.622
RFC	0.544	0.621	0.575	0.405	0.527
SVM-RBF	0.320	0.320	0.191	0.047	0.419
RvNN	0.723	0.821	0.758	0.654	0.682
Bi-GCN	0.798	0.743	0.829	**0.874**	**0.746**
DyGCN	0.806	0.871	0.790	0.748	0.752
Dy_PCRD	**0.810**	**0.876**	**0.834**	0.776	0.738

表 4 Twitter16 数据集实验结果

Table 4 Results of Twitter16 dataset

模型名称	Acc	F1			
		TR	FR	UR	NR
DTC	0.654	0.789	0.558	0.742	0.539
RFC	0.469	0.571	0.529	0.375	0.385
SVM-RBF	0.309	0.174	0.318	0.083	0.398
RvNN	0.737	0.835	0.743	0.708	0.662
Bi-GCN	0.807	0.860	0.777	0.818	0.737
DyGCN	0.814	0.787	0.760	0.736	0.688
Dy_PCRD	**0.829**	**0.897**	**0.842**	**0.827**	**0.741**

Dy_PCRD 模型在三个数据集上的准确率分别达到了 85.8%、81%和 82.9%，均高于其他模型，表明该模型在谣言检测中表现出良好的效果，通过对实验结果进行进一步分析可以得出如下结论：

基于神经网络的深度学习模型(RvNN、Bi-GCN、DyGCN 和 Dy_PCRD)效果明显要比传统机器学习模型(DTC、RFC 和 SVM-RBF)更好。一方面，传统机器学习模型只是根据谣言的内容进行分类，无法有效利用谣言的传播结构特征；另一方面，传统机器学习模型无法提取到谣言更深层次的语义特征，因此在谣言检测工作中可以利用的信息非常有限，而神经网络模型则可以捕获其所表达的语义关系，因而表现出良好的性能。

基于神经网络的深度学习模型当中，Bi-GCN、DyGCN 和 Dy_PCRD 三种基于 GNN 的模型效果整体上优于 RvNN 模型。这是因为 RvNN 模型是按照序列提取特征，无法学习到谣言传播整体的结构特征，而 GNN 则可以通过赋予不同邻居节点不同的权重有效地聚合邻居信息，实验结果也验证了 GNN 模型的有效性。但是，观察 Bi-GCN 模型的实验结果可以发现，其在 Twitter15 数据集上标签 UR 和 NR 的 F1 值高于其他模型，其在 Twitter16 数据集上除 Dy_PCRD 模型外，两类标签的 F1 值也达到了最高，这可能是因为该模型的架构设计或者参数设定更适合提取上述两类标签特征，同时，其在 Pheme 数据集上的分类准确率虽高于 RvNN，但其他评价指标均低于 RvNN，这可能

是因为该数据集的正负样本数量并不均衡，TR 的数量明显高于 FR，说明 Bi-GCN 模型在不均衡样本上的效果还有待提升。

DyGCN、Dy_PCRD 两种基于动态图的模型鲁棒性更强，而且其在标签不均衡的 Pheme 数据集上也表现出良好的效果，验证了谣言传播过程中动态特征的有效性，另外，本文提出的 Dy_PCRD 模型相较于 DyGCN 模型在三个数据集上的准确率分别提升了 2.7%、0.4%和 1.5%，同时，除 Twitter15 数据集上 NR 类别的 F1 值外，Dy_PCRD 模型的其他各评价指标也均高于 DyGCN 模型，说明在谣言传播过程中，相关帖子所表达的观点、立场等存在差别，利用社区结构特征可以有效提升谣言检测模型的效果。

3.4.2 消融实验

为验证动态性特征以及不同的社区特征融合方法对模型效果的影响，设置如下三组消融实验：

(1)w/o dynamic：移除模型的动态性特征提取模块，即针对谣言传播形成的最终状态分别提取传播结构和社区结构特征。

(2)Dy_PCRD_{dot}：动态社区结构特征提取模块中采用缩放的点积型注意力融合不同社区特征。

(3)Dy_PCRD_{add}：动态社区结构特征提取模块中采用加型注意力融合不同社区特征。

实验结果如表 5、表 6、表 7 所示，分析实验结果可以发现，首先，移除动态特征提取模块后，模型在三个数据集上的所有评价指标均有不同程度的下降，验证了动态特征在谣言检测中的有效性。

其次，对比应用注意力机制融合话题社区的几种方法，CBAM 注意力计算方法的效果优于另外两种，这可能是因为缩放点积型注意力和加型注意力通过自注意力机制更新参数，由于计算方式不同，导致计算结果存在不确定性，最终呈现的效果可能会受到数据集分布的影响，而 CBAM 通过卷积分别计算话题社区内部和不同话题社区之间的注意力权重，并依此来更新社区结构特征，对各话题重要性加以利用的同时，增强了模型的可解释性。

表 5 Pheme 数据集消融实验结果对比

Fig.5 Comparison of Ablation Results of Pheme Dataset

	Acc	Pre	Rec	F1
w/o dynamic	0.839	0.825	0.837	0.831
Dy_PCRD_{dot}	0.832	0.818	0.822	0.820
Dy_PCRD_{add}	0.829	0.816	0.816	0.816
Dy_PCRD	**0.858**	**0.847**	**0.848**	**0.847**

表 6 Twitter15 数据集消融实验结果对比

Fig.6 Comparison of Ablation Results of Twitter15 Dataset

	Acc	F1 值			
		TR	FR	UR	NR
w/o dynamic	0.787	0.852	0.811	0.734	0.736
Dy_PCRD_{dot}	0.791	0.814	0.814	0.767	0.726
Dy_PCRD_{add}	0.786	0.861	0.797	0.757	0.725
Dy_PCRD	**0.810**	**0.876**	**0.834**	**0.776**	**0.738**

表 7 Twitter16 数据集消融实验结果对比

Fig.7 Comparison of Ablation Results of Twitter16 Dataset

	Acc	F1 值			
		TR	FR	UR	NR
w/o dynamic	0.815	0.891	0.820	0.804	0.706
Dy_PCRD_{dot}	0.7831	0.8712	0.7504	0.767	0.725
Dy_PCRD_{add}	0.799	0.883	0.802	0.794	0.683
Dy_PCRD	**0.829**	**0.897**	**0.842**	**0.827**	**0.741**

3.4.3 动态性效果实验

为进一步评估检测时间对谣言检测模型效果的影响，在划分图快照的 3 个时间点基于当时谣言传播所形成的静态状态分别进行检测，三个数据集的实验结果如表 8 所示。

表 8 Twitter16 数据集消融实验结果对比

Fig.8 Comparison of Ablation Results of Twitter16 Dataset

	Pheme	Twitter15	Twitter16
t0	0.819	0.767	0.804
t1	0.819	0.793	0.811
t2	0.839	0.787	0.815
Dy_PCRD	**0.858**	**0.81**	**0.829**

Pheme 和 Twitter16 两个数据集的结果显示，随着谣言传播时间的增长，检测的准确率也会上升，说明通常情况下，谣言与非谣言的结构差异会随时间的增长变得更明显，但在 Twitter15 数据集上，虽然前期的准确率随时间的增长而提高，但是后期反而略有降低，这可能是因为随着传播时间的增长，谣言与非谣言的结构也有可能会比前期更小，因此，如果在谣言检测中直接利用当时的结构信息很可能导致检测结果不准确。

本文提出的 Dy_PCRD 模型在三个数据集上的准确率均高于直接利用前期任一时刻传播状态的准确率，这是因为 Dy_PCRD 模型利用注意力机制分别赋予谣

言传播不同阶段以不同的权重，通过训练对重要时间段的特征设置更高的权重，进而能够解决不同时间点检测结果不同可能产生不一致的问题，更好地检测出谣言。

3.4.4 参数实验

本文通过对比模型在不同学习率、batchsize 及不同聚类维度下模型的准确率对相关参数进行进一步分析。

学习率大小对于提升模型性能起着较为重要的作用[27]，在保持其他实验设置不变的情况下，本文在 0.0001 和 0.001 之间的六组数据上进行实验探究学习率大小对模型准确率的影响，实验结果如图 5 所示。

通过分析图 5 的折线图可以得出如下结论：首先，随着学习率的增加，准确率呈现先上升后下降的趋势，说明随着学习率的提高，模型能够更快地收敛到局部最优解，从而导致准确率的整体增加。然而，当学习率过高时，模型可能会陷入局部最优解附近的震荡状态，使准确率下降。其次，实验结果显示，当学习率为 0.0005 时，模型在合理的迭代次数内收敛到了较好的结果，获得了最高准确率，因此，合适的学习率能够有效提升模型性能。

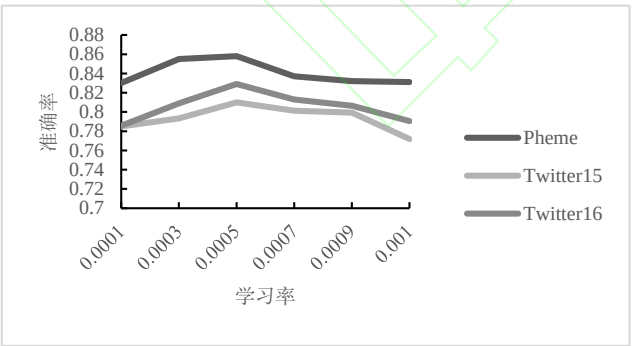

图 5 准确率和学习率的关系

Fig.5 Relationship between accuracy and learning rate.

当前计算机视觉领域也有研究发现合适的 batchs-ize 有助于提升模型效果[28][29]，为了验证 batchsize 对 Dy_PCRD 模型性能的影响，在保持其他实验设置不变的情况下，分别将 batchsize 设置为 16、32、64、 128 评估模型分类的准确率。实验结果如图 6 所示。

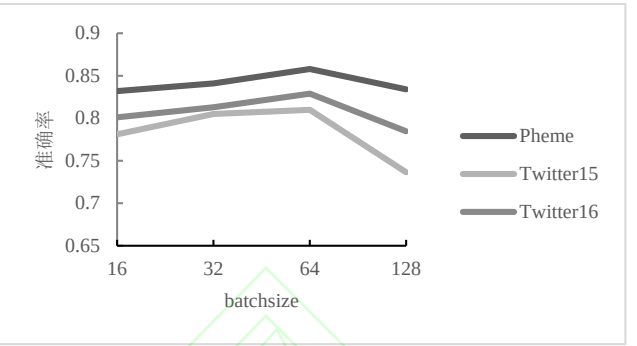

图 6 准确率和 batchsize 的关系

Fig.6 Relationship between accuracy and batchsize

通过分析图 6 的折线图可以得出如下结论：随着 batchsize 的增加，三个数据集上的准确率均呈现先上升后下降的趋势，在 batchsize 为 64 时，准确率均达到最高点。这可能是因为当 batchsize 过小时数据分布不均匀，导致模型泛化能力下降，而 batchsize 过大时使用更多的数据，又可能会导致一些重要信息被忽略掉，同时也会消耗更多的计算资源，因此合适的 batchsize 可以让数据提供更多的随机性从而帮助模型学习到更多的细节，并且可以更快地更新梯度、避免过拟合。

另外，对比三个数据集的变化趋势，当 batchsize 设置为 128 时，Twitter15 和 Twitter16 两个数据集有较大幅度的下降，但 Pheme 数据集只有轻微的下降，这可能是因为两个 Twitter 数据集的数据量较少，因此过大的 batchsize 并不合适，而 pheme 数据集的数据量更大，将 batchsize 设置为 64 和 128 均能得到较为不错的效果。因此，综合考虑上述因素，认为将 batchsize 设置为 64 对于 Dy_PCRD 模型是比较合适的。

在话题社区的划分模块，k-means 算法基于图自编码器对帖子内容进行编码获得的嵌入表示对帖子进行聚类进而得到话题社区，由于 k-means 基于欧氏距离进行聚类，其结果可能会受到特征维度影响，因此将图自编码器模块输出的特征维度分别设置为 16、32、64、128，通过比较模型在不同聚类结果下的分类准确率进而确定帖子最佳的编码维度，实验结果如图 7 所示。

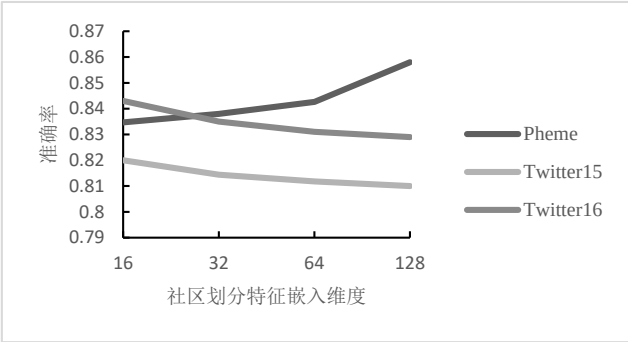

图 7　准确率和特征维度的关系

Fig.7　Relationship between accuracy and feature dimension

分析图 7 可以发现，图编码器输出的特征维度大小的确会影响模型效果，同时三个数据集的波动情况存在差别，这可能与模型原始输入有关，Pheme 数据集上的准确率随维度的升高而升高，可能是由于该数据集利用 BERT 编码帖子的原始内容，而 BERT 能够深入挖掘文本信息，且随着特征维度的升高，数据点在每个维度上的分布差异越大，聚类所获得的话题社区更精准，使得模型的效果也更好。

通过调整图编码器输出维度，在维度设置为 16 时，Twitter15 和 Twitter16 两个数据集的准确率又分别提升了 1%和 1.4%，同时图 7 显示，其准确率随维度的升高而降低，可能是由于这两个数据集将词频作为原始输入特征，且特征向量较为稀疏，同时这两个数据集本身也并不大，因此在欧氏空间中，特征维度较低时，聚类效果会更好，模型的效果也更佳。

4　结束语

本文探究了谣言传播过程中其社区结构特征的重要性，同时针对现有模型对谣言传播的时间信息利用不充分的问题，提出了一种基于动态传播和社区结构的社交媒体谣言检测模型，分别学习每个时间段内谣言的传播结构和社区结构特征，并将各时段的表示向量输入时间注意力单元学习二者的动态变化规律，最后拼接两类特征并输入分类器进行分类，另外，还应用 CBAM 注意力机制融合不同话题社区来更好地提取谣言社区结构特征。三个数据集上的实验结果也验证了该模型的有效性。

但是，有研究发现动态特征存在随机性[30]，本文的实验结果也表明动态性特征在不同数据集上的效果同样存在差别，因此在接下来的研究中，将考虑通过评估时间信息的重要性进一步提升谣言检测模型的效果，另外，本文利用 k-means 算法和图表示模型 GAT 提取话题社区结构特征，目前有学者提出相关社区表示算法[31][32]，因此在未来的工作中，将考虑应用相关模型提取到更丰富的谣言话题信息。

参考文献：

[1] YANG J, PAN Y. COVID-19 rumor detection on social networks based on content information and user response[J]. Frontiers in Physics, 2021: 570.

[2] 刘政,卫志华,张韧弦.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.

LIU Z, WEI Z H, ZHANG R X. Rumor detection model based on convolutional neural network[J].Journal of Computer Applications,2017,37(11):3053-3056+3100.

[3] 周丽娜,谭励,曹娟,等.基于卷积神经网络的食品安全领域谣言检测方法[J].计算机应用与软件, 2022, 39(03): 45-50+115.

ZHOU L N, TAN L, CAO J, et al. Rumor detection method in food safety field based on convolutional neural network[J]. Computer Applications and Software,2022,39(03):45-50+115.

[4] GHANEM B, PONZETTO S P, ROSSO P, et al. Fakeflow: fake news detection by modeling the flow of affective information.[C]// Proceedings of the 16th conference of the european chapter of the association for computational linguistics.2021:679–689.

[5] ALKHODAIR S A, DING S H H, FUNG B C M, et al. Detecting breaking news rumors of emerging topics in social media[J]. Information Processing & Management, 2020, 57(2): 102018.

[6] MONTI F, FRASCA F, EYNARD D, et al. Fake news detection on social media using geometric deep learning[J]. Cornell University - arXiv,Cornell University - arXiv, 2019.

[7] 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, 34(01): 549-556.

[8] CUI W, SHANG M. Kagn: knowledge-powered attention and graph convolutional networks for social media rumor detection[J] Journal of Big Data. 2023.

[9] WANG S, KONG Q, WANG Y, et al. Enhancing rumor detection in social media using dynamic propagation structures[C]//2019 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, 2019: 41-46.

[10] HUANG, Q., ZHOU, C., WU, J. et al. Deep spatial–temporal structure learning for rumor detection on twitter[J]. Neural Computing and Applications, 2023, 35(18): 12995-13005.

[11] MA J, GAO W, WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]. Association for Computational Linguistics, 2017.

[12] 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. IEEE, 2013: 1103-1108.
[13] BHATTACHARJEE U, SRIJITH P K, DESARKAR M S. Term specific tf-idf boosting for detection of rumours in social networks[C]//2019 11th International Conference on Communication Systems & Networks (COMSNETS). IEEE, 2019: 726-731.
[14] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[J]. International Joint Conference on Artificial Intelligence,International Joint Conference on Artificial Intelligence, 2016.
[15] WU L, LIU H. Tracing fake-news footprints: Characterizing social media messages by how they propagate. [C]// Proceedings of the eleventh ACM international conference on web search and data mining. 2018: 637–645.
[16] 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, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Melbourne, VIC, Australia, June 3, 2018, Revised Selected Papers 22. Springer International Publishing, 2018: 40-52.
[17] 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, IEEE.2015: 651–662.
[18] WANG Z, ZHANG W, TAN C W. On inferring rumor source for SIS model under multiple observations[C]//2015 IEEE International Conference on Digital Signal Processing (DSP). IEEE, 2015: 755-759.
[19] ZHOU Y, WU C, ZHU Q, et al. Rumor source detection in networks based on the SEIR model[J]. IEEE access, 2019, 7: 45240-45258.
[20] 薛海涛,王莉,杨延杰,等.基于用户传播网络与消息内容融合的谣言检测模型[J].计算机应用,2021,41(12):3540-3545.
XUE H T, WANG L, YANG Y J, et al. Rumor detection model based on user propagation network and message content[J]. Journal of Computer Applications,2021,41(12):3540-3545.
[21] MA J, GAO W, WONG K F. Rumor detection on twitter with tree-structured recursive neural networks[C]. Association for Computational Linguistics, 2018.
[22] SONG C, SHU K, WU B. Temporally evolving graph neural network for fake news detection[J]. Information Processing & Management, 2021: 102712.
[23] WEI S, WU B, XIANG A, et al. Dgtr: dynamic graph transformer for rumor detection[J]. Frontiers in Research Metrics and Analytics, 2023.
[24] CHOI J, KO T, CHOI Y, et al. Dynamic graph convolutional networks with attention mechanism for rumor detection on social media[J]. Plos one, 2021, 16(8): e0256039.
[25] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
[26] ZUBIAGA A, LIAKATA M, PROCTER R. Exploiting context for rumour detection in social media[C]//Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I 9. Springer International Publishing, 2017: 109-123.
[27] FAGBOHUNGBE O, QIAN L. Impact of learning rate on noise resistant property of deep learning models[J]. arXiv preprint arXiv:2205.07856, 2022.
[28] BROCK A, DONAHUE J, SIMONYAN K. Large scale gan training for high fidelity natural image synthesis[J]. International Conference on Learning Representations, International Conference on Learning Representations, 2018.
[29] CAI G, WANG Y, HE L. Learning smooth representation for unsupervised domain adaptation[J]. arXiv: Computer Vision and Pattern Recognition, arXiv: Computer Vision and Pattern Recognition, 2019.
[30] ZHU Y, LYU F, HU C, et al. Encoder-Decoder Architecture for Supervised Dynamic Graph Learning: A Survey[J]. arXiv preprint arXiv:2203.10480, 2022.
[31] LI M, LU S, ZHANG L, et al. A community detection method for social network based on community embedding[J]. IEEE Transactions on Computational Social Systems, 2021, 8(2): 308-318.
[32] WANG X, CUI P, WANG J, et al. Community preserving network embedding[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022.