



计算机科学

COMPUTER SCIENCE

用于谣言检测的图卷积时空注意力融合与图重构方法

陈鑫, 荣欢, 郭尚斌, 杨彬

引用本文

陈鑫, 荣欢, 郭尚斌, 杨彬. 用于谣言检测的图卷积时空注意力融合与图重构方法[J]. 计算机科学, 2024, 51(11): 54-64.

CHEN Xin, RONG Huan, GUO Shangbin, YANG Bin. [Graph Convolution Spatio-Temporal Attention Fusion and Graph Reconstruction Method for Rumor Detection](#) [J]. Computer Science, 2024, 51(11): 54-64.

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)

[基于注意力机制和深层特征优化的混凝土路面裂缝检测](#)

Crack Detection of Concrete Pavement Based on Attention Mechanism and Deep Feature Optimization
计算机科学, 2024, 51(11): 198-204. <https://doi.org/10.11896/jsjcx.240100082>

[面向自动驾驶的高精度实时语义分割算法架构](#)

High-precision Real-time Semantic Segmentation Algorithm Architecture for Autonomous Driving
计算机科学, 2024, 51(11): 174-181. <https://doi.org/10.11896/jsjcx.231000009>

[基于特征插值的深度图对比聚类算法](#)

Feature Interpolation Based Deep Graph Contrastive Clustering Algorithm
计算机科学, 2024, 51(11): 157-165. <https://doi.org/10.11896/jsjcx.231000209>

[一种基于层次超图注意力神经网络的服务推荐算法](#)

Hierarchical Hypergraph-based Attention Neural Network for Service Recommendation
计算机科学, 2024, 51(11): 103-111. <https://doi.org/10.11896/jsjcx.231100010>

[基于跨模态交互与特征融合网络的假新闻检测方法](#)

Fake News Detection Based on Cross-modal Interaction and Feature Fusion Network
计算机科学, 2024, 51(11): 23-29. <https://doi.org/10.11896/jsjcx.231200186>

用于谣言检测的图卷积时空注意力融合与图重构方法

陈鑫 荣欢 郭尚斌 杨彬

南京信息工程大学人工智能(未来技术)学院 南京 210044

(2409749761@qq.com)

摘要 互联网的快速发展给人们带来了便利的社交,同时也为谣言的产生和传播创造了条件。谣言的传播速度之快、影响之恶劣引起了广泛的关注。为了及时识别出谣言以采取截断措施,谣言检测变得尤为重要。然而,在复杂的社交网络中,谣言传播状态动态变化、传播过程中干扰信息的存在,以及传播的不确定性等均给谣言检测带来了困难。为了解决上述问题,提出了一种用于谣言检测的图卷积时空注意力融合与图重构方法(STAFRCN)。该方法对所有待检测言论进行两次检测以降低误判率,首先使用一种时间渐进卷积模块(TPC)在时间维度上整合待检测言论传播状态信息;然后分别在时间和空间两个方面使用注意力提取其主要传播特征信息并融合,对融合结果进行第一次谣言检测;随后基于 LSTM 预测和图重构方法调整待检测言论传播总图结构,将其与第一次检测结果结合进行第二次检测。实验结果表明,STAFRCN 在 Twitter15, Twitter16 和 Weibo 数据集上的检测准确率分别为 92.2%, 91.8% 和 96.5%, 与 SOTA 模型(KAGN)相比,准确率在 3 个数据集上分别提升了 3.0%, 1.5% 和 1.4%。

关键词: 谣言检测; 图神经网络; 图卷积; 注意力机制

中图分类号 TP389.1

Graph Convolution Spatio-Temporal Attention Fusion and Graph Reconstruction Method for Rumor Detection

CHEN Xin, RONG Huan, GUO Shangbin and YANG Bin

School of Artificial Intelligence(School of Future Technology), Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract The rapid development of the Internet has brought convenience to people's social life, but it also creates conditions for the generation and spread of rumors. The fast propagation speed and bad impact of rumors have attracted wide social attention. However, in complex social networks, the dynamic change of rumor propagation state, the existence of interference information in the propagation process, and the uncertainty of propagation all bring difficulties to rumor detection. In order to solve the above problems, this study proposes a graph convolution spatio-temporal attention fusion and graph reconstruction method (STAFRCN) for rumor detection, and all the speeches to be detected are detected twice to reduce the probability of misjudgment. Firstly, a temporal progressive convolution module (TPC) is used to integrate the propagation status information of the speeches to be detected in the time dimension. Then, attention is used to extract and fuse the main propagation feature information in two aspects of time and space respectively, and the fusion result is used for the first rumor detection. After that, the total graph structure of the detected speech propagation is adjusted based on long short-term memory (LSTM) prediction and graph reconstruction method. It is combined with the first detection results for the second detection. Experiments show that the detection accuracy of STAFRCN on Twitter15, Twitter16 and Weibo datasets is 92.2%, 91.8% and 96.5%, respectively. Compared with SOTA model (KAGN), the accuracy is increased by 3.0%, 1.5% and 1.4% on the 3 datasets, respectively.

Keywords Rumor detection, Graph neural network, Graph convolution, Attention mechanism

到稿日期:2024-03-29 返修日期:2024-07-24

基金项目:国家自然科学基金(62102187);江苏省自然科学基金(省基础研究计划)(BK20210639)

This work was supported by the National Natural Science Foundation of China(62102187) and Natural Science Foundation of Jiangsu Province, China(Basic Research Program)(BK20210639).

通信作者:荣欢(1227558210@qq.com)

1 引言

在互联网快速发展、社交媒体广泛普及的当下,人们的交流方式发生了翻天覆地的变化。社交媒体为人与人之间的交流提供了一个良好的平台,使言论发表变得即时、碎片化、匿名化,这在给人们带来便利的同时,也引发了虚假信息传播(即谣言)等问题。谣言一般指内容与实际不符的言论。社交网络上的信息杂乱无章,人们往往在不经意间就传播了谣言,而谣言的传播不仅可以误导公众、损害个人或组织的声誉,甚至还可能引起不必要的恐慌、造成财产损失、破坏社会稳定。因此,谣言检测作为维护信息安全和社会安定的重要手段,具有重大意义^[1]。

谣言检测是一个二分类任务,主要是通过分析待检测言论本身的内容以及与其相关的评论等来将该言论分类为谣言(Rumor)或事实(Fact)。

目前针对谣言检测任务已有许多研究,文献[2]提出了一种融合多模态信息的社交网络谣言检测方法,使用多头注意力机制提取文本和视觉特征用于谣言检测;文献[3]提出了一种基于时间的图注意网络谣言检测方法,其模型由图形注意网络、时间注意机制和门控循环单元组成,模型通过学习多个历史潜在状态感知谣言传播在时间上的动态演化过程,并以此实现谣言检测。通过研究谣言的传播结构来进行谣言检测受到了众多学者的青睐,文献[4]的研究也证明了根据传播结构来检测谣言的有效性;文献[5]提出了一种基于边推断增强对比学习的社交媒体谣言检测模型,其模型能够调整传播图的边以增强传播图数据,提高检测效果;文献[6]研究了基于传播树结构节点及路径双注意力的谣言检测模型,使用注意力提取谣言传播结构上的深层结构和语义信息并融合,进行谣言检测;文献[7]提出了一种传播路径树学习模型,在使用传播结构检测谣言时综合考虑用户的影响力、情感反馈和内容等特征。上述研究在谣言检测任务中均取得了一定的成效,这也间接表明传播结构与言论真伪之间存在紧密联系。

然而,现实世界中的谣言大多以文本的形式存在,数据获取及处理相对容易,而多模态数据包含了不同类型的信息,如文本、图像、视频等,虽然能够提供更全面的信息线索,提高模型的准确性,但数据通常需要从不同来源获取,并且需要额外的处理和融合步骤,增加了数据处理、训练的复杂性和计算成本;此外,多模态方法在面对只有文本信息的待测谣言时优势并不明显。因此,基于文本信息的谣言检测在许多场景下相较于多模态信息谣言检测更为便利和高效,综合性能更优。现实中谣言的语义、句法与真实言论相差无几,且谣言传播是一个连续的过程,谣言引发的相关评论会相互影响,并非独立存在,某一条与谣言高度相关的评论甚至能决定谣言的演变方向。此外,谣言在传播过程中引发的评论具有随机性,因而谣言传播时表现出来的结构偶然因素较多,难以准确反映其本质,故根据一个确定的传播结构来检测一个真假未知、动态演变的言论是否为谣言存在较大的缺陷。现有谣言检测方法大多只进行一次检测,而单一检测可能会受到数据质量、模型

偏见、特征选择等各种因素的影响,这也在一定程度上增加了误判、漏判的可能,进而影响最终的检测准确率。

针对上述问题,本研究提出了一种用于谣言检测的图卷积时空注意力融合与图重构方法(STAFRCN),该方法综合考虑谣言评论之间的相互依赖关系、动态演化以及传播的随机性,对待测言论进行二次检测,防止检测结果受到特定数据集、特定领域或特定算法的限制,提高了方法的泛化性能,增加了检测结果的可信度。STAFRCN由时空注意力融合网络和图重构网络构成,时空注意力融合网络由时间渐进卷积模块(Temporal Progressive Convolution)和时空注意力融合模块(Spatio-Temporal Attention Fusion)构成,图重构网络则由谣言预测模块(Rumor Prediction)和图重构模块(Graph Reconstruction)构成。本研究的主要贡献如下:

1)采用预训练语言模型 RoBERTa^[8] (Robustly Optimized BERT Pretraining Approach)生成文本嵌入表示,再通过卷积处理得到传播图表示。

2)提出了一种时间渐进卷积模块(TPC),在时间维度上逐时刻提取整合传播时动态变化的待测言论传播状态信息,建立评论之间的相互依赖关系。

3)提出了时空注意力融合模块(STAF),从时间和空间两个维度提取待测言论特征信息,既考虑了待测言论传播在时间上的动态变化性,也考虑了其在整体空间结构上的分布状态。

4)提出了图重构网络(GR),其中的谣言预测模块使用 LSTM 根据当前待测言论传播信息预测未来时刻传播信息,图重构模块据此重构传播总图,增加待测言论传播结构的不确定性,模拟真实谣言传播过程。

5)在3个数据集上的实验结果表明,STAFRCN方法在应对谣言检测任务时的表现优于现有方法,谣言检测准确率均有所提高。

2 相关工作

目前,谣言检测方法包括传统特征工程、机器学习算法以及深度学习方法。传统方法通常依赖文本特征、社交网络结构特征、用户行为特征等来进行谣言检测^[9],主要方法有支持向量机法^[10-11] (Support Vector Machine, SVM)、随机森林法^[12]、决策树法^[13-14]等。这些方法通常需要手动提取特征、标记数据,成本较高、易引入人为偏差,且传统方法提取的特征可能不足以捕获复杂语义和句法信息,导致模型在新数据集上泛化能力有限。

为了弥补传统方法的不足,基于深度学习的谣言检测方法利用神经网络强大的表征学习能力,从大规模数据中学习并挖掘出隐藏在文本背后的模式和特征以进行有效的谣言检测。文献[15]提出了一种基于图神经网络的谣言检测方法,首先将被检测帖子和相关言论表示为图,然后进行图注意层处理,最后由一个密集神经网络进行分类。文献[16]提出了基于分层注意力网络的谣言检测方法,该方法使用带有注意力机制的双向 GRU(Gate Recurrent Unit)网络学习微博上随

时间变化的文本、用户特征信息用于谣言检测。文献[17]研究了基于门控图神经网络(Gated Graph Neural Network, GGNN)的谣言检测算法,利用消息转发关系构建信息转发图,通过两个融合门控图卷积网络模块聚合邻居节点信息,得到言论传播结构,并以此检测谣言。文献[18]提出了基于传播用户代表性特征学习的早期谣言检测方法(RPPC),该方法对网络平台数据中的用户特征分布进行统计分析并选择具有代表性的用户特征,以此进行谣言检测。文献[19]提出了一种图卷积网络方法,捕获谣言传播中的局部和全局依赖关系以进行有效的谣言检测。

基于言论本身的语义特征来检测谣言能起到一定的效果,但现实世界中谣言与真实言论极为相似,仅研究言论本身的特征通常难以达到理想的效果,而研究言论的传播结构能弥补上述不足,为谣言检测提供帮助。文献[20]提出了一种基于节点级对比的传播结构融合模型 PFNC,将具有相似传播结构样本的一致和互补信息融合用于谣言检测;文献[21]提出了一种利用图卷积网络捕获用户行为以进行谣言检测的模型,该模型将根据用户行为构建的谣言传播树结构编码为具有内容语义和传播线索的向量用于检测;文献[22]使用 Tree-LSTM 多任务学习法,根据立场信息构建传播树并在树的根节点进行谣言检测。上述方法均取得了一定的效果,也进一步表明,研究传播结构对提高谣言检测准确率有很大的帮助。然而,现实中的谣言传播结构并非一成不变,环境因素、相关评论等都会影响谣言的传播,因此,谣言在现实中表现出来的传播结构只是一种偶然状态,并非常态化结果,根据待测言论传播时展现的确定结构来训练得到的模型很难有较高的泛化能力,在面对某些迷惑性较强的谣言时可能无法正确检测。

此外,基于情感计算的谣言检测近年来也受到了广泛关注。文献[23]提出了基于情感增强和心理语言学特征的在线社交媒体谣言检测方法,该方法提取帖子与其评论的情感和心理语言学特征,并获取二者的语义关系,据此进行谣言检测。文献[24]提出了一种基于迁移学习的假新闻和谣言检测的情绪感知多任务方法,该方法使用迁移学习为谣言检测数据集打上情绪类别标签,并据此获取文本真实性与其内在情绪之间的相关性,最后使用一个多任务框架,同时预测文本的情感和真实性,实现谣言检测。基于情感计算的方法的确有助于更准确地识别谣言,但通常需要对数据进行情感标注,成本较高,且某些谣言会使用一些夸张或虚假的情感,容易误导情感计算过程,导致检测难度增加。

因此,本研究综合考虑待测言论本身文本特征、其动态传播演化过程、传播结构的不确定性,提出 Spatio-Temporal Attention Fusion and Reconstruction Graph Convolution Network(STAFRCN)模型。其首先采用 RoBERTa^[8]模型预处理数据,通过卷积生成传播图表示,再使用时间渐进卷积方法充分提取言论传播时动态变化的信息;接着通过时空注意力融合方法将该言论在时间和空间上呈现的重要特征融合得到更有效的传播状态表示;然后进行第一次检测;最后使用

LSTM 方法为该言论的传播引入不确定性因素,并据此调整待测言论的传播结构,将更新后的传播结构与第一次检测结果结合进行第二次谣言检测。两次检测可有效降低误判率,提高检测准确率。

3 研究方法

3.1 问题定义

3.1.1 谣言检测问题

谣言检测指分析某待测言论本身内容及其相关评论等将该言论分类为谣言或事实的二分类任务,它可以帮助人们及时识别出谣言,采取截断措施,防止其给社会带来众多不利影响。

3.1.2 传播图序列及传播时间

本文将待测言论及其相关评论构成的图定义为内容传播图。内容传播图序列由数据集按照时间片分块获得,不同的图代表不同时间片的待测言论传播情况。每个时间片代表的时间长短可以不同,但每个时间片中节点数量相同,即每个图都由一个待测言论节点和若干该时间片的评论节点构成,其中待测节点存在于所有时刻的图中,整个动态图序列在时间维度反映待测言论自发生之时起其相关评论的发展演化过程。本研究中不同传播时间即指不同时间片。

由数据集中提取的内容传播图序列表示为 $\zeta = \{G_1, G_2, \dots, G_T\}$,其中 T 表示总的时间范围; $G_t = \{V_t, E_t, A_t\}$, G_t 为 t 时刻的传播图,由谣言与 t 时刻出现的相关评论构成; $V_t = \{v_1, v_2, \dots, v_N\}$, V_t 表示 t 时刻的节点集合。节点包含待检测言论节点和 t 时刻的相关评论节点, $X_t = [x_1^t, x_2^t, \dots, x_N^t]^T$, $X_t \in R^{N \times D}$ 为待测言论和 t 时刻所有评论的原始信息矩阵, $E_t \in R^{N \times N}$ 为 t 时刻传播图的边连接权重矩阵, $A_t \in R^{N \times N}$ 为 t 时刻传播图的邻接矩阵表示。

3.1.3 谣言检测任务

谣言检测任务构建一个分类器 $Y = f_\theta(X_{\text{text}})$,其中 X_{text} 是输入,表示待检测言论和与其相关的评论; Y 为检测结果(谣言或事实); $f_\theta(\cdot)$ 是分类器; θ 是待学习的相关参数。

3.2 模型构建

本研究提出了一个用于谣言检测的图卷积时空注意力融合与图重构方法(STAFRCN),模型包含时空注意力融合网络与图重构网络,用于对待测言论进行二次检测。

第一次检测使用时空注意力融合网络,其由两个模块构成,分别是时间渐进卷积模块和时空注意力融合模块,如图 1 所示。时间渐进卷积模块用于逐时刻提取整合待测言论的传播信息,建立依赖性;时空注意力融合模块从时间和空间两个维度对待测言论传播信息进行注意力提取并融合,融合结果用于第一次谣言检测。为了降低误判率,对所有待测言论进行第二次谣言检测。

如图 2 所示,第二次谣言检测使用图重构网络,其主要由谣言预测模块和图重构模块组成。谣言预测模块使用 LSTM 预测未来时刻待测言论的传播信息并在已有待测言论传播信息的基础上对传播总图进行重构,重构的总图结合第一次检测过程中的融合结果进行第二次谣言检测。

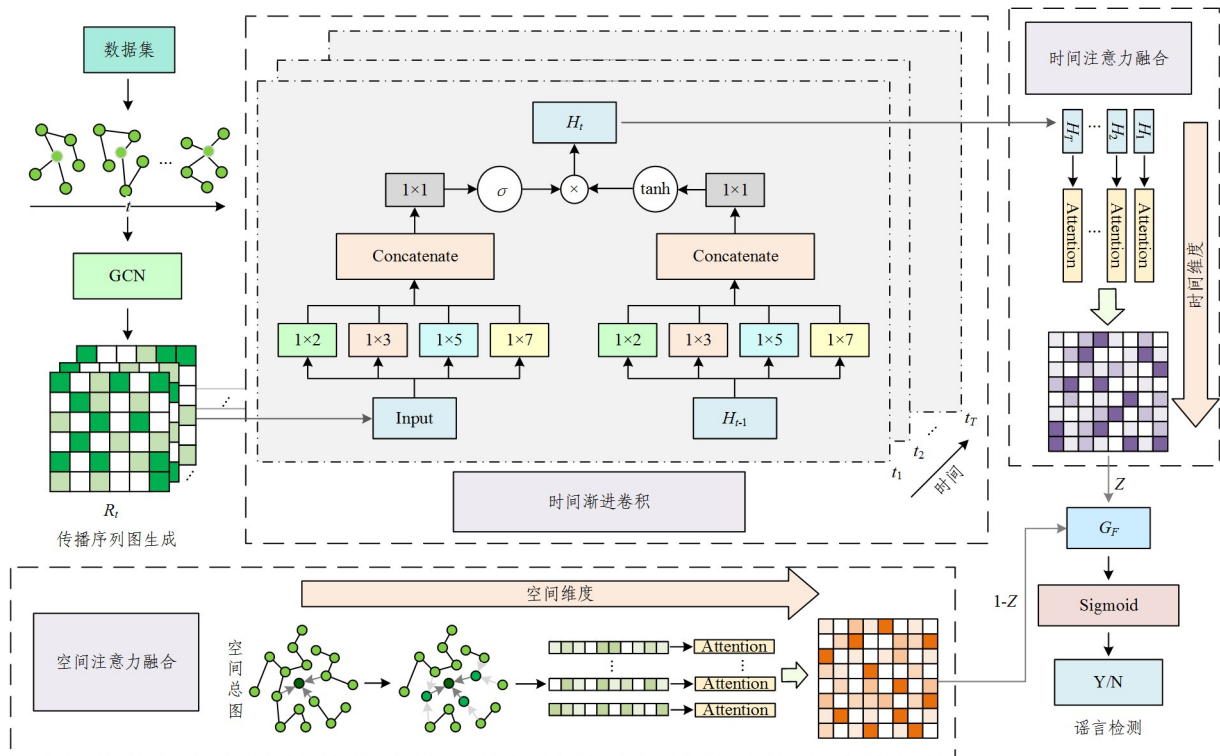


图 1 时空注意力融合框架

Fig. 1 Framework of spatio-temporal attention fusion

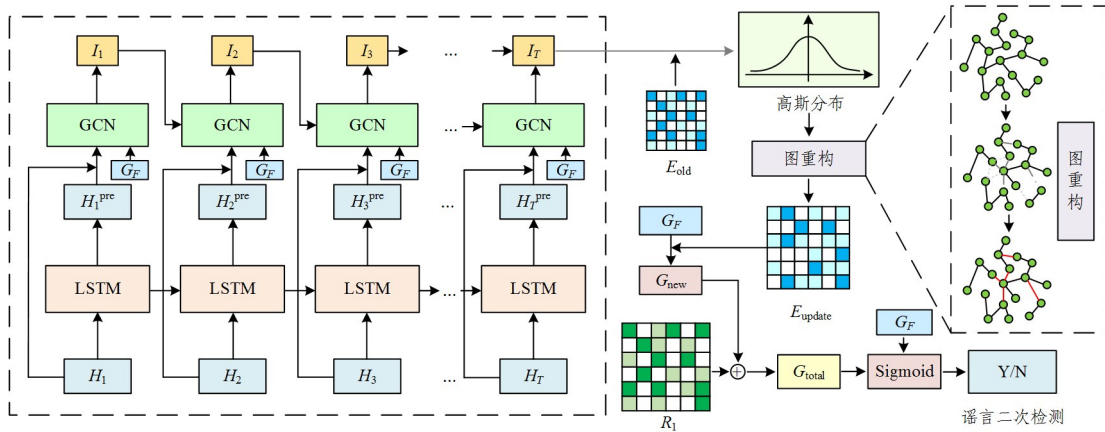


图 2 图重构框架

Fig. 2 Framework of graph reconstruction

3.2.1 传播序列图生成

首先预训练语言模型 RoBERTa^[8] 对内容传播图 G_t 中的待测言论 $TEXT$ 和其相关评论 $text'_i$ 进行词嵌入, 表示为 \mathbf{n} 和 \mathbf{m}'_i , 如式(1)所示。然后通过一个线性层将其映射到同一维度, 获取内容传播图中每个节点的表示 x'_i :

$$\begin{aligned} \mathbf{n} &= \text{RoBERTa}(TEXT) \\ \mathbf{m}'_i &= \text{RoBERTa}(text'_i) \\ \mathbf{x}'_i &= \mathbf{W}_s[\mathbf{n}; \mathbf{m}'_i] \quad (i \in [1, N]) \end{aligned} \quad (1)$$

其中, \mathbf{W}_s 为参数, 用于控制节点表示的维度相同; N 表示待测谣言和相关评论数量。

然后使用余弦相似度计算内容传播图各节点间的相关程度, 作为当下内容传播图中各节点间的边权重, 得到

边连接权重矩阵 \mathbf{E}_t , 从而确定内容传播图 G_t 。 \mathbf{E}_t 矩阵各元素的计算式如下:

$$e_{i,j} = \frac{\sum_{k=1}^n (x'_{i,k} \times x'_{j,k})}{\sqrt{\sum_{k=1}^n (x'_{i,k})^2} \times \sqrt{\sum_{k=1}^n (x'_{j,k})^2}} \quad (i, j \in [1, N]) \quad (2)$$

将权重矩阵 \mathbf{E}_t 和邻接矩阵 \mathbf{A}_t 线性组合, 对权重进行初步筛选、更新得到 \mathbf{U}_t , 去除冗余权重。接着将原始待测言论信息 \mathbf{X}_t 和更新后的权重矩阵 \mathbf{U}_t 进行卷积, 对 \mathbf{X}_t 中的特征进行重要程度区分, 处理后待测言论信息矩阵表示为 \mathbf{R}_t 。 \mathbf{U}_t 和 \mathbf{R}_t 的计算式分别如式(3)和式(4)所示:

$$\mathbf{U}_t = \mathbf{W}_e \mathbf{E}_t + \mathbf{W}_a \mathbf{A}_t + \mathbf{b}_{ea} \quad (3)$$

$$\mathbf{R}_t = \text{Sigmoid}(\mathbf{D}^{-\frac{1}{2}} \mathbf{X}_t \mathbf{U}_t \mathbf{D}^{-\frac{1}{2}} \mathbf{W}_{r1} + \mathbf{D}^{-\frac{1}{2}} (\mathbf{X}_t)^T \mathbf{U}_t \mathbf{D}^{-\frac{1}{2}})$$

$$\mathbf{W}_{r2} + \mathbf{b}_r) \quad (4)$$

其中, \mathbf{W} 为可学习矩阵, Sigmoid 为激活函数, \mathbf{b} 为偏置。

3.2.2 时间渐进卷积模块

不同时刻待测言论及其相关评论处于孤立状态, 各自提取特征则关联性不强。言论传播过程中的评论相互之间的依赖性不强, 因此需要一种方法将它们整合。

本研究提出了时间渐进卷积模块 TPC, 如图 3 所示, 它使用 4 个并行的卷积单元处理输入, 捕获不同传播范围的信息, 然后将各卷积单元的结果拼接。TPC 模块由两个相同的部分组成, 输入为两部分, 最后的输出为两个处理结果相乘。

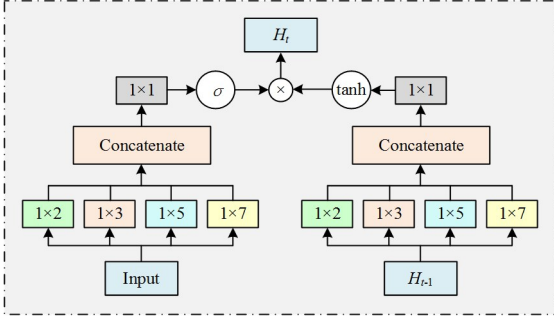


图 3 时间渐进卷积模块

Fig. 3 Temporal progressive convolution module

TPC 在时间维度上提取待测言论传播信息, 建立当前时刻待测言论传播信息和之前所有时刻待测言论传播信息的依赖关系, 得到各时刻的传播状态提取图表示 \mathbf{H}_t (隐藏状态)。 \mathbf{H}_t 集中了谣言在时间维度上的动态传播状态, 计算式如下:

$$\mathbf{H}_t = \text{Sigmoid}(\text{TPC}(\mathbf{R}_t)) \odot \tanh(\text{TPC}(\mathbf{H}_{t-1})) \quad (5)$$

3.2.3 时空注意力融合模块

谣言在时间维度上的传播状态以及在空间上的传播结构均能反映谣言的真假, 两部分各自独立用于谣言检测时表现欠佳, 而将时空传播信息提取整合能提高检测准确率。

如图 1 中空间注意力融合模块所示, 将所有时刻的传播图合并得到空间总图 $\mathbf{V} = \mathbf{V}_1 \cup \dots \cup \mathbf{V}_T$, $\mathbf{X} = \mathbf{X}_1 \cup \dots \cup \mathbf{X}_T$ 。空间总图中不同时刻节点间均建立边, 计算边权重 (如式 (2) 所示)。由于各节点间关联性不强、特征分散, 因此对空间总图进行融合邻居节点信息以更新图表示。

首先, 聚合每个节点 n_i 和一跳邻居节点 $n_i^{neighbor1}$, 更新后每个节点表示为 $\mathbf{x}_i^{\text{fir}}$, 计算式如下:

$$\mathbf{x}_i^{\text{fir}} = \text{Sigmoid}(\mathbf{W}_{f,ij} \mathbf{x}_i + \sum_{j=1}^{neighbor1} \mathbf{W}_{f,ij} e_{i,j} \mathbf{x}_j) \quad (6)$$

其中, $neighbor1$ 为一跳邻居节点集合, $\mathbf{W}_{f,ij}$ 为更新参数, Sigmoid 为激活函数, $e_{i,j}$ 为节点间连接权重, \mathbf{x}_i 和 \mathbf{x}_j 为节点表示。由于一跳邻居节点只能提供直连节点的信息, 无法涵盖更远距离节点的信息, 而谣言传播通常并不局限于节点的直接邻居, 因此在引入一跳邻居节点信息的基础上继续引入二跳节点信息, 可以在不过多地增加计算复杂度的情况下帮助节点更好地理解整个传播图的结构和特征, 减少局部噪声对检测的影响, 提高节点信息更新的鲁棒性, 也能得到更准确的空间总图特征关系。具体操作为: 将每个节点的一跳后续节点 $n_i^{neighbor2}$ 信息融合到当前节点进行二次更新, 更新后的节点表示为 \mathbf{x}_i' 。其计算式如下:

$$\mathbf{x}_i' = \text{Sigmoid}(\mathbf{W}_{s,ij} \mathbf{x}_i^{\text{fir}} + \sum_{j=1}^{neighbor2} \mathbf{W}_{s,ij} e_{i,j} \mathbf{x}_j^{\text{fir}}) \quad (7)$$

其中, $neighbor2$ 为一跳后续节点, $\mathbf{W}_{s,ij}$ 为更新参数, Sigmoid 为激活函数, $e_{i,j}$ 为节点间的连接权重, $\mathbf{x}_i^{\text{fir}}$ 和 $\mathbf{x}_j^{\text{fir}}$ 为聚合了一跳邻居节点的表示。

合并后的总图反映了待测言论在 T 时间内的整体传播结构, 其包含一个待测谣言节点和 T 时间内所有相关评论节点。

然后用注意力处理更新后的节点, 提取空间总图中融合了邻居节点的重要特征信息, 去除冗余信息, 得到传播图表示 \mathbf{G}_S :

$$\mathbf{Q}_s = \mathbf{x}_i' \mathbf{W}_{Q_s}, \mathbf{K}_s = \mathbf{x}_i' \mathbf{W}_{K_s}, \mathbf{V}_s = \mathbf{x}_i' \mathbf{W}_{V_s} \quad (8)$$

$$\text{Attention}(\mathbf{Q}_s, \mathbf{K}_s, \mathbf{V}_s) = \text{softmax}\left(\frac{\mathbf{Q}_s \mathbf{K}_s^T}{\sqrt{d}}\right) \mathbf{V}_s \quad (9)$$

$$\mathbf{G}_S^{(i)} = \text{Attention}(\mathbf{Q}_s^{(i)}, \mathbf{K}_s^{(i)}, \mathbf{V}_s^{(i)}) (i \in [1, N]) \quad (10)$$

$$\mathbf{G}_S = [\mathbf{G}_S^{(1)}, \mathbf{G}_S^{(2)}, \dots, \mathbf{G}_S^{(N)}] \quad (11)$$

其中, \mathbf{W} 为可学习参数矩阵, \mathbf{G}_S 由各节点的注意力计算结果拼接而成。

如图 1 中时间注意力融合模块所示, 在时间维度上, 对之前时间渐进卷积得到的结果也进行注意力机制的处理, 提取待测言论传播时较为突出的特征, 得到传播图的时间表示 \mathbf{G}_T 。对不同时刻的图进行注意力处理, 因为不同时刻的待测言论传播状态对待检测谣言真假的判断影响不同, 公式如下:

$$\mathbf{Q}_T = \mathbf{H}_t \mathbf{W}_{Q_T}, \mathbf{K}_T = \mathbf{H}_t \mathbf{W}_{K_T}, \mathbf{V}_T = \mathbf{H}_t \mathbf{W}_{V_T} \quad (12)$$

$$\text{Attention}(\mathbf{Q}_T, \mathbf{K}_T, \mathbf{V}_T) = \text{softmax}\left(\frac{\mathbf{Q}_T \mathbf{K}_T^T}{\sqrt{d}}\right) \mathbf{V}_T \quad (13)$$

$$\mathbf{G}_T^{(t)} = \text{Attention}(\mathbf{Q}_T^{(t)}, \mathbf{K}_T^{(t)}, \mathbf{V}_T^{(t)}) (t \in [0, T]) \quad (14)$$

$$\mathbf{G}_T' = [\mathbf{G}_T^{(1)}, \mathbf{G}_T^{(2)}, \dots, \mathbf{G}_T^{(T)}], \mathbf{G}_T = \mathbf{W}_1 \mathbf{G}_T' + \mathbf{b}_1 \quad (15)$$

其中, \mathbf{W}_1 用于调整 \mathbf{G}_T' 的维度与 \mathbf{G}_S 保持一致。

3.2.4 第一次谣言检测

将时间和空间两个维度提取到的信息进行门控时空融合, 对时空注意力处理得到的特征信息进行有选择性的保留, 得到 \mathbf{G}_F 用于第一次谣言检测, 使用 Sigmoid 函数进行谣言真伪的检测。门控融合计算式如式 (16) 所示, 融合权重计算式如式 (17) 所示:

$$\mathbf{G}_F = \mathbf{z} \odot \mathbf{G}_S + (1 - \mathbf{z}) \odot \mathbf{G}_T \quad (16)$$

$$\mathbf{z} = \text{Sigmoid}(\mathbf{G}_S \mathbf{W}_{z,1} + \mathbf{G}_T \mathbf{W}_{z,2} + \mathbf{b}_z) \quad (17)$$

其中, \mathbf{z} 为门控融合的权重参数, 由时间、空间注意力结果计算得到; \mathbf{W} 为可学习矩阵; \mathbf{b}_z 为偏置; Sigmoid 为激活函数。

检测公式为:

$$\hat{\mathbf{y}} = \text{Sigmoid}(\mathbf{G}_F \mathbf{W}_y + \mathbf{b}_y) \quad (18)$$

其中, \mathbf{W}_y 为可学习参数, \mathbf{b}_y 为偏置。

交叉熵损失函数的计算式为:

$$L_1 = - \sum_{i \in N} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (19)$$

其中, y_i 为实际值, \hat{y}_i 为预测值。

3.2.5 谣言预测模块

由于待测言论传播状态动态演变, 具有一定的不确定性, 因此内容传播动态图只能代表某时刻待测言论发展的某种状态, 具有偶然性。将其用于谣言检测会带来较大的误差, 因此第一次谣言检测对待测言论是否属于谣言的判断可能有误。

我们在在第一次谣言检测完成之后进行谣言二次检测以防止误判,首先使用谣言预测模块,即用 LSTM 预测未来时刻待测言论发展状况,并将预测结果与该时刻言论的实际传播状态结合得到 $\mathbf{H}_i^{\text{update}}$,增加不确定性因素的同时也考虑待测言论当下的现实状态,如图 4 所示。LSTM 预测的加入能减小待测言论发展不确定性对检测带来的影响,预测公式如式(20)所示,结合计算公式如式(21)所示:

$$\mathbf{H}_i^{\text{pre}} = \text{LSTM}(\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_{i-1}) \quad (20)$$

$$\mathbf{H}_i^{\text{update}} = \mathbf{W}_{h1} \mathbf{H}_i + \mathbf{W}_{h2} \mathbf{H}_i^{\text{pre}} + \mathbf{b}_h \quad (21)$$

其中, \mathbf{H}_i 为第一次检测过程中用 TPC 处理的结果, \mathbf{W} 为可学习矩阵, \mathbf{b}_h 为偏置, $\mathbf{H}_i^{\text{update}}$ 为 LSTM 预测结果与当前传播状态线性结合的结果。

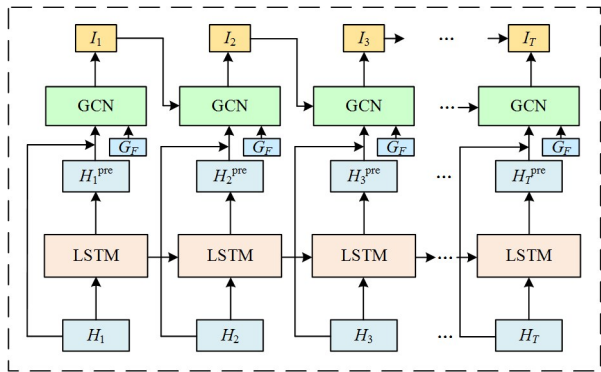


图 4 谣言预测模块

Fig. 4 Rumor prediction module

然后,在时间维度上继续对更新后的谣言表示(隐藏状态)进行逐时刻卷积,目的是逐时刻收集预测的言论传播状态的特征信息,整合第一次谣言检测中时空注意力融合的结果 \mathbf{G}_F 并加入预测成分,减小谣言传播的偶然性带来的误差。整合过程由线性组合卷积方法实现,最终得到 \mathbf{I}_T , 计算公式如下:

$$\mathbf{I}_T = \sigma(\hat{\mathbf{D}}_T^{-\frac{1}{2}} \mathbf{H}_T^{\text{update}} \hat{\mathbf{D}}_T^{-\frac{1}{2}} \mathbf{G}_F \mathbf{W}_{t1} + \hat{\mathbf{D}}_T^{-\frac{1}{2}} (\mathbf{H}_T^{\text{update}})^T \hat{\mathbf{D}}_T^{-\frac{1}{2}} \mathbf{I}_{T-1} \mathbf{W}_{t2} + \mathbf{b}_t) \quad (22)$$

其中, $\hat{\mathbf{D}}_t$ 为度矩阵, \mathbf{W} 为可学习矩阵, \mathbf{b}_t 为偏置项, σ 表示激活函数 ReLU。

3.2.6 传播图重构

如图 4 所示, \mathbf{I}_T 整合了每个时刻的待测言论信息、预测的相关评论状态以及时空注意力融合结果,基于此对原内容传播空间总图进行重构,即调整原传播空间总图的边连接情况,能够得到比较合理的传播图结构:

首先根据 \mathbf{I}_T 更新原内容传播总图的边连接权重矩阵。 \mathbf{E}_{old} 为原传播总图的边连接权重矩阵, \mathbf{E}' 为更新后的权重矩阵,计算公式为:

$$\mathbf{E}' = \mathbf{W}_{E1} \mathbf{I}_T + \mathbf{W}_{E2} \mathbf{E}_{\text{old}} \quad (23)$$

其中, \mathbf{W} 为可学习参数。

使用高斯分布^[25]对传播总图进行重构,在高斯分布空间中可以学习动态网络的节点表示来对动态信息进行建模。对于谣言传播节点图,找到对谣言具有较大影响力的节点集合对谣言检测具有关键作用。如果将任意两个节点之间的相关性看作是一个概率分布,那么使用高斯分布进行重构可以

更好地描述这种概率分布,因为高斯分布具有可控的方差和均值,可以描述概率密度的不同区域,从而更好地捕捉节点之间的影响力差异,且高斯分布具有平滑的特性,能够使重构过程更加鲁棒。

迁移到本研究中,待测言论传播结构的不确定性可以由预测得到的传播图中节点间关联程度的大小体现,而关联程度可由边连接权重计算得到,因此可以通过传播图的边连接权重的高斯分布加一个可学习的阈值来控制传播图中边的存在与否,从而实现传播图重构。具体实现过程如下:

对 \mathbf{E}' 中的边权重求均值 μ 以及协方差矩阵 Σ 得到权重高斯分布 $e_{i,j} \sim N(\mu, \Sigma)$, 由式(24)计算得到,权重矩阵中的元素更新公式如式(25)所示:

$$p(e_{i,j}) = \frac{1}{\sqrt{(2\pi)^N \Delta(\Sigma)}} e^{-\frac{1}{2}(e_{i,j} - \mu)^T \Sigma^{-1} (e_{i,j} - \mu)} \quad (24)$$

$$e_{i,j} = \begin{cases} p(e_{i,j}) \times e_{i,j}, & p(e_{i,j}) \geq \delta \\ 0, & p(e_{i,j}) < \delta \end{cases} \quad (25)$$

其中, δ 为阈值,各节点间关联程度大于阈值的边保留,小于阈值的边则舍去。

重构后的边连接权重矩阵为 $\mathbf{E}_{\text{update}}$, 重构后的传播图为 \mathbf{G}_{new} 。 \mathbf{G}_{new} 由 $\mathbf{E}_{\text{update}}$ 和 \mathbf{G}_F 逐元素相乘得到, 计算式为:

$$\mathbf{G}_{\text{new}} = \mathbf{E}_{\text{update}} \odot \mathbf{G}_F \quad (26)$$

将 \mathbf{G}_{new} 与原始传播总图的谣言信息矩阵 \mathbf{R} 拼接得到总图 $\mathbf{G}_{\text{total}}$, \mathbf{R} 表示待检测谣言最初的传播结构, \mathbf{G}_{new} 中包含了时空注意力融合结果、预测结果,以及图重构结果中的重要信息。两者拼接能有效提高谣言检测准确率, 拼接公式如下:

$$\mathbf{G}_{\text{total}} = \text{concat}(\mathbf{R}, \mathbf{G}_{\text{new}}) \mathbf{W}_{\text{total}} \quad (27)$$

3.2.7 第二次谣言检测

复用第一次检测方法得到预测值 $\hat{\mathbf{y}}_i$, 损失函数为 L_2 。总的损失 L 由 L_1 和 L_2 加权得到, 公式如下:

$$\hat{\mathbf{y}} = \text{Sigmoid}(\mathbf{G}_{\text{total}} \mathbf{W}_{y1} + \mathbf{G}_F \mathbf{W}_{y2} + \mathbf{b}_{yy}) \quad (28)$$

$$L = \alpha L_1 + \beta L_2 \quad (29)$$

其中, \mathbf{W}_{y1} 和 \mathbf{W}_{y2} 为可学习参数, \mathbf{b}_{yy} 为偏置, α 和 β 为权重系数。

4 实验结果及分析

4.1 实验环境及设置

实验中设置每个批次包含样本的数量 $batch_size=64$, 正则化参数 $reg=0.01$, 迭代次数 $epochs=50$, $dropout$ 概率设置为 0.2, 使用 Adam 优化算法训练模型参数。本实验使用 Acc. 准确率、Prec. 精确率、Rec. 召回率、F1 值这 4 个评价指标来衡量模型的检测效果。

4.2 实验数据集

为了验证本文提出模型的检测准确率,我们在 Twitter15^[26], Twitter16^[26], Weibo^[27] 这 3 个数据集上对比了 STAFRCN 与现有模型的性能差异。

Twitter15 和 Twitter16 数据集来自社交媒体平台 Twitter, 分别含有 1490 和 818 个原始帖子及其转发、评论的传播树结构, 每棵传播树由一条言论和其相关评论组成, 且所有

评论都带有距离该言论的时间。Twitter15 和 Twitter16 数据集共有 4 个标签:非谣言(NR)、验证为假的谣言(FR)、验证为真的谣言(TR)和未经验证的谣言(UR)。本实验仅使用非谣言、验证为假的谣言、验证为真的谣言,且将它们分为两类:事实和谣言。Weibo 数据集来自社交平台微博,共有 4 664 个微博原文及其转发、评论内容的传播树结构,其数据特点与 Twitter15 和 Twitter16 相似。Weibo 数据集有两个标签:谣言和非谣言。具体信息如表 1 所列。

表 1 数据集介绍

Table 1 Dataset introduction

统计量	Twitter15	Twitter16	Weibo
原帖子数量	1490	818	4664
文本数量	42914	20295	2011057
谣言数量	372	205	2351
假谣言数量	370	205	2313
未证实谣言数量	374	203	0
非谣言数量	374	205	0

在对数据进行预处理时,删去了转发帖子、用户 ID 等信息,然后以待测言论为中心,将数据集切分出不同时间片。具体操作为:根据各评论文本距离待测言论的时间顺序,将其分为多个时间片,不同时间片表示的时间长度可不相同,但每个时间片包含的评论数量相同,本研究中设置为 15(可根据实验设备性能适当增大或减小),在每个时间片构建一个传播图,且每个传播图中均包含待测言论节点。

4.3 消融实验

本研究提出的 STAFRCN 模型主要由时间渐进卷积模块、时空注意力融合模块和图重构模块构成。为了验证本文方法的有效性,分别在 3 个数据集上进行消融实验。消融实验中,“TPC”表示时间渐进卷积模块,“STAF”表示时空注意力融合模块,“GR”表示图重构模块,“None”表示只使用图卷积网络的基础模型。实验结果如表 2—表 4 所列,标粗的

数据为最佳结果,可视化结果如图 5 所示。

表 2 Twitter15 数据集上的消融实验结果

Table 2 Ablation experimental results on Twitter15

模块	Acc.	Prec.	Rec.	F_1
None	0.544	0.563	0.562	0.562
TPC	0.602	0.602	0.610	0.606
STAF	0.651	0.652	0.661	0.656
GR	0.610	0.612	0.610	0.609
TPC+STAF	0.852	0.863	0.857	0.859
TPC+GR	0.762	0.765	0.768	0.766
STAF+GR	0.821	0.825	0.831	0.827
TPC+STAF+GR	0.922	0.919	0.915	0.917

表 3 Twitter16 数据集上的消融实验结果

Table 3 Ablation experimental results on Twitter16

模块	Acc.	Prec.	Rec.	F_1
None	0.487	0.503	0.501	0.502
TPC	0.598	0.601	0.598	0.599
STAF	0.623	0.632	0.627	0.629
GR	0.613	0.612	0.612	0.612
TPC+STAF	0.818	0.820	0.817	0.818
TPC+GR	0.743	0.741	0.747	0.744
STAF+GR	0.804	0.808	0.807	0.807
TPC+STAF+GR	0.918	0.916	0.917	0.916

表 4 Weibo 数据集上的消融实验结果

Table 4 Ablation experimental results on Weibo

模块	Acc.	Prec.	Rec.	F_1
None	0.602	0.610	0.607	0.608
TPC	0.631	0.632	0.627	0.629
STAF	0.662	0.663	0.661	0.662
GR	0.643	0.642	0.642	0.642
TPC+STAF	0.875	0.875	0.876	0.875
TPC+GR	0.782	0.780	0.785	0.782
STAF+GR	0.831	0.834	0.835	0.834
TPC+STAF+GR	0.965	0.957	0.952	0.954

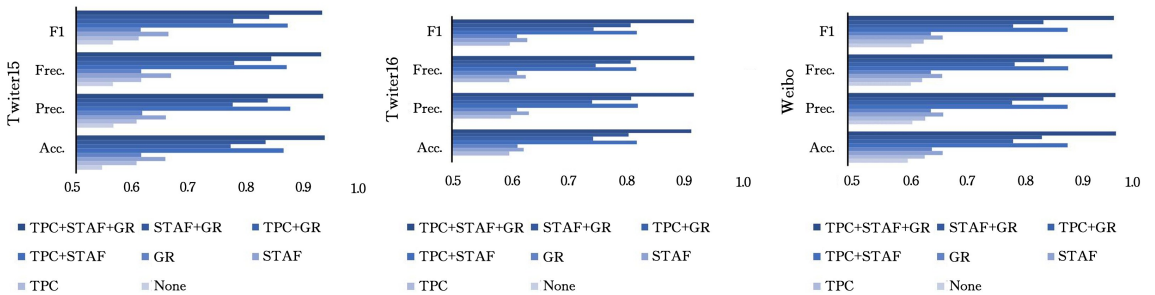


图 5 消融实验结果

Fig. 5 Ablation experimental results

根据表 2—表 4 的实验结果可知,只使用图卷积网络的基础模型用于谣言检测的准确率较低,分别单独使用本研究提出的 3 个模块时均能提高检测准确率,且单独使用时空注意力融合模块对准确率的提升效果最明显,在 Twitter15 和 Twitter16 上的准确率分别达到了 65.1%和 62.3%,在 Weibo 数据集上准确率为 66.2%,表明 STAF 在时间维度能提取待测言论传播过程中的动态演化信息,在空间维度能提取待测言论的传播结构信息,对于谣言检测有很大帮助。对比只使用本研究提出的两个模块的消融实验可以发现,TPC 结合

STAF 能显著提高检测准确度,在 3 个数据集上的准确率分别为 85.2%,81.8%,87.5%,表明了 TPC 在提取待测言论的动态演化信息及构建演化信息的相互依赖关系方面的有效性。对比使用 TPC,STAF,TPC+STAF 和添加了 GR 模块的实验结果可以发现,GR 的添加提升了模型的检测准确率,表明增加待测言论传播结构的不确定性、调整传播图的结构能有效提高模型的泛化能力,最终模型 STAFRCN 在 3 个数据集上的准确率分别为 92.2%,91.8%,96.5%,证明了本文方法的有效性。

4.4 对比实验

4.4.1 不同模型对比

本文对比了 STAFRCN 和现有模型之间的性能差异,选取的基线模型有 SVM-TS^[28], GSMRD^[29], C-GRU^[30], MGCN^[31], EPRD^[32], MGAT-ESM^[33], KAGN^[34]。在 3 个数据集上的实验结果如表 5—表 7 所列,其中粗体数据为最优数据,可视化结果如图 6 所示。

表 5 Twitter15 数据集上的实验结果

Table 5 Experimental results on Twitter15

模型	Acc.	Prec.	Rec.	F ₁
SVM-TS	0.784	0.788	0.785	0.786
GSMRD	0.891	0.887	0.883	0.885
C-GRU	0.876	0.865	0.867	0.866
MGCN	0.896	0.892	0.894	0.893
EPRD	0.897	0.886	0.890	0.888
MGAT-ESM	0.908	0.904	0.907	0.905
KAGN	0.892	0.887	0.885	0.886
STAFRCN(Ours)	0.922	0.919	0.915	0.917

表 6 Twitter16 数据集上的实验结果

Table 6 Experimental results on Twitter16

模型	Acc.	Prec.	Rec.	F ₁
SVM-TS	0.773	0.779	0.780	0.779
GSMRD	0.901	0.895	0.894	0.895
C-GRU	0.854	0.843	0.846	0.844
MGCN	0.896	0.883	0.881	0.882
EPRD	0.894	0.896	0.897	0.896
MGAT-ESM	0.901	0.898	0.897	0.897
KAGN	0.903	0.898	0.899	0.898
STAFRCN(Ours)	0.918	0.916	0.917	0.916

表 7 Weibo 数据集上的实验结果

Table 7 Experimental results on Weibo

模型	Acc.	Prec.	Rec.	F ₁
SVM-TS	0.796	0.800	0.798	0.799
GSMRD	0.879	0.868	0.866	0.867
C-GRU	0.887	0.875	0.879	0.877
MGCN	0.943	0.932	0.936	0.934
EPRD	0.901	0.903	0.902	0.902
MGAT-ESM	0.923	0.921	0.920	0.920
KAGN	0.951	0.947	0.948	0.947
STAFRCN(Ours)	0.965	0.964	0.963	0.963

根据表 5—表 7 的对比实验结果可以看出,在 3 个数据集上,STAFRCN 的谣言检测准确率分别为 92.2%,91.8%,96.5%,且均为最优,表明 STAFRCN 在谣言检测领域与其他基线模型相比具有很大的优势。

此外还可以发现,使用特征工程的传统方法 SVM-TS 在 3 个数据集上的表现都不理想,说明传统的机器学习方法难以获得待测言论的深层文本特征表示,因而在谣言检测任务中表现欠佳。与 STAFRCN 相比,GSMRD, MGAT-ESM, DAGCN 仅考虑待测言论的语义信息和传播结构而没有考虑谣言在传播过程中的不确定性,因此检测准确率略低。与 STAFRCN 相比,C-GRU, MGCN, EPRD 对待测言论的传播结构特征学习不足,准确率差距较大。而 STAFRCN 通过时间渐进卷积、时空注意力融合、谣言预测、图重构这几个步骤充分考虑了待测言论的语义特征、传播结构特征、传播过程中的动态演化信息,以及传播的不确定性,使得模型在谣言检测任务中表现优异。

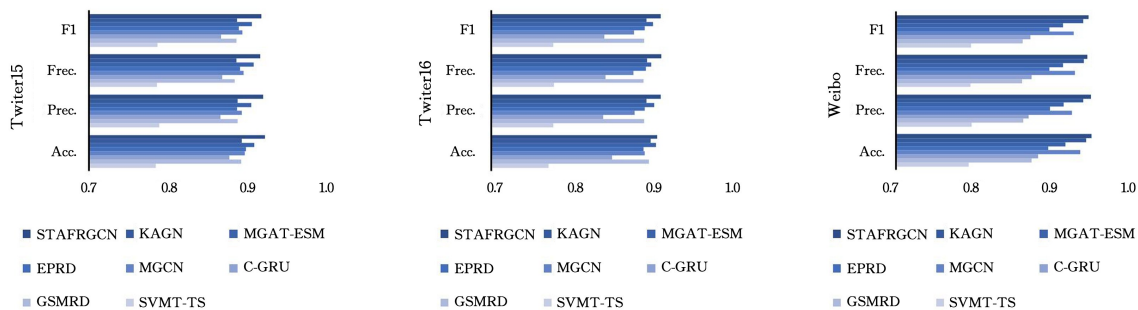


图 6 模型性能对比结果

Fig. 6 Results of model performance comparison

4.4.2 模型参数对比

本文在 STAFRCN 模型中提出了时间渐进卷积模块 (TPC), TPC 中使用了 4 个并行的卷积单元来处理输入数据,而卷积单元的数量可能会影响模型的最终性能。因此,我们针对卷积单元的数量进行了对比实验,实验选用 6 种常见大小的卷积核。图 7 给出了不同大小的卷积核在不同卷积单元数量下模型的性能变化情况。从图中的实验结果可以看出,卷积单元数量对检测效果有显著影响。卷积单元数量较少时,模型难以对输入的传播图表示进行充分的理解,检测效果不佳;随着卷积单元数量的增加,检测效果有了明显的改善,但当卷积单元数量过多时,模型容易陷入过拟合,导致泛化能力降低、检测效果下降。此外,卷积核大小对模型性能

影响也很大,分析可知,卷积核只能提取单个位置的信息,检测效果不佳,卷积核增加了参数数量和计算成本,且不能捕获足够的局部和全局信息,因此对模型的性能提升帮助不大。由实验结果可得,卷积单元数量为 4 时可以达到较好的检测效果。

针对谣言检测任务,本文使用大小为 $1 \times 2, 1 \times 3, 1 \times 5, 1 \times 7$ 的 4 个卷积核来提升准确性和效率^[35],其中 1×2 卷积核能够同时捕获相邻两个位置的信息,提取上下文特征;然后依次增大模型理解的语义信息的范围($1 \times 3, 1 \times 5, 1 \times 7$),既能提取相邻节点的信息,又能捕捉长距离节点的特征,从不同尺度进行卷积,充分理解输入传播图数据的整体结构和内在联系。

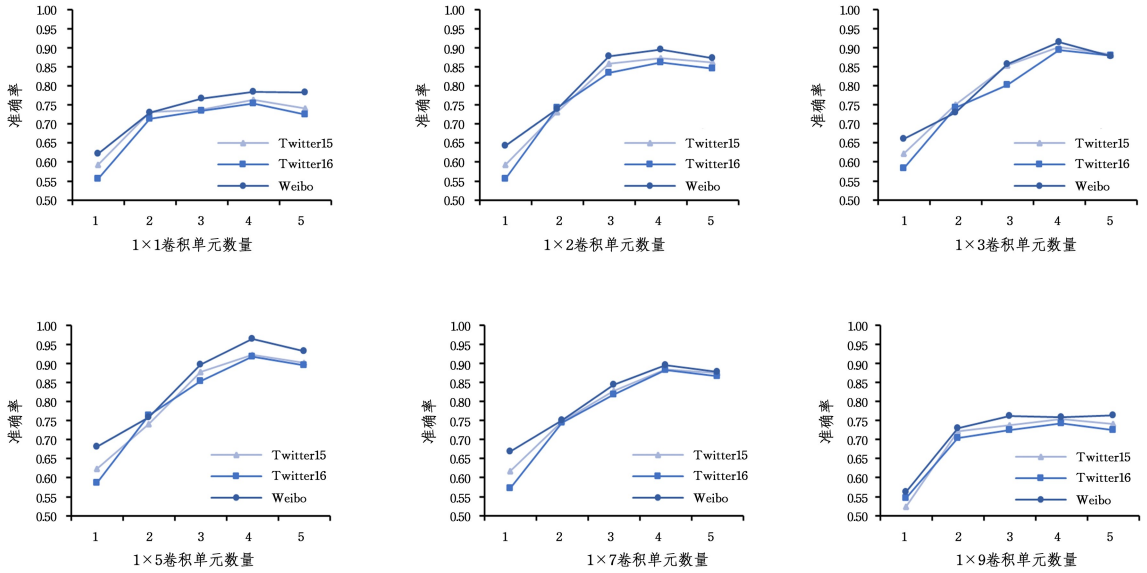


图7 卷积单元数量对检测效果的影响

Fig. 7 Effect of the number of convolutional units on detection effect

4.4.3 数据量对比

此外,本文提出的模型在3个数据集上的检测准确率有一定的差距,这可能是由数据量不同导致的,而数据量会直接影响到模型的泛化性能和预测效果。为了讨论不同数据量对模型性能的影响,我们从3个数据集分别随机抽取1000条、600条、4000条帖子作为3个新的数据集,对3个数据集按相同的比例进行实验,对比了不同的数据量对模型性能的影响,实验结果如表8—表10所列,可视化结果如图8所示。

表8 Twitter15数据量对比实验结果

Table 8 Experimental results of data volume comparison on

Twitter15				
数据量	Acc.	Prec.	Rec.	F ₁
250	0.688	0.687	0.691	0.689
500	0.829	0.812	0.817	0.814
750	0.898	0.882	0.887	0.884
1000	0.909	0.901	0.899	0.900

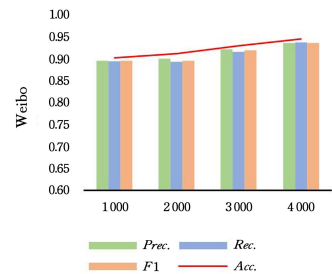
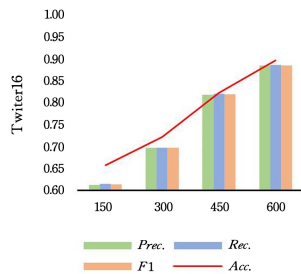
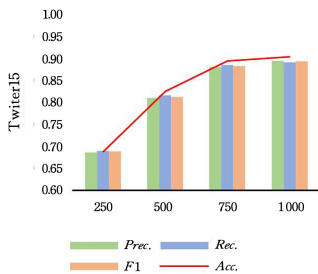


图8 数据量对准确率的影响

Fig. 8 Effect of data volume on accuracy

由实验结果可知,训练模型时数据量的多少在一定程度上会影响模型的检测准确率,因此,要获得较高的检测准确率,需要有足够的实验数据做支撑。通过比较不同数据量下模型的检测准确率变化,可以得出训练模型需要的合适的数据规模,防止数据量过少,模型出现过拟合的问题,也能预防数据量过大导致浪费计算资源的问题。由图8可以得出本研究提出的模型在帖子数量为4000左右时已能达到准确率

表9 Twitter16数据量对比实验结果

Table 9 Experimental results of data volume comparison on

Twitter16				
数据量	Acc.	Prec.	Rec.	F ₁
150	0.659	0.613	0.615	0.614
300	0.724	0.698	0.698	0.698
450	0.825	0.819	0.821	0.820
600	0.899	0.886	0.887	0.886

表10 Weibo数据量对比实验结果

Table 10 Experimental results of data volume comparison

on Weibo				
数据量	Acc.	Prec.	Rec.	F ₁
1000	0.905	0.897	0.895	0.896
2000	0.913	0.901	0.894	0.897
3000	0.933	0.923	0.917	0.920
4000	0.951	0.947	0.948	0.947

较高且模型性能较稳定的效果,因此可在此基础上适当改变训练数据量,研究模型的最优性能表现。即便训练数据量的大小会影响谣言检测准确率,本研究提出的模型与现有模型相比仍有较大优势,表明了本文模型STAFRCN在谣言检测任务上的优越性。

结束语 本文提出了一种新的谣言检测模型STAFRCN,该模型成功地提取、整合了待测言论的动态

传播信息,并从时间和空间两个维度使用注意力机制抓取待测言论传播过程中的重要特征信息,还考虑了现实中谣言传播的不确定性,使用图重构模块对待测言论的传播结构进行调整。最后,在3个数据集上的实验结果表明,本文模型的时间渐进卷积模块、时空注意力融合模块、图重构模块均能在一定程度上提高谣言检测的准确率,且本文模型的性能优于现有谣言检测模型,在谣言检测任务中具有一定的优势。实验结果也表明用于训练模型的数据集的大小也会对模型的性能产生影响,因此收集大量具有传播结构、相互之间特征有差异的数据集也具有重要意义。

本文提出的模型 STAFRCN 综合考虑待测言论的动态传播信息以及现实中谣言传播的不确定性来进行谣言检测,但缺少对转发帖子、评论用户等的分析,而转发帖子与评论用户在谣言检测中也有着重要作用。因此,在未来的研究中,我们计划针对上述不足,将转发帖子的文本特征、评论用户的特征及其相互之间的隐式关系等添加到模型的学习过程中,以给出更加有效的谣言检测方法。此外,现有数据集规模不足、特征不够全面等问题也增加了谣言检测的难度,因此关注并收集足够规模的高质量谣言检测数据集也有着重要意义。

参考文献

- [1] CHEN Y F, LI Z Y, LIANG X, et al. Review on Rumor Detection of Online Social Networks[J]. Chinese Journal of Computers, 2018, 41(7): 30.
- [2] ZHANG S Q, DU S D, ZHANG X B, et al. Social Rumor Detection Method Based on Multimodal Fusion [J]. Computer Science, 2021, 48(5): 117-123.
- [3] LV Y, SUN X, WEN Y, et al. Rumor detection based on time graph attention network[C]//2022 4th International Conference on Advances in Computer Technology, Information Science and Communications(CTISC). IEEE, 2022: 1-5.
- [4] 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.
- [5] LIU N, ZHANG F L, YIN J Q, et al. Rumor Detection Model on Social Media Based on Contrastive Learning with Edge-inference Augmentation [J]. Computer Science, 2023, 50(11): 49-54.
- [6] HAN X M, JIA C Y, LI X Y, et al. Dual-attention Network Model on Propagation Tree Structures for Rumor Detection [J]. Computer Science, 2023, 50(4): 22-31.
- [7] XU J M, SUN P, WU S F. Microblog Rumor Detection Method Based on Propagation Path Tree Kernel Learning[J]. Computer Science, 2022, 49(6): 342-349.
- [8] LIU Y, OTT M, GOYAL N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv:1907.11692, 2019.
- [9] PATEL D, DSOUZA N, GAWANDE R. Automatic Twitter Rumour Detection using Machine Learning[C]//2022 IEEE Bombay Section Signature Conference(IBSSC). IEEE, 2022: 1-5.
- [10] MA J, GAO W, WEI Z, et al. Detect rumors using time series of social context information on microblogging websites[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. 2015: 1751-1754.
- [11] 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. 2012: 1-7.
- [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] ZHAO Z, RESNICK P, MEI Q. Enquiring minds: Early detection of rumors in social media from enquiry posts[C]//Proceedings of the 24th International Conference on World Wide Web. 2015: 1395-1405.
- [14] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter [C]//Proceedings of the 20th International Conference on World Wide Web. 2011: 675-684.
- [15] HOSSEINI D, JIN R. Graph Neural Network based Approach for Rumor Detection on Social Networks [C]//2023 International Conference on Smart Applications, Communications and Networking(SmartNets). IEEE, 2023: 1-6.
- [16] LIAO X, HUANG Z, YANG D, et al. Rumor detection in social media based on a hierarchical attention network [J]. SCIENTIA SINICA Informationis, 2018, 48(11): 1558-1574.
- [17] YANG Y J, WANG L, WANG Y H. Rumor Detection Based on Source Information and Gating Graph Neural Network[J]. Journal of Computer Research and Development, 2021, 58(7): 1412-1424.
- [18] XIE X T, HU Y Y, LIU X Z, et al. Rumor Detection Based on Representative User Characteristics Learning Through Propagation[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(6): 1334-1342.
- [19] CHANG Q, LI X, DUAN Z. A Novel Approach for Rumor Detection in Social Platforms: Memory-Augmented Transformer with Graph Convolutional Networks[J]. Knowledge-Based Systems, 2024, 292: 111625.
- [20] MA J, LIU Y, HAN M, et al. Propagation structure fusion for rumor detection based on node-level contrastive learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023: 1-12.
- [21] HUANG Q, ZHOU C, WU J, et al. Deep structure learning for rumor detection on twitter[C]//2019 International Joint Conference on Neural Networks(IJCNN). IEEE, 2019: 1-8.
- [22] KUMAR S, CARLEY K M. Tree lstms with convolution units to predict stance and rumor veracity in social media conversations[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 5047-5058.
- [23] HAQUE A, ABULAIISH M. An Emotion-Enriched and Psycholinguistics Features-Based Approach for Rumor Detection on Online Social Media[C]//Proceedings of the 11th International Workshop on Natural Language Processing for Social Media. 2023: 28-37.
- [24] CHOUDHRY A, KHATRI I, JAIN M, et al. An emotion-aware multitask approach to fake news and rumor detection using transfer learning[J]. arXiv:2211.12374, 2022.
- [25] ULONG PEI G F, DU X, PECHENIZKIY M. Dynamic network

- representation learning via gaussian embedding[C]// NeurIPS 2019 Workshop on Graph Representation Learning, 2019.
- [26] 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 (Volume 1: Long Papers). Association for Computational Linguistics, 2017: 708-717.
- [27] 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, 2016.
- [28] MA J, GAO W, WEI Z, et al. Detect rumors using time series of social context information on microblogging websites[C]// Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015: 1751-1754.
- [29] YAN Y, WANG Y, ZHENG P. A graph-based pivotal semantic mining framework for rumor detection[J]. Engineering Applications of Artificial Intelligence, 2023, 118: 105613.
- [30] ZHOU Z, QI Y, LIU Z, et al. A C-GRU neural network for rumors detection[C]// 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS). IEEE, 2018: 704-708.
- [31] MA Y, XU S, DONG F. A multilevel graph convolution neural network model for rumor detection[C]// 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA). IEEE, 2022: 1225-1229.
- [32] LI J, LI R, NI S, et al. EPRD: Exploiting prior knowledge for evidence-providing automatic rumor detection[J]. Neurocomputing, 2024, 563: 126935.
- [33] RAN H, JIA C, ZHANG P, et al. MGAT-ESM: Multi-channel graph attention neural network with event-sharing module for rumor detection[J]. Information Sciences, 2022, 592: 402-416.
- [34] CUI W, SHANG M, KAGN; knowledge-powered attention and graph convolutional networks for social media rumor detection [J]. Journal of big Data, 2023, 10(1): 45.
- [35] TAN M, LE Q V. Mixconv: Mixed depthwise convolutional kernels[J]. arXiv:1907.09595, 2019.



CHEN Xin, born in 2003, undergraduate. His main research interests include data mining and machine learning.



RONG Huan, born in 1990, Ph.D, associate professor. His main research interests include knowledge graph and knowledge engineering, social network multimedia mining, and complex machine learning theory.

(责任编辑:何杨)