



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：基于时间步局部动态交互的多任务谣言检测方法
作者：杨广浩，万书振，董方敏，王梦园
网络首发日期：2024-02-01
引用格式：杨广浩，万书振，董方敏，王梦园. 基于时间步局部动态交互的多任务谣言检测方法[J/OL]. 计算机工程与应用.
<https://link.cnki.net/urlid/11.2127.TP.20240131.1549.022>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于时间步局部动态交互的多任务谣言检测方法

杨广浩^{1,2}, 万书振^{1,2}, 董方敏^{1,2}, 王梦园^{1,2}

1. 湖北省水电工程智能视觉监测重点实验室, 湖北 宜昌 443002

2. 三峡大学计算机与信息学院, 湖北 宜昌 443002

摘要：谣言检测旨鉴别社交媒体中未经官方证实或人为捏造的信息，而当今社交网络中隐含着一种难以发掘的动态关系模式，它随时间推移和不同帖子间的动态交互而变化。针对现有方法对谣言传播事件中隐含的动态特征和关联信息考虑不充分的问题，本文提出一种基于时间步局部动态交互的多任务谣言检测方法，能捕获谣言传播事件中隐含的动态关联信息；并设计了一种高效的多任务交互方式，以时间步为基本共享单元，将学习到的局部特征进行共享，极大提升了共享效率，从而形成局部动态交互，整体多任务共享的检测框架。最后利用注意力机制筛选不同任务、不同结构特征中对谣言检测更有利的信息，以提升检测效果。我们在 PHEME 和 WEIBO 数据集上进行了实验，结果表明本方法具有较先进的性能。

关键词：谣言检测；时间步局部动态交互；传播结构特征；多任务共享

文献标志码：A **中图分类号：**TP391 **doi：**10.3778/j.issn.1002-8331.2310-0333

A multi-task rumor detection method based on time-step dynamic interaction

YANG Guanghao^{1,2}, WAN Shuzhen^{1,2}, DONG Fangmin^{1,2}, WANG Mengyuan^{1,2}

1. Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang, Hubei 443002, China

2. College of Computer and Information Technology, China Three Gorges University, Yichang, Hubei 443002, China

Abstract: Rumour detection aims to identify information on social media that has not been officially verified or artificially created. Today's social networks contain a pattern of dynamic relationships that are difficult to find, changing over time and dynamic interactions between different posts. In this paper, we propose a multi-task rumour detection method based on time step local dynamic interaction, which can capture the hidden dynamic correlation information in rumour spreading events. An efficient multi-task interaction method is designed, in which time step is used as the basic sharing unit, the local features learned are shared, and the sharing efficiency is greatly improved, so that the detection framework of local dynamic interaction and whole multi-task sharing is formed. Finally, the attention mechanism is used to filter the information that is more favorable for rumour detection in different tasks and different structural characteristics to improve the detection effect. We performed experiments on PHEME and WEIBO datasets and showed that this method has advanced performance.

基金项目：国家自然科学基金新疆联合基金重点项目 U1703261 和 U61871258；三峡库区生态环境教育部工程研究中心开放基金项目 KF2023-11。

作者简介：杨广浩(1998-)，男，硕士研究生，研究方向为自然语言处理，E-mail: guanghao_y@163.com；万书振(1976-)，通信作者，男，副教授，研究方向为演化算法、自然语言处理、智能信息处理，E-mail: wanshuzhen@163.com；董方敏(1965-)，男，教授，博士生导师。研究方向为计算机图形图像处理、计算机视觉、智能信息处理；王梦园(1999-)，女，硕士研究生，研究方向为自然语言处理。

Keywords: Rumor detection; time-step local dynamic interactions; propagation structure characterization; multitask sharing

在社会心理学中,谣言被定义为未经官方证实或人为故意捏造的信息^[1]。近年来,社交媒体已经成为人们生活中不可或缺的元素,每时每刻都有大量的信息在社交媒体中传播。但由于缺少一定监管措施,为谣言的传播提供了可能。如有关 COVID-19^[2]的消息在快速传播(例如:某些自媒体声称饮用漂白剂可以治愈该疾病)这一系列帖子后来被世界卫生组织认定是谣言,若无法及时鉴别,这些谣言将会危害社会公共安全,因此,如何有效、及时的检测谣言,保持良好社会舆论环境,是当前社交网络发展的当务之急。

随着计算机硬件水平的发展,基于深度学习的研究方法逐渐代替了传统机器学习方法。在谣言检测领域,最近的研究大多从序列数据的文本内容信息入手。如研究^[3]使用基于 RNN 的方式来捕获微博中传播事件相关帖子的上下文信息。文献^[4]构建双向 LSTM-RNN 网络,通过学习基于社交和内容的特征进行谣言预测。它们都能从谣言传播中的上下文信息中学习序列化特征,这是文本内容中最基本的特征。但这种方法有一定局限性:基于上下文信息的序列化数据仅仅包含了文本内容的语义表示,虽然这种数据用上下文顺序排列的方式来表示其中的数据结构关系,但在现今的社交网络中,数据关系的形式更多元化,仅仅用序列数据不足以表示出情况多变的社交网络模型。因此,一些研究人员开始关注于其中的传播模式。

图神经网络作为处理传播模式的热门方法,近年来有广泛应用,相关研究^[5-7]把谣言事件在社交网络中传播的关系建模为传播树或传播图,将句子分类任务转变为图分类任务^[8]。这些研究都论证了一个思路:社交网络中的传播模式能更好地用于谣言检测,因为它可以学习到谣言传播事件的传播结构,这是基于文本内容^[9]的相关方法做不到的。但是该方法也有不足:它所捕获的传播结构特征是一种全局的、静态的、缺乏交互的特征,而社交网络中的帖子是不断交互和变化的,这种变化导致对应的局部空间产生细微变化。因此,本研究希望构建一个网络,不仅能捕获谣言事件的传播模式,而且在其传播过程中,能从局部特征共享的角度进行多任务交互,以适应社交网络局部空间的变化,从而学习一种更有力特征,以提高检测效果。

综上所述,我们得到启发:可以从局部特征动态交互的角度出发,将上下文的时序文本特征与事件传播结构特征进行动态交互,将两者的优点结合,以学习一种结合了传播结构特征、时序文本特征的强大表示。同时,多任务学习的相关研究^[10-13]给了我们进一步启发,它们大多定义一个 LSTM 层作为共享网络来融合其他网络特征,共同训练,通过反向传播来更新参数。但这种硬件参数共享方式^[12]的共享效果有一定局限性,缺乏交互能力,不能更好地学习到来自多个任务的特征。因此,本研究以图神经网络和循环神经网络为基准模型,用多任务的方式结合两者,为了能更充分有效共享两个任务的特征,我们设计了时间步局部动态交互层,以时间步为基本计算单元,在训练过程中,多个任务的对应模块就开始共享彼此的特征,这比仅仅利用反向传播为多任务调参,有更好的多任务共享效果。

本研究提出了基于时间步局部动态交互网络的多任务谣言检测方法,它以多任务方式结合了图神经网络与循环神经网络,在共享方式上设计了时间步动态共享层,能从局部共享的角度将传播结构特征与对应时间步的时序文本特征相结合,最后使用注意力机制过滤输出特征。最终形成局部动态交互,整体多任务共享的检测框架,用于谣言检测。试验结果表明,所提出的方法有较为出色的表现。

本文的贡献有如下几点:

1) 提出了一种新的多任务谣言检测模型,能够捕获谣言传播事件中隐含的动态关联信息,从而形成了局部动态交互、整体多任务共享的检测框架,有效提高了检测性能。

2) 提出了一种基于时间步的局部动态交互方法,用于多任务共享。它在每个时间步都对学习到的局部特征进行共享,有效提高了共享效率和最终检测效果。

3) 我们在两个真实世界数据集上进行了实验,结果表明我们提出的方法是有效的,并且性能较好。

本文的其他部分如下:在下一部分,我们回顾了谣言检测相关工作。在第2节中,我们介绍了本文所提出方法的具体实现细节。在第3节中介绍了相关数据集、基线模型、实验设置,并对实验结果进行了分析,另外还做了丰富的补充实验。最后在第4节总结了全文。

1 相关工作

在本节中,主要论述了本文涉及的相关工作,下文将以三个角度展开。

1.1 基于内容的检测方法

在当今社交媒体传播过程中,内容信息往往依附于特定载体,如音频、图像、文字、视频等。这些特定载体被神经网络理解为文本内容特征、计算机视觉特征等。文献^[14]设计了最大池化网络分支和广义均值池化网络分支,对文本内容进行建模,以进行检测。文献^[15]提出了一种基于注意力机制的多模态融合谣言检测方法,该方法在词-视觉层面进行信息交互,利用注意力机制重点关注文本内容与图片区域之间的关联,进而进行谣言检测。文献^[16]将 Word2Vec 词向量与 LDA 主题向量进行融合,利用词语对类别的贡献加权处理,对卷积层生成的特征图进行建模学习,从而进文本主题识别工作。文献^[13]使用基于注意力机制的 LSTM 网络对谣言事件的文本内容建模,利用推文文本特征和用户特征来区分要谣言与非谣言。文献^[3]基于递归神经网络(RNN)学习微博事件连续表征,以捕捉相关帖子的上下信息随时间的变化的特征,进而进行谣言检测。文献^[17]模拟人类阅读新闻、总结、查阅资料的过程,模拟出不同内容之间的依赖关系,利用其中的关联特征进行检测。

显然,基于内容的检测方案大多数仅仅利用文本内容或单一图片内容,也有少数基于多模态的研究方案,虽然可以捕获来自多个领域的特征从而提升检测效果,但仍然受限于数据集的收集、处理。因此,本研究考虑以某种更合适的方式将文本相关工作与其他工作相结合,以捕获和融合来自多方面因素的特征,以提升谣言检测效果。文献^[18]提出了基于因果干涉和反事实推理的去重架构 CCD,利用因果干涉来消除心理语言偏差,从而进行检测。

1.2 传基于传播结构的检测方法

随着信息技术不断发展,以博客、微博为代表的社交媒体呈现了不同的社交状态,针对某一帖子,不同用户之间可以互相点赞、评论、转发,转发的帖子也同样会被“转赞评”,这种逻辑上的社交网络形成了一种传播结构。而这种传播模式也能用于谣言检测,如文献^[19]提出一种用于谣言嵌入和分类的传播路径聚合网络,它

将每条谣言的传播结构建模为一组独立的传播路径,进行谣言检测。文献^[8]对数据集中的帖子进行深度关联性处理,它将帖子视作节点,边视为帖子中文本句子之间的相似性,以构建每个帖子所属的传播结构,并对传播结构建模,进而将帖子的文本分类任务转变为图分类任务。文献^[20]以另一种更好的方式构建图结构,它根据 Twitter 上帖子“谁回复谁”的关系来构建传播图,随后提出一种基于门控图神经网络的算法,对此处理,最终进行谣言检测。文献^[21]提出了一种带有事件共享模块的端到端多通道图注意力网络,它用图构建出源推文以回复的关系、源推文以及词语的关系,最后加入事件共享模块加以训练,得到最终预测结果。文献^[22]利用社交图细化的方法解决问题,将度校正随机块模型应用于社交图网络,以增强检测效果。

以上相关方法都证明了:传播结构可以进行谣言检测。但是以上相关方法构建的传播结构都展现的是传播事件在发展后期全局的、非交互的状态。然而在现实世界中,社交媒体间不同的节点间是不断变化并交互关联的。

1.3 基于多任务的检测方法

多任务学习旨联合多个相关、能共享,但不完全相同任务信息的联合学习行为^[23],次要任务起到支撑、辅助主任务的作用,通过交互和共享来补充主任务中被忽略或难以捕获的特征,在针对某些难以发掘特征的任务时,多任务学习能起到关键作用。例如文献^[24]提出了一种将谣言检测和情感分析相结合的多任务检测方法,它设计了 BERT 和 Bi-GRU 联合框架,利用权值共享的方法对两个任务进行训练,从而提升谣言检测性能。文献^[25]通过权重共享的多任务学习方式,利用一个单一的卷积神经网络架构,给定一个句子输出多种不同的预测,将多个任务联合训练。文献^[12]提出了一种多任务学习方式同时结合了谣言检测、谣言跟踪和立场分类,从而提高谣言预测的性能。文献^[26]提出了一种联合学习模型 HG-SL,它设计了基于超图的全局交互学习模块,通过全局和局部的用户传播行为差异,进行联合训练,从而进行检测。

然而,以上提出的大多数多任务实现共享的方式往往是定义一个 RNN 作为共享层,通过反向传播,用硬件参数共享的方式来实现多个任务之间的特征交互。

这种方式虽然一定程度上实现了多个任务见的交互,但是硬件参数共享过于依赖计算机硬件,当前研究缺乏一种软件上、算法上的多任务特征共享方案。因此,本研究在多任务硬件参数共享的基础之上,设计了一种在训练过程中就开始共享交互的多任务网络。

2 基于时间步局部动态交互的多任务谣言检测方法

针对现有方法对谣言传播事件中隐含的局部动态

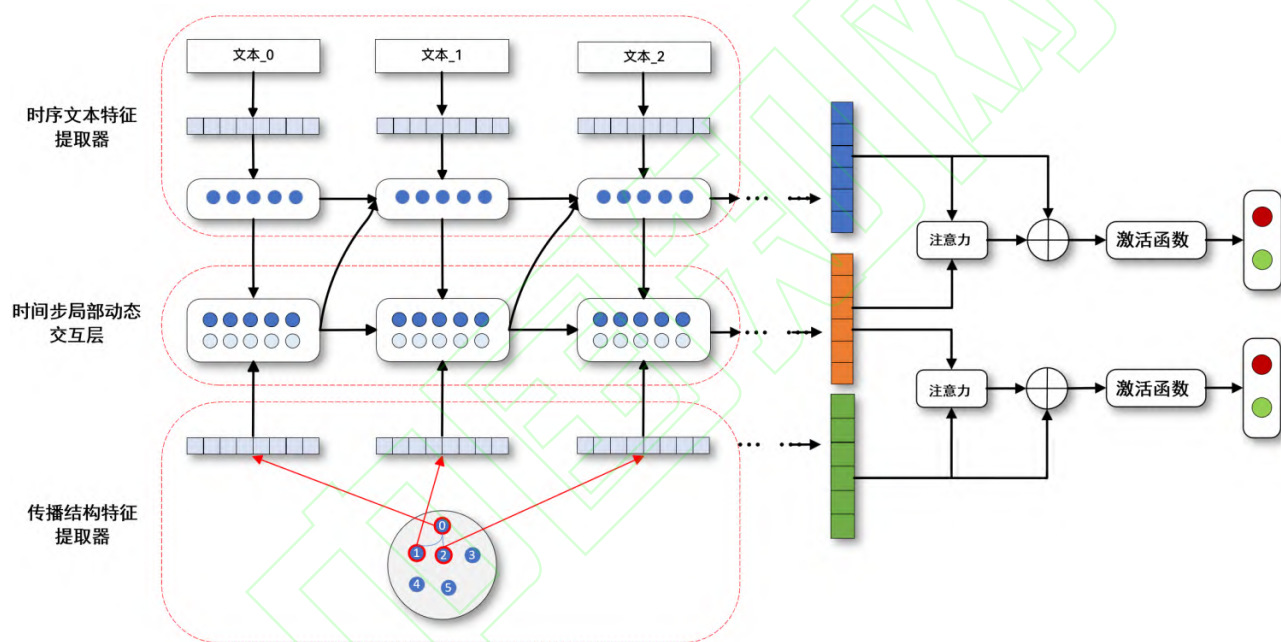


图1 时间步动态交互多任务谣言检测网络图

Fig.1 Time-step dynamic interaction multitasking rumor detection network diagram

1) 时序文本特征提取器: 输入由谣言传播事件对应帖子的文本内容构成,以 LSTM 网络为基准模型并作修改,提取文本内容特征。

2) 传播结构特征提取器: 输入由谣言传播事件的传播结构组成,以 GCN 网络为基准模型,提取与帖子对应的传播结构特征。

3) 时间步局部动态交互层: 作为共享层,以时间步为基本计算和共享单元。将上述两种特征有效结合,以实现动态信息交互,整体多任务共享的工作模式,从而充分挖掘社交网络中隐含的动态交互特征,提升谣言检测效果。

4) 注意力机制和分类网络: 利用注意力机制筛选、更新特征,重点关注对预测更有利的信息,最后经过分类网络输出预测结果。

特征和关联信息考虑不充分的问题,本文提出一种基于时间步局部动态交互的多任务谣言检测方法,综合考虑谣言传播事件中文本内容特征与传播结构特征的动态交互,能学习到更为强大的特征表示,更有利于谣言检测。

整体工作流程如图 1 所示,它主要包括时序文本特征提取器、时间步局部动态交互层、传播结构特征提取器、注意力机制和分类网络。如下:

2.1 传播结构特征提取器

图神经网络作为处理传播模式的经典方案,能获取谣言事件的传播结构特征。为了尽可能学习到谣言事件在传播中动态变化和交互的情况,本节首先设计了传播结构特征提取器,它以 GCN 为基准网络并做了修改,以捕获谣言传播事件特定时间步的传播结构特征,随后这些信息进入动态共享层与文本内容特征进行交互(2.3 中详细描述),不仅能增强了对应文本特征的全局表示,也提高了多任务信息交互的效率。如图 1 下方所示,输入数据为谣言传播事件的传播图结构,利用图卷积网络对其处理,捕获对应的传播结构特征,如公式(1):

$$H_{ir}^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H_{ir}^l W^l) \quad (1)$$

其中, \tilde{D} 为度矩阵, \tilde{A} 为邻接矩阵, H_{ir}^l 为上一层传播图所有节点的传播结构特征矩阵, 它的特征数与对应谣言传播事件的贴子数直接相关。 W^l 为权重矩阵, σ 是非线性激活函数。为了实现 2.3 节中的动态交互, 我们从当前传播结构特征矩阵 $H_{ir}^{(l+1)}$ 中截取每个时间步中对应的局部节点传播结构特征, 如公式 (2):

$$\{H_{v0}, H_{v1}, \dots, H_{vm}\} = H_{ir}^{(l+1)} \in \mathbb{R}^{\text{nodes} * F} \quad (2)$$

其中, $H_{v0} \in \mathbb{R}^F$ 表示传播结构特征矩阵 $H_{ir}^{(l+1)}$ 对应的传播过程中产生的第零个节点向量表示, 它主要包含了自身特征, 也聚合了少部分邻居节点的特征。

同理, 在每个时刻都对特征矩阵做相同处理, 都得到不同帖子对应的局部节点特征 $H_{v0}, H_{v1}, \dots, H_{vm}$, 反应了整个谣言传播事件的局部节点特征, 使之与公式 (4) 的时序文本特征相对应, 用于后续动态交互。

2.2 时序文本特征提取器

循环神经网络过作为处理时序文本数据的常用方法, 可以捕捉上下文之间的序列特征, 谣言事件在传播时呈现上下文之间的序列特性。但现实世界中, 谣言事件的时序上下文也与对应的局部传播结构相关联, 他们彼此间的交互会相互影响谣言事件整体的发展。

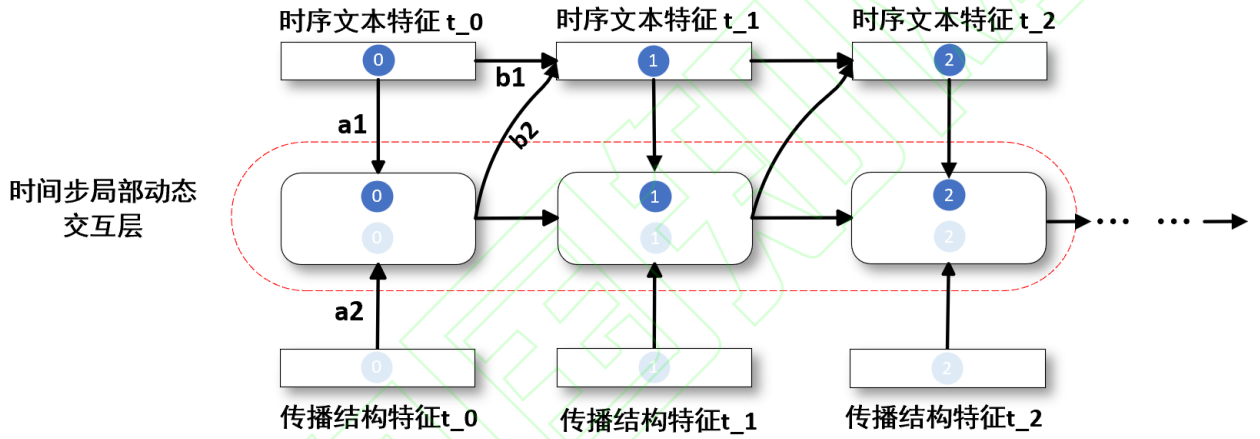


图 2 时间步动态交互层

Fig.2 Time-Step Dynamic Interaction Layer

本模块对谣言传播事件的上下文内容进行处理,以进行时间步的动态交互(2.3中详细描述),从而结合对应的局部传播结构特征。不仅能赋予对应传播特征一定的局部语义表示,也提升了多任务信息交互的效率。

如图1上方所示,输入数据为谣言传播事件对应帖子的文本内容,设某一谣言传播事件所包含的对应文本内容表示为 $X_{i_content} = \{X_{i_0}, X_{i_1}, \dots, X_{i_n}\}$, 作为时序文本特征提取器的输入, 如式(3):

$$\left. \begin{aligned} i_t &= \sigma(W_{ii} X_{i_content_t} + b_{ii} + W_{hi} h_{(t-1)} + b_{hi}) \\ f_t &= \sigma(W_{if} X_{i_content_t} + b_{if} + W_{hf} h_{(t-1)} + b_{hf}) \\ g_t &= \tanh(W_{ig} X_{i_content_t} + b_{ig} + W_{hg} h_{(t-1)} + b_{hg}) \\ o_t &= \sigma(W_{io} X_{i_content_t} + b_{io} + W_{ho} h_{(t-1)} + b_{ho}) \\ c_t &= f_t * c_{(t-1)} + i_t * g_t \\ H_{content_t} &= o_t * \tanh(c_t) \end{aligned} \right\} \quad (3)$$

其中 $X_{i_content_t}$ 为对应时间步输入到 LSTM 细胞块中的特征向量, $h_{(t-1)}$ 为上一时刻的隐藏向量, W_* 以及 b_* 为对应权值矩阵和偏置。

对所有时间步都进行上述计算,可以捕获每个时间步中对应文本内容的特征向量,从而形成一组时序文本特征 $H_{content}$, 如式(4):

$$\{H_{content_0}, H_{content_1}, \dots, H_{content_t}\} = H_{content} \quad (4)$$

需要强调的是, 时序文本特征 $\{H_{content_0}, H_{content_1}, \dots, H_{content_t}\} = H_{content}$ 与式(2)中传播结构特征 $\{H_{v0}, H_{v1}, \dots, H_{vn}\} = H_{ir}^{(l+1)}$ 源自同一谣言传播事件,在谣言传播事件的发生和处理顺序上是一一对应的,时序上的同步有助于更好地进行后续的动态交互。

2.3 时间步局部动态交互层

在上述两个小节里,传播结构特征提取器在2.1中经式(2)得到了对应传播结构特征 $\{H_{v0}, H_{v1}, \dots, H_{vn}\} = H_{ir}^{(l+1)}$; 时序文本特征提取器在2.2中经式(4)得到了时序文本特征 $\{H_{content_0}, H_{content_1}, \dots, H_{content_t}\} = H_{content}$, 它们在谣言传播中的传播顺序是有一定关系的。因此,本节构建了时间步局部动态交互层,以时间步为基本交互单元结合它们,以学习和挖掘传播事件更多元、更高级的局部动态关系,从而提升模型检测效果。

如图2所示。最上一层简要表示时序文本特征,对应图1中时序文本特征提取器;最下一层传播结构

特征,对应图1中传播结构特征提取器。工作流程如下:

第一步,构建初始时刻动态交互单元:

如图2中“a1”及“a2”标记。取 t_0 时刻的时序文本特征 $H_{content_0}$ 及传播结构特征 H_{v0} 进行拼接,并经过线性层映射,作为动态交互层第一个细胞块的输入,并经过门控单元计算,得到输出。如式(5):

$$\left. \begin{aligned} X_{share_0} &= Linear(Concat(H_{v0}, H_{content_0})) \\ i_t &= \sigma(W_{ii} X_{share_0} + b_{ii} + W_{hi} h_{(t-1)} + b_{hi}) \\ f_t &= \sigma(W_{if} X_{share_0} + b_{if} + W_{hf} h_{(t-1)} + b_{hf}) \\ g_t &= \tanh(W_{ig} X_{share_0} + b_{ig} + W_{hg} h_{(t-1)} + b_{hg}) \\ o_t &= \sigma(W_{io} X_{share_0} + b_{io} + W_{ho} h_{(t-1)} + b_{ho}) \\ c_t &= f_t * c_{(t-1)} + i_t * g_t \\ H_{share_0} &= o_t * \tanh(c_t) \end{aligned} \right\} \quad (5)$$

此计算过程与公式(3)基本一致,只是输入输出数据有所区别,此处不再一一赘述。

需要强调的是,此时,如公式(3)在时序文本特征提取器的 LSTM 细胞块中也在进行类似的操作,得到 t_0 时刻的输出 $H_{content_0}$, 如图2中的“b1”标记。

第二步,更新下一时刻隐藏状态:

如图2中“b2”及“b1”标记,将 $H_{content_0}$ 和 H_{share_0} 进行拼接,并经过线性层映射,作为下一时刻的时序文本特征提取器的输入,如式(6):

$$H_{content_0} = Linear(Concat(H_{content_0}, H_{share_0})) \quad (6)$$

第三步,重复执行一、二步,直到所有时刻的细胞块都计算完毕,最终我们得到动态交互层的输出。

此时在动态交互层中已经完成全部计算。

如图1右侧,最终我们分别得到时序文本特征提取器、动态交互层、传播结构特征提取器的输出:

$$H_{content}, H_{share}, H_{ir}$$

整体传播结构特征以局部动态共享方式,赋予了时序文本特征全局结构表示;时序文本特征以具体时间步共享方式,增强了传播结构特征的局部语义表示。两者以时间步为基本计算单位,互相交互,共同学习,学习到谣言传播模式中更隐含的动态交互信息,最终提升模型整体泛化能力。

2.4 注意力机制及分类网络

在上一小节中,已经完成了各自模块的计算过程,并得到各自输出。在本小节中,为了学习不同任务、不同时间步、不同局部传播结构特征对谣言预测有什

么重要影响,我们设计了注意力机制和分类网络,使用注意力机制筛选重要信息,以得到更强健的特征表示,最后作为分类器的输入。具体计算过程如下:

步骤一,如图1右侧所示,将动态共享层的输出 H_{share} 分别和 $H_{content}$ 、 H_{ir} 拼接,并经过线性映射层,得到注意力层的初始输入 X_1 、 X_2 ,如公式(7-8):

$$X_1 = Linear(Concat(H_{content}, H_{share})) \quad (7)$$

$$X_2 = Linear(Concat(H_{ir}, H_{share})) \quad (8)$$

步骤二, X_1 、 X_2 各自经过线性变换得到投影 (query Q; key K; value V), 作为注意力层的输入,如式(9-11):

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^o \quad (9)$$

$$head_i = Attention(XW_i^Q, XW_i^K, XW_i^V) \quad (10)$$

$$Attention(Q, K, V) = \text{soft max}(\frac{QK^T}{\sqrt{d_k}})V \quad (11)$$

$MultiHead(Q, K, V)$ 是注意力层的输出, Q、K、V 分别表示 XW_i^Q 、 XW_i^K 、 XW_i^V , W^o 是权重矩阵, W_i 表示不同多头矩阵的初始化, d_k 是 k 的维度, 当 K 较大时缩放输出。

步骤三, 用注意力层的输出分别更新原本的 $H_{content}$ 和 H_{ir} , 分别拼接, 随后经过线性层映射, 如式(12-13):

$$H_{content} = Linear(Concat(H_{share}, H_{content})) \quad (12)$$

$$H_{ir} = Linear(Concat(H_{share}, H_{ir})) \quad (13)$$

步骤四, 将更新后的特征用于分类器的输入, 设计各自的全连接层并且激活, 如式(14-15):

$$Y_C = \text{soft max}(W_C H_{content} + b_C) \quad (14)$$

$$Y_G = \text{soft max}(W_G H_{ir} + b_G) \quad (15)$$

W_* 、 b_* 是他们各自对应的权重和偏置参数, 用交叉熵损失函数作为分类损失, 如式(16-17):

$$L_C = -\sum_i Y_i \log Y_C \quad (16)$$

$$L_G = -\sum_i Y_i \log Y_G \quad (17)$$

Y_i 表示第 i 个事件的真实标签。训练时, 将两个损失函数赋予对应系数, 相加求得总损失函数, 进行反向传播, 如式(18):

$$Loss = W1 * L_C + W2 * L_G \quad (18)$$

3 实验

本节使用两个公开数据集 WEIBO^[7]和 PHEME^[12]进行对比实验, 并设计了消融实验来说明各个模块的有效性。

表1 数据集统计数据

Table 1 Statistics of Datasets

Statistic	Weibo	PHEME
#of events	4663	5473
#of Non-rumors	2312	1996
#of Rumors	2351	3477
Avg.#of posts/event	452	20
Max.#of posts/event	44764	346
Min.#of posts/event	2	3

3.1 数据集

数据集的详细描述如表1所示, 它包含来自微博¹和推特²两个社交网站的信息。每个传播事件包含对应评论、点赞、转发等互动关系, 并且每个传播事件有一个标签, 谣言或非谣言。在去除了部分样本后, 两个数据集的详细可用信息如表1所示。在尝试了若干次实验验证有效性之后, 经过结果反馈, 将 WEIBO 数据集划分为 6:2:2, PHEME 数据集划分为 8:1:1。使用准确率、精确率、召回率和 F1 分数来评价性能。计算公式分别如式(19), 其中 TP 表示真阳性率, TN 表示真阴性率, FP 表示假阳性率, FN 表示假阴性率。

$$\left. \begin{aligned} Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\ Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F1 &= \frac{2 * Precision * Recall}{Precision + Recall} \end{aligned} \right\} \quad (19)$$

表2 主要参数设置

Table 2 Main hyperparameter settings

Dataset	Hyper parameter	
PHEME	Size of feature	300
	Average subgraph nodes	20
	Max subgraph nodes	346
	Time steps	20
	Hidden Dim	128
	Batch size	128
Weibo	Size of feature	768
	Average subgraph nodes	452

Max subgraph nodes	44764
Time steps	452
Hidden Dim	128
Batch size	32

3.2 实验设置

本实验环境中使用操作系统为 Linux, GPU 为 NVIDIA 2080ti, 使用 Pytorch 框架实现该算法。

对于对比方法, 直接从原论文或其他论文中获取实验结果, 已在表 3 和表 4 中标注数据来源。对于本实验, 使用 Adma 算法优化参数, 学习率设置为 $1e-3$, 在 WEIBO 数据集上批次大小设置为 32, PHEME 数据集设置为 128。主要超参数如表 2 所示。

3.3 基线模型

为了突出本文模型的有效性, 并进行评价, 选取了若干典型模型或方法作为对比, 具体如下:

(1) DTC^[27]: 一种使用决策树分类器和手工创建特征的谣言检测方法。

(2) SVM-RBF^[28]: 一种使用带有 RBF 卷积核的 SVM 模型, 根据帖子的总体统计数据使用手工制作的特征进行分类的方法

(3) SVM-TS^[29]: 一种基于时间序列结构的 SVM 分类器, 通过模拟谣言的内容、用户、传播模式及其对社会的影响来提取特征的方法。

(5) RvNN^[7]: 一种使用递归神经网络对谣言事件的文本内容和树状传播结构建模, 提取高层表征的方法。

(6) Bi-GCN^[5]: 基于 GCN 的谣言检测模型, 利用双向传播结构获取传播和扩散特征。

(7) STS-NN^[30]: 一种时空结构神经网络, 它将空间结构和时间结构视作一个整体, 为谣言检测建立信息传播模型。

(8) DDGCN^[31]: 一种双动态图卷积网络, 在一个统一的框架内对空间结构、时间结构、外部知识和文本信息进行建模。

3.4 实验结果与分析

在本节中我们基于 WEIBO 和 PHEME 数据集与其他基线方法进行了比较分析, 表 3 表 4 对比较结果做了详细展示。第一组为传统机器学习方法, 第二组

为当前较为典型的深度学习方案。

从表 3 和表 4 中得到以下结论:

第一组的传统方案效果不好, 绝大部分指标都低于第二组深度学习方案。这是因为深度学习方案相比传统方法可以捕获更抽象更高级的特征信息, 它们可以学习到更深层次的特征, 因此分类效果更好。但第一组内互相比较也可以发现: SVM-TS 的表现较为出色, 这是因为它构建了一个动态系列的时间结构模型, 对于社会背景相关、以及时间变化等方面的信息有较好的反馈。

在众多深度学习方法中, STS-NN 比 RvNN 要稍好一点, 准确率在两个数据集中要分别高出 0.55%、3.8%, 这主要是由于 STS-NN 从时空结构的角度对信息传播进行了分类, 并且所有时空计算单元都是参数共享的, 捕捉的时空信息在训练过程中有了更强的效果。而 Bi-GCN 相较于 STS-NN 又有更好的性能, 能够体现模型整体综合性能的 F1 参数分别高出 1.4%、4.37%, 这是因为 Bi-GCN 将空间结构视为传播和散播两个过程, 它从自顶向下和自底向上两个过程训练模型, 最后将两个过程学习到的特征进行交互, 使 GCN 中每一层在交互中都利用到了两个方向的特征, 从而学习到更强的特征表示。除本文方法外, DDGCN 具有最好的效果, 它的 F1 参数比 Bi-GCN 分别高出 4%、1.07%, 这是因为它用一个时态融合单元将动态图卷积网络与知识图谱结合, 结合两类信息并在训练时动态交互, 逐步聚合, 从而更好地检测谣言。

但本文提出的方案具有更好的性能。与使用了双向图卷积的 Bi-GCN 相比, 我们从多任务动态交互的角度结合了传播模式与时序文本这两种信息, 能够学习到更多元的特征表示, 准确率比使用了 Bi-GCN 还要好 4.35%、1.18%; 与 DDGCN 相比, 具有更高效的交互方式, 我们基于时间步设计了一种多任务信息和参数的共享交互方案, 使得多个任务的特征都在对应时间步逐层动态交互。当前时间步的特征同时结合了当前传播的局部传播结构信息与对应的时序文本信息, 并作为隐藏状态传递给下一时间步, 随着时间步的动

态发展,学习到越来越强大的多维度特征。因此,本方法在能体现模型整体综合性能的 F1 参数上分别超过了 0.42%、0.71%。

表 3 WEIBO 数据集上各模型实验结果统计

Table 3 Statistics of experimental results for each model on the WEIBO dataset

Method	Acc	Pre	Re-call	F1
DTC ^[32]	0.858	0.834	0.882	0.857
SVM-RBF ^[32]	0.899	0.938	0.846	0.889
SVM-TS ^[11]	0.885	0.950	0.932	0.938
$RvNN_{BU}(ACL18)^{[7]}$	0.908	0.912	0.897	0.905
STS-NN ^[31]	0.913	0.902	0.898	0.900
Bi-GCN(AAAI20) ^[31]	0.919	0.918	0.916	0.913
DDGCN(AAAI22) ^[31]	0.948	0.953	0.948	0.950
OURS	0.959	0.955	0.955	0.954

表 4 PHEME 数据集上各模型实验结果统计

Table 4 Statistics of experimental results for each model on the PHEME dataset

Method	Acc	Pre	Re-call	F1
DTC ^[32]	0.581	0.659	0.652	0.656
SVM-RBF ^[32]	0.602	0.875	0.431	0.577
SVM-TS ^[11]	0.651	0.642	0.686	0.663
$RvNN_{BU}(ACL18)^{[31]}$	0.789	0.788	0.788	0.789
STS-NN ^[31]	0.819	0.816	0.791	0.800
Bi-GCN(AAAI20) ^[31]	0.847	0.840	0.834	0.835
DDGCN(AAAI22) ^[31]	0.855	0.846	0.841	0.844
OURS	0.857	0.848	0.851	0.849

3.5 消融实验

在本节中,为了探索模型的各个模块有何作用,我们逐步去除各个模块并进行实验分析。

定义如下对比策略:

WO/GCN: 去除图网络部分

WO/LSTM: 去除文本特征提取器

WO/SHARE: 以多任务结合两个网络,但去除动态交互层

从表 5 中显而易见:

WO/GCN 的方案性能是最差的,这是因为这种结构已经不再是多任务模式,失去了对应的共对象,仅

仅捕获了谣言事件的文本内容特征,不足以达到理想的效果。

表 5 不同策略性能综合对比

Table 5 Comprehensive comparison of the performance of different strategies

Dataset	Method	Acc	Pre	Recall	F1
Weibo	w/o GCN	0.922	0.917	0.920	0.918
	w/o LSTM	0.928	0.921	0.924	0.922
	w/o SHARE	0.945	0.948	0.944	0.945
	OURS	0.959	0.952	0.955	0.953
PHEME	w/o GCN	0.822	0.817	0.818	0.817
	w/o LSTM	0.825	0.823	0.826	0.824
	w/o SHARE	0.844	0.841	0.848	0.844
	OURS	0.857	0.858	0.855	0.856

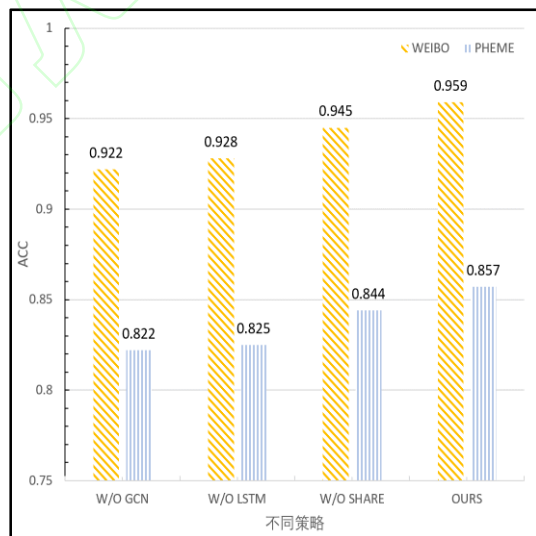


图 3 不同策略下准确率对比

Fig.3 Comparison of accuracy rates under different strategies

WO/LSTM 的方案效果稍好一点,因为它利用了谣言事件的传播结构特征。

WO/SHARE 的方案已经较为出色,因为它以多任务的方式结合了前两种网络,能够捕获到同时结合了传播模式及文本内容的有力特征,并且通过多任务的参数共享,促进了特征交互。

但本文提出的方案效果最好,因为它在 WO/SHARE 多任务的方案之上特别设计了动态交互层,以时间步为基本计算单元,在传播事件发展的对应时间步里同时融合了局部传播结构特征及对应的时

序文本内容特征。如图 3 也做了准确率的直方图对比。

3.6 早期检测

早期检测是谣言检测的重要目标之一,为了进一步探索本文提出方法的各方面性能,本节还做了早期检测的相关研究。

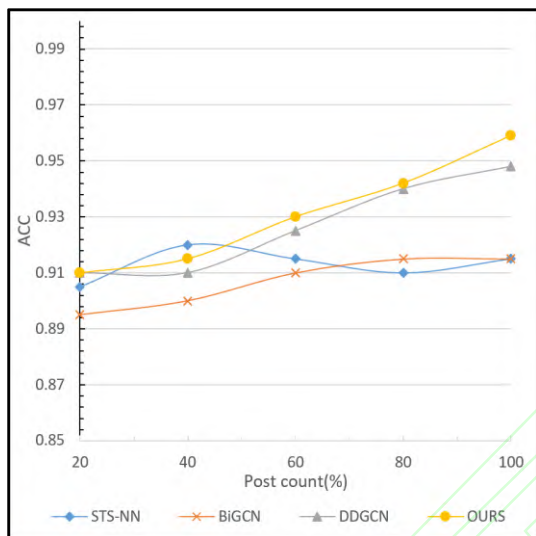


图 4 WEIBO 数据集早期检测结果

Fig.4 Early detection results from the WEIBO dataset

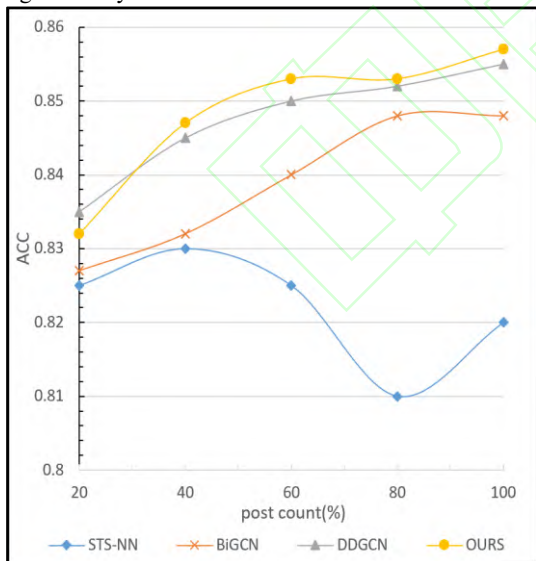


图 5 PHEME 数据集早期检测结果

Fig.5 Early detection results from the PHEME dataset

为了基于本文所提出的模型进行早期检测任务,本节通过控制时间步的数量,来模拟谣言传播事件的不同阶段。将对应数据集中不同传播事件的平均帖子数量视作最大时间步,对于非本时间阶段的时间步直接做删除处理,通过这种方式重构对应传播事件的数据集,其余操作保持一致,最后评估不同早期阶段模型的性能。如图 4 图 5 所示,分别展示了本文提出的

模型与其他方法在不同数据集上的差异。

从结果上看,本文提出的方法在早期检测的大多数阶段都有较好的表现。Bi-GCN 和 STS-NN 的效果是较差的,这是由于 Bi-GCN 在设计时仅采用了自顶向下和自底向上两个方向图,从不同的方向捕获结构特征,虽然对全局特征有更好的表现,但无法反映出动态关系的细微变化。

而 STS-NN 并没有采用图结构,无法捕捉到谣言事件早期传播结构特征,显然也不会有较好的早期检测性能。DDGCN 由于使用了双动态图卷积,在谣言发展的不同阶段,图也会有不同的状态,因此性能较为出色。但本文提出的方法综合表现更好,得益于我们使用了多任务的方式,提出的时间步局部动态交互层有更好的共享效果,并且同时捕获了来自多个方面的特征,最后通过注意力机制的过滤,使得模型能学习到谣言事件更隐含的动态联系。因此在两个数据集上,本文提出的方案都有更好的效果。

4 结束语

在本文中,我们以时间步局部动态交互的视角创新型地提出了一种全新的网络。在共享方式上额外设计了动态交互层,它极大地增强了多任务学习中特征及参数共享的效率,并同时捕获了谣言事件的文本内容特征以及传播结构特征,最后使用注意力机制重点关注不同任务、不同局部传播结构中对谣言检测有利的信息,以提升整体网络性能。实验证明,我们提出的方法能较好的进行谣言检测。除此之外还做了丰富的补充实验,以消融实验证明了本文提出方法各个模块的有效性,以早期实验验证了本方法在早期检测任务中的性能。

但本方法也有一定不足之处:比如我们虽然进行了动态交互,但对传播结构的动态图网络还需要进一步研究。

在未来,我们要进一步探索多任务动态交互中的特性,将开展基于动态图网络和多任务动态交互进行研究,以探索更符合现实生活中谣言传播的过程。

参考文献:

- [1] DiFonzo N, Bordia P. Rumor psychology: Social and organizational approaches[M]. American Psychological Association, 2007.
- [2] Cohen J, Normile D. New SARS-like virus in China triggers alarm[J]. Science, 2020, 367(6475): 234-235.

- [3] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks.(2016)[C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016). 3818-3824.
- [4] Ahmad T, Faisal M S, Rizwan A, et al. Efficient fake news detection mechanism using enhanced deep learning model[J]. Applied Sciences, 2022, 12(3): 1743.
- [5] Bian T, Xiao X, Xu T Y, 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.
- [6] Bai N, Meng F R, Rui X B, et al. Rumor detection based on a source-replies conversation tree convolutional neural net[J]. Computing, 2022: 1-17.
- [7] Ma J, Gao W, Wong K F. Rumor detection on twitter with tree-structured recursive neural networks[C]//Proceedings of the Association for Computational Linguistics, 2018.
- [8] Vaibhav V, Annasamy R M, Hovy E. Do sentence interactions matter? leveraging sentence level representations for fake news classification[J]. arXiv preprint arXiv:1910.12203, 2019.
- [9] Zhou X, Zafarani R. A survey of fake news: Fundamental theories, detection methods, and opportunities[J]. ACM Computing Surveys (CSUR), 2020, 53(5): 1-40.
- [10] Ma J, Gao W, Wong K F. Detect rumor and stance jointly by neural multi-task learning[C]//Proceedings of the Companion proceedings of the the web conference 2018. 2018: 585-593.
- [11] Wan S Z, Tang B, Dong F M, et al. A writing style-based multi-task model with the hierarchical attention for rumor detection[J]. International Journal of Machine Learning and Cybernetics, 2023: 1-16.
- [12] Kochkina E, Liakata M, Zubiaga A. All-in-one: Multi-task learning for rumour verification[J]. arXiv preprint arXiv:1806.03713, 2018.
- [13] Singh J P, Kumar A, Rana N P, et al. Attention-based LSTM network for rumor veracity estimation of tweets[J]. Information Systems Frontiers, 2020: 1-16.
- [14] 郭铃霓, 黄舰, 吴兴财, 等. 基于双分支网络联合训练的虚假新闻检测[J]. 计算机工程与应用, 2022, 58(15).
Guo L N, Huang J, Wu X C, et al. Fake News Detection Based on Joint Training Two-Branch Network [J]. Journal of Computer Engineering & Applications, 2022, 58(15).
- [15] 威力鑫, 万书振, 唐斌, 等. 基于注意力机制的多模态融合谣言检测方法[J]. 计算机工程与应用, 2022, 58(19).
Qi L X, Wan S Z, Tang B, et al. Multimodal Fusion Rumor Detection Method Based on Attention Mechanism[J]. Journal of Computer Engineering & Applications, 2022, 58(19).
- [16] 邱宁佳, 杨长庚, 王鹏, 等. 改进卷积神经网络的文本主题识别算法研究[J]. 计算机工程与应用, 2022, 58(2).
Qiu N J, Yang C G, Wang P, et al. Research on Text Topic Recognition Algorithm Based on Improved Convolutional Neural Network[J]. Journal of Computer Engineering & Applications, 2022, 58(2).
- [17] Liao H, Peng J H, Huang Z Y, et al. MUSER: A Multi-Step Evidence Retrieval Enhancement Framework for Fake News Detection[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 4461-4472.
- [18] Chen Z W, Hu L M, Li W X, et al. Causal Intervention and Counterfactual Reasoning for Multi-modal Fake News Detection[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023: 627-638.
- [19] Zhang P F, Ran H Y, Jia C Y, et al. A lightweight propagation path aggregating network with neural topic model for rumor detection[J]. Neurocomputing, 2021, 458: 468-477.
- [20] Wu Z Y, Pi D C, Chen J F, et al. Rumor detection based on propagation graph neural network with attention mechanism[J]. Expert systems with applications, 2020, 158: 113595.
- [21] Ran H Y, Jia C Y, Zhang P F, et al. MGAT-ESM: Multi-channel graph attention neural network with event-sharing module for rumor detection[J]. Information Sciences, 2022, 592: 402-416.
- [22] Wu J Y, Bryan Hooi. DECOR: Degree-Corrected Social Graph Refinement for Fake News Detection[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 2582-2593.
- [23] Zhang Y, Yang Q. A survey on multi-task learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(12): 5586-5609.
- [24] 沈瑞琳, 潘伟民, 彭成. 基于多任务学习的微博谣言检测方法[J]. 计算机工程与应用, 2021, 57(24): 192-197.
Shen R L, Pan W M, Peng C. Microblog Rumor Detection Method Based on Multi-task Learning[J]. Journal of Computer Engineering & Applications, 2021, 57(24): 192-197.
- [25] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]//Proceedings of the 25th international conference on Machine learning. 2008: 160-167.
- [26] Sun L, Rao Y, Lan Y Q, et al. HG-SL: Jointly Learning of Global and Local User Spreading Behavior for Fake News Early Detection[J]//Proceedings of the AAAI conference on artificial intelligence. 2023, 37(4): 5248-5256.
- [27] Castillo C, Mendoza M, Poblete B. Information credibility on twitter[C]//Proceedings of the 20th international conference on World wide web. 2011: 675-684.
- [28] Yang F, Liu Y, Yu X H, et al. Automatic detection of rumor on sina weibo[C]//Proceedings of the ACM SIGKDD workshop on mining data semantics. 2012: 1-7.
- [29] Ma J, Gao W, Wei Z Y, 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.
- [30] Huang Q, Zhou C, Wu J, et al. Deep spatial-temporal structure learning for rumor detection on Twitter[J]. Neural Computing and Applications, 2020: 1-11.
- [31] Sun M Z, Zhang X, Zheng J Q, et al. Ddgc: Dual dynamic graph convolutional networks for rumor detection on social

- media[C]//Proceedings of the AAAI conference on artificial intelligence. 2022, 36(4): 4611-4619.
- [32] Han S, Yu K, Su X, et al. Combining Temporal and Interactive Features for Rumor Detection: A Graph Neural Network Based Model[J]. Neural Processing Letters, 2023, 55(5): 5675-5691.

