

社交媒体假新闻检测：基本理论、方法及研究方向

赵梦凡¹, 张钰涛¹, 赵铤钊²

(1. 郑州西亚斯学院 计算机与软件工程学院, 河南 郑州 451150; 2. 石家庄学院 理学院, 河北 石家庄 050035)

摘要: 社交媒体平台的迅速发展不仅极大增强了信息的可访问性, 而且加速了假新闻的传播。假新闻的爆炸性增长不仅损害社会稳定, 还会侵蚀公众对媒体的信任。在自然语言处理领域中, 假新闻检测是一个关键而富有挑战性的任务。为此, 首先给出假新闻的定义, 深入分析其特征; 其次从新闻内容、社交语境、传播网络和混合方法4个角度对现有假新闻检测方法进行分析与评估, 介绍相关模型性能、常用数据集以及评价指标; 最后, 总结并分析目前假新闻检测研究中存在的问题, 提出后续可能的研究方向。

关键词: 社交网络; 假新闻; 自然语言处理; 早期检测; 可解释性

DOI: 10.11907/rjdk.231367

开放科学(资源服务)标识码(OSID):



中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2024)009-0031-10

Detection of Fake News on Social Media: Basic Theories, Methods and Research Prospects

ZHAO Mengfan¹, ZHANG Yutao¹, ZHAO Tingzhao²

(1. Computer and Software Engineering, Sias University, Zhengzhou 451150, China;

2. Science College, Shijiazhuang University, Shijiazhuang 050035, China)

Abstract: The rapid development of social media platforms not only greatly enhances the accessibility of information, but also accelerates the spread of fake news. The explosive growth of fake news not only damages social stability, but also erodes public trust in the media. In the field of natural language processing, fake news detection is a crucial and challenging task. To this end, first provide a definition of fake news and analyze its characteristics in depth; Then, analyze and evaluate the existing fake news detection methods from four perspectives: news content, social context, communication networks, and hybrid methods. Introduce the performance of relevant models, commonly used datasets, and evaluation indicators; Finally, summarize and analyze the current problems in fake news detection research, and propose possible future research directions.

Key Words: social network; fake news; natural language processing; early detection; explainability

0 引言

社交媒体的快速发展为公众提供了一个沟通交流、高度互联的平台, 每个人都能在社交平台上发表言论和看法, 阅读和分享各种信息。社交媒体实用便捷、开放包容、个人信息隐匿化的特点吸引了越来越多的用户加入, 成为

人们日常生活中不可或缺的一部分。然而, 信息隐匿化是一把双刃剑, 社交媒体平台在鼓励用户真实表达自我的同时也催生了很多虚假信息^[1]。

社交媒体平台用户基数大、真实身份不明, 难以进行有效管控, 导致假新闻很容易在社交网络上快速传播。假新闻的快速传播会扭曲大众价值观, 导致社会动荡。例如, 2016年英国脱欧公投事件严重削弱了政府公信力, 同

收稿日期: 2023-04-07

扫描二维码阅读全文:



基金项目: 河南省民办普通高等学校学科专业建设资助项目(教办政法[2020]162号)

作者简介: 赵梦凡(1997-), 女, 硕士, CCF会员, 郑州西亚斯学院计算机与软件工程学院教师, 研究方向为自然语言处理; 张钰涛(1998-), 男, 硕士, 郑州西亚斯学院计算机与软件工程学院教师, 研究方向为视听分离与定位; 赵铤钊(2003-), 男, 石家庄学院理学院学生, 研究方向为深度学习。

一时段的美国总统大选将假新闻一词带入全球视野,一度成为美国总统大选的代名词。Mejova等^[2]对64个国家的脸书广告进行了调查,发现5%的广告包含错误信息。许多在线新闻门户网站会故意发布误导性新闻,以提高其知名度。此外,经济也会遭受到假新闻的侵袭。例如,声称美国第44任总统巴拉克·奥巴马在白宫爆炸中受伤的假新闻使1300亿美元的股票价值凭空蒸发^[3]。2019年新冠肺炎疫情的爆发影响了世界各个国家,改变了全世界人民的生活、心理和经济状态,人们对这个话题十分感兴趣,因此通过各种平台对新冠病毒进行讨论。与此同时,关于新冠病毒各种各样的假新闻也层出不穷,一些与健康相关的假新闻向人们介绍了无效甚至有害的疗法,严重扰乱了社会秩序,如声称“酒精可以治愈COVID-19”的假新闻致使伊朗很多人住院甚至死亡^[4]。在讨论假新闻时,后真相这一概念往往同时出现。2016年,即英国脱欧和特朗普当选美国总统的同年,《牛津词典》将后真相选为年度国际词汇,其是指真相正在变得越来越不受重视或难以获得^[5]。伴随这一趋势,政治文化和公共讨论更加倾向于诉诸情感、偏见等非理性维度^[6]。为了最大程度地降低假新闻的危害,营造良好的网络环境,进行假新闻检测迫在眉睫。

假新闻检测是一门跨学科的课题,涉及机器学习、知识图谱、自然语言处理及信息检索等领域,并需要政治学、新闻学、心理学、经济学等知识的辅助。为有效遏制假新闻的传播,促进其早期检测及可解释性研究,本文首先明确假新闻的定义,深入分析其特征;其次将目前的假新闻检测方法分为4类,给出其评价指标和常用数据集,比较各方法的检测性能;最后对假新闻检测现存问题进行剖析,继而提出有前景的研究方向。

1 假新闻定义及特征

1.1 相关概念

假新闻检测面临的挑战始于对假新闻一词的界定,直到现在该词也没有通用定义。现有研究常将假新闻与欺骗性新闻(Deceptive News)^[7-9]、错误新闻(False News)^[10-11]、讽刺新闻(Satire News)^[12-14]、虚假信息(Disinformation)^[15]、错误信息(Misinformation)^[16]、选择性接受(Cherry-picking)^[17]、点击诱饵(Clickbait)^[12]和谣言(Rumor)^[18]等概念联系起来。其中,欺骗性新闻是故意且可证明为假的新闻;错误新闻是通过媒体发布和传播的包含一些非事实内容的新闻;讽刺新闻往往是结合热点时事的讽刺和夸张报道,其内容通常含有夸大其词的虚假成分;虚假信息是指被有意或无意地扭曲过的消息,或凭空捏造的消息,旨在欺骗和误导大众;点击诱饵通常是标题党,使用耸人听闻的话语吸引用户点击并进入某个网页,依靠用户点击量产生广告收入,即通过用户的活动数据获利;谣言带有强烈的个人感情色彩,当谣言被证实为假时就形成了假

新闻^[19]。

根据以上术语和概念的定义,可以从真实性、意图、危害程度3个角度对其进行辨析。其中真实性指是否包含任何非事实的陈述;意图指发布或转发假新闻的用户的目的,如误导或娱乐大众、获利等;危害程度是指对社会或大众造成负面影响的程度。表1对这些概念进行了总结。

Table 1 Terms and concepts related to fake news

表1 假新闻相关术语与概念

| 术语/概念 | 真实性 | 意图 | 危害程度 |
|-------|-------|-------------|---------------|
| 欺骗性新闻 | 非真实 | 误导大众 | 人为故意编写,影响较大 |
| 错误新闻 | 非真实 | 误传 | 有一定影响力 |
| 讽刺新闻 | 未统一 | 娱乐大众 | 影响程度较小 |
| 虚假信息 | 非真实 | 误导大众 | 恶意散布,影响社会和群众 |
| 错误信息 | 非真实 | 误传 | 有一定影响力 |
| 选择性接受 | 通常是事实 | 误导大众 | 用户偏颇吸收,影响用户个人 |
| 点击诱饵 | 通常非真实 | 通过用户点击,进行盈利 | 标题党,影响程度一般 |
| 谣言 | 非真实 | 吸引用户关注,误导大众 | 较大,影响群众身心健康 |

1.2 假新闻定义

文献[20]对假新闻这一概念施加最低程度的限制,不考虑其意图,认为广义上的假新闻就是错误新闻,同时提供了一个狭义的假新闻定义,即新闻机构故意发布的假消息,这个狭义的定义更加强调新闻的真实性和意图。随着社交网络的快速发展,多样化的沟通交流方式使得新闻不再拘泥于传统的纸质形式,也不再只由记者创建,各种演讲、文章、声明、帖子等都是广义上的新闻,每个用户都是新闻创作者。因此,假新闻一词的定义应关乎整个信息生态系统。此外,随着自媒体的发展,判定一条新闻是否为假不能简单依赖发布机构,而是根据内容的真实性,主要聚焦于互联网语境。因此,本文将假新闻定义为与事实内容不符的任何信息和言论,这一定义不考虑新闻发布者、来源及发布意图,涵盖了欺骗性新闻、错误新闻、虚假信息、错误信息、点击诱饵、谣言等概念,比较广义。

1.3 假新闻特征

假新闻具有独特的用户特征、内容特征和传播特征,通过研究这些特征可以构建更具针对性的假新闻检测模型,从而最大程度地降低其造成的危害。

1.3.1 用户特征

用户特征是指社交媒体上制造和传播新闻的用户本身所具有的个性化特点,如用户个人资料、评论或转发行为、政治偏见等。通常情况下,经常制造和分享假新闻的用户再次制造和转发的新闻大概率也是假的,因此对用户属性和行为特征进行分析能够提高假新闻检测准确率。

(1) 属性特征。属性特征包括用户的个人资料,如用户是否通过验证、粉丝数量、关注者数量、个人资料完善程度、账户注册时间、学历、工作、年龄等。通常情况下,学历

低的用户更容易制造和传播假新闻,年龄大的用户更容易受到假新闻的欺骗并进一步传播假新闻。此外,假新闻制造者通常使用新注册的账号,且大多是一次性账户^[21]。

(2) 行为特征。行为特征是指用户的评论、转发行为,以及在发布新闻后的表现,主要研究的是用户在传播假新闻后表现出的一些不同于正常用户的行为,如平均每日发文数量、转发时间间隔等^[20],可以根据用户的异常行为进行假新闻检测。

此外,政治偏见在塑造用户个人资料与影响其新闻消费选择方面起着重要作用。新闻社会学研究表明,党派偏见与新闻内容真实性之间存在相关性^[22],政治派别与人们的属性和行为相关^[23]。将用户的政治立场作为辅助特征有助于检测假新闻。

1.3.2 内容特征

(1) 情感特征。情感特征是从新闻内容中所提取的情感、语义等特征。Dickerson等^[24]研究表明与情绪相关的行为可以区分人类账户和社交机器人账户。真实的新闻立场是中立的,不掺杂个人感情色彩,假新闻通常是社交媒体用户就某一件事发表的个人看法,通常带有明显的情感极性词,有强烈的主观色彩。

(2) 文本特征。文本特征是从新闻文本中提取的句法、句法、语法、语义等特征。Rubin等^[25]分析了真假新闻的修辞结构、语篇成分及其连贯关系,然后应用向量空间模型,通过语篇特征相似度对新闻进行聚类。研究发现,假新闻的重点不在于陈述事实,通常带有很多感叹号、问号、形容词和第一人称代词,逻辑性不强,并且经常使用夸张的标题吸引大众关注;Brasoveanu等^[26]从新闻文本中提取情感、实体或事实等关系,通过分析语义特征和句法特征识别假新闻;Wang等^[27]首先从语义和文字两个方面捕获事件主题句(反映事实的句子),使用双相似度从这些事件主题句中生成一个简洁的事件摘要;然后针对句子定义4种细粒度差异,对应假新闻中常见的4种欺诈性方法,为每种差异设计定量计算方法,以衡量其差异程度;最后将差异分数转换为新闻中句子的权重,引入BERT预训练模型构建精确的新闻表示,以完成假新闻检测任务。

(3) 视听特征。视听特征是从新闻所包含的图片、音频、视频中提取的视觉和听觉等特征。传统的假新闻以文字为主。近年来,各种深度伪造技术层出不穷,换头术、真声模拟等手段使假新闻以音频和视频等动态方式呈现^[28],比静态的文字更具欺骗性。然而,由于音频、视频占用空间较大、特征提取困难,目前关于视频、音频等多模态的假新闻检测研究较少。Jing等^[29]提出一种用于多模态虚假信息检测的渐进融合网络,该网络在不同层次上捕获新闻文本和图片的特征表示,并通过混合器在同一层次与不同层次之间建立强联系,实现文本模态与视觉模态之间的融合。

1.3.3 传播特征

传播特征是指从新闻传播路径中提取的特征。通常假新闻比真新闻传播得更广泛,会产生复杂的级联效应。Friggeri等^[30]指出当谣言再次传播时,往往比上一次传播得更广泛,波及的人更多;Zhou等^[20]研究发现假新闻的传播是由少量活跃用户主导的,例如飓风桑迪期间,Twitter上转发的虚假图片中有90%是前30名用户转发的^[31]。此外,假新闻在发布早期传播速度最快,对社会和个人造成的影响也最大。Zubiaga等^[32]发现谣言的传播很大程度上发生在被揭穿之前,甚至在开始传播的前几分钟内就出现了爆发性转发。随着时间的推移,假新闻被核实后热度会逐渐下降,这也表明了实现假新闻早期检测的重要性。

2 假新闻检测方法

假新闻传播的时间越久,接触到假新闻的人越多,假新闻网会越织越密,对用户的影响也越深刻,从而使辟谣更加困难。因此,应尽早识别假新闻,最大程度地消除其带来的负面影响。目前,各国学者提出很多假新闻检测方法,总体可分为基于内容的方法、基于社交语境的方法、基于传播的方法及混合方法四大类。

2.1 基于内容的方法

基于内容的方法立足于新闻文本内容和视觉特征,通过分析语义和语言特点,以及所用图片分辨率等,利用现有事实来源挖掘假新闻的风格特征。假新闻通常带有明显的情绪极性词,且其文本内容通常具有很强的非正式性。一些类型的词语是机器学习分类器的明显特征,包括积极情感词,如love、nice;否定词,如no、not;认知动词,如ought、know;推断动词,如maybe、guess等^[33]。基于内容的方法又可分为基于知识的方法和基于风格的方法两类。

2.1.1 基于知识的方法

传统的基于知识的方法包括利用外部资源进行事实核查、面向专家的核查方法、面向众包的事实核查和面向计算的事实核查4种方式。利用外部资源进行事实核查的方法通过引入外部知识库检测假新闻,如百度、维基百科等外部知识库包含大量高质量的实体描述,可作为假新闻检测的依据;面向专家的核查方法由领域内的专家核查新闻真假;面向众包的事实核查聚合了普通人的注释,但由于普通人的政治倾向以及标注的相互冲突的标签,又进一步开发了面向计算的事实核查,该方法以外部资源为参考,由计算机执行。这些方法依靠大量人力资源,效率低下且结果不准确,已不适用于现今大体量的新闻核查。目前,基于知识的方法对新闻内容构建内部知识图和引入外部知识图,通过验证新闻内容中的知识与事实是否一致进行假新闻检测。

(1) 构建内部知识图。对新闻内容中的实体及其之间的关系构建三元组进行假新闻检测被证明是行之有效的

的方法,通常可以用(主语—谓语—宾语)的格式表示,即SPO三元组^[34]。Han等^[35]不引入任何外部知识图,仅从每条新闻中提取实体及其关系,构建了一个单独的知识图,其中一条新闻由一个子图表示,然后利用所有子图及其对应的新闻真实性标签来训练图神经网络模型,这将假新闻检测问题转化为子图分类问题,该研究为后续基于图模型的假新闻检测提供了思路;Shu等^[36]探讨了出版商偏见、新闻立场与相关用户参与度的相关性,提出一个三元关系假新闻检测框架,并在两个真实世界假新闻数据集上验证了该框架的有效性;Wu等^[37]提出一种知识图增强框架,结构如图1所示。首先从没有外部知识的语料库中提取实体关系元组,构建了基于证书的多关系知识图;然后应用复合图卷积网络从知识图中同时预训练实体和关系的嵌入,并设计框架采用结构化知识增强语义,以实现假新闻检测。基于内部知识图的方法从新闻本身提取知识进行建模,不需要与外部知识库交互,时间开销相对较小,但对新闻要求较高。若新闻中缺少实体或关系,构建的三元组或知识图将会有缺失,从而影响模型检测性能。

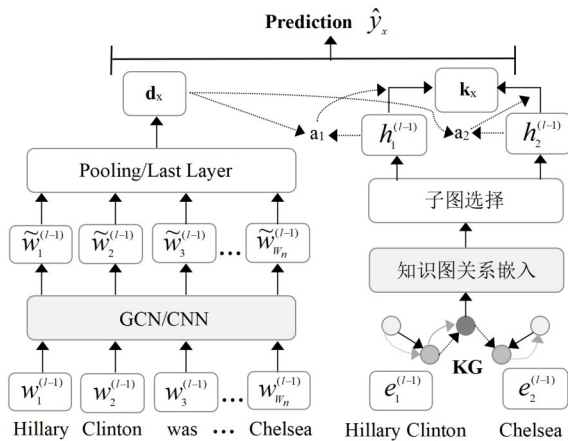


Fig. 1 Knowledge graph enhanced framework
图1 知识图增强框架

(2) 引入外部知识图。新闻内容中的知识非常有限,为了提高检测准确率,通常需要引入外部知识图^[34]。外部知识库,如百度、维基百科等,均包含大量高质量的结构化主谓宾三元组和非结构化的实体描述,可以作为假新闻检测的证据。对新闻句子构建结构化主谓宾三元组,为了充分利用外部知识库,可将实体作为连接新闻与外部知识库的桥梁,在外部知识库中对句子中的实体进行搜索核查,通过比较新闻中的实体与外部知识库中的实体判断新闻真假。Hu等^[38]为包含主题和实体的每个新闻构建一个定向异质文档图,基于该图开发了一个异质图注意力网络,用于学习主题丰富的新闻表示并对新闻内容的语义进行编码的上下文实体表示;然后通过精心设计的实体网络比较上下文实体表示与相应的基于知识库的实体表示,以捕获新闻之间的一致性。然而,引入外部知识图的方法较为复杂,由于需要时刻与外部知识库交互,对硬件环境要求较高,模型的时间复杂度是需要考量的重要因素。

2.1.2 基于风格的方法

基于风格的方法可以捕获新闻的写作风格,利用情感分析技术进行假新闻检测,常用模型包括支持向量机(Support Vector Machine, SVM)^[39]、随机森林(Random Forest, RF)^[40]等。假新闻的发布者通常是为了短期经济或政治目标^[7],而不是为了与读者建立长期稳定的关系,因此他们更喜欢使用非正式、耸人听闻、情绪化的词语^[41],这些可信度低的指标可以用来建立一个可靠的假新闻分类器。Przybyla^[42]专注于新闻写作风格的研究,即研究文本的形式而非文本的意义,通过设计基于神经网络和风格特征的检测模型实现了假新闻检测,实验结果表明该模型更加关注煽情和情感词汇,这正是假新闻的重要特征;李小艳^[43]提出一种基于多维度行文风格特征的假新闻检测模型,结构如图2所示。该模型从单词、句子、篇章3个维度提取新闻风格特征,并使用Text-CNN(Text-Convolutional Neural Network)模型提取语义信息,同时加入注意力机制对各个维度的行文风格特征加权,以此表示每种行文风格特征对假新闻检测的重要程度;Choudhary等^[44]提出一种语言模型找出产生语言驱动特征的内容属性,同时提取特定新闻的句法、语法、情感和可读性特征,使用基于神经网络的序列学习模型实现假新闻检测,平均准确率达到86%;Zhou等^[45]提出分层递归神经网络模型,将新闻文档构建为层次语言树,与神经网络相结合,捕获具有语言意义的局部顺序,并通过全局递归结构学习和分类新闻文档,通过语言风格识别假新闻,这是第一个提供层次语言树和神经网络保存树信息的研究。

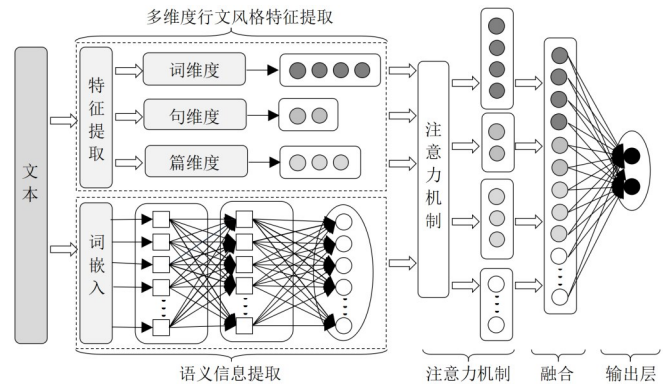


Fig. 2 Fake news detection model based on multi-dimensional style features
图2 基于多维度行文风格特征的假新闻检测模型

2.2 基于社交语境的方法

社交语境是指用户之间的互动,包括推文、转发、评论、回复、提及和关注。Vaibhav等^[46]研究表明,可信新闻和假新闻具有不同的句子交互模式,为识别社交媒体上的假新闻提供了重要信息;Jin等^[47]建立了一个姿态网络,其中边的权重表示每对帖子相互支持或否认的程度,估计与每条新闻相关的所有帖子的可信度可以将假新闻检测问题形式化为图优化问题;Yang等^[48]提出一种无监督方法,通过建立贝叶斯图模型捕获新闻有效性、用户意见与用户

可信度之间的生成过程;Mehta等^[49]认为假新闻检测是在一个图框架中推理新闻来源、用户发布的文章与用户之间的关系,在嵌入这些信息后,通过制定推理算子揭示元素之间未观察到的交互来增强图边,例如文档内容之间的相似性和用户参与模式;与将新闻信息建模为图结构的研究不同,Shu等^[50]认为新闻传播过程中的社会语境形成了发布者、新闻、用户三者之间的内在关系,这一关系能改善假新闻检测效果。基于该认知,其提出三元关系嵌入框架,该框架可以同时对接假新闻发布者—新闻关系和用户—新闻交互进行建模;Jain等^[51]通过ELMo(Embeddings from Language Models)模型捕获社交语境特征,从而获得新闻句子的语义信息,然后将上下文嵌入与单词嵌入相结合,利用注意力机制和相关元数据在LIAR数据集上实现了假新闻检测。然而,该研究未能利用新闻的多模态上下文信息,如图像,导致模型忽略了新闻的完整语义表达。为解决这一问题,Qian等^[52]提出分层多模态语境网络,联合多模态语境信息和文本语义建立了统一的深度模型,用于假新闻检测。具体来说,该模型首先使用BERT和ResNet分别学习文本和图像的特征表示;然后将获得的图像和文本特征输入到多模态语境注意网络中,融合模态间与模态内的关系;最后设计分层编码网络捕获丰富的分层语义,用于假新闻检测。该研究为利用多模态信息进行假新闻检测提供了参考。

2.3 基于传播的方法

在社交媒体上,假新闻的传播模式表现出不同于真实新闻的鲜明特征,通过构建假新闻传播图捕捉其传播风格可以实现假新闻检测^[52]。Shu等^[53]从新闻传播结构、时间和语义的角度对假新闻和真实新闻的传播网络特征进行了比较分析;Zhou等^[54]提出传播网络模式驱动假新闻检测方法,该方法基于社会心理学理论分析被传播新闻与新闻传播者之间的关系特征,利用提取出的特征模式检测假新闻;Wu等^[55]使用混合SVM捕获新闻高阶传播模式以及主题和情感等语义特征。然而,从内容中识别有用的特征具有挑战性,因为故意传播假新闻的人可能会操纵内容,使其看起来像真新闻。为解决这一问题,Wu等^[56]将长短期记忆(Long Short Term Memory, LSTM)细胞纳入循环神经网络(Recurrent Neural Network, RNN)模型,以新闻传播结构推断用户的特征嵌入,该模型可以在没有新闻内容信息的情况下实现高质量的假新闻检测;Liu等^[57]将新闻的传播路径建模为多元时间序列,构建了一个包含RNN和CNN的时间序列分类器,分别用于捕捉用户特征在传播路径上的全局和局部变化,实现了假新闻的早期检测。

深度学习方法在提取欧式空间中的数据特征方面取得了巨大成功,然而许多实际应用场景中的数据是从非欧式空间中生成的,而深度学习在处理非欧式空间数据方面的表现难以令人满意。图嵌入可以保留图拓扑结构和图节点内容信息,因此图神经网络为假新闻检测带来新的契

机。得益于图的非欧式特性,假新闻的检测准确性达到了一个新的高度。例如,Inan^[58]构建用户友谊网络和转发传播网络,并利用图注意力机制过滤潜在的假用户,同时构建新闻内容图,其中包括评论、潜在假用户个人资料描述及其共享的新闻内容。该研究利用图神经网络将新闻文本内容嵌入到传播结构中,为假新闻检测研究提供了新的思路;Silva等^[59]将新闻传播路径建模为图结构,并为传播网络中的节点和级联分配了不同程度的重要性;其还根据部分传播网络在早期检测截止日期前重建了完整的传播网络,实现了假新闻的早期检测。

2.4 混合方法

已有研究表明,利用多个特征进行假新闻检测比仅使用一个特征检测准确率高,多种检测技术结合进行假新闻检测比只利用一种技术检测效果更好。一些混合方法基于新闻内容与相关用户的交互行为进行假新闻检测。例如,Zhang等^[60]使用预先提取的单词集从新闻内容、用户个人资料和新闻主题描述中构造显式特征,同时使用RNN学习潜在特征,如新闻内容信息的不一致性和用户个人资料的潜在模式,获得这些特征后建立一个深度扩散网络来学习新闻、创作者和主题的特征,在真实世界的假新闻数据集PolitiFact上证实了该模型的有效性。还有一些方法结合不同类型的特征进行假新闻检测。例如,Castillo等^[61]提出基于消息、用户和主题的谣言检测模型,Ma等^[62]提出基于谣言生命周期的时间序列来捕捉社会背景信息在谣言传播过程中随时间变化的特征,并应用时间序列建模技术整合各种社会背景信息,该方法在谣言首次传播后的早期阶段显示出强大的检测性能。在此基础上,Ma等^[63]进一步融入新闻传播特征,提出一种具有树结构的核函数SVM分类器,通过计算传播拓扑树相似性识别谣言;Ruchansky等^[64]开发了一个三模块的假新闻检测结构,第一个模块以用户响应、新闻内容和用户特征为输入,并训练RNN以捕获文本的时间表示;第二个模块利用用户特征生成每个用户的得分和低维表示;第三个模块取前两个模块的输出训练一个神经网络来标记新闻项目。然而,从新闻文本内容、用户信息和传播模式中手动提取特征既费时又费力,并且这些特征严重依赖于数据集。此外,只利用单一方法或技术很难从不同信息中提取可区别真假新闻的多种特征,因此一些学者开始利用混合技术提取新闻的多种特征。例如,Lu等^[65]构建了图感知协同注意网络GCAN提取用户特征、源新闻中包含的单词嵌入特征,同时开发了双重共同注意机制了解源推文与转发推文之间的相关性,以及源推文与用户交互之间的共同影响;Huang等^[66]基于文本内容和谣言传播结构对推文、用户、新闻中的单词构建异质图,并使用图注意力网络从源推文内容及其与用户的交互中学习新闻特征表示。该模型具有较好的假新闻检测性能,为构建推文和用户关系图提供了思路,在Twitter数据集上的实验证明了该方法的优越性。在

此基础上,Ran等^[67]提出一个端到端的多通道图注意力网络,并行构建3个子图,同时开发了事件共享模块,通过学习事件不变特征增强模型的泛化能力,进一步提高了假新闻检测性能。然而,构建3个子图会引入更多噪声,从而影响模型性能。为解决该问题,Han等^[68]将新闻文本、用户传播网络并行构建为两个子图,并利用图卷积网络深度挖掘新闻文本特征,同时结合传播特征实现了高质量的假新闻检测。以上模型均为静态信息模型,通过使用特征挖掘算法进行假新闻检测。Go等^[69]提出一种新的动态异质

图神经网络模型,利用BERT获取新闻内容和作者表示,构建了异质新闻—作者图来反映上下文信息与关系,在LIAR数据集上的准确率达到72.26%,实现了在该数据集上的最优性能,为后续假新闻检测提供了新思路。

3 模型性能评价

表2总结了各类假新闻检测模型的基本特征、使用的数据集及性能评价结果。

Table 2 Summary of fake news detection models

表2 假新闻检测模型总结

| 类别 | 模型 | 所用特征 | 使用技术 | 数据集 | 分类类别 | 模型性能/% |
|-----------|-----------------------------------|----------------------|-----------------------------|------------------------------------|----------------------|-------------------------------------------------------------------|
| 基于内容的方法 | KGf ^[37] | 内部知识 | 图卷积网络、 卷积神经网络 | PolitiFact、 GossipCop | 二分类 | Acc: 86.0, F1: 85.3 Acc: 73.3, F1: 72.3 |
| | CompareNet ^[38] | 外部知识 | 注意力机制 | SLN、 LUN | 二分类、 四分类 | F1: 89.2 F1: 69.1 |
| | Word2vec-Text-CNN ^[43] | 语言风格 | 卷积神经网络、 注意力机制 | Weibo | 二分类 | F1: 86.95 |
| | TriFN ^[50] | 社交语境 | 关系嵌入 | PolitiFact、 BuzzFeed | 二分类 | Acc: 87.8, F1: 88.0 Acc: 86.4, F1: 87.0 |
| 基于社交语境的方法 | AENeT ^[51] | 文本内容 元数据 | ELMo、 GloVe、 注意力机制 | LIAR | 六分类 | Acc: 46.36 |
| | HMCAN ^[52] | 社交语境 | 注意力机制 | Weibo、 Twitter | 二分类 | Acc: 88.5 Acc: 89.7 |
| | HPFN ^[53] | 传播网络 | 语言模型 | PolitiFact、 GossiCop | 二分类 | Acc: 84.3, F1: 84.3 Acc: 86.1, F1: 86.2 |
| 基于传播的方法 | PPC ^[57] | 传播网络 | 循环神经网络、 卷积神经网络 | Weibo、 Twitter15、 Twitter16、 | 二分类、 四分类、 四分类、 | Acc: 92.1 Acc: 84.2 Acc: 86.3 |
| | ZoKa ^[58] | 传播网络 | 图注意力 | PolitiFact、 GossiCop | 二分类 | Acc: 87.4, F1: 88.1 Acc: 80.2, F1: 76.0 |
| | Propagation2Vec ^[59] | 传播网络 | 注意力机制 | PolitiFact、 GossiCop | 二分类 | Acc: 89.7, F1: 89.3 Acc: 89.2, F1: 87.4 |
| | CSI ^[64] | 文本内容 用户行为 源用户 | 循环神经网络 | Twitter、 Weibo | 二分类 | Acc: 89.2, F1: 89.4 Acc: 95.3, F1: 95.4 |
| 混合方法 | GCAN ^[65] | 源推文 传播网络 | 卷积神经网络、 注意力机制、 门控循环单元 | Twitter15、 Twitter16 | 二分类 | Acc: 87.7, F1: 82.5 Acc: 90.8, F1: 79.3 |
| | HGATRD ^[66] | 文本内容 传播网络 | 图注意力 | Twitter15、 Twitter16 | 四分类 | Acc: 91.1 Acc: 92.4 |
| | MGAT-ESM ^[67] | 文本内容 传播网络 用户关系 | 图注意力、 事件不变特征 | Twitter15、 Twitter16 | 四分类 | Acc: 90.8 Acc: 91.6 |
| | MSLG ^[68] | 文本内容 传播网络 | 图注意力、 图卷积网络 | Twitter15、 Twitter16 | 四分类 | Acc: 92.0 Acc: 91.3 |
| | DHGNN ^[69] | 文本内容 新闻作者 | BERT 异质图 | LIAR、 PolitiFact、 GossipCop、 | 六分类、 二分类、 二分类 | Acc: 72.3, F1: 83.8 Acc: 76.3, F1: 86.5 Acc: 79.8, F1: 87.3 |

目前常用的评价指标包括精确率(Precision)、召回率(Recall)、F1分数(F1-score)和准确率(Accuracy)。其中,精确率也被称为查准率,是正确预测为假新闻的样例的比

例,计算方法如式(1)所示;召回率也被称为查全率,是查找到的所有假新闻的比例,用于衡量分类器的敏感性,计算方法如式(2)所示;F1分数为基于查准率与查全率的调

和平均,用于综合考量查准率和查全率的性能,计算方法如式(3)所示;准确率为模型预测正确样例的比例,计算方法如式(4)所示。假新闻分类结果包括真阳性、假阳性、真阴性、假阴性 4 类,具体含义为:①真阳性(True Positive, TP)。指预测为假新闻,数据集中实际标注也为假;②假阳性(False Positive, FP)。指预测为假新闻,数据集中实际标注为真;③真阴性(True Negative, TN)。指预测为真新闻,数据集中实际标注也为真;④假阴性(False Negative, FN)指。指预测为真新闻,数据集中实际标注为假。

$$Precision = \frac{|TP|}{|TP| + |FP|} \quad (1)$$

$$Recall = \frac{|TP|}{|TP| + |FN|} \quad (2)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

从表 2 中可以看出,已有的假新闻检测工作大多为二分类和四分类任务,对于六分类任务研究不足,这主要受数据集的限制。目前公开的假新闻数据集大多是二分类任务数据集,只有 LIAR 数据集中标注的是六分类标签,从而导致在六分类任务上的模型较少。在 4 类方法中,混合方法具有较高且稳定的准确率和 F1 值,这是由于其利用

多特征或多技术,能提取到新闻不同方面的特征,这些特征具有互补性。基于社交语境的方法依靠新闻评论、用户社交等上下文信息实现假新闻检测,基于传播的方法通过挖掘新闻的传播结构信息实现假新闻检测,这两类方法也能实现较高的检测准确率。基于内容的方法重点挖掘新闻文本特征,其中引入外部知识的方法性能最佳,这是由于外部知识对新闻内容起到了补充作用。然而,考虑到模型时间效率以及设备等问题,引入外部知识的方法应用并不广泛。

通过比较不同分类任务的模型性能可以发现,大多数二分类和四分类任务的检测准确率远高于六分类任务,只有基于异质图的 DHGNN 模型在 LIAR 数据集上实现了最优值,准确率达到 72.3%。尽管 AENeT 模型使用了注意力机制和提取情感特征的 ELMo 模型,但其准确率仅为 46.36%,这是由于 LIAR 数据集中只包含新闻文本内容和用户元属性信息,数据特征较少,这也是限制六分类假新闻模型检测准确率提高的一大原因。

表 3 列出了目前假新闻检测的一些常用公开数据集。可以看出,所有数据集涵盖的信息都非常有限,不能同时包含新闻具有的所有特征,这也是目前没有模型利用到新闻全部特征的重要原因。

Table 3 Common public datasets for fake news detection

表 3 假新闻检测常用公开数据集

| 数据集 | 时间 | 来源/平台 | 包含信息 | 特征 | | | | 数据集获取链接 | |
|-------------------------|------|---------------------|-------------------------------------------|---------------|----|---------------|----|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | | | 新闻内容 文本 视觉 | 用户 | 社会背景 网络 传播 | | | |
| COVID-19 | 2021 | Facebook Twitter | 10 700 条新闻 | √ | -- | -- | -- | -- | https://github.com/parthpatwa/covid19-fake-news-detection |
| CHECKED | 2021 | 微博 | 2 104 条新闻,1 868 175 条转发, 1 185 702 条评论 | √ | √ | √ | -- | √ | https://github.com/cyang03/CHECKED |
| Weibo-COV | 2020 | 微博 | 3 亿条推文 | √ | -- | √ | √ | -- | https://github.com/nghuyong/weibo-public-opinion-datasets |
| FakeCovid | 2020 | Snopes Poynter | 5 182 个新闻文章 | √ | -- | -- | -- | -- | https://gautamshahi.github.io/FakeCovid/ |
| FakeNewsNet | 2018 | Twitter | 201 921 篇文章 | √ | √ | √ | √ | -- | https://github.com/KaiDMMML/FakeNewsNet |
| BuzzFace | 2018 | Facebook | 2 263 个新闻文章,160 万条评论 | √ | -- | -- | -- | √ | https://github.com/gsantia/BuzzFace |
| LIAR | 2017 | PolitiFact | 12 836 条短新闻 | √ | -- | √ | -- | -- | https://www.cs.ucsb.edu/~william/data/liar_dataset.zip |
| FacebookHoax | 2017 | Facebook | 15 500 条新闻 | √ | -- | √ | -- | √ | https://github.com/gabll/some-like-it-hoax |
| Twitter15、 Twitter16 | 2017 | Twitter | 1 490 条推文 818 条推文 | √ | -- | √ | -- | √ | https://www.dropbox.com/s/7ewzdrbelpmrxu/rumdetect2017.zip? dl=0 |
| SLN | 2016 | Onion Beaverton | 360 条新闻 | √ | -- | -- | -- | -- | http://victoriarubin.fims.uwo.ca/news-verification/ |
| Weibo | 2016 | 微博 | 2 313 条谣言,2 351 条非谣言 | √ | -- | √ | -- | √ | http://alt.qcri.org/~wgao/data/rumdetect.zip |
| BuzzFeedNews | 2016 | Facebook | 1 627 篇文章 | √ | -- | -- | -- | -- | https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/tree/master/data |
| CREDBANK | 2015 | Twitter | 6 000 万条推文 | √ | -- | √ | -- | -- | https://compsocial.github.io/CREDBANK-data/ |

4 现存问题及未来研究方向

4.1 多模态数据特征获取

数据是一切研究的基础。现有假新闻数据集包含的信息单一,没有涵盖新闻具有的所有特征,如内容、用户信

息、社交语境信息、传播信息等,这也是导致假新闻细粒度检测分类任务准确率不高的重要原因。此外,目前公开数据集中大多只包含文本数据,不包含图片和视频数据。尽管已有 CHECKED^[70]、FakeNewsNet^[71]等多模态数据集,但 CHECKED 为中文数据集,其中的类别分布不平衡;Fake-newsNet 为英文数据集,分为 PolitiFact 和 GossipCop 两个子

数据集,分别只有783和18 417条包含图片的新闻,数据量并不充足,且未包含传播信息,因此迫切需要开发新的数据集进行多模态假新闻检测。

4.2 假新闻细粒度分类

现有研究大多对新闻进行二分类,即将其简单地分类为真或假。目前,深度造假技术使得很多新闻成为真假陈述的混合体,如某个公众人物曾经进行了某个公开演讲,有些新闻将其引用到新发生的事件中,从而造成新闻失实,这种“移花接木”的手段对假新闻的产生和传播起到了推波助澜的作用。这些现象促使假新闻检测从二分类任务向细粒度分类任务转变。已有成熟的二分类检测技术在细粒度分类任务中的效果往往不好,需要构建新的细粒度分类模型。然而,细粒度分类任务具有很大的挑战性,因为随着类别的细化,各类别之间的界限变得越来越模糊,需要精确的多类判别函数和特征提取分析技术。后续可以尝试使用概率图^[72]、事件摘要^[27]等技术精准挖掘新闻特征,在分类时选择概率最大的类进行输出,从而实现假新闻的细粒度分类。

4.3 假新闻检测可解释性

近年来,假新闻检测已经达到了很高的准确率。然而,现有研究的关键缺失部分为假新闻的可解释性,即某条新闻为什么被检测为假。为解决这一问题,可以利用注意力机制对新闻句子中的可疑单词、新闻主题、信用度低的用户等进行重点标记,从而实现假新闻检测的可解释性。Potthast等^[73]认为基于风格进行假新闻检测几乎不起作用,或者永远不可能实现,但Przybyla^[42]通过捕捉假新闻的风格对新闻中低可信度的片段进行高亮显示,实现了假新闻的可解释性,证明了这是一个值得研究的方向。Shu等^[74]使用注意力机制捕获可疑的新闻句子和用户评论,具体来说,模型给出判断该新闻为假的评论信息,相比那些干扰和不相关的评论,模型给予可解释的评论更高权重,从而选择更多相关评论来帮助检测假新闻。这一研究在真正意义上实现了假新闻检测的可解释性,然而,受限于数据集所包含的信息,对于不包含评论信息的数据集则需要从新闻本身实现可解释性,即通过新闻中可疑的词语判断新闻为假,如情绪极性词。

4.4 假新闻的早期检测

假新闻在社交媒体上发布的时间越长,其影响就越深远,也越难以消除。因此,实现假新闻的早期检测具有重要意义,这样可以最大限度地减少给社会和大众带来的危害。目前假新闻检测大多发生在新闻发布很久之后,甚至是在社交网络上传播广泛之后,这样便错过了降低假新闻危害的最好时机。当假新闻首次发布时几乎没有任何转发或评论,此时只能充分利用源新闻的文本内容进行检测,这就要求所构建的模型能够充分提取新闻文本特征。新闻创作者有助于检测假新闻,因为经常发布和转发不实消息的用户具有很低的信用度,可以认为其发布的新闻大

多为虚假消息。学历低、年龄大的用户通常更容易制造和传播假新闻。此外,假新闻还与主题联系密切,如健康主题的新闻倾向于虚假,而经济主题的新闻则倾向于真实。因此,通过引入新闻创作者和用户的背景信息,同时结合新闻主题信息构建全面的新闻属性信息能够实现假新闻的早期检测。例如,Sun等^[75]开创性地设计了基于超图的全局交互学习模型,从用户的扩散行为中捕获其全局偏好,并引入节点中心性编码补充用户在超图学习中的影响,为验证新闻的真实性提供了额外信号。该研究实现了假新闻的早期检测,为后续研究提供了新思路。

5 结语

互联网的快速发展使得用户越来越倾向于从社交媒体获取新闻资讯,这也带来了假新闻的泛滥。如何快速检测假新闻、及时遏制其传播成为不可忽视的重要问题。本文首先明确了假新闻的相关概念,对其特征进行了详细分析;然后对目前已有的假新闻检测方法进行了分类,并对各类中具有代表性的研究方法进行了阐述,基于此介绍了模型评价指标,比较了各模型性能,并汇总了当前公开的假新闻检测数据集;最后,对假新闻检测领域存在的问题进行了剖析,并对未来研究方向进行了展望。然而,假新闻检测这一领域涉及的内容众多,本文仅对该领域中最基本的内容和已经成熟的模型进行了综述,关于超图结构在假新闻检测上的最新研究并未进行阐述。当前假新闻检测这一领域受到越来越多的关注,也在不断激发研究者构建新的模型,以应对新形式的假新闻。在后续研究中可以构建超图以更有效地建模新闻中包含的信息,并融入新闻主题、社会背景信息、用户元属性等内容。然而,融入的信息越多,提取特征时引入的噪声信息也越多,在进行特征融合时需选择有效的方法降低噪声的影响,这些都是今后值得研究的方向。

参考文献:

- [1] DANG L X. Fake news detection system in social networks [D]. Chengdu: University of Electronic Science and Technology of China, 2021. 党留新. 社交网络中的假新闻检测系统[D]. 成都:电子科技大学, 2021.
- [2] MEJOVA Y, WEBER I, FERNANDEZ-LUQUE L. Online health monitoring using Facebook advertisement audience estimates in the United States: evaluation study[J]. *JMIR Public Health and Surveillance*, 2018, 4(1): e7217.
- [3] RAPOZA K. Can “fake news” impact the stock market?[EB/OL]. www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/.
- [4] KARIMI N, GAMBRELL J. Hundreds die of poisoning in Iran as fake news suggests methanol cure for virus[EB/OL]. <https://www.timesofisrael.com/hundreds-die-of-poisoning-in-iran-as-fake-news-suggests-methanol-cure-for-virus/>.
- [5] WANG A B. “Post-truth” named 2016 word of the year by Oxford Diction-

- aries[EB/OL]. <https://www.cbsnews.com/news/post-truth-word-of-the-year-2016-oxford-dictionaries/>.
- [6] ZUO Y L. Fake news: what is it? why? what to do? [J]. *Chinese and Foreign Law*, 2021, 33(2): 544-559.
左亦鲁. 假新闻: 是什么? 为什么? 怎么办?[J]. *中外法学*, 2021, 33(2): 544-559.
- [7] ALLCOTT H, GENTZKOW M. Social media and fake news in the 2016 election[J]. *Journal of Economic Perspectives*, 2017, 31(2): 211-236.
- [8] LAZER D M J, BAUM M A, BENKLER Y, et al. The science of fake news[J]. *Science*, 2018, 359(6380): 1094-1096.
- [9] SHU K, SLIVA A, WANG S, et al. Fake news detection on social media: a data mining perspective [J]. *ACM SIGKDD Explorations Newsletter*, 2017, 19(1): 22-36.
- [10] VOSOUGHI S, ROY D, ARAL S. The spread of true and false news online[J]. *Science*, 2018, 359(6380): 1146-1151.
- [11] ADAM K. Post-truth: study epidemiology of fake news [J]. *Nature*, 2016, 540(7634): 525.
- [12] CHEN Y, CONROY N J, RUBIN V L. Misleading online content: recognizing clickbait as "false news" [C]//*Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, 2015: 15-19.
- [13] TANDOC J E, LIM Z W, LING R. Defining "fake news": a typology of scholarly definitions[J]. *Digital Journalism*, 2018, 6(2): 137-153.
- [14] RUBIN V L, CHEN Y, CONROY N K. Deception detection for news: three types of fakes[C]//*Proceedings of the 78th Association for Information Science and Technology*, 2015: 1-4.
- [15] KSHETRI N, VOAS J. The economics of "fake news" [J]. *IT Professional*, 2017, 19(6): 8-12.
- [16] CLAIRE W. newsFake. it's complicated[EB/OL]. <https://firstdraftnews.org/latest/fake-news-complicated/>.
- [17] ASUDEH A, JAGADISH H V, WU Y, et al. On detecting cherry-picked trendlines[J]. *Proceedings of the VLDB Endowment*, 2020, 13(6): 939-952.
- [18] ZUBIAGA A, AKER A, BONTCHEVA K, et al. Detection and resolution of rumours in social media: a survey[J]. *ACM Computing Surveys*, 2018, 51(2): 1-36.
- [19] WANG J, WANG Y C, HUANG M J. False information in social networks: definition, detection and control[J]. *Computer Science*, 2021, 48(8): 263-277.
王剑,王玉翠,黄梦杰. 社交网络中的虚假信息: 定义、检测及控制[J]. *计算机科学*, 2021, 48(8): 263-277.
- [20] ZHOU X Y, ZAFARANI R. A survey of fake news: fundamental theories, detection methods, and opportunities [J]. *ACM Computing Surveys*, 2020, 53(5): 1-40.
- [21] LIU Y Z, WANG J, PAN X Z, et al. Research on scale-free network rumor propagation under node influence [J]. *Small Microcomputer System*, 2018, 39(11): 2375-2379.
- [22] GENTZKOW M, SHAPIRO J M, STONE D F. Media bias in the marketplace: theory[M]. : Amsterdam: North-Holland, 2015: 623-645.
- [23] SHU K, ZHOU X, WANG S, et al. The role of user profiles for fake news detection [C]//*Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2019: 436-439.
- [24] DICKERSON J P, KAGAN V, SUBRAHMANIAN V S. Using sentiment to detect bots on Twitter: are humans more opinionated than bots? [C]//*IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2014: 620-627.
- [25] RUBIN V L, CONROY N J, CHEN Y. Towards news verification: deception detection methods for news discourse [C]//*Proceedings of the Hawaii International Conference on System Sciences*, 2015: 5-8.
- [26] BRASOVEANU A M P, ANDONIE R. Semantic fake news detection: a machine learning perspective [C]//*Proceedings of the International Workshop on Artificial Neural Networks*, 2019: 656-667.
- [27] WANG J, GAO M, HUANG Y, et al. FinD: fine-grained discrepancy-based fake news detection enhanced by event abstract generation [J]. *Computer Speech & Language*, 2023, 78: 101461.
- [28] CHESNEY B, CITRON D. Deep fakes: a looming challenge for privacy, democracy, and national security[J]. *California Law Review*, 2019, 107(6): 1753-1820.
- [29] JING J, WU H, SUN J, et al. Multimodal fake news detection via progressive fusion networks [J]. *Information Processing & Management*, 2023, 60(1): 103120.
- [30] FRIGGERI A, ADAMIC L, ECKLES D, et al. Rumor cascades [C]//*Proceedings of the International AAAI Conference on Web and Social Media*, 2014: 101-110.
- [31] GUPTA A, LAMBA H, KUMARAGURU P, et al. Faking sandy: characterizing and identifying fake images on Twitter during hurricane sandy [C]//*Proceedings of the 22nd International Conference on World Wide Web*, 2013: 729-736.
- [32] ZUBIAGA A, LIAKATA M, PROCTER R, et al. Analysing how people orient to and spread rumors in social media by looking at conversational threads[J]. *PLoS One*, 2016, 11(3): e0150989.
- [33] QAZVINIAN V, ROSENGREN E, RADEV D, et al. Rumor has it: identifying misinformation in microblogs [C]//*Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 2011: 1589-1599.
- [34] LASSILA O, SWICK R. Resource description framework (RDF) model and syntax specification [EB/OL]. <https://www.w3.org/TR/PR-rdf-syntax/>
- [35] HAN Y, SILVA A, LUO L, et al. Knowledge enhanced multi-modal fake news detection [DB/OL]. <https://arxiv.org/abs/2108.04418>.
- [36] SHU K, WANG S, LIU H. Exploiting tri-relationship for fake news detection [DB/OL]. <https://arxiv.org/abs/1712.07709v1>.
- [37] WU K, YUAN X, NING Y. Incorporating relational knowledge in explainable fake news detection [C]//*Proceedings of the 25th Advances in Knowledge Discovery and Data Mining*, 2021: 403-415.
- [38] HU L, YANG T, ZHANG L, et al. Compare to the knowledge: graph neural fake news detection with external knowledge [C]//*Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 2021: 754-763.
- [39] FENG S, BANERJEE R, CHOI Y. Syntactic stylometry for deception detection [C]//*Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, 2012: 171-175.
- [40] ZHOU X Y, JAIN A, PHOHA V, et al. Fake news early detection: a theory-driven model [J]. *Digital Threats: Research and Practice*, 2020, 1(2): 1-25.
- [41] BAKIR V, MCSTAY A. Fake news and the economy of emotions: problems, causes, solutions [J]. *Digital Journalism*, 2018, 6(2): 154-175.
- [42] PRZYBYLA P. Capturing the style of fake news [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2020: 490-497.
- [43] LI X Y. Research on fake news detection based on mixed features [D]. Guizhou: Guizhou University, 2022.

- 李小艳. 基于混合特征的假新闻检测方法研究[D]. 贵州: 贵州大学, 2022.
- [44] CHOUDHARY A, ARORA A. Linguistic feature based learning model for fake news detection and classification[J]. *Expert Systems with Applications*, 2021, 169: 114171.
- [45] ZHOU X Y, LI J Y, LI Q Z, et al. Linguistic-style-aware neural networks for fake news detection[DB/OL]. <https://arxiv.org/abs/2301.02792>.
- [46] VAIBHAV V, ANNASAMY R M, HOVY E. Do sentence interactions matter? leveraging sentence level representations for fake news classification[C]//*Proceedings of the 13th Workshop on Graph-Based Methods for Natural Language Processing*, 2019: 134-139.
- [47] JIN Z W, CAO J, ZHANG Y D, et al. News verification by exploiting conflicting social viewpoints in microblogs [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2016: 2972-2978.
- [48] YANG S, SHU K, WANG S H, et al. Unsupervised fake news detection on social media: a generative approach [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2019: 5644-5651.
- [49] MEHTA N, PACHECO M L, GOLDWASSER D. Tackling fake news detection by continually improving social context representations using graph neural networks [C]//*Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022: 1363-1380.
- [50] SHU K, WANG S H, LIU H. Beyond news contents: the role of social context for fake news detection [C]//*Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 2019: 312-320.
- [51] JAIN V, KALIYER R K, GOSWAMI A, et al. AENeT: an attention-enabled neural architecture for fake news detection using contextual features [J]. *Neural Computing and Applications*, 2022, 34(1): 771-782.
- [52] QIAN S S, WANG J G, HU J, et al. Hierarchical multi-modal contextual attention network for fake news detection[C]//*Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021: 153-162.
- [53] SHU K, MAHUESWARAN D, WANG S H, et al. Hierarchical propagation networks for fake news detection: investigation and exploitation [C]//*Proceedings of the 14th International AAAI Conference on Web and Social Media*, 2020: 626-637.
- [54] ZHOU X Y, ZAFARANI R. Network-based fake news detection: a pattern-driven approach [J]. *ACM SIGKDD Explorations Newsletter*, 2019, 21(2): 48-60.
- [55] WU K, YANG S, ZHU K Q. False rumors detection on Sina Weibo by propagation structures [C]//*2015 IEEE 31st International Conference on Data Engineering*, 2015: 651-662.
- [56] 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.
- [57] LIU Y, WU Y F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks [C]//*Proceedings of the 32th AAAI Conference on Artificial Intelligence*, 2018: 354-361.
- [58] INAN E. ZoKa: a fake news detection method using edge-weighted graph attention network with transfer models [J]. *Neural Computing & Applications*, 2022(14): 34.
- [59] 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.
- [60] ZHANG J W, DONG B, PHILIP S Y. FAKEDETECTOR: effective fake news detection with deep diffusive neural network[C]//*Proceedings of the 36th International Conference on Data Engineering*, 2020: 1826-1829.
- [61] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on Twitter [C]//*Proceedings of the 20th International Conference on World Wide Web*, 2011: 675-684.
- [62] MA J, GO W, WEI Z, et al. Detect rumors using time series of social context information on microblogging websites [C]//*Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, 2015: 1751-1754.
- [63] MA J, GAO W, WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]//*Proceedings of the Association for Computational Linguistics*, 2017: 708-717.
- [64] RUCHANSKY N, SEO S, LIU Y. CSI: a hybrid deep model for fake news detection[C]//*Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, 2017: 797-806.
- [65] LU Y J, LI C T. GCAN: graph-aware co-attention networks for explainable fake news detection on social media [C]//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020: 505-514.
- [66] HUANG Q, YU J, WU J, et al. Heterogeneous graph attention networks for early detection of rumors on Twitter [C]//*International Joint Conference on Neural Networks*, 2020: 1-8.
- [67] 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.
- [68] HAN X H, ZHAO M F, ZHANG Y T, et al. Jointly multi-source information and local-global relations of heterogeneous network for rumor detection[J]. *Frontiers in Physics*, 2023, 10: 1368.
- [69] GO J H, SARI A, JIANG J, et al. Fake news quick detection on dynamic heterogeneous information networks [DB/OL]. <https://arxiv.org/abs/2205.07039>.
- [70] YANG C, ZHOU X Y, ZAFARANI R. CHECKED: Chinese COVID-19 fake news dataset [J]. *Social Network Analysis and Mining*, 2021, 11(1): 58.
- [71] SHU K, MAHUESWARAN D, WANG S, et al. FakeNewsNet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media [J]. *Big data*, 2020, 8(3): 171-188.
- [72] LI K, GUO B, LIU J, et al. Dynamic probabilistic graphical model for progressive fake news detection on social media platform [J]. *ACM Transactions on Intelligent Systems and Technology*, 2022, 13(5): 1-24.
- [73] POTTHAST M, KIESEL J, REINARTZ K, et al. A stylistic inquiry into hyperpartisan and fake news [C]//*Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2018: 231-240.
- [74] SHU K, CUI L, WANG S, et al. DEFEND: explainable fake news detection [C]//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 395-405.
- [75] 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 [C]//*Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 2023: 395-405.