计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210300053

社交网络中的虚假信息:定义、检测及控制

王 剑 王玉翠 黄梦杰
郑州大学信息工程学院 郑州 450000

摘 要 近年来,社交网络上虚假信息传播愈演愈烈,在政治、经济、心理学等方面造成了严重的社会影响。有效检测社交网络中的虚假信息并对其实施控制,是改善社交网络生态系统质量的重要手段,能为人们营造一个安全、可信的网络环境。文中首先通过调研近年来国内外社交网络虚假信息领域的代表性研究,针对虚假信息中的假新闻和谣言,梳理并给出其定义、特征及传播模型,然后介绍了目前虚假信息检测及传播控制的各种手段及方法,最后总结并分析了目前的检测及控制方法中仍存在的问题,继而进一步探讨和提出了该领域未来的研究方向。

关键词:社交网络;假新闻;谣言;虚假信息检测;传播控制
中图法分类号 TP391

False Information in Social Networks:Definition,Detection and Control

WANG Jian,WANG Yu-cui and HUANG Meng-jie
School of Information Engineering,Zhengzhou University,Zhengzhou 450000,China

Abstract In recent years,the spread of false information on social networks has become increasingly fierce,causing serious social impact in political,economic,psychological and other aspects. Effective detection and control of false information in social networks is an important means to improve the quality of social network ecosystem and create a safe and credible network environment for people. This paper investigates the representative research in the field of false information of social networks at home and abroad in recent years,combs and gives its definition,characteristics and communication model for false news and rumors in false information,and then introduces various means and methods of detection and communication control of false information at present. Finally,this paper summarizes and analyzes the existing problems of detection and control methods,and then further discusses and puts forward the future research direction in this field.

Keywords Social network,Fake news,Rumor,False information detection,Communication control

1 引言

社交网络为人们提供了一个在世界范围、高度互联的平台,使每个人都可以在上面阅读、发布和分享信息,因此其成为人们获取信息的主要来源。美国皮尤调查中心2017年的一项调查显示,当时有超过三分之二(67%)的美国成年人通过网络渠道获取新闻,且这个数字正在呈指数级增长。然而,社交网络在为我们提供信息让我们共享便利的同时,也成为滋生和传播虚假信息的理想温床。Vosoughi等[1]调查近十年Twitter上发布的经过验证的真实和虚假信息的差异传播情况后发现,在所有类别的信息中,虚假信息比真实信息传播得更远、更快、更深、更广。同时,新媒体技术的发展也加速了信息的分享和大规模的传播,使得社交网络信息的生产成本大幅度降低,为虚假信息的传播提供了便利。

虚假信息是故意传播以误导或欺骗为目的的虚假或不准确的信息[2],其在网络上的广泛传播会给个人甚至整个社会带来深刻的负面影响,尤其是在政治、经济、心理学等方面。例如,2016年美国总统选举期间的假新闻传播风暴,乌克兰危机中社交媒体的假新闻运动,以及2017年德国联邦总统选举期间出现的社交机器人活跃和垃圾新闻频出事件等。在2020年"新冠肺炎疫情"期间,有人在社交网络上发布"多吃大蒜,喝藿香正气水、双黄连口服液可以治疗新冠肺炎"等关于"特效药"的言论,引起全民抢购热潮,造成了极坏的社会影响和市场影响。据英国天空新闻报道,位于英格兰伯明翰史帕克丘地区的一座5G信号基站被人为纵火,而原因仅仅是一段"5G会传播新冠病毒"的视频在某社交平台上广泛流传,这种虚假信息的传播掩盖了健康的行为,助长了错误的做法,增加了病毒传播的概率,最终可能导致个人出现心理问题。因

到稿日期:2021-03-04 返修日期:2021-04-20
基金项目:国家自然科学基金(61972133);中国博士后科学基金项目(2019TQ0286)
This work was supported by the National Natural Science Foundation of China(61972133) and China Postdoctoral Science Foundation(2019TQ0286).
通信作者:王剑(iejwang@zzu.edu.cn)

此，研究当前社交网络虚假信息传播的特征、规律及控制手段，对虚拟网络空间安全和有效治理意义重大，亟待深入研究。

目前，网络上的虚假信息可分为基于事实的和基于观点的两类[3]。基于事实的虚假信息包括与已证实为真实信息相矛盾的、故意捏造的或合并的信息，如假新闻、谣言等；基于观点的虚假信息是表达个人的观点，潜在地影响读者的意见或决定，这类虚假信息往往出现在电子商务平台或测评网站。本文的研究范畴是社交网络上公开的、基于事实的虚假信息，比较有代表性的是假新闻和谣言，不涉及基于观点的虚假信息，也不涉及电子商务、在线交易、协作平台上的骗局，以及恶意电子邮件、点击链接诱饵这类虚假信息。

本文基于近年来国内外针对社交网络中假新闻及谣言的研究文献的调研，对此类虚假信息的特征、检测及防控等问题进行了阐述和总结，并提出未来在此领域的有研究价值的方向及发展趋势。图 1 给出了本文后续章节行文的整体框架。本文第 2 节对社交网络虚假信息的定义及其传播过程、特征、传播模型进行了阐述；第 3 节介绍了目前社交网络中虚假信息的检测技术，包括传统检测手段及基于机器学习和深度学习的人工智能检测手段等，检测对象是虚假信息内容；第 4 节介绍了目前针对社交网络虚假信息的主要传播控制技术，包括从传播过程中以及从源头上遏制虚假信息的传播等两种传播控制策略；第 5 节总结了虚假信息检测及控制技术中存在的问题和可能的解决方法，以及这一领域研究的未来发展方向。

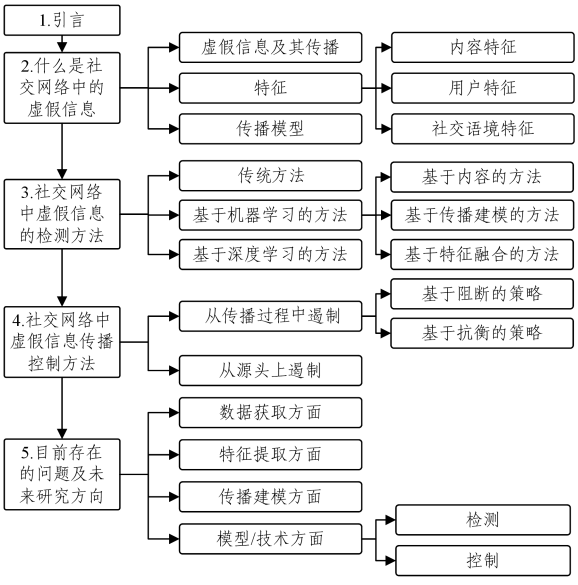

图 1 本文整体架构图

Fig. 1 Overall structure of this paper

2 什么是社交网络中的虚假信息

2.1 虚假信息及其传播

虚假信息存在的形式多种多样，如阴谋论(conspiracy theories)、标题党(clickbait)、伪科学(peudo science)，甚至捏造的“假新闻”(fake news)等。下面对社交网络中的虚假信息以及具有代表性的虚假信息(假新闻和谣言)的定义、联系及其传播过程进行介绍。

虚假信息[4]是指制造者故意误导读者，并能够通过其他来源证实其结果为假的信息，通常具有故意性(intent)和可证实性(verifiability)。传播媒介是虚假信息存在的必要条件，因此，在社交网络平台上产生并传播的虚假信息被称为社交网络中的虚假信息。假新闻是其中的一个代表，文献[1]对“新闻”的定义是任何带有断言的故事或主张，“假新闻”则是违背新闻中客观性和真实性的原则，意图误导观众和读者，使他们分不清真实和虚假的信息，从而达到制造者谋利或传达某种看法、观点的目的。随着自媒体的发展，在当前政治和媒体环境中，判断是否为假新闻的主要依据并不是发布报道的来源机构，而是发布的内容。我们对假新闻的定义是“在形式和写作风格上模仿新闻，是对某个事件的断言，为了某种目的而故意误导其他人的信息”。谣言是通过社交网络传播新闻内容或带有主观意见评论的一种社会现象。谣言内容的真伪是未经证实的，但当被证明是错误的言论时，谣言就成了虚假信息。

以假新闻和谣言形式出现的虚假信息中存在一些联系：若对假新闻发表意见一致的言论并大量传播则会形成谣言。虚假信息传播的这种动态性，使得社交网络中的假新闻和谣言通常是组合出现的。因此，在研究虚假信息的内容特征时，通常是对在整个话题内产生的所有信息内容进行分析，提取出虚假信息的内容特征。

由上述虚假信息的定义可描述出其传播过程为：首先，由带有某种意图的制造者创建虚假信息；其次，以社交网络作为传播媒介；最后，被广大网民评论并扩散。

2.2 社交网络中虚假信息的特征

本节将描述社交网络上虚假信息的特征，综合虚假信息各方面的属性，将虚假信息的特征分为内容特征、用户特征和社交上下文特征。通过研究虚假信息的特征，可以将某个特征或几个特征作为识别虚假信息的证据或线索。

2.2.1 内容特征

内容特征是从虚假信息的来源处、标题、正文部分、正文内容的图片和视频等中提取的，常用来区分真实和虚假信息或判断信息的可信度。内容特征包括文本特征和视觉特征。

(1)文本特征：由词汇(如词的总数、每个词的字符、词的频率，独特词汇等)和句法(如虚词和短语的频率、标点符号和词性标注等)等特征构成。Horne 等[5]发现，与真实推文相比，虚假推文中使用的技术词汇更少、单词更短、标点符号更少、引用更少以及词汇冗余更多。另外，虚假推文更容易阅读，其中使用的分析词汇、名词和副词较少，人称代词(如我、我们、你们)较多。

(2)视觉特征：是从视觉元素(如图像和视频)中提取的虚假信息特征，包括清晰度得分、一致性得分、相似性分布直方图、多样性得分和聚类得分等。在现实社交网络中，虚假推文中附带的图像往往清晰度不高、图文一致性得分较低，即图文意思不符的情况较为严重。

2.2.2　用户特征

用户特征指制造或传播虚假信息的用户的属性和行为特征，常用于检测恶意用户。对于虚假信息的检测任务，用户特征可作为辅助特征以提高检测的准确率，其包括用户属性特征和用户行为特征。

(1)用户属性特征：用户属性包含用户各个方面的统计资料，如用户类型、用户简介、注册时间、关注者数量/被关注者数量、已发布的推文数量等。有研究表明[6]，虚假信息创建者通常使用的是最近注册的账号，并且大多是“一次性账号”。

(2)用户行为特征：主要研究的是用户在传播虚假信息后表现出的一些不同于正常用户的异常行为，如平均每日关注人数、平均每日发布推文数量、转发推文的时间间隔等。

2.2.3　社交语境特征

社交语境指信息传播运作的整个社交环境，研究社交语境特征是为了分析传播过程中虚假信息在网络中的分布情况，用户之间的互动情况(点赞、转发、评论)等，从而对虚假信息的传播进行建模，完成虚假信息检测和控制任务。社交语境特征根据研究对象的不同可分为网络特征和传播特征。

(1)网络特征：主要表现在虚假信息传播的网络拓扑结构上的一些特征。1)网络拓扑结构类型，有友谊网络[7]和扩散网络[7]等。友谊网络显示了发布相关推文的用户之间的关系结构(互相关注还是单边关注)。扩散网络可以跟踪信息传播的轨迹，研究信息在网络中的传播情况，其中，节点代表用户，边代表用户之间的信息传播路径。2)网络度量指标，有聚类系数、扩散系数和中心性度量等。聚类系数是描述整个网络的聚类指标，是根据与某个节点相连的两个点是否也相连来定义的，文献[8]构建了一个模拟谣言传播的网络，结果表明聚类系数越低的网络越稀疏。而文献[9]表明了稀疏的网络比密集的网络更容易传播虚假信息。扩散系数可表示信息在网络中的扩散程度，文献[10]表明，与谣言相关的言论扩散系数都较高。中心性度量被用来给网络中的每个节点分配计数，其中，度中心性表示连接到节点的边的数量，这个指标反映了由哪些节点负责大部分的传播。特征向量中心性表示该节点所有相连节点或邻居的度中心性之和，具有较高特征向量中心性的节点在网络中的周围节点中具有主导影响力，我们称这些节点为具有影响力的节点。

(2)传播特征：主要表现为虚假信息与真实信息在传播过程中存在的差异。1)它比真实信息传播得更深、更远。虚假信息的传播可以描述为有一个或多个级联，信息级联表示从他人的行为中获取信息，作为自己行为选择的参照，可以简单地理解为忽略了自己的想法而去认可他人的想法，模仿他人的信息传播。级联可以用数值表示，且级联个数和每个级联值越大，表示信息被传播得越深、越远。例如，有 m 个用户对同一话题事件进行断言，并且每个断言都被转发了 n 次，那么这个谣言就会包含 m 个级联，每个级联大小为 n。Friggeri 等[11]在对 Facebook 上谣言重塑进行研究后得出结论：与非谣言相比，当谣言被再次传播时，往往比上一次传播得更深，从而接触到更多的人。类似地，文献[8-9]都表明虚假信息比真实信息传播的范围更广、速度更快。2)虚假信息的传播是由少数非常活跃的用户主导的。Gupta 等[12]发现，飓风桑迪期间 Twitter 上转发的虚假图片中，有 90％的图片都是前 30 名用户转发的。Shao 等[13]对约 130 万条虚假推文研究后发现，在这些信息中更多的是用户之间的互动对话，这表明重复和坚持在虚假信息传播中起着重要作用。3)虚假信息在传播早期呈现爆发的状态，即在早期阶段的传播尤其迅速。Zubiaga 等[14]收集了 330 个谣言线程和 4 842 条关于 9 个流行案例的推文，并研究了真/假信息在社交媒体上传播的整个生命周期，包括真实性被检验之前和检验之后，并根据虚假信息讨论线程量化了推文在 Twitter 中被支持的程度。他们发现，虚假信息的传播很大程度上发生在被揭穿之前。支持未经证实说法的推文获得了最多的转发，甚至在开始的几分钟内就出现了爆发性的转发，在谣言的真实性被核实后，人们对谣言的兴趣大幅下降。

2.3　社交网络中虚假信息的传播模型

一个社交网络可以用一个图来表示，由于信息是通过网络中不同个体之间的相互作用传播的，即社交网络拓扑结构中的节点可表示个体(用户)，边表示信息间的传播路径。目前有许多研究是对信息在网络上的传播过程进行建模。在此基础上，人们研究了虚假信息在社交网络上的传播模型，这一研究是为了解释虚假信息的传播理论依据及可视化传播过程、预测信息未来的传播路径和传播趋势，从而发现真实和虚假信息在网络传播中更深层次的差异，找出影响虚假信息传播的因素。对于广泛用于虚假信息检测任务的信息传播模型，根据模型功能的不同，可将其分为解释模型和预测模型。

2.3.1　解释模型

解释模型旨在检验信息传播过程并阐明影响信息传播的因素，试图解释虚假信息传播这一现象。由于虚假信息的传播与传染病的传播具有相似性，多数研究人员将基于数学统计的宏观数学模型——传染病模型[15]应用于虚假信息的传播并进行建模。因此，基于传染病的模型及其改进模型属于解释模型。如图 2 所示，基于传染病的模型中，用户被抽象为 3 种状态：易感染(Susceptible，S)、感染(Infected，I)和治愈(Recovered，R)，状态之间以某个概率进行转换。下面分别介绍这些模型。

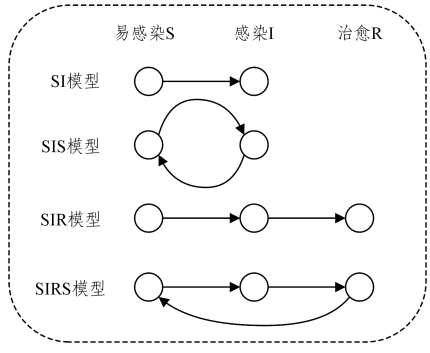

图 2　基于传染病模型的虚假信息传播模型

Fig. 2　False information dissemination model based on infectious disease model

根据状态之间的关系，可构建 SI[16]，SIS[17]，SIR[18]，SIRS[19]4 种基础模型。其中，SI 模型中 S 类中的节点都以某个固定的概率切换到 I 类，未考虑感染节点被治愈的情况；SIS 模型中 I 类中的节点再以某个固定概率切换到 S 类；SIR 模型考虑了感染节点可被治愈的情况，此时 I 类节点永久地切换到 S 类；SIRS 模型考虑了被治愈节点还会以某个概率转换为易感染节点，即 R 类节点以一定概率切换到 S 类。总的来说，这 4 种模型是通过描述网络中用户对于信息的接收状态以及节点在这些状态间的重新分配来研究信息的传播。

虚假信息在现实社交网络中的传播情况受一些因素影响（如信息内容、节点间关系强度、时间、网络结构等），使得节点处于以上 3 种状态之外的某个状态，且节点间的转换概率也会变化。Sang 等[20]研究了信息扩散的过程及其影响因素。在传统流行病模型的基础上，他们提出了 STFRD（易感、相信真相、感染、康复和免疫）模型，将全体用户分成 5 个部分，每个用户在任何时候都处于 5 个状态之一。他们假设每个用户一开始都容易受到 S（易感）的影响，若用户能主动控制屏蔽虚假信息，则该用户处于 D（免疫）状态；若该用户选择相信虚假信息，则处于 F（感染）状态；若选择相信事实，则处于 T（相信真相）状态。当用户转换为 F 状态时，是可以被治愈和免疫的，治愈后处于 R（康复）状态，免疫后处于 T 状态。实验中设置了不同的参数值来观察节点状态的转换情况，最终发现，每个状态用户的数量以及传染概率都会影响最终的传播规模。一些研究人员把用户所处的状态划分得更为细致，文献[21]基于 SIR 模型开发了动态 8 态 ICSAR 模型，把用户分成 8 种状态：无知、无知清除、谣言传播者、谣言载体、谣言倡导者、真相传播者、真相载体和真相倡导者，来分析谣言传播机制。文献[22]应用传染病模型研究了恶意节点（一旦收到虚假消息，就会转发虚假消息）对虚假信息传播的影响。分析表明，对于恶意节点的随机分布，存在一个临界数量的恶意节点，超过这个数量，虚假信息就会在网络中爆发。文献[22]进一步提出了选择恶意节点进行虚假信息传播的 3 种不同的分发策略（即基于度、基于距离和基于紧密度的分布策略）。研究结果表明，恶意节点度大或者连接紧密会扩大传播的规模。还有一些研究发现，若节点具有很大的影响力，也会加速虚假信息的传播。文献[23]应用 SIR 模型建立了一个基于自媒体的谣言传播系统——SHLR，在这个模型中，造谣者分为高速和低速。其通过对平衡点和稳定性的研究，给出了系统的数值分析。实验数据显示，自媒体会扩大谣言的传播范围和影响。

2.3.2 预测模型

预测模型旨在基于某些因素预测社交网络中未来的信息扩散过程。此类基础模型有独立级联模型（ICM）[24]、线性阈值模型（LTM）[25]和博弈论模型（GTM）[26]。

ICM 和 LTM 的网络拓扑图都可看作为有向图，其中每个节点都可以被激活（已经接收到信息并且试图传播它）或不被激活（不传播）。ICM 要求扩散概率与每条边相关联，而 LTM 要求在每条边上定义影响程度，并为每个节点定义影响阈值。对于这两种模型，扩散过程以同步方式沿着离散时间轴迭代进行，从一组最初激活的节点（最先采纳一条信息）开始扩散，然后触发其扩散的一组用户。在 ICM 情况下，对于每次迭代，新激活的节点尝试激活其邻居一次，激活概率加载在其边上；在 LTM 的情况下，在每次迭代中，如果影响程度的总和超过其自身的影响阈值，则未激活的节点将被由它们激活的邻居激活。这两种情况下，当不可能有新的传输时，即没有相邻节点可以联系时，该过程结束。ICM 通过边上的概率来描述用户之间产生影响的可能性和强度，而 LTM 的随机性完全是由节点的影响阈值决定的，一旦阈值确定，后面的传播过程就是完全确定的。这是线性阈值模型不如独立级联模型应用广泛的一个原因。之后许多研究人员对这两个模型进行了改进。文献[27]构造了一个改进的 LTM 和两个网络信息熵及其意见指标，来分析衍生虚假信息的传播规律。基于真实的社交网络数据集，他们研究了传播误差、恶意节点比例和篡改概率对谣言传播的影响。实验结果表明：1）传播误差的增大会显著提高系统的混乱程度，使观点的分布从指数变为正态；2）增加恶意节点的比例明显增加了传播虚假信息的个体数量；3）篡改概率的作用效果与恶意节点的比例密切相关。篡改具有高比例的恶意节点，将导致系统的混乱度达到峰值。

博弈论是一种利润最大化的策略。博弈论的研究仅限于多个有特定限制的个体或群体。它总是利用对手的策略来实现利润最大化。由于成本、收益和战略选择的影响，一条信息要么传播，要么不传播。文献[28]运用博弈论，将用户间关系作为一个重要变量加入到模型中，并分析了用户选择传播信息的成本、收益和策略选择。结果表明，只要信息扩散的利润大于信息成本，用户就会选择传播信息，并且用户之间的关系越密切，信息就越容易传播。在此基础上，文献[29]提出了网络群体行为的进化博弈模型，他们认为个人信息行为在微观层面上的特征比在宏观层面上复杂得多。由于社会性和随机性，网络中的群体行为往往表现出很大的不确定性。进化博弈模型适用于解决社会网络中信息传播的动态问题。

事实上，解释和预测模型并不是完全相互独立的，前者是后者的研究基础，后者是前者的工具或方法，两者是相辅相成的。总之，通过构建虚假信息的传播模型，可分析出信息传播的影响因素和下一步传播路径，为实现虚假信息控制奠定了基础。

3 社交网络中虚假信息的检测方法

虚假信息检测的目标是在传播的早期阶段有效地识别出虚假信息，而识别出虚假信息可看作是将信息分类为真实和虚假的过程。为了实现这一过程，研究人员先从信息中提取一些相关特征，然后使用分类器来区分信息的真假。可提取的特征有内容特征、网络特征、传播特征等，根据检测技术的发展历程，社交网络中虚假信息检测方法可分为传统的方法和基于机器学习及基于深度学习的方法。

3.1 传统的方法

由于虚假信息试图在内容中散布虚假的陈述或观念，因此传统的检测虚假信息的方式是利用事实进行手动核查，将信息内容与专家核查的真实信息进行比较，以评估信息的准确性。

手动事实核查可分为基于专家和众包的方式。在基于专家的方式中，验证信息真实性的过程是由这个领域的一组专家进行的。为了使基于专家的事实核查能更好地为公众服务，许多提供这种服务的网站应运而生。文献[30]提到了一个名为 PolitiFact. com 的网站，该网站对民选官员、权威人士、专栏作家、博客作者、政治分析人士和其他媒体人士的声明或声明的准确性进行评级，是一个独立的、无党派的政治新闻和信息的在线事实核查系统。网站编辑们人工检查声明的具体词语和整个上下文，验证声明的可靠性，并将声明的真实性分为对、大部分对、一半对、大部分错、错误和严重错误 6 个级别。PolitiFact 网站还为用户提供 API，允许用户访问已被检查的声明、故事、承诺和更新的全文。专家核查的方法的优点是易于管理、准确性很高；缺点是成本昂贵，并且在待检测信息量增加时，准确性变得难以衡量。

基于众包的方式是利用群众的智慧在众包市场进行事实核查，如亚马逊机械土耳其公司，其聚集了一个庞大的事实核查人员群体，并创建了可公开获得的假新闻数据集“CREDBANK”[31]。同时，也有一些网站利用众包的方式进行事实核查，但这些网站还处于早期开发阶段。如 Fiskkit，在这个网站上，用户可以上传文章并对文章中的句子进行评级，给文章加上一些最能描述其内容的标签，有助于区分内容的类型（新闻或非新闻）以及确定其可信度。与基于专家的事实核查相比，众包事实核查的成本较低，但相对难以管理，并且由于事实核查者的政治偏见以及相互冲突的标签，会降低核查结果的准确性。因此，在众包事实核查中，人们通常要过滤掉不可信的用户，并且要解决核查结果相互冲突的事件，这本身会对检测效率及准确性带来一定的影响。

3.2　基于机器学习的方法

基于机器学习的方法旨在研究怎样利用数据使计算机的学习过程自动化，一般是使用基于文本的算法或自然语言处理技术将文本信息转换成一个巨大的特征向量，并将该向量输入到有监督的学习模型中，以自动识别虚假信息。本文按照提取特征的不同，将此类方法分为基于内容的方法、基于传播建模的方法和基于特征融合的方法。

表 1 列出了现有基于机器学习的方法中使用的特征、模型及检测性能。

表 1　基于机器学习的检测方法总结

Table 1　Summary of detection methods based on machine learning

Thesis	Type of feature			Classifier/model	Dataset	Evaluation index and result/%
	Content	Social context	User and other			
文献[32]	Capitalization, unigral-exical features, bigram-based lexical features, hashtags, URL, etc.	—	the log-likelihood ratio that user is under a positive user model, etc.	a KL divergence retrieval model with Dirichlet smoothin(KL)	Twitter	*Acc*:94.1
文献[33]	Linguistic features: posemo, negate, social, cogmech, excl, insight, tentat, see, hear, etc.	—	Total population of available users, Probability of infection, Starting time of breaking news Background noise, Strength of interaction periodicity, etc.	DT, RF, SVM	Twitter	*Acc*:90; *F*1:89.3
文献[34]	Linguistic features: Modal Adverb, Action Adverb, 1st pers singular (I), Manner Adverb, Superlatives, etc.	—	—	LSTM	PolitiFactData	*F*1:58
文献[35]	Average length of microblogs, # of positive (negative) words in microblogs, Average sentiment score of microblogs, % of microblogs with URL, % of positive (negative) microblogs, etc.	Average # of retweets, Average # of comments for Weibo posts, # of microblogs, etc.	% of users that provide personal description, % of users that provide personal picture in profile, % of verified users, % of male (female) users, etc.	SVM	Twitter/ Sina Weibo	*Acc*:89.6/ *Acc*:84.6
文献[36]	has multimedia, sentiment, has URL, time span, topic type, etc.	num of coments, num of reposts, repost time score, etc.	is verified, has description, gender, location, num of followers, etc.	混合 SVM	Weibo	*Acc*:91.3
文献[37]	—	Message dissemination records, etc.	Age, Registration time, Number of followers, Number of friends, Weibo level, Active days, etc.	SVM	Weibo	*Acc*:81.3
文献[38]	Tweet With Most Frequent User Mentioned, Length In Chars, Tweet With Most Freq URL, etc.	Number of referrals, etc.	Profile Image, Age, Follower Count, Status Count, Location, etc.	SVM, NB, KNN, DT	Twitter	*Acc*:86
文献[39]	# of source tweets, # of false rumors, # of true rumors, etc.	# of threads, etc.	# of users, etc.	PTK, cPTK	Twitter15/ Twitter16	*Acc*:75/ *Acc*:73.2
文献[40]	Style, complexity, readability, Syntax, Biased language, Connotations, etc.	—	moral foundations, psycholinguistic signals, etc.	ME, RF	Summaries/ NewsPages/ Tweet	*F*1:92

3.2.1 基于内容的方法

对于一个话题事件，社交网络上发出的帖子通常是一段文字描述，再关联几个图片或视频。基于内容的方法主要是利用内容特征（文本特征）研究虚假信息特定的语言用词和写作风格。

许多研究利用文本特征来检测虚假信息。例如，Qazvinian[32]表明词汇、词性是一个有效的文本特征，为此他们手动标注了一个数据集（包含来自 5 个不同争议话题的 10 000 条推文）。特别地，他们认为单词大写是一个重要属性，因此在处理出现在用户时间线中的推文时，没有做任何预处理。Kwon 等[33]发现，一些类型的情感词是机器学习分类器的明显特征，包括积极情感词（如 love，nice，sweet）、否定词（如 no，not，never）、认知动词（如 cause，know，ought）和推断动词（如 may be，perhaps，guess）。继而，他们提出了一个周期性时间序列模型来识别 Twitter 中谣言和非谣言之间的关键语言差异。类似地，Rashkin 等[34]总结了不可信新闻内容的语言风格，发现第一/第二人称代词以及夸张的词在低可信度信息中使用更频繁，进而他们以此作为判断信息可信度的依据。

3.2.2 基于传播建模的方法

基于传播建模的方法是借助信息的社交语境特征（网络特征和传播特征）来构建信息传播网络模型，从整体上评价某个话题事件或帖子的可信度。

一些研究通过区分真、假信息的传播模式来检测虚假信息。例如，文献[35]表明信息传播的特征会随着时间的推移而逐渐改变，并提出一个动态序列-时间结构（DSTS）模型来表征虚假信息传播特征的时间模式。类似地，文献[36]联合建模消息传播结构、主题信息、用户属性等提出了一个混合 SVM 分类器，其在谣言扩散的早期阶段具有很强的检测性能，但比较耗时。文献[37]基于用户的异质性将用户分组，定量地模拟了不同用户群体之间的谣言传播情况，以此识别虚假信息的特殊传播结构，其同样更适用于谣言早期阶段的检测。该方法的缺点是需要获取用户的各种属性信息，只适用于像 Facebook 这种存储了用户详细资料的网站。为了能够利用不同类型的数据来增加模型的灵活性并提高其预测能力，Jooyeon 等[41]联合文章主题和用户兴趣提出了一种贝叶斯非参数可扩展模型来描述新闻文章的传播，并且可将用户的传播级联和用户的信息内容合并到模型中，以提高检测能力。

此外，许多研究还通过构建特定的网络结构或树来检测虚假信息。例如，Manish 等[38]构建了一个包含用户、帖子和事件的可信度传播网络，以模拟虚假信息的传播过程。与之不同的是，文献[39]基于内核的方法将谣言相关微博的传播建模为传播树，通过评估传播树结构之间的相似性来捕获区分不同类型谣言的高阶模式。由于这种方法比单纯基于特征的方法涵盖更多的结构信息，因此具有更好的性能。

3.2.3 基于特征融合的方法

基于内容的检测方法主要识别写作风格和词汇句法特征方面的真实和虚假信息之间的差异，而基于传播建模的检测方法主要使用从网络结构及信息传播过程中提取的特征。但这两种方法在应用场景上存在局限性，如文献[42]表明，不能仅通过写作风格等特征来解决虚假信息的检测问题，以及基于传播建模的方法大多只适用于谣言扩散早期阶段的检测。因此许多研究人员开始研究新的基于特征融合的方法。例如，Vedova 等[43]利用了用户和帖子之间的交互信息（转发、评论等），以及帖子的文本信息。具体来说，他们对社交帖子进行词干分析，并将每个帖子表示为单词的 TF-IDF 向量。然后，他们利用用户的相似行为来描述信息传播特征，并最终通过整合这两种信号来识别虚假信息。Volkova 等[40]通过整合新闻内容的心理语言学信号和文本信息信号来检测社交网络中的虚假信息。此外，文献[44]考虑了信息传播过程中信息发布者、推文内容和普通用户之间的三重关系，提出了一个名为 TriFN 的通用检测框架，该框架通过非负矩阵分解（Nonnegative Matrix Factorization，NMF）算法对它们之间的内在关系进行建模，以识别低可信度信息。

表 1 列出了基于机器学习的检测方法中使用的特征、分类器/模型及检测出虚假信息性能的总结。主要性能评价指标如下。

（1）Acc（Accuracy）：准确率，其计算公式如式（1）所示：

$$Acc=\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

其中，TP 为被正确划分为正例的个数，FP 为被错误划分为正例的个数，FN 为被错误划分为负例的个数，TN 为被正确划分为负例的个数。

（2）F1-score：F1 分数，它是精确率 P 和召回率 R 的调和平均数，其计算式如式（2）－式（4）所示：

$$P=\frac{TP}{TP+FP} \tag{2}$$

$$R=\frac{TP}{TP+FN} \tag{3}$$

$$\frac{2}{F1}=\frac{1}{P}+\frac{1}{R} \tag{4}$$

（3）AUC：AUC 值为 ROC 曲线下与坐标轴围成的面积。ROC 曲线是以 FPR 为横轴、TPR 为纵轴所绘制的。计算公式如式（5）、式（6）所示：

$$FPR=\frac{FP}{FP+TN} \tag{5}$$

$$TPR=\frac{TP}{TP+FN} \tag{6}$$

从表 1 中可以看出，在检测虚假信息时，不仅需要提取关于虚假信息的特征，还需要提取其他特征（如用户信息、用户好友数、关注数等）来辅助检测。目前常用的分类器有朴素贝叶斯（Naive Bayes，NB）、K 近邻（K Nearest Neighbors，KNN）、支持向量机（Support Vector Machines，SVM）、决策树（Decision Trees，DT）、随机森林（Random Forest，RF）、最大熵（Maximum Entropy，ME）等。它们都能有效地识别出虚假信息，且融合越多有用的特征，性能越好。

3.3 基于深度学习的方法

基于深度学习的方法旨在自动提取虚假信息数据的高级表示。在这一领域使用的深度学习的方法有很多种，其中最常用的方法是深度神经网络（DNNs），尤其是卷积神经网络（CNNs）[45]和递归神经网络（RNNs）[46]。下面介绍几种广泛使用的深度学习模型。

(1)卷积神经网络(CNNs)：它是典型的前馈神经网络，具有卷积层、汇集层和全连接层。通过多重卷积和汇集操作，CNNs可以从输入中捕获局部和全局特征，并输出分类结果。例如，文献[47]结合卷积网络和循环网络构建了一个分类器，它可以捕获沿传播路径的用户特征的全局和局部变化，通过对新闻传播路径进行分类，可在社交媒体上于传播早期检测出假新闻。另外，基于CNNs的虚假信息检测模型可以灵活地提取分散在输入序列中的关键特征，并在重要特征之间形成高级交互，这有助于有效地识别虚假信息并实现早期检测。如文献[48]考虑到帖子的时间序列特征有助于对事件进行精确建模，进而提出了ACAMI模型，该模型利用event2vec(用来学习事件的分布式表示)和一种注意机制提取事件的时间和语义特征，然后利用CNNs提取高级特征对虚假微博帖子进行分类。

(2)图形卷积网络(GCNs)：与CNNs不同的是，GCNs处理的数据是图结构，可有效捕捉具有图特性的社交网络的结构特征以及信息在网络中的传播结构特征，然后为分类器提取信息传播模式的高级表示。例如，文献[49]提出的基于GCNs的模型可根据原始推文和所有相关推文，即评论和转发，以每条推文为节点、推文传播路径和用户关系为边构建事件特定图。并且其使用具有两个卷积层(每个层64维输出特征图)和两个完全连接层(分别产生32维和2维输出特征)的4层GCNs来预测信息真假的概率。

(3)递归神经网络(RNNs)：也是一种前馈神经网络，主要用于处理可变长度的序列或时间序列数据。社交网络上发布的信息具有明显的时序特征，可以被划分为连续的片段，并通过RNNs捕捉序列特征。例如，文献[50]提出了一种用于捕捉用户评论连续流的时间-语言特征模型。文献[51]提出的CSI模型由捕捉、评分和集成3个模块组成。第一个模块基于用户响应和文本，它使用RNNs来捕捉给定新闻文章中用户活动的时间模式；第二个模块根据用户的行为学习源特征，两者与第三个模块相结合，将一篇二新闻文章分类为真或假。

由于RNNs自身的局限性(可能遭受梯度消失或爆炸)，随后研究者们提出了具有门控机制的RNNs——长短期记忆(LSTM)[52]和门控循环单元(GRU)[53]，它们可直接提取出更多有效特征，提高虚假信息检测的准确率和时间效率。例如，文献[54]基于LSTM构建的模型可用来辨别信息标签的真伪。文献[55]使用GRU中的复位门和更新门来共同控制和更新单词嵌入层，通过这种方式，模型可以逐步学习输入文本中单词的分布式向量，从中捕获单词的深层潜在特征及其相关性，以此构建虚假信息检测模型。LSTM与GRU在功能效果上相似，但LSTM参数更多，在训练过程中需要更多的计算能力。因此，考虑到硬件的计算能力和时间成本，目前在虚假信息检测领域研究人员更倾向于使用GRU。

(4)注意力机制(attention mechanism)：常用于描述神经网络对输入序列的注意分布。它计算当前输入序列和输出向量之间的匹配度，目的是捕捉输入的关键信息，匹配度越高，关注度分数就越高。因此，它能够从更少的数据和更多的噪声中挖掘出关键特性，适用于虚假信息检测的早期阶段。例如，文献[48]提出内容注意和时间注意，用来学习话题事件的内容和时间信息的重要性权重，即：注意力机制选择性地关注事件的重要内容和时间特征，更有针对性地和高效率地检测虚假信息。

表2列出了现有基于深度学习的检测方法使用的模型、输入的数据以及模型性能。可以看出，许多研究通过将相关帖子建模为时间序列数据来学习虚假信息的潜在文本内容表示；一些方法将帖子文本信息、传播信息及用户信息等结合起来作为深层神经网络的数据输入；还有一些方法利用视觉信息来检测虚假信息，如文献[56]利用频率域和像素域的视觉信息，在物理和语义层面上有效地捕捉了假新闻图像的特征，再利用注意力机制动态融合了频率域和像素域的特征表示。

表2　基于深度学习的检测方法总结

Table 2　Summary of detection methods based on deep learning

Thesis	Input data						Model	Dataset	Evaluation index and result/%
	Textual data	Interaction data	User info	User position	Propagation path	Visual data			
文献[47]	√	—	√	—	√	—	CNNs+RNNs	Twitter/Weibo	Acc:85/Acc:92
文献[48]	√	√	—	—	—	—	CNNs+Attention	Twitter/Weibo	F1:94.6/F1:79.4
文献[49]	√	—	√	√	√	—	GCN	Twitter	AUC:92.7
文献[50]	√	√	—	—	—	—	RNNs	Twitter/Weibo	Acc:83.9/Acc:89
文献[57]	√	√	—	—	—	—	双向GRU	BuzzfeedNews/PolitiFact	Acc:76.42;F1:76
文献[54]	√	—	—	—	—	—	LSTM	PHEME	F1:71
文献[55]	√	—	—	—	—	—	GRU+LSTM	ISOT	Acc:99.8
文献[58]	√		√	—	—	—	双向LSTM+Attention	Twitter/Weibo	Acc:84.4/Acc:94.3
文献[51]	√	—	√	√	—	—	RNNs	Twitter/Weibo	F1:89.4/F1:95.4
文献[56]	√	—	—	—	—	√	RNNs+Attention	Weibo	Acc:90.1;F1:90.6

总的来说，利用深度神经网络将异构数据或多模态数据进行融合，自动提取更高级的特征，能够更加准确地检测出社交网络中的虚假信息。表3列出了目前应用于虚假信息检测的公开标记数据集，显然这些数据集都不能同时包含检测时所需的所有特征。因此，基于机器学习和基于深度学习的检测方法都存在的一个关键性问题：需要收集足够强大、丰富和可靠的标记数据集，以便在其上训练和测试算法或模型。

表 3 社交网络中公开的虚假信息检测数据集

Table 3 Public false information detection data set on social networks

Dataset	Source/platform	Information contained	Feature coverage					Data connection
			Content		User	Social context		
			Text	Visual		Network	Propagation	
BuzzFeedNews	Facebook	1627 articles	√	—	—	—	—	https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/tree/master/data
LIAR	PolitiFact	12836 short statements	√	—	—	—	—	https://www.cs.ucsb.edu/~william/software.html
BS Detector	BS detector	—	√	—	—	—	—	https://github.com/bs-detector/bs-detector
CREDBANK	Twitter	60 million tweets	√	—	√	—	—	http://compsocial.github.io/CREDBANK-data/
BuzzFace	Facebook	2263 news articles and 1.6 million comments	√	—	—	—	√	https://github.com/gsantia/BuzzFace
FacebookHoax	Facebook	15500 posts	√	—	√	—	√	https://github.com/gabll/some-like-it-hoax
NELA-GT-2018	194 news and media outlets	713000 articles	√	—	—	—	—	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ULHLCB
FakeNewsNet	Twitter	201921 articles	√	√	√	√	—	https://github.com/KaiDMML/FakeNewsNet

4 社交网络中虚假信息传播的控制方法

针对社交网络中虚假信息传播的控制策略有两个，一个是在虚假信息的传播过程中控制其传播，或打压其传播势力；另一个是从源头上遏制其传播，即建立一个预防系统，防止虚假信息的发布和传播。

4.1 在传播过程中遏制虚假信息传播

此类方法主要分为两种策略：一种是阻断策略，阻断虚假信息传播，减小传播范围；另一种是抗衡策略，通过扩大真实信息的传播力度和范围来打击虚假信息，使得信息在传播的过程中最小化相信虚假信息的用户数量。在网络中，节点代表现实系统中的实体，节点之间的连边代表实体间的联系。信息在真实的网络系统中传播，将节点作为载体，连边作为媒介，也就是节点代表用户，连边代表传播路径。通过对信息传播模型/结构中的节点或边进行控制，从而实现对社交网络中虚假信息传播动力学的干预。

4.1.1 基于阻断策略的控制方法

表 4 列出了基于阻断策略的控制方法，控制方法的原理可归纳为两种。

其一，阻塞(保护)或删除网络中的一组节点，最小化虚假信息的传播规模及影响(最小化虚假信息接触到的用户量)。阻塞(保护)节点，即将该节点免疫，使它不受周围节点的影响；删除节点，即将传播虚假信息的节点删除。解决这类问题的关键在于如何识别出这样的一组关键节点。文献[59]提出了一种最大边际增益的规则，它定义了节点选择规则：当所选节点被添加到阻塞节点集时，该规则使最终感染节点数的减量最大化，以此达到感染范围最小的效果。Yan 等[60]继续了文献[59]的研究，选择一个具有 k 个节点的阻断集 B，使得用户被谣言源集 S 激活的总概率最小。为了有效地解决镜像问题，他们又提出了一种生成候选集和为一般网络选择阻塞器的两阶段方法。文献[61]通过识别一个极小的个体子集作为初始保护器，来最小化在真实和虚假信息扩散过程结束时的感染人数，所提出的贪婪算法具有 $1-1/e$ 的近似比。在现实的社交网络中，通常还要考虑时间和成本的问题，在有限时间和预算内最大限度地控制虚假信息的传播。文献[62]结合时间因素，在独立级联模型和线性阈值模型下，提出了两个新的社会网络竞争影响扩散模型，这两个模型考虑了信息扩散的时间期限、信息交换的随机时间延迟和个人对信息接受的兴趣，并使用近似比为 $1-1/e$ 的贪婪算法找出了最初保护节点集。Pham 等[63]提出一个具有 $1-1/\sqrt{e}$ 的近似比的加速贪婪算法(SG)，用于求解待删除节点集；同时又提出了一个适用于大型网络的启发式算法(PR-DAG)，且实验结果表明 PR-DAG 比 SG 运行得更快(高达 45 倍)。文献[64]在研究识别网络中的关键节点时，证明了这个问题在线性阈值下是 P-难的，在独立级联扩散模型下是 NP-难的。

其二，切断网络中用户之间传播某条虚假信息的连边(实际上是消除节点之间的相互作用，阻断传播路径)。研究人员通常致力于寻找最小的边集合进行删除，文献[65]将该问题定义为最小化虚假信息传播污染区。目前大多数边缘阻断研究侧重于简单的传染性，文献[66]基于传染病模型，提出了根据边的重要性来删除边的策略，并且表明定向删除边策略比随机删除边策略更有效。文献[67]提出了适用于线性阈值模型的一种贪婪算法。在此基础上，Kuhlman 等[68]研究了简单和复杂的传染，确定了最小成本的边子集并进行删除，然后证明了对于简单和复杂的传染，保存所有可挽救节点(即所有可避免受影响的节点)的基于边的阻塞问题是有效可解的。Yao 等[69]提出了具有精度保证的两种启发式算法及贪婪算法，用于寻找待删除边集合的近似解。并且通过实验证明了贪婪算法在最小化负面影响方面比其他两种算法更有效，而基于中间度和超出度的启发式方法在运行时间上要比贪婪算

法快几个数量级。同样，文献[70]提出了一种近似比为$1-1/e-\varepsilon$的贪婪算法，该算法在时间和空间上的复杂性与网络的大小呈线性关系。由于边是两个用户之间交互的渠道，因此文献[71]将感染最小化问题定义为“最小化用户之间的虚假信息交互量”，且证明了该问题是 NP-困难的、问题的目标函数计算是#P-困难的，以及该目标函数既不是次模函数也不是超模函数。为了更高效地找到这个边集合，文献[72]通过跟踪信息的发展和传播，提出了一个关键边缘检测器(CED)，通过免疫这些边，能够在当前网络上最大程度地遏制虚假信息的传播。

表 4　基于阻断策略的控制方法总结

Table 4　Summary of control methods based on blocking strategy

Thesis	Network dataset	Information dissemination model	Method	Node/edge selection basis	Algorithm	
					Greedy algorithm	Heuristic algorithm
文献[59]	Enron email Network	ICM	保护k个未感染节点	节点的出度、节点中心性、最大边际增益规则	Greedy	—
文献[60]	Synthetic(SYN)/ Synthetic Tree(SYN-T)/ Wiki Vote(WV)/ Google+(G+)	ICM	找到一个具有k个节点的阻断集	最大边际增益规则	GCSSB	—
文献[61]	Enron email Communication Network/ Collaboration Network	OOAO，DOAM	确定一个最小的个体子集作为初始保护者	社区边界的节点、节点的度	SCBG	—
文献[63]	Gnutella/ Oregon/ Epinions/ EU Email	RPIC-M，RPLT-M	找到一组节点进行移除	时间、成本预算，边权重，传播阈值	SG	PR-DAG
文献[65]	Blog/ Wikipedia	ICM	移除一组边	节点的影响度、传播概率	Greedy	—
文献[68]	MONT-VA/ FB-1/ FB-2	SyDS	移除一组边	传播阈值、种子节点个数、成本预算、节点的度	—	ECH
文献[70]	COERPE RIPHERT/ ERDOSRENYI/ HIERARCHI/HEPPH/ EPINIONS/ MEMETRACKER	LTM	删除部分边	删除边的预算、扩散概率、节点的度、节点中心性、网络特征值	Greedy	—
文献[71]	CGSCol/ PClimate/ Higgs	ICM	移除一组边	节点间信息交互量、影响概率	TSGA	—

4.1.2　基于抗衡策略的控制方法

表 5 列出了基于抗衡策略的控制方法，控制方法的原理可归纳为两种。

其一，在网络中找到具有最大传播影响力的节点集，将它们保持为传播真相的节点。因此，此类方法中需要模拟两类(真实和虚假)信息在网络中的传播情况。Budak 等[73]对“如何找到具有最大传播影响力的节点子集”的问题进行了定义，并证明了这个优化问题是 NP 难的，为这个问题的各种定义的贪婪解提供$1/(1-e)$的近似保证。文献[74]证明了这个问题的目标函数模函数，并利用这个结果分别在线性阈值模型和独立级联模型下设计了一个近似比为$1-1/e$的贪婪近似算法。虽然这些贪婪算法在识别最具影响力节点集时有所突破，但非常耗时，并不适合大型网络。为了使算法扩展到大型网络中，文献[73]通过实验证明了启发式算法比贪婪算法耗时短，继而文献[75]和[76]分别提出了启发式的方法 SL-CRB 算法和 DI 算法，但它们都不能保证找到的节点集是具有最大影响力的。在这个问题中，还有一种比较便捷高效的方法，即直接寻找有公信力的媒体和有影响力的用户，如文献[23]提到了自媒体会扩大谣言的传播范围和影响。因此，类似于这样有影响力的媒体在第一时间发布权威公告，可有效避免谣言的产生和传播。

其二，在网络中播种若干传播真实信息的节点，增强真实信息在网络中的影响，使原本相信虚假信息的节点转变为相信真相的节点。He 等[77]将这个问题转化为“影响阻塞最大化(IBM)”问题，设计了一个近似比为$1-1/e$的贪婪算法和一个可扩展算法，构建了一个有向无环图(DAG)来估计节点的局部竞争影响，并为节点提供了有关传入影响扩散的决策能力。文献[78]提出了两种比贪婪算法耗时短、比其他启发式算法计算准确率高的算法 ContrId 和 ProxContrId。但它们的模型中只考虑了节点仅受网络中文本信息的影响，事实上，一些其他的现实社会因素也会影响用户的决策，如邻居节点之间的关系、亲密度等。

表 5 基于抗衡策略的控制方法总结

Table 5 Summary of control methods based on counterbalance strategies

Thesis	Network dataset	Information dissemination model	method	Node/edge selection basis	Algorithm	
					Greedy algorithm	Heuristic algorithm
文献[73]	Facebook Monterey Bay 2008 network	MCICM	识别有影响力的节点来传播真相	节点的状态、初始激活节点数量、边活跃度、节点的影响度	PHCA	—
文献[76]	Gnutella/ Facebook/ Amazon	CLTM	在被感染前使一些节点相信真相	节点的状态、阈值、边权重、边际收益最大化规则	DI	CELF
文献[77]	Mobile/ NetHEPT/ NetPHY	CLTM	寻找具有自身影响力的若干种子节点以传播真相	节点状态，边权重、阈值，正、负激活概率，正、负传播概率	CLDAG	—
文献[78]	Scale-free network/ Small-work network/ NetScience/US-power	LT1DT	在谣言节点附近播种若干真相节点	边权重、节点状态、影响阈值、传播阈值	—	ContrId/ ProxContrId

总的来说，这两种策略都是有效的，但阻断策略在实践中是不可行的。因为阻塞/删除部分节点/边会破坏网络的结构，并且移除关键节点会影响用户的体验，甚至违反道德标准，所以，目前比较推崇的策略是通过在网络中传播真实信息来打击虚假信息，最小化虚假信息的传播范围。

4.2 从源头上遏制虚假信息传播

目前，社交网络中虚假信息存在和泛滥的根本原因是，信息传播缺乏透明性、可追溯性以及用户没有有效的工具来评估信息的可信度和准确性，用户可能在不知情的情况下助长了虚假信息的传播。如果这些问题能够得到解决，从根本上改善社交网络的生态系统，就可以从源头上遏制虚假信息的传播。而区块链技术[79]的可追溯性、防篡改性及去中心化特征可有效地解决以上问题。目前，已有一些研究者提出利用区块链技术进行虚假信息传播控制的思想，大体可分为以下两类。

(1)通过有效地追踪虚假信息的来源，从而扼杀虚假信息，并通过对虚假新闻来源进行问责来使网络环境更加健康。文献[79]提出一个将区块链与分布式结构、一致性算法、智能契约相结合的系统，该系统使用源评估的方法确保新闻文章没有被篡改，读者可以使用区块链网络追溯文章的来源以评估文章。并且，该系统使用多节点验证来确保只有真实的新闻文章在网上传播。文献[80]利用区块链的信任机制，建立了一个可以显示任何数字媒体来源，包括虚假信息在试图误导用户时脱离上下文使用的图像的应用程序。该应用程序是一个早期的原型，当查找相似信息源头时能力有限，但研究者们提出可以使用 fisher 向量和 k-means 聚类方法及感知散列来改进原型。文献[81]将人工智能与区块链相结合，描述了一个社交网络接口及遏制虚假信息传播的模型，实现了一个检测和跟踪虚假新闻的系统。文献[82]创建了一个透明的新闻追踪系统，该系统具有区块链的所有优点，并由智能契约管理。但是，该系统存在几个局限性：一是它在一定程度上依赖于人，如果半数以上的参与者报告帖子为真，则该系统不能删除虚假文章，且无法阻止虚假信息实体的产生。因此，系统需要实现一个强大的过滤器来决定允许哪些审计人员进入。二是其只能追踪系统中创建的文章，无法追踪来自互联网的文章。三是每次创建交易时维护契约状态都很困难，并且该系统尚未在真正的区块链网络上部署，因此部署后的复杂性尚不清楚。

(2)使用激励机制来鼓励用户和新闻机构发布真实的信息。这种方法需要为每个新闻机构、普通用户或每条信息设置一个可信度分数，以激励用户创造真实的信息。目前多数研究是针对新闻机构的，文献[83]提出一个基于区块链的新闻机构，以及一个定制的授权证明(POA)算法，该算法采用动态加权排名评估分数的形式来产生可信度分数。首先，新闻机构请求加入该系统，然后在对当前映射进行初步检查之后，判断新闻机构是否成为认证节点。其次，用户被授予一个公钥和密钥对，该公钥和密钥对的状态为已认证或未认证的发布者，并且初始可信度分数将基于所发布的新闻而演变。信息的完整性和真实性通过语义相似度和默克树方法[83]来验证。类似地，文献[84]提出了授权证明(POA)算法，在新闻机构成为认证节点后，新闻机构可以选择请求发布信息，根据计算出的可信度分数，其中一些节点(如可信度高的新闻机构)可能会成为验证节点，负责验证。当信息内容被提交审查时，交易进入验证阶段，验证者指出其为真实或虚假，若满足系统设置的判定质量的特定阈值，则交易的散列被提交给区块链，从而为是否发布的决策提供了灵活性。

事实上，区块链技术由于自身的局限性以及现实社会中相关数据缺乏等问题，大多数模型还处于研发和实验阶段，因此，还需要更新的思想和方法来推动区块链技术在虚假信息控制这一领域的应用。

5 目前存在的问题及未来的研究方向

上文陈述了虚假信息的定义、特征、模型以及已有的检测和传播控制技术，本节将对前面研究中所存在的问题进行总结，并对未来的研究方向做出展望。具体地，从虚假信息研究领域所涉及的数据获取、特征提取、模型/技术等几个方面进行阐述。

5.1 数据获取方面

数据是一切研究的基础，社交网络中虚假信息检测及控制方法的研究必须建立在足够多的、有效的样本基础上。目前在数据获取方面存在的问题如下。

(1)收集数据耗时。社交网络平台和技术应用编程接口

的限制,使得收集数据的过程需要大量的等待时间。

(2)数据存储及处理困难。多媒体社交网络数据具有异构性、多样性及复杂性,且网络爬取数据质量较低并存在各种噪声(如广告类无用信息、个人生活类无核查价值的信息等)。

(3)现有公开数据集(如 BuzzFeedNews,LIAR,BS Detector,CREDBANK 等)均不能包含所有虚假信息样本特征,高质量的标记数据集的可访问性有限。

因此,缺乏高效的数据收集策略及包含完整特征的标记数据集成为该领域研究的一个重要难题。如何获取并创建一个样本特征全面、数据量足够多的高质量的虚假信息标记数据集以供该领域研究人员使用,是亟待解决的一个问题。

5.2　特征提取方面

特征提取通常是虚假信息检测及控制的首要工作。社交网络上的信息内容主要由文本、图像、视频这 3 类信息单独或以多模态形式呈现。目前在提取这几类信息特征时,存在的问题及未来的研究方向如下。

(1)文本信息。对特定领域提取的文本信息不适用于其他领域。因此,亟需研究能够捕捉虚假信息写作风格、跨话题、跨领域及跨语言等共性特征的技术。例如迁移学习技术能有效地学习可以进行知识迁移的信息,提取领域特有的特征和领域间共享的特征。

(2)图像及视频信息。由于这类信息主要是从物理、语义层面及基于神经网络来提取特征,因此还需要提取其他辅助特征(如用户传播时的行为特征)来提高检测准确率。

(3)多模态信息。目前无法预知虚假信息将包含哪些类型的图像或视频在社交网络中传播,因此,如何提取适应于标记多模态信息的普适特征来进行有效的检测是值得研究的问题。

5.3　传播建模方面

信息传播建模的主要目标是分析说明信息扩散过程,前文介绍了信息解释模型和信息预测模型并不是相互独立的,而是相辅相成的。为了使信息传播模型更加符合真实社交网络中虚假信息的传播情况,建模过程中还需要考虑以下影响因素。

(1)用户的个人情感。社交网络为用户提供了一种情感的表达方式,具有情感内容的信息传播会对现实社会产生巨大的影响。目前仅有少量的传播模型带有情感分析,且很不完善,只分析了快乐、愤怒等类型的情感信息。未来,还应分析更深层次的情感,构建一个良好、自适应的基于情感分析的系统以监控信息传播,从而实现虚假信息的早期防控。

(2)社交群体状态与网络结构。为了解决现有传播模型中节点只关注周围节点的群体状态以及节点只受到网络结构中某个阈值影响的问题,需要将两种因素结合起来,构建兼顾用户主观性及信息特性的信息传播模型。

5.4　模型/技术方面

5.4.1　虚假信息检测

虚假信息检测领域的一个关键挑战是:如何在传播早期阶段高效地检测出虚假信息。虽然目前已有对虚假信息早期阶段检测的研究,但碍于数据获取及特征提取方面的问题,此类检测模型的时间效率和准确率还有待提高。具体存在的问题、改进方法及未来发展方向如下:

(1)对于检测性能较低的早期虚假信息分类器,可借助集成学习的思想将一类或几类分类器(如都是决策树分类器,或训练集、测试集使用不同的分类器)使用迭代算法(如 Boosting、Bagging 系列算法)集合构成一个强分类器,加大分类误差率小(如小于平均分类误差率)的弱分类器权重,从而提高检测的准确率。

(2)对于多维特征的处理,可使用聚类的方法对多维特征加以简化提取突出特征。例如,使用层次聚类的方法将这些特征抽象为在一个 k 维空间里的 n 个点,首先使用距离公式计算每个点到其他点的距离;其次将具有最短距离的两点归为一类,此距离即为聚类图中两点的共同平台高度;然后计算各类间的距离;最后循环计算各特征、类间距离,直至所有特征聚为一类,输出聚类图。在聚类图中选取一个固定平台高度即可完成对聚类类别的选取,即完成简化提取突出特征。

(3)对于判别部分真实、部分虚假的信息时,使用概率图模型来预测信息类别标签的概率分布,即信息为假的概率,比直接判为虚假更有意义。例如,在基本概率图模型的基础上,融合信息文本、社交网络等特征构建贝叶斯图模型,通过可观测到的单词、文档等变量来推测类别标签的概率分布。贝叶斯图模型可忽略数据缺失、文本短小、语法不规则等影响,增强了文本单词间的关联性。

(4)对于多源多模态数据的检测,研究基于多模态机器学习以实现跨平台的虚假信息检测也是未来的一个重要研究方向。

5.4.2　虚假信息控制

基于阻断和抗衡策略的控制方法的关键问题是如何识别出关键节点集或边集,目前存在的问题及解决方法如下:

(1)关键节点/边的选取依据未考虑到更深层、更广泛的特征。目前多数研究人员将节点的度,节点的中心性、边权重、传播阈值、网络特征值等组合作为选取依据,未来可增添一些更具有实际意义的因素,如节点传递性,节点间关系强度,节点间交互信息(时间点、次数)等可作为关键节点/边的选取依据。

(2)求解此类问题的最优解已被证明是 P-难或 NP-难的。目前多数研究人员使用贪婪算法求解近似最优解来解决此类问题,也有研究人员提出启发式的方法来缓解计算时间和空间的问题,但效果并不理想,且不适用于大型社交网络。有效的解决方案是通过降维的手段降低问题的复杂度;或将以往对节点/边赋予一个准确度量值作为最优解的判断方式,转变为基于概率的方法对节点/边的关键性进行排序的方式,以避开问题复杂度的问题。如使用信息融合技术中的证据理论方法,赋予节点/边关键性更合理、准确的序列。对于如何改进其中的组合规则[85],使其在现有特征条件较弱的情况下给出决策结果,以提出更有效的虚假信息控制方法是我们下一步的研究计划。

另一类基于区块链技术的控制方法目前存在的问题如下:

(1)存在人为作弊的可能。使用区块链技术构建的系统在一定程度上依赖于人工的审核和界定,若利用大数据和机

器学习来替代人工,也必然面临检测准确性的问题,无法从源头上遏制虚假信息的传播。

(2)应用上存在局限性。这类方法仅能对区块链系统中创建的新闻进行有效控制,无法追踪来自更广阔的互联网的新闻。

(3)数据的存储及交互效率不高。区块链架构的运行机制将导致此类问题在社交网络的大数据流量下变得愈发突出。

未来,底层基于区块链架构的社交媒体平台也许可以进一步解决网络空间中虚假信息的产生和传播问题,但在此架构下,依然需要解决人工智能替代人工进行信息审核以及效率的问题,并且需要进行量子计算时代的区块链加密技术的研究,以保证区块链的安全性。

结束语 随着社交网络的普及,虚假信息传播造成的社会影响越来越不可忽视,如何准确监测并有效遏制其传播是很多研究人员致力的方向。本文以虚假信息中最具有代表性的假新闻和谣言为例,给出了其定义、传播过程、特征及传播模型;继而阐述分析了已有研究中传统的和基于机器学习、深度学习的检测方法,以及从传播过程和源头上遏制虚假信息传播的不同种类的控制方法;最后,分别从数据获取方面、特征提取方面、传播建模方面以及模型/技术方面对该领域的研究成果进行了总结,表明了构建全样本特征的数据库,研究虚假信息传播早期阶段的检测模型及融合多模态特征的检测模型,以及将区块链技术与人工智能技术相结合以设计出更安全、更值得信任的社交网络系统等是未来的重要研究方向。

参 考 文 献

[1] VOSOUGHI S, ROY D, ARAL S. The spread of true and false news online[J]. Science, 2018, 359(6380): 1146-1151.

[2] FALLIS D. What is Disinformation? [J]. Library Trends, 2015, 63(3): 401-426.

[3] KUMAR S, SHAH N. False Information on Web and Social Media: A Survey [EB/OL]. (2018-04-23) [2021-02-25]. https://arxiv.org/pdf/1804.08559.

[4] BONDIELLI A, MARCELLONI F. A survey on fake news and rumour detection techniques[J]. Information Sciences, 2019, 497: 38-55.

[5] HORNE B D, ADALI S. This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news[EB/OL]. (2017-03-28) [2021-02-25]. https://arxiv.org/pdf/1703.09398.

[6] KUMAR S, WEST R, LESKOVEC J. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes [C]// Proceedings of the 25th International Conference on World Wide Web. 2016.

[7] MATSUBARA Y, SAKURAI Y, PRAKASH B A, et al. Rise and fall patterns of information diffusion: Model and implications[C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12). New York, NY: ACM Press, 2012: 6-14.

[8] LIU Y Z, WANG J, PAN X Z, et al. Research on scale-free network rumor propagation under the influence of nodes[J]. Small and Microcomputer System, 2018, 39(11): 2375-2379.

[9] FOSTER E K, ROSNOW R L. Gossip and Network Relationships[M]// Relating difficulty: The processes of constructing and managing difficult interaction. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2006: 161-180.

[10] ZENG L, STARBIRD K, SPIRO E S. Rumors at the Speed of Light? Modeling the Rate of Rumor Transmission During Crisis [C]// Proceedings of the 49th Annual Hawaii International Conference on System Sciences (hicss 2016). Los Alamitos, CA: IEEE Computer Society, 2016: 1969-1978.

[11] FRIGGERI A, ADAMIC L, ECKLES D, et al. Rumor Cascades [C]// Proceedings of the International AAAI Conference on Web and Social Media. Palo Alto, CA: AAAI Press, 2014: 101-110.

[12] 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-WWW'13 Companion. New York, NY: ACM Press, 2013: 729-736.

[13] SHAO C, CIAMPAGLIA G L, FLAMMINI A, et al. Hoaxy: A Platform for Tracking Online Misinformation[C]// Proceedings of the 25th International Conference Companion on World Wide Web-WWW'16 Companion. New York, NY: ACM Press, 2016: 745-750.

[14] ZUBIAGA A, LIAKATA M, PROCTER R, et al. Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads [J]. PLOS ONE, 2016, 11(3): e0150989.

[15] SUDBURY A. The proportion of the population never hearing a rumour[J]. Journal of applied probability, 1985, 22(2): 443-446.

[16] HURLEY M, JACOBS G, GILBERT M. The basic SI model [J]. New Directions for Teaching and Leaming, 2006, 106(6): 11-22.

[17] PASTOR-SATORRAS R, VESPIGNANI A. Epidemic dynamics and endemic states in complex networks[J]. Physical Review E, 2001, 63(6): 066117.

[18] MORENO Y, PASTOR-SATORRAS R, VESPIGNANI A. Epidemic outbreaks in complex heterogeneous networks[J]. The European Physical Journal B, 2002, 26(4): 521-529.

[19] JIN Y, WANG W, XIAO S. An SIRS model with a nonlinear incidence rate[J]. Chaos Solitons Fractals, 2007, 34, 1482-1497.

[20] SANG C, LI T, TIAN S, et al. SFTRD: A novel information propagation model in heterogeneous networks: Modeling and restraining strategy[J]. Physica A: Statistical Mechanics and its Applications, 2019, 524: 475-490.

[21] ZHANG N, HUANG H, SU B, et al. Dynamic 8-state ICSAR rumor propagation model considering official rumor refutation [J]. Physica A: Statistical Mechanics and its Applications, 2014, 415: 333-346.

[22] RUAN Z, YU B, SHU X, et al. The impact of malicious nodes on the spreading of false information[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2020, 30(8): 083101.

[23] LI J, SONG B, LUO C, et al. Considering Self-media Influence

Network Rumor Propagation Model and Control Strategy[C]// 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). New York, NY: IEEE, 2020: 1407-1411.

[24] GOLDENBERG J, LIBAI B, MULLER E. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth[J]. Marketing Letters, 2001, 12(3): 211-223.

[25] GRANOVETTER M. Threshold Models of Collective Behavior [J]. American Journal of Sociology, 1978, 83(6): 1420-1443.

[26] CAMERER C F. Behavioral game theory: Experiment in strategic interaction[J]. Socio-Econom, 2003, 32: 135-146.

[27] ZHAO Z, CHEN X. Propagation Model of Derivative Rumor Considering Propagation Error and Malicious Tampering[C]// 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA). New York, NY: IEEE, 2019: 241-245.

[28] HANG Q F, ZHU J M, SONG B, et al. Game model of information transmission in social networks[J]. Journal of Chinese Computer Systems, 2014, 35: 473-477.

[29] WANG Y, YU J, QU W, et al. Evolutionary game model and analysis methods for network group behavior[J]. Chin. J. Comput. 2015, 38: 282-300.

[30] ZHOU X, ZAFARANI R. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities[J]. ACM Computing Surveys, 2020, 53(5): 1-40.

[31] MITRA T, GILBERT E. CREDBANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations[C]// Proceedings of the International AAAI Conference on Web and Social Media. Palo Alto, Calif: AAAI Press, 2015.

[32] 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. Stroudsburg, PA: ACL, 2011: 1589-1599.

[33] 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. Piscataway, NJ: IEEE, 2013: 1103-1108.

[34] RASHKIN H, CHOI E, JANG J Y, et al. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2017: 2931-2937.

[35] 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. New York, NY: ACM, 2015: 1751-1754.

[36] 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. Piscataway, NJ: IEEE, 2015: 651-662.

[37] LIU Y, XU S. Detecting rumors through modeling information propagation networks in a social media environment[J]. IEEE Transactions on Computational Social Systems, 2016, 3(2): 46-62.

[38] MANISH G, ZHAO P X, HAN J W. Evaluating event credibility ontwitter[C]// Proceedings of the 2012 SIAM International Conference on Data Mining. Philadelphia, PA: SIAM, 2012: 153-164.

[39] MA J, GAO W, WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]// 55th Annual Meeting of the Association-for-Computational-Linguistics (ACL). 2017: 708-717.

[40] VOLKOVA S, JANG J Y. Misleading or falsification: Inferring deceptive strategies and types in online news and social media [C]// Companion Proceedings of the The Web Conference 2018. New York, NY: ACM, 2018: 575-583.

[41] JOOYEON K, DONGKWWA K, ALICE O. Homogeneity-based transmissive process to model true and false news in social networks[C]// Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. New York, NY: ACM, 2019: 348-356.

[42] POTTHAST M, KIESEL J, REINARTZ K, et al. A Stylometric Inquiry into Hyperpartisan and Fake News[EB/OL]. (2017-02-18) [2021-02-25]. https://arxiv.org/pdf/1702.05638.

[43] VEDOVA M L D, TACCHINI E, MORET S, et al. Automatic online fake news detection combining content and social signals [C]// 2018 22nd Conference of Open Innovations Association (FRUCT). Piscataway, NJ: IEEE, 2018: 272-279.

[44] SHU K, WANG S, 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. New York, NY: ACM, 2019: 312-320.

[45] POUYANFAR S, SADIQ S, YAN Y, et al. A Survey on Deep Learning: Algorithms, Techniques, and Applications[J]. ACM Comput. Surv., 2018, 51(5).

[46] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. NATURE, 2015, 521(7553): 436-444.

[47] LIU Y, WU Y F. Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks[C]// Thirty-Second AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018: 354-361.

[48] YU F, LIU Q, WU S, et al. Attention-based Convolutional Approach for Misinformation Identification from Massive and Noisy Microblog Posts[J]. Computers & Security, 2019, 83: 106-121.

[49] MONTI F, FRASCA F, EYNARD D, et al. Fake News Detection on Social Media using Geometric Deep Learning[EB/OL]. (2019-02-10) [2021-02-25]. https://arxiv.org/pdf/1902.06673.

[50] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]// International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI

Press, 2016: 3818-3824.

[51] RUCHANSKY N, SEO S, LIU Y. CSI: A hybrid deep model for fake news detection[C]// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. New York, NY: ACM, 2017: 797-806.

[52] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[53] CHO K, VAN M B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB/OL]. (2014-09-03) [2021-02-25]. https://arXiv.org/pdf/1406.1078.

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

[55] KULA S, CHORAŚ M, KOZIK R, et al. Sentiment Analysis for Fake News Detection by Means of Neural Networks[C]// Computational Science (ICCS 2020). Cham: Springer International Publishing, 2020: 653-666.

[56] QI P, CAO J, YANG T, et al. Exploiting Multi-domain Visual Information for Fake News Detection[C]// 2019 IEEE International Conference on Data Mining (ICDM). Piscataway, NJ: IEEE, 2020: 8-11.

[57] UPPAL A, SACHDEVA V. Fake news detection using discourse segment structure analysis[C]// 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). Piscataway, NJ: IEEE, 2020: 751-756.

[58] GUO H, CAO J, ZHANG Y Z, et al. Rumor detection with hierarchical social attention network[C]// The 27th ACM International Conference on Information and Knowledge Management. New York, NY: ACM, 2018: 943-951.

[59] WANG S, ZHAO X, CHEN Y, et al. Negative Influence Minimizing by Blocking Nodes in Social Networks[C]// Proceedings of the 17th AAAI Conference on Late-Breaking Developments in the Field of Artificial Intelligence. Palo Alto, CA: AAAI Press, 2013: 134-136.

[60] YAN R, LI D, WU W, et al. Minimizing Influence of Rumors by Blockers on Social Networks: Algorithms and Analysis[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(3): 1067-1078.

[61] FAN L, LU Z, WU W, et al. Least Cost Rumor Blocking in Social Networks[C]// 2013 IEEE 33rd International Conference on Distributed Computing Systems. Piscataway, NJ: IEEE, 2013: 540-549.

[62] FAN L, WU W, ZHAI X, et al. Maximizing rumor containment in social networks with constrained time[J]. Social Network Analysis and Mining, 2014, 4(1): 214.

[63] PHAM C V, THAI M T, DUONG H V, et al. Maximizing misinformation restriction within time and budget constraints[J]. Journal of Combinatorial Optimization, 2018, 35(4): 1202-1240.

[64] PHAM C V, PHU Q V, HOANG H X, et al. Minimum budget for misinformation blocking in online social networks[J]. Journal of Combinatorial Optimization, 2019, 38(4): 1101-1127.

[65] KIMURA M, SAITO K, MOTODA H. Minimizing the Spread of Contamination by Blocking Links in a Network[C]// Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2008: 1175-1180.

[66] ZHANG H F, LI K Z, FU X C, et al. An Efficient Control Strategy of Epidemic Spreading on Scale-Free Networks[J]. Chinese Physics Letters, 2009, 26(6): 068901.

[67] KIMURA M, SAITO K, MOTODA H. Solving the contamination minimization problem on networks for the linear threshold model[C]// Pacific Rim International Conference on Artificial Intelligence. Heidelberg, Berlin: Springer, 2008: 977-984.

[68] KUHLMAN C J, TULI G, SWARUP S, et al. Blocking simple and complex contagion by edge removal[C]// 2013 IEEE 13th International Conference on Data Mining. Piscataway, NJ: IEEE, 2013: 399-408.

[69] YAO Q, ZHOU C, XIANG L, et al. Minimizing the negative influence by blocking links in social networks[C]// International Conference on Trustworthy Computing and Services. Heidelberg, Berlin: Springer, 2014: 65-73.

[70] KHALIL E B, DILKINA B, SONG L. Scalable Diffusion-Aware Optimization of Network Topology[C]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM Press, 2014: 1226-1235.

[71] NI P K, ZHU J M. Study on the Edge Blocking Strategy of Minimizing the amount of False Information Interaction in Online Social Networks[EB/OL]. (2020-05-18). https://doi.org/10.16381/j.cnki.issn1003-207x.2019.Y-01.

[72] BASARAS P, KATSAROS D, TASSIULAS L. Dynamically blocking contagions in complex networks by cutting vital connections[C]// 2015 IEEE International Conference on Communications (ICC). Piscataway, NJ: IEEE, 2015: 1170-1175.

[73] BUDAK C, AGRAWAL D, ELABBADI A. Limiting the Spread of Misinformation in Social Networks[C]// Proceedings of the 20th International Conference on World Wide Web. New York, NY: ACM Press, 2011: 665-674.

[74] LI S, ZHU Y, LI D, et al. Rumor restriction in Online Social Networks[C]// 2013 IEEE 32nd International Performance Computing and Communications Conference (IPCCC). Piscataway, NJ: IEEE, 2014: 6-8.

[75] PING Y, CAO Z, ZHU H. Sybil-aware least cost rumor blocking in social networks[C]// 2014 IEEE Global Communications Conference. Piscataway, NJ: IEEE, 2014: 692-697.

[76] ZHANG H, ZHANG H, LI X, et al. Limiting the Spread of Misinformation While Effectively Raising Awareness in Social Networks[C]// Computational Social Networks. Cham: Springer International Publishing, 2015: 35-47.

[77] HE X, SONG G, CHEN W, et al. Influence Blocking Maximization in Social Networks under the Competitive Linear Threshold

Model[C]// Proceedings of the 2012 SIAM International Conference on Data Mining. Philadelphia,PA:SIAM,2012:463-474.

[78] YANG L,LI Z,GIUA A. Containment of rumor spread in complex social networks[J]. Information Sciences,2020,506:113-130.

[79] SHANG W,LIU M,LIN W,et al. Tracing the source of news based on blockchain[C]// 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS). Piscataway,NJ:IEEE,2018:377-381.

[80] HUCKLE S,WHITE M. Fake News:A Technological Approach to Proving the Origins of Content,Using Blockchains [J]. Big Data,2017,5(4):356.

[81] SHAE Z,TSAI J. AI blockchain platform for trusting news [C]// 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS). Piscataway, NJ: IEEE, 2019: 1610-1619.

[82] BABAR A,SHUKLA A,JAGTAP N,et al. News Tracing System Using Blockchain [J]. International Journal of Engineering Applied Sciences and Technology,2020,5(2):554-562.

[83] GUEGAN D. Public blockchain versus private blockhain [J]. Université Paris Panthéon-Sorbonne (Post-Print and Working Papers),HAL,2017:halshs-01524440.

[84] CHEN Q,SRIVASTAVA G,PARIZI R M,et al. An incentive-aware blockchain-based solution for internet of fake media things[J]. Information Processing & Management,2020,57(6):102370.

[85] WANG J,QIAO K Y,ZHANG Z Y,et al. An improvement for combination rule in evidence theory[J]. Future Generation Computer Systems,2019:1-9.

WANG Jian, born in 1978, Ph.D, professor, is a member of China Computer Federation. Her main research interests include multimedia social networks, information security, trusted computing and usage control.