

# 社交媒体中错误信息的检测方法研究述评

吴诗苑<sup>1</sup>, 董庆兴<sup>2,3</sup>, 宋志君<sup>1</sup>, 张 斌<sup>4</sup>

(1. 华中师范大学信息管理学院, 武汉 430079; 2. 武汉大学新闻与传播学院, 武汉 430072;

3. 武汉大学大数据研究院, 武汉 430072; 4. 南京大学信息管理学院, 南京 210023)

**摘 要** 社交媒体改变了人们获取信息的方式, 但也助长了错误信息在网络上的产生和传播。如何准确且快速地检测出社交媒体上的错误信息以净化网络环境, 是一个重要的研究议题。本文从信息生态理论出发, 分别从内容、用户和传播三个角度来阐述当前错误信息检测所关注的问题以及对应的检测方法, 对近年来国内外的相关研究成果进行了系统梳理。现有的检测方法已经利用深度学习等技术取得了较好的检测结果。但是, 由于错误信息爆发初期相关数据较少, 有关早期检测的研究尚不多见; 能够实现有效迁移和预训练的大规模基准数据集仍有待构建; 从用户入手的信息挖掘有待进一步深入研究。

**关键词** 社交媒体; 错误信息; 深度学习; 特征融合; 自动检测

## Misinformation Detection in Social Media

Wu Shiyuan<sup>1</sup>, Dong Qingxing<sup>2,3</sup>, Song Zhijun<sup>1</sup> and Zhang Bin<sup>4</sup>

(1. School of Information Management, Central China Normal University, Wuhan 430079; 2. School of Journalism and Communication, Wuhan University, Wuhan 430072; 3. Big Data Institute, Wuhan University, Wuhan 430072;

4. School of Information Management, Nanjing University, Nanjing 210023)

**Abstract:** Social media has dramatically improved the efficiency of information access, but it has also contributed to the generation and dissemination of misinformation on the Internet. The accurate and quick detection of misinformation to improve the online information environment is an important issue. Inspired by the Information Ecology Theory, this paper expounds on the current problems and relevant detection methods of misinformation from the three perspectives of content, users, and dissemination. Existing detection methods have achieved state-of-the-art results by deep learning methods. However, because of the lack of relevant data in the early stages, studies on the early detection of misinformation are still rare. Additionally, large-scale benchmark datasets for transfer learning and pre-training tasks are yet to be constructed. Moreover, information mining from users needs to be further evaluated.

**Key words:** social media; misinformation; deep learning; feature fusion; automatic detection

## 1 引 言

在“万物皆媒”的新媒体时代, 人们在社交媒

体上的活动不限于接收与传播信息, 同时也在生产、发布信息, 其中, 一些不负责任的用户会有意无意地散布一些错误信息<sup>[1]</sup>。错误信息的泛滥增加

收稿日期: 2021-08-23; 修回日期: 2022-03-31

基金项目: 国家自然科学基金面上项目“面向群智感知大数据的群体评价模型与方法研究”(71871102), “基于注意力机制的学术信息动态推荐研究”(72074109)。

作者简介: 吴诗苑, 女, 1998年生, 硕士研究生, 研究方向为自然语言处理; 董庆兴, 男, 1988年生, 副教授, 研究方向为计算传播, E-mail: qxdong@whu.edu.cn; 宋志君, 女, 1997年生, 硕士研究生, 研究方向为用户信息行为; 张斌, 男, 1984年生, 副教授, 研究方向为大数据与数据挖掘。

了信息环境中的“噪音”和“污染”，降低了信息环境中的信噪比，给信息生态的平衡、稳定和有序造成了严重的威胁<sup>[2]</sup>。然而，与生产和传播错误信息相比，验证一条信息真伪的成本高昂，成本之间的不对称导致错误信息在互联网上很容易蔓延开来。因此，如何利用已有的技术和方法在新媒体环境下低成本、高效率地检测出错误信息是学界和业界都广泛关注的一个重要议题。图书情报领域的学者已经对错误信息的传播、演化和模型仿真有了深入的研究<sup>[3]</sup>，近年来开始在错误信息检测这一方向上取得突破<sup>[4-6]</sup>。综合考虑网络环境、用户和错误信息的传播机制，学者们运用机器学习和深度学习等方法提出了错误信息甄别、信息危机预警和早期检测等一系列具有实用价值的模型，并在测试中取得了不错的效果，给错误信息的检测提供了新思路。

所谓错误信息（misinformation）有狭义和广义之分，如图1所示。狭义的错误信息是指无意传播的假的或者不准确的信息<sup>[7]</sup>。这类信息的传播动机不是为了欺骗接收者，可能只是传播者出于对信息源的信任，这种被信任的信息源通常是传播者的家人、朋友或社交网络上的意见领袖<sup>[8]</sup>。传播者通常持分享、关心或娱乐的动机将潜在的问题或情况告知他们的朋友，以此起到提醒和共鸣的作用<sup>[9]</sup>。因为传播者是出于好意，所以，这类错误信息通常也被称作“诚实的错误”<sup>[10]</sup>。与狭义的错误信息相对应的一个术语叫“虚假信息（disinformation）”，也就是刻意制造并故意传播的假的或者不准确的信息<sup>[11-13]</sup>，这类信息是恶意用户为了获得不正当利益而故意制造的，专门用来误导大众，引起公众的焦虑情绪<sup>[14]</sup>。虽然错误信息和虚假信息都是假的、不准确的，但它们之间最大的区别在于主观意图——虚假信息是被恶意制造并传播的，但错误信息的制造与传播并非出于恶意<sup>[15]</sup>。而广义的错误信息则是指所有假的、不准确的或未经证实的信息<sup>[16-17]</sup>。广义定义并不区分传播者的动机，只要信息与真实情况不符或与现实证据冲突都可以被视为错误信息<sup>[18]</sup>。因此，在广义层面上，错误信息作为一个总括术语包含了虚假信息在内的谣言、假新闻、阴谋论等所有可能造成严重后果的不实信息<sup>[19]</sup>。错误信息检测的目的在于找到一种方法来有效地辨别信息内容的真假，而对于传播者的意图，不论是研究者还是平台管理员通常都难以确定。因此，广义定义在错误信息检测研究中更加常用<sup>[20]</sup>。本文主要关注散布在社交媒体上的不准确信息，循例使用错误信

息的广义定义。

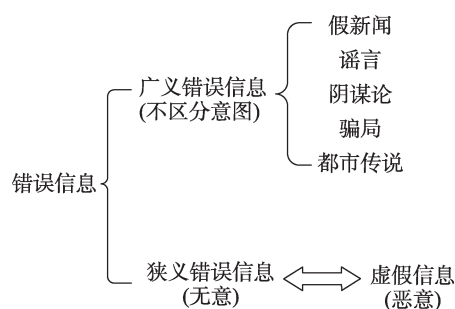


图1 错误信息的分类

错误信息的检测方法可以从不同的角度划分成不同的种类。由于错误信息的产生与传播不仅限于信息本身，更是一个牵涉到人和环境的复杂事件，因此有必要从一个整体的框架来梳理当前错误信息的检测方法。信息生态学则提供了在一个完整的信息生态中分析信息、信息人、信息环境三者互相影响、共同演化的视角，如图2所示。信息生态理论指出，信息生态是由信息、信息人和信息环境中各种要素和资源共同构成的整体<sup>[21]</sup>。各个组成部分之间存在强大的相互联系和依赖：信息在传播时，越来越多信息人的参与会不断地向信息中加入新内容或改变原内容，在这个过程中，信息不仅会受到信息人的价值和偏好的影响，还会受到信息环境的干扰，使得信息量增减并存并产生价值变异<sup>[22]</sup>；信息人在受到信息环境的刺激后通过自身的能动性加工发生了信息的生产、组织、传递、消费、分解等基本行为，这些行为会反过来作用于信息环境<sup>[23]</sup>；信息环境是一切信息人发生信息交互行为的场所，信息环境的改变会影响信息人的行为和信息传播的方式<sup>[24]</sup>。由此可见，信息生态中每种要素的变化都是系统性的，一个要素发生变化就会影响整个信息生态系统<sup>[25]</sup>。因此，从信息生态的角度来讨论错误信

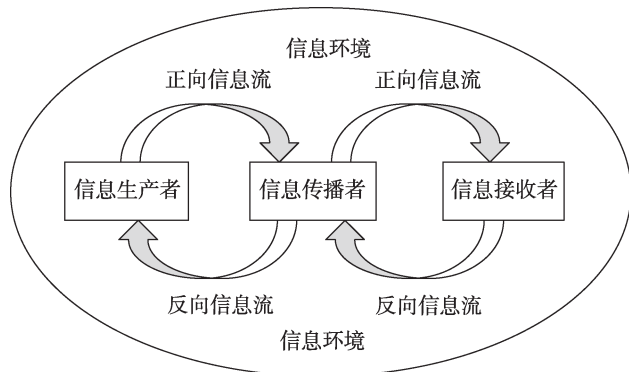


图2 信息生态基本结构<sup>[23]</sup>

息检测可以得到一个相对完整的分析框架。

本文对照信息生态中的各个要素,从内容、用户和传播这三个角度对错误信息检测方法进行综述。具体地,基于内容的检测方法主要从信息内容本身入手,直接捕获信息传达的思想和情感,并以此为依据来进行识别。基于用户的检测是以信息人在信息生态中扮演的角色以及信息人的行为为依据进行检测。基于传播的检测则是根据错误信息传播的时间特征和网络分布特征进行检测。这种在信息生态背景下分析错误信息检测问题的论述方式既涵盖了信息本身,又考虑了信息人与信息环境之间的交互背景,有助于更加系统、全面地了解各种因素对错误信息产生的影响,从而更有针对性地梳理当前关于错误信息检测方法的研究。

本文把从中国知网、万方数据、Web of Science、DBLP (DataBase Systems and Logic Programming) 以及 Google Scholar 中检索到的论文作为主要数据源。中文论文检索以“错误信息/虚假信息/假新闻/阴谋论/谣言”为主要关键词,分别将这些关键词与“检测/识别”搭配检索;英文论文检索以“misinformation/disinformation/false information/fake news/conspiracy/rumor”为主要关键词,并分别将这些关键词与“detection/recognition/identification”搭配检索。检索时除了考虑关键词之外,还考虑了其同义词、近义词、相关词、下位词等,不限制检索的相关起始时间,将截止时间设定为2021年3月3日,综合考虑主题相关性后共得到1012篇文献。基于内容、基于用户和基于传播的三种检测方法下的相关论文数量对比如图3所示。由于一些论文同时使用了多种检测方法,因此三种检测方法的论文数量总和大于1012篇。总体而言,在文献数量方面,基于内容检测的文献数量要远高于另外两种检测方法的文献数量,这可能是因为文本内容是最直观的,内容特征是最容易获取和处理的。

## 2 基于内容的错误信息检测

在社交媒体上识别错误信息的一个最直接的方法是检测包括实体和非实体在内的信息内容<sup>[26]</sup>。其中,实体内容是信息的载体和形式,包括标题、正文、图像等。由于实体内容具有直观性,它是判断信息真实性的重要依据,检测者可以根据检测需要对其进行一系列的筛选、提取、编码等操作<sup>[27]</sup>。而非实体内容则是作者表达的观点、态度和情感等隐含信息,它是判断信息真实性的主要着力点<sup>[28]</sup>。错

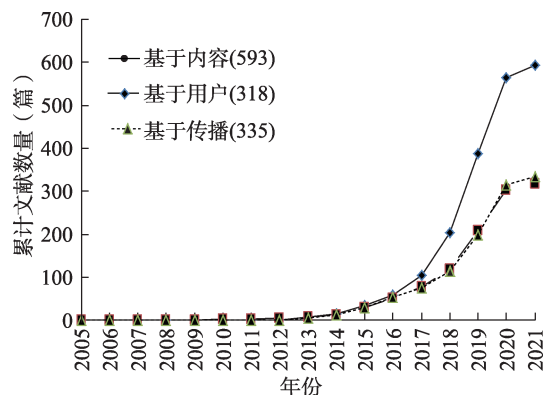


图3 三种检测方法下的论文数量对比

误信息大多是为了误导公众和引起人们的注意而编造的,往往与真实信息有着不同的行文模式<sup>[29]</sup>。比如,内容真实的新闻往往含有更多的链接且文本长度通常大于假新闻<sup>[30]</sup>;假新闻中通常含有更多的否定词、消极词和夸张表达,而内容真实的新闻中则有更多的认知行为词和推断行为词<sup>[31]</sup>。

根据这些特点,Castillo等<sup>[29]</sup>利用统计特征,如标点符号、表情、超链接等出现的次数,使用决策树模型来简单评估推文真实性。考虑到假新闻和垃圾邮件在行文模式上的相似性,使用垃圾邮件分类技术(如朴素贝叶斯分类器)也可有效检测假新闻<sup>[32]</sup>。随着语言特征提取技术的发展,更多精准、高效的提取技术被有效地利用于该领域<sup>[33]</sup>。例如,Rashkin等<sup>[34]</sup>使用一种复杂的语法信息与心理语言特征工具LIWC (linguistic inquiry and word count) 设计了多种语言特征,结合长短期记忆网络(long short term memory, LSTM)取得了不错的检测效果。Ahmed等<sup>[35]</sup>通过TF-IDF提取新闻文本特征并使用支持向量机(support vector machine, SVM)对社交媒体上的新闻进行分类,准确率高达92%。但是,上述模型过于依赖训练数据中的先验分布,对新话题和新事件的应用效果不理想。针对该问题,学界开始研究更具有话题迁移性的检测方法<sup>[36]</sup>。例如,将假新闻检测视为一种链接预测任务,或是通过分析知识图上相关概念之间的路径长度来评估陈述的真实性,两种方法都可在新话题和新事件上正常运行<sup>[37-38]</sup>。Guacho等<sup>[39]</sup>则通过半监督方法,在构建的K-NN (k-nearest neighbor) 图上运用FaBP (fast belief propagation) 传播已知标签以确定未知标签,比完全监督模型实现了更高的检测精度。

伴随着深度学习的热潮,深度学习强大的学习能力和特征表示能力引起了广泛的关注,逐渐被应用到错误信息检测中。Ma等<sup>[40]</sup>首次使用循环神经



网络 (recurrent neural network, RNN) 对特征进行学习, 在 Twitter 和 Weibo 数据集上均达到 80% 以上的准确率。楼靓<sup>[41]</sup>改进卷积神经网络 (convolutional neural network, CNN) 得到 TextCNN 模型, 该模型可结合字粒度的文本语义向量和句粒度的文本统计特征, 提升了模型的识别精度。针对传统检测模型需要大量统计特征以及难以实现及时检测的问题, 李悦晨等<sup>[5]</sup>使用 BERT (bidirectional encoder representation from transformers) 对微博原文进行向量表示, 然后将获取的语义特征输入 RCNN (region-CNN) 模型中, 得到了较高的检测速度和准确率。深度学习模型在错误信息自动检测领域中的应用解决了传统特征工程方法的不足, 提高了检测结果的可信度。

然而, 上述检测方法都是针对文本的检测, 无法识别由图像构成的信息。在同一则消息中, 图像和文本作为不同模态的信息联系紧密, 同时使用不同模态的数据可以提高数据的利用率<sup>[42]</sup>, 由此产生了基于多模态数据的错误信息检测方法<sup>[43-46]</sup>。例如, 将与事件相关的图像特征和文本特征进行融合检测<sup>[47]</sup>; 或是将图像映射为语义标签, 通过计算图像与文本的语义一致性来检测图文不匹配类型的谣言<sup>[48]</sup>。基于多模态检测的思路, 刘金硕等<sup>[49]</sup>通过提取文本、图像和图像内嵌文本特征设计出多模态网络谣言检测模型 MRSD, 与基准模型相比, 该模型在准确率和 F1 值上都有显著提高。然而, 仅实现特征融合是不够的, 模型应更具有泛化能力, 可以学习与事件无关的特征。Wang 等<sup>[50]</sup>提出的基于对抗网络的端到端模型将 VGG19 (visual geometry group-19) 提取的视觉特征和 TextCNN 提取的文本特征相融合, 可在判断新闻真实性的同时判断新闻来自哪一类事件。为了提高模型检测的准确率, Khattar 等<sup>[51]</sup>将文本、视觉的拼接特征编码为一个中间表达, 利用重构损失保证编码后的中间表达可以解码回原状态, 并用该中间表达的向量做虚假新闻检测。

综上所述, 基于内容的检测主要是借助 SVM、决策树等判别模型以及 CNN、RNN 等神经网络模型从语法和语义层面展开的<sup>[52]</sup>。与纯文本的错误信息相比, 带有多媒体内容的错误信息具有更快的传播速度<sup>[53]</sup>。由此提出的多模态错误信息检测方法在提高数据利用率的同时还提高了模型的鲁棒性。此外, 即使其中一种模态数据缺失, 多模态检测方法仍然可以运行。因此, 在未来的研究中应当更加重视对多模态信息的利用并不断提高多模态特征的提取效率和准确率, 以更有效地检测错误信息。

### 3 基于用户的错误信息检测

社交媒体的开放性让许多用户从信息的被动接收者变成内容的主动贡献者。如果将用户理解为数据感知器, 那么社交媒体就是由用户所见、所想、所听、所说组成的<sup>[54]</sup>。因此, 利用用户画像抽取用户特征有利于提高错误信息检测的准确率。从不同的粒度看, 用户特征可以分为两大类: 个体特征和群体特征<sup>[55]</sup>。针对错误信息检测这一特定任务来说, 个体特征一般是指从单个用户的各项统计指标中抽取的用来分析用户可信度的一系列特征, 包括用户的 ID、性别、年龄、粉丝数、注册时间、地理位置等<sup>[56]</sup>; 群体特征则是指在信息传播的过程中具有相似属性的用户所拥有的整体特征, 通常由个体特征聚合而来<sup>[57]</sup>, 如平均粉丝数、认证用户的比例等, 群体特征的基本假设是传播错误信息的群体和传播真实信息的群体具有不同的特点<sup>[31]</sup>, 这些特点主要体现在用户简档、用户可信度、用户行为和情感立场四个方面<sup>[26]</sup>。

以用户简档作为输入数据的一部分, Wang<sup>[58]</sup>使用混合 CNN 模型取得了较好的检测结果。Long 等<sup>[59]</sup>将用户简档和带有注意力机制的 LSTM 模型结合起来, 与当时最先进的模型相比, 准确率提高了 14.5%。从用户行为的角度出发, Liang 等<sup>[60]</sup>将用户行为视为隐藏线索, 以指示谁可能是谣言制造者。Yang 等<sup>[61]</sup>则将新闻可信度和用户信誉作为潜在变量, 从用户的社会参与中提取用户对新闻可信度的立场, 以此为依据来检测假新闻。此外, 融合文本特征和用户特征的方法备受关注。Zhang 等<sup>[62]</sup>通过融合文本流行度、文本情感极性、用户影响力和博文转发率等特征来构造微博谣言检测器。吴树芳等<sup>[63]</sup>提出了融合微博用户交互行为和博文内容的用户可信度评估算法 HITS-UC (hyperlink-induced topic search - user's credibility), 并验证了算法的可行性和有效性。按照信息生态中信息、信息人、信息环境的划分, Shu 等<sup>[64]</sup>设计的半监督多特征融合模型结合了文本特征、相关用户特征和媒体特征来检测假新闻, 与其他同类型半监督模型相比, 该模型获得了更高的分类准确率。基于深度神经网络, Xu 等<sup>[65]</sup>通过 RNN 对文本编码, 并通过点赞数进行回复文本过滤, 取得了不错的分类效果。柳先觉等<sup>[66]</sup>提出了结合自注意力和 TextCNN 的二级编码和个人资料信誉编码的谣言检测方法, 发现文本内容结合个人资料和微博统计数据构成的特征取得了非常好

的检测结果。受此启发,尹鹏博等<sup>[67]</sup>通过对用户历史行为特征进行分析并结合用户属性和微博文本,使用C-LSTM(convolution-long short-term memory)深度学习模型实现了微博谣言的早期检测。由于该方法不使用评论转发信息,能在谣言产生之初将其识别出来,因此,可以比现有的方法在更“早”的阶段实现检测。

上述方法都是基于真实用户的检测,但是社交媒体上还充斥着大量的电子人和社交机器人,其中一些机器人是专门用来传播谣言、阴谋论等网络噪音的<sup>[68]</sup>。从发布的内容来看,与人类用户相比,社交机器人产生的推文缺乏原创性,转发的推文占比很高,而且推文中含有更多的外部链接<sup>[69]</sup>;从发文时间来看,社交机器人发文时间不规律,短时间内发文数量多,但长时间来看不活跃<sup>[70]</sup>。根据这些特征,Khaund等<sup>[71]</sup>提出了一种基于图的检测方法,该方法的基本假设是社交机器人也有集群的特点,他们彼此交互紧密而与人类交互有限,可利用这个特点检测出网络上的社交机器人。

此外,近年来一些基于众包和群体智慧的检测方法也不断涌现。Wang等<sup>[72]</sup>设计了一个可扩展的检测系统来识别恶意用户,该系统的众包层标注经过过滤层过滤得到的可疑用户,在合理的成本下实现了准确性和可扩展性。Wu等<sup>[73]</sup>从历史众包数据中获得用户信息,通过建立稀疏表示模型选择共享特征来训练事件无关分类器。Cai等<sup>[74]</sup>则从群众响应的角度出发抽取了转发和评论中的文本特征,使用分层聚类算法来发现样本中更多的谣言。为应对传统谣言检测中因人力资源不足而带来的检测周期过长的问題,杨文太等<sup>[75]</sup>利用群众智慧挖掘和度量待检测信息和微博用户的领域相关性,把待检测谣言推送给领域最相关的微博用户(即领域专家)作信息真实性甄别;该方法不仅缩短了检测时间,还能够较好地应用于实际的微博谣言检测。为实现假新闻检测的自动化,Qian等<sup>[76]</sup>提出了一种生成式条件变分自动编码器,该编码器可以从历史用户对真假新闻的评论中获取用户响应模式,进而在假新闻传播的早期阶段,当用户与文章的交互数据不可用时,利用群体智慧来生成对新文章的响应,以提高模型的检测能力。

总体来说,基于用户的检测方法主要是寻找在社交媒体上高度活跃、带有明显倾向性的恶意账号,这些恶意账号既包括真实用户,也包括伪装成真实用户的社交机器人,成功识别这类用户有助于提高

错误信息的检测效率,降低错误信息的干预成本。

## 4 基于传播的错误信息检测

网络是错误信息滋生的温床和传播的渠道,了解网络特征有助于从传播的角度对错误信息进行检测。网络特征是通过在发布相关帖子的用户之间构建特定的网络来提取的<sup>[55]</sup>,可分为分布特征和时间特征。分布特征有助于捕捉网络信息的独特传播模式,可以通过构建一棵传播树来描述<sup>[77]</sup>。例如,基于核函数的传播树检测模型对微博帖子的传播过程进行建模,通过评估传播树结构之间的相似性来区分不同类型的谣言<sup>[78-79]</sup>。时间特征则是以时间序列的方式描述网络信息各项指标的变化情况,是可用于检测可疑发布行为的良好属性<sup>[80]</sup>。常用的时间特征包括两个帖子发布的时间间隔、某个帐户的发布、回复和评论的频率、在固定时间内帖子被重新发布和转发的频率等<sup>[81]</sup>。研究发现,未经证实的新闻极有可能被重新报道,随着时间的推移会出现周期性的讨论高峰,而已经证实的新闻往往只有一个显著的高峰<sup>[82]</sup>。对比真新闻的传播过程,假新闻整体上要传播得更远、更快、更广,因为相同时间里假新闻的级联深度更深,最大宽度更大,且假新闻级联达到任何深度和宽度所需的时间均小于真新闻,这一点在政治类假新闻上表现得尤为明显<sup>[83]</sup>。

基于上述特点,Kwon等<sup>[31]</sup>提出时序、结构、语言三类特征,将谣言和非谣言的数目-时间图与传播网络图之间的差异作为检测特征,在随机森林算法上得到了92%的精确率。Ma等<sup>[40]</sup>将事件的数据流根据时间算法进行切片后得到的数据通过RNN进行学习,并将学习结果与传统的谣言分类结果进行比较,发现使用神经网络可以大大提高谣言检测的准确性。与上述两种方法类似,任文静等<sup>[84]</sup>将帖子的评论信息按照时间节点展开后利用注意力机制衡量每个时间节点对最终语义表示的重要程度,发现在加入评论信息和注意力机制后,准确率可以达到92.66%。

随着研究的深入,特征组合检测的方式渐渐成为主流。最常见的做法是从总体数据中挖掘文本、用户、传播结构特征的动态时序模型以捕捉这些特征随时间的变化<sup>[85]</sup>。毛二松等<sup>[86]</sup>据此提出了微博情感倾向性特征、意见领袖影响力特征、转发数差值和评论数差值等特征,并使用集成分类器来检测微博上的谣言;结果表明,分类特征和集成分类器都有效地提高了谣言检测的准确率。针对假新闻检测

模型的检测性能与泛化性能无法兼顾的问题,何韩森等<sup>[87]</sup>提出了一种基于特征聚合的假新闻检测器CCNN(center cluster neural network),与传统的SVM、朴素贝叶斯和随机森林模型相比,该模型的整体性能有明显优势。刘勘等<sup>[88]</sup>则提出了一种基于深度迁移网络的跨领域谣言检测模型,在源领域拥有标注数据、目标领域无标注数据的情况下,通过双层双向LSTM实现对文本特征的深度提取,并与用户特征及传播特征相融合,结果证明,深度迁移网络能够有效迁移源领域标注数据,帮助目标领域构建谣言检测模型,显著提升目标领域谣言的检测准确率。

此外,一些学者根据信息传播和传染病传播之间的相似性,把SI、SIS和SIR(S=susceptible, I=infectious, R=recovered)等传染病模型运用到谣言传播模型的构建中<sup>[89]</sup>。其中,Zanette<sup>[90]</sup>使用SI模型模拟了小世界网络中谣言传播的动态过程,以定量的方法证实网络谣言的传播过程可被SI模型解释。Nekovee等<sup>[91]</sup>通过SIR模型和马尔可夫链建立了平均场方程来研究网络拓扑结构中的谣言传播过程;陈一新等<sup>[92]</sup>在SIR模型的基础上加入“辟谣者”状态构建了基于社交网络中节点交互作用的SIDR(susceptible infected denied removal)模型,利用Beam Search搜索算法在谣言传播初期Top 5节点的识别上达到了83%的准确率。

错误信息检测的最终目的是及时找到并有效阻断

错误信息的传播,防止其可能对社会造成的不良影响。上述基于传播的检测方法可以满足该需求,但是,由于许多新闻在发布的时候只有有限的,甚至根本没有社会语境信息,导致这类方法在没有良好输入的情况下准确率很难得到保证,优势难以发挥。

## 5 总结与展望

### 5.1 已有研究述评总结

检测出社交媒体中的错误信息是净化网络环境、维护信息生态平衡的关键。鉴于信息生态是由信息、信息人、信息环境组成的一个密不可分的整体,本文在整理和分析相关文献后,对照信息生态的三个组成要素将检测方法从内容、用户和传播三个方面进行了阐述,如图4所示。从当前的研究趋势看,在检测方法上,研究者们从早期的人工检测、实证研究向人工智能和机器学习转变;训练数据正在从静态、滞后、小规模数据向动态、早期、大规模数据转变,但是这种转变对模型的迁移能力、特征提取技术和数据集规模提出了较高的要求。尤其是早期检测,相关数据的稀缺、用户数据的提取困难导致效果不够理想。为了充分满足错误信息检测的需求,需要对各项检测特征进行深度挖掘,对检测方法进行进一步探索和优化,从而设计出更加准确、高效、可解释、可扩展的模型。

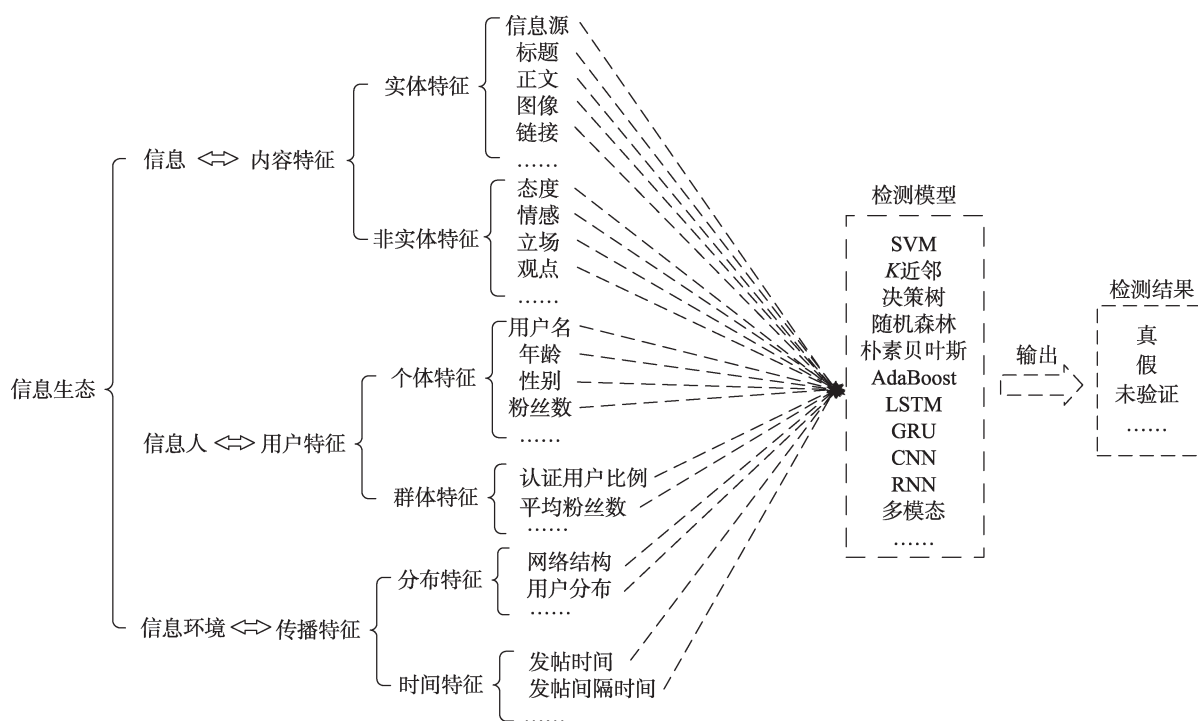


图4 基于信息生态理论的错误信息检测框架



5.2 未来研究展望

首先，当前的错误信息检测主要依赖机器学习和深度学习技术<sup>[93-94]</sup>，并由大规模文本数据集训练而来。而在某一特定的错误信息传播过程中，与其相关的数据是随着事态发展而增多的，传统模型只有在积累到一定量的数据以后才能进行有效训练，这就导致检测方法的开发滞后于检测需求。因此，在数据匮乏的情况下完成早期检测是一个亟待解决的任务。其次，无论是为了更有效地评价错误信息检测模型还是提升模型的迁移能力，当前都需要构建规模更大、更全面的数据集。此外，通过分析用户的简档和历史行为来提取用户画像是检测错误信息的重要手段。但是，不少恶意用户会刻意伪装成正常用户，且用户特征一般具有无序性和复杂性，这给用户信息挖掘带来了困难。面对当前业界实践以及科学研究中的具体难题，错误信息检测领域存在以下几个方面的挑战。

1) 错误信息早期检测

错误信息早期检测是一个极具社会现实意义的

研究方向。从信息生态的角度看，在错误信息传播早期，信息大多分布在位于信息生态链上游的生产者以及与这些生产者联系较为密切的小部分传播者之间。如果能在错误信息到达中下游之前将其识别出来并加以控制，那么会大幅降低错误信息的治理成本和可能产生的负面影响。现有的检测模型在特征提取方面已表现得较为出色，但仍需要大量特征信息作为输入。在错误信息传播的早期，文本信息、用户评论、传播信息等相关特征非常少，并不能满足基于多特征检测模型的需要，即早期检测存在“冷启动”问题。此外，现有的检测方法更倾向于学习特定事件的特征，而这些特征在不同的事件之间往往不能迁移。因此，基于这种不可迁移性，对暂未发生的事件目前还很难做到有效的早期检测。

2) 大规模基准数据集构建

错误信息检测模型的开发和测评都需要依赖数据输入。虽然当下已有许多公开的数据集可供开发和检验错误信息检测模型，但这些数据集都存在一定程度上的不足，如表 1 所示。

表 1 现有的主要公开数据集

数据集	年份	数据量	数据集描述	适用领域	局限
PHEME	2014	6425	包含了 9 个重要事件下的 6425 条推文，注释标签分为真、假、未验证 3 种类型	谣言检测	包含用户信息但缺乏多媒体内容
CREDBANK	2014	6 千万	涵盖了 1000 多个事件的大规模众包数据集，每个事件都由 30 名注释者进行可信度评估	假新闻检测	缺少视觉信息
BS Detector	2016	12999	包含来自 244 个网站的文本和元数据，标签是 BS Detector 输出的，而不是人工标注	可信度检测	只能进行基于文本的分析
BuzzFeedNews	2017	2282	收集了 2016 年美国大选期间 9 家新闻机构在脸书上发布的 2282 篇帖子，每个帖子中都由 5 名 BuzzFeed 记者进行事实核查	假新闻检测	只能进行基于文本的分析
LIAR	2017	12836	通过 API 接口从事实核查网站 PolitiFact 中收集得到，包含 2007 年至 2016 年不同背景下的带有标签的简短陈述，这些标签被分为真、大部分真、半真等 6 个类别	假新闻检测	只能进行基于文本的分析
MisInfo Text	2019	1725	包含了带有标签的新闻原文和源头链接	错误信息检测	话题主要针对美国大选，不含视觉信息
CoAID	2020	302926	包含了与新冠肺炎相关的错误信息和用户参与信息	错误信息检测	话题单一，主要针对新冠肺炎的错误信息
FakeNewsNet	2020	23921	包含了从 BuzzFeed 和 PolitiFact 网站上搜集的假新闻文章的标题和正文，以及从 Twitter 上收集到的与这些假新闻有关的用户参与信息	假新闻检测	数据集并未完全公开
Fakeddit	2020	1063106	包含了文本和图像，根据不同级别的粒度给每个样本打上了多个标签	假新闻检测	数据集并未完全公开

随着相关研究的深入，学界需要建立更全面的大规模基准数据集以便更好地构建和测试错误信息的检测模型。首先，就数据获取方法来说，目前错误信息数据的来源主要有社交媒体平台提供的 API (application programming interface)、通用爬虫、第三方公开数据集。然而，由于社交媒体平台对本平

台的数据都有相应的保护策略，通过平台 API 获取数据在爬取速度和数量上都受到严格控制，无法满足检测的需求。基于爬虫的方式不仅技术复杂，还可能面临法律风险。第三方公开数据虽然容易获得，但这些数据是由数据提供者根据自己的需要收集后公开的，显然无法满足所有用户的需求<sup>[95]</sup>。上

述三种方式都是以研究人员为中心的数据获取。为了提高错误信息检测问题的研究效率,将研究者从琐碎繁重的数据收集工作中解放出来,政府、社交平台与学术界应当协同合作,共同构建大规模数据集,这可以让研究者把更多的精力放在检测方法的探索上。其次,就数据标记来说,许多数据集通常是通过手工标记的,这种标记方式成本高昂且需要从业人员具有丰富经验。因此,如何准确、高效地构建大规模数据集是一大挑战。此外,错误信息的产生往往来源于最新事件,新事件的特点就是缺乏确凿的证据,现有的数据库无法对这些新事件进行及时的检验与核实,导致无法获取完全有效的特征并影响模型建立。因此,建立一个全面、大规模、实时更新的基准数据集对推动错误信息检测方法的开发与测试有着巨大的帮助。

### 3) 恶意用户识别

错误信息泛滥与网络上充斥着大量的恶意账户密切相关。水军、社交机器人等在社交媒体上开展舆论引导、恶意评论、诽谤和意识形态渗透等一系列活动,这一小部分恶意用户是网络中大部分错误信息的贡献者,他们的存在是网络环境和信息生态面临的最大威胁,较早识别出恶意用户对错误信息检测具有重要意义。但是,由于用户是否存在恶意很难有一个明确的判断标准,这类恶意用户不仅特征复杂,还会伪装成正常用户以避免被发现。现有的许多研究只从单一的角度提取社交机器人的特征,并不能得到相对详细的恶意用户的画像。所以如何全面、有效地捕获恶意用户的特征仍是一个有待解决的问题。此外,不同的社交媒体平台在语言、功能、特性方面存在差异,适用于某一社交媒体平台的恶意用户识别技术不一定适用于另一个社交媒体平台。因此,这种难移植性对基于用户信息的错误信息检测方法提出了很高的要求。

## 参 考 文 献

- [1] 陈燕方,李志宇,梁循,等. 在线社会网络谣言检测综述[J]. 计算机学报, 2018, 41(7): 1648-1677.
- [2] 邸子桓,吕明臣. 信息生态视域下网络语言信息传播机制研究[J]. 情报科学, 2021, 39(1): 28-34.
- [3] 位志广,宋小康,朱庆华,等. 基于随机森林的健康谣言分享意愿研究[J]. 现代情报, 2020, 40(5): 78-87.
- [4] 江成,刘室辰. 谣言网络多级传播路径下关键引爆点识别模型和算法研究[J]. 情报杂志, 2020, 39(6): 152-158.
- [5] 李悦晨,钱玲飞,马静. 基于BERT-RCNN模型的微博谣言早期检测研究[J]. 情报理论与实践, 2021, 44(7): 173-177, 151.
- [6] 王晰巍,张柳,黄博,等. 基于区块链的网络谣言甄别模型及仿真研究[J]. 情报学报, 2021, 40(2): 194-203.
- [7] Wu L, Morstatter F, Hu X, et al. Mining misinformation in social media[M]// Big Data in Complex and Social Networks. London: Chapman and Hall/CRC, 2016: 123-152.
- [8] 滕婕,夏志杰,罗梦莹,等. 基于Multi-Agent的网络谣言传播事件中信息主体信任识别研究[J]. 情报杂志, 2020, 39(3): 105-114.
- [9] Guo B, Ding Y S, Yao L N, et al. The future of misinformation detection: new perspectives and trends[OL]. (2019-09-09). <https://arxiv.org/pdf/1909.03654.pdf>.
- [10] Shu K, Bhattacharjee A, Alatawi F, et al. Combating disinformation in a social media age[J]. WIREs Data Mining and Knowledge Discovery, 2020, 10(6): e1385.
- [11] Araujo R F, de Oliveira T M. Disinformation about hydroxychloroquine on Twitter: from political pressure to scientific dispute[J]. AtoZ, Novas Práticas em Informação e Conhecimento, 2020, 9(2): 196-205.
- [12] Furnival A C M, Santos T. Disinformation and fake news: development, detection and ways of fighting it[J]. Conexão: Comunicação e Cultura, 2019, 18(36): 94-113.
- [13] Galitsky B. Detecting rumor and disinformation by web mining [C]// Proceedings of the 2015 AAAI Spring Symposium Series. Palo Alto: AAAI Press, 2015.
- [14] Sharma K, Qian F, Jiang H, et al. Combating fake news: a survey on identification and mitigation techniques[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(3): Article No.21.
- [15] Fallis D. A functional analysis of disinformation[C]// iConference 2014 Proceedings. iSchools, 2014: 621-627.
- [16] Krause N M, Freiling I, Beets B, et al. Fact-checking as risk communication: the multi-layered risk of misinformation in times of COVID-19[J]. Journal of Risk Research, 2020, 23(7/8): 1052-1059.
- [17] Su Q, Wan M Y, Liu X Q, et al. Motivations, methods and metrics of misinformation detection: an NLP perspective[J]. Natural Language Processing Research, 2020, 1(1/2): 1-13.
- [18] Barua Z, Barua S, Aktar S, et al. Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation[J]. Progress in Disaster Science, 2020, 8: 100119.
- [19] Sharma K, Seo S, Meng C Z, et al. Coronavirus on social media: analyzing misinformation in Twitter conversations[OL]. (2020-03-26). <https://arxiv.org/pdf/2003.12309v1.pdf>.
- [20] Wu L, Morstatter F, Carley K M, et al. Misinformation in social media: definition, manipulation, and detection[J]. ACM SIGKDD Explorations Newsletter, 2019, 21(2): 80-90.
- [21] Davenport T H, Prusak L. Information ecology: mastering the information and knowledge environment[M]. New York: Oxford University Press, 1997.



- [22] 曹海军, 侯甜甜. 信息生态视角下政务短视频的内生逻辑与优化路径[J]. 情报杂志, 2021, 40(2): 189-194.
- [23] 娄策群, 周承聪. 信息生态链: 概念、本质和类型[J]. 图书情报工作, 2007, 51(9): 29-32.
- [24] 靖继鹏. 信息生态理论研究发展前瞻[J]. 图书情报工作, 2009, 53(4): 5-7.
- [25] 陈曙. 信息生态研究[J]. 图书与情报, 1996(2): 12-19.
- [26] Zhang X C, Ghorbani A A. An overview of online fake news: Characterization, detection, and discussion[J]. *Information Processing & Management*, 2020, 57(2): 102025.
- [27] Ahn Y C, Jeong C S. Natural language contents evaluation system for detecting fake news using deep learning[C]// *Proceedings of the 2019 16th International Joint Conference on Computer Science and Software Engineering*. IEEE, 2019: 289-292.
- [28] Devitt A, Ahmad K. Sentiment polarity identification in financial news: a cohesion-based approach[C]// *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Association for Computational Linguistics, 2007: 984-991.
- [29] Castillo C, Mendoza M, Poblete B. Information credibility on Twitter[C]// *Proceedings of the 20th International Conference on World Wide Web*. New York: ACM Press, 2011: 675-684.
- [30] Castillo C, Mendoza M, Poblete B. Predicting information credibility in time-sensitive social media[J]. *Internet Research*, 2013, 23(5): 560-588.
- [31] Kwon S, Cha M, Jung K, et al. Prominent features of rumor propagation in online social media[C]// *Proceedings of the 2013 IEEE 13th International Conference on Data Mining*. IEEE, 2013: 1103-1108.
- [32] Granik M, Mesyura V. Fake news detection using naive Bayes classifier[C]// *Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering*. IEEE, 2017: 900-903.
- [33] Rubin V L, Conroy N J, Chen Y M. Towards news verification: deception detection methods for news discourse[C]// *Proceedings of the Hawaii International Conference on System Sciences*, 2015.
- [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: Association for Computational Linguistics, 2017: 2931-2937.
- [35] Ahmed H, Traore I, Saad S. Detection of online fake news using *N*-gram analysis and machine learning techniques[C]// *Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*. Cham: Springer, 2017: 127-138.
- [36] Rajdev M, Lee K. Fake and spam messages: detecting misinformation during natural disasters on social media[C]// *Proceedings of the 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. IEEE, 2015: 17-20.
- [37] Ciampaglia G L, Shiralkar P, Rocha L M, et al. Computational fact checking from knowledge networks[J]. *PLoS One*, 2015, 10(6): e0128193.
- [38] Shi B X, Weninger T. Fact checking in heterogeneous information networks[C]// *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web. Republic and Canton of Geneva: Conferences Steering Committee, 2016: 101-102.
- [39] Guacho G B, Abdali S, Shah N, et al. Semi-supervised content-based detection of misinformation via tensor embeddings[C]// *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2018: 322-325.
- [40] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks[C]// *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2016: 3818-3824.
- [41] 楼靓. 社交网络虚假新闻识别方法[J]. 浙江交通职业技术学院学报, 2020, 21(2): 106-110.
- [42] Kumar S, Mahanti P, Wang S J. Intelligent computational techniques for multimodal data[J]. *Multimedia Tools and Applications*, 2019, 78(17): 23809-23814.
- [43] Qi P, Cao J, Yang T Y, et al. Exploiting multi-domain visual information for fake news detection[C]// *Proceedings of the 2019 IEEE International Conference on Data Mining*. IEEE, 2019: 518-527.
- [44] Rössler A, Cozzolino D, Verdoliva L, et al. FaceForensics++: learning to detect manipulated facial images[C]// *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. IEEE, 2019: 1-11.
- [45] Sabir E, Cheng J X, Jaiswal A, et al. Recurrent convolutional strategies for face manipulation detection in videos[C]// *Proceedings of the Workshop on Applications of Computer Vision and Pattern Recognition to Media Forensics at CVPR 2019*, 2019: 80-87.
- [46] Singhal S, Shah R R, Chakraborty T, et al. SpotFake: a multimodal framework for fake news detection[C]// *Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data*. IEEE, 2019: 39-47.
- [47] Jin Z W, Cao J, Guo H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]// *Proceedings of the 25th ACM International Conference on Multimedia*. New York: ACM Press, 2017: 795-816.
- [48] 张国标, 李洁. 融合多模态内容语义一致性的社交媒体虚假新闻检测[J]. 数据分析与知识发现, 2021, 5(5): 21-29.
- [49] 刘金硕, 冯阔, Jeff Z. Pan, 等. MSRD: 多模态网络谣言检测方法[J]. 计算机研究与发展, 2020, 57(11): 2328-2336.

- [50] Wang Y Q, Ma F L, Jin Z W, et al. EANN: event adversarial neural networks for multi-modal fake news detection[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 849-857.
- [51] Khattar D, Goud J S, Gupta M, et al. MVAE: multimodal variational autoencoder for fake news detection[C]// Proceedings of the World Wide Web Conference. New York: ACM Press, 2019: 2915-2921.
- [52] Jiang Y, Liu Y J, Yang Y L. LanguageTool based university rumor detection on Sina Weibo[C]// Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing. IEEE, 2017: 453-454.
- [53] Yang Q H, Sangalang A, Rooney M, et al. How is marijuana vaping portrayed on YouTube? Content, features, popularity and retransmission of vaping marijuana YouTube videos[J]. Journal of Health Communication, 2018, 23(4): 360-369.
- [54] 张少钦, 杜圣东, 张晓博, 等. 融合多模态信息的社交网络谣言检测方法[J]. 计算机科学, 2021, 48(5): 117-123.
- [55] Shu K, Sliva A, Wang S H, et al. Fake news detection on social media[J]. ACM SIGKDD Explorations Newsletter, 2017, 19(1): 22-36.
- [56] Shu K, Wang S H, Liu H. Understanding user profiles on social media for fake news detection[C]// Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval. IEEE, 2018: 430-435.
- [57] 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. New York: ACM Press, 2012: Article No.13.
- [58] Wang W Y. "Liar, liar pants on fire": a new benchmark dataset for fake news detection[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017: 422-426.
- [59] Long Y F, Lu Q, Xiang R, et al. Fake news detection through multi-perspective speaker profiles[C]// Proceedings of the Eighth International Joint Conference on Natural Language Processing. Asian Federation of Natural Language Processing, 2017: 252-256.
- [60] Liang G, He W B, Xu C, et al. Rumor identification in microblogging systems based on users' behavior[J]. IEEE Transactions on Computational Social Systems, 2015, 2(3): 99-108.
- [61] Yang S, Shu K, Wang S H, et al. Unsupervised fake news detection on social media: a generative approach[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 5644-5651.
- [62] Zhang Q, Zhang S Y, Dong J, et al. Automatic detection of rumor on social network[C]// Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing. Cham: Springer, 2015: 113-122.
- [63] 吴树芳, 徐建民. 基于HITS算法的微博用户可信度评估[J]. 山东大学学报(工学版), 2016, 46(5): 7-12.
- [64] Shu K, Wang S H, Liu H. Exploiting tri-relationship for fake news detection[OL]. (2017-12-20). <https://arxiv.org/pdf/1712.07709v1.pdf>.
- [65] Xu Y C, Wang C, Dan Z P, et al. Deep recurrent neural network and data filtering for rumor detection on sina weibo[J]. Symmetry, 2019, 11(11): 1408.
- [66] 柳先觉, 徐义春, 董方敏. 结合文本及用户资料数据的微博谣言检测[J]. 信息通信, 2020, 33(12): 39-43.
- [67] 尹鹏博, 潘伟民, 彭成, 等. 基于用户特征分析的微博谣言早期检测研究[J]. 情报杂志, 2020, 39(7): 81-86.
- [68] Đukić D, Keča D, Stipičić D. Are you human? Detecting bots on twitter using BERT[C]// Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics. IEEE, 2020: 631-636.
- [69] Ratkiewicz J, Conover M, Meiss M, et al. Detecting and tracking political abuse in social media[J]. Proceedings of the International AAAI Conference on Web and Social Media, 2011, 5(1): 297-304.
- [70] Lee S, Kim J. Early filtering of ephemeral malicious accounts on Twitter[J]. Computer Communications, 2014, 54: 48-57.
- [71] Khaund T, Al-Khateeb S, Tokdemir S, et al. Analyzing social bots and their coordination during natural disasters[C]// Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. Cham: Springer, 2018: 207-212.
- [72] Wang G, Mohanlal M, Wilson C, et al. Social Turing tests: crowdsourcing sybil detection[C]// Proceedings of the 20th Annual Network Distributed System Security Symposium, 2013.
- [73] Wu L, Li J D, Hu X, et al. Gleaning wisdom from the past: early detection of emerging rumors in social media[C]// Proceedings of the 17th SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics Publications, 2017: 99-107.
- [74] Cai G Y, Wu H, Lv R. Rumors detection in Chinese via crowd responses[C]// Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE, 2014: 912-917.
- [75] 杨文太, 梁刚, 谢凯, 等. 基于突发话题和领域专家的微博谣言检测方法[J]. 计算机应用, 2017, 37(10): 2799-2805.
- [76] Qian F, Gong C Y, Sharma K, et al. Neural user response generator: fake news detection with collective user intelligence[C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: 3834-3840.
- [77] Watts D J, Peretti J, Frumin M. Viral marketing for the real world[M]. Boston: Harvard Business School Publication, 2007.
- [78] Ma J, Gao W, Wong K F. Detect rumors in microblog posts using

- propagation structure via kernel learning[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017: 708-717.
- [79] Wu K, Yang S, Zhu K Q. False rumors detection on Sina Weibo by propagation structures[C]// Proceedings of the 2015 IEEE 31st International Conference on Data Engineering. IEEE, 2015: 651-662.
- [80] Ji F, Tay W P. Identifying rumor sources with different start times [C]// Proceedings of the 2016 IEEE Statistical Signal Processing Workshop. IEEE, 2016: 1-5.
- [81] Kotteti C M M, Dong X S, Qian L J. Rumor detection on time-series of tweets via deep learning[C]// Proceedings of the 2019 IEEE Military Communications Conference. IEEE, 2019: 1-7.
- [82] Gao H Y, Mao J H, Zhou J, et al. Are you talking to a machine? Dataset and methods for multilingual image question answering [C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 2296-2304.
- [83] Jin Z W, Cao J, Jiang Y G, et al. News credibility evaluation on microblog with a hierarchical propagation model[C]// Proceedings of the 2014 IEEE International Conference on Data Mining. IEEE, 2014: 230-239.
- [84] 任文静, 秦兵, 刘挺. 基于时间序列网络的谣言检测研究[J]. 智能计算机与应用, 2019, 9(3): 300-303.
- [85] 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. New York: ACM Press, 2015: 1751-1754.
- [86] 毛二松, 陈刚, 刘欣, 等. 基于深层特征和集成分类器的微博谣言检测研究[J]. 计算机应用研究, 2016, 33(11): 3369-3373.
- [87] 何韩森, 孙国梓. 基于特征聚合的假新闻内容检测模型[J]. 计算机应用, 2020, 40(8): 2189-2193.
- [88] 刘勘, 杜好宸. 基于深度迁移网络的 Twitter 谣言检测研究[J]. 数据分析与知识发现, 2019, 3(10): 47-55.
- [89] 贾硕, 张宁, 沈洪洲. 网络谣言传播与消解的研究进展[J]. 信息资源管理学报, 2019, 9(3): 62-72.
- [90] Zanette D H. Critical behavior of propagation on small-world networks[J]. Physical Review E, 2001, 64(5): 050901.
- [91] Nekovee M, Moreno Y, Bianconi G, et al. Theory of rumour spreading in complex social networks[J]. Physica A: Statistical Mechanics and Its Applications, 2007, 374(1): 457-470.
- [92] 陈一新, 陈馨悦, 刘奕, 等. 基于SIDR模型的谣言传播与源头检测研究[J]. 数据分析与知识发现, 2021, 5(1): 78-89.
- [93] Chen W L, Zhang Y, Yeo C K, et al. Unsupervised rumor detection based on users' behaviors using neural networks[J]. Pattern Recognition Letters, 2018, 105: 226-233.
- [94] Goodfellow I J, Bengio Y, Courville A C. Deep learning[M]. MIT press Cambridge, 2015.
- [95] 高玉君, 梁刚, 蒋方婷, 等. 社会网络谣言检测综述[J]. 电子学报, 2020, 48(7): 1421-1435.

(责任编辑 魏瑞斌)