



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331, CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目：基于局部和全局特征聚合的虚假新闻检测方法  
作者：杨书新，丁祺伟  
网络首发日期：2024-04-21  
引用格式：杨书新，丁祺伟. 基于局部和全局特征聚合的虚假新闻检测方法[J/OL]. 计算机工程与应用. <https://link.cnki.net/urlid/11.2127.TP.20240419.1103.006>



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

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

## 基于局部和全局特征聚合的虚假新闻检测方法

杨书新, 丁祺伟

江西理工大学 信息工程学院, 江西 赣州 341000

**摘要：**在虚假新闻检测的问题上，现有的方法大多通过捕捉上下文语义特征、对照外部知识库来判断新闻的真实性，但很多研究人员忽略了新闻在发布时期各大社交媒体的新闻生态。针对上述问题，提出一种基于局部和全局特征聚合的虚假新闻检测方法。具体来说，设计了全局环境感知模块和局部环境感知模块，其中全局环境感知模块通过均值池化、全连接、注意力机制等操作获得全局特征，局部环境感知模块通过将文本向量进行聚类、均值池化、哈达玛乘积、多层感知器、注意力机制等操作获得局部特征。然后利用融合机制将这些特征进行聚合，最后与其它虚假新闻检测器协同工作，用以验证新闻的真实性。本文在微博和 Twitter 两个新闻文本数据集上进行了对比实验，实验结果表明本文提出的方法可以有效提取全局和局部环境中的新闻文本特征，从而提高检测精度。

**关键词：**虚假新闻检测；注意力机制；特征提取；社交媒体

**文献标志码：**A      **中图分类号：**TP391      **doi：**10.3778/j.issn.1002-8331.2312-0108

### Fake news detection method based on local and global feature aggregation

YANG Shuxin, DING Qiwei

School of Information Engineering, Jiangxi University of Science & Technology, Ganzhou, Jiangxi 341000, China

**Abstract:** On the issue of fake news detection, existing works rely on capturing contextual semantic features and referencing external knowledge bases to judge the authenticity of news, while neglecting the public opinion orientation at the time of news publication. To fill the gap, a fake news detection method based on local and global feature aggregation is proposed. In a more specific context, both a global environment perception module and a local environment perception module have been designed. The global environment perception module utilizes operations such as average pooling, fully connected layers, and attention mechanisms to extract global features, The local environment perception module obtains local features by performing operations such as clustering, average pooling, hadamard product, multi-layer perceptron, and attention mechanism on text vectors. Subsequently, these features are consolidated using a fusion mechanism and, in the end, work in conjunction with other fake news detectors to authenticate the news. This study conducted comparative experiments on news text datasets from both Weibo and Twitter. The experimental results demonstrate that the proposed method is effective in extracting news text features from both global and local environments, consequently enhancing detection accuracy.

**Key words:** fake news detection; attention mechanism; feature extraction; social media

**基金项目：**国家自然科学基金（72261018）；江西省自然科学基金（20212BAB202014）；江西省教育厅科技（GJJ2200830）；江西省教育厅青年项目（GJJ2200868）。

**作者简介：**杨书新(1978—)，男，教授，CCF 会员，主要研究方向为社交网络信息和生物信息；丁祺伟(1998—)，通信作者，男，硕士，CCF 会员，研究方向为自然语言处理、数据挖掘，E-mail: 1207890438@qq.com。

5G 时代下, 随着物联网 (Internet of Things)、社交网络等信息技术的快速发展, 人们获取信息的渠道逐步从报纸、杂志等传统纸质媒介替换为微博、推特等社交平台。然而上述社交平台注册门槛通常过低, 对用户发布的信息缺乏严格的审核机制, 从而使某些别有用心的人或机构通过社交媒体发布虚假新闻, 以达到引导公众舆论的目的。虚假新闻是指未经证实或蓄意编纂的新闻, 它往往结合时事热点, 把握读者心理, 相较于真实新闻, 虚假新闻往往具备传播速度快、传播覆盖面广, 辟谣成本高等特性。若虚假新闻在发布初期未能及时被识别和遏制, 将对公众造成重大误导并可能为不法群体带来经济利益<sup>[1]</sup>。这不仅影响社会和谐与稳定, 还可能威胁到国家机构的安全。

为检测虚假新闻, 国内外许多科研机构提出了多种理论和方法。比如一些研究将循环神经网络<sup>[2]</sup> (Recurrent Neural Networks, RNN) 引入检测过程中, 这项技术允许模型更好的理解文本中的时间序列特性, 被广泛运用于自然语言处理领域。还有些研究采用了外部知识库作为验证和参考的来源, 此类方法允许将帖子的内容和已有的知识库进行对照, 从而验证帖子的真实性。随着虚假新闻检测技术的不断演化, 越来越多的策略被不断提出旨在提高模型的性能和鲁棒性。然而这些方法一方面依赖深度学习的分类模型本身, 另一方面对外部的先验知识储备要求很高, 却忽略了假新闻帖子很大一部分是由真实帖子改编而来。由于虚假新闻只有在大范围曝光和病毒式传播的基础上才可以起到引导舆论的作用, 所以假新闻编辑者有一定的动机会对当下的热点新闻进行改编, 捏造出读者可能感兴趣的帖子, 达到吸引读者注意力的目的。本文的出发点是将新闻的发布背景引入到虚假新闻检测过程中, 将重点放在研究待检测新闻与流行新闻的相似性与差异性上, 再通过配合已有模型进一步提升虚假

新闻的检测精度。本文的主要贡献有以下三点:

(1) 提出一种基于局部和全局特征聚合的虚假新闻检测方法 (Local and Global Aggregation, LGA), 利用不同算法计算目标新闻与流行新闻的相似性与差异性, 为虚假新闻检测提供了新的影响因子。

(2) 利用特征融合器将新闻所处环境的局部特征和全局特征有效聚合, 利用两者的关联性提高虚假新闻检测的精度。

(3) 本文进行相关实验来评估我们提出的方法, 在中文和英文数据集上验证了提出方法的有效性。

## 1 相关工作

虚假新闻检测旨在通过检测模型判断社交媒体中发布新闻的真实性, 在自然语言处理领域中属于二分类问题<sup>[3]</sup>。在对虚假新闻的文本内容进行检测时, 现有的研究主要集中于分析新闻文本的相关信息以及外部知识来源的引入<sup>[4]</sup>。

### 1.1 基于新闻内容的虚假新闻检测方法

早期的虚假新闻检测方法主要通过分析文本内容和用户个人信息中的各类特征信息来判断帖子的真实性。Kumar 等<sup>[5]</sup>提出了多任务学习框架, 该框架融合了虚假新闻检测以及立场检测, 将用户对于新闻事件的看法和态度等因素融入了虚假新闻检测任务中。在之后的研究中, Kumar 等<sup>[6]</sup>设计了 Tree-LSTM 模型, 实现了将新闻的后续反馈用于虚假新闻检测任务。除此之外, 新闻的各种特征对于新闻的真实性也是很重要的考量因素。Castillo 等<sup>[7]</sup>通过特殊字符、表情符号、新闻主题等特征来研究 Twitter 平台上新闻的可信度。Yang 等<sup>[8]</sup>通过加入符号特征、关键字特征完善了 Castillo 等的工作。Kwon 等<sup>[9]</sup>提出了一种基于虚假新闻传播时间、结构、语言属性的虚假新闻检测方法。在深度学习流行的今天, 研究人员也将这项技术引入了谣言检测工作中。Ma 等<sup>[2]</sup>利用深度学习研究虚假新闻检测, 提出了一种动态时间序列结构模型, 该模型使用 RNN 捕捉上下文特征, 抽取相关帖子组成新闻事件。

Chen 等<sup>[10]</sup>在 RNN 模型中使用了注意力机制,给时间序列赋予权重。Yu 等<sup>[11]</sup>利用卷积神经网络 (Convolutional Neural Network, CNN) 捕获文本重要特征之间的高级交互。范昊等<sup>[12]</sup>设计了一种新闻标题文本分类模型,该模型首先使用 BERT 来生成新闻标题的文本向量,接着通过 TextCNN 提取这些向量的特征,并将这些特征传递给 Bi-LSTM 以捕获题中的上下文信息。最后,通过 softmax 函数确定最终的分类结果。

## 1.2 引入知识来源的虚假新闻检测方法

仅凭借新闻内容,通过提取文本各种特征进行虚假新闻检测,就会忽略掉其他重要的参考因素,例如专家对于目标新闻的重要评价、官方的辟谣或是相关文献等外部的知识来源。这些参考因素对于虚假新闻检测均有一定的正反馈作用。于是在对新闻文本进行特征提取的前提下,相关人员引入外部知识来源用于辅助检测器进行虚假新闻检测任务。Hu 等<sup>[13]</sup>通过将待检测新闻与外部知识库中的信息进行比对,再用图神经网络进行虚假新闻检测工作。针对以三元组形式(例如主语、谓语、宾语)表达的虚假新闻检测任务 Shiralkar 等<sup>[14]</sup>提出了无监督网络流方法。Pan 等<sup>[15]</sup>在此基础上提出了基于知识图谱嵌入的政治选举类虚假新闻检测模型。该团队首先捕获了政治选举类的新闻知识(以三元组形式),然后比较三元组之间的相似度从而完成虚假新闻检测任务。Cui 等<sup>[16]</sup>通过引入医学知识图谱学习知识引导的信息嵌入式表示,并通过知识图谱路径传播获取节点表示来检测医学虚假信息。郭秋实等<sup>[17]</sup>通过引入知识图谱作为额外先验知识表示,提出了生成式对抗网络的知识图谱嵌入模型。该团队在知识图谱嵌入的模型基础上,针对社交网络的异构图结构使用了图卷积网络挖掘节点的传播行为获得了文本的实体表示,并进行了实体嵌入和文本词嵌入,最后融合多头注意力机制实现了虚假新闻检测。当研究者

考虑将外部知识来源引入虚假新闻检测后,一定程度的提高了虚假新闻的检测精度。同理,在虚假新闻检测的过程中,若能考虑其它有意义的外部因素,检测结果就有更加精确的可能。因此,聚合局部与全局特征对于虚假新闻检测是一个新的视角。本文提出的基于局部与全局特征聚合的虚假新闻检测方法,与上述两类检测方法不同的是,LGA 在特征感知阶段不需要考虑文本的上下文特征或根据知识库证明目标帖子的真实性,而是通过判断目标新闻与热点新闻在内容上的关联程度和差异程度,并结合其他有效的虚假新闻检测器验证新闻的真实性。

## 2 基于局部和全局特征聚合的虚假新闻检测方法

在以往的工作中,研究人员将大量精力放在优化训练模型和方法上。为改进此类方法的局限性,本研究试图结合局部和全局特征以提高虚假新闻检测的准确性。LGA 的总体框架如图 1 所示。具体来说,LGA 由三个模块组成,即新闻环境分类模块、新闻环境感知模块和虚假新闻分类模块。首先,新闻环境分类模块负责构建能够反映当下新闻特点的全局环境(Global Environment),并且通过计算目标新闻与全局环境中所有新闻的相似度用以区分出局部环境(Local Environment),为后续的特征提取和判定新闻真伪奠定基础;其次,环境感知模块负责捕获目标新闻与全局环境和局部环境中新闻的交互模式,其中全局环境感知模型侧重于识别目标新闻与主流媒体发布的热点新闻之间的关联性,并得到新闻的全局特征向量  $v^{Nf, glo}$ ,局部环境感知模型则聚焦于识别目标新闻和与其内容相似的新闻之间在内容上的微妙差异,并得到新闻的局部特征向量  $v^{Nf, loc}$ ;最后,虚假新闻分类模块负责在维持目标新闻与新闻环境之间敏感度的同时结合新闻分类模型对新闻的真伪进行判定,验证新闻的真实性。



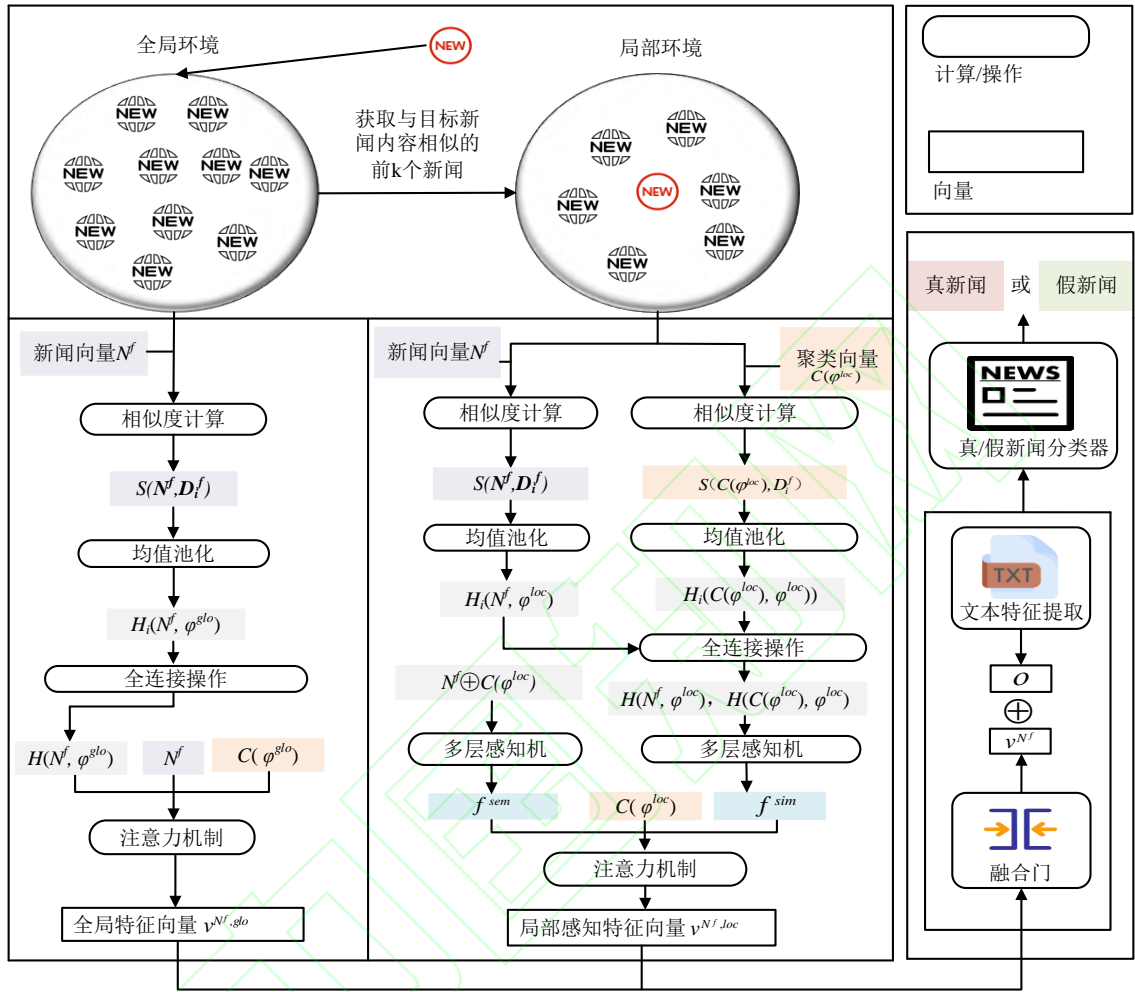


图1 模型架构图

Fig.1 Overall Structure of the Model

## 2.1 新闻全局环境和局部环境

新闻环境指当下各大媒体发布的时事热点以及在社会各个领域发生的重大事件。本文提出的 LGA 方法基于此构建了一个可以反映新闻主流焦点和读者阅读兴趣的新闻环境。设  $\varphi$  为在目标新闻  $N$  之前发布的新闻集合。在两个数据集中对局部环境和全局环境进行分类，定义如下：

全局环境  $\varphi^{glo}$  是指在  $N$  发布前  $T$  天内各大社交媒体产生的新闻项目合集。

$$\varphi^{glo} = \{D : D \in \varphi, 0 < t_N - t_{D_i} \leq T\} \quad (1)$$

其中  $D = [D_1, D_2, \dots, D_n]$  表示在  $N$  发布的前  $T$  天内收集到的新闻集合， $D_i$  表示  $D$  中的第  $i$  条新闻， $t_N$  和  $t_{D_i}$  分别表示新闻项目  $N$  和  $D_i$  的发布日期。

局部环境  $\varphi^{loc}$  是指在全局环境  $\varphi^{glo}$  中与新闻  $N$  密切相关的新闻集合。本研究从全局环境  $\varphi^{glo}$  中挑选了与新闻  $N$  内容相似度较高的前  $k$  个新闻，以构建局部

环境。相关公式描述如下：

$$\varphi^{loc} = \{D_i : D_i \in Top-k(N, \varphi^{glo})\} \quad (2)$$

其中  $k = \lceil r|\varphi^{glo}| \rceil$ ， $r$  为设定的超参数，范围为  $(0, 1)$ 。公式(2)将全局环境  $\varphi^{glo}$  中的新闻数量与超参数  $r$  相乘，再对结果向上取整。

新闻的全局环境  $\varphi^{glo}$  表示近期新闻媒体根据发生的各种社会事件发布的新闻集合，而局部环境  $\varphi^{loc}$  是由与待检测新闻相类似的新闻事件的集合组成。本文使用预训练模型 BERT<sup>[18]</sup> 提取特征向量来得到新闻表示。对于全局/局部环境中的任意新闻  $D_i$  和待检测新闻  $N$  的文本向量表示为：

$$D_i^f = BERT\{[CLS], D_i, [SEP]\} \quad (3)$$

$$N^f = BERT\{[CLS], N, [SEP]\} \quad (4)$$

其中  $D_i^f$  和  $N^f$  是通过 BERT 编码以后得到的文本向量。

## 2.2 环境感知模型

环境感知模型的作用是从待检测新闻  $N$  所处的新闻环境中捕获有用的信号。这些信号反映了虚假新闻与新闻环境特殊的互动模式。LGA 改变了虚假新闻检测的视角, 不直接提取新闻的语义特征、符号特征等代表性特征, 也不直接利用外部知识来源, 而是从假新闻创作者发布虚假信息的动机出发, 从新闻发布者的视角引导模型感知待检测新闻与真实新闻的相似性与差异性。

### 2.2.1 全局环境感知

假新闻创作者为了让编造的新闻在短时间内获得足够的影响力并快速传播开来, 会选择在编辑帖子时借鉴近期的时事热点。具体操作为选取主流媒体的热点新闻, 再对原本的新闻进行加工篡改, 使篡改后的新闻内容能够更好的抓住读者眼球从而达到吸引读者的目的。本文考虑目标新闻在全局环境中与其他新闻的相似程度, 将待检测新闻  $N$  与收集到的每个新闻项目之间进行相似度估计, 如果在全局环境中有许多新闻项目与  $N$  相似, 那么待检测新闻在这样的环境里也有很大可能受读者欢迎。本文通过计算  $N$  在全局环境中与每个新闻项目之间的欧几里得距离 (Euclidean Distance) 来得到目标新闻与各个新闻的相似度估计:

$$S(N^f, D_i^f) = \sqrt{\sum_{i=1}^n (N^f - D_i^f)^2} \quad (5)$$

其中  $D_i^f$  为全局环境中任意新闻的文本向量,  $N^f$  为待检测新闻的文本向量。此外, 在计算目标新闻与其他新闻的相似度时, 我们并不直接采用公式(5)得到的相似性估计值作为基本输入, 这是由于该公式计算的结果具有较高的维度, 因此需要对数据进一步处理。在数据处理过程中, 我们期望变换后的环境感知向量既能够有效地反映  $N$  与全局环境之间的相似程度, 又能够最小化信息损失。为达到这一目标, 我们对获得的欧几里得距离列表进行了均值池化 (Average Pooling), 该过程可以形式化表示为:

$$H_i = \text{AvgP}(S(N^f, D_i^f)) \quad (6)$$

$$H_i(N^f, \varphi^{glo}) = \sum_{i=1}^{|\varphi^{glo}|} H_i \quad (7)$$

在进行均值池化后, 再对得到的结果进行全连接

操作:

$$H(N^f, \varphi^{glo}) = FC\left(\bigoplus_{i=1}^n H_i(N^f, \varphi^{glo})\right) \quad (8)$$

全连接操作的目的是将均值池化后的特征进一步处理, 具体而言是将池化层得到的矩阵平铺为一个列向量以便进行更高级别的特征提取, 有利于提高模型的泛化能力, 其中  $H(N^f, \varphi^{glo})$  表示经过处理后的环境感知向量,  $\oplus$  是拼接操作符 (Concat)。

通过全连接层的输出  $H(N^f, \varphi^{glo})$  可以获得待检测新闻与全局环境之间的相似性分布, 若待检测新闻  $N$  与全局环境中的新闻具有较高相似性, 则说明  $N$  在某些特征或属性上与它们非常接近。为了获得能够代表新闻全局环境的特征向量, 我们通过注意力机制<sup>[19]</sup>对经过处理的特征向量进行加权处理, 从而让模型更加关注于输入向量的关键信息。

$$v^{N^f, glo} = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

其中  $Q$  表示待检测新闻的文本嵌入 ( $Q = N^f$ ),  $V$  表示经过池化和全连接处理后得到的环境感知向量 ( $V = H(N^f, \varphi^{glo})$ ),  $K$  表示对全局环境中每条新闻进行聚类操作得到的质心新闻的嵌入 ( $K = C(\varphi^{glo})$ ),  $d_k$  是质心新闻的维度,  $C(\varphi^{glo})$  表示对数据进行聚类操作得出的聚类向量, 该向量反映了全局环境中新闻的总体特征。

### 2.2.2 局部环境感知

局部环境感知模块主要研究对象是与目标新闻内涵较为相似的新闻项目。通过对全局环境的研究可以知道, 假新闻编辑者会参考当下的时事热点来编辑假新闻。然而, 若站在阅读者的角度, 如果编辑出的帖子与其他帖子在内容上过于相似, 阅读者可能对这些新闻的关注程度不会超过真实新闻。Vosoughi 等<sup>[20]</sup>发现, 读者感兴趣的新闻往往比真实新闻更加夸张。读者在阅读完这些新闻后也更乐于对此进行点评、点赞和转发。这说明虚假新闻通常比真实新闻影响力更大, 传播范围更广。因此 LGA 在局部环境感知模块重点研究目标新闻  $N$  在局部环境中的差异程度。

在研究中, 若目标新闻对比与它内容相似的新闻仍存在差异, 我们就将它视为此类新闻中的异常类别。局部环境感知过程的操作步骤如下: 首先利用公式(5)到(8)得到目标新闻与局部环境中真实新闻的相似程度, 与全局环境感知模块不同的是, 在局部环境感知

模块, 我们同时计算出两个相似特征  $H(N^f, \varphi^{loc})$  和  $H(C(\varphi^{loc}), \varphi^{loc})$ , 其中  $H(C(\varphi^{loc}), \varphi^{loc})$  是辅助性质的环境感知向量, 利用后者作为前者的参考, 以便模型能够更加精准的“校准”感知。生成目标新闻  $N$  的局部环境感知向量的相关公式描述如下:

$$f^{sem} = MLP(N^f \oplus C(\varphi^{loc})) \quad (10)$$

$$f^{sim} = MLP(f(H(N^f, \varphi^{loc}), H(C(\varphi^{loc}), \varphi^{loc}))) \quad (11)$$

其中  $f(H(N^f, \varphi^{loc}), H(C(\varphi^{loc}), \varphi^{loc}))$  为哈达玛乘积算子, 强调了两个特征向量之间的相似性和差异性, 有助于模型更好的理解文本之间的关系, 提高分类性能。在局部环境感知向量的生成过程中, 由于 MLP 机制可以进一步学习和提取数据特征, 我们首先将文本向量  $N^f$  以及  $C(\varphi^{loc})$  进行聚合操作后输入到 MLP 中得到语义信息  $f^{sem}$ 。将哈达玛算子的结果输入 MLP 中得到相似性信息  $f^{sim}$ 。最后把代表新闻语义信息的向量  $f^{sem}$  和代表新闻相似性的向量  $f^{sim}$  与聚类向量  $C(\varphi^{loc})$  引入注意力机制, 得到局部特征向量  $v^{N^f, loc}$ 。

$$v^{N^f, loc} = Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

其中,  $Q = C(\varphi^{loc})$ ,  $K = f^{sem}$ ,  $V = f^{sim}$ 。

当我们将目标新闻  $N$  与局部环境中的新闻进行对比时, 若目标新闻在内容上经过一定程度的修改或篡改, 就会被局部环境感知模块敏锐的识别到。

## 2.3 感知环境下的预测

LGA 不仅适用于特定的虚假新闻检测模型, 还旨在确保其能够与各种虚假新闻检测模型协同工作。为实现这一目标, 我们引入了门融合机制。该机制能有效调整输入向量  $v^{N^f, glo}$  和  $v^{N^f, loc}$  之间的权重分配, 根据自适应的方式聚合新闻的全局环境特征和局部环境特征。

$$v^{N^f} = f \odot v^{N^f, glo} + (1 - f) \odot v^{N^f, loc} \quad (13)$$

其中  $v^{N^f}$  为经过门机制处理后得到的特征向量。 $f = sigmoid(Linear(o \oplus v^{N^f, glo}))$  表示门控向量, 用于将每个元素的值约束在  $[0, 1]$  之间,  $o$  表示使用其他虚假新闻检测器得到的特征向量。将  $o$  和  $v^{N^f}$  进行特征聚合后进一步输入到 MLP 中再通过分类函数 Softmax 进行最终的预测。

$$\hat{y} = softmax(MLP(o \oplus v^{N^f})) \quad (14)$$

若要利用 LGA 方法配合其他模型完成虚假新闻

检测任务时, 可以方便的聚合上述式子中提出的这些特征向量。例如, 若需要配合的虚假新闻检测模型是引入知识来源这种类型, 我们就可以把  $v^{N^f}$  和依赖知识来源的虚假新闻检测器得到的特征向量  $o$  进行聚合, 最终通过 softmax 分类器验证新闻的真实性。

## 3 实验结果与分析

### 3.1 数据集

为评估本文设计的 LGA 对于虚假新闻的检测效果, 实验使用了来自微博和推特两个平台的数据集, 统计信息如表 1 所示。

表 1 数据集统计

统计信息	微博	推特
待检测新闻数量	39,066	6,483
全局环境新闻数量	583,208	1,003,646

实验所用数据集涵盖了中文和英文两个不同语言环境的内容, 旨在验证模型在不同语言环境中的检测精度。中文数据集由经过验证的微博帖子组成, 其中引用了[2]、[21]、[22]使用到的数据集, 内容涵盖了广泛的社会和文化背景。英文数据集则汇集了不同新闻杂志(如 Daily、NPR)中的新闻标题, 其中引用了[23]、[24]、[25]使用到的数据集。为提高数据集的可信度, 实验删除了内容真实性难以辨别以及观点模糊的帖子。

### 3.2 实验设置

本节使用了 numpy、torch、pandas 以及 sklearn 等库中的 API 进行编码。在基准方法的参数设置中, LGA 在处理中文数据时将隐藏层的维度设置为 128, 在处理英文数据时将隐藏层的维度设置为 256, 每个 token 表示的维度设置为 768, Adam 优化器的学习率为  $10^{-5}$ 。在数据集处理中, 中文数据集使用 jieba14 进行分段, 文本最大长度为 100, Dropout 率为 0.3。英文数据集使用 Natural Language Toolkit (NLTK) 进行标记, 文本最大长度设置为 256, Dropout 率为 0.1。本文的实验环境为 PyTorch1.6.0、python3.8, 其中 GPU 为 RTX3080。

### 3.3 评价指标

本文使用了多个评价指标来综合评估 LGA 在虚假新闻检测任务中的性能, 这些评价指标包括以下几个方面。

(1) 准确率 (Accuracy, Acc): Acc 用于衡量模型正确分类的样本数量与总样本数量之间的比例, 相关公式描述如下:



$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

其中 TP 为真正例 (True Positive), TN 为真负例 (True Negative), FP (False Positive) 为假正例, FN (False Negative) 为假负例。

(2)宏 F1 分数 (Macro-F1, MacF1): MacF1 通过计算每个类别的 F1 分数, 并对获得的 F1 分数取平均值, 能够综合评估提出的方法在不同类别上的性能, 相关公式描述如下:

$$MacF1 = \frac{1}{n} \sum_{i=1}^n \frac{2 \cdot Precision_i \cdot Recall_i}{Precision_i + Recall_i} \quad (16)$$

(3)针对虚假类别的 F1 分数 (F1 for Fake Class, F1fake): F1fake 用于评估模型在检测虚假新闻时的性能, 相关公式描述如下:

$$F1_{fake} = \frac{2 \cdot Precision_{fake} \cdot Recall_{fake}}{Precision_{fake} + Recall_{fake}} \quad (17)$$

(4)针对真实类别的 F1 分数 (F1 for Real Class, F1real): F1real 用于评估模型在检测真实新闻时的性能, 相关公式描述如下:

$$F1_{real} = \frac{2 \cdot Precision_{real} \cdot Recall_{real}}{Precision_{real} + Recall_{real}} \quad (18)$$

在公式 (16) (17) (18) 中:

$$Precision = \frac{TP}{TP + FP} \quad (19)$$

$$Recall = \frac{TP}{TP + FN} \quad (20)$$

### 3.4 比较算法

实验选取了 7 个具有代表性的虚假新闻检测器作为实验的基础模型。

(1)Bi-LSTM<sup>[26]</sup>: Bi-LSTM 是一个双向长短时记忆网络模型, 通常用于将文本序列转换为具有上下文信息的向量表示, 常用于对社交媒体帖子进行编码。

(2)EANN<sup>[27]</sup>: 该模型采用对抗训练的方式, 可以有效提取文本特征以及语义结构。

(3)BERT<sup>[18]</sup>: 该模型是一种用于预训练的深度学习模型, 能够有效地提取文本中的语义和关系特征。

(4)BERT-Emo<sup>[28]</sup>: 该模型在 BERT 编码的基础上融合了一系列情感特征, 用于情感分类任务。

(5)S<sup>2</sup>GC<sup>[29]</sup>: S<sup>2</sup>GC 模型结合马尔可夫扩散核在文

本分类中强化上下文关联, 注重于局部与全局层面上文本特征的捕捉, 旨在深入揭示文本间的复杂关系。这种方法提供了一种有效手段, 用于分析文本数据的结构性和上下文之间的相关性。

(6)DeClarE<sup>[30]</sup>: DeClarE 是一个用于实体链接的模型, 它利用帖子和检索到的文档作为证据, 从而验证待检测新闻的真实性。

(7)MAC<sup>[31]</sup>: 该模型构建了一个多头注意力网络, 用于感知文本中的证据。它能够捕捉文本中的关键信息, 有助于提高虚假新闻检测的性能。

为突出本文提出的模型在结合纯文本模型和结合引入知识表示的模型均能够提升检测精度, 实验将上述 7 个用于对比的虚假新闻检测模型分为两类。其中纯文本模型有: Bi-LSTM、EANN、BERT、BERT-Emo、S<sup>2</sup>GC。引入知识表示的模型有: DeClarE、MAC。

## 3.5 实验结果与分析

### 3.5.1 对比实验

为全面评估提出方法的有效性, 分别使用中文数据集和英文数据集进行实验。通过 LGA 在中文数据集和英文数据集上的表现进行综合比较和分析, 我们能够更好地理解提出的方法在不同数据集和不同模型中的性能表现, 从而确保其有效性和通用性。实验结果如表 2、表 3 所示。

表 2 对比实验性能比较 (中文数据集)

Table 2 Comparative Experiment Results (Chinese Dataset)

Model	Chinese			
	Acc	MacF1	F1 <sub>fake</sub>	F1 <sub>real</sub>
Bi-LSTM	0.722	0.709	0.653	0.750
Bi-LSTM+LGA	<b>0.750</b>	<b>0.760</b>	<b>0.735</b>	<b>0.800</b>
EANN	0.729	0.702	<b>0.732</b>	0.787
EANN+LGA	<b>0.769</b>	<b>0.767</b>	0.724	<b>0.806</b>
BERT	0.781	0.789	0.738	0.823
BERT+LGA	<b>0.810</b>	<b>0.809</b>	<b>0.773</b>	<b>0.831</b>
BERT-Emo	0.808	0.804	0.775	0.839
BERT-Emo+LGA	<b>0.826</b>	<b>0.826</b>	<b>0.806</b>	<b>0.859</b>
S <sup>2</sup> GC	0.709	0.695	<b>0.682</b>	0.703
S <sup>2</sup> GC+LGA	<b>0.718</b>	<b>0.704</b>	0.681	<b>0.716</b>
DeClarE	0.755	0.748	0.771	<b>0.822</b>
DeClarE+LGA	<b>0.801</b>	<b>0.792</b>	<b>0.779</b>	0.821
MAC	<b>0.750</b>	0.755	0.719	0.782
MAC+LGA	0.739	<b>0.760</b>	<b>0.727</b>	<b>0.783</b>



表3 对比实验性能比较(英文数据集)

Table 3 Comparative Experiment Results (English Dataset)

Model	English			
	<i>Acc</i>	<i>MacF1</i>	<i>F1<sub>fake</sub></i>	<i>F1<sub>real</sub></i>
Bi-LSTM	0.708	0.704	0.686	0.710
Bi-LSTM+LGA	<b>0.711</b>	<b>0.710</b>	<b>0.725</b>	<b>0.712</b>
EANN	<b>0.748</b>	0.692	0.669	0.681
EANN+LGA	0.719	<b>0.723</b>	<b>0.729</b>	<b>0.726</b>
BERT	0.700	0.690	0.702	0.672
BERT+LGA	<b>0.712</b>	<b>0.709</b>	<b>0.724</b>	<b>0.716</b>
BERT-Emo	0.719	0.717	0.715	0.715
BERT-Emo+LGA	<b>0.720</b>	<b>0.729</b>	<b>0.721</b>	<b>0.725</b>
S <sup>2</sup> GC	0.694	0.692	0.691	0.657
S <sup>2</sup> GC+LGA	<b>0.712</b>	<b>0.711</b>	<b>0.693</b>	<b>0.669</b>
DeClarE	0.711	0.706	<b>0.713</b>	0.709
DeClarE+LGA	<b>0.713</b>	<b>0.719</b>	0.705	<b>0.711</b>
MAC	0.701	0.693	0.703	0.703
MAC+LGA	<b>0.711</b>	<b>0.717</b>	<b>0.715</b>	<b>0.716</b>

从表2、表3可以观察到7个基线模型与LGA合作后在*Acc*、*macF1*、*F1<sub>fake</sub>*和*F1<sub>real</sub>*方面均有相应的提升。这既说明本文提出的LGA兼容性强,可以与各种虚假新闻检测器进行合作,又证明本文提出的方法可以有效提高检测精度。此外,实验结果表明LGA方法在纯文本的虚假新闻检测任务中,*F1<sub>fake</sub>*的识别率比*F1<sub>real</sub>*更高,这表明LGA可以更敏锐的识别虚假新闻的特征,更好的达到识别虚假新闻的目的。在引入知识表示的虚假新闻检测方法中,LGA同样可以提升基础模型的检测精度。

### 3.5.2 消融实验

为验证局部环境感知模块和全局环境感知模块对虚假新闻检测精度的双重影响,本节设计了以下两组消融实验。

表4 消融实验性能比较

Table 4 Ablation Experiment Performance Comparison

Model	Chinese			
	<i>Acc</i>	<i>MacF1</i>	<i>F1<sub>fake</sub></i>	<i>F1<sub>real</sub></i>
Global Environment	0.689	0.651	0.552	0.760
Local Environment	0.669	0.623	0.510	0.741
LGA	<b>0.691</b>	<b>0.667</b>	<b>0.561</b>	<b>0.764</b>

Model	English			
	<i>Acc</i>	<i>MacF1</i>	<i>F1<sub>fake</sub></i>	<i>F1<sub>real</sub></i>
Global Environment	0.689	0.687	0.697	0.682
Local Environment	<b>0.699</b>	0.692	0.697	0.694
LGA	0.698	<b>0.697</b>	<b>0.700</b>	<b>0.688</b>

表4代表了LGA方法脱离了其他的虚假新闻检测器,分别使用全局特征向量、局部特征向量和两者的聚合向量进行实验得到的结果。实验结果表明LGA可以脱离其他虚假新闻检测器独立进行工作,将其中的全局环境感知模块或局部环境感知模块单独从LGA中剥离出来也可以进行工作,但实验结果同时表明,仅把它们作为唯一的输出会导致检测精度略低,因此使用LGA和其他模型进行配合是更好的选择。

表5使用到的消融组旨在深入研究虚假新闻检测任务中局部和全局环境对检测精度的影响,通过对BERT-Emo+LGA模型和DeClarE+LGA模型进行消融实验可以更好的理解在不同的情境下(即考虑文本语境和引入外部知识库的情况下),LGA如何应对虚假新闻检测任务。实验分别删除这两个模型中的局部和全局环境感知模块,以便更清楚地了解它们对模型的重要性,以及它们如何协同工作从而提高虚假新闻检测的效果。实验结果表明,删除任意一个环境特征向量都导致检测精度下降。这说明无论是全局环境感知模块还是局部环境感知模块对于提高检测精度都起着必要的作用。它们在虚假新闻检测任务中相互合作,共同提高了实验的精度。

表5 环境检测模块对检测精度的影响

Table 5 Impact of Environment Detection Module on Accuracy

Model	Chinese			
	<i>Acc</i>	<i>MacF1</i>	<i>F1<sub>fake</sub></i>	<i>F1<sub>real</sub></i>
BERT-Emo+LGA	<b>0.838</b>	<b>0.826</b>	<b>0.807</b>	<b>0.857</b>
W/O Global Environment	0.829	0.816	0.792	0.844
W/O Local Environment	0.829	0.826	0.794	0.846
DeClarE+LGA	<b>0.799</b>	<b>0.798</b>	<b>0.776</b>	<b>0.822</b>
W/O Global Environment	0.779	0.772	0.732	0.807
W/O Local Environment	0.769	0.774	0.734	0.798

Model	English			
	<i>Acc</i>	<i>MacF1</i>	<i>F1<sub>fake</sub></i>	<i>F1<sub>real</sub></i>
BERT-Emo+LGA	0.720	<b>0.725</b>	<b>0.718</b>	<b>0.726</b>
W/O Global Environment	<b>0.750</b>	0.691	0.688	0.689
W/O Local Environment	0.715	0.723	0.716	0.723
DeClarE+LGA	<b>0.710</b>	<b>0.718</b>	0.712	<b>0.715</b>
W/O Global Environment	0.709	0.707	0.706	0.707
W/O Local Environment	0.700	0.700	<b>0.714</b>	0.697

### 3.6 新闻时间间隔对于全局环境感知的影响

为验证目标新闻的发布时间对于虚假新闻检测精度的影响,本实验在BERT-Emo+LGA模型上分别验

证了发布时间为  $T=1、3、5、7、9$  (天) 时的新闻识别准确度, 目的是寻找合适的  $T$  值, 使得 LGA 模型可以发挥最大的作用。寻找合适的  $T$  值对于感知模型中的相似度运算, 以及输入多层感知器的向量数量都有影响, 特别是在局部环境感知模型中直接关系到局部环境中的新闻数量, 故该过程具有重要意义。

图 2 和图 3 分别反映了全局环境中的新闻数量和新闻识别准确率随着不同  $T$  值的变化。由图 2 可知, 全局环境中的新闻随着  $T$  值的增加在逐渐增多。由图 3 可知, 当  $T=3$  时准确度最高。这个结果也符合大众的认知, 因为刚发布的新闻需要时间的发酵, 才会进入公众的视野, 流行程度才会提高, 但发布时间过长的新闻也会会被大众慢慢忘记。

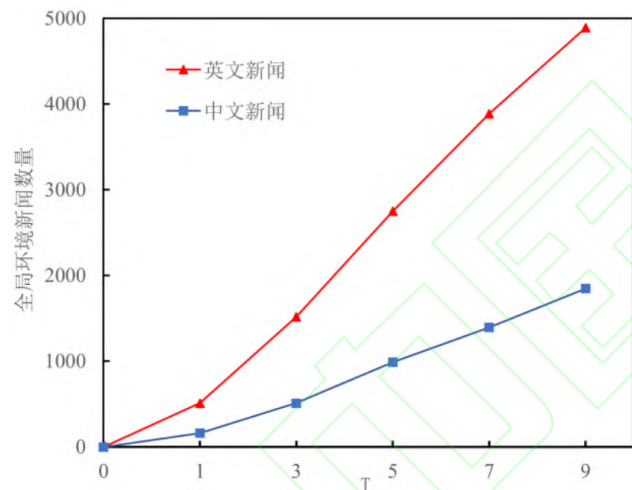


图 2 不同  $T$  值对全局环境中新闻数量的影响

Fig.2 Influence of different  $T$  values on the number of news in global environment

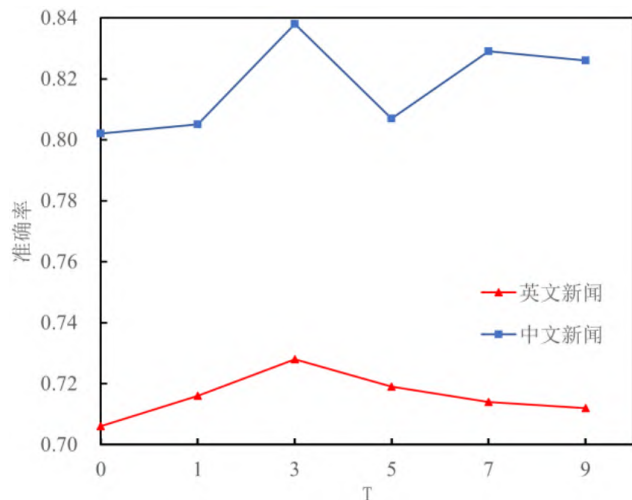


图 3 不同  $T$  值对准确率的影响

Fig.3 Influence of different  $T$  values on accuracy

### 3.7 超参数实验

在上一节中我们得出当  $T=3$  时, 实验效果最好, 故本节将  $T$  设置为 3。在研究讨论超参数  $r$  对于实验的影响时。实验将  $r$  的初始值设置为 0.05, 步长设置为 0.05, 到 0.30 时中止。实验结果如图 4、图 5 所示。

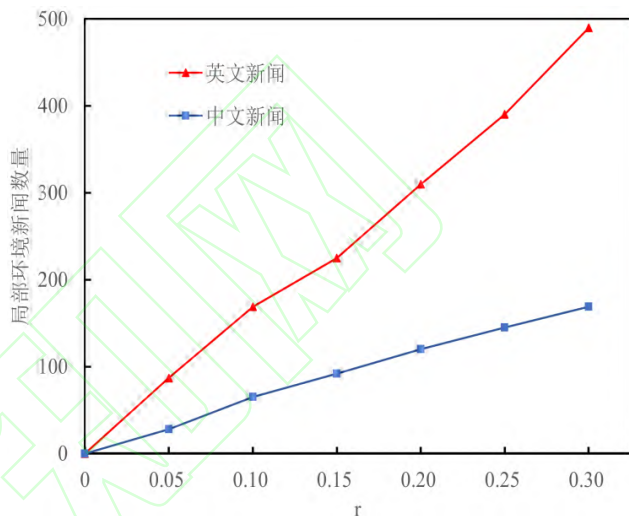


图 4 不同  $r$  值对局部环境中新闻数量的影响

Fig.4 Influence of different  $r$  values on the number of news in local environment

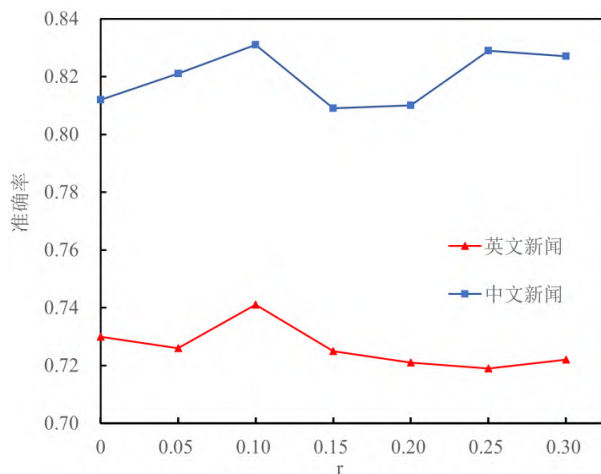


图 5 不同  $r$  值对准确率的影响

Fig.5 Influence of different  $r$  values on accuracy

实验表明当  $r$  增大时, 局部环境中的新闻数量逐渐变多, 但是实验得出的虚假新闻检测准确度并不是线性增长, 而是随着  $r$  的增长而波动。基于此进行分析, 当  $r$  值过小时, 局部环境中的新闻数量较小, 导致与目标新闻  $N$  内容相似的新闻覆盖率不高, 因此效果不太理想。但当  $r$  值过高时, 局部环境中新闻数量变多也带来了许多与目标新闻关系不明显的干扰新闻。

因此过高的  $r$  在带来极少收益的同时, 也会降低模型的性能。由图可知, 当  $r=0.1$  时实验效果在给定的范围内最好。

## 4 总结

通过将当前社会的时事热点、用户言论等元素融合进了虚假新闻检测的研究中, 本文提出的 LGA 方法分析了目标新闻在当前新闻环境中的相似程度和差异程度来辅助其他虚假新闻检测模型提高检测精度。实验表明, 提出的方法可以提高现有模型的检测精度。消融实验也表明, LGA 方法中的全局环境感知模块以及局部环境感知模块均可以不同程度的提高检测准确度。在未来的研究中, 我们将设计更加精确的数据收集策略, 以减少数据集与实际情况间的差异, 从而获得更具有代表性的全局环境和局部环境。同时, 我们将致力于进一步改进提出的方法, 尝试引入其他特征, 并研究更加高效的特征融合方法, 以提高虚假新闻的识别准确率。

## 参考文献

- [1] 毛震东, 赵博文, 白嘉萌, 等. 基于传播意图特征的虚假新闻检测方法综述[J]. 信号处理, 2022, 38(6):1155-1169. MAO Z D, ZHAO B W, BAI J M, et al. Review of fake news detection methods based on the features of propagation intention[J]. Journal of Signal Processing, 2022, 38(6):1155-1169.
- [2] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence, 2016:3818-3824.
- [3] 毕蓓, 潘慧瑶, 陈峰, 等. 基于异构图注意力网络的微博谣言监测模型[J]. 计算机应用, 2021, 41(12):3546-3550. BI B, PAN H, CHEN F, et al. Microblog rumor detection model based on heterogeneous graph attention network[J]. Journal of Computer Applications, 2021, 41(12):3546-3550.
- [4] 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. GAO Y J, LIANG G, JIANG F T, et al. Social network rumor detection: a survey[J]. Acta Electronica Sinica, 2020, 48(7):1421-1435.
- [5] KUMAR A, SANGWAN S R. Rumor detection using machine learning techniques on social media[C]//Proceedings of the Conference on Innovative Computing and Communications Berlin, 2019: 213-221.
- [6] KUMAR S, CARLEY K M. Tree lstrms with convolution units to predict stance and rumor veracity in social media conversations[C]//Proceedings of the 57th Annual Meeting of The Association for Computational Linguistics, 2019: 5047-5058.
- [7] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C]//Proceedings of the 20th International Conference on World wide web, 2011: 675-684.
- [8] YANG F, LIU Y, YU X, et al. Automatic detection of rumor on sina weibo[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data, 2012: 1-7.
- [9] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media[C]//Proceedings of the 13th International Conference on Data Mining, 2013: 1103-1108.
- [10] CHEN T, YIN H, CHEN H, et al. Air: attentional intention-aware recommender systems[C]//Proceedings of the 35th International Conference on Data Engineering, 2019: 304-315.
- [11] YU F, LIU Q, WU S, et al. A convolutional approach for misinformation identification[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017:3901-3907.
- [12] 范昊, 何灏. 融合上下文特征和 BERT 词嵌入的新闻标题分类研究[J]. 情报科学, 2022, 40(6):90-97. FAN H, HE H. News title classification based on contextual features and bert word embedding[J]. Information Science, 2022, 40(6): 90-97.
- [13] HU Y, PAN Q, HOU W, et al. Rumor spreading model with the different attitudes towards rumors[J]. Physica A: Statistical Mechanics and its Applications, 2018, 502: 331-344.
- [14] SHIRALKAR P, FLAMMINI A, MENCZER F, et al. Finding streams in knowledge graphs to support fact checking[C]//Proceedings of the IEEE International Conference on Data Mining, 2017: 859-864.
- [15] PAN J Z, PAVLOVA S, LI C, et al. Content based fake news detection using knowledge graphs[C]//Proceedings of the 17th International Semantic Web Conference, 2018: 669-683.
- [16] CUI L, SEO H, TABAR M, et al. Deterrent: knowledge guided graph attention network for detecting healthcare misinformation[C]//Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, 2020: 492-502.
- [17] 郭秋实, 李晨曦, 刘金硕, 等. 引入知识表示的图卷积网络谣言检测方法[J]. 计算机应用研究, 2022, 39(7):2032-2036. GUO Q S, LI C X, LIU J S, et al. Rumor detection

- with knowledge representation and graph convolutional network.[J]Application Research of Computers, 2022,39(7):2032-2036.
- [18] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019: 4171-4186.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31th International Conference on Neural Information Processing, 2017: 6000- 6100.
- [20] VOSOUGH S, ROY D, ARAL S. The spread of true and false news online[J]. science, 2018, 359(6380): 1146-1151.
- [21] SONG C, YANG C, CHEN H, et al. Ced: credible early detection of social media rumors[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(8): 3035-3047.
- [22] SHENG Q, CAO J, ZHANG X, et al. Article reranking by memory-enhanced key sentence matching for detecting previously fact-checked claims[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Proceedings, 2021:5468-5481.
- [23] KOCHKINA E, LIAKATA M, ZUBIAGA A. All-in-one: Multi-task learning for rumor verification[C]//Proceedings of the 27th International Conference on Computational Linguistics, 2018:3402-3413.
- [24] AUGENSTEIN I, LIOMA C, WANG D, et al. Multifc: a real-world multi-domain dataset for evidence-based fact checking of claims[C]//Proceedings of the 9th International Joint Conference on Natural Language Processing, 2019: 4685-4697.
- [25] SHENG Q, CAO J, ZHANG X, et al. Zoom out and observe: news environment perception for fake news detection[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022: 4543- 4556.
- [26] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional lstm and other neural network architectures[J]. Neural Network, 2005.18(5-6): 602-610.
- [27] WANG Y, MA F, JIN Z, 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, 2018: 849-857.
- [28] ZHANG X, CAO J, LI X, et al. Mining dual emotion-for fake news detection[C]//Proceedings of the web conference, 2021: 3465-3476.
- [29] ZHU H, KONIUSZ P. Simple spectral graph convolution[C]//International Conference on Learning Representation, 2021:151-163.
- [30] POPAT K, MUKHERJEE S, YATES A, et al. Declare: debunking fake news and false claims using evidence-aware deep learning[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2018: 22-32.
- [31] VO N, LEE K. Hierarchical multi-head attentive network for evidence-aware fake news detection[C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, 2021: 965-975.