

## 基于增强对抗网络和多模态融合的谣言检测方法

唐 樾, 马 静

(南京航空航天大学 经济与管理学院, 江苏 南京 211106)

**摘要:**【目的/意义】随着社交网络的复杂化,当前谣言往往是由描述事件的文本、对应的图片或者视频组成,多种模态的谣言更容易给用户传达一种错误的认知。现有谣言检测的研究往往只使用谣言文本特征,且未能充分挖掘谣言与事件存在的联系。【方法/过程】因此本文提出一种基于增强对抗网络和多模态融合的谣言检测方法,使用BERT和Text-CNN提取文本特征,使用VGG-19网络提取图像特征,再通过注意力机制捕捉多个模态的特征交互,最后使用增强对抗网络来挖掘谣言和事件之间联系。【结果/结论】在公开的微博多模态数据集上进行对比实验,实验结果表明该方法检测的准确率达到92.5%,相较于传统单模态和现有多模态模型,提升了约10%~20%。【创新/局限】本文将对抗网络和多模态特征融入谣言检测中,有效提升了谣言检测的效果,但目前仅尝试了文本和图像两种模态的结合,如何融合更多模态的特征后续有待研究。

**关键词:**多模态融合;对抗网络;谣言检测;深度学习;自然语言处理

**中图分类号:**G206.3      **DOI:**10.13833/j.issn.1007-7634.2022.06.014

## 1 引言

随着多媒体技术的不断发展,微博、Twitter等社交媒体已逐渐成为热点事件和新闻的主要载体。与传统的纯文本内容相比,以文本、图片等多种模态呈现的信息可以更好地表达事件的内容,吸引读者的注意力。同时,由真实事件与虚假内容组合的谣言,更具有迷惑性,也很容易煽动人们广泛传播,使得网络平台很难进行有效监管,对社会造成负面的影响。

谣言通常是指在网络平台产生、传播、影响的某阶段或全过程起到过关键作用的,内容未经证实的,且造成了一定社会舆论影响的阐释<sup>[1]</sup>。谣言所阐述的事件并不一定是假的,大多数谣言往往是由真实事件和未经证实的虚假内容衍生而来。例如2011年日本福岛核电站泄露为真实事件,但是与伴随该事件衍生的“吃碘盐可以防辐射”的言论为谣言。

谣言会引起不明公众的恐慌,严重的甚至会导致社会短暂的动荡,带来巨大的经济和社会影响。因此及时有效检测谣言对维持社会的安定和健康发展而言具有重要意义。目前,微博通常以用户举报结合人工审核方式进行谣言检测,会极大地耗费人力并且及时性不强,因此谣言自动检测具有极大的社会意义。

现阶段的谣言检测方法大多数都是基于单模态的文本

内容<sup>[2]</sup>,通常是从相关的文本信息中提取特征进行谣言的甄别,使用多模态方法的谣言检测研究较少<sup>[3]</sup>。现有模型倾向于捕获针对某特定事件的特征,而这类特征在不同的事件中是无法共享的。这种基于事件的特性,虽然能够帮助对已知谣言进行检测,但会影响对未知谣言的检测。

因此,本文提出一种基于增强对抗网络和多模态融合的谣言检测模型,通过并行网络提取各模态的特征,并引入事件对抗网络来学习所有事件之间的共享特征,从而实现挖掘各模态和事件之间的关联。

本文的主要贡献和创新点:①提出一种基于多模态融合的谣言检测模型,该模型通过注意力机制捕获模态之间的特征关联,从而实现模态间的信息交互;②将基于事件的增强对抗网络与多模态网络相结合,不仅能捕捉特定事件的特征,并且可以学习到模态特征与事件之间的关联;③在公开的微博多模态数据集上进行了大量的实验,实验结果表明,所提出的模型能够有效地识别谣言,在各指标效果上优于传统检测模型。

## 2 相关工作

国内外许多学者都针对谣言检测领域进行了深入的研究。早期学者们致力于研究如何有效提取谣言文本特征和传播特性<sup>[4]</sup>。

收稿日期:2021-07-03

**基金项目:**国家自然科学基金“基于多模态特征融合的网络舆情智能发现与检测及演化研究”(72174086);南京航空航天大学前瞻性发展策略研究基金项目“基于网络舆情大数据的我国跨境电商营商环境优化策略研究”(NW2020001);研究生开放基金“基于BERT模型的短文本分类方法与应用研究”(kfj20200907)。

**作者简介:**唐樾(1997-),男,湖北黄冈人,硕士生,主要从事自然语言处理和机器学习研究;马静(1968-),女,重庆人,教授,博士生导师,主要从事大数据网络舆情分析、企业信息化研究,通讯作者:majing5525@126.com。

早期谣言检测主要基于人工提取的有效特征,包括谣言的语言特征,上下文特征以及传播特征等等。Castillo(2011)等<sup>[5]</sup>统计了标点符号、特殊字符、词频分布等文本特征来进行谣言检测。Qazvinian(2011)等<sup>[6]</sup>通过统计信息、用户、主题和传播四个维度的特征来分析谣言的可信度。Kwon(2013)等<sup>[7]</sup>提出了基于时间周期的谣言检测模型,引入时间特征提升谣言检测的效果。Takahashi(2012)等<sup>[8]</sup>对 Twitter 谣言进行分析,发现谣言转发率远高于正常文本,而且两者关键词分布也存在明显差异。

由于人工提取特征较为复杂,学者们逐渐开始使用深度学习方法来提取特征,目前经典的深度学习模型主要分为以下三大类:卷积神经网络 CNN<sup>[9]</sup>、循环神经网络 RNN<sup>[10]</sup>及其变式 LSTM<sup>[11]</sup>和 GRU<sup>[12]</sup>、结合注意力的衍生模型。刘政(2017)等<sup>[13]</sup>提出了基于卷积神经网络(CNN)的谣言检测模型,对谣言事件进行向量化处理,通过隐含层的学习训练来挖掘文本深层特征。Ma(2016)等<sup>[14]</sup>提出利用递归神经网络(RNN)来捕捉潜在的时空语义特征,在 Twitter 数据集上取得较好效果。Zhou(2018)等<sup>[15]</sup>提出 C-GRU 谣言检测模型,使用 CNN 提取微博的深层文本特征,利用 GRU 关联事件下相关微博的时间序列信息。Ma(2018)等<sup>[16]</sup>又提出一种将树结构和递归神经网络结合的模型,分析推文内容和评论信息,取得了很好的效果,但是该方法依赖于一定的评论和转发信息,因此存在一定的延后性。Liu(2018)等<sup>[17]</sup>通过对传播路径进行分类,并引入 CNN 和 RNN,建立通过时间序列来提取传播路径中用户特征的谣言检测模型。Liao(2018)等<sup>[18]</sup>将微博分割成不同间隔,通过两层带有注意力机制的 GRU 网络来分别获取微博和时间段序列的潜在特征。

随着 BERT<sup>[19]</sup>等预训练模型的发展,国内学者尝试将模型特征迁移到谣言检测中。李悦晨(2021)等<sup>[20]</sup>尝试用 BERT 模型对微博原文进行向量表示,并使用 RCNN 模型进行谣言检测,取得了较好检测效果。梁兆君(2021)等<sup>[21]</sup>提出基于 BERT 模型的增强混合神经网络的谣言检测方法,使用 BERT 模型将推文向量化,并卷积池化后的特征输入到

BiLSTM 中学习序列特征,同时引入 Attention 机制计算注意力分值,实现谣言检测。

以上工作都是基于单模态的谣言检测研究,如今社交媒体的内容形式通常由文本、图像和语音等多模态组成。

多模态研究主要有两个拟解决的问题:一是如何有效提取各模态的特征;二是如何进行模态间的特征充分融合。现有研究已证实视觉内容可以提供有效信息,Jin(2016)等<sup>[22]</sup>根据视觉特征和统计特征来识别谣言,发现假新闻事件和真实新闻事件具有不同的图像分布模式。随后,Jin(2017)等<sup>[23]</sup>又提出了一种具有注意机制(att-RNN)的递归神经网络,用于融合多模式特征以进行有效的谣言检测。该方法在检测多模态谣言上取得了良好的效果。Wang(2018)等<sup>[24]</sup>提出了一个名为事件对抗神经网络(EANN)的端到端框架。它消除了对数据集中特定事件的紧密依赖,挖掘了谣言本身和事件的关联,提高了对未知事件的泛化能力。Khattar(2019)等<sup>[25]</sup>为解决谣言检测的方法无法学习多模式信息的共享表示的缺点,提出了一种端到端网络 MVAE,该网络将双模态可变自动编码器与二进制分类器结合使用以进行谣言检测,在微博和 Twitter 两个数据集上进行了实验,在 F1 值和准确率上超越了 SOTA 方法。

国内学者也开始了多模态谣言检测的研究,威力鑫(2021)等<sup>[26]</sup>为解决多模态谣言检测方法没有关注词与图片区域对象之间的特征融合的问题,提出了一种基于注意力机制的多模态融合网络 AMFNN 应用于谣言检测,利用注意力机制捕捉与关键词语相关的视觉特征。陶霄(2020)等<sup>[27]</sup>提出了基于注意力机制与 Dempster 组合规则的混合融合方法 DHF。在前期融合和后期融合均加入注意力机制,实现特征和决策的自动加权,并使用 Dempster 组合规则实现混合融合。刘金硕(2021)等<sup>[28]</sup>提出了一种基于多模态网络谣言检测方法 MSRD,有效解决忽略了谣言图像内容以及图像中的内嵌文本内容的问题。张少钦(2021)等<sup>[29]</sup>提出了一个基于深度学习的端到端的多模态融合网络 MFN,弥补了没有充分挖掘出视觉特征和融合文本与视觉的联合表征特征的

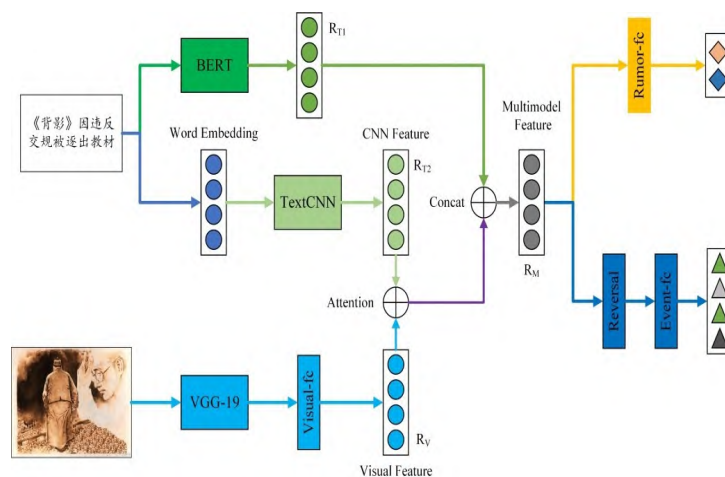


图1 EANMF模型结构图

Figure 1 The structure of the enhance adversarial network and multimodal fusion (EANMF) model

不足。

基于以上相关研究,本文提出一种基于增强对抗网络和多模态的谣言检测方法(EANMF),该模型能自动学习多模态特征表示,充分挖掘文本特征和视觉特征的联合表征,还能利用对抗网络生成事件的特征表示。

### 3 基于增强对抗网络和多模态融合的谣言检测方法

本文提出的基于增强对抗网络和多模态融合的谣言检测方法(EANMF),其主要包含以下4个模块,其模型结构如图1所示。

**文本特征提取器:**该模块采用Text-CNN和BERT提取文本特征,将Text-CNN特征向量与图像特征进行注意力融合,BERT特征向量作为文本特征的加强,与融合后的多模态特征进行拼接。

**图像特征提取器:**该模块采用VGG-19网络提取图像特征,经过一个Visual-fc的全连接层后,与文本特征进行注意力融合。

**事件对抗网络:**该网络为事件鉴别器,它由两个全连接层组成,并具有相应的激活函数。它的目的是基于多模态特征表示对K个事件的后分类。

**谣言检测网络:**该网络为谣言判别器,通过一个全连接层和相应的激活函数后,可实现谣言检测的目的。

#### 3.1 文本特征提取器

该模块的输入为由单词序列组成的原始文本,为了有效提取文本语义特征,本文采用Text-CNN和BERT作为文本特征提取器的两个核心模块。其中BERT模型能有效提取文本深层语义特征,Text-CNN通过不同大小的滑动窗口,能关注到局部视野和细粒度的特征。

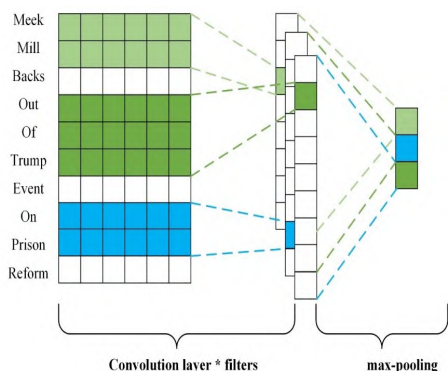


图2 Text-CNN模型结构图

Figure 2 The structure of Text-CNN model

对于句子中的第*i*个单词,对应的第*k*维单词嵌入向量记为 $T_i \in \mathbb{R}^k$ 。因此,由*n*个单词组成的句子可以表示为公式(1):

$$T_{1:n} = T_1 \oplus T_2 \oplus \dots \oplus T_n \quad (1)$$

Text-CNN模块其结构如图2所示,通过Word2vec将文

本进行词向量化。每个词采用搜狗公开预训练好的模型,嵌入到对应数据集上进行初始化。

卷积神经网络(CNN)的核心思想是捕捉局部特征,对于文本来讲,局部特征就是由若干单词组成的滑动窗口,类似于N-gram。CNN的优势在于能够自动地对不同粒度特征进行组合和筛选,从而获得不同抽象层次的语义信息。假设一个窗口大小为*h*的卷积滤波器将句子中*h*个单词的连续序列作为输入输出一个特征, $A[i:j]$ 表示*A*的第*i*行到第*j*行,那么卷积操作可以用如下公式(2)表示:

$$t_i = \sigma(W_c \cdot T_{i:i+h-1}) \quad (2)$$

其中 $\sigma$ 为表示激活函数,通常用到的激活函数有Sigmoid、Tanh和RELU等, $W_c$ 表示过滤器的数量,过滤器的参数为全局共享,从而得到整个句子的特征向量,如公式(3)所示:

$$T = [t_1, t_2, \dots, t_{n-h+1}] \quad (3)$$

对于每个特征向量*t*,采用最大池化的方式,从而提取关键的信息,得到了特定过滤器的对应特征。为了提取不同粒度的文本特征,设置了(2、3、4)不同的窗口大小的卷积核。对于一个特定的窗口大小,使用 $n_h$ 个不同的过滤器。假设有*c*个可能的窗口大小,总共有 $c \cdot n_h$ 过滤器。经过最大池化操作后的文本特征表示成 $R_T \in \mathbb{R}^{c \cdot n_h}$ ,如公式(4)所示。

$$R_T = \sigma(W_{fc} \cdot R_T) \quad (4)$$

BERT模型是Google在2018年公开推出的一个语言理解模型,其结构由双向Transformer<sup>[30]</sup>的编码器结构堆叠而成,已在命名实体识别文本分类、情感分析、智能问答等多个NLP领域中取得出色效果。BERT可以学习每个单词前后两侧的上下文信息,从而获得更全面的词向量表示,其结构如图3所示。

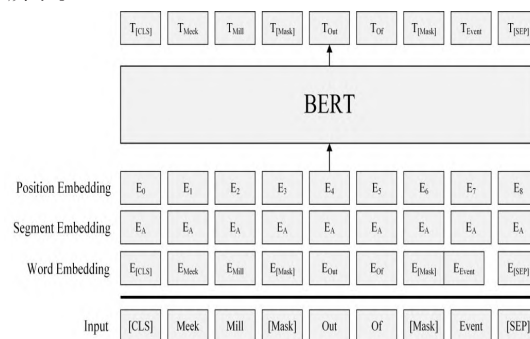


图3 BERT模型结构图

Figure 3 The structure of BERT model

其主要结构为自注意力机制,通过计算单词间的相互关系,得到每个词与其他词的权重来达到表现词之间的关联程度,可以实现一词多义的区分,计算公式(5)如下:

$$Attention(Q, K, V) = \text{soft max} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (5)$$

#### 3.2 图像特征提取器

文本内容相关的图像是该模块的输入,可记为*V*。为了



有效地提取视觉特征,采用了预先训练的VGG19来提取图像特征。在VGG19网络的最后一层之上,添加了一个全连接层,将最终视觉特征表示的维度调整为 $q$ 。在图像单独训练的时候将VGG19网络的参数进行更新和迭代,以适应谣言图像的数据集,但在与文本特征提取器联合训练的过程中,将预先训练的VGG19网络的参数进行freeze冻结处理,避免融合过程中的过拟合。将视觉特征表示为 $R_v \in \mathbb{R}^q$ ,则图像特征提取器中最后一层的操作可以表示为公式(6):

$$R_v = \sigma(W_{vf} \cdot R_{vgg}) \quad (6)$$

其中, $R_{vgg}$ 为预先训练的VGG19得到的视觉特征表示, $W_{vf}$ 为视觉特征提取器中全连通层的权重矩阵。将文本特征表示 $R_t$ 和视觉特征表示 $R_v$ 进行注意力加权计算,形成多模态特征表示为 $R_f = R_t \oplus R_v \in \mathbb{R}$ ,这是多模态特征提取器的输出。

### 3.3 多模态特征融合

多模态融合指的是“深度学习中的多模态融合技术是模型在完成分析和识别任务时处理不同形式的数据的过程”,能够获得更全面的特征,挖掘出模态之间的更深层次的关联。多模态特征融合主要分为前期、中期和后期融合,Baltrušaitis(2018)<sup>[31]</sup>和何俊等(2020)<sup>[32]</sup>已经有研究对多模态的融合技术研究进行了综述。早期融合是在将多个模态的数据融合成一个单一的特征向量,然后输入到分类器中。中间融合是指将不同的模态数据先转化为高维特征表达,再于模型的中间层进行融合。后期融合是将每种模态信息单独进行分类训练,最后将输出打分进行融合决策。

一般的融合方式将各个模态的特征进行简单的拼接,没有考虑各个模态的重要性和模态之间的特征关联。本文采用注意力机制将文本特征和视觉特征进行融合,该方法可以计算文本中的每个单词与图像各候选区域的潜在联系,从而对两者进行加权联合,使得各模态特征之间相互补充和关联,例如文本内容可能表现积极乐观,而图像的主体区域表现一种嘲讽或者不甘,两者之间存在矛盾,说明该信息为谣言的可能性更高。采用自注意力机制,分别对词语和视觉特征进行加权,能有效捕捉这种特征之间的矛盾。其计算过程如式(7-8)所示:

$$Atten(T, V) = \text{soft max}(W_h^*(R_t \otimes R_v) + b) \quad (7)$$

$$R_M = F_T(R_t \otimes Atten) \oplus F_V(R_v \otimes Atten) \quad (8)$$

其中, $Atten(T, V)$ 为文本特征, $R_t$ 与视觉特征, $R_v$ 计算的联合注意力权值,通过soft max函数进行权值归一化, $R_M$ 为两个模态的特征与联合注意力的点乘,作为融合特征的输出。

### 3.4 事件对抗网络

该模块的主要结构参考EANN<sup>[24]</sup>,事件对抗网络由两个完全连接层组成,并采用相应的激活函数进行转换。该模块的作用是基于多模态特征表示完成对 $K$ 个事件的分类,通过

参数共享来挖掘事件与谣言之间的关联,消除了对所收集数据集中特定事件的严格依赖性,并实现了对未见事件的更好的泛化能力。将事件鉴别器定义为 $G_e(R_f; \theta_e)$ , $\theta_e$ 表示其学习的参数,采用交叉熵(Cross-Entropy)定义事件鉴别器的损失函数,用 $Y_e$ 表示事件的标签集,其损失函数可由公式(9)表示,最小化损失函数表示为公式(10):

$$L_e(\theta_f, \theta_e) = -\mathbb{E}_{(m, y) \sim (M, Y_e)} \left[ \sum_{f=1}^K P_{[k=y]} \log(G_e(G_f(m; \theta_f); \theta_e)) \right] \quad (9)$$

$$\hat{\theta}_e = \arg \min_{\theta_e} L_e(\theta_f, \theta_e) \quad (10)$$

损失函数 $L_e(\theta_f, \theta_e)$ 可以用来估计不同事件分布的差异性。损失越大意味着不同事件表示的分布是相似的,学习的事件特征是具有事件不变性。因此,为了挖掘事件与谣言的关联,需要通过梯度下降法来寻找最优参数 $\theta_f$ 来最大化损失函数 $L_e(\theta_f, \theta_e)$ 。

该损失可用来估计不同事件分布的差异性。损失越大意味着不同事件的表示形式的分布是相似的,并且获悉的特征是事件不变的。因此,为了消除每个事件的唯一性,让模型无法区分事件来达到消除事件特定的特征的目的,需要通过寻找最佳参数来使鉴别损失最大化。一方面,两个特征提取器试图欺骗事件鉴别器以使鉴别损失最大化;另一方面,事件鉴别器的目的是发现特征表示中包含的事件特定信息来识别事件。

### 3.5 谣言检测网络

该模块为实现谣言检测的网络,将两个特征提取器的特征融合之后输入到该网络,经过一个soft max全连接层来判断是否为谣言事件。该网络是建立在多模态特征提取器的基础上,以多模态特征表示 $R_f$ 作为输入,将谣言判别器表示为 $G_d(R_d; \theta_d)$ ,如公式(11)所示:

$$y_\theta(m_i) = G_d(R_d; \theta_d) \quad (11)$$

该网络的目的是识别特定的事件是否为谣言,其中 $\theta_d$ 表示该网络的所有参数, $y_\theta(m_i)$ 为当前事件为谣言的概率。用 $y_d$ 表示谣言数据的标记集合,用sigmoid交叉熵来计算检测损失,公式(12)如下:

$$L_d(\theta_f, \theta_d) = -\mathbb{E}_{(m, y) \sim (M, Y_d)} \left[ y \log(y_\theta(m)) + (1 - y) (\log(1 - y_\theta(m))) \right] \quad (12)$$

其中, $M$ 表示由文本和图像组成的消息集合, $Y_d$ 表示谣言数据的标记集合通过寻找最优参数来最小化谣言分类的损失,公式(13)表示如下:

$$(\hat{\theta}_f, \hat{\theta}_d) = \arg \min_{\theta_f, \theta_d} L_d(\theta_f, \theta_d) \quad (13)$$

谣言检测网络的难点在于如何训练数据集未涉及的事件,这要求模型能够学习更多通用特征表示,以捕获所有事件之间的共同特征。这样的表示应该是事件不变的,并且不包括任何事件特定的功能,而事件对抗网络与谣言检测网络

的结合能有效解决该问题。

## 4 实验与分析

### 4.1 实验数据集

本文采用<sup>[23]</sup>中公开的微博谣言数据集,该数据集的真实新闻来源于官方权威新闻媒体,如新华社等,包含从2012年5月到2016年1月间微博官方辟谣系统审核并验证的新闻;并去除重复和低质量的图像,以确保整个数据集的质量。为保证对比实验的合理性,仅使用同时具有文本和图像两个模态的数据部分,从而避免单模态和多模态实验的数据分布偏差。

本文将数据集按照7:1:2的比例拆分为训练集、验证集和测试集,其详细统计情况如表1所示。

表1 微博数据集分布

Table 1 Weibo dataset distribution

微博	谣言	非谣言	总数
训练集	2898	2517	5415
验证集	454	389	843
测试集	756	709	1465
总数	4108	3615	7723

### 4.2 对比实验设计

为了验证所提模型的有效性,本文设计三种对比模型:单模态模型、多模态模型和本文提出的EANMF模型。

单模态模型仅使用文本或者图像其中之一的单模态特征,实验模型如下:

Text-CNN(2014)<sup>[9]</sup>:采用所提出的经典模型Text-CNN作为文本特征提取模块,词向量使用搜狗实验室预训练的300维Word2vec模型。模型使用CNN提取文本特征,经过一个的全连接层来实现谣言检测。

BERT(2018)<sup>[19]</sup>:BERT-BASE使用谷歌官方提供的中文基础预训练模型的chinese\_L-12\_H-768\_A-12.zip。模型参数大小为110M,同样经过一个soft max的全连接层来实现谣言检测。

Visual:采用预先训练的VGG-19(2014)<sup>[33]</sup>和全连接层提取视觉特征 $R_v$ 。将提取的视觉特征输入全连接层进行谣言检测,设置的全连接层的隐藏大小为128。

多模态谣言检测模型同时使用文本内容和配套的图像数据,采用多模态提取器提取特征并进行融合,达到谣言检测的目的,实验模型如下:

VAQ(2015)<sup>[34]</sup>:该视觉问答模型的任务是根据给定的图像回答问题。原始的VQA模型为多分类任务过将最后一层的soft max分类层进行修改,以适应谣言检测任务。

att-RNN(2017)<sup>[23]</sup>:采用LSTM提取文本特征,采用VGG-19模型提取视觉特征,利用注意力机制融合文本和视觉特征,将两者进行合并以实现谣言检测任务。

EANN(2018)<sup>[24]</sup>:通过Text-CNN模型提取文本特征,图片特征使用预训练的VGG-19模型提取,并加入了对抗网络以学习事件的不变表示,有效提高谣言检测的准确性。

BEANN:将EANN中的Text-CNN模型替换为BERT模型,其他结构保持不变。

MVAE(2019)<sup>[25]</sup>:提出一种端到端的多模态变分自编码器,使用双峰的变分自编码器二元分类器进行谣言检测任务,可以从学习到的共享表示中重构出两种模态,进而发现跨模态之间的关联。

EANMF:本文所提出的基于增强对抗网络和多模态融合的谣言检测方法,引入对抗网络来学习事件特性,通过BERT模型进行文本语义增强;并尝试采用拼接、加权求和以及注意力三种方式进行多模态特征融合。

实验过程如下,其中部分模型的参数和特征提取的概况如表2所示。

(1)单模态谣言检测实验。文本模态分别采用Text-CNN和BERT构建谣言检测模型,Text-CNN中采用Word2vec模型作为词嵌入,卷积核为常规的(2,3,4)大小,过滤器的数量设置为128;BERT模型对small(12层+768维)的中文base预训练模型进行fine-tune,开放所有层的权重更新。图像模态采用经典VGG-19网络,每张图像的维度会统一成224\*224大小,输出为标准的7\*7\*512的特征向量,经过拉平处理作为最后的输出。

(2)多模态谣言检测实验。五组对比模型与本文采用相同中文微博数据集,因此可保证数据层面的一致性,并且对比模型的结果取原论文表格中最佳效果;本文所提出的EANMF模型采用Text-CNN和BERT提取文本特征,采用VGG-19提取图像特征,与单模态模型中参数保持完全一致,保证与单模态对比的一致性。

表2 模型参数及特征概况

Table 2 Overview of model parameters and characteristics

模态 模型	文本			图像	
	Text-CNN	BERT		VGG-19	
参数	最大长度	32	-	32	输入维度 224*224
	批次大小	64	-	64	-
	卷积核	2,3,4	层数	12	卷积核 3,4,5
	过滤器	128	隐层大小	768	过滤器 128/256/512
	学习率	1e-3	-	2e-5	-
特征	输出维度	128*3	-	768	-
	全连接维度	128	-	128	-
	特征融合				128
维度	类别数				2

### 4.3 实验结果分析

为了验证模型的检测效果,本文采用二分类问题中常用的评价指标,准确率(Accuracy)、精确率(Precision)、召回率(Recall)和F1分值(F1-Score)对模型进行评价,实验对比结

果如表3和图4所示。

表3 微博数据集的实验结果

Table 3 Experimental results of Weibo dataset

模态	模型	Accuracy	Precision	Recall	F1
单模态	Text-CNN	0.763	0.827	0.683	0.748
	BERT	0.867	0.874	0.870	0.868
	Visual	0.615	0.615	0.677	0.648
	VAQ	0.773	0.780	0.782	0.781
	att-RNN	0.779	0.778	0.799	0.789
多模态	EANN	0.827	0.847	0.812	0.829
	MVAE	0.824	0.854	0.769	0.809
	BEANN	0.889	0.891	0.890	0.892
	EANMF(concat)	0.912	0.917	0.910	0.920
	EANMF(linear)	0.921	0.918	0.916	0.920
	EANMF(atten)	0.925	0.926	0.926	0.925

从表3来看,单模态模型中,文本的检测效果明显优于图像,这是因为在谣言检测任务中,文本能表达事件的主要内容,其涵盖的语义信息能直接帮助鉴别谣言,而单单只看图片是很难鉴别谣言的。文本模态中,BERT模型相对Text-CNN模型效果显著,提升了约10%,甚至优于大多数的多模态模型效果,这说明预训练模型对于文本语义特征的提取有十分显著的增强效果。

多模态模型中,EANN模型引入了事件对抗网络,相较于VAQ和att-RNN模型检测效果更好。并且EANN模型的文本特征是采用Text-CNN模型进行提取,较单模态的Text-CNN模型提升约6%,说明图像特征在一定程度上能对单文本特征进行补充,提升谣言检测的效果。BEANN将EANN模型中的Text-CNN替换为BERT模型,效果较EANN提升了约6%,说明BERT能增强整个对抗网络和检测网络的语义信息,帮助捕捉谣言文本的特征。

其中,本文提出的EANMF模型在各项指标均优于其他模型,三种特征融合方式略有差异,其中拼接(concat)的方法最为简单,其效果略差;线性加权(linear)的方法能在一定程度上自动学习文本和图像两种模特的特征权重;注意力(attention)能全局捕捉模特之间的信息交互,使得文本和图像特征能有效融合,取得了最优的效果。与单模型BERT模型相比提升约5%,较BEANN模型提升约3%,这是由于EANMF模型通过注意力机制能捕捉各模态间的信息交互,有效融合了多模态特征,使得模态内的特征组合更加丰富,同时BERT模型能对文本特征进一步加强,从而学习到事件特征和模态特征之间的关联。

综合来说,与单模态模型相比,多模态融合能在一定程度上提高谣言检测的效果,其中文本特征在谣言检测中起关键作用,图像特征对单文本特征进行一些补充。BERT模型对文本语义特征的增强效果显著;注意力机制能有效捕捉模态间的信息交互,充分挖掘了文本与图像之间的潜在联系;事件对抗网络能挖掘模型特征和事件间的关联,两者使得谣言检测的效果都具有显著提升。基于这些点,本文所提出的基于增强对抗网络和多模态融合的谣言检测模型,在微博数据集上的检测效果有了显著提升。

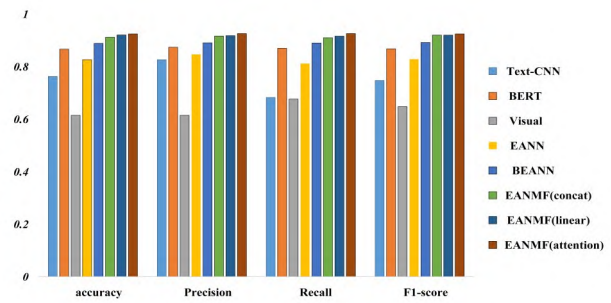


图4 对比实验结果

Figure 4 Comparison of experimental results

## 5 总结和展望

为解决社交媒体中多模态难以检测的问题,本文提出一种基于增强对抗网络和多模态融合的谣言检测模型,该模型通过注意力机制充分挖掘模态间的信息交互,并增强事件对抗网络来学习谣言与事件存在的关联,增强模型对未知事件的适应性。在微博数据集上进行对比实验,检测效果均优于单模态和其他多模态方法,表明模态间的信息交互和事件对抗信息,对于谣言检测具有一定的辅助作用。

但是目前仅尝试了文本和图像两种模型的结合,社交媒体上还有着许多短视频和语音数据,如何融合更多模态的特征,挖掘它们之间的潜在联系,是后未来工作中需要去深度研究和尝试的一个点。

## 参考文献

- 1 陈燕方,李志宇,梁循,齐金山.在线社会网络谣言检测综述[J].计算机学报,2018,41(7):1648-1677.
- 2 李奥,但志平,董方敏,刘龙文,冯阳.基于改进生成对抗网络的谣言检测方法[J].中文信息学报,2020,34(9):78-88.
- 3 金志成,曹娟,王博,王蕊,张勇东.融合多模态特征的社会多媒体谣言检测技术研究[J].南京信息工程大学学报(自然科学版),2017,9(6):583-592.
- 4 Kai Shu,Amy Sliva,Suhang Wang,Jiliang Tang,and Huan Liu.Fake news detection on social media:A data mining perspective[J].ACM SIGKDD Explorations Newsletter,2017(90):22-36.
- 5 Castillo C,Mendoza M,Poblete B.Information credibility on twitter[C]//Proceedings of the 20th international conference on World wide web,2011:675-684.
- 6 Qazvinian V,Rosengren E,Radev D,et al.Rumor has it: Identifying misinformation in microblogs[C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing,2011:1589-1599.
- 7 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. IEEE,



- 2013:1103-1108.
- 8 Takahashi T,Igata N.Rumor detection on twitter[C]// The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems. IEEE, 2012:452-457.
- 9 KIM Y.Convolutional Neural Networks for Sentence Classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).ACM,2014:1746-1751.
- 10 Mikolov T,Karafiát M,Burget L,et al.Recurrent neural network based language model[C]//Eleventh annual conference of the international speech communication association,2010: 1045-1048.
- 11 Hochreiter S,Schmidhuber J.Long short-term memory[J]. Neural computation,1997,9(8):1735-1780.
- 12 Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J].arXiv preprint arXiv:1406.1078,2014.
- 13 刘政,卫志华,张初弦.基于卷积神经网络的谣言检测[J]. 计算机应用,2017,37(11):3053-3056.
- 14 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.2016:3818-3824.
- 15 Zhou Z, Qi Y, Liu Z, et al. A C-GRU neural network for rumors detection[C]//2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS). IEEE,2018:704-708.
- 16 Ma J, Gao W, Wong K F. Rumor Detection on Twitter with Tree-structured Recursive Neural Networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 1980-1989.
- 17 Liu Y, Wu Y F B. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks[C]//32nd AAAI Conference on Artificial Intelligence, AAAI 2018. AAAI press, 2018: 354-361.
- 18 Liao X, Huang Z, Yang D, et al. Rumor detection in social media based on a hierarchical attention network[J]. SCIENTIA SINICA Informationis, 2018, 48(11): 1558-1574.
- 19 Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- 20 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.
- 21 梁兆君,但志平,罗衍潮,李奥.基于BERT模型的增强混合神经网络的谣言检测[J]. 计算机应用与软件, 2021, 38(3): 147-152, 189.
- 22 Jin Z, Cao J, Zhang Y, et al. Novel visual and statistical image features for microblogs news verification[J]. IEEE transactions on multimedia, 2016, 19(3): 598-608.
- 23 Jin Z, 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., 2017: 795-816.
- 24 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.
- 25 Khattar D, Goud J S, Gupta M, et al. Mvae: Multimodal variational autoencoder for fake news detection[C]//The World Wide Web Conference, 2019: 2915-2921.
- 26 威力鑫,万书振,唐斌,徐义春.基于注意力机制的多模态融合谣言检测方法[J/OL]. 计算机工程与应用: 1-11[2021-08-24].<http://kns.cnki.net/kcms/detail/11.2127.TP.20210430.1407.008.html>.
- 27 陶霄,朱焱,李春平.基于注意力和多模态混合融合的谣言检测方法[J/OL]. 计算机工程: 1-7[2021-08-24].<https://doi.org/10.19678/j.issn.1000-3428.0059683>.
- 28 刘金硕,冯阔,Jeff Z. Pan,邓娟,王丽娜. 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.
- 29 张少钦,杜圣东,张晓博,李天瑞.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.
- 30 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: 5998-6008.
- 31 Baltrušaitis T, Ahuja C, Morency L P. Multimodal machine learning: A survey and taxonomy[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(2): 423-443.
- 32 何俊,张彩庆,李小珍,张德海.面向深度学习的多模态融合技术研究综述[J]. 计算机工程, 2020, 46(5): 1-11.
- 33 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- 34 Antol S, Agrawal A, Lu J, et al. Vqa: Visual question answering [C]//Proceedings of the IEEE international conference on computer vision, 2015: 2425-2433.

(责任编辑:赵红颖)

(下转第131页)

## Feature Mining and Relationship Prediction of Author Cooperation in the Field of Domestic Smart Library from SNA Perspective

PANG Yu-jing, LU Xin-yuan, ZHANG Heng

(School of Information Management, Central China Normal University, Wuhan 430079, China)

**Abstract:** [Purpose/significance] This paper aims to explore the current situation of scientific cooperation and the composition of research team among domestic scholars in the field of Smart Library, and tap the potential cooperative relationship between authors, in order to promote the cooperative research and practice in the field of Smart Library. [Method/process] Based on the perspective of SNA, using ucinet6 and coc9.9 as analysis tools, this paper analyzes 117 authors in the core author group in the field of Smart Library in China, and reveals the external and potential cooperation relations of scholars in the field of Smart Library from five dimensions: subnet, network density, centrality, cohesive subgroup and structural hole. [Result/conclusion] The research shows that most domestic scholars in the field of Smart Library conduct cooperative research in a small team of 2-3 authors, and there is a lack of communication between teams. The analysis of potential cooperative relations shows that there is still great potential and space for cooperation in this field. In the future, scholars in this field can strengthen cooperation around intelligent services, university libraries, the Internet of things, big data, artificial intelligence and other directions, promote the formation of large-scale interdisciplinary and cross regional teams and closer contacts among scholars, and promote the development of this field. [Innovation/limitations] It is the first time to explore the existing and potential cooperation relationships of scholars in the field of Smart Library from the perspective of SNA, and the potential relationships need to be further studied.

**Keywords:** smart library; author cooperation relationship; potential cooperation relationship; social network analysis; cooperation network

---

(上接第 114 页)

## A Rumor Detection Method Based on Enhance Adversarial Network and Multimodal Fusion

TANG Yue, MA Jing

(College of Economic and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

**Abstract:** [Purpose/significance] With the complexity of social networks, current rumors are often composed of text describing events, corresponding pictures or videos, and multimodal rumors are more likely to convey a wrong cognition to users. The existing research on rumor detection only uses the text features of rumor, and fails to fully explore the relationship between rumor and event. [Method/process] Therefore, this paper proposes a rumor detection method based on enhanced adversary network and multimodal fusion. It uses BERT and Text CNN to extract text features, VGG-19 network to extract image features, and then uses attention mechanism to capture multi-modal feature interaction. Finally, it uses enhanced adversary network to mine the relationship between rumors and events. [Result/conclusion] The experimental results show that the detection accuracy of this method reaches 92.5%, which is about 10~20% higher than that of traditional single-mode and existing multi-mode models. [Innovation/limitation] At present, only the combination of text and image modes has been tried. There are still many short video and voice data on social media. How to integrate more modal features needs to be studied in the future.

**Keywords:** multimodal fusion; adversarial network; rumor detection; deep learning; natural language processing