

## 多源异构数据渐进式融合虚假新闻检测

于泳欣, 纪科, 高源, 陈贞翔, 马坤, 赵晓凡

### 引用本文

于泳欣, 纪科, 高源, 陈贞翔, 马坤, 赵晓凡. 多源异构数据渐进式融合虚假新闻检测[J]. 计算机科学, 2024, 51(11): 30-38.

YU Yongxin, JI Ke, GAO Yuan, CHEN Zhenxiang, MA Kun, ZHAO Xiaofan. [Multi-source Heterogeneous Data Progressive Fusion for Fake News Detection](#) [J]. Computer Science, 2024, 51(11): 30-38.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于注意力机制和深层特征优化的混凝土路面裂缝检测](#)

Crack Detection of Concrete Pavement Based on Attention Mechanism and Deep Feature Optimization  
计算机科学, 2024, 51(11): 198-204. <https://doi.org/10.11896/jsjx.240100082>

#### [基于多模态自适应融合的短视频虚假新闻检测](#)

Multimodal Adaptive Fusion Based Detection of Fake News in Short Videos  
计算机科学, 2024, 51(11): 39-46. <https://doi.org/10.11896/jsjx.240700062>

#### [基于联合增强图像对的红外可见光深度展开图像融合网络](#)

Infrared and Visible Deep Unfolding Image Fusion Network Based on Joint Enhancement Image Pairs  
计算机科学, 2024, 51(10): 311-319. <https://doi.org/10.11896/jsjx.230800069>

#### [基于多尺度跨模态特征融合的图文情感分类模型](#)

Image-Text Sentiment Classification Model Based on Multi-scale Cross-modal Feature Fusion  
计算机科学, 2024, 51(9): 258-264. <https://doi.org/10.11896/jsjx.230700163>

#### [艺术美感增强的图像任意风格迁移](#)

Image Arbitrary Style Transfer via Artistic Aesthetic Enhancement  
计算机科学, 2024, 51(9): 129-139. <https://doi.org/10.11896/jsjx.230800098>

# 多源异构数据渐进式融合的虚假新闻检测

于泳欣<sup>1,2</sup> 纪科<sup>1,2</sup> 高源<sup>1,2</sup> 陈贞翔<sup>1,2</sup> 马坤<sup>1,2</sup> 赵晓凡<sup>3,4</sup>

1 济南大学信息科学与工程学院 济南 250022

2 山东省网络环境智能计算技术重点实验室 济南 250022

3 中国人民公安大学信息安全学院 北京 102623

4 安全防范技术与风险评估公安部重点实验室 北京 102623

(3106615421@qq.com)

**摘要** 社交媒体平台上充斥着大量未经验证的信息,这些信息大多为不同来源的异构数据,其传播范围之广、速度之快,对个人和社会造成了严重危害。因此,有效检测和防范虚假新闻至关重要。针对当前虚假新闻检测模型局限于从单一数据来源获取新闻文本及视觉信息,导致新闻报道主观性较强、数据覆盖不全面的问题,提出了一种多源异构数据渐进式融合的虚假新闻检测模型。首先,进行多源异构数据的收集、筛选和清洗,由此构建了一个多源多模态数据集,其中包含关于每个事件的多个不同角度的报道;接着,通过将文本特征提取器和视觉特征提取器获取的特征输入多源融合模块,实现了不同来源特征之间的渐进式融合;同时,引入文本的情感特征和图像的频域特征,以实现多层次的特征提取;最后,采用软注意力机制进行特征集成。实验结果和分析表明,与已有的流行方法相比,所提模型有较好的检测效果,为大数据时代的虚假新闻检测提供了一种有效的解决方案。

**关键词:** 虚假新闻检测;数据扩增;多源异构数据;特征融合;情感特征;频域特征

**中图分类号** TP391

## Multi-source Heterogeneous Data Progressive Fusion for Fake News Detection

YU Yongxin<sup>1,2</sup>, JI Ke<sup>1,2</sup>, GAO Yuan<sup>1,2</sup>, CHEN Zhenxiang<sup>1,2</sup>, MA Kun<sup>1,2</sup> and ZHAO Xiaofan<sup>3,4</sup>

1 School of Information Science and Engineering, University of Jinan, Jinan 250022, China

2 Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Jinan 250022, China

3 School of Information and Cyber Security, People's Public Security University of China, Beijing 102623, China

4 Key Laboratory of Security Prevention Technology and Risk Assessment of the Ministry of Public Security, Beijing 102623, China

**Abstract** Social media platforms are inundated with a vast amount of unverified information, much of which originates from heterogeneous data from multi-source, which spreads so widely and quickly that it poses a significant threat to individuals and society. Therefore, it is crucial to effectively detect and prevent fake news. Targeting the current limitations of fake news detection models, which typically rely on single data sources for news textual and visual information, resulting in strong subjective news reports and incomplete data coverage, a model is proposed for detecting fake news by progressively fusing multi-source heterogeneous data. Firstly, multi-source heterogeneous data collection, screening, and cleaning are conducted to create a multi-source multimodal dataset containing reports about each event from diverse perspectives. Next, by inputting the features obtained from the textual feature extractor and visual feature extractor into the multi-source fusion module, a progressive fusion of features from various sources is achieved. Additionally, sentiment features extracted from text and frequency domain features extracted from images are incorporated into the model to enable multi-level feature extraction. Finally, this paper adopts the soft attention mechanism for feature integration. Experimental results and analysis show that the proposed model has better detection performance compared to existing popular methods, providing an effective solution for fake news detection in the era of big data.

**Keywords** Fake news detection, Data augmentation, Multi-source heterogeneous data, Feature fusion, Sentiment feature, Frequency domain feature

到稿日期:2024-07-01 返修日期:2024-09-06

基金项目:山东省重点研发计划(2021CXGC010103, 2018CXGC0706);山东省自然科学基金(ZR2022LZH016)

This work was supported by the Shandong Provincial Key R&D Program of China(2021CXGC010103, 2018CXGC0706) and Natural Science Foundation of Shandong Province, China(ZR2022LZH016).

通信作者:纪科(ise\_jik@ujn.edu.cn)

## 1 引言

随着互联网和社交媒体的蓬勃发展,人们养成了在社交媒体平台上迅速获取新闻信息并分享个人观点的习惯。随着互联网用户数量的不断增加,信息扩散的速度和范围也显著提升。然而,由于信息的真实性难以即时核实,一些不良分子受利益驱使,故意利用这一点传播虚假新闻,以达到个人或政治目的。通过混淆事实真相,可能使公众产生错误的认知,从而对政治、社会和文化产生负面影响。因此,不能简单地将虚假新闻问题视为信息失真,而应当认识到它可能对整个社会的稳定和合理发展构成严峻威胁。在这一背景下,确保信息的准确性和可信度变得尤为迫切。因此,研究虚假新闻检测技术具有积极而深远的影响。

虚假新闻检测的发展经历了一个持续演进的过程。早期的虚假新闻检测方法依赖于手工提取文本特征,但这种方法需要大量人工参与且无法充分捕捉文本的复杂性。随后,特征工程和监督学习逐渐成为主流,并利用标记数据集训练模型来自动识别虚假新闻。然而其面临着数据不平衡、特征选择难和泛化能力不足等挑战。信息形式的多样化,使得多模态特征的提取和融合成为研究热点。深度学习技术如卷积神经网络(Convolutional Neural Network, CNN)<sup>[1]</sup>、循环神经网络(Recurrent Neural Network, RNN)<sup>[2]</sup>,以及预训练模型如BERT(Bidirectional Encoder Representations from Transformers)<sup>[3]</sup>和GPT(Generative Pre-trained Transformer)<sup>[4]</sup>,在虚假新闻检测中得到广泛应用。研究人员致力于学习更高阶的特征表示,创造了许多功能强大的特征提取器,如Text-CNN(Text Convolutional Neural Network)<sup>[5]</sup>,VGG(Visual Geometry Group)<sup>[6]</sup>和Swin Transform<sup>[7]</sup>等,并探索多模态数据的融合方法,包括简单拼接和注意力机制等方式。此外,一些数据扩增技术,如文本反向翻译<sup>[8]</sup>、生成图像描述<sup>[9]</sup>以及提取图像上文字<sup>[10]</sup>等方法,被用于增加可用数据的数量和提升数据质量,从而提升多模态学习的性能。

尽管虚假新闻检测技术发展迅速,但如何整合多来源和多模态数据的问题仍未得到充分探讨。目前,用于检测虚假新闻的数据集不仅数量有限,而且大多来源于单一渠道。过度依赖单一数据源的文本和视觉信息可能导致新闻报道的主观性较强,其难以提供全面、多角度的信息,从而难以适应虚假新闻多样的形式和来源情形。其次,传统的数据扩增方法受限于单一数据模态,并且其实现及效果依赖于另一模型。此外,在处理多模态信息时,当前方法存在模态信息利用不足以及特征提取不充分等问题。最后,在多模态虚假新闻检测任务中,如何融合不同模态的信息一直是一个经典的问题。

为解决上述问题,本文提出了一种多源异构数据渐进式融合模型(Multi-source Heterogeneous Data Progressive Fusion, MHDPF)。首先,该模型利用网络爬虫技术在3个数据量丰富的大型网站上扩增原始数据,获取每个事件的多个不同角度的报道,从而实现文本-图片的全面数据扩增效果。其次,采用文本匹配领域中的SimBERT<sup>[11]</sup>模型清洗和筛选数据,构建一个涵盖多源多模态信息的数据集。然后,设计渐进式特征融合网络,以整合同一事件的不同报道。通过提取

文本情感特征和图像频域特征,实现多层次特征的提取。最后,通过软注意力机制为每个特征表示分配权重,加权求和生成一个联合表示,用于最终的虚假新闻预测。本文的主要贡献如下:

1)提出了多源异构数据渐进式融合的虚假新闻检测模型。这种综合考虑多个数据源和特征的方法,有效提升了虚假新闻的检测性能。

2)设计渐进式特征融合网络对不同来源的信息进行建模,并逐级整合这些信息,有效避免了简单拼接多源异构数据引入更多噪声的问题。

3)采用软注意力机制进行特征集成。根据各特征的重要性和贡献程度,为其分配不同的权重。

## 2 相关工作

### 2.1 单模态虚假新闻检测

早期的虚假新闻检测研究主要基于单一数据模态(文本或图像)进行分析。对于新闻文本,采用文本特征分析来检测其中的虚假信息,包括挖掘语法结构、情感色彩、词汇选择等方面的特征。Ma等<sup>[12]</sup>基于双向长短期记忆网络(Bidirectional Long Short-Term Memory, Bi-LSTM)来捕获上下文信息随时间变化的隐藏表示。Yu等<sup>[13]</sup>采用卷积神经网络对文本进行特征提取,利用多个不同大小的卷积核来捕捉文本中不同层次的信息。Ma等<sup>[14]</sup>使用多任务学习框架来同时检测谣言和分析立场。Vaibhav等<sup>[15]</sup>将虚假新闻检测任务转化为图分类问题,考虑句子间的相互作用。Jian等<sup>[16]</sup>通过对文本数据进行处理和分析,提高了模型对关键信息的关注度,从而增强了对虚假新闻的辨别能力。Sun等<sup>[17]</sup>通过引入新闻主题信息并结合外部知识和新闻内容,提高了假新闻检测的性能。该方法通过构建异构图获取新闻主题的背景信息,并设计了不同的特征融合策略以适应不同写作风格下的假新闻特征。图像在虚假新闻检测中也是关键特征之一,用于验证图像真实性,并检测编辑痕迹以及与文本的一致性。Qi等<sup>[18]</sup>通过分析图像频率变化和分布来鉴别图片是否为伪造的,并利用注意力机制融合频域和空域特征。

### 2.2 多模态虚假新闻检测

随着互联网和社交媒体的普及,多种媒体类型(包括文本、图像、视频和音频)的使用使得虚假新闻的传播呈现出多模态的特点。单一模态的检测方法已不足以应对,因此研究者们开始探索多模态虚假新闻检测方法。

Wang等<sup>[19]</sup>创建了一个事件对抗性神经网络模型(Event Adversarial Neural Networks, EANN)来学习具有判别性和可迁移性的特征表示,以便在新兴事件中检测虚假新闻。Khattar等<sup>[20]</sup>提出了一个多模态变分自编码器(Multimodal Variational Autoencoder, MVAE),编码器将文本和图像信息编码为共享的潜在向量,解码器则将潜在向量映射回文本和图像空间,以重建原始输入。Zhou等<sup>[21]</sup>提出了一个相似性感知的虚假新闻检测模型(Similarity-Aware Multi-Modal Fake News Detection, SAFE),用于解决图文不符的问题。Wu等<sup>[22]</sup>提出了一种多模态融合模型(Multimodal Co-Attention Networks, MCAN),通过堆叠多个协同注意力层来融合

多模态特征,模拟人类阅读新闻时文本与图像的相互理解,以学习它们间的相互依赖关系。Qian 等<sup>[23]</sup>提出了一种层级多模态上下文注意网络(Hierarchical Multi-modal Contextual Attention Network, HMCAN),该网络先通过层级注意力机制进行模态内自注意力操作,再通过跨模态注意力机制捕捉模态间关联。Guo 等<sup>[24]</sup>提出了一种双分支多模态虚假新闻检测模型(Multimodal Bilinear Pooling and Attention Mechanism, MBPAM)。该模型通过双分支结构学习文本和图像的隐藏信息,并利用多模态双线性池化和自注意力机制来融合特征并捕捉模态间的关系。Jing 等<sup>[25]</sup>设计了一个多级视觉和文本融合模型(Multimodal Progressive Fusion Network, MPFN),细粒度融合每种模态的特征表示。鉴于模态间单向关注的不足,Guo<sup>[26]</sup>提出了基于互注意力机制的模型(Mutual Attention Neural Network, MANN),通过双向流动操作在文本和视觉特征间进行融合。Liu 等<sup>[27]</sup>提出了一个基于分层双线性融合和多模态一致性的虚假新闻检测模型(Hierarchical Bilinear Fusion and Multimodal Consistency for Fake News Detection, BC-FND)。该模型采用对比学习方法对齐文本与图像特征,通过一致性损失函数学习真实与虚假新闻间的差异,并通过设计的分层网络实现文本与图像特征的融合。Peng 等<sup>[28]</sup>提出了一种新方法,即上下文语义表示学习(Contextual Semantic Representation Learning for Multimodal Fake News Detection, CSFND),通过上下文语义表示学习和

无监督上下文特征聚合,有效区分了不同语义上下文中的虚假新闻和真实新闻。Wang 等<sup>[29]</sup>提出了一种基于三级特征匹配距离方法的多模态融合模型(Three-Level Feature Matching Distance for Fake News Detection, TLFND)。该模型结合了新闻文章中两个层次的文本信息(标题和正文)和图像数据(多图像融合),利用切比雪夫距离度量方法计算特征之间的匹配距离,并采用自适应进化算法优化损失函数。

针对可用数据量不足的难题,现有技术尝试采用多种方法来增加可用数据的数量并提升其质量。例如,采取反向翻译的方式获得逻辑结构存在差异的前提下仍然能传达相同语义的文本数据;或提取图像中的文字,还可以利用图像生成文本模型生成图像描述。但这些方法局限于单一数据模态的扩增,无法实现全面的数据扩增,并且依赖于另一模型的实现及效果。

### 3 多源异构数据渐进式融合模型

虚假新闻检测旨在将社交媒体上的新闻分类为真实或虚假。本文将一个新闻实例  $I = \{\text{Ori}_{\text{text}}, \text{Ori}_{\text{image}}, \text{S}_{\text{text}}, \text{S}_{\text{image}}\}$  定义为由 4 种不同内容组成的元组:原始新闻文本  $\text{Ori}_{\text{text}}$ 、原始新闻图片  $\text{Ori}_{\text{image}}$ 、多源新闻文本  $\text{S}_{\text{text}}$  以及多源新闻图片  $\text{S}_{\text{image}}$ 。模型总体结构如图 1 所示,主要由多源异构数据扩增、渐进式特征融合、多层次特征提取以及特征集成分类器组成。通过这些组件的协同作用,用于检测以多源多模态内容为输入的虚假新闻。

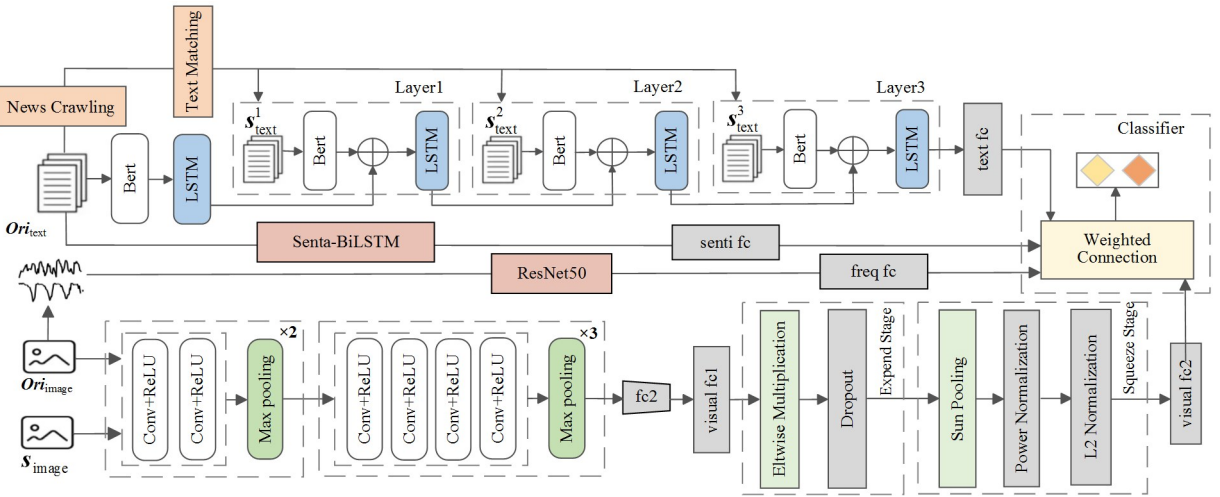


图 1 多源异构数据渐进式融合模型

Fig. 1 Multi-source heterogeneous data progressive fusion model

#### 3.1 多源异构数据扩增

在真实的新闻报道中,不同来源的报道对同一事件的描述和细节通常是相似的,这反映了报道的一致性。相反,虚假新闻往往包含着矛盾和错误的信息。通过比较不同来源对同一事件的描述,可以检测到其中的不一致之处。多源异构数据扩增指以每一个新闻样本为中心,采集描述同一事件但不同角度的报道。

具体来说,使用关键词提取技术 TextRank4ZH<sup>[30]</sup>从每条新闻中提取至少 5 个长度不小于 2 的关键词;然后利用网络爬虫技术在 3 个数据量丰富的大型网站中采集与这些关键词相关的报道,以实现文本-图片多模态数据的扩增。针对

无法采集到相关报道的情况,采用零填充策略,从而构建了初步的多源多模态数据集。

由于新闻的宏观大环境下可供选择的多元数据数量非常多,数据的质量层次也不相同,因此本节使用 SimBERT 模型对数据进行清洗和筛选。该模型基于 BERT 结构,在文本匹配领域表现出色,用于捕捉文本之间的语义关系。模型主要分为 3 个部分:词嵌入、编码器和预测。输入为原始新闻文本以及 3 个多源新闻文本,模型旨在计算它们之间的相似度来获取预测结果  $p(p \in [0, 1])$ 。

具体来说,首先对输入的原始新闻文本和多源新闻文本进行分词,分别表示为  $\text{Ori}_{\text{text}} = [\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_m]$  和  $\text{S}_{\text{text}}^Q = [\mathbf{s}_1,$

$s_2, \dots, s_n$ ], 其中  $n$  和  $m$  代表分词后句子中的单词数,  $j(j \in \{1, 2, 3\})$  代表新闻来源的数量。接着, 在每个句子的开头添加 [CLS] 标记, 生成  $token$  序列  $\mathbf{T}_{Ori} = [o_{[CLS]}, o_1, \dots, o_m]$  和  $\mathbf{T}_S^{(j)} = [s_{[CLS]}, s_1, \dots, s_n]$ 。然后, 将  $\mathbf{T}_{Ori}$  和  $\mathbf{T}_S^{(j)}$  输入 BERT 模型中。BERT 模型由 12 个编码器组成, 每个编码器层的隐藏层大小为 768。在每个编码器层中, 通过自注意力机制来捕捉单词之间的依赖关系, 并通过前馈网络将上一个编码器结果传递给下一个编码器, 从而获得单词嵌入向量, 表示为  $e_{Ori}$  和  $e_S^{(j)}$ 。接下来, 在自注意力网络中计算每个单词的权重, 对  $e_{Ori}$  和  $e_S^{(j)}$  进行加权, 从而突出更重要的单词对句子的影响。随后, 通过池化层对所有单词的加权嵌入向量进行处理, 该步骤能够将单词级别的表示转换为句子级别的表示  $\mathbf{E}_{Ori}$  和  $\mathbf{E}_S^{(j)}$ 。

通过计算原始新闻文本与每个多源新闻文本之间的相似度, 可以衡量新闻之间的相关性。图 2 给出了数据集中原始新闻文本与每个多源新闻文本之间的相似度分布情况。X 轴表示相似度区间, Y 轴表示数据量。

通过将阈值  $\theta$  设置为 0.5, 本节对多源数据应用了一个过滤器。仅保留相似性得分高于该阈值的预测, 以确保数据的质量和相关性, 由此构建最终的多源多模态数据集。过滤函数可以表示为:

$$Filter(data) = \{d \mid similarity(\mathbf{E}_{Ori}, \mathbf{E}_S^{(j)}) > \theta\} \quad (1)$$

其中,  $\forall j \in \{1, 2, 3\}, d \in \{data\}$  表示数据集中的一个样本数据。该过程确保了构建的数据集具备多模态性和语义一致性。

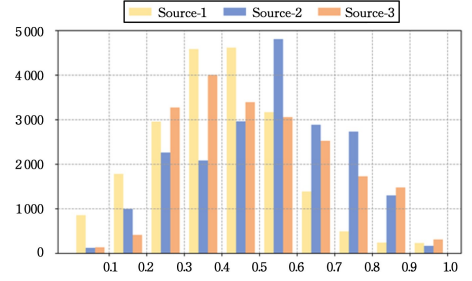


图 2 相似度结果分布情况

Fig. 2 Distribution of similarity results

### 3.2 渐进式特征融合

本节分别采用不同方法对文本和图像数据进行向量化表示、特征提取以及特征融合。

#### 3.2.1 新闻文本特征融合

为实现多源新闻文本的渐进式融合, 使用多层网络与传统的 LSTM 模型相结合, 融合方法如图 3 所示。

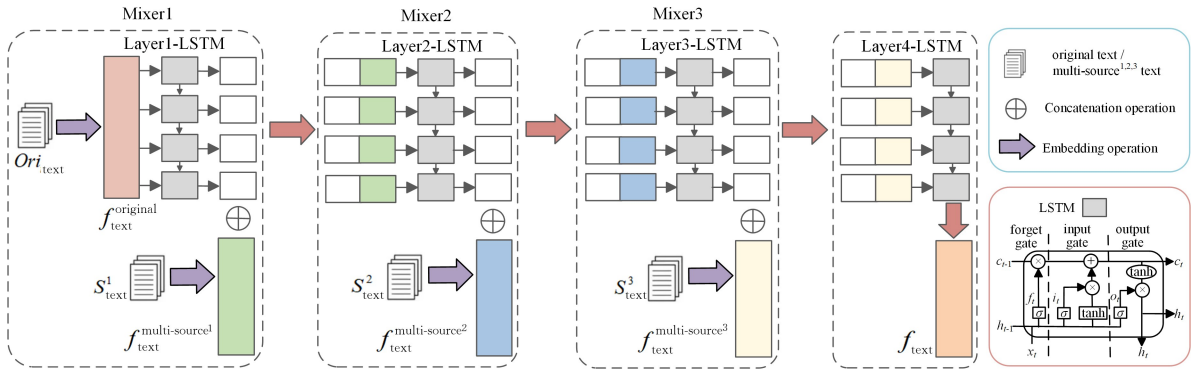


图 3 渐进式融合方法

Fig. 3 Progressive fusion method

文本特征的提取质量将直接影响虚假新闻检测的准确率。相比只能提取低阶语义特征的传统模型, 预训练模型 BERT 在大规模数据上进行了训练, 内部包含了大量的语法知识和句子间的隐含信息, 极大地增强了模型的语义表达能力。通过将  $\mathbf{Ori}_{text}$  和  $\mathbf{S}_{text}^{(j)}$  输入到预训练模型 BERT 中, 单词被映射到高维向量空间中的表示:  $f_{text}^{original}$  和  $f_{text}^{multi-source^j}$ 。接着, 通过搭建 3 个混合器逐步融合每一个来源的新闻文本向量。整个融合过程可以划分为 4 个层次, 其中每一层都将当前输入的文本向量与前一层的隐藏状态结合。

具体来说, 首先将原始新闻文本向量  $f_{text}^{original}$  输入 LSTM 层中, 以获取其隐藏状态  $f_t^1$ 。随后, 在第一个混合器中, 将第一个多源新闻文本向量  $f_{text}^{multi-source^1}$  与上一阶段得到的隐藏状态  $f_t^1$  相拼接, 从而将不同句子中的上下文信息整合在一起, 这样 LSTM 网络可以在一个连续的序列中学习文本的整体结构和语义。接着, 将拼接后的特征输入 Layer1\_LSTM, 得到新的隐藏状态  $f_t^2$ 。这一过程一直持续到第三个混合器, 用于融合另外两个来源的新闻文本向量。在 Layer3\_LSTM

之后, 引入一个全连接层 (text fc), 用来调整文本特征的维度, 生成文本语义特征联合表示  $f_{text}$ 。整个过程可以表示为:

$$f_t^1 = LSTM(f_{text}^{original}) \quad (2)$$

$$f_t^2 = Layer1\_LSTM(\text{concat}(f_t^1, f_{text}^{multi-source^1})) \quad (3)$$

$$f_t^3 = Layer2\_LSTM(\text{concat}(f_t^2, f_{text}^{multi-source^2})) \quad (4)$$

$$f_t^4 = Layer3\_LSTM(\text{concat}(f_t^3, f_{text}^{multi-source^3})) \quad (5)$$

$$f_{text} = \varphi(\mathbf{W}_{tf} \cdot f_t^4) \quad (6)$$

其中,  $f_t^*$  表示每一层 LSTM 处理后的隐藏状态,  $\mathbf{W}_{tf}$  为 text fc 的权重矩阵,  $\varphi(\cdot)$  为双曲正切函数。

#### 3.2.2 新闻视觉特征融合

在社交媒体中, 图像作为一种直观生动的视觉呈现方式, 容易迅速吸引用户关注, 但也易被用于传播虚假信息以追求热度。因此, 在当前的虚假新闻检测中引入对图像真实性的判断同样至关重要。

该模块接收  $\mathbf{Ori}_{image}$  和  $\mathbf{S}_{image}$  作为输入, 利用在 ImageNet 数据集上预训练的 VGG19 模型来提取图像的语义特征。VGG19 通过一系列卷积层和池化层逐步提取图像的层次化

特征。该网络由 19 个层组成,其中包括 16 个卷积层和 3 个全连接层。通过增加网络的深度,VGG19 模型提升了对图像语义信息的学习能力。输出层选择了 VGG19 网络的 fc2 层,在该层之后,引入了一个全连接层(visual fc1),用来调整视觉特征的维度。视觉特征可以表示为:

$$f_{ov} = \varphi(W_{vf1} \cdot R_{VGG}^{\text{original}}) \quad (7)$$

$$f_{sv} = \varphi(W_{vf1} \cdot R_{VGG}^{\text{multi-source}}) \quad (8)$$

其中, $R_{VGG}^{\text{original}}$  和  $R_{VGG}^{\text{multi-source}}$  为 VGG19 网络提取的原始新闻视觉特征和多源新闻视觉特征, $W_{vf1}$  为 visual fc1 的权重矩阵。

为了更有效地融合原始新闻视觉特征和多源新闻视觉特征,本节采用矩阵分解双线性池化方法<sup>[31]</sup>。该方法将两路特征提取器的输出特征  $f_{ov}$  和  $f_{sv}$  进行融合,能够更好地捕捉原始新闻视觉特征和多源新闻视觉特征之间的相关性,从而增强特征的表征能力。其计算式如下:

$$v_i = f_{ov}^T W_i f_{sv} \quad (9)$$

其中, $W_i \in \mathbb{R}^{m \times n}$  为投影矩阵, $v_i \in \mathbb{R}$  为新闻视觉特征融合模块的输出。为了获取融合后的  $o$  维视觉特征向量  $v$ ,需要学习  $W = [W_i, \dots, W_o] \in \mathbb{R}^{m \times n \times o}$ 。双线性池化虽然有效地捕捉了特征维度之间的成对交互,但同时也引入了大量的参数,这会导致较高的计算成本和过拟合的风险。因此,为了减少参数量, $W_i$  被分解成两个低秩矩阵:

$$\begin{aligned} v_i &= f_{ov}^T U_i V_i^T f_{sv} \\ &= \sum_{d=1}^k f_{ov}^T U_d V_d^T f_{sv} \\ &= \mathbf{1}^T (U_i^T f_{ov} \circ V_i^T f_{sv}) \end{aligned} \quad (10)$$

其中, $k$  为分解矩阵  $U_i = [u_1, \dots, u_k] \in \mathbb{R}^{m \times k}$  和  $V_i = [v_1, \dots, v_k] \in \mathbb{R}^{n \times k}$  的潜在维数, $\circ$  表示两个向量的元素相乘, $\mathbf{1} \in \mathbb{R}_k$  为一个全 1 向量。为了得到输出特征  $v \in \mathbb{R}^o$ ,需要学习的权重是两个三阶张量  $U = [U_1, \dots, U_o] \in \mathbb{R}^{m \times k \times o}$  和  $V = [V_1, \dots, V_o] \in \mathbb{R}^{n \times k \times o}$ 。接下来使用 reshape 操作将  $U$  和  $V$  进一步转换为二维矩阵  $\tilde{U} \in \mathbb{R}^{m \times ko}$  和  $\tilde{V} \in \mathbb{R}^{n \times ko}$ 。因此,式(10)可改写为:

$$v = \text{SumPooling}(\tilde{U}^T f_{ov} \circ \tilde{V}^T f_{sv}, k) \quad (11)$$

其中, $\text{SumPooling}(x, k)$  表示使用大小为  $k$  的一维非重叠窗口在  $x$  上执行求和池化。在该模块之后,引入一个全连接层(visual fc2)用来调整  $v$  的维度,生成视觉语义特征联合表示  $f_{\text{image}}$ :

$$f_{\text{image}} = \varphi(W_{vf2} \cdot v) \quad (12)$$

其中, $W_{vf2}$  为 visual fc2 的权重矩阵。

### 3.3 多层次特征提取

本文的研究重点仍然放在原始数据上,这是本研究的基础和最初的信息来源。为了从多个层面挖掘特征信息,本节选择在原始数据上提取情感特征和频域特征。这类特征能够揭示虚假新闻中可能存在的情感操纵和图像篡改现象。

#### 3.3.1 情感特征提取

新闻文本通常蕴含着丰富的情感信息,而情感分析技术可以识别新闻文本中的情感倾向,帮助区分虚假新闻和真实新闻。虚假新闻通常使用过于情感化或夸张的语言,旨在吸引读者并迅速传播;而真实新闻则更注重客观呈现事实。本节使用 Senta-BiLSTM(Senta-Bidirectional Long Short-Term

Memory)模型对原始新闻文本进行情感分类预测,得到文本情感特征  $s$ :

$$s = \text{Senta-BiLSTM}(\text{Ori}_{\text{text}}) \quad (13)$$

随后,引入一个全连接层(senti fc),用于调整情感特征  $s$  的维度,以生成最终的文本情感特征  $f_{\text{senti}}$ :

$$f_{\text{senti}} = \varphi(W_{sf} \cdot s) \quad (14)$$

其中, $W_{sf}$  为 senti fc 的权重矩阵。

#### 3.3.2 频域特征提取

虚假新闻往往通过篡改或者合成图像来支持虚假的陈述。大量研究表明,被篡改过的图像通常会在频域空间中留下不同于原始图像的痕迹。因此,通过分析这些频域特征,可以有效检测篡改图像,揭露虚假信息。本节应用二维离散余弦变换(2D-DCT)将图像从原始的空间域表示转化为频域表示,并使用 ResNet50<sup>[32]</sup> 模型作为特征提取器。这一模型因引入了残差学习机制而著名,能够更有效地捕捉图像的抽象语义信息。通过 2D-DCT 将图像进行转换,并与 ResNet50 模型的多尺度特征提取能力相结合,增强了特征的表达能力。随后引入一个全连接层(freq fc),以调整频域特征的维度,从而得到图像频域特征  $f_{\text{freq}}$ :

$$f_{\text{freq}} = \varphi(W_{ff} \cdot R_{\text{ResNet50}}) \quad (15)$$

其中, $R_{\text{ResNet50}}$  为 ResNet50 提取的图像频域特征, $W_{ff}$  为 freq fc 的权重矩阵。

### 3.4 特征集成分类器

文本特征和视觉特征在虚假新闻检测中都发挥着重要作用,但并不是每个表示对该任务都有着同样的贡献。本节引入软注意力机制来区分不同表示在虚假新闻检测中的贡献程度。该机制通过计算各个表示的得分为其分配不同的权重占比。最终,通过加权求和生成一个联合表示  $r$ 。其计算过程如下:

$$\text{Attn}(f_i) = v^T \varphi(W f_i + b) \quad (16)$$

$$a_i = \frac{\exp(\text{Attn}(f_i))}{\sum_{j=1}^N \exp(\text{Attn}(f_j))} \quad (17)$$

$$r = \sum_{i=1}^N a_i f_i \quad (18)$$

其中, $N$  表示特征的总数, $f_i$  表示第  $i$  个特征, $a_i$  表示第  $i$  个特征的注意力得分, $v \in \mathbb{R}^d$  是可学习的向量, $W \in \mathbb{R}^{d \times d}$  表示参数矩阵, $b$  则是偏置项,联合表示  $r$  是通过将注意力权重  $a_i$  用于对每个特征  $f_i$  进行加权得到的。

随后,采用一个带有 softmax 激活函数的全连接层,用于输出联合表示  $r$  代表新闻是真实或虚假的概率。具体形式如下:

$$\hat{y}_k = \text{softmax}(W \times r + b) \quad (19)$$

其中, $W$  表示全连接层的权重矩阵, $b$  是偏置项。

本节使用二元交叉熵(Binary Cross Entropy, BCE)作为损失函数,通过最小化预测标签与真实标签之间的差异来优化二分类模型。其计算式如下:

$$L = - \sum_{k=1}^K [y_k \log(\hat{y}_k) + (1 - y_k) \log(1 - \hat{y}_k)] \quad (20)$$

其中, $K$  表示新闻样本的总数, $y_k$  代表样本的真实标签, $\hat{y}_k$  代表模型的预测标签。

## 4 实验过程及结果分析

### 4.1 数据集

本文所用的数据集来自人工智能竞赛平台 Biendata,由智源研究院与中科院计算所联合发布。原始数据中的新闻收集截止于 2019 年。为确保数据的实时性和全面性,首先选取了原始数据中的一部分无重复数据,并通过网络爬虫技术不断添加最新的新闻内容。随后进行数据扩增,获取关于每一事件的多角度报道。数据集按 8:2 的比例划分为训练集和测试集。表 1 列出了数据集的统计信息。

表 1 数据集统计信息  
Table 1 Dataset statistics

| 划分   | 训练集    | 测试集   |
|------|--------|-------|
| 虚假新闻 | 8 039  | 1 961 |
| 真实新闻 | 8 230  | 2 106 |
| 多源新闻 | 19 527 | 5 509 |

### 4.2 实验设置

本文实验环境如表 2 所列。

表 2 实验环境  
Table 2 Experimental environment

| 名称      | 版本                                                    |
|---------|-------------------------------------------------------|
| 处理器     | 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz |
| 显卡      | RTX 3090(24 GB)                                       |
| 开发平台    | Anaconda                                              |
| 深度学习框架  | Tensorflow, Keras                                     |
| 开发语言及平台 | Python, PyCharm                                       |

### 4.3 参数设置

本文以准确率(Accuracy)、精准率(Precision)、召回率(Recall)以及 F1 值(F-measure)作为评价指标来评估模型性能。

在模型训练过程中,将 epoch 设置为 10, batch\_size 设置为 128, 学习率设置为 0.0001, 同时采用 dropout 等正则化技术防止过拟合。对于文本数据,利用基于 Keras 的 keras-bert 库,并加载 bert-base-chinese 预训练模型,以生成 768×1 维的文本向量,并通过 LSTM 模型来提取文本特征,其中隐藏层单元大小设定为 64。图像大小统一调整为 224×224×3,经 VGG19 模型处理,输出维度为 4096×1 的视觉特征,并冻结 VGG19 参数以避免过拟合。在每个特征向量后添加了一个

全连接层,将特征维度调整为 32,并使用 softmax 激活函数进行分类。模型参数优化采用了 Adam 优化器。

### 4.4 基线

本文选择两类基线模型与所提出的 MHDPF 模型进行比较,分别是单模态模型以及多模态模型。由于基线模型不涉及对多源数据的处理和融合,因此实验在原始数据上进行。

#### 4.4.1 单模态模型

1)Textual-Word2Vec:该文本模型仅使用原始新闻文本进行分类。使用 Word2Vec<sup>[39]</sup>将新闻文本转换为词向量,然后使用 LSTM 提取文本语义信息。

2)Textual-BERT:不同于另一文本模型 Textual-Word2Vec,该文本模型采用预训练模型 BERT 提取词向量。

3)Visual:视觉模型仅使用原始新闻图像进行分类。采用预训练的 VGG19 网络提取图像的视觉特征。

#### 4.4.2 多模态模型

1)Simple Fusion:通过全连接层调整两种单模态特征的维度并进行拼接,然后输入分类器中得出预测结果。

2)Original and Multi-source News(O-M):该模型在 Simple Fusion 的基础上继续添加多源文本特征以及多源视觉特征以整合多源、多模态数据。随后,经过全连接层将这些特征统一转换至相同的维度并进行拼接,最后输入假新闻分类器中进行结果预测。

3)att-RNN<sup>[34]</sup>:该模型通过引入注意力机制来融合文本、图像以及社交上下文特征。在本实验中,移除了处理社交上下文信息的部分。

4)EANN:该模型由多模态特征提取器、假新闻检测器和事件鉴别器 3 个关键组件构成。在本实验中采用了不包含事件鉴别器的 EANN 变体。

5)MVAE:该模型由 3 个关键组件构成。编码器负责将文本和图像转换为向量表示,解码器用于重构原始图像和文本内容,虚假新闻检测器利用学习到的共享表示进行检测。

6)MANN:该模型使用了一种基于互注意力机制的融合方法,通过计算文本和图像之间的相互注意力权重,将两种模态的特征进行融合。

### 4.5 实验结果及分析

表 3 列出了基线模型和本文模型在数据集上的性能结果。表中详细记录了假新闻检测器的准确率,以及在虚假新闻和真实新闻上的精确度、召回率和 F1 分数。

表 3 MHDPF 模型与基线的对比结果

Table 3 Comparison results between MHDPF and baselines

| Method           | Accuracy     | Fake News    |              |              | Real News    |              |              |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                  |              | Precision    | Recall       | F-Measure    | Precision    | Recall       | F-Measure    |
| Textual-Word2Vec | 0.864        | 0.881        | 0.831        | 0.855        | 0.850        | 0.896        | 0.872        |
| Textual-BERT     | 0.938        | 0.946        | 0.924        | 0.934        | 0.930        | 0.951        | 0.940        |
| Visual           | 0.634        | 0.593        | 0.772        | 0.670        | 0.704        | 0.506        | 0.589        |
| Simple Fusion    | 0.883        | 0.915        | 0.834        | 0.873        | 0.857        | 0.928        | 0.891        |
| O-M              | 0.896        | 0.889        | 0.896        | 0.892        | 0.902        | 0.896        | 0.899        |
| att-RNN          | 0.903        | 0.933        | 0.861        | 0.896        | 0.879        | 0.943        | 0.910        |
| EANN             | 0.911        | 0.909        | 0.907        | 0.908        | 0.913        | 0.916        | 0.915        |
| MVAE             | 0.895        | 0.919        | 0.857        | 0.887        | 0.875        | 0.929        | 0.901        |
| MANN             | 0.917        | 0.928        | 0.898        | 0.913        | 0.908        | 0.935        | 0.921        |
| MHDPF            | <b>0.969</b> | <b>0.973</b> | <b>0.962</b> | <b>0.967</b> | <b>0.965</b> | <b>0.975</b> | <b>0.970</b> |

通过实验对比可以看出,在单模态模型方面,Visual 模型的准确率低于 Textual 模型(Textual-Word2Vec 和 Textual-BERT)的准确率。这表明在虚假新闻检测任务中,相较于图像,文本所提供的语义信息更为关键,在新闻表达中占据主导地位。另外,基于 BERT 的文本模型效果要优于基于 Word2Vec 的文本模型,从而证明了在虚假新闻检测任务中采用更为先进的特征提取器对于增强模型性能的重要性。使用 BERT 模型进行词嵌入,这一方法相比传统的词嵌入技术可以带来更好的效果提升。因为 BERT 模型能更好地捕捉上下文信息,这在理解虚假新闻中的语义信息时至关重要。

实验结果表明,引入经过仔细数据筛选和清洗的同一事件多角度的报道有助于提升模型的训练效果,准确率从 88.3%提升到了 89.6%。这说明增加数据量和提高数据的多样性可以有效提升模型的性能。数据扩增策略有助于提升模型的泛化能力和鲁棒性。而数据清洗和筛选有助于剔除无用信息和可能带来误导的样本,提高数据的质量。

MHDPF 模型采用多模态特征加权拼接策略,能够调整不同模态信息在融合中的权重,更好地控制各个特征对最终分类结果的影响程度。相较于 att-RNN 和 EANN 模型的简单拼接方式,MHDPF 模型能更有效地利用多模态信息,性能更优。MVAE 模型的准确率比 O-M 模型低 0.1%,原因在于数据集中含有图像的新闻文本只占总量的 55.2%,导致 MVAE 无法充分学习到图像和文本的共享表示。MANN 模型通过双向流动操作实现文本和视觉特征的互注意力机制融合,但在图像样本少的数据集上表现不佳,数据不均衡导致难以有效学习图像和文本特征间

的关联性,从而影响特征融合的效果。

综上所述,MHDPF 模型通过多源异构数据扩增,实现了特征的渐进式融合和多层次的特征提取,同时考虑了多模态特征对模型的不同贡献,相比基线方法表现出更优的性能。这些结果进一步证实了本文提出的方法在提高分类性能和可靠性方面所取得的成果。

#### 4.6 消融实验及结果分析

为验证本文方法的有效性,构造了若干变体模型进行分析,每个变体通过移除 MHDPF 的某个组件来评估其作用。

1)MHDPF(-BERT):移除 MHDPF 模型中的 BERT 方法,改用 Word2Vec 方法进行词嵌入。相比 BERT 获取更高维度的语义表示,轻量级模型 Word2Vec 计算效率高且更简单。

2)MHDPF(-1):在 MHDPF 模型的渐进式融合基础上,对融合过程进行拆分。通过移除一个新闻文本特征融合模块 Layer3 和视觉特征融合模块,验证多个来源数据的引入对模型的有效性。

3)MHDPF(-2):在 MHDPF(-1)的基础上,再次移除一个多源数据融合模块 Layer2。

4)MHDPF(-3):在 MHDPF(-2)的基础上,再次移除一个多源数据融合模块 Layer1。

5)MHDPF(-S):移除 MHDPF 模型中的情感特征提取器,仅融合文本语义特征、图像的频域特征以及空间域特征。

6)MHDPF(-F):移除 MHDPF 模型中的频域特征提取器,仅融合文本语义特征以及情感特征、图像的空间域特征。

实验结果如表 4 所列。

表 4 消融实验结果的对比

Table 4 Comparison of ablation experiment results

| Method       | Accuracy     | Fake News    |              |              | Real News    |              |              |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|              |              | Precision    | Recall       | F-Measure    | Precision    | Recall       | F-Measure    |
| MHDPF        | <b>0.969</b> | <b>0.973</b> | 0.962        | <b>0.967</b> | 0.965        | <b>0.975</b> | <b>0.970</b> |
| MHDPF(-BERT) | 0.923        | 0.927        | 0.913        | 0.920        | 0.920        | 0.933        | 0.926        |
| MHDPF(-1)    | 0.961        | 0.949        | 0.972        | 0.960        | 0.974        | 0.951        | 0.962        |
| MHDPF(-2)    | 0.959        | 0.953        | 0.963        | 0.958        | 0.965        | 0.955        | 0.960        |
| MHDPF(-3)    | 0.956        | 0.948        | 0.963        | 0.955        | 0.965        | 0.951        | 0.958        |
| MHDPF(-S)    | 0.958        | 0.960        | 0.952        | 0.956        | 0.956        | 0.963        | 0.960        |
| MHDPF(-F)    | 0.960        | 0.943        | <b>0.975</b> | 0.959        | <b>0.976</b> | 0.945        | 0.960        |

MHDPF 模型相比使用 Word2Vec 方法的 MHDPF(-BERT)模型,准确率高出 4.6%。表明使用 BERT 替代传统的文本表示方法 Word2Vec,可以获得更为准确和有用的特征表示,从而提升模型的性能。当使用 Word2Vec 替代 BERT 作为词向量时,MHDPF 的效果仍然好于其他的多模态基线模型。此外,通过比较 MHDPF(-1),MHDPF(-2)和 MHDPF(-3)模型得出,通过逐步移除融合阶段的各个模块,MHDPF 模型的性能可以得到不同程度的提升。因此,保留多源数据融合模块有助于模型更好地利用不同信息源的特征。相反,当融合模块被移除后,模型可能无法充分利用多源数据的信息,从而导致性能下降。

在移除文本情感模块和图像频域模块后,检测准确率分别降低了 1.1%和 0.9%。这表明多层次特征提取模块的

加入能够有效提升模型性能。

本文研究了多层 LSTM 模型中隐藏单元数量对 MHDPF 模型性能的影响。通过设置隐藏单元数为 16,32,64,128,可以观察到改变 LSTM 隐藏单元数对模型性能的影响趋势。如图 4 所示,当隐藏层单元数设置为 64 时,MHDPF 模型展现出最优的性能。

为验证特征加权拼接方法的有效性,将其与逐元素相加(Addition)和张量拼接(Concat)两种替代方式进行比较。从图 5 中可以看出,使用特征加权拼接方法的模型相比逐元素相加方法,检测准确率提高了 1.6%;相比张量拼接方法,检测准确率提高了 0.7%。这是因为简单拼接方法不能有效融合多模态特征,反而可能引入冗余信息或噪声,导致性能下降。

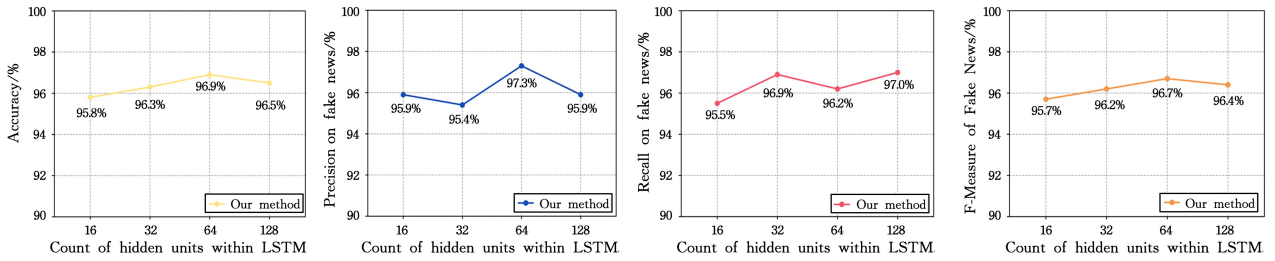


图4 LSTM隐藏单元数对模型性能的影响分析

Fig. 4 Analysis of the effect of LSTM hidden units on model performance

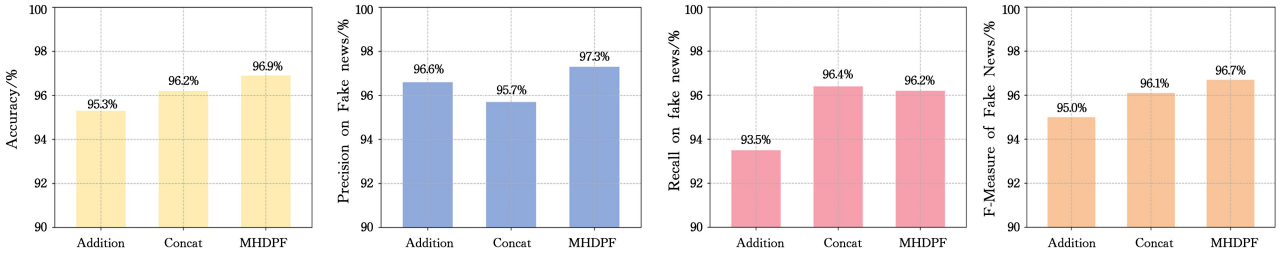


图5 不同融合方法的性能对比

Fig. 5 Performance comparison of different fusion methods

**结束语** 当前虚假新闻检测技术难以适应信息时代的快速发展。应对这些挑战的关键在于创造性地扩展可用数据集,并整合不同来源的异构数据,以更好地应对虚假信息的多样性和复杂性。本文致力于解决当前虚假新闻检测领域中数据来源单一以及数据扩增受限于单一模态的问题。针对这一挑战,本文提出了多源异构数据渐进式融合的虚假新闻检测模型来检验新闻的真实性。通过对同一事件进行多源异构数据扩增,获得多角度的报道,并由此构建多源多模态数据集。引入渐进式融合策略构建文本和图像融合模块,实现了不同来源特征的逐级融合。同时,多层次地考虑了文本的情感信息和图像的频域信息,并通过软注意力机制方法实现特征集成。通过在数据集上进行对比实验和消融实验,验证了本文模型的可行性和有效性。

另外,本文目前的工作主要集中在实现文本与图像的一对一关系。然而,在实际的新闻报道中,通常包含多张携带关键线索的图片,本文方法并不能有效整合这些丰富的视觉资源。此外,本文也尚未对图文之间的语义关联进行深入研究。在未来的工作中,我们可以将研究重点扩展到文本与图像的一对多关系,即一个文本对应多个相关图像的情况。这将有助于我们更全面地理解和分析新闻报道中的视觉信息。同时,还可以探索如何将图文关联特征进行增强,以提高文本与图像之间的关联性。

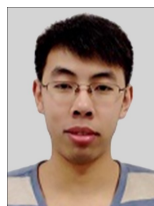
## 参考文献

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. *Communication of the ACM*, 2017, 60(6): 84-90.
- [2] ELMAN J L. Finding structure in time[J]. *Cognitive Science*, 1990, 14(2): 179-211.
- [3] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [4] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[J/OL]. [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf).
- [5] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv:1408.5882, 2014.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556, 2014.
- [7] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 10012-10022.
- [8] HUA J, CUI X, LI X, et al. Multimodal fake news detection through data augmentation-based contrastive learning[J]. *Applied Soft Computing*, 2023, 136: 110125.
- [9] MAHAJAN S, ROTH S. Diverse image captioning with context-object split latent spaces[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 3613-3624.
- [10] LI Y, JI K, MA K, et al. Fake News Detection Based on the Correlation Extension of Multimodal Information[C]// *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*. Cham: Springer Nature Switzerland, 2022: 443-450.
- [11] SU J. Simbert: Integrating retrieval and generation into bert[J/OL]. Tech. Rep. <https://github.com/Zhuiyi Technology/simbert>.
- [12] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]// *Proceedings of IJCAI*. 2016: 3818-3824.
- [13] YU F, LIU Q, WU S, et al. A Convolutional Approach for Misinformation Identification[C]// *IJCAI*. 2017: 3901-3907.

- [14] MA J,GAO W,WONG K F. Detect rumor and stance jointly by neural multi-task learning[C]// Companion Proceedings of the Web Conference 2018. 2018;585-593.
- [15] VAIBHAV V,ANNASAMMY R M,HOVY E. Do sentence interactions matter? leveraging sentence level representations for fake news classification[J]. arXiv;1910. 12203.2019.
- [16] JIAN W,LI J P,AKBAR A, et al. SA-Bi-LSTM; Self Attention With Bi-Directional LSTM based Intelligent Model for Accurate Fake News Detection to ensured information integrity on social media platforms[J]. IEEE Access,2024,12:48436-48452.
- [17] SUN L,WANG H. Topic-Aware Fake News Detection Based on Heterogeneous Graph [J]. IEEE Access, 2023, 11: 103743-103752.
- [18] 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). IEEE,2019;518-527.
- [19] 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.
- [20] 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.
- [21] ZHOU X,WU J,ZAFARANI R. Safe; Similarity-aware multi-modal fake news detection. arxiv[J]. arXiv;2003. 04981,2020.
- [22] WU Y,ZHAN P,ZHANG Y, et al. Multimodal fusion with co-attention networks for fake news detection[C]// Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021;2560-2569.
- [23] QIAN S,WANG J,HU J, et al. Hierarchical multi-modal contextual attention network for fake news detection[C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021; 153-162.
- [24] GUO Y,GE H,LI J. A two-branch multimodal fake news detection model based on multimodal bilinear pooling and attention mechanism [J]. Frontiers in Computer Science, 2023, 5: 1159063.
- [25] JING J, WU H, SUN J, et al. Multimodal fake news detection via progressive fusion networks[J]. Information Processing & Management,2023,60(1):103120.
- [26] GUO Y. A mutual attention based multimodal fusion for fake news detection on social network[J]. Applied Intelligence,2023, 53(12);15311-15320.
- [27] LIU Y,BING W,REN S, et al. BC-FND: An Approach Based on Hierarchical Bilinear Fusion and Multimodal Consistency for Fake News Detection[J]. IEEE Access,2024,12:62738-62749.
- [28] PENG L, JIAN S, KAN Z, et al. Not all fake news is semantically similar: Contextual semantic representation learning for multimodal fake news detection[J]. Information Processing & Management,2024,61(1):103564.
- [29] WANG J,ZHENG J,YAO S, et al. TLFND: A Multimodal Fusion Model Based on Three-Level Feature Matching Distance for Fake News Detection[J]. Entropy,2023,25(11):1533.
- [30] MIHALCEA R, TARAU P. Textrank: Bringing order into text [C]// Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. 2004;404-411.
- [31] YU Z, YU J, FAN J, et al. Multi-modal factorized bilinear pooling with co-attention learning for visual question answering [C]// Proceedings of the IEEE International Conference on Computer Vision. 2017;1821-1830.
- [32] HE K,ZHANG X,REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;770-778.
- [33] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [J]. arXiv: 1301. 3781,2013.
- [34] 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.



**YU Yongxin**, born in 2000, postgraduate, is a member of CCF (No. N4386G). Her main research interests include machine learning and natural language processing.



**JI Ke**, born in 1989, Ph. D, associate professor, is a member of CCF (No. 78936M). His main research interests include machine learning and recommendation systems.

(责任编辑:柯颖)