基于多通道特征增强与图文相似度感知的虚假新闻检测

张仕斌[1,2,3],蔡松睿[1,2,3,4],杨 敏[2,3*],陈世航[2,3]

(1. 成都信息工程大学人工智能学院(芯谷产业学院),四川成都 610225;
2. 成都信息工程大学网络空间安全学院(芯谷产业学院),四川成都 610225;
3. 先进密码技术与系统安全四川省重点实验室,四川成都 610225;
4. 先进微处理器技术国家工程研究中心(工业控制与安全分中心),四川成都 610225)

摘 要: 人工智能(Artificial Intelligence,AI)技术的快速发展在丰富互联网内容生态的同时,也加剧了多模态虚假新闻的广泛传播,特别是深度伪造技术的应用使得虚假信息在视觉与语义层面呈现出较高的逼真性,严重威胁网络公共空间的信任体系. 尽管现有的多模态虚假新闻检测技术已利用跨模态注意力机制及大语言模型(Large Language Models,LLMs)实现了多模态语义对齐与推理增强,但这些方法在特定场景下仍面临挑战. 一方面,通用大模型存在"幻觉"风险,且多局限于粗粒度的语义融合,难以精准捕捉图文实体间的不匹配冲突;另一方面,现有模型往往忽略了对图像频域物理伪影及文本情感操纵信号的挖掘,导致其在面对生成式AI高保真的伪造内容时鉴别力受限. 针对上述问题,本文提出了一种基于多通道特征增强与图文相似度感知的图注意力网络(Multimodal Similarity-aware Graph Attention Network,MS-GAT). 该方法首先设计了多通道特征提取模块,其利用双向编码器表征(Bidirectional Encoder Representations from Transformers,BERT)模型提取文本的深层语义与情感特征,并结合视觉Transformer(Vision Transformer,ViT)获取图像空间特征,同时引入快速傅里叶变换(Fast Fourier Transform,FFT)捕捉图像频域中的异常伪影,并通过自适应门控单元实现多通道特征的加权融合. 在此基础上,本文构建了一个包含图文实体节点与模态枢纽节点的相似度感知异构图,利用对比语言-图像预训练(Contrastive Language-Image Pre-training,CLIP)模型计算各节点在共享语义空间中的相似度,并以此显式地建模图文间的细粒度关联. 最后,模型利用图注意力网络(Graph Attention Network,GAT)聚合邻域信息,通过注意力权重动态调整节点间的关联强度以聚焦图文不一致特征,并配合自适应的多任务损失函数解决联合学习中的优化不平衡问题. 所提方法在Weibo17和CFND数据集上的准确率分别达到94.5%和87.6%,各项关键性能指标均优于现有主流基线. 研究结果表明,本方法通过融合图文多通道特征与结构化推理机制,实现了对图文深层语义冲突的捕捉,为提升多模态虚假新闻检测的可解释性与鲁棒性提供了新的视角与技术支撑.

关键词: 虚假新闻检测;多模态融合;图文相似度感知;多通道特征提取;图注意力网络;异构图

基金项目: 国家重点研发计划(No.2022YFB3103103);先进密码技术与系统安全四川省重点实验室开放课题(No.SKLACSS-202404);成都市重点研发项目(No.2023-XT00-00002-GX);成都市重点研发支撑计划项目(No.2024-YF05-01227-SN)

中图分类号: TP391.1 **文献标识码:** A **文章编号:** 0372-2112(2025)12-4614-16

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250650

Fake News Detection via Multi-Channel Feature Enhancement and Visual-Textual Similarity Awareness

ZHANG Shi-bin[1,2,3], CAI Song-rui[1,2,3,4], YANG Min[2,3*], CHEN Shi-hang[2,3]

(1. *College of Artificial Intelligence (Xin Gu Industrial College), Chengdu University of Information Technology, Chengdu, Sichuan* 610225, *China*;
2. *School of Cybersecurity (Xin Gu Industrial College), Chengdu University of Information Technology, Chengdu, Sichuan* 610225, *China*;
3. *Advanced Cryptography and System Security Key Laboratory of Sichuan Province, Chengdu, Sichuan* 610225, *China*;
4. *SUGON Industrial Control and Security Center, Chengdu, Sichuan* 610225, *China*)

Abstract: The rapid development of artificial intelligence (AI) technology has enriched the Internet content ecosystem while simultaneously exacerbating the widespread propagation of multimodal fake news. In particular, the application

收稿日期:2025-07-24;录用日期:2025-12-17;责任编辑:覃怀银
*通讯作者:杨敏

of deepfake technology renders false information highly realistic at both visual and semantic levels, posing a severe threat to the trust system of the online public sphere. Although existing multimodal fake news detection techniques have utilized cross-modal attention mechanisms and large language models (LLMs) to achieve multimodal semantic alignment and reasoning enhancement, these methods still face challenges in specific scenarios. On one hand, general-purpose large models are prone to "hallucination" risks and are often limited to coarse-grained semantic fusion, making it difficult to accurately capture mismatch conflicts between visual and textual entities. On the other hand, existing models often overlook the mining of physical artifacts in the image frequency domain and emotional manipulation signals in the text, resulting in limited discrimination capability when facing high-fidelity fake content generated by generative AI. To address the aforementioned issues, this paper proposes a multimodal similarity-aware graph attention network (MS-GAT) based on multi-channel feature enhancement. The method first designs a multi-channel feature extraction module, utilizing the bidirectional encoder representations from transformers (BERT) model to extract deep semantic and emotional features of the text, combined with the vision transformer (ViT) to acquire image spatial features. Simultaneously, it introduces the fast Fourier transform (FFT) to capture anomalous artifacts in the image frequency domain and implements weighted fusion of multi-channel features through an adaptive gating unit. Building upon this, this paper constructs a similarity-aware heterogeneous graph containing visual-textual entity nodes and modality hub nodes. It utilizes the CLIP model to calculate the similarity of each node in a shared semantic space and thereby explicitly models the fine-grained associations between images and text. Finally, the model employs the graph attention network (GAT) to aggregate neighborhood information, dynamically adjusting the association strength between nodes via attention weights to focus on visual-textual inconsistency features, and incorporates an adaptive multi-task loss function to resolve the optimization imbalance problem in joint learning. The proposed method achieves accuracies of 94.5% and 87.6% on the Weibo17 and CFND datasets, respectively, with all key performance indicators outperforming existing mainstream baselines. Research results indicate that by integrating multi-channel visual-textual features with structured reasoning mechanisms, the proposed method successfully captures deep semantic conflicts between images and text, providing a new perspective and technical support for enhancing the interpretability and robustness of multimodal fake news detection.

Key words: fake news detection; multimodal fusion; visual-textual similarity awareness; multi-channel feature extraction; graph attention network; heterogeneous graph

Foundation Item(s): National Key Research and Development Program of China (No. 2022YFB3103103); Open Fund of Advanced Cryptography and System Security Key Laboratory of Sichuan Province (No.SKLACSS-202404); Chengdu Key Research and Development Project (No. 2023-XT00-00002-GX); Chengdu Key Research and Development Support Program Project (No.2024-YF05-01227-SN)

1 引言

在当前的数字化浪潮中,社交媒体以其即时性与开放性重塑了信息传播的方式,然而,这也为虚假新闻的滋生与蔓延提供了便捷. 虚假新闻通常指受特定利益驱动而故意编造的欺骗性内容[1]. 此类信息的广泛传播严重侵蚀了社会信任体系,并对公共秩序与社会稳定构成了严峻挑战[2]. 因此,研究有效的虚假新闻检测技术,对于及时阻断虚假新闻传播、塑造健康的网络舆论生态具有重要的现实意义.

随着多媒体技术的普及,图文结合的多模态形式已逐渐成为虚假新闻传播的主要载体. 视觉元素的介入契合了受众"眼见为实"的认知心理,在增强内容煽动性的同时,也加快了其在社交网络中的扩散[3]. 图1展示了多模态虚假新闻的四种典型形态. 其中,图1(a)~图1(c)属于内容的直接伪造:图1(a)为图文双重虚假;图1(b)虽文本属实,但图像包含明显的物理篡改痕迹;图1(c)的图像则来源真实,但配以虚假的文本描述. 相比之下,图1(d)呈现出更为隐蔽的图文语义错配,即图文素材各自虽未被篡改,但二者的组合却构建了虚假的事实关联. 这种"张冠李戴"式的跨模态语义冲突,往往具有更高的隐蔽性与欺骗性. 面对上述多样化的伪造手段,早期主要依赖单一模态分析的检测方法[4,5]显露出明显的局限性. 具体而言,仅依赖文本语义的分析方法[6,7]无法感知图1(a)和图1(b)中的视觉篡改痕迹;而侧重图像取证的技术[8,9]不仅难以判定图1(a)和图1(c)中的文本虚构,在面对图1(d)中素材真实但图文语义逻辑冲突的隐蔽伪造时,更是难以奏效.

鉴于单模态检测方法的固有限制,学术界和工业界开始积极探索基于多模态融合的虚假新闻检测技术[10~13],并取得了显著进展. 然而,随着生成式人工智能和深度伪造技术的快速演进,虚假新闻的生成手段日益复杂化和高保真化,现有的多模态检测方法在深层特征挖掘与跨模态推理机制上显露出明显的局限性,主要体现在以下两个方面.

(*a*) 文本和图片都是虚假的

(*b*) 带有正常文本的虚假图片

(*c*) 带有正常图片的虚假文本

(*d*) 图文语义不匹配

图1 多模态虚假新闻的四种典型形态

首先,单纯依赖表层语义特征的融合策略已难以应对高保真的伪造内容. 现有主流方法[14,15]多侧重于图文语义的拼接与加权,在面对细节造假的深度伪造样本时鉴别力有限. 研究表明,关键的伪造线索常隐匿于底层的物理信号与情感特征之中[16,17]. 例如图像频域中的异常伪影及文本中违背事实的情感煽动. 然而,当前多数方法尚未能有效整合这些多通道特征,导致其在复杂语义混淆场景下的鲁棒性受限.

其次,现有的跨模态融合策略缺乏显式的结构化推理与细粒度对齐机制,难以捕捉不同模态实体间的深层逻辑冲突. 例如,验证"图像中的车辆"与"文本中的事故描述"是否一致,依赖于模型对细粒度特征的推理能力. 尽管当前的注意力机制能够建立隐式关联,但其缺乏明确的拓扑结构以显式建模模态内及模态间的复杂交互. 这种结构化推理的缺失不仅限制了检测精度,也导致模型缺乏可解释性,难以满足实际应用对判决依据透明度的需求.

针对上述挑战,本文提出了一种基于多通道特征增强与图文相似度感知的图注意力网络(Multimodal Similarity-aware Graph Attention Network, MS-GAT). 该方法首先设计了多通道特征提取机制,分别从图像空间域与频域、文本语义与情感维度挖掘特征,旨在捕捉隐藏的物理伪影与情感操纵痕迹. 其次,针对跨模态推理能力的不足,该方法通过计算图文实体的语义相似度,构建了包含模态枢纽节点的异构图,将多模态融合转化为节点间的信息聚合过程,并利用图注意力机制自适应聚焦于图文语义不一致的关键路径. 此外,模型采用基于可学习参数的自适应多任务损失函数,自动平衡各任务分支的优化贡献. 实验结果显示,MS-GAT在Weibo17[18]和CFND[19]数据集上的准确率分别达到94.5%和87.6%,虚假新闻F1分数分别为94.2%和84.5%,均优于现有基线方法. 研究表明,该框架成功融合了多通道深层特征与结构化推理机制,为多模态虚假新闻检测提供了新的技术视角.

本文的主要贡献归纳如下:

(1)提出了一种融合频域与情感维度的多通道特征表示方法. 该方法通过门控机制融合图像空间/频域与文本语义/情感特征,有效弥补了传统方法仅依赖表层语义的不足,为模型提供了更为全面且鲁棒的判别依据.

(2)构建了基于相似度感知的多模态异构图与MS-GAT检测框架. 通过引入实体相似性连接与模态枢纽节点,该框架实现了跨模态细粒度关联的显式推理,在提升检测精度的同时显著增强了模型的可解释性.

(3)引入了自适应的多任务损失加权机制. 该机制利用可学习参数动态平衡不同模态信息与分类任务的贡献,有效避免了人工调参的复杂性与潜在偏差,提升了模型的泛化能力与训练效率.

2 相关工作

伴随社交媒体的普及,虚假新闻的泛滥及其带来的社会挑战日益凸显. 为应对这一挑战,学术界与工业界对虚假新闻检测技术展开了深入探索. 本文遵循技术演进脉络,从单模态与多模态检测两个维度对相关研究进行综述.

2.1 单模态虚假新闻检测

传统虚假新闻检测主要基于单模态数据,根据信息载体可分为文本、视觉及社交上下文三类技术路线.

在文本检测领域,研究重点在于挖掘语言学特征与语义表征[20]. 早期研究多采用支持向量机(Support Vector Machine, SVM)、朴素贝叶斯等模型处理词汇分布及情感极性等手工特征[21,22]. 随后,循环神经网络(Recurrent Neural Network, RNN)、长短期记忆网络(Long Short-Term Memory, LSTM)及Transformer等深度学习架构被广泛应用于捕捉长程语义依赖[23~25]. 例如,Bahad等人[25]利用双向长短期记忆网络(Bidirectional Long Short-Term Memory, Bi-LSTM)模型证明了序列建模在上下文捕捉上优于卷积神经网络(Convolutional Neural Network, CNN). 然而,纯文本方法难以验证内容与视觉证据或外部事实的一致性.

视觉检测技术主要涵盖语义表征与视觉取证两个方向. 前者利用视觉几何组(Visual Geometry Group, VGG)、残差网络(Residual Network, ResNet)等提取图像的高层语义,以识别特定的伪造模式;后者则通过分析纹理、颜色直方图或误差水平来检测物理篡改痕迹[8,9]. 目前,现有的视觉方法多侧重于空间域特征,在

应对生成式人工智能内容(Artificial Intelligence Generated Content,AIGC)这类高保真的图像时,鉴别能力仍有待提升.

社交上下文分析侧重于挖掘新闻传播的拓扑结构,通常利用图神经网络(Graph Neural Network,GNN)聚合用户交互与传播动力学特征[21]. 例如,Chang等人[26]通过构建全局注意力记忆网络来提取传播特征,取得了良好的检测效果. 尽管此类方法具有参考价值,但其高度依赖社交互动数据,在新闻传播初期常面临数据稀疏问题,且难以直接判断内容本身的真实性.

综上所述,单模态方法虽在特定场景下具备成效,但其局限在于难以捕捉跨模态的关联特征与矛盾信号. 面对图文冲突或高保真虚假内容时,此类系统的检测性能受限. 因此,如何融合多源异构数据并构建跨模态一致性分析框架,已成为当前研究的核心挑战.

2.2　多模态虚假新闻检测

鉴于社交媒体中大量虚假新闻以图文结合的形式传播,且文本与图像之间的不一致性往往成为识别虚假新闻的重要线索,多模态检测方法逐渐成为研究热点. 该类方法旨在联合分析来自不同模态的信息,以提升虚假新闻检测的准确性和鲁棒性.

2.2.1　基础融合策略与语义对齐

多模态虚假新闻检测的研究最初侧重于基础融合策略. 早期融合通过在输入层直接拼接文本与图像特征构建统一向量,并交由分类器处理[27];而晚期融合则在各模态独立预测的基础上进行结果集成[28]. 虽然这些方法逻辑直观,但早期融合难以消除特征空间的异质性差异,晚期融合则缺乏模态间的深度交互,导致信息利用效率受限. 因此,中间层融合技术逐渐占据主流,其中注意力机制被广泛应用于捕捉模态间的隐式关联. 例如,Wu等人[15]提出的多模态协同注意力网络,通过跨模态注意力机制整合图像空间特征与文本语义特征,其检测效果优于传统的拼接方式.

近年来,基于预训练模型的深层语义对齐技术得到了广泛应用. 随着双向编码器表征(Bidirectional Encoder Representations from Transformers,BERT)模型与视觉Transformer(Vision Transformer,ViT)等单模态模型的成熟,研究者利用此类架构提升了多模态特征的提取效果. 此外,CLIP模型通过在海量图文对上进行对比学习,展现了较强的跨模态对齐能力[29]. 大语言模型(Large Language Model,LLM)与大规模视觉语言模型(Large Vision-Language Model,LVLM)则通过指令微调进一步增强了模型的逻辑推理能力[30~32]. 然而,此类方法在虚假新闻检测场景中仍存在局限性:模型固有的幻觉风险可能产生与事实不符的错误联想;同时,现有方法多侧重于粗粒度语义对齐,难以精准捕捉图文间深层的逻辑冲突,且对高保真伪造内容的鉴别能力有待提升.

2.2.2　多通道深层特征挖掘与表示

为了克服仅依赖表层语义特征的局限,当前研究前沿正转向挖掘多通道深层特征,重点关注视觉频域与文本情感两个维度,以捕捉伪造内容的物理痕迹与心理倾向.

在视觉频域分析方面,生成对抗网络(Generative Adversarial Network,GAN)与扩散模型生成的图像在空间域已达到高度逼真的效果,CNN难以有效识别其伪造特征[33,34]. 然而,此类图像在频域往往残留特定的物理伪影或压缩痕迹[35]. Jing等人[36]提出了多模态渐进式融合网络,利用Transformer架构提取并融合频域、空间及文本特征,增强了模型捕捉微细伪造痕迹的能力. 此外,Wu等人[15]提出的多模态协同注意力网络同样引入了图像频域信息,用以学习多模态特征间的相互依赖关系.

在情感特征分析方面,虚假新闻常利用强烈的煽动性加速传播[37]. 研究发现,虚假内容往往包含较高比例的负面情感词[38]. Ajao等人[39]利用情感词典构建特征以辅助检测新闻真实性. Zhang等人[40]则通过建模新闻发布者的情感与社交评论情感之间的关联,提升了检测性能. 然而,当前多数工作仍将情感分析视为独立的下游任务或补充特征[41~43],尚未建立起将情感信息作为核心特征与语义、视觉信息深度协同的多通道框架.

2.2.3　基于图结构的细粒度推理与知识增强

为了克服现有方法在捕捉细粒度关联方面的不足,基于GNN的显式推理已成为当前的研究重点,其核心在于通过建模实体间的交互来揭示逻辑冲突.

在外部知识增强方面,研究者通常利用实体对齐技术引入语义关联以增强检测能力[44]. 例如,Fu等人[45]提出了一种基于知识图谱的多领域检测框架,通过改进BERT并注入实体三元组来丰富新闻背景. 然而,此类方法高度依赖外部知识库的质量,面临实体链接错误、知识库不完备及噪声干扰等挑战[46]. 为了突破外部知识库的局限,近期研究转向构建基于内容的异构图进行闭环推理. 例如,Qian等人[47]提出了知识感知多模态自适应图卷积网络,将文本概念与视觉信息联合建模;Qi等人[48]提出的实体增强的多模态假新闻检测框架(Entity-cnhanced Multimodal FakE News Detection,EM-FEND)则引入视觉实体识别技术,通过增强跨模态对齐提升了模型对噪声样本的鲁棒性.

尽管上述方法在显式推理方面有所进展,但现有的图结构建模仍存在局限性:一是多侧重于模态间的

粗粒度连接，缺乏对多模态相似度差异的感知机制，导致在面对传播噪声时推理能力受限；二是异构图尚未实现多通道深层特征的融合. 针对上述问题，本文提出一种新的多模态虚假新闻检测框架. 该框架通过构建显式的跨模态实体关联图、融合多通道单模态特征并引入相似度感知注意力机制，旨在实现更具可解释性的信息融合与不一致性建模，从而提升检测性能.

3 相关理论及问题定义

3.1 图神经网络理论

GNN是一类以图结构数据为输入、通过节点间信息传递实现特征学习的深度模型[49]. 给定图$G=(V,E)$，其中V为节点集合，E为边集合. 节点特征矩阵表示为$\boldsymbol{X}\in\mathbb{R}^{|V|\times F}$，其中$|V|$是节点数量，$F$是特征维度. 图的连接拓扑结构通常由邻接矩阵$\boldsymbol{A}\in\{0,1\}^{|V|\times|V|}$表示.

GNN的核心在于消息传递机制[50]，在第k层网络($k\geqslant 1$)中，节点$v\in V$首先聚合其一跳邻居节点$u\in N(v)$在第$k-1$层的隐藏表示$\{\boldsymbol{h}_u^{(k-1)}\}$生成聚合信息$\boldsymbol{a}_v^{(k)}$，其中$N(v)$表示节点$v$的一跳邻居集合；然后结合自身先前表示$\{\boldsymbol{h}_v^{(k-1)}\}$(对于$k=0$，$\{\boldsymbol{h}_v^{(0)}\}=\boldsymbol{X}_v$，即节点$v$的原始特征)通过更新函数得到新的隐藏表示$\{\boldsymbol{h}_v^{(k)}\}$. 聚合函数通常包括求和、平均或最大操作，而更新函数一般为带有非线性激活$\sigma(\cdot)$的仿射变换：

$$\boldsymbol{h}_v^{(k)}=\sigma\left(\boldsymbol{W}^{(k)}[\boldsymbol{h}_v^{(k-1)}\|\,\boldsymbol{a}_v^{(k)}]+\boldsymbol{b}^{(k)}\right) \tag{1}$$

其中，$\boldsymbol{W}^{(k)}$和$\boldsymbol{b}^{(k)}$是可学习的权重矩阵和偏置向量，$\|$表示向量连接操作.

通过堆叠k层消息传递，GNN能够捕获不同跳数的拓扑特征. 对于本研究涉及的多模态虚假新闻检测的图级分类任务，模型需要将所有节点的最终表示汇聚成一个单一的图级表示$\boldsymbol{h}_G$. 这一过程通常通过对所有节点表示求和、求平均等方式来实现.

最终，图级表示$\boldsymbol{h}_G$被输入分类器以输出预测结果. GNN框架通过局部结构建模与全局信息汇聚，为处理复杂图结构数据及实现多模态信息融合提供了理论支撑.

3.2 跨模态表示学习模型CLIP

CLIP是一种基于对比学习的跨模态预训练模型，旨在共享嵌入空间中学习文本与图像的联合表示. 该模型由文本编码器(text encoder)与图像编码器(image encoder)组成，通过在大规模图文对数据集上进行端到端训练，实现跨模态特征的深度对齐[51]. 其训练目标是最大化匹配图文对在嵌入空间中的相似度，同时最小化非匹配对的相似度.

经过预训练后，CLIP模型能够为输入的文本T和图像I分别生成固定维度的嵌入向量$\boldsymbol{v}_T=\text{TextEncoder}(T)$和$\boldsymbol{v}_I=\text{ImageEncoder}(I)$. 这两个嵌入向量位于同一个语义共享空间$\mathbb{R}^D$中，其中$D$为嵌入空间的维度. 文本与图像间的跨模态相似度通常通过计算其对应嵌入向量的余弦相似度来衡量：

$$\text{Similarity}(T,I)=\frac{\boldsymbol{v}_T\cdot\boldsymbol{v}_I}{\|\boldsymbol{v}_T\|\ \|\boldsymbol{v}_I\|} \tag{2}$$

其中，$\boldsymbol{v}_T\cdot\boldsymbol{v}_I$表示向量点积；$\|\boldsymbol{v}_T\|$和$\|\boldsymbol{v}_I\|$表示向量的L2范数. CLIP模型计算的跨模态相似度能够量化文本与图像内容在语义层面的一致性，这为本研究构建基于相似度感知的图注意力网络提供了计算依据.

3.3 问题定义

本文研究的核心任务是多模态虚假新闻检测. 该任务旨在通过联合分析给定的文本与图像信息，判定该信息实例是否为虚假内容.

具体而言，给定包含N个样本的数据集D，其中第i个实例由文本内容T_i和相关的图像内容I_i组成. 假定文本内容T_i来自文本空间T，图像内容I_i来自图像空间I. 数据集D可以形式化表示为$D=\left\{(T_i,I_i)\right\}_{i=1}^{N}$. 每个信息实例$(T_i,I_i)$关联一个真实的二元标签$y_i=\{0,1\}$，表示该实例的真伪性. 在本研究中，$y_i=1$表示该信息为虚假新闻，$y_i=0$表示该信息为真实新闻.

假设存在一个未知但固定的数据分布$P\left(T,I,y\right)$，数据集D中的实例$\left(T_i,I_i,y_i\right)$从中独立同分布采样获得. 多模态虚假新闻检测的目标是学习一个分类模型f，该模型接收多模态输入(T_i,I_i)并输出一个预测标签$\hat{y}_i\in\{0,1\}$或属于虚假类别的概率分数$\hat{p}_i\in[0,1]$. 形式上，该任务的目标是优化映射函数$f:T\times I\to\{0,1\}$或$f:T\times I\to[0,1]$，使得未见过的测试数据，其预测结果与真实标签尽可能一致，从而实现分类性能的最优化. 后续章节将基于此定义，详细阐述本文提出的模型架构.

4 模型介绍

4.1 整体框架

本文提出一种名为多模态相似度感知图注意力网络MS-GAT的多模态虚假新闻检测框架. 该框架的核心在于构建并处理一种能够显式建模图文细粒度语义关联的异构图，以实现跨模态信息的深层交互. 此外，MS-GAT结合了基于多通道特征的单模态检测分支，构建了一个联合学习框架. 该框架通过跨模态图融合与辅助单模态分析的协同作用，旨在提升检测系统的性能及判别结果的可解释性. 图2展示了MS-GAT的总体架构.

MS-GAT框架主要由四个核心模块组成：多模态数据预处理、多通道特征提取、多模态相似度感知异构图构建及虚假新闻检测模块. 后续各小节将对各模块的功能原理与实现细节进行详细阐述.

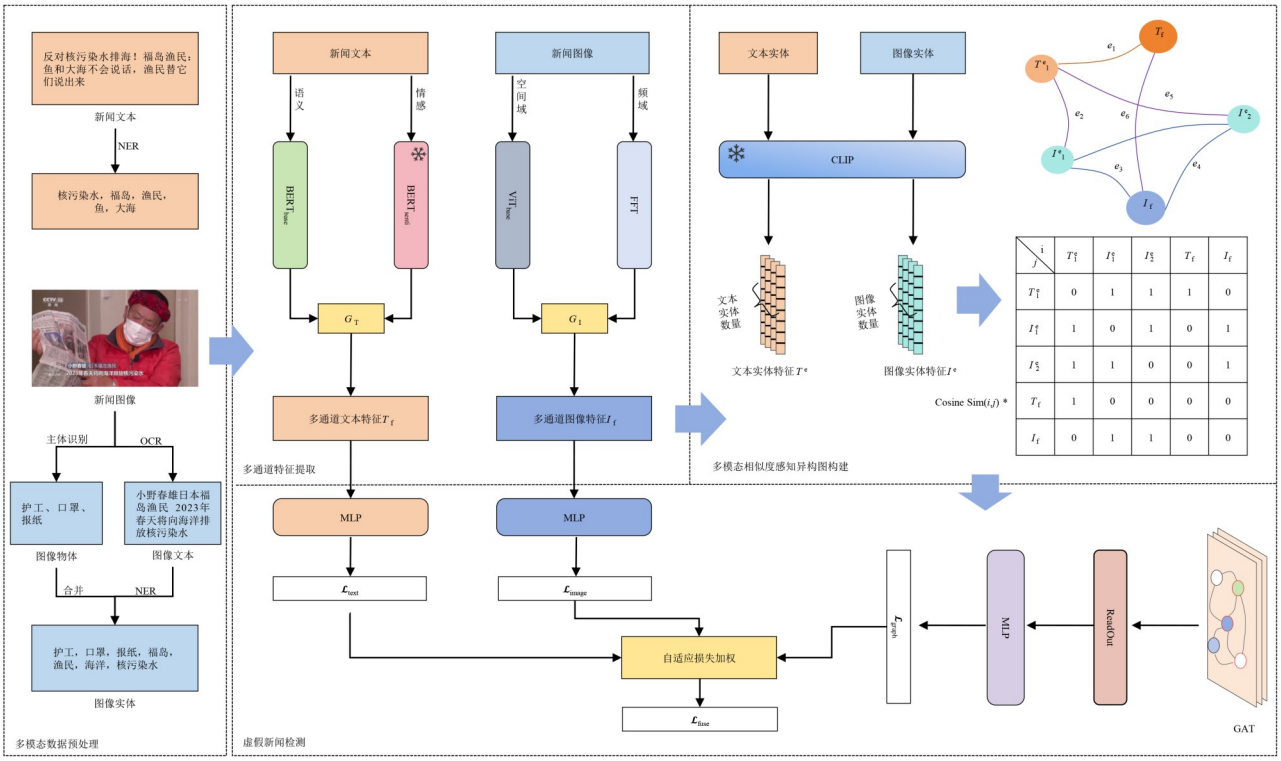


图2　多模态相似度感知的图注意力网络结构图

4.2　多模态数据预处理

本模块负责对原始多模态数据进行清洗、格式化及实体识别.

对于输入的文本内容T,该模块首先执行分词与清洗等标准预处理操作,随后利用命名实体识别(Named Entity Recognition,NER)工具提取文本实体集合E_T. 对于输入的图像内容I,该模块首先完成尺寸调整与像素归一化,并利用光学字符识别(Optical Character Recognition,OCR)技术提取图像中的嵌入文本,同时通过目标检测识别图像的主要视觉对象. 随后,NER 工具被应用于OCR 提取的文字结果以识别图像中的文本实体. 最后,该模块将识别出的文本实体与视觉对象取并集,构成图像实体集合E_I.

经过此阶段,模型获得了标准化的多模态数据,以及用于后续构建图结构的文本实体集合E_T和图像实体集合E_I.

4.3　多通道特征提取

多通道特征提取模块旨在从文本与图像中挖掘关键表征信息,并通过门控机制进行初步融合,为后续图结构构建提供丰富的节点特征.

在文本模态处理中,该模块分别提取语义特征与情感特征. 语义特征利用预训练的Bert-base[52]模型获取,用以捕捉上下文语义信息;情感特征则通过在情感标注数据集上微调的BERT模型提取,捕捉文本所表达的情感倾向.

在图像模态处理中,该模块提取空间域特征与频域特征. 空间域特征采用预训练的ViT-base[53]获取,侧重于图像空间特征的捕捉;频域特征通过快速傅里叶变换(Fast Fourier Transform,FFT)处理图像得到,旨在识别由于编辑或压缩产生的物理伪影.

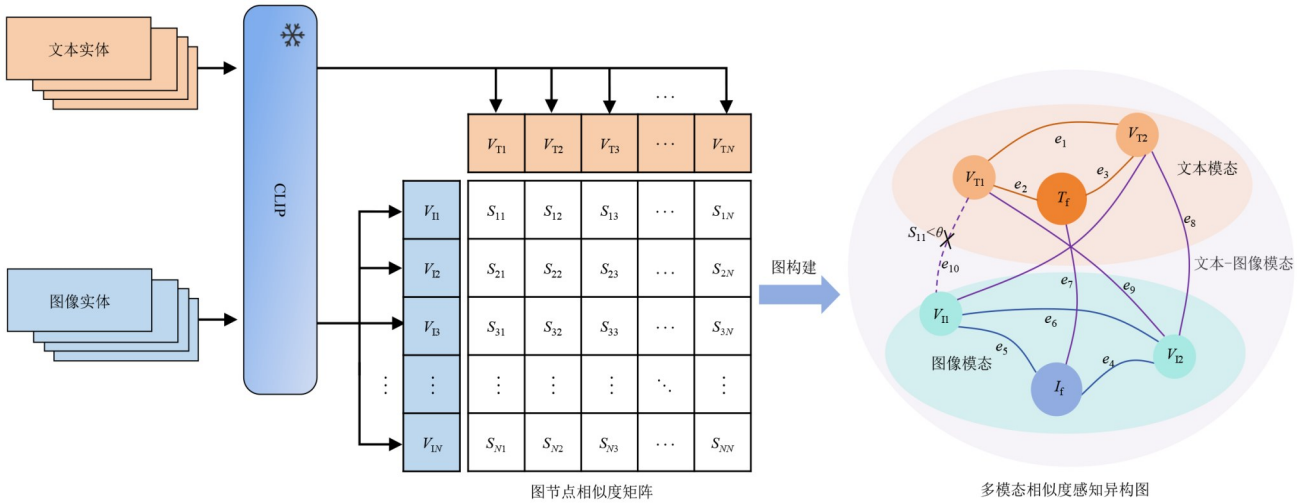
为整合各模态内部的多通道特征,本方法引入了门控融合机制. 该机制通过一个包含多层感知机与Sigmoid激活函数的门控单元,学习生成动态权重以平衡不同通道的贡献. 形式上,设文本语义特征为$\boldsymbol{T}_{\text{sem}}$,情感特征为$\boldsymbol{T}_{\text{senti}}$,门控单元输出权重为$\boldsymbol{g}\in[0,1]^D$($D$为特征维度),融合后的文本特征$\boldsymbol{T}_{\text{f}}$可表示为

$$\boldsymbol{T}_{\text{f}}=\boldsymbol{g}\odot\boldsymbol{T}_{\text{sem}}+(1-\boldsymbol{g})\odot\boldsymbol{T}_{\text{senti}} \tag{3}$$

其中,$\odot$表示元素乘法. 该模块采用类似的门控机制融合图像的空间域特征I_{spa}与频域特征I_{freq},得到融合图像特征I_{f}. 这种自适应融合策略能够根据输入内容动态调整特征通道的重要性. 融合后的特征向量不仅作为异构图中模态枢纽节点的初始属性,还被直接输入至后续单模态检测分支参与推理.

4.4　多模态相似度感知异构图构建

本模块负责将多模态信息转化为能够显式捕捉图文细粒度跨模态关联的图结构. 该异构图作为图注意力网络(Graph Attention Network,GAT)[54]的输入,用于指导信息在不同节点间的传播与聚合,其构建过程如图3所示.



注：图节点相似度矩阵中的S_{ij}表示图像实体i与文本实体j特征的余弦相似度.

图3 多模态相似度感知异构图构建过程

首先，该模块定义了图的节点集合V. 节点代表多模态信息中的关键信息点，分为三类：文本实体节点V_{T}、图像实体节点V_{I}及模态枢纽节点$V_{\mathrm{M}}=\{\boldsymbol{T}_{\mathrm{f}},\boldsymbol{I}_{\mathrm{f}}\}$. 图的总节点数$|V|=|V_{\mathrm{T}}|+|V_{\mathrm{I}}|+2$. 文本实体和图像实体的初始特征由预训练的CLIP模型进行提取，而模态枢纽节点的初始特征则采用4.3节生成的融合特征向量.

其次，该模块通过边集合E反映节点间的关联强度. 边的权重w_{ij}在GAT中用于指导信息聚合，其构建核心在于捕捉跨模态相似性关联. 对于任意两个实体节点$i,j\in V_{\mathrm{T}}\cup V_{\mathrm{I}}$，该模块计算其初始特征$\boldsymbol{h}_i^{(0)}$和$\boldsymbol{h}_j^{(0)}$之间的余弦相似度：

$$\mathrm{Similarity}(i,j)=\frac{\boldsymbol{h}_i^{(0)}\cdot\boldsymbol{h}_j^{(0)}}{\left\|\boldsymbol{h}_i^{(0)}\right\|\left\|\boldsymbol{h}_j^{(0)}\right\|} \tag{4}$$

该计算在CLIP的共享嵌入空间中进行. 基于此相似度，该方法构建实体相似性边：对于文本实体对$i,j\in V_{\mathrm{T}}$，若$\mathrm{Similarity}(i,j)>\theta_{\mathrm{t}}$，则在$i$和$j$之间添加权重为$w_{ij}=\mathrm{Similarity}(i,j)$的无向边；对于图像实体对，采用阈值$\theta_{\mathrm{i}}$进行类似处理. 对于跨模态实体对，计算任意文本实体$i\in V_{\mathrm{T}}$和图像实体$j\in V_{\mathrm{I}}$，若$\mathrm{Similarity}(i,j)>\theta_{\mathrm{m}}$，则建立跨模态相似性边，权重设为$w_{ij}=\mathrm{Similarity}(i,j)$.

此外，该图结构还包含连接实体与模态枢纽的边，旨在将模态内的实体信息汇聚至枢纽节点. 对于每个文本实体$i\in V_{\mathrm{T}}$，该模块将添加一条连接至文本枢纽$\boldsymbol{T}_{\mathrm{f}}$的无向边；对于每个图像实体$j\in V_{\mathrm{I}}$，该模块将添加连接至图像枢纽$\boldsymbol{I}_{\mathrm{f}}$的无向边. 此类边的权重均设置为固定值$w_{\mathrm{m}}$.

构建完成后，该模块获得包含节点特征$\boldsymbol{X}=\{\boldsymbol{h}_v^{(0)}|v\in V\}$、边索引及边权重$W=\left\{w_{ij}|(i,j)\in E\right\}$的多模态相似度感知异构图$G$. 该图通过细粒度语义相似性构建连接，并通过模态枢纽加强信息汇聚，具体的构建步骤如算法1所示.

算法1 多模态相似度感知异构图的构建

输入： 文本实体集合$E_{\mathrm{T}}=\{T_1,T_2,\cdots,T_N\}$，图像实体集合$E_{\mathrm{I}}=\{I_1,I_2,\cdots,I_N\}$，融合后的模态枢纽特征$\boldsymbol{T}_{\mathrm{f}}$、$\boldsymbol{I}_{\mathrm{f}}$，CLIP模型，相似度阈值$\theta_{\mathrm{t}}$、$\theta_{\mathrm{i}}$、$\theta_{\mathrm{ti}}$，枢纽边权重$w_{\mathrm{m}}$

输出： 多模态相似度感知异构图$G=(V,E)$，图节点特征矩阵$\boldsymbol{X}$

// 步骤1：节点初始化与特征提取
$V_{\mathrm{T}}\leftarrow$为E_{T}中的每个实体T_i创建节点$V_{\mathrm{T}i}$
$V_{\mathrm{I}}\leftarrow$为E_{I}中的每个实体I_j创建节点$V_{\mathrm{I}j}$
$V\leftarrow V_{\mathrm{T}}\cup V_{\mathrm{I}}\cup\{\boldsymbol{T}_{\mathrm{f}},\boldsymbol{I}_{\mathrm{f}}\}$
for 每个节点$v\in V$ do
  if $v\in V_{\mathrm{T}}$ then $\boldsymbol{X}[v]\leftarrow$CLIP_text_encoder
  else if $v\in V_{\mathrm{I}}$ then $\boldsymbol{X}[v]\leftarrow$CLIP_image_encoder
  else $\boldsymbol{X}[v]\leftarrow\{\boldsymbol{T}_{\mathrm{f}},\boldsymbol{I}_{\mathrm{f}}\}$
end for
// 步骤2：基于相似度的边构建
$E\leftarrow\varnothing$ // 将边集合E初始化为空集
for 每对节点(u,v) in $(V_{\mathrm{T}}\times V_{\mathrm{T}})\cup(V_{\mathrm{I}}\times V_{\mathrm{I}})\cup(V_{\mathrm{T}}\times V_{\mathrm{I}})$ do
  $s\leftarrow$cosine_similarity($\boldsymbol{X}[u],\boldsymbol{X}[v]$)
  if $(u,v\in V_{\mathrm{T}}$ and $s>\theta_{\mathrm{t}})$ or $(u,v\in V_{\mathrm{I}}$ and $s>\theta_{\mathrm{i}})$ or $(u\in V_{\mathrm{T}},v\in V_{\mathrm{I}}$ and $s>\theta_{\mathrm{ti}})$ then
    $E\leftarrow E\cup\{(u,v,\mathrm{weight}=s)\}$
  end if
end for
// 步骤3：实体-模态枢纽边构建
for 每个实体节点$v\in V_{\mathrm{T}}\cup V_{\mathrm{I}}$ do
  $E\leftarrow E\cup\left\{\left(v,\{\boldsymbol{T}_{\mathrm{f}},\boldsymbol{I}_{\mathrm{f}}\},\mathrm{weight}=w_{\mathrm{m}}\right)\right\}$
end for
return $G=(V,E)$, $\boldsymbol{X}$

4.5 虚假新闻检测

本节详细阐述模型并行处理的两个核心检测分支:图检测分支和单模态检测分支. 两个分支通过联合损失函数进行端到端训练,以实现最终的分类预测.

4.5.1 图检测分支

该分支利用多层图注意力网络在异构图 G 上学习节点表征. GAT通过注意力机制为邻居节点分配差异化权重,以实现信息的高效聚合. 对于图中的任意节点 v,其在第 l 层的特征表示为 $\boldsymbol{h}_v^{(l)}$. 从第 l 层到第 $l+1$ 层的更新过程如下.

首先,该方法计算节点 v 与其邻居节点 $j \in \mathbb{N}_v$ 之间的注意力系数 $e_{vj}^{(l)}$,用以反映邻居节点对中心节点的重要性:

$$e_{vj}^{(l)} = \text{LeakyReLU}\left(\boldsymbol{a}^{(l)\text{T}}[\boldsymbol{W}^{(l)}\boldsymbol{h}_v^{(l)} \| \boldsymbol{W}^{(l)}\boldsymbol{h}_j^{(l)}]\right) \tag{5}$$

其中,$\boldsymbol{W}^{(l)}$ 是可学习的线性变换矩阵;$\boldsymbol{a}^{(l)\text{T}}$ 为前馈网络权重向量;$\|$ 表示向量拼接操作. 随后,该方法将预先构建的相似性边权重 w_{ij} 与注意力系数相加,赋予相似度较高的邻居以更高的先验权重,并利用Softmax函数完成归一化,得到最终权重 $\alpha_{vj}^{(l)}$. 最后,根据计算出的权重加权聚合邻居节点特征,并通过激活函数得到下一层特征表示 $\boldsymbol{h}_v^{(l+1)}$:

$$\boldsymbol{h}_v^{(l+1)} = \sigma\left(\sum_{j \in \mathbb{N}_v} \alpha_{vj}^{(l)} \boldsymbol{W}^{(l)} \boldsymbol{h}_j^{(l)}\right) \tag{6}$$

经过 L 层传播后,模型获得最终节点表示矩阵 $\boldsymbol{H}(L)=\{\boldsymbol{h}_v^{(L)} | v \in V\}$. 随后,模型通过全局池化操作提取图级特征 $\boldsymbol{h}_\text{G}$,并将其输入由两层全连接网络构成的多层感知机(MultiLayer Perceptron, MLP)分类器,通过Softmax函数输出图预测概率分布 $\hat{y}_\text{G}$.

4.5.2 单模态检测分支

为了捕捉在图构建过程中可能忽略的全局信息或特定模态线索,本框架并行设置了文本与图像检测分支. 文本分支接收4.3节生成的融合文本特征 $\boldsymbol{T}_\text{f}$,通过独立MLP与Softmax处理输出预测概率 $\hat{y}_\text{T}$. 类似地,图像分支接收多通道图像特征 $\boldsymbol{I}_\text{f}$,由独立MLP分类器输出图像预测概率 $\hat{y}_\text{I}$.

4.5.3 联合学习与自适应加权损失

为动态平衡三个检测分支的贡献,本文采用基于可学习参数的自适应加权损失函数. 对于各分支输出,模型计算其与真实标签 y 之间的交叉熵损失 $\mathcal{L}_\text{text}$, $\mathcal{L}_\text{image}$, $\mathcal{L}_\text{graph}$. 总训练损失 $\mathcal{L}_\text{total}$ 定义为

$$L_\text{total} = \sum_{j \in \{\text{text, image, graph}\}} \exp\left(-\lambda_j\right) L_j + \lambda_j \tag{7}$$

其中,λ_j 是与各损失项相关联的可学习参数,代表对应任务的对数方差. 该机制通过优化 λ_j 自动调整权重,当对数方差较大时,该损失项的权重降低. 额外的 $\sum_j \lambda_j$ 项起到了正则化作用. 通过最小化 $\mathcal{L}_\text{total}$,模型能够自动评估并调整各分支在训练中的权重,从而优化多模态虚假新闻检测性能,避免了手动调参带来的不确定性.

5 实验与分析

本节在多个公开数据集上对MS-GAT框架的检测性能进行评估. 通过与基线方法的对比实验验证模型的有效性,利用消融实验探究关键组件的贡献,并结合可视化案例分析模型的工作机制与可解释性.

5.1 数据集介绍

本研究选取Weibo17与CFND两个公开数据集进行性能评估. 这些数据集包含丰富的图文结合样本,且标签标注过程严谨,能够代表现实网络环境下的虚假新闻分布.

Weibo17数据集源自新浪微博平台. 为了确保标签的客观性与权威性,该数据集通过“微博辟谣”官方系统收集了2012年5月至2016年1月期间的已验证谣言;真实新闻则选取自新华社等权威机构发布的推文.

CFND是一个大规模、细粒度的中文多模态数据集. 其虚假新闻采集6个主流事实核查网站及社交平台,真实新闻则源于4个官方新闻门户,确保了数据来源的可靠性. 此外,该数据集通过专家标注与交叉验证,将样本划分为疫情、健康等五大领域,反映了谣言在不同领域的传播特征.

在实验设置阶段,本研究将上述两个数据集按照6∶2∶2的比例随机划分为训练集、验证集和测试集,分别用于模型的训练、验证和最终性能评估. Weibo17和CFND数据集的详细统计信息如表1所示.

表1 数据集统计信息

数据集	新闻总数	虚假新闻数目	真实新闻数目	图片数目
Weibo17	9 402	4 748	4 654	9 402
CFND	26 665	10 271	16 394	26 665

5.2 实验设置与评价指标

5.2.1 实验设置

本研究基于PyTorch深度学习框架实现MS-GAT模型. 所有实验均在搭载NVIDIA GeForce RTX 4090D GPU的硬件环境下运行. 模型采用自适应矩估计(Adaptive Moment Estimation, Adam)优化器进行参数更新,初始学习率设为 5×10^{-3},训练过程总共迭代3个epoch.

模型训练的关键超参数设置如下:GAT层数设置为1;注意力头数设置为4;节点嵌入维度设置为768;Dropout比率设置为0.3;相似度阈值 θ_t、θ_i、θ_ti 均设置为0.6;模态连接权重 w_m 设置为1.0;批次大小设置为20.

5.2.2 评价指标

为了全面评估模型的性能,本研究采用涵盖总体性能与类级性能的二分类评价指标,包括准确率(Accu-

racy,Acc),以及针对每一类别的精确率(Precision,P)、召回率(Recall,R)和F1分数(F1-score,F1).所有指标均基于测试集上的真阳性(true positives,TP)、真阴性(true negatives,TN)、假阳性(false positives,FP)和假阴性(false negatives,FN)统计结果进行计算.

准确率定义为模型正确预测的样本数占总样本数的比例,用于反映模型的整体预测效能.针对真实新闻与虚假新闻两类标签,本研究分别计算其类级指标:精确率衡量模型预测为特定类别的样本中实际属于该类别的比例;召回率反映该类别实际样本中被模型正确识别的比例;F1分数则是精确率与召回率的调和平均数,用于综合评价模型在特定类别上的性能平衡.实验结果将报告测试集上的总体准确率,以及各类别对应的精确率、召回率与F1分数.

5.3 对比试验

本节将MS-GAT模型与现有基线方法在多模态虚假新闻检测任务中进行对比,旨在客观评估MS-GAT的性能表现.

5.3.1 基线方法

为评估本文所提模型的有效性,本研究选取了以下5类代表性的基线方法.

(1)单模态基线

此类方法仅利用文本或图像的单一维度信息,用于验证多模态融合的必要性.

(a)BERT[52]:该方法利用预训练的BERT提取文本特征,并结合分类层进行判定.

(b)ViT[53]:该方法利用预训练的ViT提取图像特征,并结合分类层进行判定.

(2)多模态融合基线

此类方法通过不同的对齐或融合策略结合图文信息,是当前多模态检测的主流方法.

(a)UniSMMC[55]:该方法结合聚合式与对齐式融合策略构建联合表征,并提出一种基于弱监督的模态对齐机制,通过将各单模态表征向具有正确预测的模态对齐来学习可信特征.

(b)MIAN[56]:该方法利用分层学习模块增强单模态表征,通过跨模态交互模块与协同注意力机制建模语义依赖,并引入逆向注意力机制以显式提取模态内及模态间的不一致特征.

(3)图融合基线

此类方法利用图神经网络显式建模新闻内容或社交上下文间的复杂逻辑.

(a)GAT[54]:该方法通过学习邻居节点的权重实现图信息的动态聚合.

(b)基于视觉推理提示的跨模态对齐(Cross-Modal Alignment with Visual Reasoning Prompting,CMA-VRP)[57]:该方法从新闻实体构建多模态图,并利用LLM与LVLM获取推理特征,最后通过图对比学习与融合技术整合特征.

(4)多通道特征融合基线

此类方法在跨模态融合前,预先从模态内部挖掘多维度信息,以捕捉虚假新闻的特定信号.

(a)基于分层社会注意力的双重情感特征(Hierarchical Social Attention-Dual Emotion Features,HSA-DEF)[40]:该方法利用Bi-LSTM编码语义,采用层次化社会注意力机制建模社交结构,并引入双重情感特征捕捉发布者与受众间的心理差异.

(b)多模态协同注意力网络(Multimodal Co-Attention Networks,MCAN)[15]:该方法同时提取图像的空间域与频域特征,通过多个协同注意力层实现视觉特征融合及其与文本特征的深度关联.

(c)多模态渐进式融合网络(Multimodal Progressive Fusion Network,MPFN)[36]:该方法利用Swin Transformer提取空间信息,并结合VGG19提取频域信息,最后通过多级融合策略处理图文表征.

(5)大模型基线

此类方法利用大型预训练模型的上下文理解能力进行推理,代表了当前的先进技术水平.

(a)LLMs:本研究选取Doubao1.6与Qwen3[31]作为基线.该实验将新闻内容处理为纯文本格式,并利用模型在海量数据中预训练所获得的上下文学习能力,直接对新闻的真实性进行推理和判断.

(b)NLIN[19]:该方法在预处理阶段通过OCR实现图文转换,并结合实体链接构建背景知识;随后采用低秩自适应(Low-Rank Adaptation,LoRA)微调的LLM编码信息,并通过推理解码器输出判定结果.

5.3.2 实验结果与分析

本研究将MS-GAT模型与5.3.1节所述基线方法在Weibo17和CFND数据集上进行对比实验.表2展示了各方法在测试集上的性能结果.

由表2可知,本文所提出的MS-GAT模型在两个数据集上的准确率及各类别F1分数均优于对比方法,验证了该框架在多模态虚假新闻检测任务中的有效性.

在Weibo17数据集上,MS-GAT的准确率达到0.945,虚假新闻类别的F1分数为0.942;在CFND数据集上,其准确率为0.876,虚假新闻F1分数为0.845.相比之下,仅依赖单一模态的BERT与ViT模型表现一般,证明了单模态特征在识别复杂伪造内容时的局限性,而MS-GAT通过融合图文互补信息提升了检测性能.

在与多模态融合及图融合方法的对比中,MS-GAT体现了细粒度推理的优势.尽管MIAN与UniSMMC利

表2　各方法在数据集上的性能对比

数据集	方法	准确率	虚假新闻			真实新闻		
			精确率	召回率	F1分数	精确率	召回率	F1分数
Weibo17	BERT	0.857	0.854	0.858	0.856	0.861	0.857	0.859
	ViT	0.778	0.771	0.785	0.778	0.786	0.772	0.779
	UniSMMC	0.935	0.928	0.927	0.927	0.929	0.930	0.929
	MIAN	0.936	0.950	0.920	0.935	0.923	0.952	0.937
	GAT	0.871	0.875	0.878	0.877	0.881	0.877	0.879
	CMA-VRP	0.938	0.933	**0.942**	0.941	0.929	0.906	0.938
	HSA-DEF	0.913	0.911	0.913	0.912	0.915	0.913	0.914
	MCAN	0.899	0.913	0.889	0.901	0.884	0.909	0.897
	MPFN	0.838	0.857	0.894	0.889	0.873	0.863	0.876
	Doubao 1.6	0.770	0.792	0.763	0.777	0.747	0.778	0.762
	Qwen3	0.639	0.839	0.357	0.501	0.583	0.929	0.717
	NLIN	0.922	0.905	0.941	0.923	**0.940**	0.903	0.921
	本文方法	**0.945**	**0.956**	0.928	**0.942**	0.931	**0.958**	**0.944**
CFND	BERT	0.828	0.812	0.721	0.764	0.837	0.895	0.865
	ViT	0.782	0.721	0.709	0.715	0.819	0.828	0.824
	UniSMMC	0.858	**0.872**	0.740	0.801	0.851	**0.932**	0.890
	MIAN	0.859	0.799	0.780	0.829	0.833	0.823	0.847
	GAT	0.823	0.764	0.781	0.773	0.861	0.849	0.855
	CMA-VRP	0.866	0.795	0.776	0.785	0.861	0.868	0.863
	HSA-DEF	0.839	0.840	0.820	0.830	0.850	0.830	0.840
	MCAN	0.845	0.774	0.845	0.808	0.897	0.845	0.870
	MPFN	0.824	0.854	0.769	0.809	0.802	0.875	0.837
	Doubao 1.6	0.779	0.738	0.868	0.798	0.839	0.691	0.758
	Qwen3	0.675	0.731	0.553	0.630	0.641	0.796	0.710
	NLIN	0.874	0.813	0.873	0.842	0.917	0.874	0.895
	本文方法	**0.876**	0.816	**0.891**	**0.845**	**0.918**	0.891	**0.897**

注:最好的结果用粗体表示,次好的结果用下划线表示.

用注意力机制取得了良好效果,但MS-GAT通过构建异构图实现了更深层的跨模态交互. 此外,与同类图方法CMA-VRP相比,MS-GAT引入的相似度感知机制与模态枢纽节点能更精准地衡量模态间的语义一致性,增强处理异构关系时的鲁棒性.

与多通道特征基线的对比进一步验证了特征提取策略的有效性. 实验结果显示,侧重于情感特征的HSA-DEF或侧重于频域特征的MCAN与MPFN均未达到最优性能. 这表明单一维度的深层特征挖掘存在局限性:HSA-DEF忽略了图像物理痕迹,而MCAN与MPFN未能充分利用文本情感倾向. MS-GAT通过整合图像频域信息与文本情感特征,弥补了单一视角的不足,从而能够识别更多类型的虚假新闻.

最后,与大语言模型基线的对比验证了该方法的优势. 通用大模型在零样本设置下表现受限,难以捕捉细微的伪造痕迹. 与经过微调的NLIN相比,MS-GAT在多个指标上仍具备优势. 这可能是由于NLIN将多模态信息统一映射至文本空间,导致图像细节与文本情感特征在转换过程中发生损耗;而MS-GAT保留了多通道原始特征并在图中进行显式交互,避免了过度简化带来的信息缺失.

5.4　消融分析

本节通过在Weibo17与CFND数据集上进行消融实验,评估MS-GAT框架中关键模块的贡献. 本研究以完整的MS-GAT模型为基准,通过构建6种变体来考察图结构推理、多通道特征、融合机制及优化策略对性能的影响.

本节设计了以下6种消融变体进行对比分析.

(1)w/o Graph(移除图模块):该变体移除了图构建与GAT模块,直接将融合后的多模态特征输入全连接层进行分类.

(2)w/o Multi-channel(移除多通道特征):该变体移除了文本情感特征和图像频域特征,仅保留基础的文本语义和图像空间域特征.

(3)w/ Concat(移除门控融合机制):该变体将单模态内部的多通道特征融合方式从门控机制替换为简单的特征拼接.

(4)w/ GMU(替换为多模态门控单元):该变体将门控机制替换为经典的多模态门控单元(Gated Multimodal Unit,GMU)[58],以对比不同门控机制对噪声特征的过滤能力.

(5)w/ Late Fusion(决策层融合):该变体移除了所有跨模态交互模块,独立利用文本和视觉分支进行预测,并将输出概率的平均值作为最终判定结果.

(6)w/o Adaptive Loss(移除自适应加权损失):该变体移除了基于可学习参数的损失加权机制,采用固定的平均权重(1∶1)来组合图损失、文本损失与图像损失.

表3展示了MS-GAT及其6种消融变体在Weibo17与CFND数据集上的性能表现.实验结果表明,移除或修改框架中的关键模块均会导致不同程度的性能下降,这验证了多模态异构图结构、多通道深层特征、门控融合机制及自适应优化策略在虚假新闻检测任务中的必要性.

表3 消融实验结果

数据集	方法	准确率	虚假新闻			真实新闻		
			精确率	召回率	F1分数	精确率	召回率	F1分数
Weibo17	本文方法	**0.945**	**0.956**	**0.928**	**0.942**	**0.931**	**0.958**	**0.944**
	w/o Graph	0.918	0.925	0.903	0.914	0.908	0.932	0.920
	w/o Multi-channel	0.912	0.918	0.898	0.908	0.902	0.925	0.913
	w/ Concat	0.906	0.910	0.892	0.901	0.896	0.919	0.907
	w/ GMU	0.939	0.948	0.924	0.936	0.927	0.954	0.940
	w/ Late Fusion	0.868	0.871	0.855	0.863	0.860	0.879	0.869
	w/o Adaptive Loss	0.925	0.932	0.911	0.921	0.915	0.938	0.926
CFND	本文方法	**0.876**	**0.816**	**0.891**	**0.845**	**0.918**	**0.891**	**0.897**
	w/o Graph	0.837	0.760	0.842	0.800	0.894	0.834	0.863
	w/o Multi-channel	0.843	0.769	0.847	0.806	0.898	0.840	0.868
	w/ Concat	0.840	0.779	0.817	0.797	0.882	0.854	0.868
	w/ GMU	0.860	0.795	0.857	0.825	0.906	0.862	0.883
	w/ Late Fusion	0.830	0.806	0.735	0.769	0.843	0.889	0.865
	w/o Adaptive Loss	0.843	0.782	0.821	0.801	0.884	0.857	0.870

注:最好的结果用粗体表示,次好的结果用下划线表示.

其中,w/ Late Fusion变体的性能下降最为明显,证明了简单的决策层集成难以捕捉细粒度的跨模态语义冲突,而通过异构图实现的中间层深度融合对识别伪造线索至关重要.移除图模块导致CFND数据集准确率由0.876降至0.837,说明仅依靠特征提取并不充分,利用GAT建立实体的显式逻辑关联能有效放大不一致性信号.移除多通道特征导致两个数据集的准确率均下降约3.3%,证实了频域与情感特征在物理及心理维度上能够弥补基础语义特征的盲区.在内部融合策略对比中,w/ Concat变体由于直接拼接异质特征引入了噪声干扰,表现最差;而相比结构复杂的GMU,本文设计的门控机制通过减少冗余参数降低了过拟合风险,提升了模型泛化能力.最后,移除自适应加权损失导致性能下降,验证了动态调整多任务分支贡献在优化模型收敛与平衡各项任务中的优势.

5.5 T-SNE可视化

为分析模型学习到的决策空间,本研究采用t-分布随机邻域嵌入(t-distributed Stochastic Neighbor Embedding,t-SNE)[59]非线性降维技术,该技术将模型在测试集上输出的高维对数几率映射至二维空间进行可视化.图4展示了Weibo17与CFND数据集测试样本输出Logits的可视化结果.

由图4可见,尽管数据来源于两个不同的数据集,经过MS-GAT模型处理后的样本输出对数几率在低维空间就实现了对虚假新闻与真实新闻样本的区分.这种映射过程形成了相对集中的簇结构,该分布情况表明模型学习到了具备判别性的表征特征.同时,通过分析不同数据集样本在簇内的分布情况,可以初步评估模型的判别模式在跨数据集场景下的泛化能力.

5.6 参数分析

本节旨在探究MS-GAT模型中关键超参数对性能的影响,以确定参数配置并解析其调控机制.本研究在Weibo17数据集的验证集上采用控制变量法进行实验,评估不同参数取值下准确率及F1分数的变化趋势.图5展示了各项参数对模型性能的影响.

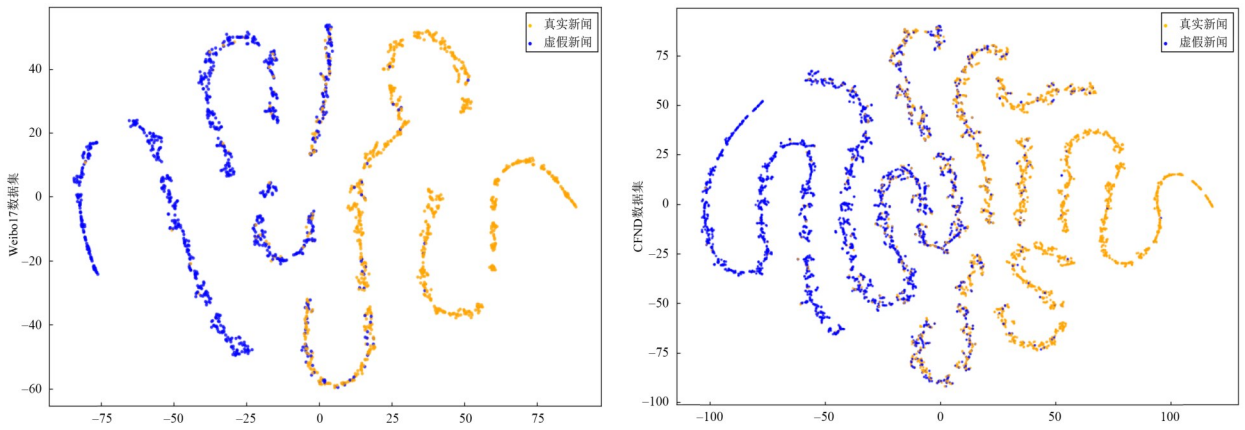


图4　t-SNE可视化结果

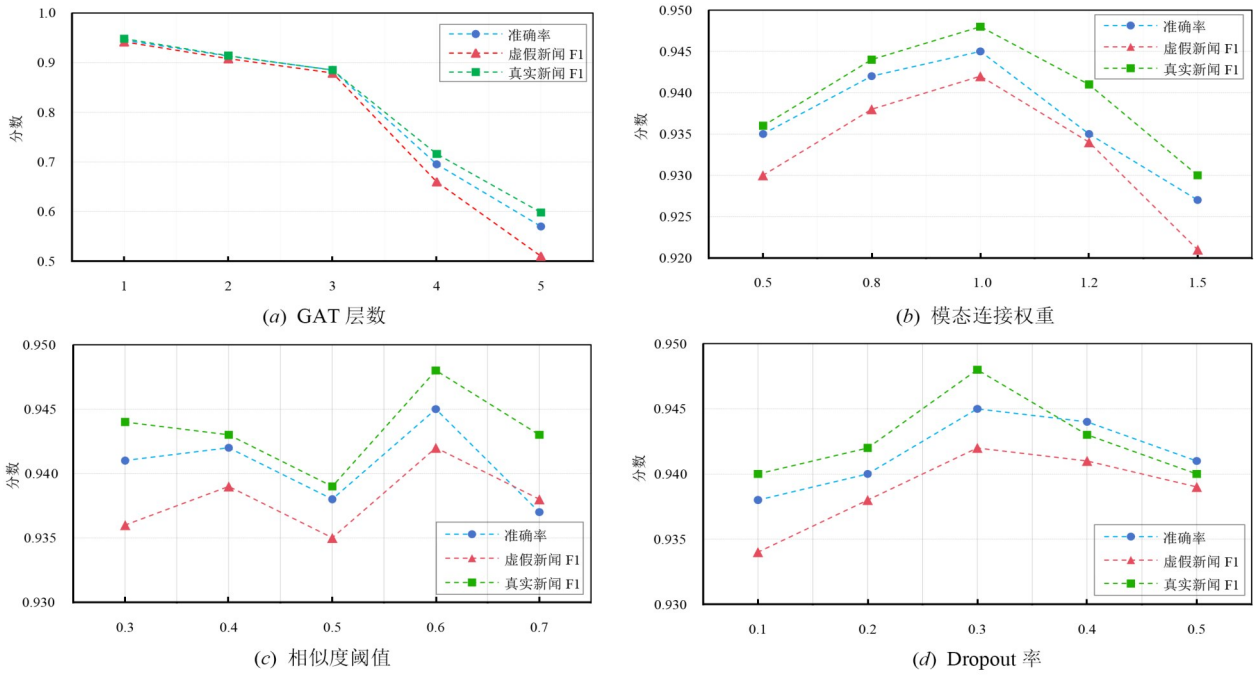


图5　MS-GAT模型在Weibo17验证集上的参数分析

5.6.1　GAT层数

首先,本研究考察了GAT层数对性能的影响,结果如图5(*a*)所示. 当GAT层数为1时,模型性能达到最优. 随着网络层数增加,性能呈现下降趋势;当层数达到4层与5层时,准确率分别降至0.695与0.570. 这种现象主要归因于GNN中的过平滑问题:随着层数增加,节点表征在多次聚合后趋于一致,导致特征区分度降低. 此外,过深的网络增加了参数量,提升了过拟合风险. 因此,本研究将GAT层数设定为1.

5.6.2　模态连接权重

其次,本研究分析了模态连接权重w_{m}的影响,该参数决定了实体节点与模态枢纽节点间的信息交互强度. 由图5(*b*)可见,随着权重从0.5增加到1.0,模型性能稳步提升并在w_{m}取1时达到最佳;当权重继续增至1.5时,各项指标均出现下降. 这表明权重较低时模态间交互不足,难以利用互补信息校正偏差;而权重过高则会导致模态噪声过度传播,干扰真实性判断. 因此,本研究将w_{m}设定为1.

5.6.3　相似度阈值

此外,本研究探讨了相似度阈值θ_{t}、θ_{i}、θ_{ti}对模型的影响,该阈值用于过滤低置信度的边,决定了图结构的密度. 由图5(*c*)可见,随着阈值从0.3增加至0.6,模型性能呈波动上升趋势并在0.6时达到最优. 当阈值较低时,稠密的图结构引入了无关实体的语义噪声;而阈值过高则导致图结构过度稀疏,阻碍了关键信息的跨模态传递. 因此,本研究将相似度阈值统一设定为0.6.

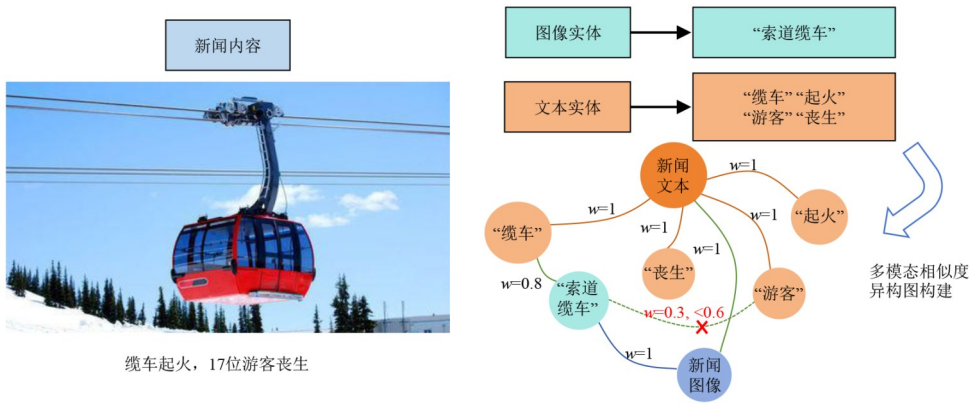
5.6.4 Dropout率

最后，本研究考察了Dropout率对模型泛化能力的影响，由图5(d)可见，随着Dropout率从0.1增加到0.3，模型性能逐步提升并在0.3时达到最优；当比例继续增加时，性能出现下滑．该结果说明适度的神经元丢弃能有效抑制过拟合，但过高的比例会导致信息丢失．因此，本研究将该参数设置为0.3.

综合以上参数分析结果，本文确定了MS-GAT模型在Weibo17数据集上的最优超参数配置(GAT层数为1，模态连接权重为1，相似度阈值为0.6，Dropout率为0.3)，并在最终的对比实验和消融分析中统一采用了这组参数．

5.7 案例分析

本节通过具体案例展示MS-GAT框架的工作原理与优势，重点分析其利用图结构识别图文冲突的能力．本研究从CFND数据集中选取一个典型样本进行深入分析．图6展示了该虚假新闻案例的内容及其对应的异构图结构示意图．



注：节点(以实体名称标注)代表该实体经过多通道特征提取和CLIP编码后的特征向量．

图6 多模态虚假新闻案例分析示意图

如图6所示，该样本的文本内容为“缆车起火，17位游客丧生”，属于具有煽动性的描述；然而，配图仅展示了正常的索道缆车，并未出现火灾或事故迹象．此类案例对单模态检测方法构成了挑战，因为文本在语义上具有误导性，而图像在单一模态分析下表现正常．该样本的真实标签为虚假新闻．

MS-GAT模型准确地将该样本判定为虚假新闻．在预处理阶段，模型首先识别出文本实体(如“缆车”“起火”“丧生”)与图像实体(如“索道缆车”主体)．在图构建阶段，模型利用CLIP模型对实体特征进行编码，并计算跨模态实体间的余弦相似度．若相似度低于阈值0.6，则不建立对应的关联边．在本案例中，描述灾难的文本实体(“起火”“丧生”)与图像实体(“索道缆车”)间的语义相似度较低，导致跨模态关联缺失，从而产生不一致性信号．在GAT处理阶段，模型通过聚合邻居节点信息放大并融合该不一致性信号，使得多模态图检测分支能够准确判定该案例为虚假内容．

该案例验证了MS-GAT模型在捕捉细粒度图文不一致性方面的能力．通过构建显式的跨模态实体关联图并引入相似度感知机制，模型能够超越表层语义信息，分析跨模态语义的一致性，从而识别此类典型的图文不符虚假新闻．

6 结论

多模态虚假新闻在社交平台上的快速传播对信息环境治理构成了严峻挑战．针对现有方法在跨模态语义关联建模与多模态信息融合方面的局限，本文提出一种基于多通道特征增强与图文相似度感知的图注意力网络MS-GAT．该模型通过构建显式建模图文细粒度语义关联的异构图，利用多通道特征提取模块捕获图文深度特征，并引入相似度感知的图注意力机制实现跨模态信息的动态融合，提升了检测性能与模型判别结果的可解释性．在Weibo17与CFND两个公开数据集上的实验结果表明，MS-GAT在准确率、F1分数等指标上优于单模态模型、传统多模态融合方法及近期先进模型．消融实验与可视化分析验证了模型关键组件的有效性及其在跨数据集场景下的泛化能力．

本文通过图结构建模为多模态虚假新闻检测提供了新的研究视角，其显式的跨模态关联建模有助于揭示图文内容间的语义一致性与矛盾性．然而，当前方法仍受限于实体识别精度对图构建质量的影响，以及在大规模数据场景下的计算效率问题．未来的研究工作将重点探索更为鲁棒的跨模态实体链接方法，优化图构建与处理技术以提升模型的扩展性，并尝试将本框架应用于其他多模态内容分析任务，为社交平台信息治理提供技术支撑．

参考文献

[1] LAZER D M J, BAUM M A, BENKLER Y, et al. The science of fake news[J]. Science, 2018, 359(6380): 1094-1096.

[2] HU L M, YANG T C, ZHANG L H, et al. Compare to the knowledge: Graph neural fake news detection with external knowledge[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Kerrville: Association for Computational Linguistics, 2021: 754-763.

[3] GUO B, DING Y S, YAO L N, et al. The future of false information detection on social media: New perspectives and trends[J]. ACM Computing Surveys (CSUR), 2021, 53(4): 1-36.

[4] LUO D L, LIU Y L, YANG R, et al. Toward real text manipulation detection: New dataset and new solution[J]. Pattern Recognition, 2025, 157: 110828.

[5] MRIDHA M F, KEYA A J, HAMID M A, et al. A comprehensive review on fake news detection with deep learning[J]. IEEE Access, 2021, 9: 156151-156170.

[6] ESSA E, OMAR K, ALQAHTANI A. Fake news detection based on a hybrid BERT and LightGBM models[J]. Complex & Intelligent Systems, 2023, 9(6): 6581-6592.

[7] BALSHETWAR S V, RS A, R D J. Fake news detection in social media based on sentiment analysis using classifier techniques[J]. Multimedia Tools and Applications, 2023, 82(23): 35781-35811.

[8] STEINEBACH M, LIU H J, GOTKOWSKI K. Fake news detection by image montage recognition[J]. Journal of Cyber Security and Mobility, 2020, 9(2): 175-202.

[9] ZHOU P, HAN X T, MORARIU V I, et al. Learning rich features for image manipulation detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 1053-1061.

[10] ZHANG T, WANG D, CHEN H H, et al. BDANN: BERT-based domain adaptation neural network for multimodal fake news detection[C]//2020 International Joint Conference on Neural Networks. Piscataway: IEEE, 2020: 9206973.

[11] WANG L Z, ZHANG C, XU H B, et al. Cross-modal contrastive learning for multimodal fake news detection[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM, 2023: 5696-5704.

[12] FU L F, PENG H X, MA C J, et al. Fake news detection based on text-modal dominance and fusing multiple multi-model clues[J]. Computers, Materials & Continua, 2024, 78(3): 4399-4416.

[13] WU L W, LIU P S, ZHAO Y Q, et al. Human cognition-based consistency inference networks for multi-modal fake news detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(1): 211-225.

[14] DUC TUAN N M, QUANG NHAT MINH P. Multimodal fusion with BERT and attention mechanism for fake news detection[C]//2021 RIVF International Conference on Computing and Communication Technologies. Piscataway: IEEE, 2021: 9642125.

[15] WU Y, ZHAN P W, ZHANG Y J, et al. Multimodal fusion with co-attention networks for fake news detection[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Stroudsburg: ACL, 2021: 2560-2569.

[16] CUI L M, WANG S H, LEE D. SAME: Sentiment-aware multi-modal embedding for detecting fake news[C]//2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Piscataway: IEEE, 2020: 41-48.

[17] LAO A, ZHANG Q, SHI C Y, et al. Frequency spectrum is more effective for multimodal representation and fusion: A multimodal spectrum rumor detector[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(16): 18426-18434.

[18] 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, 2017: 795-816.

[19] ZHANG Q,LIU J,ZHANG F,et al. Natural language-centered inference network for multi-modal fake news detection[C]//Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24. California: IJCAI, 2024: 2542-2550.

[20] ZHANG X C, GHORBANI A A. An overview of online fake news: Characterization, detection, and discussion[J]. Information Processing & Management, 2020, 57(2): 102025.

[21] ASHISH, SONIA, ARORA M, et al. An analysis and identification of fake news using machine learning techniques[C]//2024 11th International Conference on Computing for Sustainable Global Development. Piscataway: IEEE, 2024: 634-638.

[22] TIAN Z Y, BASKIYAR S. Fake news detection using machine learning with feature selection[C]//2021 6th In-

ternational Conference on Computing, Communication and Security. Piscataway: IEEE, 2021: 9776346.

[23] 蒋凌云, 鞠金恒, 徐佳, 等. 一种基于改进CRNN的轻量化乐谱识别方法[J]. 电子学报, 2023, 51(11): 3167-3175.
JIANG L Y, JU J H, XU J, et al. A lightweight music recognition method based on improved CRNN[J]. Acta Electronica Sinica, 2023, 51(11): 3167-3175. (in Chinese)

[24] 苏兆品, 张羚, 张国富, 等. 基于多特征融合和BiLSTM的语音隐写检测算法[J]. 电子学报, 2023, 51(5): 1300-1309.
SU Z P, ZHANG L, ZHANG G F, et al. A speech steganalysis algorithm based on multi-feature fusion and BiLSTM[J]. Acta Electronica Sinica, 2023, 51(5): 1300-1309. (in Chinese)

[25] BAHAD P, SAXENA P, KAMAL R. Fake news detection using bi-directional LSTM-recurrent neural network[J]. Procedia Computer Science, 2019, 165: 74-82.

[26] CHANG Q, LI X, DUAN Z. Graph global attention network with memory: A deep learning approach for fake news detection[J]. Neural Networks, 2024, 172: 106115.

[27] HUANG Y, LIN J Y, ZHOU C, et al. Modality competition: What makes joint training of multi-modal network fail in deep learning? (provably)[EB/OL]. (2022-03-23)[2025-10-10]. https://arXiv.org/abs/2203.12221.

[28] VARALAKSHMI K, ASHOK KUMAR P M. A late fusion framework using whale optimization technique and attention-BiLSTM for fake news detection[J]. International Journal of Data Science and Analytics, 2024, 18(3): 275-294.

[29] NASIR S, WASIM M, REHMAN A, et al. FACT-CLIP: Fake news detection via CLIP-based cross-modal attention and transformer fusion[C]//2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication. Piscataway: IEEE, 2025: 11070224.

[30] LIU A X, FENG B, XUE B, et al. DeepSeek-V3 technical report[EB/OL]. (2025-02-18)[2025-10-10]. https://arXiv.org/abs/2412.19437.

[31] YANG A, LI A F, YANG B S, et al. Qwen3 technical report[EB/OL]. (2025-05-14)[2025-10-10]. https://arXiv.org/abs/2505.09388.

[32] XU P, SHAO W Q, ZHANG K P, et al. LVLM-EHub: A comprehensive evaluation benchmark for large vision-language models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(3): 1877-1893.

[33] CHOI Y, UH Y, YOO J, et al. StarGAN v2: Diverse image synthesis for multiple domains[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 8185-8194.

[34] TRAN N T, TRAN V H, NGUYEN N B, et al. Self-supervised GAN: Analysis and improvement with multi-class minimax game[EB/OL]. (2020-01-08)[2025-10-10]. https://arXiv.org/abs/1911.06997.

[35] FRANK J, EISENHOFER T, SCHÖNHERR L, et al. Leveraging frequency analysis for deep fake image recognition[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 3247-3258.

[36] JING J, WU H C, SUN J, et al. Multimodal fake news detection via progressive fusion networks[J]. Information Processing & Management, 2023, 60(1): 103120.

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

[38] PASCHEN J. Investigating the emotional appeal of fake news using artificial intelligence and human contributions[J]. Journal of Product & Brand Management, 2019, 29(2): 223-233.

[39] AJAO O, BHOWMIK D, ZARGARI S. Sentiment aware fake news detection on online social networks[C]//ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2019: 2507-2511.

[40] ZHANG X Y, CAO J, LI X R, et al. Mining dual emotion for fake news detection[C]//Proceedings of the Web Conference 2021. New York: ACM, 2021: 3465-3476.

[41] GHANEM B, PONZETTO S P, ROSSO P, et al. FakeFlow: Fake news detection by modeling the flow of affective information[EB/OL]. (2021-01-24)[2025-10-10]. https://arXiv.org/abs/2101.09810.

[42] WAN M Y, ZHONG Y, GAO X F, et al. Fake news, real emotions: Emotion analysis of COVID-19 infodemic in weibo[J]. IEEE Transactions on Affective Computing, 2024, 15(3): 815-827.

[43] GIACHANOU A, ROSSO P, CRESTANI F. The impact of emotional signals on credibility assessment[J]. Journal of the Association for Information Science and Technology, 2021, 72(9): 1117-1132.

[44] LI P G, SUN X, YU H F, et al. Entity-oriented multi-modal alignment and fusion network for fake news detection[J]. IEEE Transactions on Multimedia, 2022, 24: 3455-3468.

[45] FU L F, PENG H X, LIU S. KG-MFEND: An efficient

knowledge graph-based model for multi-domain fake news detection[J]. The Journal of Supercomputing, 2023, 79(16): 18417-18444.

[46] MA Z H, LUO M N, GUO H, et al. Event-radar: Event-driven multi-view learning for multimodal fake news detection[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2024: 5809-5821.

[47] QIAN S S, HU J, FANG Q, et al. Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2021, 17(3): 1-23.

[48] QI P, CAO J, LI X R, et al. Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 1212-1220.

[49] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2009, 20(1): 61-80.

[50] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.

[51] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[EB/OL]. (2021-02-26)[2025-10-10]. https://arXiv.org/abs/2103.00020.

[52] 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: Human Language Technologies, Volume 1 (Long and Short Papers). Kerrville: Association for Computational Linguistics, 2019: 4171-4186.

[53] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03) [2025-10-10]. https://arXiv.org/abs/2010.11929.

[54] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2018-02-04) [2025-10-10]. https://arXiv.org/abs/1710.10903.

[55] ZOU H Q, SHEN M, CHEN C, et al. UniS-MMC: Multimodal classification via unimodality-supervised multimodal contrastive learning[EB/OL]. (2023-05-16) [2025-10-10]. https://arXiv.org/abs/2305.09299.

[56] ZHANG T L, YU E, SHAO Y, et al. Multimodal inverse attention network with intrinsic discriminant feature exploitation for fake news detection[EB/OL]. (2025-05-29) [2025-10-10]. https://arXiv.org/abs/2502.01699.

[57] LI G Y, HU D, FU X M, et al. Entity graph alignment and visual reasoning for multimodal fake news detection[C]//Proceedings of the 33rd ACM International Conference on Multimedia. New York: ACM, 2025: 2486-2495.

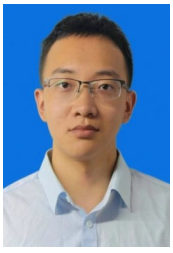
[58] AREVALO J, SOLORIO T, MONTES-Y-GÓMEZ M, et al. Gated multimodal units for information fusion[EB/OL]. (2017-02-07)[2025-10-10]. https://arXiv.org/abs/1702.01992.

[59] MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9: 2579-2605.

作者简介

张仕斌　男,1970年2月出生于重庆市丰都县. 现为成都信息工程大学人工智能学院教授、博士生导师. 主要研究方向为网络与信息安全、人工智能安全.
E-mail: cuitzsb@cuit.edu.cn

杨　敏　女,1994年8月出生于四川省达州市. 现为成都信息工程大学网络空间安全学院讲师. 主要研究方向为数据安全与隐私保护.
E-mail: youngm24@cuit.edu.cn



蔡松睿　男,2001年10月出生于山西省运城市. 现为成都信息工程大学网络空间安全学院硕士研究生. 主要研究方向为虚假新闻检测.
E-mail: cuitcsr@163.com

陈世航　男,2005年8月出生于河南省焦作市. 现为成都信息工程大学网络空间安全学院本科生. 主要研究方向为虚假信息检测.
E-mail: 2541367758@qq.com