双分支线索深度感知与自适应协同优化的多模态虚假新闻检测

钟善男[1,2] 彭淑娟[1,3] 柳 欣[1,2] 王楠楠[4] 李太豪[2]

[1](华侨大学计算机科学与技术学院 福建 厦门 361021)

[2](之江实验室 杭州 311121)

[3](华侨大学福建省大数据智能与安全重点实验室 福建 厦门 361021)

[4](西安电子科技大学空天地一体化综合业务网全国重点实验室 西安 710126)

摘 要 深度学习方法促使多模态虚假新闻检测领域快速发展,现有的检测模型通常从全局角度学习新闻图文间的跨模态语义关联,并利用共享语义内容获取检测的关键信息.然而,新闻内部的局部语义差异可能会限制模型有效利用跨模态语义关联的能力,其中潜在的非共享语义内容作为重要线索能够有效揭示虚假新闻的篡改意图和目的.为了解决上述问题,本文提出了一种双分支线索深度感知与自适应协同优化的多模态虚假新闻检测模型.该模型首先从图像显著区域和文本语义单词中提取细粒度的新闻特征,并使用跨模态加权残差网络从中学习共享语义线索.同时,根据所有图像区域和文本单词之间的语义相关性,双分支图文线索感知模块显式地建模共享与非共享语义内容的语义关联.其中,线索关联优化分支对两类语义内容的关联边界持续迭代优化,促使模型准确区分非共享语义线索;线索关联分析分支刻画两类语义内容的可信程度,并在此基础上引导模型实现线索的自主融合.通过上述自适应协同优化框架,本文提出的模型能够在复杂新闻语境下进行线索的深度感知与融合,实现更准确、更可解释的多模态虚假新闻检测.在广泛使用的中英文真实数据集上的实验结果表明,本文提出的模型明显优于基线方法,在准确率和虚假新闻检测精确率上分别平均提高了4.85%和4.50%.

关键词 多模态虚假新闻检测;局部语义差异;跨模态语义关联;非共享语义线索;自适应协同优化

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2023.02612

Multimodal Fake News Detection via Two-Branch Deep Clue Perception and Adaptive Collaborative Optimization

ZHONG Shan-Nan[1,2] PENG Shu-Juan[1,3] LIU Xin[1,2] WANG Nan-Nan[4] LI Tai-Hao[2]

[1](*College of Computer Science and Technology, Huaqiao University, Xiamen, Fujian* 361021)

[2](*Zhejiang Lab Hangzhou* 311121)

[3](*Fujian Key Laboratory of Big Data Intelligence and Security, Huaqiao University, Xiamen, Fujian* 361021)

[4](*State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an* 710126)

Abstract Deep learning methods are able to learn the high-level semantic features and significantly promote multimodal fake news detection performances. In the literature, existing multimodal fake news detection models usually learn the cross-modal semantic correlations across news image and text from a global perspective, and utilize their shared information to infer the key clues for detection. Although such approaches are able to detect the obvious multimodal fake

收稿日期:2023-01-16;在线发布日期:2023-08-14.本课题得到之江实验室开放课题(2021KH0AB01)、国家自然科学基金联合基金重点项目(U22A2096)、福建省自然科学基金项目(2020J01083,2020J01084)资助.**钟善男**,硕士研究生,主要研究领域为多媒体数据挖掘与多模态理解.E-mail:snzhong@stu.hqu.edu.cn. **彭淑娟**(通信作者),博士,副教授,中国计算机学会(CCF)会员,主要研究领域为模式识别、机器学习与多媒体分析.E-mail:pshujuan@hqu.edu.cn. **柳 欣**,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为人工智能与多媒体分析. **王楠楠**,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为计算机视觉与机器学习. **李太豪**,博士,研究员,主要研究领域智能计算与情绪识别.

news，these global modeling methods cannot well differentiate the local semantic differences within the news and therefore may degrade the detection performance. Indeed，the unshared semantic content serves as an important clue that can directly reveal their tampering intentions and purposes. Inspired by these findings，this paper proposes a multimodal fake news detection model based on two-branch deep clue perception and adaptive collaborative optimization，which can well mine the non-shared semantic clues for efficient detections. Specifically，the model first extracts fine-grained features from image salient regions and text semantic words to capture the semantic news content. Then，the heterogeneous news features are semantically aligned using a cross-modal weighted residual network，featuring on learning the shared semantic clues. For deep inference of non-shared semantic clues，the model designs an adaptive two-branch clue perception strategy to learn the cross-modal semantic correlations within the multimodal news. Specifically，the model automatically constructs an image-text clue correlation matrix based on all image regions and text words. Accordingly，a two-branch clue perception module is explicitly designed to model the probability distributions for shared and non-shared semantic correlations of all content in the matrix. On the one hand，the clue correlation optimization branch embeds an optimization algorithm to continuously update the semantic correlation boundaries of the different semantic correlation distributions iteratively，whereby the non-shared semantic clues can be well differentiated from the clue correlation matrix. On the other hand，the clue correlation analysis branch portrays the credibility of the different news content and guides the model to achieve automatic multimodal feature fusion on the basis of the correlation scores. Finally，the learning framework can well generates the non-shared semantic clue features by aggregating the optimized image-text clues. With the above adaptive co-optimization framework，the proposed model is capable of mining the deep clues to achieve more accurate and interpretable multimodal fake news detection. Experimental results on public English and Chinese datasets show that the proposed detection model significantly outperforms the baseline methods，with an average improvement of 4.85% and 4.50%，respectively，in accuracy and fake news detection precision.

Keywords multimodal fake news detection；local semantic difference；cross-modal semantic correlation；non-shared semantic clues；adaptive collaborative optimization

1 引 言

随着互联网时代的深入发展，图文并茂的多模态新闻逐渐成为社交媒体中主流的信息形式，从而虚假新闻的形态也从单一文本扩展到图像、音频和视频等多模态领域，这对虚假新闻检测提出了新的挑战[1]. 作为中国最大的全媒体内容形式的社交媒体平台，微博在2022年度共有效处置不实信息82274条，辟除新增谣言及引导争议事件1355例，标记不实信息9286条. 一些社会热点事件经过虚假新闻的包装，迅速在社交媒体平台中演化发酵，持续激荡的网络舆情对社会稳定和经济发展都有不利影响. 在新冠肺炎疫情期间，诸如“连花清瘟能预防新冠感染”、“饮高度酒可以对抗新冠病毒”等网络谣言层出不穷. 这类虚假新闻导致相关商品一度脱销，不仅误导群众，更扰乱正常的市场秩序. 在网络舆情日趋复杂的背景下，对社交媒体平台中的多模态虚假新闻进行检测具有重要意义.

为了实现多模态虚假新闻的自动化检测，以卷积神经网络[2]和循环神经网络[3]为代表的深度学习方法被提出，这类方法[4-5]一般使用预训练模型分别提取多模态新闻的文本特征与视觉特征，然后通过拼接或注意力机制增强的方式进行融合得到多模态分类特征. 通过最大化跨模态语义相关性，这种检测模式激励模型学习新闻图文间的共享语义内容，有助于检测通用的虚假新闻类型. 然而，新闻内部的局部语义差异可能会限制模型有效利用跨模态语义关联的能力，导致其难以感知隐藏于多模态新闻中的非共享语义内容. 如图1所示，虚假新闻中通常

图1 社交媒体平台中的多模态虚假新闻示例

同时包含共享与非共享的语义内容，其中阴影区域的非共享语义内容作为重要线索有效揭示了虚假新闻的篡改意图和目的. 因此，深度挖掘非共享语义线索为多模态虚假新闻检测提供了一种新的途径，有利于模型面对复杂多变的新闻语境时获得更加可靠的检测结果.

新闻图文间丰富的局部语义交互为虚假新闻检测提供了探索非共享语义线索的可能，同时也对模型如何把握跨模态语义关联提出了更高的要求. 由于异构的图文特征间存在跨模态语义鸿沟，现有工作[6-7]通常利用若干层神经网络将图像与文本的全局特征映射至一个潜在公共空间，并通过约束特征间距离的方式实现语义对齐. 这种粗粒度的跨模态语义关联学习方式难以充分建模图像区域和文本单词之间的细粒度语义交互，更可能掩盖真正的虚假线索. 同时，如何对不同的高层特征进行融合将影响多模态虚假新闻检测的精度. 现有工作[8-9]利用各类注意力机制进行多模态特征融合，从而最大限度地提升跨模态语义相关性. 这类依赖于共享语义内容的融合方法能够帮助模型初步检出部分虚假新闻，但对于存在局部语义差异的情况，往往会出现错检、漏检的现象. 面对复杂语境下的虚假新闻，非共享语义线索能够帮助模型有效捕捉多模态内容中潜在的局部语义差异，为融合过程提供关键信息. 因此，准确理解跨模态语义关联对于建模非共享语义线索起到重要的作用，但也需要认识到新闻存在局部语义差异的客观事实.

综合上述问题，本文发现多模态虚假新闻检测方法还存在一些挑战，包括3方面：(1)现有方法缺乏对新闻非共享语义线索的深度感知与挖掘，新闻内部存在的局部语义差异限制了模型精度的进一步提升. (2)面向全局的跨模态语义关联学习方式无法充分建模图像区域和文本单词之间的细粒度语义交互，这导致检测模型难以实现异质模态间有效的推理判断. (3)传统的多模态特征融合策略忽视了不同高层语义特征的可信程度. 在处理存在局部语义差异的虚假新闻类型时，依赖共享语义内容的融合方式可能带来缺乏解释性的结果. 为了满足实际的检测需求，多模态虚假新闻检测方法应当在充分建模跨模态语义关联交互的前提下，深度感知新闻内部的非共享语义线索，实现自适应协同优化的线索融合与检测.

为了解决上述挑战，本文提出了一种双分支线索深度感知与自适应协同优化的多模态虚假新闻检测方法，在深度挖掘非共享语义线索的基础上，对新闻的不同线索进行有效的融合与检测. 在特征提取阶段，本文通过特定的特征提取网络分别表征新闻的图像显著区域和文本语义单词. 接着，使用跨模态加权残差网络对输出的异构新闻特征进行聚合与对齐，并生成共享语义线索. 在线索推理阶段，本文设计了一个自适应优化的双分支图文线索感知模块，通过学习跨模态语义关联以实现图文线索的深度感知. 具体来说，模型首先根据新闻的图像区域-文本单词组合的语义相关性自动构建图文线索关联矩阵. 进一步，对其中的所有组合进行依次采样以显式地建模共享与非共享语义内容的语义关联概率分布. 针对得到的两类分布，线索关联优化分支利用优化算法自主调整两类新闻内容的最优区分边界，同时更新线索关联矩阵；线索关联分析分支通过度量两类分布间的差异来刻画两类新闻内容的可信程度，并提出一个关联分数来指导多模态特征的融合过程. 模型根据更新的线索关联矩阵得到线索注意力矩阵，并进一步聚合生成非共享语义线索. 在线索融合阶段，模型利用关联分数自适应地融合共享语义线索和非共享语义线索，最终得到新闻的多模态特征表示.

本文的主要贡献包括3个方面：

(1)提出了一种双分支线索深度感知与自适应协同优化的多模态虚假新闻检测模型，首次挖掘并利用非共享语义内容作为重要线索进行虚假新闻检测，并实现不同线索的自主融合.

(2)设计了一个自适应优化的双分支图文线索感知模块，该模块明确地驱动新闻的跨模态语义关联学习以最大程度地区分非共享语义线索，在保证挖掘精度的同时，得到更全面的语义线索.

(3)在真实世界的2个中英文新闻数据集上对本文提出的方法进行验证. 与现行的基线方法相

比，本文提出的模型能够大幅提高检测准确率以及虚假新闻检测精确率，效果得到了全面提升.

2　相关工作

虚假新闻检测技术的发展与新闻形态的演变密切相关，因此本节从基于单模态的虚假新闻检测、基于多模态融合的虚假新闻检测以及基于跨模态语义关联的虚假新闻检测三个方面介绍本文的相关工作.

2.1　基于单模态的虚假新闻检测

随着深度学习的发展，深度神经网络能够学习新闻的高层语义特征并实现自动分类. 早期的虚假新闻通常以单模态的形式在社交媒体中传播[10]，因此该阶段的研究大多利用新闻的文本、图像以及传播过程中产生的社交上下文进行检测.

相比于传统的手工特征模型[11]，Ma等人[12]首次使用循环神经网络(RNN)对文本内容进行虚假新闻检测，其模型具有更优的检测性能和可扩展性. 随后，经过大规模语料训练的预训练模型在自然语言处理领域取得巨大成功，基于变换器的双向编码器表示模型(BERT)作为文本特征提取器被广泛应用于虚假文本检测任务中[13-14]. 这类基于预训练模型的单模态检测方法依靠训练语料的优势能够初步检出虚假新闻，但由于其无法深入理解虚假新闻的内在特点，始终存在检测盲区.

在社交媒体的富媒体化趋势下，面向文本的检测方法难以满足新闻图像的检测需求，虚假新闻检测技术得到进一步扩展. Qi等人[15]明确指出社交媒体中的篡改图像在区域单位上存在明显的伪造痕迹，并在高层语义层面反映出一定的造假意图. 受启发于计算机视觉的研究，以基于视觉几何组的卷积神经网络(VGG)为代表的预训练模型被用于底层图像特征抽取，并开发新闻图像的检测框架[16-17]. 尽管图像语义特征能够在一定程度上反映新闻的伪造痕迹与造假意图[18]，但社交媒体中的图像复杂多样，这类方法在虚假新闻检测任务上的作用仍然有限. 除此之外，如评论、转发等交互数据为虚假新闻检测提供了丰富的参考信息，这类信息作为新闻的社交上下文能够用来提升模型的检测性能. 面向社交上下文的检测方法主要包括基于用户信息的方法[19]和基于传播网络的方法[20]. 然而，来源于用户交互行为的社交上下文信息庞杂且质量参差不齐，难以直接用来评估新闻内容的可信程度.

2.2　基于多模态融合的虚假新闻检测

多模态新闻的文本与图像为虚假新闻检测提供互相补充、各有侧重的丰富信息，融合多模态内容进行虚假新闻检测成为目前的研究热点之一. 常用的多模态融合检测框架是从VGG预训练模型中提取浅层视觉特征，再将其与文本特征进行简单拼接. 在这种框架下，Singhal等人[21]使用BERT预训练模型学习新闻的文本特征，在不依赖其他辅助任务的情况下实现良好的检测性能. 此类方法本质上是将图像的浅层语义特征建模成为文本的补充信息，容易使模型学习到与数据集高度相关的特征，泛化能力存在上限. 因此，一些研究考虑将辅助任务加入检测过程，指导模型充分建模多模态特征. Wang等人[22]引入事件鉴别模块作为虚假新闻检测的辅助任务，以对抗的方式学习不以事件转移的通用特征；Khattar等人[23]借鉴变分自动编码器的思想，提出双分支网络学习新闻的共享语义特征；Silva等人[24]提出两个独立的潜在空间来学习新闻的不同特征，并通过特征解构与重构增强模型的泛化能力. 这些辅助任务能够在多模态虚假新闻检测上发挥一定作用，但是其对新闻的文本与图像单独建模，没有考虑异质模态间的语义鸿沟问题，导致模型难以把握新闻内部的跨模态语义关联.

同时，一些研究提出通过多模态相互增强机制进行数据融合. Jin等人[25]利用注意力机制将图像特征融合进入文本和社交上下文的联合特征中，从而生成可信的多模态特征；Song等人[26]提出一种跨模态注意力残差机制和多通道卷积神经网络，实现文本和图像之间的双向增强建模，从而提高模型的检测性能；Qi等人[27]提出一种基于实体增强的多模态融合框架，尝试提取视觉实体来理解图像的高层语义，并借助视觉实体对多模态特征进行相互增强建模. 上述方法都利用了注意力机制来促进对新闻语义的模态间交叉理解，旨在通过最大化新闻的跨模态语义相关性以增强模型的泛化性. 然而，新闻内部的跨模态局部语义交互可能制约了模型利用跨模态语义相关性进行特征融合的能力. 此外，这类方法忽视了非共享语义线索的重要性，从而影响其面对复杂新闻语境时的检测性能.

2.3　基于跨模态语义关联的虚假新闻检测

一些研究开始意识到跨模态语义关联对于虚假新闻检测任务的重要性，尝试通过度量和约束新闻图文的语义一致性来提升多模态模型的检测性能.

Xue等人[28]认为图文不匹配的新闻更可能是虚假新闻，因此从全局相似度出发，利用新闻图文的语义相似程度进行检测；Zhou等人[29]将文本特征和视觉特征利用权重共享的方式转换到公共空间，在此基础上利用余弦相似度定义新闻的语义关联程度．这类基于全局的度量方法尝试利用相似性度量函数来确定新闻的跨模态语义关联，但在图文局部语义交互的捕捉上仍有所欠缺，难以深刻把握细粒度的跨模态语义关联．

为了实现多模态内容的细粒度建模，Shang等人[30]观察到新闻图像中通常包含多个显著对象，由此提出了一种基于目标感知的特征提取方法．该方法从图像中学习局部语义内容，有利于准确评估跨模态语义关联程度．Jin等人[31]提出捕捉单词级的细微线索能够使模型更好地反映人类的思维过程，实现对虚假新闻检测的细粒度推理．遵循人类信息处理模式，他们利用一种双通道核图网络来模拟局部内容之间的细微差异．这些方法在细粒度推理方面已有初步尝试，但对于跨模态语义交互的建模过程较为模糊，这阻止了人们更好地理解和信任检测模型，更难以引导模型实现性能优化．

因此，针对现行工作的不足，本文提出一种双分支线索深度感知与自适应协同优化的多模态虚假新闻检测方法．通过准确刻画跨模态语义关联，该方法能够细粒度地捕捉新闻图文中可疑的语义线索，进而实现有效的多模态融合与检测．

3 虚假新闻检测方法

本文提出通过双分支线索深度感知与自适应协同优化模型来解决多模态虚假新闻检测问题．如图2所示，检测模型主要由多模态特征提取网络、跨模态加权残差网络、双分支图文线索感知模块、多模态特征融合与检测模块4部分组成．

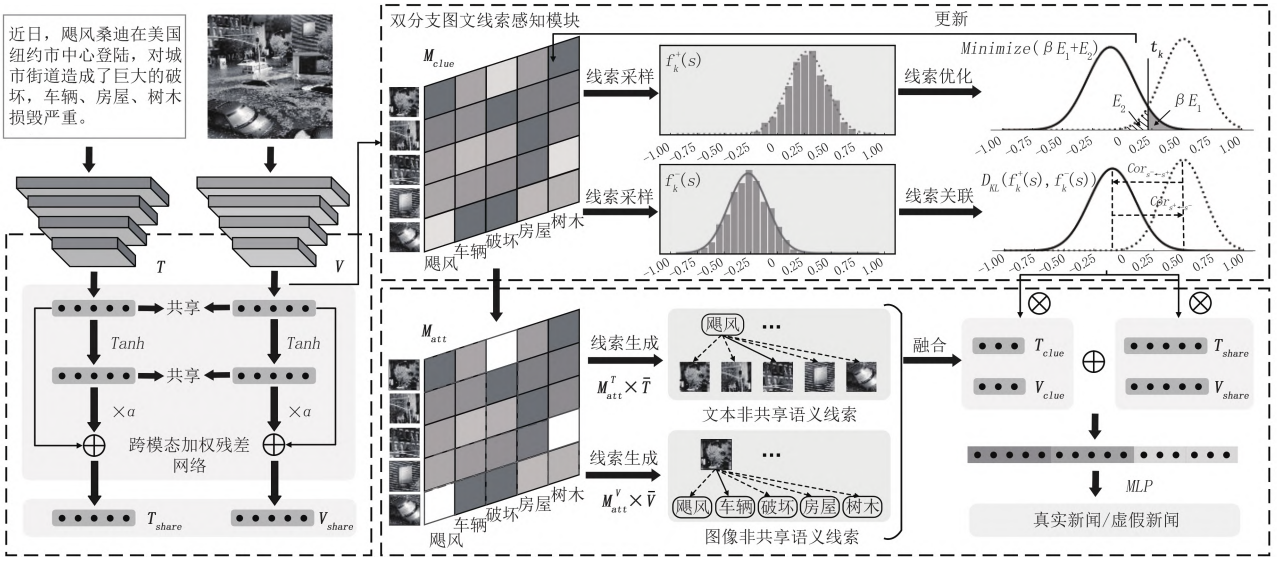


图2 双分支线索深度感知与自适应协同优化的多模态虚假新闻检测模型框架图

3.1 多模态特征提取网络

通过大规模语料训练得到的语言模型能够实现对多模态新闻内容的初步建模，提取检测任务所需的重要特征．对于新闻图片来说，目前常用VGG、深度残差神经网络(ResNet)等视觉预训练模型能够深入挖掘全局特征，但在关注图片的局部显著性特征方面表现不足，难以准确捕捉人们阅读新闻时所关注的重点区域．针对传统特征提取方法存在的局限性，本文使用自下而上的注意力模型[32](Bottom-up Attention)作为图像模态的特征提取子网络，细粒度地捕捉图像内容中的显著特征．与人们总会自发地关注醒目内容的习惯类似，Bottom-up Attention机制的优势在于能够根据任务的需求将图像表征为一组显著性的区域特征．具体来说，首先使用基于区域的卷积神经网络(Fast-RCNN)检测新闻图像，得到所有显著对象和区域．随后，从中选择前m个显著区域输入ResNet模型，得到m个图像区域的特征集合 $V=\{v_1,v_2,\cdots,v_m\}$．

为了使新闻文本与其对应图像区域之间能够实现局部语义对齐，本文将预处理后的文本词序列输入文本模态的特征提取子网络进行表征．具体来说，文本词序列中的每个语义单词都被映射为一个

固定的词嵌入向量，进一步将每个被表征的词嵌入向量通过给定数据集上的预训练词嵌入特征进行初始化操作. 为了整合上下文信息并进行特征维度对齐[14]，将所有词嵌入向量继续输入双向门控循环神经网络(Bi-GRU)，最终得到n个文本语义单词的特征集合$T=\{t_1,t_2,\cdots,t_n\}$.

3.2　跨模态加权残差网络

对于一般的检测任务而言，共享语义内容作为被广泛使用的通用线索能够在一定程度上反映多模态新闻的真实性. 挖掘新闻内部的共享语义线索不仅能够促进模型学习异构特征的通用表示，也能引导模型深入理解虚假新闻的共性特征，确保模型的泛化性能. 然而，由于不同模态的底层构成不同，多模态特征提取网络输出的图像特征和文本特征之间存在巨大的跨模态语义鸿沟[33]，这导致模型难以捕捉新闻的共享语义内容，更无法准确把握跨模态语义相关性. 为了解决这个困难，本文设计了一种跨模态加权残差网络来学习异构特征的通用表示，并生成文本与图像的共享语义线索. 该网络结合双流深度网络[34]和残差网络结构[35]的优势，在确保图像特征与文本特征完整性的同时，能够有效学习两类特征间的非线性相关性.

具体来说，跨模态加权残差网络结构主要由2层共享残差单元组成，并使用跨层跳跃连接等方式实现信息共享. 为了保持特征不变性，首先对多模态特征提取网络输出的两个特征集合进行全局平均池化，得到的两个输入特征分别用$\overline{T}$和$\overline{V}$表示. 随后，输入特征$\overline{T}$和$\overline{V}$经过2层共享残差单元，得到权重共享的中间特征Res($\overline{T}$)和Res($\overline{V}$). 该过程分别表示为：

$$\begin{aligned} Res(\overline{T}) &= d(\omega_2\sigma(\omega_1\overline{T})) \\ Res(\overline{V}) &= d(\omega_2\sigma(\omega_1\overline{V})) \end{aligned} \tag{1}$$

其中ω_1和ω_2为共享残差单元的网络参数，$\sigma(\cdot)$为双曲正切激活函数，$d(\cdot)$表示权重丢弃函数.

对于通过共享权重的相同网络结构处理后的中间特征Res($\overline{T}$)和Res($\overline{V}$)，进一步使用恒等映射的方式与输入特征$\overline{T}$和$\overline{V}$进行跳跃连接，最终得到文本与图像的共享语义线索特征T_{share}与V_{share}. 该过程分别表示为：

$$\begin{aligned} T_{share} &= \sigma(d(\overline{T})+\alpha\cdot\mathrm{Res}(\overline{T})) \\ V_{share} &= \sigma(d(\overline{V})+\alpha\cdot\mathrm{Res}(\overline{V})) \end{aligned} \tag{2}$$

其中，α是可学习的缩放参数[36].

跨模态加权残差网络结构将中间特征与输入特征进行跳跃连接，这种方式在缓解网络梯度消失的同时，又保留了输入特征的特性. 通过2层共享残差单元进行权值共享，模型在减少自身网络参数数量的同时，也有利于其捕捉新闻的共享语义线索. 基于此，跨模态加权残差网络能够帮助模型有效地学习异构图文特征的通用表示，实现对新闻内部共享语义线索的有效提取.

3.3　双分支图文线索感知模块

新闻内部的局部语义交互信息为多模态虚假新闻检测提供了大量的线索，其中潜在的非共享语义内容作为重要线索能够有效解释虚假新闻的篡改意图与目的. 为了有效区分两类新闻内容以实现对非共享语义线索的深度感知，本模块采取双分支结构以优化与分析新闻图像区域和本文单词间的跨模态语义关联.

结合图2所示，双分支图文线索感知流程如下：(1)计算输入特征T和V之间的语义关联程度s_i，构造图文线索关联矩阵M_{clue}. 随后，对M_{clue}进行采样，显式建模两类新闻内容的语义关联概率分布$f_k^+(s)$和$f_k^-(s)$. (2)线索关联优化分支利用优化算法调整区分阈值，并确定为最优关联边界t_k；同时，线索关联分析分支度量两类语义关联概率分布之间的差异，并得到两类新闻内容的关联分数Cor_S. (3)利用最优关联边界t_k得到线索注意力矩阵M_{att}^T与M_{att}^V，最终与处理后的输入特征$\overline{T}$和$\overline{V}$聚合生成非共享语义线索T_{clue}和V_{clue}.

3.3.1　分支一：线索关联优化

使用固定数值来区分两类新闻内容难以准确提取非共享语义线索，因此本文期望能够通过跨模态语义关联学习，自适应地找到最优的区分边界. 为了实现这个目的，线索关联优化分支首先根据输入特征来自动构建新闻的图文线索关联矩阵$M_{clue}=T\cdot V^{\mathrm{T}}/|T|\times|V|$，其中图像区域-文本单词组合之间的语义关联程度$s_{ij}$的计算方式为：

$$s_{ij}=\frac{t_i v_j^T}{\|t_i\|\|v_j\|},\ t_i\in T,v_j\in V \tag{3}$$

在图文线索关联矩阵中，语义关联程度s_{ij}直接反映了图文片段间的语义共享程度. 当s_{ij}越小时，图文片段间的语义共享程度越低，其被区分为非共享语义线索的可能性越大. 由于缺乏线索的先验信息，且新闻图像中存在较多重叠区域，模型从文本的角度出发，对每个文本单词对应的语义关联程度s_i依次采样. 该过程分别表示为：

$$S_k^+=\left[s_1^+,s_2^+,s_3^+,...,s_i^+,...\right],s_i^+=\max_j\left(\left\{s_{ij}\right\}_{j=1}^m\right)$$
$$S_k^-=\left[s_1^-,s_2^-,s_3^-,...,s_i^-,...\right],s_i^-=\min_j\left(\left\{s_{ij}\right\}_{j=1}^m\right) \tag{4}$$

其中，S_k^+和S_k^-表示共享语义内容的语义关联程度s_i^+和非共享语义内容的语义关联程度s_i^-的集合. 两个集合在训练中迭代更新，k为当前迭代次数.

根据集合S_k^+和S_k^-，分别建立共享语义内容和非共享语义内容关于语义关联s的概率模型$f_k^+(s)$和$f_k^-(s)$，具体形式如下：

$$f_k^+(s)=\frac{1}{\sigma_k^+\sqrt{2\pi}}e^{\left[-\frac{(s-\mu_k^+)^2}{2(\sigma_k^+)^2}\right]}$$
$$f_k^-(s)=\frac{1}{\sigma_k^-\sqrt{2\pi}}e^{\left[-\frac{(s-\mu_k^-)^2}{2(\sigma_k^-)^2}\right]} \tag{5}$$

其中，μ_k和σ_k分别是两类分布的均值和标准差.

在建模两类新闻内容的语义关联分布后，模型使用一个关联边界t_k从中准确区分出非共享语义线索. 其中，仅当$s<t_k$时，这类语义关联程度小于边界的内容能够被视作非共享语义线索. 如图3所示，其中存在两种区分错误：(1)漏分错误E_1：将实际上的非共享语义线索漏分为共享语义内容；(2)错分错误E_2：将实际上的共享语义内容错分为非共享语义线索.

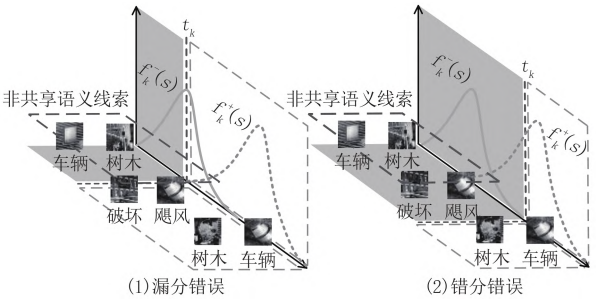


图3 错分错误与漏分错误示意图

为了最大程度地区分出非共享语义线索，模型需要对关联边界t_k进行优化，使得两类区分错误的概率最低. 因此，非共享语义线索的关联边界优化问题可以描述为：

$$\min\beta\int_{t_k}^{+\infty}f_k^-(s)\mathrm{d}s+\int_{-\infty}^{t_k}f_k^+(s)\mathrm{d}s$$
$$\text{s.t.}\quad t_k\geqslant 0 \tag{6}$$

其中，t_k是决策变量；β是惩罚参数；$t_k\geqslant 0$是关联边界的约束条件.

在约束条件下对关联边界优化问题进行一阶导数零点搜索，能够求得最优关联边界t_k：

$$t_k=\max\left(\left[\left(\left(\left(\gamma_2^k\right)^2-\gamma_1^k\gamma_3^k\right)^{\frac{1}{2}}-\gamma_2^k\right)/\gamma_1^k\right],0\right) \tag{7}$$

其中，γ_1^k、γ_2^k、γ_3^k分别表示为：

$$\gamma_1^k=\left(\sigma_k^+\right)^2-\left(\sigma_k^-\right)^2$$
$$\gamma_2^k=\mu_k^+\left(\sigma_k^-\right)^2-\mu_k^-\left(\sigma_k^+\right)^2$$
$$\gamma_3^k=\left(\mu_k^-\sigma_k^+\right)^2-\left(\mu_k^+\sigma_k^-\right)^2+2\left(\sigma_k^+\sigma_k^-\right)^2\ln\frac{\sigma_k^-}{\beta\sigma_k^+} \tag{8}$$

3.3.2 分支二：线索关联分析

通过评估语义关联分布$f_k^+(s)$和$f_k^-(s)$之间的KL散度[37]，线索关联分析分支能够得到两类新闻内容的语义关联程度. 进一步，模型提出一个关联分数来刻画不同新闻内容的可信程度，并引导两类线索特征在检测网络中实现自主融合. 当语义关联程度较低时，这表明共享语义内容的可信程度较低，检测模块应当优先考虑非共享语义线索. 反之，表明非共享语义内容的可信程度较低，检测模块应当优先考虑共享语义线索. 由于KL散度具备不对称性的特点[38]，模型在度量两类关联分布间的差异时还需要区分基准分布和估计分布.

根据先前的采样过程，两类新闻内容的语义关联概率模型$f_k^+(s)$和$f_k^-(s)$的统一表达如下：

$$f_k^+(s)=N(s^+|\mu_k^+,\sigma_k^+)$$
$$f_k^-(s)=N(s^-|\mu_k^-,\sigma_k^-) \tag{9}$$

当以$f_k^-(s)$为基准来估计$f_k^+(s)$时，模型得到共享语义内容的基准关联分数$Cor_{s^+\leftarrow s^-}$. 反之，则得到非共享语义内容的基准关联分数$Cor_{s^-\leftarrow s^+}$. 该过程分别表示为：

$$Cor_{s^+\leftarrow s^-}=D_{KL}\left(f_k^+(s)\|f_k^-(s)\right)$$
$$Cor_{s^-\leftarrow s^+}=D_{KL}\left(f_k^-(s)\|f_k^+(s)\right) \tag{10}$$

其中$D_{KL}(\cdot)$表示KL散度.

最后，利用sigmoid激活函数将两类基准关联分数的均值进行归一化处理，得到关联分数Cor_S：

$$Cor_S=sigmoid\left(\frac{1}{2}\left(Cor_{s^+\leftarrow s^-}{}^2+Cor_{s^+\leftarrow s^-}{}^2\right)\right) \tag{11}$$

随着两类语义关联概率模型的不断迭代更新，语义关联分数Cor_S能够有效刻画两类新闻内容的可信程度，并进一步指导模型在多模态特征融合过程中两类语义线索的融合与检测.

3.3.3 非共享语义线索生成

最优关联边界的学习过程以及关联分数的度量过程被整合进模型的训练中，从而创建了一个持续迭代的协同优化过程. 对于多模态虚假新闻检测任

务，该过程实现了跨模态语义关联的优化与分析，为非共享语义线索的生成提供重要的依据. 图4具体展现了图文线索关联矩阵的更新以及非共享语义线索生成的过程.

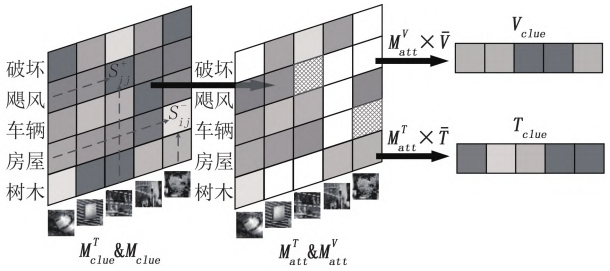


图4　非共享语义线索生成示意图

通过最优关联边界t_k对线索关联矩阵$\boldsymbol{M}_{clue}$持续更新与转换，模型能够得到线索注意力矩阵$\boldsymbol{M}_{att}$，再利用该矩阵与其对应的原始特征聚合生成非共享语义线索. 具体来说，首先计算$\boldsymbol{M}_{clue}$中的语义关联程度s_{ij}与最优关联边界t_k的差值，再从中提取非共享语义线索的语义相关性s'_{ij}：

$$s'_{ij}=(s_{ij}-t_k)\odot Mask(s_{ij}-t_k) \tag{12}$$

其中，$Mask(\cdot)$为掩码函数，当$s_{ij}-t_k$为负数时输出为1，否则为0；$\odot$表示点积运算.

更新后的线索关联矩阵重点关注非共享语义线索的跨模态语义关联，进一步通过可学习参数λ以及softmax函数将其转换为图像和文本对应的线索注意力矩阵$\boldsymbol{M}_{att}^{V}$与$\boldsymbol{M}_{att}^{T}$：

$$\begin{aligned}\boldsymbol{M}_{att}^{V}&=softmax\left(\boldsymbol{M}_{clue}/\lambda\right)\\ \boldsymbol{M}_{att}^{T}&=softmax\left(\boldsymbol{M}_{clue}^{\mathrm{T}}/\lambda\right)\end{aligned} \tag{13}$$

最后，将得到的跨模态线索注意力矩阵分别与平均池化后的输入特征$\overline{V}$和$\overline{T}$聚合，得到图像和文本最终的非共享语义线索表示$\boldsymbol{T}_{clue}$与$\boldsymbol{V}_{clue}$：

$$\begin{aligned}\boldsymbol{V}_{clue}&=\boldsymbol{M}_{att}^{V}\times\overline{V}\\ \boldsymbol{T}_{clue}&=\boldsymbol{M}_{att}^{T}\times\overline{T}\end{aligned} \tag{14}$$

3.4　多模态特征融合与检测模块

至此，模型得到了非共享语义线索$\boldsymbol{T}_{clue}$和$\boldsymbol{V}_{clue}$以及共享语义线索$\boldsymbol{T}_{share}$和$\boldsymbol{V}_{share}$，再利用关联分数Cor_S来对两类线索特征进行融合. 该过程表示为：

$$\begin{aligned}\tilde{\boldsymbol{x}}=&(Cor_s\times(\boldsymbol{T}_{clue}\oplus\boldsymbol{V}_{clue}))\\ &\oplus((1-Cor_s))\times(\boldsymbol{T}_{share}\oplus\boldsymbol{V}_{share}))\end{aligned} \tag{15}$$

其中，$\oplus$代表连接操作.

然后，将多模态融合特征$\tilde{x}$输入进一个全连接网络来获得新闻的预测标签$\tilde{y}_{cls}$：

$$\tilde{y}_{cls}=softmax(MLP(\tilde{\mathbf{x}})) \tag{16}$$

由于虚假新闻检测是一项二分类任务，模型对新闻样本的真实标签与模型获得的预测标签应用交叉熵损失函数：

$$L_{cls}=-y_{cls}\log(\tilde{y}_{cls})-(1-y_{cls})\log(1-\tilde{y}_{cls}) \tag{17}$$

4　实验与分析

4.1　数据集

本文采用2个从真实社交媒体平台收集的Weibo、Twitter数据集进行虚假新闻检测模型的性能评估，并分析模型应对不同新闻语境时的检测能力. 数据集的详细统计信息如表1所示.

表1　数据集详细统计信息

数据集	训练集		测试集	
	虚假新闻数量	真实新闻数量	虚假新闻数量	真实新闻数量
Weibo	3783	3749	620	544
Twitter	4774	3958	467	333

Weibo数据集[39]是首次由Jin等人基于中文新浪微博平台构建的虚假新闻数据集. 该数据集包含新浪微博官方谣言举报平台上从2012-05至2016-01经过验证的虚假新闻，以及从新华社的热点新闻发现系统采集的同一时期的真实新闻.

Twitter数据集[27]是首次由多媒体评估组织MediaEval Benchmarking在2015年举办的多模态信息验证任务中提出的. 该数据集主要来源于推特社交平台中从2012年至2016年所有已被官方验证的虚假新闻和真实新闻.

4.2　实验设置

本文使用准确率(Accuracy)和真、假类别的精确率(Precision)、召回率(Recall)以及F1值(F1 Score)作为模型检测性能的评估指标. 在多模态特征提取网络中，图像显著区域的维度为2048，文本词向量的维度为300. 在线索推理过程中，共享语义线索维度和非共享语义线索的维度均为64，关联边界优化问题中的惩罚参数β值为1.0. 跨模态加权残差网络中的可学习缩放参数α以及线索关联矩阵更新时所用的可学习参数λ均通过密集连接层进行学习.

在训练过程中，模型使用自适应矩估计优化器(Adam)进行损失优化，初始学习率设置为10^{-4}，丢弃率(Dorpout)设置为0.3，批大小(Mini-batch)设置为256，迭代轮次(Epoch)设置为100.

4.3 实验1:虚假新闻检测性能比较

4.3.1 对比方法

为了验证本文提出方法的有效性,本文对虚假新闻检测的各类基线方法进行性能对比.

(1)Bi-LSTM[12]:使用双向循环神经网络建模新闻文本,并利用一层全连接层进行分类.

(2)BERT[40]:使用BERT预训练模型提取新闻文本特征,并利用一层全连接层进行分类.

(3)VGG19[41]:使用VGG预训练模型提取新闻图像特征,并使用一层全连接层进行分类.

(4)Att-RNN[25]:通过长短期记忆网络(LSTM)对新闻文本进行建模,并结合跨模态注意力机制将文本特征与视觉特征相互融合.

(5)EANN[22]:使用文本卷积神经网络(TextCNN)和VGG预训练模型提取新闻的文本特征与视觉特征,将两类特征拼接后输入到新闻分类器进行分类,同时利用事件鉴别器判断事件.

(6)MVAE[23]:使用LSTM时序模型和VGG预训练模型提取新闻的文本特征与视觉特征,两者的拼接特征被编码成为多模态特征后,用于重构输入特征及虚假新闻分类.

(7)SAFE[29]:使用LSTM时序模型和VGG预训练模型提取文本特征与视觉特征,再通过计算语义相似度来表征新闻的跨模态语义关联,以实现共同学习来预测虚假新闻.

(8)Spotfake[21]:使用BERT和VGG预训练模型提取新闻的文本特征与视觉特征,对两类模态的特征进行拼接后,直接进行新闻分类.该方法不考虑任何其他的辅助任务.

(9)CARMN[26]:使用一种基于交叉注意力残差和多通道卷积神经网络的多模态虚假新闻检测框架,在同时提取原始信息和融合信息的特征表示,减轻跨模态融合特征可能产生的噪声影响.

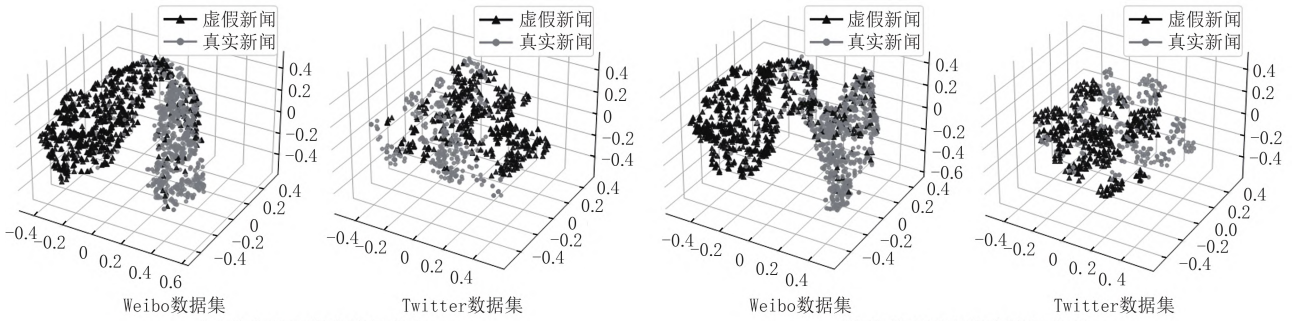
4.3.2 结果分析

表2列出了对比实验的性能比较结果,图5展现了共享语义线索与非共享语义线索在不同数据集下特征分布的t-SNE[42]三维可视化结果,通过观察与分析得到以下结论.

(1)本文提出的方法使用中英文数据集进行验证,并在准确率、精确率、召回率、F1值上显著超过其他对比方法.在准确率上,分别超出对比方法中最好结果4.3%和5.4%;在虚假新闻检测精确率上,分别超出对比方法中最好结果3.9%和5.1%.这说明本文提出的模型能够有效提升虚假新闻检测的性能,检出现行方法中遗漏、错分的虚假新闻.

表2 对比实验性能比较

数据集	方法	准确率	虚假新闻			真实新闻		
			精确率	召回率	F1值	精确率	召回率	F1值
Weibo	Bi-LSTM	0.774	0.796	0.758	0.776	0.754	0.793	0.773
	BERT	0.804	0.798	0.831	0.814	0.811	0.776	0.793
	VGG19	0.635	0.630	0.706	0.666	0.641	0.559	0.597
	Att-RNN	0.781	0.802	0.765	0.783	0.761	0.801	0.781
	EANN	0.810	0.831	0.792	0.812	0.789	0.829	0.809
	MVAE	0.824	0.854	0.769	0.809	0.802	0.875	0.837
	SAFE	0.763	0.757	0.799	0.777	0.772	0.726	0.749
	Spotfake	0.892	0.902	0.904	0.932	0.847	0.656	0.739
	CARMN	0.853	0.891	0.814	0.851	0.818	0.894	0.854
	Ours	**0.935**	**0.941**	**0.923**	**0.932**	**0.920**	**0.939**	**0.930**
Twitter	Bi-LSTM	0.551	0.553	0.679	0.609	0.548	0.413	0.471
	BERT	0.599	0.611	0.468	0.530	0.591	0.720	0.649
	VGG19	0.567	0.699	0.442	0.542	0.491	0.738	0.590
	Att-RNN	0.681	0.758	0.659	0.705	0.603	0.712	0.653
	EANN	0.678	0.765	0.641	0.698	0.597	0.729	0.657
	MVAE	0.598	0.697	0.543	0.610	0.518	0.676	0.587
	SAFE	0.643	0.676	0.506	0.579	0.625	0.772	0.691
	Spotfake	0.701	0.730	0.606	0.663	0.682	0.789	0.731
	CARMN	0.735	0.778	0.652	0.709	0.704	0.817	0.756
	Ours	**0.789**	**0.829**	**0.746**	**0.786**	**0.755**	**0.836**	**0.794**



(a)共享语义线索特征分布　　(b)非共享语义线索特征分布

图5　共享语义线索和非共享语义线索的t-SNE三维可视化结果

(2)基于文本的单模态检测方法明显优于基于图像的单模态检测方法,这说明虚假新闻检测主要依赖于文本内容,缺乏对图像语义线索的深度感知.基于多模态融合的方法优于相同子网络的单模态方法,这说明文本模态和视觉模态能够为虚假新闻检测提供互补的语义线索,提高模型检测精度.

(3)在多模态方法中,CARMN、Att-RNN的表现不够稳定,其利用的跨模态注意力增强机制未能在检测过程中得到具备良好区分性的分类特征.尽管此类方法能够一定程度地促进新闻的模态间交叉语义理解、增强跨模态共享语义相关性,但仅仅依靠共享语义内容的检测方式仍然制约了模型面对复杂新闻语境的泛化性能,导致其难以全面应对虚假新闻.此外,SAFE的检测性能不足,其使用的全局度量方法无法捕捉新闻内部的跨模态语义交互,缺乏对多模态新闻内在特点的深入理解.

(4)图5展现了共享语义线索和非共享语义线索在两个数据集下的特征分布情况,直观地说明了模型依靠两类语义线索来应对不同新闻语境的检测能力.整体来看,本文提出的模型在微博新闻语境中能够有效学习到可区分度高的语义线索,呈现出色的检测性能.其中,共享语义线索的可区分度略优于非共享语义线索.其次,模型在推特新闻语境中得到具备可区分性的非共享语义线索.尽管共享语义线索的区分度略有不足,模型利用关联分数调整两类线索后,仍能呈现出良好的检测性能.

4.4　实验2:消融分析

4.4.1　对比方法

为验证模型中每个组件的有效性,本文设计了5种模型的变体,对模型进行消融分析.

(1)去除局部显著特征.分别使用Bert模型以及VGG模型代替原先的特征提取网络.

(2)去除共享语义线索.使用独立的2层全连接层代替原先的跨模态加权残差网络.

(3)去除非共享语义线索.仅使用跨模态加权残差网络生成的共享语义线索进行检测.

(4)去除关联边界优化方法.忽略可能出现的两类区分错误,将线索的区分阈值固定为0.

(5)去除关联分数.忽略新闻不同内容的可信程度,直接拼接共享语义线索与非共享语义线索进行多模态特征融合.

4.4.2　结果分析

表3列出了移除模型各个重要模块或方法进行消融分析的实验结果,观察得到以下结论.

表3　消融实验性能比较

	方法	准确率	精确率	召回率	F1值
Weibo	去除局部特征	0.901	0.921	0.884	0.902
	去除共享线索	0.912	0.911	0.920	0.915
	去除非共享线索	0.908	0.910	0.913	0.911
	去除边界优化	0.911	0.920	0.907	0.913
	去除关联分数	0.918	0.924	0.918	0.921
	本文方法	**0.935**	**0.941**	**0.923**	**0.932**
Twitter	去除局部特征	0.734	0.767	**0.775**	0.771
	去除共享线索	0.774	0.795	0.757	0.776
	去除非共享线索	0.763	0.775	0.722	0.749
	去除边界优化	0.764	0.772	0.726	0.748
	去除关联分数	0.772	0.812	0.696	0.751
	本文方法	**0.789**	**0.829**	0.746	**0.786**

(1)移除模型的任何重要部分或重要方法,模型性能都会出现不同程度的下降,这说明了本文所提出模型中的各模块在多模态虚假新闻检测任务中的有效性.

(2)根据准确率的下降程度,去除局部显著特征对于模型性能的影响最为明显,本文所使用的细粒度特征提取方法是保证虚假新闻检测精度的重要基础.通过去除非共享语义线索和关联边界优化的消融实验发现,捕捉新闻内部的跨模态语义交互能够帮助模型有效挖掘非共享语义线索,这在提升模型性能的同

时，增强了检测过程的可解释性．融合过程中使用的关联分数对于模型性能提升有限，这说明了两类线索在虚假新闻检测中互有补充，共同发挥重要作用．

(3)对比在不同新闻语境下移除各模块后模型性能的下降程度，去除双分支图文线索感知模块后，仅仅依靠共享语义线索的检测方式在微博和推特的新闻语境下均表现不足，难以全面检出虚假新闻．这强调了不同类型的新闻内部都包含大量潜在的非共享语义线索，促使模型全面感知这类线索是提升多模态虚假新闻检测性能的关键．同时，这也反映出本文所提出的双分支图文线索感知模块能够准确挖掘和利用非共享语义线索，有效弥补了依靠共享语义内容进行检测的不足．

4.5 线索感知与融合的定量分析

4.5.1 非共享语义线索的相关性分析

模型通过训练来学习具备区分度的线索特征是解决多模态虚假新闻检测的关键，这些区分性特征具体表现为同类新闻之间的相关性较强，异类新闻之间的差异性较大．为了形象展现非共享语义线索支持虚假新闻检测的能力，本文使用相关性热力图[33]对不同数据集下非共享语义线索之间的语义相关性进行可视化．

如图6所示，非共享语义线索特征具有明显的类内相关性和类间差异性，特别是在图6(a)中的微博新闻语境下，图像线索和文本线索都呈现出了较为明确的区分边界．在图6(b)中的推特新闻语境下，尽管文本线索之间的区分度不足，但图像线索具有相对明确的区分边界，两者相互补充后仍能有效推动多模态虚假新闻的检测．因此，模型所学习得到的非共享语义线索在图像与文本的单模态特征区分度中已经初具优势，从而也保证了多模态虚假新闻检测的精度．

4.5.2 最优关联边界与关联分数的有效性分析

最优关联边界的学习过程以及关联分数的度量过程是模型获取非共享语义线索的重要基础．随着模型的训练，最优关联边界t_k与关联分数Cor_S的更新共同体现了新闻跨模态语义关联的协同优化与分析过程．为了明确展现两值在非共享语义线索线索感知与融合过程中的有效性，本文在模型训练过程中对两值的优化分析过程进行了同步记录．

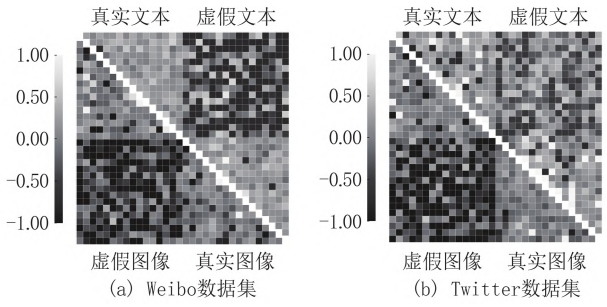


图6 非共享语义线索的图文相关性矩阵

如图7所示，最优关联边界t_k在模型训练初期较不稳定，在零值附近动态变化．随着模型的训练，两类新闻语境下的最优关联边界t_k都逐渐增大并趋向稳定，其中微博新闻的关联边界更高．对于图文内容丰富的微博新闻语境来说，更高的关联边界促使模型更加全面地感知非共享语义线索，避免漏检虚假新闻．关联分数Cor_S在不同新闻语境的变化趋势都较为稳定，其中推特新闻的关联分数更高．关联分数Cor_S反映了两类新闻内容的可信程度，较高的关联分数意味着新闻内部存在较大的局部语义差异，非共享语义线索能够有效帮助模型进行检测．通过对上述两值的定量分析，本文提出的线索感知策略的有效性得到进一步的验证．基于此，模型能够深入理解多模态虚假新闻的内在特点，实现更准确、更可解释的检测过程．

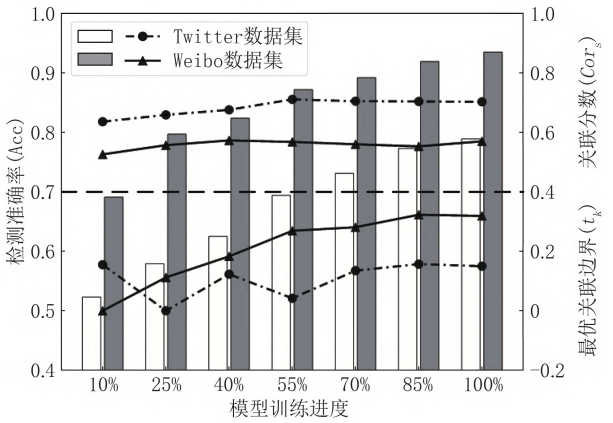


图7 最优关联边界与关联分数的优化分析过程

4.6 案例分析

为了直观展现模型的检测效果，本文从实验数据集中选取了两则具有代表性的多模态虚假新闻进行案例分析．图8分别展现了模型对新闻图文内容的局部特征选择和线索感知效果．在线索感知效果中，深色阴影区域的新闻内容突出非共享语义线索，其他区域的新闻内容突出共享语义线索．

通过观察发现，多模态虚假新闻内部存在丰富的非共享语义线索，模型在检测过程中重点关注了其中的人物、地点以及数量等细节信息，能够准确把

图8　多模态虚假新闻案例分析示意图

握新闻的主体内容与涵义. 结合新闻的实际语境加以验证，非共享语义线索感知方法能够帮助模型深入理解多模态虚假新闻的篡改意图，从而进行有效的推理判断. 基于此，本文所提出的模型能够深度感知共享语义线索和非共享语义线索，有效应对复杂语境下的多模态虚假新闻检测任务.

5　总　结

针对现有方法缺乏对非共享语义线索的有效利用的问题，本文提出了一种双分支线索深度感知与自适应协同优化的多模态虚假新闻方法. 通过细粒度的模式对图像显著区域和文本单词进行表征，并使用跨模态加权残差网络学习共享语义线索. 同时，通过双分支图文线索感知模块挖掘与优化非共享语义线索，进一步根据跨模态语义关联程度调整多模态融合过程. 实验结果表明：本文提出的方法在准确率和虚假新闻检测精确率上大幅超越基线方法，证明了本文所提出方法的有效性.

在未来的工作中，本文将继续在线索深度感知与自适应协同融合层面对多模态虚假新闻检测模型进行优化. 为了完善虚假新闻局部线索的推理过程，本文将深入探索新闻线索的本质特点，分析不同线索的可信程度，并进一步提升多模态虚假新闻检测模型的性能.

参　考　文　献

[1] Shu K, Sliva A, Wang S, et al. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 2017, 19(1): 22-36

[2] Zhang Shun, Gong Yi-Hong, et al. The development of deep convolution neural network and its applications on computer vision. Chinese Journal of Computers, 2019, 42(3): 453-482 (inChinese)
(张顺，龚怡宏，等. 深度卷积神经网络的发展及其在计算机视觉领域的应用. 计算机学报，2019，42(3)：453-482)

[3] Sherstinsky A. Fundamentals of recurrent neural network (RNN)and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 2020, 404: 132306

[4] Nasir J A, Khan O S, Varlamis I. Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights, 2021, 1(1): 100007

[5] Sheng Q, Cao J, Zhang X, et al. Zoom out and observe: News environment perception for fake news detection//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 2022: 4543-4556

[6] Ghorbanpour F, Ramezani M, Fazli M A, et al. FNR: a similarity and transformer-based approach to detect multi-modal fake news in social media. Social Network Analysis and Mining, 2023, 13(1): 56

[7] Zhou Y, Yang Y, Ying Q, et al. Multimodal fake news detection via clip-guided learning//Proceedings of the IEEE International Conference on Multimedia and Expo, Brisbane, Australia, 2023: 2825-2830

[8] Dun Y, Tu K, Chen C, et al. Kan: Knowledge-aware attention network for fake news detection//Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 81-89

[9] Qian S, Wang J, Hu J, et al. Hierarchical multi-modal contextual attention network for fake news detection//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021: 153-162

[10] Shu K, Cui L, Wang S, et al. defend: Explainable fake news detection//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Alaska, USA, 2019: 395-405

[11] Castillo C, Mendoza M, Poblete B. Information credibility on twitter//Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 2011: 675-684

[12] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks//Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, USA, 2016: 3818-3824

[13] Zhang W, Gui L, He Y. Supervised contrastive learning for multimodal unreliable news detection in covid-19 pandemic//Proceedings of the 30th ACM International Conference on Information& Knowledge Management, Gold Coast, Australia, 2021: 3637-

3641

[14] Yuan H, Zheng J, Ye Q, et al. Improving fake news detection with domain-adversarial and graph-attention neural network. Decision Support Systems, 2021, 151: 113633

[15] Qi P, Cao J, Yang T, et al. Exploiting multi-domain visual information for fake news detection//Proceedings of the IEEE International Conference on Data Mining, New York, USA, 2019: 518-527

[16] Zhou P, Han X, Morariu V I, et al. Learning rich features for image manipulation detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 1053-1061

[17] Singh B, Sharma D K. Predicting image credibility in fake news over social media using multi-modal approach. Neural Computing and Applications, 2022, 34(24): 21503-21517

[18] Cao J, Qi P, Sheng Q, et al. Exploring the role of visual content in fake news detection. Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities, 2020: 141-161

[19] Dou Y, Shu K, Xia C, et al. User preference-aware fake news detection//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021: 2051-2055

[20] Shu K, Mahudeswaran D, Wang S, et al. Hierarchical propagation networks for fake news detection: Investigation and exploitation//Proceedings of the International AAAI Conference on Web and Social Media, Atlanta, USA, 2020: 626-637

[21] Singhal S, Shah R R, Chakraborty T, et al. Spotfake: A multi-modal framework for fake news detection//Proceedings of the IEEE 5th International Conference on Multimedia Big Data, New Delhi, India, 2019: 39-47

[22] Wang Y, Ma F, Jin Z, et al. Eann: Event adversarial neural networks for multi-modal fake news detection//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 2018: 849-857

[23] Khattar D, Goud J S, Gupta M, et al. Mvae: Multimodal variational autoencoder for fake news detection//Proceedings of the World Wide Web Conference, San Francisco, USA, 2019: 2915-2921

[24] Silva A, Luo L, Karunasekera S, et al. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data//Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 557-565

[25] Jin Z, Cao J, Guo H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs//Proceedings of the 25th ACM International Conference on Multimedia, California, USA, 2017: 795-816

[26] Song C, Ning N, Zhang Y, et al. A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. Information Processing & Management, 2021, 58(1): 102437

[27] Qi P, Cao J, Li X, et al. Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues//Proceedings of the 29th ACM International Conference on Multimedia, Chengdu, China, 2021: 1212-1220

[28] Xue J, Wang Y, Tian Y, et al. Detecting fake news by exploring the consistency of multimodal data. Information Processing & Management, 2021, 58(5): 102610

[29] Zhou X, Wu J, Zafarani R. Safe: Similarity-aware multi-modal fake news detection//Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Singapore, 2020: 354-367

[30] Shang L, Kou Z, Zhang Y, et al. A duo-generative approach to explainable multimodal covid-19 misinformation detection//Proceedings of the ACM Web Conference, Lyon, France, 2022: 3623-3631

[31] Jin Y, Wang X, Yang R, et al. Towards fine-grained reasoning for fake news detection//Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, USA, 2022: 5746-5754

[32] Connor C E, Egeth H E, Yantis S. Visual attention: bottom-up versus top-down. Current Biology, 2004, 14(19): 850-852

[33] Chen Y, Li D, Zhang P, et al. Cross-modal ambiguity learning for multimodal fake news detection//Proceedings of the ACM Web Conference, Texas, USA, 2022: 2897-2905

[34] Zhen L, Hu P, Wang X, et al. Deep supervised cross-modal retrieval//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, California, USA, 2019: 10394-10403

[35] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778

[36] Chung J S, Zisserman A. Out of time: automated lip sync in the wild//Proceedings of the Asian Conference on Computer Vision. Springer, Cham, Las Vegas, USA, 2016: 251-263

[37] Goldberger J, Gordon S, Greenspan H. An efficient image similarity measure based on approximations of KL-Divergence between two gaussian mixtures//Proceedings of the IEEE International Conference on Computer Vision, Nice, France, 2003: 487-493

[38] Zhang W, Xie R, Wang Q, et al. A novel approach for fraudulent reviewer detection based on weighted topic modelling and nearest neighbors with asymmetric Kullback-Leibler divergence. Decision Support Systems, 2022, 157: 113765

[39] Ma J, Gao W, Wong K F. Detect rumors in microblog posts using propagation structure via kernel learning//Proceedings of the Association for Computational Linguistics, Vancouver, Canada, 2017: 708 - 717

[40] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv Preprint ArXiv:1810.04805, 2018

[41] Sengupta A, Ye Y, Wang R, et al. Going deeper in spiking neural networks: VGG and residual architectures. Frontiers in Neuroscience, 2019, 13: 95

[42] Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008, 9(11): 2579-2605

ZHONG Shan-Nan, M. S. candidate. His research interests include multimedia data mining and multimodal comprehension.

PENG Shu-Juan, Ph. D., associate professor. Her research interests include pattern recognition, machine learning and multimedia analytics

LIU Xin, Ph. D., professor. His research interests include artificial intelligence and multimedia analytics.

WANG Nan-Nan, Ph. D., professor. His research interests include computer vision and machine learning.

LI Tai-Hao, Ph. D., professor. His research interests include intelligent computing and emotion recognition.

Background

In recent years, multimodal news has gradually become the mainstream form of information in social media, but it also poses great challenges to fake news detection technology. Benefiting from the deep neural network to automatically extract the high-level feature representation of fake news, the deep learning method represented by convolution neural network and recurrent neural network has achieved great success in the task of fake news detection. At present, multimodal fake news detection methods generally adopt a general pretrained model to capture the text features and visual features of fake news, and then obtain multimodal features by attention mechanism. However, current cross-modal global alignment methods ignore the fine-grained interactions between local salient regions of images and words. This results in models that lack the ability to capture cross-modal semantic correlation, hindering the improvement of fake news detection.

In order to solve the above problems, this paper proposes a multimodal fake news method based on two-branch clue deep perception and adaptive collaborative optimization, which achieves deep mining of non-shared semantic clues. Through the two-branch clue perception module, the detection model proposed in this paper can accurately mine potential non-shared semantic clues of news and adaptively adjust the multimodal fusion process. The performance of the detection model is compared with nine representative fake news detection baseline, and the results of experiment show that it outperforms the competing baseline methods in many aspects, which fully proves the effectiveness of the proposed detection model.