



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

《计算机工程与应用》网络首发论文

题目：多模态特征自适应融合的虚假新闻检测
作者：王腾，张大伟，王利琴，董永峰
网络首发日期：2023-08-25
引用格式：王腾，张大伟，王利琴，董永峰. 多模态特征自适应融合的虚假新闻检测[J/OL]. 计算机工程与应用.
<https://link.cnki.net/urlid/11.2127.TP.20230824.1644.004>



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

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

多模态特征自适应融合的虚假新闻检测

王腾¹, 张大伟², 王利琴¹, 董永峰¹

1.河北工业大学 人工智能与数据科学学院,天津 300401

2.中国科学院自动化研究所 模式识别国家重点实验室,北京 100190

摘要：为解决社交媒体新闻中多模态新闻检测难以充分利用图文信息问题以及探索高效的多模态信息交互方法，提出了一种多模态特征自适应融合的虚假新闻检测模型。首先，分别对新闻文本语义特征、文本情感特征和图文语义差异特征进行提取和表示；然后，通过添加自适应权重参数的方式对多种特征进行加权拼接融合，以减少模型拼接时引入的冗余信息；最后，将融合特征送入分类器中进行新闻的真假分类。实验结果表明，所提出的模型在F1值等评价指标上都优于当前先进的模型。有效提升了虚假新闻检测性能，为社交媒体中虚假新闻的检测提供了有力支持。

关键词：虚假新闻检测；情感特征；图像描述；自适应融合

文献标志码：A 中图分类号：TP391 doi: 10.3778/j.issn.1002-8331.2303-0316

Multimodal Feature Adaptive Fusion for Fake News Detection

WANG Teng¹, ZHANG Dawei², WANG Liqin¹, DONG Yongfeng¹

1.School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300401, China

2.National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract: In order to solve the problem that it is difficult to make full use of graphic and text information in multimodal news detection in social media news and to explore efficient multimodal information interaction methods, an adaptive multimodal feature fusion model for fake news detection is proposed. First, the model extracts and represents news text semantic features, text emotional features, and image-text semantic difference features; then, weighted splicing and fusion of various features are performed by adding adaptive weight parameters to reduce the redundancy introduced by model splicing; finally, the fusion feature is sent to the classifier. Experimental results show that the proposed model outperforms the current state-of-the-art models in evaluation indicators such as F1 score. It effectively improves the performance of fake news detection and provides strong support for the detection of fake news in social media.

Key words: fake news detection; emotional feature; image caption; adaptive fusion

基金项目：国家自然科学基金资助项目（61806072）；河北省高等学校科学技术研究项目（ZD2022082；QN2021213）；河北省自然科学基金（F2020202008）。

作者简介：王腾(1993-)，男，硕士研究生，CCF 学生会员，研究方向为虚假新闻检测、自然语言处理；张大伟(1989-)，男，副研究员，研究方向为模式识别、自然语言处理、知识推理；王利琴(1980-)，女，博士，实验师，CCF 专业会员，研究方向为智能信息处理、知识图谱；董永峰(1977-)，男，博士，教授，CCF 高级会员，研究方向为人工智能、知识图谱，E-mail: dongyf@hebut.edu.cn。

随着网络与社交媒体的飞速发展,信息的分享与获取达到前所未有的便利,虚假新闻也趁势泛滥。2022年中国互联网络信息中心(CNNIC)发布的《第50次中国互联网络发展状况统计报告》^[1]显示,截至2022年6月,我国网民规模达10.51亿,其中网络新闻用户规模达7.88亿,较2021年12月增长1698万,占网民整体的75.0%。互联网用户每天自发地在网络上发布和分享最新的新闻和见解,却很少仔细检查他们所分享信息的真实性,这意味着大量的虚假信息可能会在社会媒体平台出现和传播。

不同于传统媒体的传播模式,社交媒体的传播特点和方式使得新闻扩散愈加迅速,其所带来的不良影响更广,社会危害也更大^[2]。尤其在新冠肺炎疫情期间,互联网上充斥着捏造的错误信息,例如茶叶、醋和盐水可以有效治疗新冠肺炎疫情^[3]。显而易见的是,社交媒体与新闻的联系日趋复杂,如何阻止虚假新闻在社交媒体平台造成社会危害已成为一个值得深入探究的社会问题。因此,开展虚假新闻检测研究,及时发现和阻止社交媒体虚假信息的传播,对净化网络空间环境、维护社会和谐稳定、保障国家信息安全等方面意义重大。

虚假新闻检测在早期主要依靠人工进行事实核查,依靠专家的领域知识进行判断,不仅耗时耗力,而且时效性和扩展性较差。为了解决早期方法的局限性,研究人员通过使用专家知识从新闻文本内容中手动提取特征,使用朴素贝叶斯^[4]、决策树^[5]或支持向量机^[6]等传统的机器学习算法进行虚假新闻分类。但是,特征的选取会直接影响分类器的性能,这不但增大了检测开销,而且也对检测效率造成影响。随着深度学习的持续发展与应用以及新闻内容愈发多样化,很多针对多模态特征的研究开始利用深度神经网络学习模态之间的高阶表示。在虚假新闻检测方面,最具代表性的是通过一些辅助任务或简单拼接等方式融合多模态特征的。例如,早期研究主要是通过简单拼接来增强特征的提取和融合多模态信息。其他一些研究利用辅助任务,如特征重构和事件判别来提高虚假新闻检测的可靠性和泛化能力。

虽然人们在多模态虚假新闻检测中付出了巨大的努力,但是现有方法依然存在问题:一方面,在处理文本信息时,仅仅聚焦文本语义特征,而忽略了文本深层的情感特征,这对识别煽动性极强的虚假新闻有很明显的帮助。另一方面,现有方法在处理图文多模态信息时,没有充分挖掘图文的关联性信息,充分理解这种隐含的一致性关系,对于识别“图文不符”

的虚假新闻有较高的价值。最后,融合多模态特征时,采用向量拼接的方式会引入噪声,产生过多冗余信息,导致多模态虚假新闻的检测效果难有提升。

为解决上述问题,本文提出了一种多模态特征自适应融合的虚假新闻检测模型(Multimodal Feature Adaptive Fusion for Fake News Detection, MFAF)。模型首先分别对新闻的文本特征、文本情感特征和图片文字语义差异特征进行提取和表示,其中,在获取图片文字语义差异特征时,由于社交媒体中的虚假新闻图片大都是图文共存的形式,即存在大量聊天、视频或者新闻报道截图等。而在提取此类图像特征时,目前绝大多数方法使用的图像特征提取模型是在标准的自然图像上进行大规模训练的,这种方式提取到的图像特征会受到图像中文字干扰,导致难以对图像进行准确表达,为此,模型采用图片描述预训练模型获取图像内容的文字化描述,同时使用光学字符识别(Optical Character Recognition, OCR)工具提取图片中的文本信息来对图像描述进行拼接补充,随后通过本文提出的多模态语义特征相似度比较网络,将其与新闻文本特征进行比较,获得图像文本语义差异特征向量。同时为更好地融合上述特征,采用设置和学习多个自适应加权参数的方式对多种特征进行加权拼接,最终进行虚假新闻的结果预测。该模型有效提升了社交媒体中虚假新闻检测任务的准确率。

本文贡献如下:

- 1.提出了多模态特征自适应融合的虚假新闻检测模型。该模型在利用新闻文本语义特征的基础上引入了文本情感特征和图像文本语义差异特征。其中,图像文本语义差异特征是通过将图像描述信息与图像中的文字信息进行拼接,然后与新闻文本语义特征一同送入多模态语义特征相似度比较网络计算得到的。模型最后采用自适应加权拼接的方式进行特征融合,从而学习到新闻内容更为丰富的语义信息,实现了高性能的虚假新闻检测;

- 2.设计了多模态语义特征相似度比较网络。为了获取蕴含新闻图像和文本之间关联信息,更好地对图像与文本语义差异性进行表示,通过多模态语义特征相似度比较将图像和文本的语义特征进行计算,获得了更为合理的图文语义相似度的特征表示,同时也避免了使用传统图像特征提取模型易受到图像中文字干扰,导致图像特征中噪声大、结果偏差大的问题;

- 3.采用了自适应的加权拼接的方式进行特征融合。通过在特征融合时引入可学习参数,将不同特征表示进行自适应加权拼接,有效降低了简单拼接文本图像

特征引入的噪声问题,提高了虚假新闻检测的效果。

1 相关工作

1.1 单模态虚假新闻检测

一篇新闻中包括许多类型的内容,如文本、图像和社交场景信息。在早期,许多方法仅仅是基于新闻的文本内容来检测文章的真实性,应用的主要方法是基于统计的文本特征或者语义级别的文本特征,例如文本的段落数^[7],词语占文本的百分比^[8]、符号数量^[9]、写作风格^[10]和语言风格^[11]等。随着机器学习的不断发展,一些传统的机器学习算法开始应用于虚假新闻的检测。但是,从文本中通过手动的方式提取特征,不仅在设计上困难且耗时耗力,而且无法充分利用文本的内容。为了解决这个问题,许多研究人员使用深度学习方法来识别假新闻,因为深度学习技术具有强大的表征学习能力。Ma 等人^[12]基于递归神经网络(Recursive Neural Network, RNN)提出了一种以时间序列中的文本特征为代表的文章学习模型。Yu 等人^[13]在模型中应用了卷积神经网络(Convolutional Neural Network, CNN),检查基本特征之间更深层次的相互作用。Chen 等人^[14]提出了一种基于 RNN 的虚假新闻检测模型,用于选择性学习连续新闻文本序列的时间隐藏表示,可以得到具有某一特定焦点的不同特征向量,并产生其隐藏表示,从而捕捉到相关新闻内容随时间的上下文变化。沈瑞琳等人^[15]在虚假新闻检测任务中运用了迁移学习的思想,通过分析和使用新闻的评论数据,并将模型的特征提取网络进行微调,使得提取出来的新闻评论文本可以应用到后续的检测中。

随着新闻内容多模态形式的兴起,利用视觉特征检测虚假信息的研究正在迅速增加。许多研究^[16-17]使用视觉特征仔细调查了伴随图像和图像类型等信息。然而,这些方法的学习模式较为简单,并且有很大的局限性。Qi 等人^[18]提出了一种基于卷积神经网络的图像模式捕获模型,该模型最初使用 CNN 提取频域模式,随后使用递归神经网络对照片真实性进行语义检测,最后使用注意力机制融合图像模式和频域模式。尽管图像的视觉特征可以提供丰富的视觉信息,但是目前针对视觉特征在虚假新闻检测中的作用的理 解仍然存在局限性^[19]。

1.2 多模态虚假新闻检测

随着虚假新闻的多样化发展,许多研究人员开始研究多模态虚假新闻检测问题。文本数据和图像数据的表示维度不同,这为充分利用二者之间的互补关系

增加了难度。目前常用的做法是对文本和图片采取两种不同的特征提取器,将得到不同模态的特征进行一系列处理后拼接到一起然后放入不同的分类器中进行分类。

Jin 等人^[20]提出了一种基于注意力机制的循环神经网络来融合多模态特征,将图片、文本和背景特征进行融合,获得了比单模态检测方法更好的效果,这也是最早将基于深度学习的多模态特征学习方法引入虚假新闻检测中的方法。Kumari 等人^[21]提出了一个多模态假新闻检测框架,该框架最大限度地提高了文本和图像特征之间的相关性,实现更好的多模态信息表示。Wang 等人^[22]提供了一个虚假新闻检测框架,该框架根据少量经过验证的帖子有效地学习和进行突发新闻的检测。大规模预训练模型的发展极大地推动了虚假新闻检测的研究。Tuan 等人^[23]通过引入 BERT^[24](Bidirectional Encoder Representations from Transformers)和注意力机制有效地学习和融合帖子中的多模态特征。Wang 等人^[25]构建了一个事件对抗性神经网络(Event Adversarial Neural Network, EANN),使用事件鉴别器学习文章中文本和图像的特征表示。Zhuang 等人^[26]分别从采用了词表示的全局向量^[27](Global Vectors for Word Representation, GloVe)和 VGG16^[28]作为文本和图像的特征提取器,将得到的文本特征放入 BiLSTM 层和全连接层处理后,与经过全连接层处理后图像特征拼接再放入多层感知机进行分类。Khattar 等人^[29]为解决现有虚假新闻检测方法无法学习多模态信息的共享表示,提出了一个多模态变分自动编码器(Multimodal Variational Autoencoder, MVAE),可以学习多模态信息特征的共享表示,同时为了更好地利用多模态信息, MVAE 还设计了一个用于新闻重构的辅助任务,可以更好地融合新闻的多模态信息。Qian 等人^[30]提出了一种分层多模态上下文注意力网络(Hierarchical Multi-modal Contextual Attention Network, HMCAN)用于多模态虚假新闻检测,很好地针对虚假新闻检测任务中容易忽略文本内容完整层次语义、多模态信息以及高阶互补信息的问题,该网络可以在建模多模态信息的同时建模多层次语义关系。Li 等人^[31]则在上述工作的基础上,进一步对图像和文字的实体对应关系进行了探索,提出了一个面向实体的多模态对齐和融合网络(Entity-oriented Multi-modal Alignment and Fusion network, EMAF)。该网络采用了以实体为中心的跨模态交互方式,这样可以保留语义完整性,同时能够捕捉到多模态实体的更多细节。Zhou 等人^[32]认为新闻中的图像和文字之

间存在强相关性,针对图文不符的虚假新闻检测,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),该模型首先使用 image2sentence 将图片转化为文本的表述方式,接着用 TextCNN 分别提取文本和图片的特征,并通过全连接层和模态交互层计算新的相似性函数和损失函数对文本图片进行分类。

综上所述,现有的虚假新闻检测研究已经开始综合考虑新闻的图像与文本信息,尽管在虚假新闻检测任务上取得了很好的性能,但仍存在以下问题:

(1) 文本深层情感特征未充分利用。上述研究在利用文本语义特征时,往往只是考虑采用合适的模型提高文本语义特征的精度,而忽略了文本深层的隐藏特征,如本文引入的情感特征,可以获取到新闻文本更为丰富的语义信息。

(2) 图像语义特征不能精准获取和利用。在提取图像特征时,由于社交媒体中的虚假新闻图片大都是图文共存甚至只有文字的形式,而目前大多数方法,如 MVAE 模型使用的图像特征提取器,是在标准的自然图像上训练而成的,在处理图片时,很容易受到图像中文字干扰,导致难以对图像进行准确表达。同时,获取到的图像语义特征需要得到充分利用以进一步提升虚假新闻检测精确度。如 SAFE 模型,通过计算文

字和图像语义特征的一致性作为相似度损失函数,与分类损失函数一起对虚假新闻检测模型进行监督。然而,虽然该方法在图像特征利用上提供了新的思路,但仍然存在由于图像语义描述提取错误导致引入监督噪声的问题。

(3) 融合后的特征噪声信息过多。在多模态特征融合时,现有方法大都通过向量拼接的手段,如 EANN 模型通过将图像和文本特征简单拼接,往往会引入噪声,导致多模态虚假新闻的检测效果难有提升。

为了解决上述问题,本文在新闻内容的特征提取、特征表示和特征融合等方面进行了研究、实验设计与验证,设计了表现效果更为优异的虚假新闻检测模型。

2 多模态特征自适应融合模型

虚假新闻检测任务通常被建模为二元分类问题,即真实新闻和虚假新闻两类。模型的输入数据是新闻内容,包含文字与图像等多种模态信息,输出是真实新闻和虚假新闻的标签,其中 1 为虚假新闻,0 为真实新闻。

本文提出了一种多模态特征自适应融合的虚假新闻检测模型,结构如图 1 所示,模型主要由输入层、特征表示层、特征融合层和输出层组成。

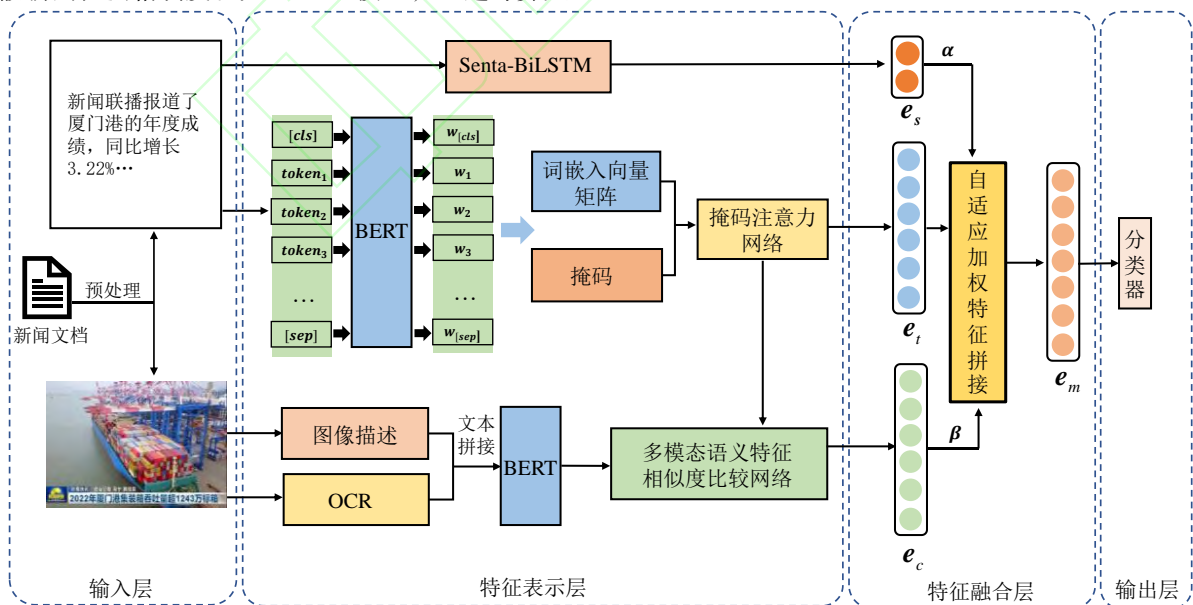


图 1 多模态特征自适应融合模型

Fig.1 Multimodal feature adaptive fusion model

2.1 输入层

在输入层中,主要实现对新闻数据中图像内容和文字内容的预处理操作。

首先对新闻中的文本进行预处理,主要是对文本

中的信息进行过滤,文本数据预处理过程包括:文本标准化、文本清洗、繁体转换和拼写纠错等步骤,获得最终的新闻文本 S 。

然后对新闻中的图片进行预处理。由于数据集中

存在少量长宽比失衡和分辨率较低的图片,剔除这些图片后,考虑到图像描述模型和 OCR 提取工具的处理效果,本文将图片尺寸调整至 512×512,得到最终的新闻图像 I 。

将预处理后的文本和图像数据,作为输入数据。

2.2 特征表示层

针对输入层预处理后的数据,分别采用不同方法来实现数据的向量化表示。

(1) 情感特征提取

新闻文本中通常蕴含大量的情感信息,新闻发布者为了吸引读者,使虚假新闻迅速传播,会添加一些具有情感煽动性的词汇,例如“震惊”、“千万不要”等,从而引起读者的情感共鸣,进而干扰读者的主观判断。

为了充分利用文本的深层情感特征,有效地捕捉更多语义信息,本节使用 Senta-BiLSTM 提取新闻文本的情感特征,详见图 1 的特征表示层上半部分。Senta-BiLSTM 是以 BiLSTM 为基础,使用了预训练模型和 SKEP^[33] (Sentiment Knowledge Enhanced Pre-training) 算法而形成的专门用于进行情感分类的方法。由于该算法是在海量数据上训练而成的,学习到了丰富的情感语义。

具体地,首先在虚假新闻分类任务的数据集上对模型进行微调,随后使用模型预测新闻文本的情感分类,结果返回文本的积极情绪和消极情绪的预测概率,最终通过构建文本情感特征 e_s , 得到一个 2×1 维的文本情感极性特征表示:

$$e_s = \text{Senta_BiLSTM}(S) \quad (1)$$

(2) 文本特征提取

文本作为新闻的主体内容,其包含了丰富的信息,所以文本特征便成为了文本分类任务中的重要一环,提取特征的好坏将对虚假新闻检测的准确率产生很大影响。以 BERT 为代表的预训练语言模型,由于是在大规模数据上进行训练,其内部也包含了大量语法知识和语句间的隐含信息,这大大增强了模型的语义表示能力,因此,本文使用 BERT 预训练模型作为新闻文本模态的特征提取器,具体流程详见图 1 特征表示层中间部分。

具体地,首先对输入的文本语句 $S = [s_1, s_2, \dots, s_n]$ 进行编码,其中 s_i 代表分词后句子中第 i 个单词,添加 [CLS] 和 [SEP] 标记后,获得 token 列表 $T = [[CLS], token_1, \dots, token_n, [SEP]]$ 其中 $token_i$ 为第 i 个单词的 token, 公式如下:

$$T = \text{BertTokenizer}(S) \quad (2)$$

然后将 token 列表 T 送入 BERT 模型,获得单词的嵌入向量 $W = [w_{cls}, w_1, \dots, w_n, w_{sep}]$, 其中 w_i 为第 i 个单词的嵌入表示, 公式如下:

$$W = \text{BERT}(T) \quad (3)$$

最后,将所有单词的嵌入向量送入掩码注意力网络,通过计算每个单词的权重,对其向量表示进行加权,从而突出更重要单词对句子的影响。得到维度为 768×1 的新闻文本句子的语义特征表示 e_t , 公式表示如下:

$$e_t = \text{Mask_Attention}(W) \quad (4)$$

(3) 图文语义差异特征提取

图片作为新闻内容的一部分,其蕴含的语义信息往往与新闻的文本信息联系紧密。为解决 MVAE 等模型不能精准提取图像语义特征的问题,本文采用图像描述模型提取新闻图像的描述信息,同时,采用光学字符识别模型对图像中包含的文字信息进行提取,作为图像描述信息的补充,将两者拼接后获得图像的文本表达信息,最终挖掘到了新闻图像的高阶语义信息。具体流程详见图 1 特征表示层下半部分。

针对待检测新闻的一张图像 I 可能同时包含图像的描述以及文本信息,使用基于 VIT-GPT2 的图像描述预训练模型和 CnOCR 文字识别工具分别提取新闻图像中文描述以及图像中的文字信息,然后将两者拼接,公式表示如下:

$$D_{i2t} = \text{image_to_text}(I) \quad (5)$$

$$D_{ocr} = \text{Cn_OCR}(I) \quad (6)$$

$$D = \text{concat}(D_{i2t}, D_{ocr}) \quad (7)$$

其中, D_{i2t} 为图像描述的文本, D_{ocr} 为 OCR 提取的图像中文本, D 为最终图像描述文本和图像中文本的拼接文本, $\text{image_to_text}(\cdot)$ 为 VIT-GPT2 模型的图像描述计算, $\text{Cn_OCR}(\cdot)$ 为 CnOCR 的核心计算函数,用于提取图像中的文字信息, $\text{concat}(\cdot)$ 为特征向量拼接操作。

随后,将拼接的文本输入到 BERT 模型中,获得图像的文本描述特征表示,其公式表示如下:

$$T_d = \text{BertTokenizer}(D) \quad (8)$$

$$W_d = \text{BERT}(T_d) \quad (9)$$

$$e_d = \text{Mask_Attention}(W_d) \quad (10)$$

其中 D 为图像描述文本和图像中的文本的拼接文本, T_d 为拼接文本单词的 token 列表, W_d 为单词的词嵌入向量, e_d 表示拼接文本句子的语义特征表示, $\text{BertTokenizer}(\cdot)$ 是 BERT 模型的编码函数, $\text{BERT}(\cdot)$ 为 BERT 模型的核心计算函数, $\text{Mask_Attention}(\cdot)$ 为

掩码注意力网络计算函数,用于获取文本句子的语义特征表示。

随后,在得到上述图像的文本表达信息基础上,设计了一种多模态语义特征相似度比较网络,以获得图像文本语义差异比较特征,与SAFE容易引入监督

噪声问题不同的是,MFAF则是将其加入到特征集合中,并在后续为其添加可学习参数,从而在有效利用图文语义特征的情况下,避免噪声的产生。网络结构如图2所示:

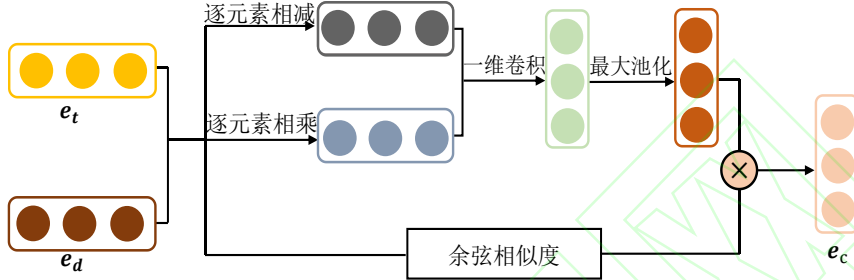


图2 多模态语义特征相似度比较网络

Fig.2 Multimodal semantic feature similarity comparison network

e_t 表示新闻的文本句子的语义特征向量, e_d 表示图像描述文本与图像中文本拼接后的文本句子语义特征向量,为了衡量两个嵌入的贴近度与相关性,受到 Shen 等人^[34]的启发,使用逐个元素减法和乘法的组合设计了比较函数 $f_{cmp}(\cdot)$, 公式如下:

$$f_{cmp}(e_d, e_t) = \text{MaxPool}(\text{conv1D}[e_d - e_t, e_d \odot e_t]) \quad (11)$$

其中, \odot 为 Hadamard 乘积, $\text{conv1D}[\cdot]$ 为一维卷积计算, $\text{MaxPool}(\cdot)$ 为最大池化计算。

在此基础上,使用向量余弦相似度计算新闻图像和文本的语义一致性,其公式表示如下:

$$\text{cosine_similarity}(e_d, e_t) = \frac{e_d \cdot e_t}{\max(\|e_d\|_2 \cdot \|e_t\|_2, \varepsilon)} \quad (12)$$

最后,将两者相乘,获得最终的新闻图像文本语义差异比较特征表示 e_c , 其公式如下:

$$e_c = f_{cmp}(e_d, e_t) * \text{cosine_similarity}(e_d, e_t) \quad (13)$$

2.3 特征融合层

特征融合层将输入的多种特征经过计算后,输出融合后的特征向量。为了解决 EANN 等模型将多模态的特征向量简单拼接,从而引入过多的噪声的问题,本文对不同特征赋予不同的可学习参数,进行自适应加权特征拼接操作,得到最终的新闻多模态特征表示。

具体地,将获得的新闻文本特征表示 e_t 前的参数固定为 1,将新闻情感极性特征表示 e_s 、和新闻图文语义差异比较特征表示 e_c 前分别添加不同的可学习参数 α 、 β 。之后将添加了可学习参数的不同特征表示进行拼接,获得最终的新闻多模态特征表示 e_m , 具体公式如下:

$$e_m = e_t + \alpha e_s + \beta e_c \quad (14)$$

2.4 输出层

输出层首先包括两组线性变换层、批归一化层和 ReLU 激活函数层,以对 e_m 进行非线性变换,得到 e'_m , 然后使用一个带有 softmax 激活函数的全连接层,用以输出 e'_m 代表真实新闻和虚假新闻的概率,具体公式如下:

$$y = \text{softmax}(W' \times e'_m + b) \quad (15)$$

其中 y 为预测概率, W' 为全连接层的参数, b 为偏置项。

根据预先设定的阈值 thr , 得到最终的预测结果:

$$\text{预测结果} = \begin{cases} \text{虚假新闻}, & y > thr \\ \text{真实新闻, 其他情况} \end{cases} \quad (16)$$

由于虚假新闻检测任务属于一个二元分类问题,所以本文采用二元交叉熵(Binary Cross Entropy, BCE)作为损失函数,如下公式所示:

$$L = -\sum_{i=1}^N (y^i \log y^i + (1 - y^i) \log(1 - y^i)) \quad (17)$$

其中 y^i 代表真实标签的值, y^i 表示模型预测的值。

3 实验过程及结果分析

3.1 数据集和评价标准

(1) 数据集

为了更好地评价模型性能,采用 Weibo21, MCG-FneWS 和 DataFountain 平台疫情期间互联网虚假新闻检测竞赛数据集(以下简称 DF-NEWS)。

Weibo21 数据集包含了 2014 年 12 月至 2021 年 3 月共计 9128 条数据。其中有 4488 条虚假新闻和 4640 条真实新闻。

MCG-FNeWS 是智源研究院与中科院计算所开源

的已标注好的虚假新闻数据集, 包含文本和图片两种模态。共含有 38471 条微博新闻和 34096 张对应图片, 其中包括真实新闻 19186 条和对应图片 20461 张, 虚假新闻 19285 条和对应图片 13635 张。

DF-NEWS 是一个疫情期间互联网虚假新闻检测竞赛数据集, 包含了文本和图片两种模态的信息, 共含有 48876 条新闻和 61432 张图片。

本文在保持真实新闻和虚假新闻数量平衡的前提下, 对数据集的文字和图片进行筛选和调整, 最终获得的数据统计信息如表 1 所示。

表 1 数据集的数据统计信息

Table 1 Data distribution of datasets

数据集	原始新闻数量	含图片的新闻数量	最终数量
Weibo21	9128	6344	5678
MCG-FNeWS	38471	19597	18439
DF-NEWS	48876	28338	26774

在本文实验中, 每条新闻都与一张图片对应, 每个数据集的训练集、验证集和测试集以 6:2:2 的比例进行划分。

(2) 评价指标

采用分类任务常用的准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1 值 (F1-score) 指标评价虚假新闻检测模型的性能, 同时使用 AUC 值 (Area Under Curve) 评价模型对于正负样例的辨别能力, 首先给出以下几个定义:

TP (true positive) 和 TN (true negative) 分别代表被模型正确预测的正类样本数量和负类样本数量;

FP (false positive) 和 FN (false negative) 分别代表被模型错误预测为正类的负类样本数量和被模型错误预测为负类的正类样本数量;

准确率表示预测正确的样本数量占所有样本数量的比例, 计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

精确率的定义是表示被模型预测为正类的样本占所有实际正类样本中的比例, 计算公式如下:

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

召回率表示正类样本占模型正确预测的样本比例, 计算如下式:

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

F1 值是精确率和召回率的调和平均, 综合考虑了以上两指标, 计算公式如下:

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (21)$$

AUC 是衡量二分类模型优劣的一种评价指标, 表示在数据集中任取一个正样本和负样本, 预测的正例排在负例前面的概率。在通常情况下, AUC 越大, 说明模型把正例放在前面的可能性越大, 可以衡量模型的排序能力, 计算公式如下:

$$AUC = \frac{\sum_{i \in positive} rank_i - \frac{n_{pos}(1 + n_{pos})}{2}}{n_{pos}n_{neg}} \quad (22)$$

其中, $rank_i$ 为一个正例在整体中从低分到高分排序的位次值, n_{pos} 为正样本个数, n_{neg} 为负样本个数。

3.2 实验设置

本文的实验在 Ubuntu16.4 和 Python3.8 的环境下进行, 采用的 Pytorch 深度学习框架构建虚假新闻检测模型并进行模型训练, 具体参数如表 2 所示。

表 2 实验参数设置

Table 2 Experimental parameter setting

参数	值
Epoch	100
学习率	0.0005
Dropout	0.3
BatchSize	32
优化器	Adam

3.3 对比模型

为了验证提出模型的有效性, 采用以下模型作为对比模型, 进行实验验证并分析实验结果。

BERT: 使用 BERT 预训练模型提取新闻文本特征并分类。

EANN: EANN 由多模态特征提取器、虚假新闻检测器和事件鉴别器组成。为了保证本实验的公平性, 实验中去除了 EANN 模型的事件鉴别器, 只保留了虚假新闻检测模块。

MVAE: 分别使用 Bi-LSTM 和 VGG-19 提取文本和图像特征后, 采用两种特征进行拼接的方式获得融合特征, 然后使用变分自编码器来学习多模态数据的分布, 最后通过将多模态特征重构后, 将其用于新闻分类预测。

SAFE: 首先使用 image2sentence 将图片转化为文本的表述方式, 然后使用 Text-CNN 来提取新闻的文本信息特征和图片描述信息特征, 随后计算新闻文

本和图像信息之间的相似性，最后进行分类预测。

MDFEND^[35]: MDFEND 是当前虚假新闻检测领域中的 SOTA (state-of-the-art) 模型。该模型利用领

域门来聚合由专家混合提取的多个表示，进而提高检测效果。本文对所有专家采用相同的 TextCNN 结构。

表 3 MFAF 模型与基线模型实验对比结果

Table 3 Comparison results of MFAF model and baseline models

数据集	方法	Accuracy	Precision	Recall	F1-score	AUC
Weibo21	BERT	0.8808	0.8628	0.9004	0.8812	0.8811
	EANN	0.8639	0.8331	0.9040	0.8671	0.8646
	MVAE	0.7749	0.7189	0.8895	0.7951	0.7769
	SAFE	0.8514	0.8279	0.8804	0.8534	0.8519
	MDFEND	0.9110	0.9248	0.8913	0.9077	0.9107
	MFAF	0.9146	0.8838	0.9511	0.9162	0.9152
MCG-FNeWS	BERT	0.9664	0.9658	0.9668	0.9663	0.9664
	EANN	0.9341	0.9402	0.9286	0.9343	0.9341
	MVAE	0.9314	0.9343	0.9277	0.9310	0.9314
	SAFE	0.9574	0.9561	0.9587	0.9574	0.9574
	MDFEND	0.9764	0.9782	0.9744	0.9763	0.9764
	MFAF	0.9770	0.9771	0.9766	0.9769	0.9770
DF-NEWS	BERT	0.9699	0.9559	0.9575	0.9567	0.9670
	EANN	0.9399	0.9266	0.8977	0.9119	0.9300
	MVAE	0.9289	0.9091	0.8831	0.8959	0.9181
	SAFE	0.9660	0.9554	0.9461	0.9508	0.9614
	MDFEND	0.9785	0.9729	0.9650	0.9689	0.9754
	MFAF	0.9796	0.9729	0.9682	0.9706	0.9770

3.4 实验结果分析

本节将本文提出的模型 MFAF 和基准模型进行对比比较，实验结果如表 3 所示。

通过观察对比结果，可以看出，本文提出的 MFAF 模型在多项指标上均优于其他对比模型，取得良好的分类效果。其中，所有模型在 MCG-FNeWS 和 DF-NEWS 数据集上的表现优于 Weibo21 数据集，这是因为 Weibo21 数据集的训练样本量较少，可供模型学习的数据多样性不足，从而降低了模型的泛化能力，在测试集上的检测效果差于后两种数据集。

针对只考虑了新闻文本的 BERT 模型，依然可以达到较为不错的检测结果，这说明新闻文本中包含了绝大部分的信息内容和含义，是虚假新闻检测中的主要依据。可以看到，在 Weibo21 数据集上，BERT 模型 F1-score 达到 0.8812，与本文提出的 MFAF 模型相比，约有 4% 的性能差距。而在 MCG-FNeWS 和 DF-NEWS 数据集上，差距明显缩小，这说明增加数据样本量，提高训练数据的多样性，可以有效提升模型的检测精度。

然而，融合了多模态数据特征的 EANN 模型和 MVAE 模型却在性能方面有明显下降。这很有可能是因为，社交媒体新闻数据集的图片不够规范，通常包含大量文字信息，而 VGG19 是在自然图像数据上训练得到的，使得其在社交媒体图片上提取到的特征无法准确描述其深层语义，从而降低了分类性能。

SAFE 模型使用图像描述技术，将图像转换为文本表述形式，从而探索输入文本和对应图像语义描述之间的相似性。SAFE 首先利用 Text-CNN 架构分别对文本特征和图像描述特征进行提取，之后利用两者的拼接特征进行分类预测，同时计算文本特征和图像特征的相似度，以作为损失函数进行监督训练，该模型分别受到两个损失函数的监督：分类损失函数和相似度损失函数。与 SAFE 模型不同，本文提出的 MFAF 模型，将提取到的文本特征与图像描述特征输入到多模态语义特征相似度比较网络中，生成差异向量，再对该差异向量进行自适应加权拼接，作为原始文本特征的补充信息，对新闻进行分类预测。该方法的好处是可以在有效利用图文语义特征的情况下，减少由于

图像语义提取错误导致的引入监督噪声的问题,从而得到更好的检测效果。MFAF 模型较 SAFE 在 F1 分数上提升了 7%以上。

MDFEND 模型在所有数据集上均取得了不错的效果,与引入多模态融合的方式不同,该模型提出了“领域门”的概念,充分提取了不同领域的新闻文本特征,取得了 SOTA 结果。与 EANN 模型相比,本文提出的 MFAF 模型通过度量图像描述信息和图像文字信息与原始文本信息特征相似度的方式,弥补了直接使用图像提取特征的不足,在实验结果的所有度量指标上均获得了良好的表现,超越 MDFEND,达到 SOTA 水平,进一步证明了 MFAF 模型的有效性。

为了进一步验证模型的稳定性和泛化性能,本文分别统计了所有模型在三个数据集的验证集上的 F1 值。每进行一个回合的训练,就将模型在验证集上进行测试,得到的统计结果如图 3、图 4 和图 5 所示。可以看到,几种模型的 F1 值在约 20 回合后趋于稳定。在 Weibo21 数据集上, MDFEND 模型表现更优,但波动较大,在测试集上的结果也差于 MFAF 模型。这也表明,在数据量较小的情况下,验证集和测试集的数据分布存在一定差异,导致在验证集上表现良好的模型在测试集上没有取得较好的结果。而在 MCG-FNeWS 和 DF-NEWS 数据集上, MFAF 模型无论在验证集上还是测试集上均表现更优,这也证明了 MFAF 模型具有良好的泛化性能。

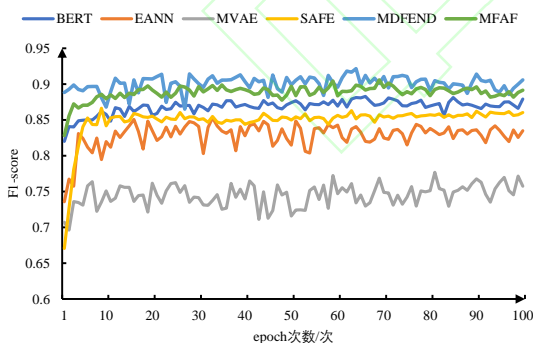


图3 Weibo21 数据集中各模型的 F1 值

Fig.3 F1 score of each model on Weibo21 dataset

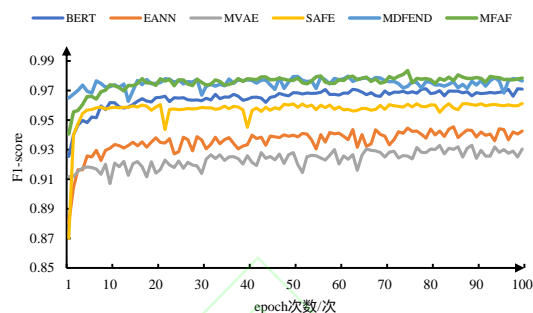


图4 MCG-FNeWS 数据集中各模型的 F1 值

Fig.4 F1 score of each model on MCG-FNeWS dataset

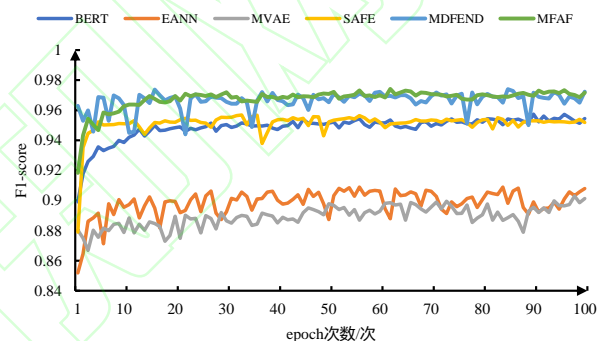


图5 DF-NEWS 数据集中各模型的 F1 值

Fig.5 F1 score of each model on DF-NEWS dataset

3.5 消融实验与结果分析

为了进一步验证本文提出的模型的有效性,同时探索每种特征对实验结果的影响程度,设计了以下消融实验进行分析,鉴于模型在 Weibo21 数据集总体质量高于其他数据集,故展示在 Weibo21 数据集上的消融实验结果来进行分析和讨论,结果见表 4 所示,其中 R_s 代表模型中添加文本情感特征, R_c 代表模型中添加图像文本语义差异特征, C_d 代表模型在融合特征时,直接将多模态特征进行拼接, C_a 代表模型在融合特征时,采用自适应加权特征拼接的操作。在模块选择一栏中通过添加“√”标记代表在本次融合实验中使用了该模块。

分析表 4 中的实验结果,可以得到如下结论:

(1) 实验 1 和后续实验对比可以看出,仅考虑新闻文本单模态信息的模型,检测性能已经可以达到较高水平,这说明文本信息是虚假新闻检测的主要影响因素,而后续实验中融入的文本情感特征、图文语义

差异性特征可以在一定程度上辅助文本特征, 进行补充;

(2) 实验 1 与实验 2 对比可以看出, 融合了文本深层情感特征后, 对比仅有文本特征的检测效果, 各项指标均有提升, 证明了加入文本深层情感特征的有效性。

(3) 实验 2 与实验 3 对比可以看出, 融合特征时, 自适应的加权拼接的方式与传统向量拼接方式相比, 除了在召回率上表现相当, 其余指标均有较明显提升, 证明了采用自适应的加权拼接的方式的有效性。

(4) 实验 2 与实验 4 对比可以看出, 采用特征直接拼接的方式融合了图像文本语义差异特征后, 模型的表现效果有所下降, 这可能是由于将图像语义差异特征进行直接拼接, 会对原始文本特征造成干扰, 从而导致模型检测效果不佳。

(5) 实验 4 与实验 5 对比可以看出, 采用自适应加权特征拼接的方式融合图像文本语义差异特征后, 各项指标均有较大提升, 证明了使用自适应加权特征拼接的方式, 能够有效减少信息冗余, 降低噪声干扰。

表 4 消融实验结果对比

Table 4 Comparison of ablation results

	模块选择				Accuracy	Precision	Recall	F1-score	AUC
	R_s	R_c	C_d	C_a					
实验 1					0.9048	0.8843	0.9275	0.9054	0.9052
实验 2	√		√		0.9066	0.8887	0.9257	0.9068	0.9069
实验 3	√			√	0.9110	0.8965	0.9250	0.9109	0.9113
实验 4	√	√	√		0.9039	0.8922	0.9149	0.9034	0.9041
实验 5	√	√		√	0.9146	0.8838	0.9511	0.9162	0.9152

3.6 模型优化器和学习率选择分析

为了择优选择模型训练时的优化器, 并为其设置合适的学习率参数, 本文对优化器的种类和学习率的大小进行实验探索。由图 6、7 所示。

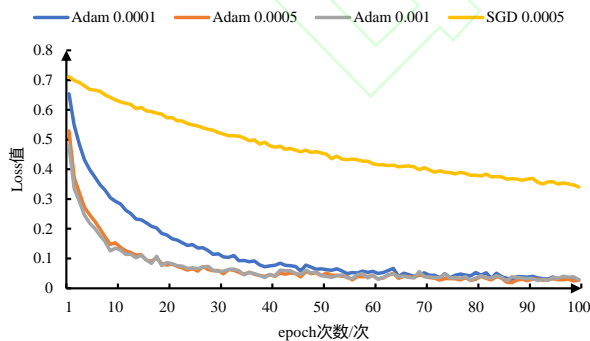


图 6 训练过程中的 Loss 函数变化曲线

Fig.6 The change curve of Loss function during training

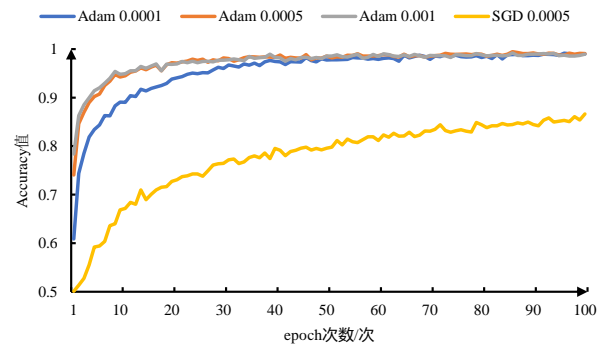


图 7 训练过程中的预测准确率变化曲线

Fig.7 The change curve of Prediction accuracy during training

分析图像可得: SGD 优化器收敛速度较慢, 经过 100 回合训练后, Loss 依然处于 0.4 左右的水平, 而使用 Adam 优化器的 Loss 可以收敛到 0.1 以下。针对模型预测准确率来说, 使用 SGD 优化器经过 100 回合训练达到 0.85 左右的分类精度, 而 Adam 优化器的分类精度则接近于 1.0。使用 Adam 优化器时, 学习率对 Loss 收敛和分类精度的影响不大, 学习率设置为 0.0001 时前期收敛速度较慢, 但后期与 0.0005 和 0.001 的学习率基本持平。而 0.0005 学习率和 0.001 学习率的 Adam 表现几乎一致, 0.001 学习率的模型波动

稍大。

表 5 不同优化器下的 F1 值

Table 5 F1 score of different optimizers

优化器	学习率	F1-score
SGD	0.0005	0.8684
Adam	0.0001	0.9073
Adam	0.0005	0.9162
Adam	0.001	0.9102

从表 5 中可以看到, 在 weibo21 数据集上, 使用 0.0005 学习率的 Adam 优化器可以达到实现最佳测试结果, 这与训练阶段的表现较为一致, 说明 0.0005 学习率下, 模型收敛较为稳定, 泛化能力更好, 因此本文使用 0.0005 学习率的 Adam 进行所有模型的训练。

3.7 模型结构及参数分析

为探究模型结构和参数选择对于模型性能的影响, 以 Weibo21 数据集为例, 设计了相关实验并对结果进行分析。实验具体设置如下:

(1) 实验 1: 将多模态语义特征相似度比较网络中的卷积层的 kernel 从 3 扩大为 5;

(2) 实验 2: 将多模态语义特征相似度比较网络中的卷积层改为全连接层, 引入更多参数;

(3) 实验 3: 取消网络结构中所有的 dropout 层;

(4) 实验 4: 去除分类器中的多个全连接层, 直接利用拼接的多模态特征进行分类。

(5) 实验 5: 提出的 MFAF 模型结构。

实验结果如表 6 所示。

表 6 模型结构及参数分析实验结果

Table 6 Model structure and parameter analysis experimental results

	Accuracy	Precision	Recall	F1-score	AUC
实验 1	0.9101	0.9063	0.9112	0.9088	0.9102
实验 2	0.9119	0.8912	0.9348	0.9125	0.9123
实验 3	0.8923	0.8684	0.9203	0.8936	0.8928
实验 4	0.8737	0.8372	0.9221	0.8776	0.8745
实验 5	0.9146	0.8838	0.9511	0.9162	0.9152

通过对表 6 中的实验结果进行分析, 可以得到以下结论:

(1) 实验 1 中, 通过将多模态语义特征相似度比较网络中的卷积层的 kernel 从 3 扩大为 5 之后, 检测结果除了在精确率上有所提升之外, 其余评价指标的表现均有一定下降, 说明了卷积核的尺寸设置为 3 较为合适;

(2) 实验 2 中, 将多模态语义特征相似度比较网

络中的卷积层改为全连接层, 实验的各项指标均无较大变化, 总体稍差于卷积层的设置, 这说明增加参数量并不能提升网络的预测性能;

(3) 实验 3 中, 取消网络结构中所有 dropout 层的操作在 F1 分数上有了两个点的降低, 这说明 dropout 层可以有效避免网络的过拟合, 提升网络的泛化性能;

(4) 实验 4 中, 去除分类器中的多个全连接层, 直接利用拼接的多模态特征进行分类的方式, 对实验结果的影响最大, 只取得了 0.8776 的 F1 分数, 对比提出的 MFAF 结构性能降低 4.2%。说明在输出分类结果前进行几次非线性变化有助于提升模型的检测效果。

4 结束语

本文针对社交媒体中的虚假新闻检测任务, 综合考虑了新闻的文本信息、新闻文本的情感信息和图像文本语义差异性信息, 并通过提出的自适应加权方法, 构建了多模态特征自适应融合的虚假新闻检测模型, 有效提升了虚假新闻检测效果。通过在多个数据集上进行对比实验、消融实验等, 实验结果表明了本文模型的可行性和有效性。

另外, 本文提出的自适应融合多模态特征的虚假新闻检测模型目前仅从新闻内容入手, 考虑了新闻的文本和图像信息, 后续可尝试加入新闻的社交上下文信息, 例如, 新闻的传播路径、新闻的用户特征和新闻的数据分布等; 如何获取高质量的有标签数据集, 仍是目前虚假新闻检测的主要挑战之一, 未来的工作可以扩展到半监督学习或者无监督学习领域; 此外, 受限于当前的图像描述预训练模型的精度, 导致无法完全正确地描述图像内容, 因此, 如何有效度量图文语义一致性, 将会是未来的另一个研究方向。

参考文献:

- [1] 中国互联网络信息中心. 第 50 次中国互联网络发展状况统计报告[R]. 北京: 中国互联网络信息中心, 2022. China Internet Network Information Center. The 50th statistical report on China's internet development [R]. Beijing: China Internet Network Information Center, 2022.
- [2] MEEL P, VISHWAKARMA D K. Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities[J]. Expert Systems with Applications, 2019, 153(1):112986.
- [3] SATU M S, KHAN M I, MAHMUD M, et al. TClustVID: a novel machine learning classification model to investigate topics and sentiment in COVID-19 tweets[J]. Knowledge-

- Based Systems, 2021, 226: 107126.
- [4] RISH I. An empirical study of the naive Bayes classifier[J]. Journal of Universal Computer Science, 2001, 1(2):127.
 - [5] SAFAVIAN S R, LANDGREBE D. A survey of decision tree classifier methodology[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1991, 21(3): 660-674.
 - [6] NOBLE W S. What is a support vector machine? [J]. Nature Biotechnology, 2006, 24(12): 1565-1567.
 - [7] VOLKOVA S, SHAFFER K, JIN Y J, et al. Separating facts from fiction: linguistic models to classify suspicious and trusted news posts on twitter[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (volume 2: Short papers). 2017: 647-653.
 - [8] POTTHAST M, KIESEL J, REINARTZ K, et al. A stylistic inquiry into hyperpartisan and fake news[J]. ArXiv Preprint ArXiv:1702.05638, 2017.
 - [9] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C]//Proceedings of the 20th International Conference on World Wide Web. 2011: 675-684.
 - [10] CHEN Y, CONROY N J, RUBIN V L. Misleading online content: recognizing clickbait as false news[C]//Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection. 2015: 15-19.
 - [11] FENG S, BANERJEE R, CHOI Y. Syntactic stylometry for deception detection[C]//Meeting of the Association for Computational Linguistics: Short Papers. 2012:171-175
 - [12] JING M, WEI G, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]// International Joint Conference on Artificial Intelligence. 2016.
 - [13] FENG Y, QIANG L, SHU W, et al. A convolutional approach for misinformation identification[C]//Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017: 901-907.
 - [14] CHEN T, LI X, YIN H, et al. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection[C]//Lecture Notes in Computer Science: Trends and Applications in Knowledge Discovery and Data Mining PAKDD 2018 Workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Melbourne, VIC, 2018: 40-52
 - [15] 沈瑞琳,潘伟民,彭成,尹鹏博. 基于多任务学习的微博谣言检测方法[J]. 计算机工程与应用, 2021, 57(24): 192-197.
 - SHEN R L, PAN W M, PENG C, YIN P B. Microblog rumor detection method based on multi-task learning[J]. Computer Engineering and Applications, 2021, 57(24): 192-197.
 - [16] JIN Z, CAO J, ZHANG Y, et al. Novel visual and statistical image features for microblogs news verification[J]. IEEE Transactions on Multimedia, 2017, 19(3):598-608.
 - [17] KE W, SONG Y, ZHU K Q. False rumors detection on sina weibo by propagation structures[C]//IEEE International Conference on Data Engineering. IEEE, 2015:651-662.
 - [18] QI P, CAO J, YANG T, et al. Exploiting multi-domain visual information for fake news detection[C]//IEEE International Conference on Data Mining. IEEE, 2019: 518-527.
 - [19] CAO J, QI P, SHENG Q, et al. Exploring the role of visual content in fake news detection[J]. Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities, 2020: 141-161.
 - [20] 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.
 - [21] KUMARI R, EKBAL A. AMFB: attention based multimodal factorized bilinear pooling for multimodal fake news detection[J]. Expert Systems with Application, 2021(Dec.):184.
 - [22] WANG Y, MA F, WANG H, et al. Multimodal emergent fake news detection via meta neural process networks[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 3708-3716.
 - [23] TUAN N M D, MINH P Q N. Multimodal fusion with BERT and attention mechanism for fake news detection[C]//2021 RIVF International Conference on Computing and Communication Technologies. IEEE, 2021: 1-6.
 - [24] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2019:4171-4186.
 - [25] 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.
 - [26] ZHUANG Y, ZHANG Y. Yet at memotion 2.0 2022: hate speech detection combining bilstm and fully connected layers[C]//Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR. 2022.
 - [27] PENNINGTON J, SOCHER R, MANNING C D. Glove:

- Global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),2014: 1532-1543.
- [28] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv Preprint ArXiv:1409.1556, 2014.
- [29] KHATTAR D, GOUD J S, GUPTA M, et al. Mvae: multi-modal variational autoencoder for fake news detection[C]//The World Wide Web Conference. 2019: 2915-2921.
- [30] 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.
- [31] LI P, SUN X, YU H, et al. Entity-oriented multi-modal alignment and fusion network for fake news detection[J]. IEEE Transactions on Multimedia, 2021, 24: 3455-3468.
- [32] ZHOU X, WU J, ZAFARANI R. Safe: similarity-aware multi-modal fake news detection[C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2020: 354-367.
- [33] TIAN H, GAO C, XIAO X, et al. SKEP: sentiment knowledge enhanced pre-training for sentiment analysis[J]. ArXiv Preprint ArXiv:2005.05635, 2020.
- [34] SHEN D, ZHANG X, HENAO R, et al. Improved semantic-aware network embedding with fine-grained word alignment[J]. ArXiv Preprint ArXiv:1808.09633, 2018.
- [35] NAN Q, CAO J, ZHU Y, et al. MDFEND: multi-domain fake news detection[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021: 3343-3347.