文章编号：1003-0077(2023)09-0131-09

基于一对多关系的多模态虚假新闻检测

袁 玥，刘永彬，欧阳纯萍，田纹龙，方文泷

（南华大学 计算机学院，湖南 衡阳 421001）

摘 要：面向多模态的虚假新闻检测工作大部分是利用文本和图片之间的一对一关系，将文本特征和图片特征进行简单融合，忽略了帖子内多张图片内容的有效特征，对帖子间的语义关联建模不足。为了克服现有方法的局限性，该文提出了一种基于文图一对多关系的多模态虚假新闻检测模型。利用跨模态注意力网络筛选多张图片的有效特征，通过多模态对比学习网络动态调整帖子间高层次的语义特征关联，增强融合图文特征的联合表示。在新浪微博数据集上的实验结果表明，该模型能充分利用文图一对多关系的有效信息和帖子之间的语义特征关系，比基线模型准确率提升了 3.15%。

关键词：虚假新闻检测；跨模态注意力机制；多模态对比学习

中图分类号：TP391 **文献标识码**：A

Multi-modal Fake News Detection Based on the One-to-many Relationship

YUAN Yue, LIU Yongbin, OUYANG Chunping, TIAN Wenlong, FANG Wenlong

(School of Computer, University of South China, Hengyang, Hunan 421001, China)

Abstract: Most of the existing works for multi-modal fake news detection simply fuse textual and image features in a one-to-one manner, while ignoring the information of multiple images in news posts as well as the relationship between different news posts. To overcome these limitations, this paper proposes a model employing the one-to-many relationship of text and images for multi-modal fake newsdetection(OMMFN). In our method, the cross-modal attention network(CMA) is used to extract the effective features of multiple images. Then, the multi-modal contrast learning network(MCL) is used to dynamically adjust the semantic feature relationship between different news posts to improve multi-modal joint feature representation of text and images. Experiments on Sina Weibo dataset show that our model can capture the effective information of text and images with the one-to-many relationship and make full use of the semantic feature relationship between different news posts. The performance in accuracy is improved by 3.15% over the state of the art significantly.

Keywords: fake news detection; cross-modal attention; multi-modal contrast learning

0 引言

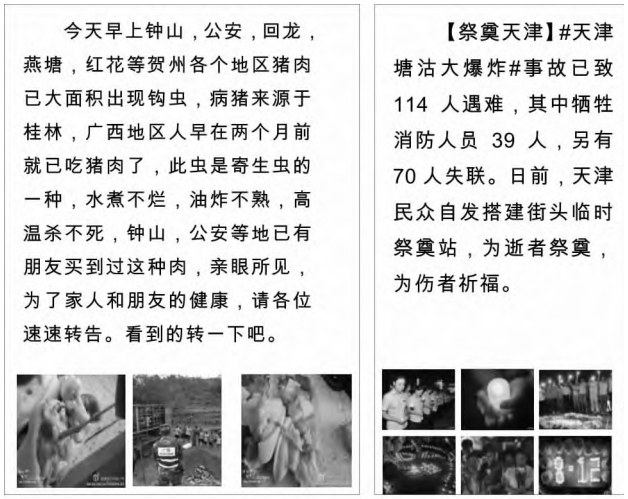
根据多模态的定义[1-2]“不同信息的来源或形式称为不同的模态”，图片、语音、视频、文本等被认为是不同的模态。随着社交网络的迅速发展，新浪微博、短视频等新兴的媒体平台成为了大众获取新闻的主要渠道。以新浪微博平台为例，一条新闻帖子通常包含两种模态即文本模态和视觉模态。本文致力于研究融合文本模态和视觉模态的互补信息，进行虚假新闻检测。

不同于传统媒体时代，新兴媒体平台的用户可以发布或者转载未经核实的帖子，并且在帖子中使用了多张图片作为“证据”，这种带有图片证据的虚假新闻容易被迅速传播，会带来巨大的经济和社会舆论影响。图 1(a)是一条来自新浪微博的新闻帖

收稿日期：2022-10-15 **定稿日期**：2023-02-16

基金项目：国家自然科学基金(N061402220)；湖南省教育厅重点科研项目(19A49)；湖南省自然科学基金(2020JJ4525，2022JJ30495)

子，文本描述了一条关于“猪肉中存在钩虫”的虚假新闻帖子，并配图证明此新闻的真实性，成功误导了大众并被迅速传播[3]。针对这一新闻，当地政府称大众所认为的“钩虫”并非真虫，而是猪的神经纤维、血管等结构，仅仅看起来像虫子。相关研究显示，图片会刺激人大脑中的“伪证据”，其直观性和易读性更能吸引公众的注意力。因此，基于文图的虚假新闻自动检测至关重要。



(a) 猪肉钩虫 (b) 多图片新闻

图1 来自新浪微博数据集的新闻帖子

针对大量的文字和图片混合的虚假新闻，学者们开始关注多模态虚假新闻自动检测。现有的面向多模态的虚假新闻检测方法主要是利用了文图特征的互补性来提升检测性能。例如，EANN[4]，MVAE[5]等方法简单融合了提取的单模态特征，用于虚假新闻分类。随后，研究学者通过研究文图模态间的相似性特征，辅助虚假新闻检测任务。例如，SAFE[6]学习了图片和文本的语义相似性；文献[7-8]学习了图片和文本中实体对的相似性。与单模态相比，这些方法表现出了更好的性能，但它们只是抽取新闻帖子中一张图片的信息作为文本内容的补充。如图1(b)所示，一段新闻文本对应了多张图片，但是若只使用其中一张带有“蜡烛”实体的图片，而在新闻文本中又没有对应的文本实体，则容易误导检测的结果。通过统计新浪微博数据集中每个帖子图片数量占比(表1)，我们发现大多数的新闻帖子内包含多张图片，包含1张图片的帖子仅占比37.8%。因此，如何充分利用视觉信息并挖掘新闻帖子中多张图片的有效特征进行虚假新闻检测是当前亟需解决的一个问题。

表1 新浪微博数据集中每条新闻图片数量统计

图片数量	占比/%
1	37.80
2	9.50
3	7.70
4	7.90
5	2.90
6	6.80
7	1.40
8	1.70
≥9	24.30

最近，一些研究学者还关注到利用一个帖子内的模态交互关系可以增强融合图文特征的联合表示。例如，MFN[9]、HMCAN[10]等方法利用了注意力机制融合帖子内的文图特征，在虚假新闻检测任务上取得了良好的表现。但是这些方法重点关注在一个帖子内融合文图特征，忽略了相同标签新闻帖子之间的关联。

新闻帖子1：【震惊！五个孩子是被割肾?】……让李元龙记者出来！垃圾箱内如何烧东西？还有五个人呢？最先爆料者李元龙失踪了，这又是在掩盖什么？(假新闻)

新闻帖子2：#塘沽爆炸真相#……怎么可能就50人遇难，对面就是居民楼，方圆三十公里都有人，政府以为这样就能瞒得住群众吗？太让人心寒。(假新闻)

我们发现相同真(假)新闻的帖子具有共同特征，假新闻帖子间的语义表达更具有倾向性和主观性。以上例子可以看出，帖子1和中的2的发布者均主观性地虚构了新闻事实，并使用疑问语气渲染气氛放大矛盾。其中，帖子1主观性地认为爆料者在掩盖真相；帖子2倾向性地将问题矛头对准政府。当不明真相的网络用户看到这些帖子，他们的情绪瞬间被带动，并转载不实的帖子。因此，学习真新闻帖子间的共同特征和假新闻帖子间的共同特征，对于增强融合文图特征的多模态表示十分重要。

针对上述问题，我们提出了一个基于文图一对多关系的多模态虚假新闻检测方法。主要贡献如下：

(1) 利用新闻帖子中的一对多关系的文图特征，捕捉文本和视觉内容的完整语义。同时，利用跨

模态注意力网络，增强视觉内容的语义表达。

（2）基于多模态对比学习网络，动态调整帖子间的关联程度，学习真新闻帖子间的共同特征和假新闻帖子间的共同特征，加强融合文本和视觉特征的多模态联合表示。

（3）在新浪微博数据集上的多组对比实验结果表明，本模型同时利用了帖子内和帖子间的多模态特征表示，比现有的多模态虚假新闻检测模型准确率提升大约 3.15%。

1　相关研究

1.1　单模态虚假新闻检测

面向单模态的虚假新闻检测任务侧重于提取文本或图片的单模态语义特征。

基于文本的虚假新闻检测任务主要研究新闻帖子的单词符号和句子表示等文本内容。例如，Ma 等人[11]首次将深度学习应用到文本虚假新闻检测任务中，利用循环网络的隐藏层向量表示句子的语义特征，用于文本分类；Cheng 等人[12]使用变分自动编码器辅助编码文本信息，用于新闻二分类任务，提升了模型的效果。另一方面，基于图片的虚假新闻检测则主要研究虚假新闻中的图片特征。如 Jin 等人[13]利用统计学方法，提取了虚假新闻中图片的特征（图片的清晰度、图片间的相似度），用于虚假新闻的自动检测。Qi 等人[14]设计了一个基于 CNN[15]的模型，利用频域和像素域的视觉信息，获取了假新闻图片在物理和语义层面的特征。由于文本或图片只关注了某一层面的信息，所以限制了单模态虚假新闻检测方法的性能。研究学者发现不同模态间的互补性能够提供更多有效的信息，提高虚假新闻检测效果。

1.2　多模态虚假新闻检测

面向多模态的虚假新闻检测任务侧重于关注模态间的联系，构建更为有效的文图模态特征融合表示。

一部分学者侧重于引入辅助增强功能提升模型的检测效果。例如，Wang 等人[4]提出使用对抗神经网络去除不同事件的特有特征，来学习不同领域新闻的共享特征；Khattar 等人[5]提出利用变分自动编码器来学习多模态表示；Zhou 等人[6]认为文本和视觉信息不匹配的新闻更容易被伪造，从而利用跨模态相似性计算来分析新闻文本和视觉信息之间的相关性。这些方法将不同模态特征进行简单融合，相比利用单模态特征进行虚假新闻检测效果有所提升。所以后续有学者尝试利用深度学习技术来增强各个单模态的表示效果，用于提升多模态联合特征表示。2019 年，Singhal 等人[16]利用 BERT[17]和 VGG19[18]模型分别提取文本和图片特征，再拼接各个单模态特征，用于虚假新闻检测；2020 年，该团队又使用了一种新型的预训练模型 XLNET[19]抽取文本特征，提升了文本模态的表示效果。上述方法关注了增强某一种模态的特征表示来提升虚假新闻检测效果，但是均未考虑不同模态间的特征关联，导致多模态联合特征表示效果不佳。

有部分学者则将问题聚焦于如何利用一个新闻帖子内模态间的关联，提升多模态融合特征表示。例如，Jin 等人[20]首次引入注意力机制融合一个帖子内的多模态特征；Qi 等人[8]提出用实体连接文图信息，并使用多模态协同注意力 Transformer 来对齐文本与视觉模态；张少钦等人[9]利用多头注意力来融合不同模态特征；Qian 等人[10]提出了一种基于层次化的多模态上下文注意模型。以上方法虽然利用了一个帖子内的图文间丰富的分层语义特征及一个帖子内不同模态间的交互关系，取得了不错的效果，但是均忽略了多个新闻帖子间的特征关联。

1.3　对比学习

近年来，对比学习被广泛应用于各大研究领域，如自然语言处理[21-22]、计算机视觉[23-24]领域等。其核心思想是：使得相似样本距离更近，反之距离更远[25]。目前，关于多模态对比学习，现有的方法大多是进行图片和文本模态之间的对比，学习更好的单模态表示。例如，Jia 等人[26]基于对比学习损失，训练模型将匹配的文图对融合，不匹配的文图对分散，用来对齐图片和文本表示；Li 等人[27]提出了一种对比损失，计算图文特征表示的相似性，并动态构造负样本将多模态表示对齐。基于此，我们发现对比学习是解决模态之间的语义特征关联发现的有效方法，可以利用不同帖子的多模态联合表示作为正负样本，进行对比学习，拉大真假新闻帖子样本的差距，从而增强融合文图信息的多模态联合特征表示。

由于新闻帖子中大量存在文图一对多的关系结构特点，基于以上相关研究，本文提出了一个面向虚

假新闻检测的多模态深度学习模型(OMMFN),不仅能够利用帖子内的一对多的文图语义信息,也能利用帖子间的高层次的语义特征关联,增强融合不同模态特征的联合表示。

2 方法

2.1 总体框架

本文提出了一种基于文图一对多关系的多模态虚假新闻检测模型(OMMFN),整体框架如图 2 所示。它由多模态特征提取、多模态增强功能和分类三个模块组成。

(1) 多模态特征提取。该模块由文本编码器和图片编码器组成,文本编码器首先利用 BERT 模型获取最后一层的特征向量,然后利用 CNN 增强模型的泛化能力,将特征向量转化为 32 维,用于表示文本特征;图片编码器首先获取该帖子的多张图片数据,使用 VGG19 模型对每一张图片进行编码,并转化为与文本相同的 32 维的特征向量,最后联合所有图片作为视觉特征表示。

(2) 多模态增强功能。一方面,利用多模态对比学习网络增大正负样本之间的差距,在跨模态注意力前动态调整多模态的特征表示;另一方面,利用跨模态注意力网络寻找文本和图片模态间的联系,赋予每张图片不同的权重,最后将不同模态的特征拼接,得到新闻帖子的多模态表示。

(3) 分类。将多模态表示输入新闻检测器进行二分类预测,检测新闻真假。

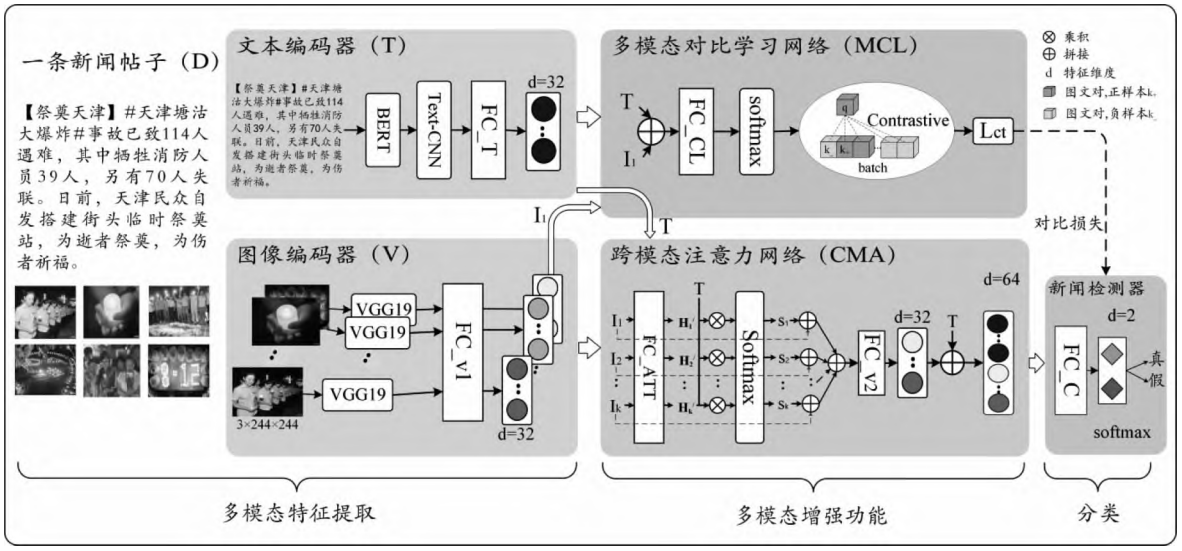


图 2 基于文图一对多关系的多模态虚假新闻检测模型结构图

2.2 多模态特征提取

从新浪微博数据集中获取新闻帖子的多模态数据,用三元组 $D=(T,V,L)$ 表示。其中,T、V、L 分别代表文本、图片和新闻的真假标签。

2.2.1 文本编码器

为了加强语义表示,首先使用 BERT 中文基础模型(BERT-base-Chinese)为新闻帖子中的文本建立一个字序列 $\{a_{i,1},a_{i,2},\cdots,a_{i,200}\}$(最大长度为 200),然后转化为对应的字向量。文本表示如(1)式所示。

$$W=x_1\oplus x_2\oplus\cdots\oplus x_{200} \tag{1}$$

相应的 Mask 值如式(2)所示。

$$W\_\text{Mask}=(m_1,\cdots,m_r,\cdots,m_{200}) \tag{2}$$

其中,$\oplus$ 是串联(Concatenation)运算符,m_r 满足式(3):

$$m_r=\begin{cases}0, & x_r=0\\ 1, & x_r\neq 0\end{cases} \tag{3}$$

由于 BERT 具有超强的特征提取能力,可能使得训练结果陷入局部最优。于是,我们利用了 Text-CNN 的稀疏特性,经 BERT 编码后的特征向量使用不同窗口大小的滤波器,过滤掉一部分噪声,捕捉到不同颗粒度的文本特征,同时为了防止梯度消失同时增强模型的泛化能力,使用了 LeakyReLU 激活函数并随机遮蔽了部分参数,最后,用一个全连接层提取到 32 维的文本特征向量,如式(4)所示。

$$T=e_t=(e_{t1}\oplus\cdots\oplus e_{t32}) \tag{4}$$

2.2.2　图片编码器

使用 VGG19 网络并增加一个全连接层，将特征转换为与文本相同维度的特征序列。则第 j 张图片的特征表示如式(5)所示。

$$I_j = b_{j,1} \oplus b_{j,2} \oplus \cdots \oplus b_{j,32} \tag{5}$$

帖子中的视觉特征表示如式(6)所示。

$$V = \{I_1, \cdots, I_j, \cdots, I_k\} \tag{6}$$

其中，k 为图片数量，$k \in \{1,2,3,4,5\}$。相应的 Mask 值如式(7)所示。

$$V\_\text{Mask} = \{I_1\_\text{Mask}, \cdots, I_j\_\text{Mask}, \cdots, I_k\_\text{Mask}\} \tag{7}$$

I_j _Mask 如式(8)所示。

$$I_j\_\text{mask} = \begin{cases} 0, & \text{位置 } j \text{ 无图片} \\ 1, & \text{位置 } j \text{ 有图片} \end{cases} \tag{8}$$

2.3　多模态增强功能

2.3.1　多模态对比学习网络

在引入多模态对比学习网络之前，简单拼接单模态特征，并归一化文图对，得到帖子 i 的多模态特征，如式(9)所示。

$$q = \text{Softmax}(\text{linear}(T \oplus I_i)) \tag{9}$$

我们发现相同真(假)新闻的帖子具有共同特征。为了使真假新闻帖子更好地被区分，一方面，我们需要拉近相同真(假)新闻帖子的距离；另一方面，增大真假新闻帖子间的差距。而对比学习能够使得相似样本距离更近，使得不相似样本距离更远。

因此，在一个帖子数量为 N 的 batch 内，对于第 i 条真(假)新闻帖子，我们将其他的 $R(>0)$ 个真(假)新闻帖子视为正样本，记为 k_+；将 $M(M=N-R-1)$ 个假(真)新闻帖子视为负样本，记为 k_-。通过数据增强动态地增加一倍的负样本，同时排除与自身的相似度，计算 q_i 与正样本的点乘相似度。对比损失定义如式(10)所示。

$$L_q = -\sum_{k=1}^{R} \log \frac{\exp(q \cdot k_+/\tau)}{\sum_{k=1|[k\neq q]}^{R} \exp(q \cdot k_+/\tau) + \sum_{k=1}^{M} \exp(q \cdot k_-/\tau)} \tag{10}$$

其中，τ 是一个可以学习的温度系数。对比损失表示如式(11)所示。

$$L_{cl} = \frac{1}{R} L_q \tag{11}$$

2.3.2　跨模态注意力网络

为了有效融合图片特征，缓解多张图片带来的噪声问题，首先使用一层全连接层将图片特征转化为与文本对应的维度 $\boldsymbol{H}$，再利用跨模态注意力的打分机制，批量计算帖子中 k 张图片与文本对应的相似度分数 S，如式(12)所示。

$$S = H' \otimes T \tag{12}$$

其中，H' 代表图片特征的转置。重新获取图片的特征向量如式(13)所示。

$$I = I \otimes S \tag{13}$$

根据 I_j _mask 位置标记统计有效的图片数量 g，则过滤后的有效图片特征可表示如式(14)所示。

$$V = I_1 \oplus I_2 \oplus \cdots \oplus I_g, \quad (g \leqslant k) \tag{14}$$

通过全连接层调整视觉特征向量，使得与文本向量维度一致，得到的视觉表示如式(15)所示。

$$V = e_v = (e_{v1}, \cdots, e_{v32}) \tag{15}$$

拼接文本和视觉特征，最终的多模态特征表示如式(16)所示。

$$e = e_t \oplus e_v \tag{16}$$

2.4　虚假新闻检测器

使用 Softmax 分类，将多模态特征向量 e 映射到真实和虚假两类目标空间中，帖子的概率分布如式(17)所示。

$$P = \text{Softmax}(We_i + b) \tag{17}$$

其中，W 代表对应的权重，b 代表偏置项，帖子的概率 P 的取值范围为[0,1]，然后取最大值作为最终的二分类预测类别标签。

在模型训练过程中，选用 Adam 优化器，以及交叉熵函数(Cross Entropy Loss)，虚假新闻预测的分类损失表示如式(18)所示。

$$L_{ce}(y,P) = -\frac{1}{n}\sum_{n}[y\log(P) + (1-y)\log(1-P)] \tag{18}$$

其中，n 代表训练集中的样本总和，y 代表每条帖子的真实类别。最终的损失定义如式(19)所示。

$$L = L_{ce} + L_{cl} \tag{19}$$

3　实验

3.1　实验设置

3.1.1　数据集

本文使用由 Jin 等人[20]构建的新浪微博数据集，该数据集是多模态虚假新闻检测领域的公开数据集。在该数据集中，真实的新闻帖子由中国的官方新闻来源(如新华社)收集，假新闻帖子经过了微博

官方辟谣平台验证。本文将整个数据集划分为训练集、验证集和测试集，具体划分形式如表 2 所示。

表 2 数据集统计信息 （单位：条）

数据集划分	训练集	验证集	测试集
假新闻	2 898	454	756
真新闻	2 517	389	709
总计	5 415	843	1 465

3.1.2 参数设置和评价标准

设置每个帖子的文本最大长度为 200，且最多包含 5 张图片。在模型训练过程，选取 LeakyReLU 为非线性激活函数。根据文献[28]调整参数初始值，参数的设置如表 3 所示。

表 3 参数设置

参数名	数值
d（特征维度）	32
Dropout	0.5
学习率	1e−5
小批量(Batch)	32
迭代次数(Epoch)	30
τ(温度系数)	0.3
Text-CNN(卷积核大小)	[1，2，3，4]×256

本文在训练集和验证集上采取常用的评判指标，即准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F_1 值(F_1-score)，并将测试集上的输出值使用混淆矩阵可视化。

3.2 基线模型

为了验证 OMMFN 的有效性，我们同时对比了单模态和多模态的基线模型。其中，EM-FEND＊模型的结果由文献[7]作者提供，其他模型结果均是在上述实验环境下复现所得。

3.2.1 单模态模型

(1) CNN[15]，利用 CNN 模型提取文本特征，将卷积后的特征分类。

(2) BERT[17]，使用大规模预训练的 BERT 模型，得到的 12 层特征向量代表文本特征，并使用全连接层进行分类。

(3) VGG19[18]，微调 VGG19 模型，生成 32 维的图片特征，进行分类。

3.2.2 多模态模型

(1) EANN[4]，使用 Text-CNN 和预训练的 VGG19 模型分别提取文本和图片特征，将两种特征拼接后送入分类器。

(2) MVAE[5]，分别利用 VGG-19 模型提取视觉模态特征，并使用 Bi-LSTM 提取文本模态特征，将其编码后再重构出原始的单模态特征向量，使用学习到的隐向量来预测新闻是否为假。

(3) SAFE[6]，分别利用 Text-CNN 和 VGG-19 模型来提取文图单模态特征，然后利用余弦相似度计算文图特征的相关性，最后将两种模态的特征拼接作为分类器的输入，用于虚假新闻检测。

(4) MFN[9]，首先利用 FasterRCNN 模型提取图片中多个区域的特征，然后使用多头注意力以及权重拼接简单融合单模态特征，得到包含文图特征的多模态表示。

(5) HMCAN[10]，将利用 BERT 和 ResNet 分别学习到的文图模态表示，输入一个上下文注意网络，捕捉多模态的语义信息来进行新闻检测。

(6) EM-FEND[8]，提取图片中的嵌入文本，并根据视觉实体与文本实体的一致性以及一个 Transformer 为基础的多模态编码器进行建模。

(7) OMMFN-，由于基线模型均使用了一张图片的信息，在 OMMFN 模型的基础上，我们仅保留一张图片进行分类。

3.3 实验结果及分析

3.3.1 主实验

将 OMMFN 模型与上述介绍的所有基线比较，实验结果如表 4 所示。

表 4 不同模型的性能比较

模型	模态	准确率	精确率	召回率	F_1 值
CNN	文本	0.765 2	0.766 5	0.763 5	0.764 0
BERT	文本	0.890 7	0.909 0	0.880 0	0.890 6
VGG19	图片	0.620 0	0.617 0	0.615 0	0.626 0
EANN	文本＋图片	0.792 5	0.795 0	0.790 0	0.790 0

续表

模型	模态	准确率	精确率	召回率	F_1 值
MVAE	文本＋图片	0.815 0	0.828 0	0.806 0	0.823 0
SAFE	文本＋图片	0.796 7	0.795 0	0.820 0	0.790 0
MFN	文本＋图片	0.810 0	0.804 0	0.828 0	0.816 0
HMCAN(BERT-based)	文本＋图片	0.885 0	0.890 0	0.846 0	0.895 0
EM-FEND*(BERT-based)	文本＋图片	0.904 0	0.897 0	0.904 0	0.901 0
OMMFN-(Ours)	文本＋图片	0.932 1	0.931 9	0.924 8	0.924 5
OMMFN(Ours)	文本＋图片	**0.935 5**	**0.935 6**	**0.936 0**	**0.935 5**

从表 4 的对比结果可以得出以下结论：

(1) 在单模态模型中，单一文本模态比单一图片模态的检测效果好。单一图片模态在所有的对比模型中表现最差，其证明了文本能够提供比图片更丰富的特征，而单张图片所含有的特征不足以识别虚假新闻。

(2) 与单模态相比，基于多模态的虚假新闻检测模型大多有更好的表现，表明了多模态特征的互补性能有效提升虚假新闻检测效果。

(3) 使用 BERT 模型提取文本特征来检测虚假新闻的方法均取得了较好的效果，说明 BERT 模型能够有效捕捉文本的语义信息。但是使用了 BERT 模型的 HMCAN 和 EM-FEND 两个方法虽然融合了文图两个模态特征，而效果并没有比使用单一文本模态特征的 BERT 模型具有优势，说明多模态融合过程中视觉模态的特征表示还有待加强。

(4) 我们提出的 OMMFN-和 OMMFN 模型效果表现较为优越，显著超过其他所有基线模型。其中 OMMFN-模型仅在一对一文图关系上使用了对比学习网络，说明对比学习网络能更好地增强多模态的特征表示。OMMFN 模型既考虑了文图一对多关系，也考虑了帖子间的语义关系，说明本文提出的模型确实能够有效捕捉到多幅图片中被忽视的重要特征，而加强图片模态的特征表示确实是提升多模态虚假新闻检测效果的有效途径。

3.3.2　辅助实验

(1) 分析图片数量对虚假新闻检测准确率的影响。如图 3 所示，随着图片数目的增长，虚假新闻检测的准确率不断升高。实验结果表明，图片中的视觉内容可以持续补充虚假新闻检测需要的重要信息，使用五张图片与使用一张图片的视觉信息相比，准确率提升了 2.41%。

(2) 消融实验。在不考虑图片数量对于模型效果影响的条件下，选择性屏蔽不同模块，分析在一对一文图关系上各模块的有效性，结果如表 5 所示。

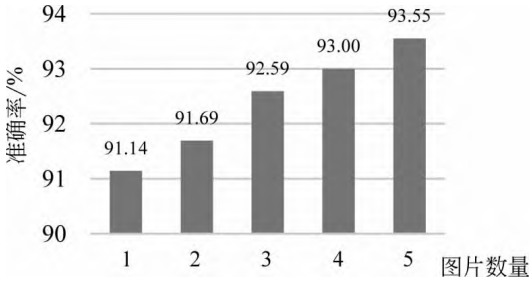


图 3　图片数量对虚假新闻检测准确率的影响

表 5　不同模块的性能分析

方法	准确率	精确率	召回率	F_1 值
-V	0.906 2	0.905 0	0.910 0	0.903 0
-MCL	0.911 4	0.912 7	0.910 6	0.911 1
-CMA	0.920 3	0.915 0	0.910 0	0.920 0
OMMFN	0.932 1	0.931 9	0.924 8	0.924 5

去掉视觉(-V)　在 BERT 模型的基础上增加 Text-CNN 网络，并将提取的文本特征用于虚假新闻检测。

去掉多模态对比学习网络(-MCL)　使用一对一的文图数据，采用注意力机制用于新闻检测。

去掉跨模态注意力网络(-CMA)　使用一对一的文图数据，分析多模态对比学习对新闻分类性能的影响。

实验结果表明，屏蔽模型的任意一个模块，虚假新闻检测的效果都会出现一定程度的降低，说明了模型中各模块的有效性。其中，与直接将 BERT 的输出转化为 32 维相比，-V 准确率提高了 1.55%，说明模型能够利用 Text-CNN 捕捉不同颗粒度的文本特征，增强模型的泛化能力。另外，在使用一对一

关系的图文数据时，去掉 MCL 模块比去掉 CMA 模块的准确率降低 2.07%，说明对比学习能有效提升虚假新闻检测效果。

(3) 在对比学习模型中，温度系数 τ 被用于调节正负样本间的距离，增强样本间的区分度。

为了找到合适的 τ 值，我们将 τ 值设置为 1，0.3，0.07，分别测试其对虚假新闻检测准确率的影响，结果如图 4 所示。

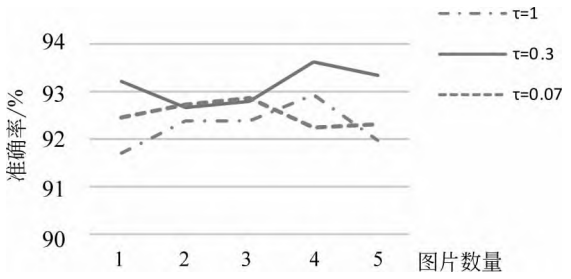


图 4 τ 值对虚假新闻检测准确率的影响

图 4 中，当 τ 值从 1 降低到 0.3 时，模型的准确率不断提高，从 0.3 降低到 0.07 时，模型的准确率开始降低。结果表明，随着 τ 值的减小，模型逐渐加强了对困难样本(即与当前新闻帖子的相似度非常大但真假类别不同的的帖子)的关注度，但是 τ 值的进一步减小使得模型忽略了与大部分帖子的相关性，只关注与最困难样本间的区分度。因此，我们在新浪微博数据集上的实验选取了 $\tau=0.3$。

4 总结

充分利用多模态数据以及帖子间的关系，提取新闻帖子中多张图片的有效特征并增强完整丰富的多模态特征联合表示，是目前面向多模态的虚假新闻检测任务的关键性问题。本文针对以上问题提出了一种基于文图一对多关系的多模态虚假新闻检测模型。该模型首先提取文本和多张图片的特征，通过跨模态注意力网络聚焦于多张图片的有效特征，并利用多模态对比学习网络学习帖子间的特征关联关系，拉大真假新闻之间的差距，动态调整融合后的多模态联合表示，用于虚假新闻检测。在新浪微博数据集上的实验结果表明，该模型能捕捉文图一对多关系的有效信息，提升图片模态的特征表示，学习到不同帖子间的特征关联关系，增强多模态特征表示能力，虚假新闻检测准确率比基线模型提升了 3.15%。

在未来的工作中，我们将考虑如何更有效地利用多种来源的信息提取不同图片的有效特征。另外，随着抖音、微视、快手等短视频平台的发展，基于视频和文本的多模态虚假新闻检测也是未来的研究方向之一。

参考文献

[1] O'HALLORAN K L. Interdependence, interaction and metaphor in multisemiotic texts[J]. Social Semiotics, 1999, 9(3): 317-354.

[2] MORENCY L P, BALTRUSAITIS T. Tutorial on multimodal machine learning [C]//Proceedings of NAACL, 2022: 33-38.

[3] 贺雅文. 从“猪肉钩虫”事件看微博谣言的传播及应对策略[J]. 新闻世界, 2014 (10): 123-124.

[4] 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.

[5] KHATTAR D, GOUD J S, GUPTA M, et al. MVAE: Multimodal variational autoencoder for fake news detection[C]//Proceedings of the World Wide Web Conference, 2019: 2915-2921.

[6] ZHOU X, WU J, ZAFARANI R. Similarity-aware multi-modal fake news detection[C]// Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer International Publishing, 2020: 354-367.

[7] 亓鹏, 曹娟, 盛强. 语义增强的多模态虚假新闻检测[J]. 计算机研究与发展, 2021, 58(7): 1456.

[8] QI P, CAO J, LI X, 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, 2021: 1212-1220.

[9] 张少钦，杜圣东，张晓博，等.融合多模态信息的社交网络谣言检测方法[J].计算机科学，2021，48(05)：117-123.

[10] 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.

[11] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks [C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence. 2016: 3818-3824.

[12] CHENG M, NAZARIAN S, BOGDAN P. Vroc: Variational autoencoder-aided multi-task rumor clas-

sifier based on text[C]//Proceedings of the Web Conference,2020：2892-2898.
[13] JIN Z, CAO J, ZHANG Y, et al. Novel visual and statistical image features for microblogs news verification[J]. IEEE Transactions on Multimedia, 2016, 19(3)：598-608.
[14] QI P, CAO J, YANG T, et al. Exploiting multi-domain visual information for fake news detection[C]//Proceedings of the IEEE International Conference on Data Mining. IEEE, 2019：518-527.
[15] CHEN Y. Convolutional neural network for sentence classification[D]. University of Waterloo, 2015.
[16] SINGHAL S, SHAH R R, CHAKRABORTY T, et al. Spotfake：A multi-modal framework for fake news detection[C]//Proceedings of the 15th International Conference on Multimedia Big Data. IEEE, 2019：39-47.
[17] KENTON J D M W C, Toutanova L K. BERT：Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of NAACL-HLT. 2019：4171-4186.
[18] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]Proceedings of the ICLR 2015：1-14.
[19] SINGHAL S, KABRA A, SHARMA M, et al. Spotfake+：A multimodal framework for fake news detection via transfer learning (student abstract)[C]//Proceedings of the AAAI Conference on Artificial Intelligence,2020, 34(10)：13915-13916.
[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] FANG H, WANG S, ZHOU M, et al. Cert：Contrastive self-supervised learning for language understanding[J]. arXiv preprint arXiv:2005.12766, 2020.
[22] WU X, GAO C, ZANG L, et al. ESimCSE：Enhanced sample building method for contrastive learning of unsupervised sentence embedding[C]//Proceedings of the 29th International Conference on Computational Linguistics. 2022：3898-3907.
[23] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the International Conference on Machine Learning. PMLR, 2020：1597-1607.
[24] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020：9729-9738.
[25] CHEN X, HE K. Exploring Simple siamese representation learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021：15750-15758.
[26] JIA C, YANG Y, XIA Y, et al. Scaling up visual and vision-language representation learning with noisy text supervision[C]//Proceedings of the International Conference on Machine Learning. PMLR, 2021：4904-4916.
[27] LI J, SELVARAJU R, GOTMARE A, et al. Align before fuse：Vision and language representation learning with momentum distillation[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems, 2021：9694-9705.
[28] HE T, ZHANG Z, ZHANG H, et al. Bag of tricks for image classification with convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019：558-567.

袁玥(1998—)，硕士研究生，主要研究领域为自然语言处理和多模态深度学习。
E-mail：uscyuanyue@gmail.com

欧阳纯萍(1979—)，通信作者，博士，教授，主要研究领域为语义网技术、知识图谱、自然语言处理、领域数据分析。
E-mail：ouyangcp@126.com

刘永彬(1978—)，博士，副教授，主要研究领域为自然语言处理和知识图谱。
E-mail：yongbinliu03@gmail.com