CN 43-1258/TP
ISSN 1007-130X

计算机工程与科学
Computer Engineering & Science

第 45 卷第 6 期 2023 年 6 月
Vol. 45, No. 6, Jun. 2023

文章编号:1007-130X(2023)06-1087-10

多层 CNN 特征融合及多分类器混合预测的多模态虚假信息检测*

梁　毅[1,2],吐尔地·托合提[1,2],艾斯卡尔·艾木都拉[1,2]

(1.新疆大学信息科学与工程学院,新疆 乌鲁木齐 830017;2.新疆信号检测与处理重点实验室,新疆 乌鲁木齐 830017)

摘　要:针对现有的多模态虚假信息检测方法很少对多模态特征在特征层面进行融合,同时忽略了多模态特征后期融合作用的问题,提出了一种基于 CNN 多模态特征融合及多分类器混合预测的虚假信息检测模型。首次将多层 CNN 应用于多模态特征融合,模型首先用 BERT 和 Swin-transformer 提取文本和图像特征;随后通过多层 CNN 对多模态特征在特征层面进行融合,通过简单拼接对多模态特征在句子层面进行融合;最后将 2 种融合特征输入到不同的分类器中得到 2 个概率分布,并将 2 个概率分布按比例进行相加得到最终预测结果。该模型与基于注意力的多模态分解双线性模型(AMFB)相比,在 Weibo 数据集和 Twitter 数据集上的准确率分别提升了 6.1%和 4.3%。实验结果表明,所提模型能够有效提高虚假信息检测的准确率。

关键词:虚假信息检测;多模态;后期融合;多层 CNN;多分类器

中图分类号:TP391.1　　**文献标志码**:A

doi:10.3969/j.issn.1007-130X.2023.06.016

Multi-modal false information detection via multi-layer CNN-based feature fusion and multi-classifier hybrid prediction

LIANG Yi[1,2], Turdi Tohti[1,2], Askar Hamdulla[1,2]

(1. School of Information Science and Engineering, Xinjiang University, Urumqi 830017;
2. Xinjiang Key Laboratory of Signal Detection and Processing, Urumqi 830017, China)

Abstract: Aiming at the problem that the existing multi-modal false information detection methods rarely fuse multi-modal features at the feature level and ignore the late fusion effect of multi-modal features, a false information detection method based on CNN multi-modal feature fusion and multi-classifier hybrid prediction is proposed. This method applies multi-layer CNN to multi-modal feature fusion for the first time. The model first uses BERT and Swin-transformer to extract text and image features, and then uses multi-layer CNN to fuse multi-modal features at the feature level. Modal features are fused at the sentence level. Finally, the two fusion features are input into different classifiers to obtain two probability distributions, and the two probability distributions are added proportionally to obtain the final prediction result. Compared with the attention-based multi-modal factorization bilinear model (AMFB), the accuracy of this model is improved by 6.1% and 4.3% on the Weibo dataset and Twitter dataset, respectively. The experimental results show that the proposed model can effectively improve the accuracy of false information detection.

Key words: false information detection; multi-modal; late fusion; multi-layer CNN; multi-classifier

* **收稿日期**:2022-08-31;**修回日期**:2022-10-28
基金项目:国家自然科学基金(62166042,U2003207);新疆维吾尔自治区自然科学基金(2021D01C076);国防科技基金加强计划(2021-JCJQ-JJ-0059)
通信作者:吐尔地·托合提(turdy@xju.edu.cn)
通信地址:830017 新疆乌鲁木齐市水磨沟区华瑞街 777 号新疆大学信息科学与工程学院
Address: School of Information Science and Engineering, Xinjiang University, 777 Huarui Street, Shuimogou District, Urumqi 830017, Xinjiang, P. R. China

1 引言

随着社交网络的发展,信息的传播速度飞快地提升,在方便人们进行社交的同时也为虚假信息的扩散提供了一定的便利[1,2]。虚假信息严重损害了媒体的公信力,侵害了大众的知情权、参与权和监督权,严重时还会扰乱社会秩序,造成人们的财产损失,引发公众的恐慌,对社会造成严重的不良影响。例如,2011 年因日本福岛核泄漏事件而产生的含碘食物可预防核辐射的谣言,该谣言使得大家疯狂抢购食盐,不仅造成资源浪费还扰乱了社会的安稳。因此,如何在早期对虚假信息进行检测成为了近期的研究热点。

早期的信息多为纯文本的形式,因此早期的方法主要是通过从文本内容中提取文本特征来对虚假信息进行检测[3,4]。随着社交媒体的发展,信息已经从纯文本的形式向多媒体的形式转变[5-7],现有的信息多为多媒体的形式,同时有研究发现,将不同模态的特征进行融合能有效提升检测效率。因此,最近在虚假信息检测领域的研究以多模态的方法为主。然而现有的多模态方法存在一定的局限性。首先,常用的融合方法有简单拼接、注意力机制、双线性池和自编码器 4 种,其中,简单拼接、双线性池和自编码器是对多模态特征在句子层面的融合,注意力机制是对多模态特征在单词层面的融合,现有的模型仅仅对多模态特征进行句子层面的融合或者单词层面的融合,没有同时从 2 个层面对多模态特征进行融合,并且没有对多模态特征在特征层面进行融合。其次,对于文本进行特征提取,大多数依赖于双向门控循环单元 Bi-GRU(Bi-Gated Recurrent Unit)每一个时间步的拼接输出,然而,由于特征提取过程缺少相应事实知识的参与,这类方法对帖子文本中命名实体的理解能力有限,进而难以充分捕捉虚假信息语义层面的线索[8]。图像特征的提取多依赖于 VGG19(Visual Geometry Group 19)[9],该方法需要的参数量较大,训练时需要大量的计算资源。同时,通过这 2 个模型提取到的特征与现有一些模型相比质量较低。

本文的主要工作如下所示:

(1)同时对多模态的特征在句子和特征层面进行融合,并且探究了新的多模态特征在特征层面的融合方法。使用 CNN(Convolutional Neural Network)对多模态特征在特征层面进行融合,并证明使用 CNN 来对多模态特征进行融合是有效的。

(2)针对从单模态信息中提取出的单模态特征的质量较低的问题,本文在文字特征提取方面使用 BERT(Bidirectional Encoder Representations for Transformers)[10]来代替 Bi-GRU。BERT 已经在多个领域证明了其在该方面的有效性。在图像特征提取方面使用 SWTR(Shifted Window TransformeR)[11]来代替 VGG19。SWTR 一经发布就在目标检测、实例分割和语义分割等多种任务上取得了良好的表现。本文将 SWTR 应用于多模态的虚假信息检测任务,将其作为图像特征提取模块,实验结果表明,SWTR 在虚假信息检测领域也具有很好的表现。

(3)提出了一种基于 CNN 的多模态融合模型,并在中文和英文 2 种语言的数据集上进行了对比实验及消融实验,实验结果表明了该模型在虚假信息检测方面的有效性。

2 相关工作

虚假信息被定义为未经过证实或故意捏造的故事或陈述[12]。多模态数据是指对同一个对象不同角度的描述,每一个角度的描述就是一个模态[13]。本文按照所使用的模态的数量将目前的方法分为 2 类:基于单模态的方法和基于多模态的方法[14]。

2.1 基于单模态的方法

内容特征指的是可以直接从发布的信息中提取的特征。内容特征主要包括文本特征和视觉特征[15]。

在早期的研究中,人们对于文本内容的处理主要通过手动的方式来提取文本特征并使用机器学习的分类算法来对虚假信息进行检测。Pérez-Rosas 等[16]基于语言特征构建了假新闻检测器。作者从文本中提取 N_grams、标点符号、心理语言学特征、表明文本可理解性的特征和语法特征,最后使用支持向量机 SVM(Support Vector Machine)分类器和 5 次交叉验证来进行检测。Kwom 等[17]利用结构信息和语言特征,使用支持向量机、随机森林分类器和决策树 3 种分类模型对谣言传播的多峰现象进行捕捉。Jin 等[18]证明了图像在假新闻检测中的重要性。由于缺乏专业知识,传统的基于机器学习的虚假信息检测模型难以获得手工特征。

随着深度学习的广泛应用,Ma等[19]使用了循环神经网络来获取文本的隐藏特征,使用该特征来对谣言进行检测。Qi等[20]通过注意力机制动态融合图像频域和像素域的特征进行虚假信息检测。虽然深度学习能自动提取特征,但是其经常受到噪声影响,导致其模型性能降低。受生成对抗网络GAN(Generative Adversarial Network)的启发,Ma等[21]提出了一种基于GAN网络的模型,通过对抗训练可以去除噪声和无关特征,获得判别性更强的特征。

虽然单独使用文本特征或者视觉特征对虚假信息进行检测已经被证明是有效的,但是随着多媒体时代的到来,信息的存在形式已经由过去的纯文本形式转变为同时包含文本、图像和视频等多模态数据的形式。因此,如果仅仅从文本或图像的角度来对虚假信息进行检测,不仅信息利用率和检测准确率较低,同时很难适应已经到来的多媒体时代。

2.2　基于多模态的方法

现有的基于多模态的方法大多使用文本和图像这2种模态信息来检测虚假信息。Singhal等[22]利用BERT提取文本特征,使用VGG19提取图像特征,将其进行拼接作为融合特征进行虚假信息检测。Kumari等[23]提出了一个基于多模态分解双线性池的多模态融合模型,能解决不同特征简单拼接无法确定特征边界以及无法发现图像和文本特征表示之间的相关性的问题。

虽然上述研究都在融合模块提出了不同的想法,但是输入都只有文本和图像这2种模态,忽略了其余模态的作用。Giachanou等[24]在从文本中提取特征的同时还从文本中提取出情感,在提取图像特征时还提取出图像的标签,将图像的标签和文本特征进行相似度计算。通过上述方法,能从文本和图像中获得更多的信息,增加信息的利用率。但是,其仅仅是通过简单的拼接进行特征融合,未能充分考虑不同模态之间的相关性。亓鹏等[8]提出了语义增强的多模态虚假信息检测,通过卷积网络获得图像的特征、标签和图中的文字,并通过将提取的图中文字添加到文本信息中的方式来对文本信息进行更新。图标签、原文字信息和更新后的文字信息通过艾尼ERNIE(Enhanced Representation from kNowledge IntEgration)[25]来提取特征。使用从原文本中提取的特征,通过注意力机制来更新图标签和图的特征向量。最后将更新后的文字特征、更新后的图标签特征和更新后的图像特征进行拼接作为最后的融合特征进行检测。

孟杰等[14]提出了一个基于注意力机制的多模态融合模型,该模型首先使用Bi-GRU提取语义特征,使用多分支卷积-循环神经网络CNN-RNN(Convolutional-Recurrent Neural Network)提取图像不同层次的特征;然后采用模间注意力和模内注意力来融合多模态信息,获得融合特征;最后,通过注意力机制将文本特征、图像特征和融合特征进行融合,增强原信息的作用。

上述研究都是在提取出文本和图像的特征后,对多模态的特征在句子层面或单词层面进行融合,以获得多模态特征的融合特征,未能同时对多模态特征在2个层面进行融合,并且没有对多模态特征在特征层面进行融合。本文同时在句子和特征层面对多模态特征进行融合,同时对新的特征层面融合方法进行研究,发现多层CNN的模型基于其共享卷积核的特点使其能轻松地对高维数据进行处理,并随着残差网络的提出,多层CNN模型的退化问题也在一定程度上得以解决,其在计算机视觉的多个不同的任务上都获得了很好的表现。由于多模态特征融合是对高维数据进行处理,同时需要深度神经网络来提取深层的融合特征,因此本文对多层CNN模型进行了适当的修改使其可以在多模态任务上使用,同时使用后期融合的方法来应对池化层会丢失大量信息的问题。本文提出了一种基于CNN的多模态虚假信息检测方法。

3　基于多层CNN的虚假信息检测方法

问题定义:假设$H=\{h_1,h_2,h_3,\cdots,h_m\}$是一个有关社交网络多模态帖子的数据集,其中$h_i$代表第$i$个帖子。$T=\{t_1,t_2,t_3,\cdots,t_m\}$为文本集合,$t_i$代表第$i$个帖子中的文字内容。$V=\{v_1,v_2,v_3,\cdots,v_m\}$为图像集合,$v_i$代表第$i$个帖子中的图像。$L=\{l_1,l_2,l_3,\cdots,l_m\}$为标签集合,$l_i$是第$i$个帖子的标签。虚假信息检测的任务可以描述成学习一个函数$f(T,V)=Y$,$Y=\{y_1,y_2,y_3,\cdots,y_m\}$,$y_i$为第$i$个帖子预测的标签值,$y_i\in\{0,1\}$,0代表真实信息,1代表虚假信息。

本文模型主要由4个部分组成:文本特征提取、图像特征提取、特征融合和分类器。图1所示为本文提出的多层CNN特征融合及多分类器混合预测的多模态虚假信息检测方法。本文模型对多模态的特征进行了早期融合和后期融合,早期融合是在特征提取后对多模态特征进行融合。后期融合是在做出决策后再进行融合,即对不同分类器

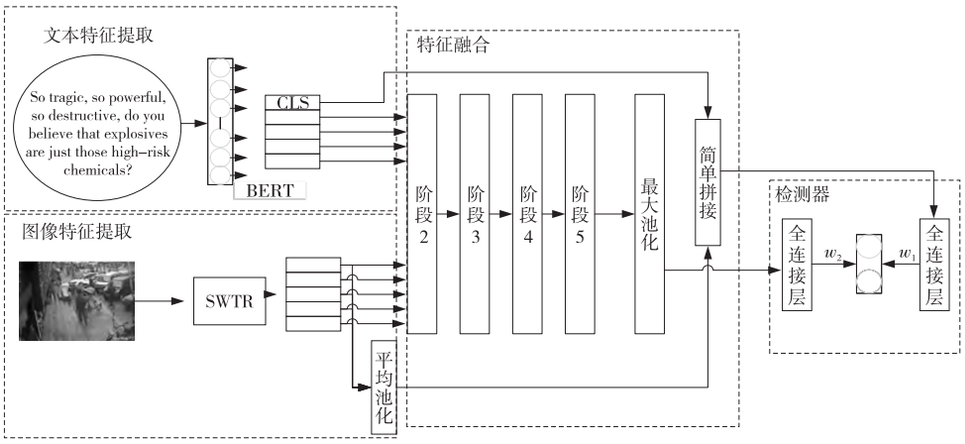

Figure 1 Overall framework of the model

图 1 模型总体框架

的预测结果进行融合。图 2 所示为本文提出的基于多层 CNN 的早期融合模块。

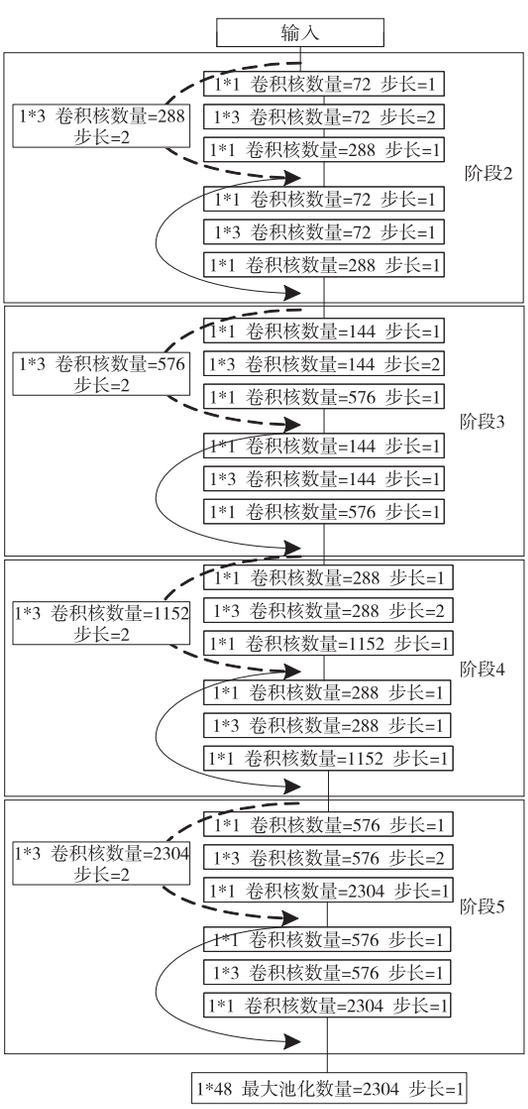

Figure 2 Fusion module

图 2 融合模块

模型首先使用 BERT 和 SWTR 提取出文本和图像特征;随后将除[CLS]位置以外的文本特征和图像特征输入基于多层 CNN 的特征融合模块进行特征层面的融合,得到特征层面的融合特征,并将文本特征中[CLS]位置的特征和经过平均池化后的图像特征进行简单拼接得到句子层面的融合特征;随后将 2 种融合特征输入不同的分类器得到 2 种预测结果;最后将 2 个预测结果按照不同的权重进行相加得到最终的预测结果。

3.1 文本特征提取

文本是帖子的主要组成部分,包含了发布者想要表达的主要信息,同时也能展现出发布者的情感等信息。因此,本文采用具有强大建模能力的 BERT 预训练模型来对这一重要部分进行处理。

BERT 是一种基于 Transformer[26] 的预训练模型,其特别的无监督任务使得其可以学习到上下文信息,其内部只使用了 Self-Attention 机制而没有使用 RNNs(Recurrent Neural Networks),使得其可以通过并行计算来加速训练过程,并且在大规模预训练语料上学习到了某些句法知识和常识知识。在最近的一些类似任务中,通常基于 BERT 的模型比一些基于 RNNs 和 CNN 搭建的网络有更好的表现,这些实验表明了使用 BERT 可以提取出更高质量的文本特征。因此,本文使用 BERT 来提取文本特征。

BERT 是将文本以‘[CLS]’+句子+‘[SEP]’的形式输入,输出是一组单词的嵌入表示。文本特征计算如式(1)所示:

$$\boldsymbol{T}_i = Bert(t_i) \tag{1}$$

其中,t_i 为一段文字的集合,$\boldsymbol{T}_i = \{\boldsymbol{f}_{\text{cls}}^{\text{t}}, \boldsymbol{f}_1^{\text{t}}, \boldsymbol{f}_2^{\text{t}}, \cdots, \boldsymbol{f}_n^{\text{t}}\}$,$\boldsymbol{T}_i \in \mathbf{R}^{(n+1)*d_i}$ 为文本的特征向量,n 为一个句子中单词的个数,d_i 为每个单词特征向量的维

度。本文将[CLS]标记位以外的所有输出作为局部特征融合模块的输入,即将 $\boldsymbol{T}_{i-1}=\{\boldsymbol{f}_1^{\mathrm{t}},\boldsymbol{f}_2^{\mathrm{t}},\cdots,\boldsymbol{f}_n^{\mathrm{t}}\}$ 输入到局部特征融合模块。

3.2　图像特征提取

图像有对文本信息进行补充、增加内容可信度的作用。对图像部分进行处理能更充分地理解多模态帖子的语义,从而更好地对虚假信息进行检测。本文采用 SWTR 来对图像部分进行处理。

SWTR 也是一种基于 Transformer 的图像特征提取模型。该模型扩展了 Transformer 的适用性,将 Transformer 的高性能迁移到视觉领域,解决了 CNN 对于全局信息特征提取的不足,同时凭借其特有的窗口机制大大降低了自注意力的计算复杂度,解决了 token 尺度固定的问题,现已成为计算机视觉的通用模型。自提出以来在图像分类和分割等任务上都取得了比较好的结果。因此,本文使用 SWTR 来提取图像特征,SWTR 的输入为一幅图像,输出为该幅图像的特征向量。本文删除 SWTR 中最后的平均池化层,将未经平均池化层处理的特征输入局部特征融合模块。具体计算过程如式(2)所示:

$$\boldsymbol{V}_i=SWTR(\boldsymbol{I}_i) \tag{2}$$

其中,$\boldsymbol{I}_i$ 为第 i 个帖子包含的一幅图像,$\boldsymbol{V}_i=\{\boldsymbol{f}_1^{\mathrm{v}},\boldsymbol{f}_2^{\mathrm{v}},\cdots,\boldsymbol{f}_z^{\mathrm{v}}\}$,$\boldsymbol{V}_i\in\mathbf{R}^{z*d_i}$ 为图像经过 SWTR 提取出的特征向量,z 为提取的特征数量,d_i 为特征向量的维度。

3.3　特征层面的融合

至此获得了文本特征 $\boldsymbol{T}_{i-1}$ 和图像特征 $\boldsymbol{V}_i$。为了得到文本和图像之间的特征层面的融合关系,本文使用 CNN 网络来对文本和图像的特征进行融合,以得到多模态的特征层面的融合特征。CNN 网络可以利用不同的卷积核从不同的角度来对 2 种模态的深层特征进行融合。该模块通过使用残差网络来防止深度神经网络的退化。与传统的用于文本的 Text-CNN 不同,本文所使用的 CNN 为传统的一维卷积。Text-CNN 通过固定卷积核的长度使得每次卷积都是对几个完整的词向量进行卷积,这样可以在单词层面进行融合,但在更深层的特征层面却无法进行充分的融合,同时图像特征不存在词向量的概念,因此本文采用传统的一维卷积来对多模态的特征在更深层的特征层面进行融合。

该模块首先将文本和图像的特征进行拼接;随后依次通过阶段 2、阶段 3、阶段 4 和阶段 5 这 4 个卷积块来对特征进行融合;接着经过一层最大池化层得到最终的融合特征,如图 2 所示。上述过程的公式表达如式(3)和式(4)所示:

$$\boldsymbol{F}_1=concate(\boldsymbol{T}_{i-1},\boldsymbol{V}_i) \tag{3}$$

$$\boldsymbol{F}_{\mathrm{part}}=Fusion\_Module(\boldsymbol{F}_1) \tag{4}$$

其中,$Fusion\_Module(\cdot)$ 代表图 1 中的特征融合模块。

本文在搭建 CNN 融合网络时借鉴 ResNet50 网络,每个阶段都包括 2 个部分,对这 2 个部分要采取不同的方法进行残差连接。如图 2 所示,在进行虚线部分的残差连接时,需要先将输入的特征向量通过一层卷积核来改变其维度,随后与输出的特征向量按维度进行相加来实现残差连接。其具体计算过程如式(5)~式(9)所示:

$$\boldsymbol{C}_1=relu(batch\_Norm(Conv(\boldsymbol{F}))) \tag{5}$$

$$\boldsymbol{C}_2=relu(batch\_Norm(Conv(\boldsymbol{C}_1))) \tag{6}$$

$$\boldsymbol{C}_3=relu(batch\_Norm(Conv(\boldsymbol{C}_2))) \tag{7}$$

$$\boldsymbol{C}'_1=relu(batch\_Norm(Conv(\boldsymbol{F}))) \tag{8}$$

$$\boldsymbol{C}_{\mathrm{out}}=Add(\boldsymbol{C}_3,\boldsymbol{C}'_1) \tag{9}$$

在对实线部分进行残差连接时将输入的特征向量直接与输出的特征向量按维度进行相加。具体计算过程如式(10)~式(13)所示:

$$\boldsymbol{D}_1=relu(batch\_Norm(Conv(\boldsymbol{F}))) \tag{10}$$

$$\boldsymbol{D}_2=relu(batch\_Norm(Conv(\boldsymbol{D}_1))) \tag{11}$$

$$\boldsymbol{D}_3=relu(batch\_Norm(Conv(\boldsymbol{D}_2))) \tag{12}$$

$$\boldsymbol{D}_{\mathrm{out}}=Add(\boldsymbol{D}_3,\boldsymbol{F}) \tag{13}$$

其中,$Conv(\cdot)$ 代表卷积操作,$Add(\cdot)$ 代表按维度进行相加,$\boldsymbol{F}$ 代表输入特征。

本文在每个阶段的第 2 个部分,即图 2 中的实线部分,会重复卷积多次,具体设置在 4.2 节中展示。本文在此处的设置与 ResNet 网络不同,本文每个阶段的第 2 个部分在进行多次重复卷积时共用一组卷积核。在每个阶段,首先通过一组卷积核将输入特征的维度降低一半,随后通过另外一组卷积核对其重复卷积 2 次。这样设置可以通过减少参数量来防止过拟合问题的发生,从而得到更好的训练效果。

3.4　句子层面的融合

由于 BERT 内部使用了 Self-Attention 机制,如果使用有具体含义的词来代表整个句子,这个词向量就会受到该词本身的影响,难以客观表示整个句子的特征,[CLS]是一个标记位不包括任何语义信息,所以能更好地作为整个文本的特征。对提取出的图像特征进行平均池化和展平后,用其代表整个图像的特征。

将文本整体特征和图像整体特征进行拼接得到

句子层面的融合特征。通过简单拼接来对多模态特征进行句子层面的融合,可以有效减少融合过程中的信息损失。公式表达如式(14)~式(16)所示:

$$\boldsymbol{V}_{\text{averagepool}}=averagepool(\boldsymbol{V}_i) \tag{14}$$

$$\boldsymbol{V}_{\text{entirety}}=flatten(\boldsymbol{V}_{\text{averagepool}}) \tag{15}$$

$$\boldsymbol{F}_{\text{entirety}}=concat(\boldsymbol{f}^{\text{t}}_{\text{cls}},\boldsymbol{V}_{\text{entirety}}) \tag{16}$$

3.5 虚假信息检测

将句子层面的融合特征向量输入激活函数为Softmax的全连接层得到检测结果,如式(17)所示:

$$\boldsymbol{P}_{\text{entirety}}=softmax(Linear(\boldsymbol{F}_{\text{entirety}})) \tag{17}$$

将经过CNN网络融合后的特征层面的特征向量通过一层激活函数为Softmax的全连接层处理得到检测结果。随后将通过句子层面融合特征得到的预测结果和通过特征层面融合特征得到的预测结果按不同的权重进行相加,得到最终的预测结果。公式表达如式(18)~式(20)所示:

$$\boldsymbol{P}_{\text{part}}=softmax(Linear(\boldsymbol{F}_{\text{part}})) \tag{18}$$

$$\boldsymbol{P}_{\text{final}}=W_1\boldsymbol{P}_{\text{entirety}}+W_2\boldsymbol{P}_{\text{part}} \tag{19}$$

$$y_i=\text{argmax}(\boldsymbol{P}_{\text{final}}) \tag{20}$$

其中,y_i 为第 i 个帖子通过模型处理后预测的标签值,W_1 和 W_2 分别代表不同模型对最终检测结果的影响程度。

不同融合层面的模型对最终检测结果的影响程度是不一样的,因此需要根据不同模型的影响程度进行加权求和来获得最终的检测结果。

损失函数定义为预测概率分布和真实标签之间交叉熵损失函数,如式(21)所示:

$$L=-\sum_{i=1}^{m}[l_i\log p_i+(1-l_i)\log(1-p_i)] \tag{21}$$

其中,m 为帖子的个数;$l_i\in\{0,1\}$为真实标签值,1表示虚假信息,0表示真实信息;p_i 表示预测为虚假信息的概率。

4 实验与结果分析

4.1 数据集

本文使用2个公开的数据集进行实验,即Twitter[27]和Weibo[16]。这是研究人员设计的高质量的用于多模态虚假信息检测的2个数据集。为了与之前的工作进行比较,在这2个公开的数据集上训练本文模型。

Twitter数据集是由Boididou等(2015)[27]发布的。该数据集由训练集和测试集2部分组成,包含文本、相关图像和上下文信息。测试集不包含幽默类别,所以忽略这部分的信息[11]。

Weibo数据集是一个多模态中文数据集,收集了2012年5月~2016年6月的Weibo。为了保证数据集的质量,Jin等[18]已经删除了非常小的和重复的图像,以及不包含图像的帖子,使其在本质上完全为多模态信息。

表1所示为2个数据集的完整数据分布。为了方便与基准模型进行对比,本文对文本和图像进行了相同的预处理。对于文本部分删除句子中的标点符号、URL及表情,对于图像部分使所有图像的大小相等。

Table 1 Datasets used in the experiment

表1 实验中所使用的数据集

Dataset	Train		Test		Image
	False	Real	False	Real	
Twitter	6 841	5 009	2 564	1 217	410
Weibo	3 784	3 783	1 000	996	13 274

4.2 实验设置

本文实验的硬件配置和软件环境:CPU:Intel Xeon Gold 6130H,内存为63 GB,内核为8核,GPU:NVIDIA GeForce RTX 3080Ti,PyTorch(1.7.1),Python(3.8),Cuda(10.2)。

本文的融合模块具体参数设置如表2所示。表3列出了用于训练模型的所有超参数。其中,s 为步长,lr 为学习率。

Table 2 Parameter setting of CNN fusion module

表2 CNN融合模块参数设置

模块名	输出尺寸	
阶段2	1*384	$\begin{cases}1*1,s=1,72\\1*3,s=2,72*1\\1*1,s=1,28\end{cases}$ $\begin{cases}1*1,s=1,72\\1*3,s=1,72*2\\1*1,s=1,288\end{cases}$
阶段3	1*192	$\begin{cases}1*1,s=1,144\\1*3,s=2,144*1\\1*1,s=1,576\end{cases}$ $\begin{cases}1*1,s=1,144\\1*3,s=1,144*3\\1*1,s=1,576\end{cases}$
阶段4	1*96	$\begin{cases}1*1,s=1,288\\1*3,s=2,288*1\\1*1,s=1,1152\end{cases}$ $\begin{cases}1*1,s=1,288\\1*3,s=1,288*5\\1*1,s=1,1152\end{cases}$
阶段5	1*48	$\begin{cases}1*1,s=1,576\\1*3,s=2,576*1\\1*1,s=1,2304\end{cases}$ $\begin{cases}1*1,s=1,576\\1*3,s=1,576*2\\1*1,s=1,2304\end{cases}$
最大池化	1*1	1*48

Table 3　Model super parameter setting

表 3　模型超参数设置

Parameter	Twitter	Weibo
Text length	33	95
Image size	(224,224,3)	(224,224,3)
Batch size	70	70
Optimizer	Adam(lr=0.0001)	Adam(lr=5e−5)
Epochs	100	100
W_1	0.26	0.573
W_2	0.74	0.427

4.3　基准模型

与同一个数据集上的以下不同多模态方法的预测结果进行比较,以验证本文模型的有效性。

(1) att-RNN[28]:att-RNN(Recurrent Neural Network with an attention mechanism)使用带有注意力机制的 RNN,融合文本、图像和社交上下文特征,以实现谣言检测。

(2)EANN[29]:EANN(Even Adversarial Neural Network) 有 3 个主要组件:多模态特征提取器、虚假信息检测器和事件鉴别器。模型中,特征提取器会在事件鉴别器的帮助下提取事件的文本和视觉特征,然后将得到的文本和视觉特征进行拼接,最后使用虚假信息检测器检测新闻帖子的真假。

(3) MVAE[30]:MVAE(Multimodal Variational AutoEncoder)通过训练 3 个子网络来检测假新闻。在这里,训练一个变分自动编码器,以获得更好的文本和视觉特征表示。共享潜在特征被进一步用于分类。

(4)AMFB[23]:AMFB(Attention based Multimodal Factorized Bilinear)网络由 3 部分组成:特征提取模块、特征融合模块和虚假信息检测模块。该模型利用多模态分解双线性池来融合文本和图像的特征。

(5)OUR:本文所提出的虚假信息检测模型,该模型由特征提取器、图像特征提取器、早期融合模块和后期融合模块组成。该模型的后期融合模块将 3 个分类器得到的概率分布按比例进行相加,得到最终的概率分布,并以相加后的概率分布来确定结果。

4.4　实验结果分析

基准模型和本文模型在 2 个数据集上的实验结果如表 4 所示。评价指标为准确率 *Accuracy*、正确率 *Precision*、召回率 *Recall* 和 *F*1 分数。实验结果表明,本文模型在 Weibo 数据集和 Twitter 数据集上的准确率均优于基准模型的。

在 Twitter 数据集上,本文模型的准确率比 AMFB 的准确率高 4.3%,比 MVAE 的高 18.1%,比 EANN 的高 18.5%。在 Weibo 数据集上,本文模型的准确率比 AMFB 的高 6.1%,比 MVAE 的高 6.9%,比 EANN 的高 10.2%。

4.5　消融实验

本文设置了消融实验来进一步验证所提出的模型的有效性。

通过对删除不同的模块后的模型进行对比实验来验证各个模块的有效性,并在保持其他模块不变的情况下,与未复用卷积核的多层 CNN 融合模型进行对比,以验证复用卷积核的有效性,其结果如表 5 和表 6 所示。

Table 4　Experimental results on two datasets

表 4　2 个数据集上的实验结果

Dataset	Model	*Accuracy*	Fake News			Real News		
			Precision	*Recall*	*F*1-*Score*	*Precision*	*Recall*	*F*1-*Score*
Twitter	att-RNN	0.664	0.749	0.615	0.676	0.589	0.728	0.651
	EANN	0.741	0.690	0.550	0.610	0.760	0.850	0.810
	MVAE	0.745	0.801	0.719	0.758	0.689	0.777	0.730
	AMFB	0.883	0.890	0.950	0.920	0.870	0.760	0.810
	OUR	0.926	0.979	0.901	0.939	0.853	0.968	0.907
Weibo	att-RNN	0.772	0.797	0.713	0.692	0.684	0.840	0.754
	EANN	0.791	0.840	0.720	0.780	0.760	0.860	0.800
	MVAE	0.824	0.854	0.769	0.809	0.802	0.875	0.837
	AMFB	0.832	0.820	0.860	0.840	0.850	0.810	0.830
	OUR	0.893	0.880	0.919	0.899	0.909	0.866	0.887

Table 5 Ablation experimental results on Weibo datasets

表 5 在 Weibo 数据集上的消融实验

Model	*Accuracy*	Fake News			Real News		
		Precision	*Recall*	*F1-Score*	*Precision*	*Recall*	*F1-Score*
CNN_FM	0.893	0.880	0.919	0.899	0.909	0.866	0.887
Text	0.868	0.841	0.917	0.878	0.902	0.815	0.856
Visual	0.718	0.767	0.651	0.704	0.679	0.789	0.730
Delete_CNN	0.890	0.870	0.927	0.897	0.916	0.851	0.882
Delete_concate	0.884	0.882	0.895	0.888	0.886	0.872	0.878
ResNet_Fusion	0.879	0.860	0.915	0.887	0.903	0.841	0.871

Table 6 Ablation experimental results on Twitter datasets

表 6 在 Twitter 数据集上的消融实验

Model	*Accuracy*	Fake News			Real News		
		Precision	*Recall*	*F1-Score*	*Precision*	*Recall*	*F1-Score*
CNN_FM	0.926	0.979	0.901	0.939	0.853	0.968	0.907
Text	0.831	0.944	0.777	0.852	0.709	0.922	0.802
Visual	0.834	0.849	0.895	0.871	0.804	0.731	0.766
Delete_CNN	0.880	0.915	0.890	0.903	0.823	0.860	0.841
Delete_concate	0.923	0.979	0.897	0.936	0.847	0.968	0.904
ResNet_Fusion	0.855	0.865	0.910	0.887	0.833	0.760	0.795

其中,(1)CNN_FM:包含所有模块;(2)Text:仅用文本特征来对信息进行检测;(3)Visual:仅用图像特征来对信息进行检测;(4)Delete_CNN:将局部融合模块去掉,仅进行整体融合;(5)Delete_concate:将整体融合去掉,仅进行局部融合;(6)ResNet_Fusion:使用传统的多层 CNN 来进行融合,在进行重复卷积时,每次重复卷积时都会采用一组新的卷积核,不共用一组卷积核。

从表 5 和表 6 中可以看出,模型在去掉任何一个模块后,性能均有下降。Text 模型与原模型相比,在 Weibo 数据集上准确率降低了 2.5%,在 Twitter 数据集上的准确率降低了 9.5%,Visual 模型与原模型相比,在 Weibo 数据集上准确率降低了 17.5%,在 Twitter 上准确率降低了 9.2%。这说明利用多模态特征对信息进行检测相对于利用单模态特征进行检测而言更有效。单模态模型只能从一个角度观察信息,而多模态模型可以同时从多个角度来对信息进行观察,因此可以提高模型的分类准确率。Delete_CNN 模型和 Delete_concate 模型与原模型相比,在 Weibo 数据集和 Twitter 数据集上准确率均有下降,说明了对多模态特征同时进行整体和局部融合能提升特征融合程度,提高分类准确率。在 Weibo 和 Twitter 数据集上,ResNet_Fusion 模型在准确率上对比原模型都有一定程度的下降,说明在进行重复卷积时复用一组卷积核可以提高模型的检测准确率。对于小数据集而言,如果模型太大容易引起过拟合的问题,从而导致模型的表现下降。本文通过复用卷积核的方式来减少模型的参数量,以防止过拟合。

5 结束语

本文提出的检测模型具有以下优点:(1)可以在早期对不同模态的数据进行融合,利用其相关性。(2)能同时对多模态特征进行局部和整体融合。同时也存在如下几个缺点:(1)除了文本和图像信息以外,无法有效地使用其他信息。(2)特征提取困难,不同模态的特征必须具有相同的格式。

本文提出了一种利用多模态信息进行虚假信息检测的模型。该模型使用混合融合方法对多模态信息进行融合,先采用多层 CNN 和简单拼接的方法对多模态特征从局部和整体进行融合,得到 2 种融合特征,最后将其输入后期融合模块进行进一步的融合。在 Weibo 数据集和 Twitter 数据集上的实验结果表明,本文模型的检测准确率均优于基准模型的。同时,本文提出的模型也有局限性:(1)对于包含多幅图像的帖子,只能使用其中一幅图像进行检测,不能同时使用所有图像。(2)该模型结构复杂,参数众多,无法在小型设备上运行。在未来的工作中,主要对以下几个问题进行研究:(1)如

何降低模型的复杂性,以便在小型设备上部署。(2)如何充分利用帖子中的所有信息。(3)如何检测在社交平台上已经开始传播的虚假信息。

参考文献:

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

[2] Song C G,Ning N W,Zhang Y L,et al. A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks[J]. Information Processing & Management,2021,58(1):102437-1-102437-14.

[3] Rashkin H,Choi E,Jang J Y,et al. Truth of varying shades:Analyzing language in fake news and political fact-checking[C]//Proc of the 2017 Conference on Empirical Methods in Natural Language Processing,2017:2931-2937.

[4] Popat K,Mukherjee S,Strötgen J,et al. Credibility assessment of textual claims on the web[C]//Proc of the 25th ACM International Conference on Information and Knowledge Management,2016:2173-217.

[5] Alonso-Bartolome S,Segura-Bedmar I. Multimodal fake news detection[J]. arXiv:2112.04831,2021.

[6] Peng X,Bao X T. An effective strategy for multi-modal fake news detection[J]. Multimedia Tools and Applications,2022,81:13799-13822.

[7] Choi H,Ko Y. Effective fake news video detection using domain knowledge and multimodal data fusion on YouTube[J]. Pattern Recognition Letters,2022,154:44-52.

[8] Qi Peng,Cao Juan,Sheng Qiang. Semantics-enhanced multimodal fake news detection[J]. Journal of Computer Research and Development,2021,58(7):1456-1465.(i1n Chinese)

[9] Simonyan K,Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556,2014.

[10] Devlin J,Chang M W,Lee K,et al. BERT:Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805,2018.

[11] Liu Z,Lin Y,Cao Y,et al. Swin transformer:Hierarchical vision transformer using shifted windows[J]. arXiv:2103.14030,2021.

[12] Gupta M,Zhao P,Han J. Evaluating event credibility on Twitter[C]//Proc of the 2012 SIAM International Conference on Data Mining,2012:153-164.

[13] Zhao Liang. Research on multimodal data fusion methods[D]. Dalian:Dalian University of Technology,2018.(in Chinese)

[14] Meng Jie,Wang Li,Yang Yan-jie,et al. Multi-modal deep fusion for false information detection[J]. Journal of Computer Applications,2021,42(2):419-425.(in Chinese)

[15] Wang Jian,Wang Yu-cui,Huang Meng-jie. False information in social networks:Definition,detection and control[J]. Computer Science,2021,48(8):263-277.(in Chinese)

[16] Pérez-Rosas V,Kleinberg B,Lefevre A,et al. Automatic detection of fake news[J]. arXiv:1708.07104,2017.

[17] Kwon S,Cha M,Jung K,et al. Prominent features of rumor propagation in online social media[C]//Proc of 2013 IEEE 13th International Conference on Data Mining,2013:1103-1108.

[18] 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.

[19] Ma J,Gao W,Mitra P,et al. Detecting rumors from microblogs with recurrent neural networks[C]//Proc of the 25th International Joint Conference on Artificial Intelligence,2016:1-7.

[20] Qi P,Cao J,Yang T,et al. Exploiting multi-domain visual information for fake news detection[C]//Proc of 2019 IEEE International Conference on Data Mining,2019:518-527.

[21] Ma J,Gao W,Wong K F. Detect rumors on Twitter by promoting information campaigns with generative adversarial learning[C]//Proc of the World Wide Web Conference,2019:3049-3055.

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

[23] Kumari R,Ekbal A. AMFB:Attention based multimodal factorized bilinear pooling for multimodal fake news detection[J]. Expert Systems with Applications,2021,184:115412.

[24] Giachanou A,Zhang G,Rosso P. Multimodal fake news detection with textual,visual and semantic information[C]//Proc of International Conference on Text,Speech,and Dialogue,2020:30-38.

[25] Sun Y,Wang S,Li Y,et al. ERNIE:Enhanced representation through knowledge integration[J]. arXiv:1904.09223,2019.

[26] Vaswani A,Shazeer N,Parmar N,et al. Attention is all you need[C]//Proc of International Conference on Neural Information Proces- sing Systems,2017:5998-6008.

[27] Boididou C,Andreadou K,Papadopoulos S,et al. Verifying multimedia use at mediaeval 2015[J]. MediaEval,2015,3(3):1-4.

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

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

[30] Khattar D,Goud J S,Gupta M,et al. MVAE:Multimodal variational autoencoder for fake news detection[C]//Proc of the World Wide Web Conference,2019:2915-2921.

附中文参考文献：

[8] 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):1456-1465.

[13] 赵亮.多模态数据融合算法研究[D].大连:大连理工大学,2018.

[14] 孟杰,王莉,杨延杰,等.基于多模态深度融合的虚假信息检测[J].计算机应用,2021,42(2):419-425.

[15] 王剑,王玉翠,黄梦杰.社交网络中的虚假信息:定义,检测及控制[J].计算机科学,2021,48(8):263-277.

作者简介：

梁毅(1998-),男,山西运城人,博士生,研究方向为自然语言处理。**E-mail**:1049809183@qq.com

LIANG Yi, born in 1998, PhD candidate, his research interest includes natural language processing.

吐尔地·托合提(1975-),男,新疆乌鲁木齐人,博士,教授,研究方向为自然语言处理、知识问答及对话系统、多模态交互与情感计算、多模态知识图谱构建与应用。**E-mail**:turdy@xju.edu.cn

Turdi Tohti, born in 1975, PhD, professor, his research interests include natural language processing, Q&A and dialogue system, multimodal interaction & affective computing, multimodal knowledge graph construction & application.

艾斯卡尔·艾木都拉(1972-),男,新疆乌鲁木齐人,博士,教授,研究方向为智能信息处理。**E-mail**:askar@xju.edu.cn

Askar Hamdulla, born in 1972, PhD, professor, his research interest includes intelligent information processing.

2023 CCF 全国高性能计算学术年会征文通知

由中国计算机学会主办,中国计算机学会高性能计算专业委员会、崂山实验室、中国海洋大学、青岛海洋科技中心、齐鲁工业大学(山东省科学院)共同承办,青岛国实科技集团有限公司、山东省计算中心(国家超级计算济南中心)、中北大学、北京并行科技股份有限公司共同协办的“2023 CCF 全国高性能计算学术年会(CCF HPC CHINA 2023)”将于 2023 年 8 月 24 日至 26 日在青岛召开。全国高性能计算学术年会是中国一年一度高性能计算领域的盛会,为相关领域的学者提供交流合作、发布最前沿科研成果的平台,将有力地推动中国高性能计算的发展。

征文涉及的领域包括但不限于:高性能计算机体系结构、高性能计算机系统软件、高性能计算环境、高性能微处理器、高性能计算机应用、并行算法设计、并行程序开发、大数据并行处理、科学计算可视化、云计算和网格计算相关技术及应用、AI+Science、量子计算、State of Practice 最佳实践,以及其他高性能计算相关领域。本次大会设置“超算最佳应用”Track,评选 2023 年度超算最佳应用。

会议录用的中文论文将分别推荐到《计算机研究与发展》(EI)、《计算机学报》(EI)、《计算机科学与探索》(正刊)、《计算机工程与科学》(正刊)、《计算机科学》(正刊)、《国防科技大学学报》(EI 正刊)和《数据与计算发展前沿》(正刊)等刊物上发表,英文论文推荐到 CCF Transactions on High Performance Computing(CCF THPC)、Algorithms 或拟由 Springer 出版。会议还将评选优秀论文和优秀论文提名奖各 5 名。

投稿须知：

本届大会接收中英文投稿。作者所投稿件必须是原始的、未发表的研究成果、技术综述、工作经验总结或技术进展报告。

请登录 https://easychair.org/conferences/?conf=hpcchina2023 的会议投稿系统链接进行投稿,首次登录请注册。

投稿要求：

论文模版下载地址为 https://gitee.com/hpcchina/template,其中,中文/英文 word 模版为 word-cn-en.doc,中文/英文 latex 模版为 latex-cn-en.zip。

“超算最佳应用”Track 请单独使用 best-application-latex-cn.zip 中文模版或者 best-application-latex-en.zip 英文模版。

论文提交截止日期:2023 年 6 月 30 日

论文录用通知日期:2023 年 7 月 31 日

正式论文提交日期:2023 年 8 月 05 日

联系人:袁良、李希代

联系电话:010-6260 0662

电子邮箱:hpcchina@gmail.com、lixidai@ict.ac.cn