Journal of Computer Applications
计算机应用,2022,42(2):419－425
ISSN 1001-9081
CODEN JYIIDU
2022-02-10
http://www.joca.cn

文章编号:1001-9081(2022)02-0419-07
DOI:10.11772/j.issn.1001-9081.2021071184

基于多模态深度融合的虚假信息检测

孟　杰[1],王　莉[1*],杨延杰[1],廉　飚[2]
(1.太原理工大学 大数据学院,太原 030600;　2.北方自动控制技术研究所,太原 030006)
(*通信作者电子邮箱 wangli@tyut.edu.cn)

摘　要:针对虚假信息检测中图片特征提取不充分,以及忽视了单模内关系以及单模与多模之间交互作用的问题,提出一种基于文本和图片信息的多模态深度融合(MMDF)模型。首先,用双向门控循环单元(Bi-GRU)提取文本的丰富语义特征,用多分支卷积-循环神经网络(CNN-RNN)提取图片的多层次特征;然后,建立模间和模内的注意力机制以捕获语言和视觉领域之间的高层交互,并得到多模态的联合表征;最后,将各模态原表征与融合后的多模态联合表征依据注意力权重进行再融合,以加强原信息的作用。该模型与多模态变分自动编码器(MVAE)模型相比,在中国计算机学会(CCF)竞赛和微博数据集上的准确率分别提升了1.9个百分点和2.4个百分点。实验结果表明,所提模型能够充分融合多模态信息,有效提高虚假信息检测的准确率。

关键词:虚假信息检测;多模态融合;双向门控循环单元;注意力机制;联合表征
中图分类号:TP391　　**文献标志码**:A

Multi-modal deep fusion for false information detection

MENG Jie[1], WANG Li[1*], YANG Yanjie[1], LIAN Biao[2]
(1. *College of Data Science, Taiyuan University of Technology, Taiyuan Shanxi* 030600, *China*;
2. *North Automatic Control Technology Institute, Taiyuan Shanxi* 030006, *China*)

Abstract: Concerning the problem of insufficient image feature extraction and ignorance of single-modal internal relations and the interactions between single-modal and multi-modal, a text and image information based Multi-Modal Deep Fusion (MMDF) model was proposed. Firstly, the Bi-Gated Recurrent Unit (Bi-GRU) was used to extract the rich semantic features of the text, and the multi-branch Convolutional-Recurrent Neural Network (CNN-RNN) was used to extract the multi-level features of the image. Then the inter-modal and intra-modal attention mechanisms were established to capture the high-level interaction between the fields of language and vision, and the multi-modal joint representation was obtained. Finally, the original representation of each modal and the fused multi-modal joint representation were re-fused according to their attention weights to strengthen the role of the original information. Compared with the Multimodal Variational AutoEncoder (MVAE) model, the proposed model has the accuracy improved by 1.9 percentage points and 2.4 percentage points on the China Computer Federation (CCF) competition and the Weibo datasets respectively. Experimental results show that the proposed model can fully fuse multi-modal information and effectively improve the accuracy of false information detection.

Key words: false information detection; multi-modal fusion; Bi-directional Gated Recurrent Unit (Bi-GRU); attention mechanism; joint representation

0　引言

社交媒体在给人们带来便利的同时,促进了虚假信息的广泛传播,对社会稳定造成了巨大的威胁。例如,在2019新型冠状病毒暴发之后,各种虚假信息在社交媒体广泛传播[1],引起民众极大的恐慌。因此,迫切需要使用技术手段自动化检测虚假信息,防止引发严重负面影响。

早期方法主要从文本内容中提取语言特征[2-4]来检测虚假信息,后来研究发现图片也包含丰富的信息,能有效提高模型检测准确率。因此,最近很多研究工作将文本和图片这两种模态信息相结合,采用基于多模态的方法检测虚假信息。然而,现有多模态方法存在一定不足。首先,对图片表征时,大多数工作依赖于VGG19[5]的最终输出,忽略了图片的不同层次特征;其次,在学习多模态联合表征时只考虑到

收稿日期:2021-07-09;**修回日期**:2021-07-18;**录用日期**:2021-07-21。　**基金项目**:国家自然科学基金资助项目(61872260)。
作者简介:孟杰(1994—),男,山西长治人,硕士研究生,主要研究方向:自然语言处理、虚假信息检测;　王莉(1971—),女,山西太原人,教授,博士,CCF高级会员,主要研究方向:大数据计算与分析、数据挖掘;　杨延杰(1995—),男,山西原平人,硕士研究生,主要研究方向:自然语言处理、数据挖掘;　廉飚(1987—),男,山西太原人,硕士,主要研究方向:软件开发、数据挖掘。

不同模态之间的交互关系，忽略了单模态内部的复杂关系以及单模与多模的交互关系。

为了解决上述问题，本文提出了一种基于文本和图片信息的多模态深度融合（Multi-Modal Deep Fusion，MMDF）模型来识别虚假信息，首先，用双向门控循环单元（Bi-Gated Recurrent Unit，Bi-GRU）提取文本特征，并借鉴 Qi 等[6]的方法，用多分支卷积-循环神经网络（Convolutional Neural Network-Recurrent Neural Network，CNN-RNN）提取图片的不同层次特征；然后采用模间和模内注意力机制动态融合多模态信息，得到多模态联合表征；最后，通过注意力机制将各模态原表征与融合后的多模态联合表征进行再融合，加强原信息的作用。实验结果表明，本文模型可以有效融合多模态信息。

本文的工作主要包括以下几个方面：

1）提出了一种基于 VGG19 的多分支网络 CNN-RNN 来提取图片的不同层次特征；

2）提出了模间和模内注意力机制，在学习不同模态交互关系的基础上，进一步捕获了单模内的复杂关系；

3）将各模态原表征与融合后的多模态联合表征通过注意力网络相结合，加强原信息的作用。

1 相关工作

虚假信息可以定义为故意捏造且验证为假的信息[7-8]，而模态是指每一种信息的来源或形式。本文按照模态数将目前研究大致分为两类：基于单模态的方法和基于多模态的方法。

1.1 基于单模态的方法

基于单模态的方法分为基于文本的方法和基于图片的方法。

基于文本的方法在早期主要通过提取各种手工特征并结合机器学习方法进行虚假信息检测[9]，但易受到数据集的限制，可扩展性较差。随着深度学习的广泛应用，Ma 等[4]引入递归神经网络学习文本的隐藏表征；刘政等[10]通过卷积神经网络隐含层的学习训练来挖掘表示文本深层的特征；受生成对抗网络（Generative Adversarial Network，GAN）的启发，Ma 等[11]提出了一种基于 GAN 的模型，通过对抗性训练可以捕获低频但判别性更强的特征。

除了文本之外，图片已被证明在虚假信息检测中起着非常重要的作用[8,12]。早期研究主要使用图像的基本统计特征，但无法完整提取图片内包含的大量信息。最近研究通常使用经过预训练的深层 CNN 来提取图片特征。Qi 等[6]通过注意力机制动态融合图片频域和像素域的特征进行虚假信息检测。

虽然虚假信息检测从单模态的角度已经取得了一定进展，但仅仅从文本或图片的角度来研究问题，信息利用率和检测性能较低。

1.2 基于多模态的方法

基于多模态的方法使用文本和图片这两种模态信息来检测虚假信息。Singhal 等[13]利用 BERT（Bidirectional Encoder Representations from Transformers）提取文本特征，用 VGG19 提取图片特征，并将这些特征拼接作为联合表征进行分类；Jin 等[14]提出了一种具有注意力机制的循环神经网络，以融合文本和图片特征进行虚假信息检测；Song 等[15]则利用跨模态注意力残差网络从源模态中选择性地提取与目标模态相关的信息；为了排除特定事件对信息真假判别的干扰，Wang 等[16]提出了一种利用事件对抗神经网络来检测虚假信息的方法，该方法可以学习不同事件之间的共性，取得了较好的效果；Zhang 等[17]利用事件记忆网络来捕获与特定事件无关的潜在话题信息，对新出现的事件获得了更好的泛化能力；为了学习跨模态的共享潜在表示，Khattar 等[18]提出了一种多模态可变自动编码器，通过将可变分自动编码器与分类器相结合以进行虚假信息检测；Zhou 等[19]则通过比较图片和文本之间的相似性来检测虚假信息。在比较相似性的基础上，Xue 等[20]引入了误差等级分析（Error Level Analysis，ELA）算法，使用卷积神经网络对图片在物理层面的真实性进行判断。

2 问题定义

设 $P=\{p_1,p_2,\cdots,p_m\}$ 是一个虚假信息数据集，其中 p_i 为第 i 个帖子，m 为数据集中帖子的个数。对于任意帖子 $p=\{T,V\}$，T 和 V 分别代表其对应的文本和图片。虚假信息检测任务可以描述为学习一个函数 $f(T,V)\rightarrow y$，其中标签值 $y\in\{0,1\}$，0 代表真实信息，1 代表虚假信息。

3 本文模型

本文模型主要由四部分组成：文本特征提取器、图片特征提取器、多模态融合器和虚假信息检测器，如图 1 所示。

3.1 文本特征提取器

为了捕获句子中的长期依赖关系和单词的上下文信息，本文采用 Bi-GRU 作为文本特征提取器的核心模块。用经过 Word2Vec[21]预训练的词嵌入来初始化每个单词向量。对于文本 T，第 i 个单词初始化向量为 $\boldsymbol{T}_i\in\mathbf{R}^k$（$k$ 为维度）。

因此，具有 n 个单词的文本表示为：$T=\{\boldsymbol{T}_1,\boldsymbol{T}_2,\cdots,\boldsymbol{T}_n\}$。Bi-GRU 计算如下：

$$\overrightarrow{\boldsymbol{h}_i}=\overrightarrow{\mathrm{GRU}}(\boldsymbol{T}_i);\ i\in[1,n] \tag{1}$$

$$\overleftarrow{\boldsymbol{h}_i}=\overleftarrow{\mathrm{GRU}}(\boldsymbol{T}_i);\ i\in[1,n] \tag{2}$$

对于第 i 个时间步，$\overrightarrow{\boldsymbol{h}_i}\in\mathbf{R}^k$ 表示 $\boldsymbol{T}_i$ 通过前向 GRU 得到的隐藏表征，$\overleftarrow{\boldsymbol{h}_i}\in\mathbf{R}^k$ 则表示 $\boldsymbol{T}_i$ 通过后向 GRU 得到的隐藏表征，隐藏表征 $\boldsymbol{h}_i$ 由 $\overrightarrow{\boldsymbol{h}_i}$ 和 $\overleftarrow{\boldsymbol{h}_i}$ 拼接而成，即 $\boldsymbol{h}_i=[\overrightarrow{\boldsymbol{h}_i},\overleftarrow{\boldsymbol{h}_i}]$，$\boldsymbol{h}_i\in\mathbf{R}^{2k}$。按顺序将 n 个时间步的隐藏表征堆叠起来得到文本特征矩阵 $\boldsymbol{T}_{\mathrm{m}}\in\mathbf{R}^{n\times 2k}$。将前向 GRU 中最后一个时刻的隐藏层向量 $\overrightarrow{\boldsymbol{h}_n}$

和后向 GRU 中第一个时刻的隐藏层向量表征 $\overleftarrow{\boldsymbol{h}_1}$ 拼接的结果作为整个文本的原表征 $\boldsymbol{T}_{\mathrm{f}} \in \mathbf{R}^{2k}$,即 $\boldsymbol{T}_{\mathrm{f}} = [\overrightarrow{\boldsymbol{h}_n}, \overleftarrow{\boldsymbol{h}_1}]$。

3.2　图片特征提取器

图片特征提取器的核心是多分支 CNN-RNN 模块,如图 2 所示,包括 5 个 CNN 分支,每个分支的 Block 与 VGG19 相对应,将每个分支提取出来的特征分别依次通过卷积层、平铺层和全连接层,得到对应特征向量 $\boldsymbol{v}_t \in \mathbf{R}^k (t \in [1,5])$,代表从局部到全局不同层次的特征。

不同层次的特征之间有很强的依赖性。例如,中层特征纹理由底层特征线组成,又构成了高层特征对象,这可以视为一种序列关系,所以用 Bi-GRU 对这些特征之间的顺序依赖性进行建模:

$$\overrightarrow{\boldsymbol{l}_t} = \overrightarrow{\mathrm{GRU}}(\boldsymbol{v}_t);\ t \in [1,5] \tag{3}$$

$$\overleftarrow{\boldsymbol{l}_t} = \overleftarrow{\mathrm{GRU}}(\boldsymbol{v}_t);\ t \in [1,5] \tag{4}$$

类似于文本特征提取器,得到图片特征矩阵 $\boldsymbol{V}_{\mathrm{m}} \in \mathbf{R}^{5\times 2k}$ 和图片原表征 $\boldsymbol{V}_{\mathrm{f}} \in \mathbf{R}^{2k}$,即 $\boldsymbol{V}_{\mathrm{f}} = [\overrightarrow{\boldsymbol{l}_5}, \overleftarrow{\boldsymbol{l}_1}]$。

3.3　多模态融合器

3.3.1　模间注意力模块

为捕获文本和图片之间的交互关系,首先使用注意力机制计算不同模态之间的相关性,然后根据学习到的相关性权重更新文本和图片特征矩阵。注意力机制公式如下:

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V} \tag{5}$$

其中:Attention(·)为注意力模块运算函数;$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别为 query 矩阵、key 矩阵和 value 矩阵;d 作为防止分子点积值过大的比例因子,其值为输入特征的维度。

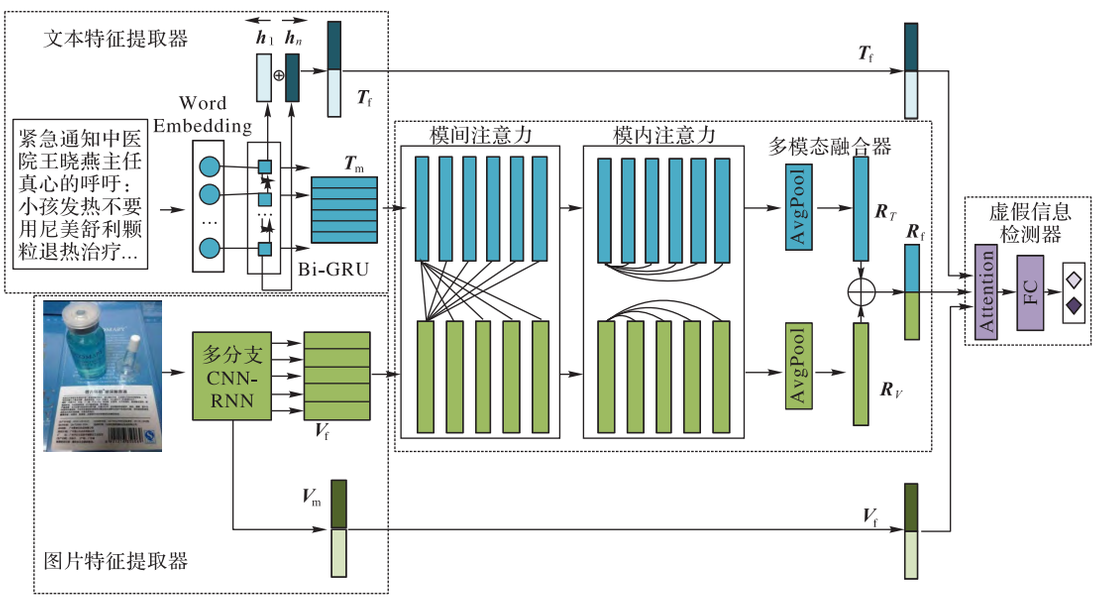


图 1　本文模型总体框架

Fig. 1　Overall framework of the proposed model

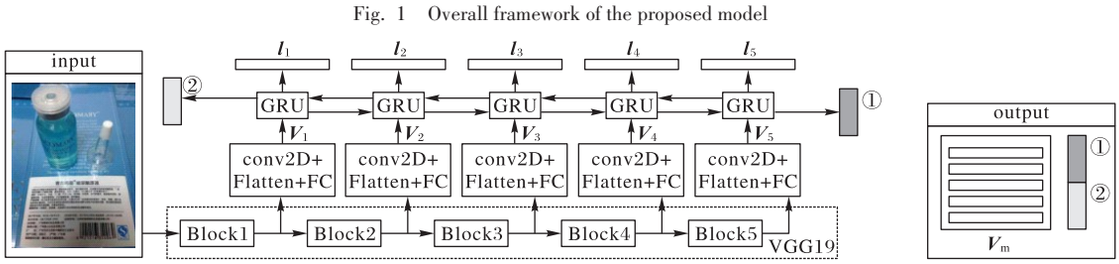


图 2　多分支 CNN-RNN 结构

Fig. 2　Multi-branch CNN-RNN structure

经过模间注意力模块得到的文本更新特征矩阵 $\boldsymbol{T}_{\mathrm{update}}$ 和图片更新特征矩阵 $\boldsymbol{V}_{\mathrm{update}}$ 分别为:

$$\boldsymbol{T}_{\mathrm{update}} = \mathrm{Attention}(\boldsymbol{T}_{\mathrm{m}}\boldsymbol{W}_{\mathrm{q1}}, \boldsymbol{V}_{\mathrm{m}}\boldsymbol{W}_{\mathrm{k1}}, \boldsymbol{V}_{\mathrm{m}}\boldsymbol{W}_{\mathrm{v1}}) \tag{6}$$

$$\boldsymbol{V}_{\mathrm{update}} = \mathrm{Attention}(\boldsymbol{V}_{\mathrm{m}}\boldsymbol{W}_{\mathrm{q2}}, \boldsymbol{T}_{\mathrm{m}}\boldsymbol{W}_{\mathrm{k2}}, \boldsymbol{T}_{\mathrm{m}}\boldsymbol{W}_{\mathrm{v2}}) \tag{7}$$

其中:$\boldsymbol{T}_{\mathrm{update}} \in \mathbf{R}^{n\times 2k}$;$\boldsymbol{V}_{\mathrm{update}} \in \mathbf{R}^{5\times 2k}$;$\boldsymbol{W}_{\mathrm{q1}}$、$\boldsymbol{W}_{\mathrm{k1}}$、$\boldsymbol{W}_{\mathrm{v1}}$、$\boldsymbol{W}_{\mathrm{q2}}$、$\boldsymbol{W}_{\mathrm{k2}}$、$\boldsymbol{W}_{\mathrm{v2}} \in \mathbf{R}^{2k\times 2k}$。

将 $\boldsymbol{T}_{\mathrm{m}}$ 与 $\boldsymbol{T}_{\mathrm{update}}$ 拼接作为新的文本特征矩阵 $\boldsymbol{T}_{\mathrm{m1}} \in \mathbf{R}^{n\times 4k}$:

$$\boldsymbol{T}_{\mathrm{m1}} = [\boldsymbol{T}_{\mathrm{m}}, \boldsymbol{T}_{\mathrm{update}}] \tag{8}$$

同理,可得新的图片特征矩阵 $\boldsymbol{V}_{\mathrm{m1}} \in \mathbf{R}^{5\times 4k}$:

$$\boldsymbol{V}_{\mathrm{m1}} = [\boldsymbol{V}_{\mathrm{m}}, \boldsymbol{V}_{\mathrm{update}}] \tag{9}$$

3.3.2　模内注意力模块

单模内关系是不同模态间交互关系的补充,本文利用模内注意力模块对单模内关系建模,其计算过程如下:

$$\boldsymbol{T}_{\mathrm{m2}} = \mathrm{Attention}(\boldsymbol{T}_{\mathrm{m1}}\boldsymbol{W}_{\mathrm{q11}}, \boldsymbol{T}_{\mathrm{m1}}\boldsymbol{W}_{\mathrm{k11}}, \boldsymbol{T}_{\mathrm{m1}}\boldsymbol{W}_{\mathrm{v11}}) \tag{10}$$

$$\boldsymbol{V}_{\mathrm{m2}} = \mathrm{Attention}(\boldsymbol{V}_{\mathrm{m1}}\boldsymbol{W}_{\mathrm{q21}}, \boldsymbol{V}_{\mathrm{m1}}\boldsymbol{W}_{\mathrm{k21}}, \boldsymbol{V}_{\mathrm{m1}}\boldsymbol{W}_{\mathrm{v21}}) \tag{11}$$

其中:$\boldsymbol{T}_{\mathrm{m2}} \in \mathbf{R}^{n\times 4k}$ 和 $\boldsymbol{V}_{\mathrm{m2}} \in \mathbf{R}^{5\times 4k}$ 分别是最终得到的文本和图

片特征矩阵；$\boldsymbol{W}_{q11}, \boldsymbol{W}_{k11}, \boldsymbol{W}_{v11}, \boldsymbol{W}_{q21}, \boldsymbol{W}_{k21}, \boldsymbol{W}_{v21} \in \mathbf{R}^{4k \times 4k}$。

3.3.3 融合模块

对上面得到的 $\boldsymbol{T}_{m2}$ 和 $\boldsymbol{V}_{m2}$ 分别做平均池化，得到文本和图片的最终表征 $\boldsymbol{R}_T, \boldsymbol{R}_V \in \mathbf{R}^{4k}$：

$$\boldsymbol{R}_T = \text{AvgPool}(\boldsymbol{T}_{m2}) \tag{12}$$

$$\boldsymbol{R}_V = \text{AvgPool}(\boldsymbol{V}_{m2}) \tag{13}$$

其中：AvgPool(·)为平均池化。将文本表征 $\boldsymbol{R}_T$ 和图片表征 $\boldsymbol{R}_V$ 拼接起来，得到文本和图片联合表征 $\boldsymbol{R}'_f \in \mathbf{R}^{8k}$，即 $\boldsymbol{R}'_f = [\boldsymbol{R}_T, \boldsymbol{R}_V]$。然后将其线性变换，得到多模态联合表征 $\boldsymbol{R}_f \in \mathbf{R}^{2k}$。

3.4 虚假信息检测器

在不同模态信息融合过程中，原文本和原图片信息存在一定程度的丢失。建立注意力机制将各模态原表征 $\boldsymbol{T}_f, \boldsymbol{V}_f$ 与融合后的多模态联合表征 $\boldsymbol{R}_f$ 进行再融合，加强原信息的作用，其计算过程如下：

$$\boldsymbol{u}_t = \tanh(\boldsymbol{W}_w \boldsymbol{h}_t + \boldsymbol{b}_w);\ t \in [1,3] \tag{14}$$

$$\alpha_t = \frac{\exp(\boldsymbol{u}_t^{\mathrm{T}} \boldsymbol{u}_w)}{\sum_t \exp(\boldsymbol{u}_t^{\mathrm{T}} \boldsymbol{u}_w)} \tag{15}$$

$$\boldsymbol{s} = \sum_t \alpha_t \boldsymbol{h}_t \tag{16}$$

其中：$\boldsymbol{W}_w$ 表示权重矩阵；$\boldsymbol{b}_w$ 表示偏置项；$\boldsymbol{h}_1$、$\boldsymbol{h}_2$、$\boldsymbol{h}_3$ 分别表示 $\boldsymbol{R}_f$、$\boldsymbol{V}_f$、$\boldsymbol{T}_f$；$\boldsymbol{u}_1$、$\boldsymbol{u}_2$、$\boldsymbol{u}_3$ 分别是 $\boldsymbol{h}_1$、$\boldsymbol{h}_2$、$\boldsymbol{h}_3$ 经过非线性变换之后的结果；上下文向量 $\boldsymbol{u}_w$ 在训练过程中随机初始化并共同学习；α_t 表示第 t 个表征的标准化权重；$\boldsymbol{s}$ 即输入帖子的高级表征。

使用激活函数为 softmax 的全连接层将高级表征 $\boldsymbol{s}$ 投射到二分类目标空间，得到概率分布 p：

$$p = \text{softmax}(\boldsymbol{W}_c \boldsymbol{s} + \boldsymbol{b}_c) \tag{17}$$

其中：$\boldsymbol{W}_c$ 表示权重参数；$\boldsymbol{b}_c$ 表示偏置项。损失函数定义为预测概率分布和真实标签之间的交叉熵误差：

$$L = -\sum_{i=1}^{m} [y_i \log p_i + (1 - y_i) \log(1 - p_i)] \tag{18}$$

其中：m 是帖子的个数；$y_i \in \{0, 1\}$ 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 实验与分析

4.1 数据集

4.1.1 Weibo 数据集

微博(Weibo)数据集[14]真实信息从中国权威信息来源收集，虚假信息则通过微博官方辟谣系统获得。本文使用类似于文献[5]的方法进行数据预处理，删除重复图像和低质量图像，以确保整个数据集的均匀性。然后将整个数据集按 7∶1∶2 的比例划分为训练集、验证集和测试集，并确保它们不包含任何相同事件。

4.1.2 CCF 竞赛数据集

该数据集来自中国计算机学会(China Computer Federation，CCF)举办的“疫情期间互联网虚假信息检测”竞赛，包含 8 个领域：健康、经济、技术、娱乐、社会、军事、政治和教育。数据预处理过程和 Weibo 数据集类似。表 1 列出了这两个数据集的统计信息。

4.2 实验设置

软硬件环境为：Intel i7 2.20 GHz CPU，64 GB 内存，RTX-3090 GPU，Python3.7.6，scikit-learn0.22.1，Pytorch1.4.0。

在整个网络的训练中，batchsize 设置为 100，epoch 设置为 100，学习率为 10^{-3}，dropout 设置为 0.3。优化器为 Adam。

表 1 数据集的统计信息

Tab. 1 Dataset statistics

数据集	虚假信息数	真实信息数	图片数
Weibo	4 749	4 779	9 528
CCF 竞赛	4 324	5 521	9 845

4.3 基准模型

为了验证本文所提出模型的性能，将其与两类基准模型进行了比较：单模态模型和多模态模型。

4.3.1 单模态模型

1)Text：首先将预训练的词嵌入输入 Bi-GRU 以提取文本特征，然后将其输入带有 softmax 激活函数的全连接层进行分类。

2)Visual：将图像输入经过预训练的 VGG-19 网络来提取视觉特征，然后将其输入全连接层并通过 softmax 函数进行分类。

4.3.2 多模态模型

1)VQA(Visual Question Answering)[22]：旨在根据给定的图片回答相应问题。实验中，将文本和图片特征拼接作为联合表征，多分类层替换为二分类层，长短期记忆(Long Short-Term Memory，LSTM)网络层数设置为 1。

2)Neural Talk[23]：根据给定图片生成对应字幕的模型。将循环神经网络(RNN)在每个时间步隐藏表征的平均值作为文本和图片的联合表征，然后通过全连接层进行预测。

3)att-RNN[14]：利用含有注意力机制的循环神经网络融合文本表征和图片表征进行虚假信息检测。

4)EANN(Event Adversarial Neural Network)[16]：是一种端到端的事件对抗神经网络。事件判别器用于度量不同事件之间的差异性，并进一步学习事件之间的共享特征。

5)多模态变分自动编码器(Multimodal Variational AutoEncoder，MVAE)[18]：用于学习各模态之间的相关性，然后与分类器相结合以检测虚假信息。

4.4 实验结果分析

基准模型和本文模型在两个数据集上的实验结果如表 2 所示。实验结果表明，本文模型的准确率优于基准模型。

在两个数据集上，Text 的准确率均高于 Visual，原因是帖子内容大多数以文本为核心，辅以相应图片，所以文本包含更丰富的语义信息。

从单模态和多模态的角度来看,多模态模型的准确率均高于单模态,说明不同模态之间的信息为互补关系,多模态信息相结合可以有效提高虚假信息检测准确率。

在CCF竞赛数据集上MMDF模型准确率比EANN高2.7个百分点,比MVAE高1.9个百分点。在Weibo数据集上比EANN高4.5个百分点,比MVAE高2.4个百分点。说明本文模型能够较好地融合多模态信息,具有较好的虚假信息检测性能。

4.5　t检验

本节使用t检验来验证本文模型相对于基准模型的改善显著性。表3显示了在10次实验的基础上,MMDF模型分别相对于MVAE、EANN和att-RNN在两个数据集上的t检验结果。表3中的p值均小于0.05,说明本文模型相对于基准模型的改善具有显著性。作为验证,还根据自由度和置信水平查找t值,以验证结论是否正确。通过查询t分布表,在自由度为18,置信度为95%的情况下,t值为1.734 1,而表3中t值均大于该值,证明其结论是正确的。

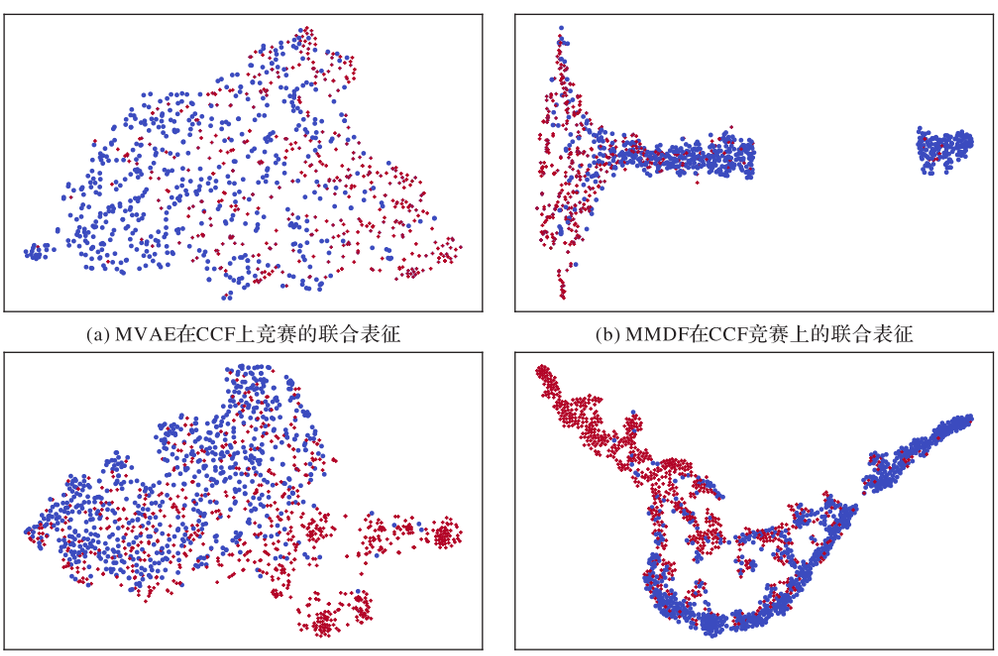
4.6　多模态联合表征可视化

为了进一步证明MMDF模型的优越性,图3是运用t-SNE(t-distributed Stochastic Neighbor Embedding)算法可视化MMDF模型和MVAE模型在两个数据集上的多模态联合表征的结果。图3中“·”和“+”分别对应虚假信息和真实信

表2　两个数据集上的实验结果

Tab. 2　Experimental results on two datasets

数据集	基准模型	准确率	虚假信息			真实信息		
			Precision	Recall	F1	Precision	Recall	F1
CCF竞赛	Text	0.613	0.664	0.532	0.538	0.601	0.66	0.659
	Visual	0.571	0.517	0.712	0.599	0.659	0.455	0.539
	VQA	0.715	0.814	0.622	0.706	0.636	0.832	0.719
	NeuralTalk	0.681	0.733	0.661	0.695	0.634	0.674	0.633
	att-RNN	0.739	0.788	0.651	0.712	0.682	0.805	0.741
	EANN	0.766	0.812	0.623	0.705	0.741	0.822	0.806
	MVAE	0.774	0.808	**0.736**	0.771	0.743	0.814	0.777
	MMDF	**0.793**	**0.821**	0.689	**0.779**	**0.776**	**0.877**	**0.823**
Weibo	Text	0.643	0.662	0.578	0.617	0.609	0.685	0.647
	Visual	0.608	0.610	0.605	0.607	0.607	0.611	0.609
	VQA	0.736	0.797	0.634	0.706	0.695	0.838	0.760
	NeuralTalk	0.726	0.794	0.713	0.692	0.684	0.840	0.754
	att-RNN	0.772	**0.854**	0.656	0.742	0.72	**0.889**	0.795
	EANN	0.793	0.796	0.806	0.801	0.790	0.780	0.785
	MVAE	0.814	0.765	0.874	0.833	0.863	0.734	0.775
	MMDF	**0.838**	0.815	**0.886**	**0.849**	**0.866**	0.786	**0.824**



(a) MVAE在CCF上竞赛的联合表征　(b) MMDF在CCF竞赛上的联合表征

(c) MVAE在Weibo上的联合表征　(d) MMDF在Weibo上的联合表征

图3　用t-SNE可视化在两个数据集测试数据上的多模态联合表征

Fig. 3　t-SNE visualization of multi-modal joint representations on the test data of two datasets

息。可以观察到，在两个数据集上，MMDF模型均学习到可判别性更强的多模态联合表征，这进一步验证了本文模型的优越性。

表3 多个模型在两个数据集上的t检验对比结果

Tab. 3 Comparison results of t-tests of multiple models on two datasets

对比模型	CCF竞赛	Weibo
MMDF和MVAE	t=9.016，p=1.721E-5	t=12.079，p=2.548E-6
MMDF和EANN	t=12.961，p=2.398E-6	t=24.382，p=4.381E-9
MMDF和att-RNN	t=18.365，p=1.964E-6	t=28.896，p=2.333E-8

4.7 消融实验

为了进一步研究本文所提MMDF模型中各个模块的作用，通过删除某些模块，简化模型进行对比实验，其结果如图4所示。图4中：

1）MMDF：包含所有模块；

2）w/o T：删除文本原表征，只保留多模态联合表征和图片原表征；

3）w/o V：删除图片原表征，只保留多模态联合表征和文本原表征；

4）w/o inter-att：删除模间注意力，保留模内注意力；

5）w/o intra-att：删除模内注意力，保留模间注意力。

从图4可以看出：在两个数据集上，w/o inter-att和w/o intra-att准确率最低，表明了挖掘不同模态间关系和单模内关系对虚假信息检测的重要性。在CCF竞赛数据集上w/o inter-att准确率高于w/o intra-att；但在Weibo数据集上情况却相反，说明不同模态间关系和单模内关系在不同情况下重要性有所差异。在CCF竞赛数据集上，w/o T的准确率为77.9%，w/o V的准确率为78.3%，均比MMDF模型略低；在Weibo数据集上，w/o T的准确率为82.3%，w/o V的准确率为82%，同样低于MMDF模型的准确率。这说明原文本和原图片融合过程中确实存在一定程度的信息丢失，将各模态原表征与融合后的多模态联合表征进行再融合，加强原信息的作用，能有效提高模型检测准确率。

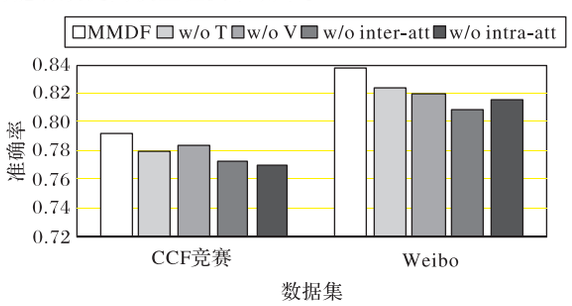


图4 在两个数据集上的消融实验结果

Fig. 4 Ablation experiment result on two datasets

5 结语

本文提出了一种多模态信息深度融合模型用于虚假信息检测，该模型利用模间和模内注意力机制捕获语言和视觉领域之间的高层交互，并建立注意力机制将各模态原表征与融合后的多模态联合表征进行再融合，加强原信息的作用。在两个数据集上的实验结果表明，本文模型的检测准确率优于基准模型。在同一个帖子中，有时候会附加多张不同的图片，从不同的角度向用户传达信息。在未来的工作中，会进一步考虑如何将文本信息和多张不同的图片信息相结合来检测虚假信息。

参考文献（References）

[1] KOUZY R, ABI JAOUDE J, KRAITEM A, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter [J]. Cureus, 2020, 12(3): No. e7255.

[2] RASHKIN H, CHOI E, JANG J Y, et al. Truth of varying shades: analyzing language in fake news and political fact-checking [C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2017: 2931-2937.

[3] POPAT K, MUKHERJEE S, STRÖTGEN J, et al. Credibility assessment of textual claims on the Web [C]// Proceedings of the 25th ACM International Conference on Information and Knowledge Management. New York: ACM, 2016: 2173-2178.

[4] 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. [S. l.]: IJCAI Organization, 2016: 3818-3824.

[5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10) [2021-03-10]. https://arxiv.org/pdf/1409.1556.pdf.

[6] QI P, CAO J, YANG T Y, et al. Exploiting multi-domain visual information for fake news detection [C]// Proceedings of the 2019 IEEE International Conference on Data Mining. Piscataway: IEEE, 2019: 518-527.

[7] RUCHANSKY N, SEO S, LIU Y. CSI: a hybrid deep model for fake news detection [C]// Proceedings of the 2017 ACM Conference on Information and Knowledge Management. New York: ACM, 2017: 797-806.

[8] SHU K, SLIVA A, WANG S H, et al. Fake news detection on social media: a data mining [J]. ACM SIGKDD Explorations Newsletter, 2017, 19(1): 22-36.

[9] 毛二松，陈刚，刘欣，等. 基于深层特征和集成分类器的微博谣言检测研究[J]. 计算机应用研究，2016，33(11)：3369-3373.（MAO E S, CHEN G, LIU X, et al. Research on detecting micro-blog rumors based on deep features and ensemble classifier [J]. Application Research of Computers, 2016, 33(11): 3369-3373.）

[10] 刘政，卫志华，张韧弦. 基于卷积神经网络的谣言检测[J]. 计算机应用，2017，37(11)：3053-3056，3100.（LIU Z, WEI Z H, ZHANG R X. Rumor detection based on convolutional neural network [J]. Journal of Computer Applications, 2017, 37(11): 3053-3056, 3100.）

[11] MA J, GAO W, WONG K F. Detect rumors on Twitter by

promoting information campaigns with generative adversarial learning [C]// Proceedings of the 2019 World Wide Web Conference. New York: ACM, 2019: 3049-3055.

[12] JIN Z W, CAO J, ZHANG Y D, et al. Novel visual and statistical image features for microblogs news verification [J]. IEEE Transactions on Multimedia, 2017, 19(3): 598-608.

[13] SINGHAL S, SHAH R R, CHAKRABORTY T, et al. SpotFake: a multi-modal framework for fake news detection [C]// Proceedings of the IEEE 5th International Conference on Multimedia Big Data. Piscataway: IEEE, 2019: 39-47.

[14] JIN Z W, CAO J, GUO H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs [C]// Proceedings of the 25th ACM International Conference on Multimedia. New York: ACM, 2017: 795-816.

[15] 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 and Management, 2021, 58(1): No. 102437.

[16] WANG Y Q, MA F L, JIN Z W, 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 and Data Mining. New York: ACM, 2018: 849-857.

[17] ZHANG H W, FANG Q, QIAN S S, et al. Multi-modal knowledge-aware event memory network for social media rumor detection [C]// Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 1942-1951.

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

[19] ZHOU X Y, WU J D, ZAFARANI R. SAFE: similarity-aware multi-modal fake news detection [C]// Proceedings of the 2020 Pacific-Asia Conference on Knowledge Discovery and Data Mining, LNCS 12085. Cham: Springer, 2020: 354-367.

[20] XUE J X, WANG Y B, TIAN Y C, et al. Detecting fake news by exploring the consistency of multimodal data [J]. Information Processing and Management, 2021, 58(5): No. 102610.

[21] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [EB/OL]. (2013-09-07) [2021-03-10]. https://arxiv. org/pdf/1301. 3781. pdf.

[22] ANTOL S, AGRAWAL A, LU J S, et al. VQA: visual question answering [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 2425-2433.

[23] VINYALS O, TOSHEV A, BENGIO S, et al. Show and tell: a neural image caption generator [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3156-3164.

This work is partially supported by National Natural Science Foundation of China (61872260).

MENG Jie, born in 1994, M. S. candidate. His research interests include natural language processing, false information detection.

WANG Li, born in 1971, Ph. D, professor. Her research interests include big data computing and analysis, data mining.

YANG Yanjie, born in 1995, M. S. candidate. His research interests include natural language processing, data mining.

LIAN Biao, born in 1987, M. S. His research interests include software development, data mining.