DOI:10.19652/j.cnki.femt.2406090

基于度量学习的多模态谣言检测*

李 娜[1] 余晓栋[1,2] 朱节中[2,3]
(1.南京信息工程大学自动化学院 南京 210044;2.无锡学院物联网工程学院 无锡 241405;
3.南京信息工程大学软件学院 南京 210044)

摘 要:目前主流的多模态谣言检测模型,主要侧重于建模过程中模态的特征提取与拼接方法研究,而各模态局部特征关系、模态内与模态间的信息交互往往被忽略,这在一定程度上影响到了谣言检测的效果。针对该问题,提出了一种基于度量学习的多模态谣言检测方法。考虑到各模态局部特征关系对模态整体特征表示的影响,采用了句法分析和注意力机制技术分别挖掘文本和图片的局部特征关系;同时,将度量学习应用到谣言检测中,通过三元组学习和对比学习找出模态内与模态间的关联信息。在 Twitter 和 Weibo 两个公开的数据集上进行了性能测试实验,准确率分别达到 92.8%和 85.2%,结果表明将各模态局部特征关系、模态内与模态间的信息交互加入谣言检测模型中能够进一步提升谣言检测的精准度。

关键词:谣言检测;度量学习;多模态;三元组学习;对比学习
中图分类号: TP183 **文献标识码**:A **国家标准学科分类代码**: 520.604

Metric learning-based multimodal rumor detection

Li Na[1] Yu Xiaodong[1,2] Zhu Jiezhong[2,3]
(1. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China;
2. School of IoT Engineering, Wuxi University, Wuxi 214105, China;
3. School of Software, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: At present, the mainstream multi-modal rumor detection models mainly focus on the feature extraction and splicing methods of the modes in the modeling process, while the local feature relationship of each mode and the information interaction within and between modes are often ignored, which affects the effect of rumor detection to a certain extent. To address this issue, we propose a metric learning-based multimodal rumor detection method. Considering the influence of local feature relationships within each modality on the overall representation of modalities, we employed the technology of syntactic analysis and attention mechanism to exploring the local feature relationships of text and images, respectively. Additionally, metric learning is applied to rumor detection, where triplet learning and contrastive learning are utilized to identify the associated information within and between modalities. Performance testing experiments conducted on publicly available datasets from Twitter and Weibo demonstrate accuracy rates of 92.8% and 85.2%, respectively. These results indicate that incorporating local feature relationships within each modality and the interaction between modalities into the rumor detection model can further enhance the accuracy of rumor detection.

Keywords: rumor detection; deep metric learning; multimodal; triplet metric learning; contrastive learning

0 引 言

与传统谣言相比,网络谣言存在传播速度快、影响范围广、呈现方式多的特点[1]。谣言一旦扩散将会带来社会动荡,引起人们恐慌,甚至造成不必要的经济损失。因此,及时有效识别网络谣言对维护社会秩序、保障人民财产安全有重要的意义。

为了提高谣言检测的效率,早期的机器学习主要使用

收稿日期:2024-03-29
* **基金项目**:教育部人文社会科学研究规划基金(18YJA820035)、江苏高校哲学社会科学研究项目(2022SJYB0982)资助

新闻语义特征和新闻传播特征[2]进行谣言检测。后来随着深度学习的广泛应用,研究者们开始使用神经网络挖掘更复杂的特征表示。陈林威等[3]使用异质图神经网络建模事件间的结构关系和信息传播的时序关系,使模型提取更全面的特征从而提高谣言检测的准确性。郭秋实等[4]引入知识图谱技术丰富文本内容表示,并通过图卷积神经网络进行特征提取以提高模型的识别能力。

随着多媒体技术的成熟,网络上的信息开始以文本+图片这类多模态的形式呈现,仅仅依靠单模态信息已经无法准确有效地对谣言进行甄别。因此,梁毅等[5]使用多层卷积神经网络(convolutional neural networks, CNN)同时从句子和特征层面进行特征融合,有效提取多模态信息间的特征关系。为了提高模型的泛化能力,孟佳娜等[6]在原有的多模态模型基础上,引入对抗神经网络学习数据的公共特征,以确保模型在新数据上仍有较高的准确率。考虑到多模态信息间存在相互关联的现象,强子珊等[7]通过构建多模态社交媒体异质图的方法挖掘不同模态间的关系。为了解决数据集稀缺问题,高国鹏等[8]为多模态谣言检测研究提供了更加丰富的数据内容,有效扩充谣言的研究内容。

然而以上的多模态谣言检测研究只考虑到不同模态的特征提取方法以及部分特征之间的关系,忽略了各模态局部特征关系、模态内的特征联系和模态间的特征约束关系共同作用时对模型的影响。多模态谣言检测的关键就在于对模态特征的提取和各模态特征之间的关系挖掘,只有提取到文本和图片的有效特征以及它们之间的相互联系,才能实现更加精准的谣言识别。

本文提出了一种基于度量学习的多模态谣言检测方法,可以有效地捕捉各模态局部特征关系、模态内和模态间的特征联系。首先对文本和图片分别采用预训练的BERT(bidirectional encoder representations from transformers)和VGG-19(visual geometry group)模型进行特征提取,然后用句法分析和自适应注意力挖掘各模态的局部特征关系。接着在空间映射中将两个模态特征转变成同维度向量,使用度量学习中的对比学习和三元组学习依次确认模态间和模态内的特征关系。其中,对比学习计算同一标签下文本特征和视觉特征的相似度并不断优化对比损失函数从而训练模型区分文本+图片组合形式的谣言和非谣言信息,三元组学习计算模态内谣言和非谣言距离,并使用三元组损失函数不断优化两者之间的距离从而训练模型区分同类模态的谣言和非谣言信息。最后,将加入局部特征关系的各模态特征向量输入特征融合单元中进行特征融合,融合后的特征用于谣言检测器对信息的检测。

1 相关研究

1.1 单模态谣言检测

谣言检测技术主要包括机器学习和深度学习两类方法。基于机器学习的检测方法主要依赖人工建立特征工程,再将提取到的特征输入到分类模型中进行训练,最后将训练好的模型用于检测未标记的数据从而判断是否为谣言。Wu等[9]针对谣言内容提出了主题类型特征、用户类型特征、平均情感特征以及转发时间特征等新的特征,通过平均情感得分建立谣言和情感之间的联系,最终实验结果表明提出的特征可以提高谣言检测的准确率。由于机器学习技术的效果取决于人工提取的特征质量,且谣言的传播特征是随着时间变化的,因此机器学习技术存在成本高、耗时多、效率低的弊端。

近年深度学习技术逐渐出现在大众的视野中,其凭借着强大的特征学习能力在人工智能领域中广受欢迎。Ma等[10]首次将循环神经网络应用在谣言检测中,通过对传播数据在时间特征上建模得到谣言的内容随时间变化的隐藏特征,最终准确率高达91%。Chen等[11]在此基础上将循环神经网络和变分自编码器结合(variation auto-encoder)使用无监督学习模型学习文本数据。后来,Xu等[12]提出一种融合神经谣言检测模型,从源帖内容、转发贴的扩散和用户信息3个方面分别使用基于内容的注意力机制和基于扩散的注意力机制以及用户特征编码器提取出高级的特征表示,通过这些特征进行模型训练得到的模型在测试集上的准确率达到94.4%。目前社交平台上的信息不再以单一文本的形式传播,仅考虑文本内容提取到的信息不足以作为谣言判别的充分依据。

1.2 多模态谣言检测

谣言以多模态方式传播比单模态更具有迷惑性,多模态谣言包含的信息量要比单一文本谣言大很多,因此仅依靠文本信息对谣言进行检测的方法在多模态谣言的检测中已经失效。针对这一局限性,Wang等[13]提出了多模态假新闻检测的事件对抗神经网络,该方法首先使用单模态特征提取器分别对文本和图片进行特征提取,接着将各模态特征拼接采用极大极小博弈生成对抗网络提取不变的特征,最后将不变特征应用于新消息进行谣言判断。该方法虽然完成了多模态信息的融合,但是忽略了模态间的信息联系,导致模态信息提取不充分,检测准确率较低。注意力机制使模型能够动态地分配权重给不同的输入信息,从而提高了模型的性能和表现力。Jin等[14]将注意力机制融入递归神经网络,使用长短期记忆网络(long short-term memory, LSTM)获得文本和上下文的联合表示,接着对联合表示加入注意力机制与视觉特征融合提取图片中与文本信息有关的特征,实验结果显示该方法可以有效提取模态间的相关特征,但是仍然存在简单拼接的弊端。Wu等[15]使用Co-Attention分层融合不同模态特征,使模态间的信息得到充分的利用,为了提取更多的图片信息,将图片分为空间域和频域特征提取,然后将其进行协同注意特征融合。戚力鑫等[16]考虑到关键词和图片区域之间的联系,提出了基于注意力机制的多模态融合谣言检测方

法，利用注意力机制挖掘图片中与关键词相关的特征，提出了自适应注意力机制约束模态内部信息。以上方法虽然实现了多模态特征的融合，但是忽略了模态内局部特征之间的关系以及模态间有效信息互相约束作用，容易出现跨模态特征提取重点信息丢失和特征冗余的情况。

1.3 度量学习谣言检测

根据数据之间的距离来判断数据之间的相似性，从而完成数据的分类，该思想被称为最近邻分类。通过将待测数据分类为距离它最近的样本类型，这种思想促进了距离度量学习的产生[17]。度量学习旨在学习嵌入空间中的距离函数使正样本之间的距离不断减小，负样本之间的距离不断扩大。度量学习首次被 Zhang 等[18]应用在多模态不可靠新闻检测上，提出了一种基于 BER 的多模态不可靠新闻检测方法，利用对比学习策略从不可靠新闻中捕获文本和视觉信息。对比学习机制使不可靠新闻分类器去驱使相似的可信信息更近，同时使内容相似可信度标签相反的新闻远离，完成不可信新闻的分类任务。该分类方法在考虑模态特征的同时也考虑了文章之间的关系。由于多模态特征的简单拼接会造成特征的关键信息被忽略，为了同时考虑模态内和模态间的特征关系对多模态谣言检测的影响，Peng 等[19]提出了基于深度度量学习的多模态谣言检测，该方法使用三元组学习提取每个模态中的谣言和非谣言之间的关系，以及对比学习来捕获跨模态的模态间的关系，其实验结果表明模态内和模态间的关系对谣言检测的准确率均有影响。关于度量学习在少样本谣言检测方面的应用，Ran 等[20]提出一种用于小样本交叉事件谣言检测的度量学习方法，该模型主要包括 Base-Classifier 预训练模块和 Base-Meta 训练模块，通过 Base-Classifier 预训练模块在旧事件上训练得到分类模型，然后去除预测层将剩余部分作为编码器，Base-Meta 训练模块用于对编码器进行微调，同时在旧事件的情景上训练度量学习模型，最后将训练好的模型用于对输入信息的分类。以上模型虽然利用度量学习完成了分类工作，但是在进行特征提取时忽略了局部特征之间的联系以及多模态特征之间的融合。

2 基于度量学习的多模态谣言检测

本文提出一种基于度量学习的多模态谣言检测模型，该模型的工作流程如图 1 所示。信息以文本＋图片的方式输入模型，分别采用预训练模型对文字和图片进行特征提取。为了提取出完整的语义表达，需要对文本特征和视觉特征进行局部关系提取。考虑到多模态信息间的信息互补对检测结果的影响，将输入信息划分为谣言和非谣言两类并通过优化对比损失函数实现模态间的信息交互，同时根据同模态内谣言和非谣言的标签优化三元组损失函数实现模态内的信息流动，充分挖掘文本和视觉提供的有效信息。度量学习的多模态谣言检测模型由如下 6 个部分组成。

1）特征提取器，使用 BERT 和 VGG-19 分别对文本信息和图片信息进行特征提取。

2）局部关系提取器，使用句法依存分析和图片自适应技术分别找到文本特征和视觉特征的局部关系，实现模态内部的信息流动。

3）映射空间，将不同模态下的特征表示映射到同一空间，便于数据的距离计算。

4）度量学习，建模模态内和模态间的距离函数，实现信息之间的交互，确定数据之间的联系。

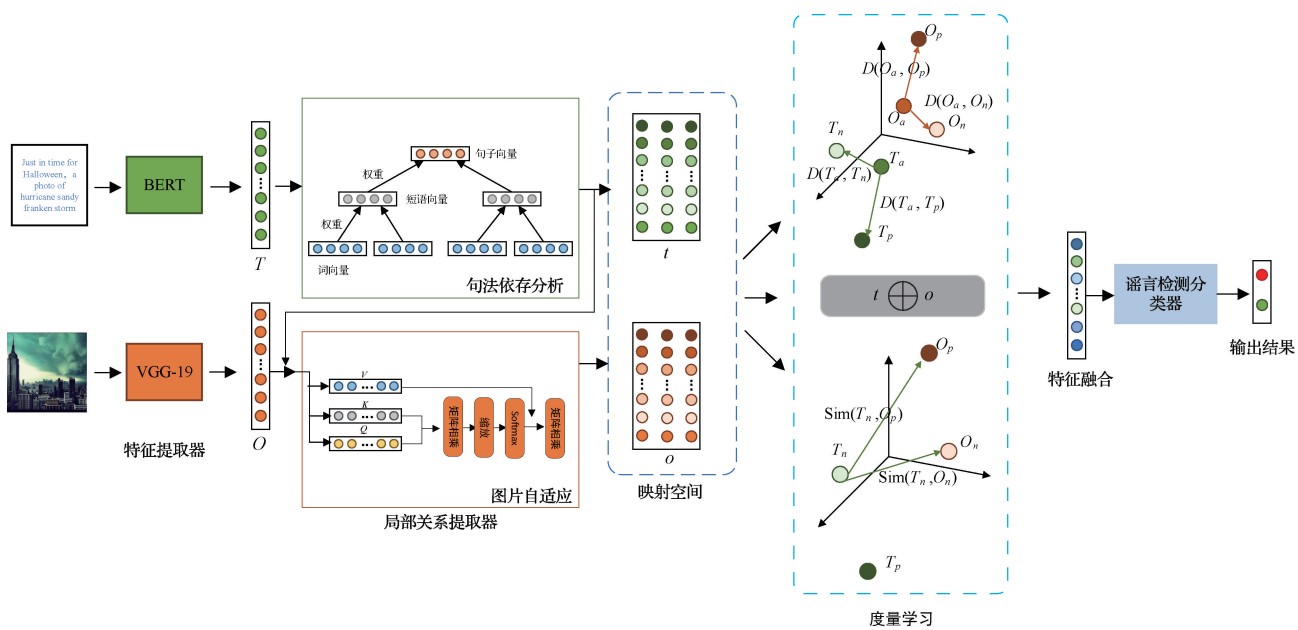

图 1 模型总体图

Fig. 1 Model overview diagra

5)特征融合,将处理后的文本特征和视觉特征进行融合,充分利用输入信息提供的有效数据。

6)谣言检测,通过全连接层神经网络学习分类器,对输入信息的真假性进行判断。

2.1 特征提取器

1)文本特征提取器

BERT[21]编码器作为文本语言模型的核心模块,已经成功应用于问答、翻译、阅读理解和文本分类等任务中,本文使用该模型提取文本特征。首先对文本中每个语句进行分词操作,则句子可由词的组合表示 $W=\{w_1,w_2,w_3,\cdots,w_n\}$($n$ 表示每句话中词的数量)然后通过 BERT 内部的 Encoder 操作输出文本的特征表示。记输出文本特征为 T:

$$T=BERT(W_1,W_2,\cdots,W_n) \tag{1}$$

2)视觉特征提取器

首先将图片转换成 RGB 格式,接着输入 VGG-19[22]模型,通过 16 个卷积层和 3 个全连接层提取图片的表示特征,将最后一层得到的特征矩阵 $\boldsymbol{V}$,接入一个全连接层,通过 ReLU 激活函数最终得到图片的视觉特征,将输出视觉特征记为 O:

$$O=\sigma(\boldsymbol{W}_O\cdot\boldsymbol{V}+b_O) \tag{2}$$

式中:$\boldsymbol{W}_0$ 为全连接层的系数矩阵;b_0 表示偏移量;σ 表示 ReLU 激活函数。

2.2 局部关系提取

考虑到各模态内局部特征之间存在隐性的联系,分别使用句法分析实现文本特征的局部关系提取,自适应注意力挖掘视觉特征的局部关系。

1)句法分析的局部关系提取

句法分析可以对输入的文本信息进行句法结构处理,因此针对文本局部关系提取问题,本文采用句法分析生成词向量之间的相互依存关系。使用 Google 开源的 SyntaxNet 系统构建输入文本的语法树,然后根据不同词的贡献程度为每个词赋不同的权重。最后结合词向量和注意权重更新词的向量表示。将更新后的词向量表示为 $\boldsymbol{T}_{wi}$:

$$\boldsymbol{T}_{w_i}=BERT(W)\cdot Attentions(w_i) \tag{3}$$

式中:$Attentions(w_i)$表示第 i 个词经过句法分析处理后得到的权重。

因为词是组成句子的最小单元,所以在词向量基础上组合短语表示(phrase),根据句法依存分析获得的关系树构建短语特征,其表示为带有注意力权重词向量的加权平均,然后再利用短语向量组合成句子向量表示(sentence),整个过程如图 2 所示。

$$phrase_m=\frac{1}{k}\cdot\sum_j\boldsymbol{T}_{w_j} \tag{4}$$

$$\boldsymbol{T}_s=[phrase_1,phrase_2,\cdots,phrase_p] \tag{5}$$

式中:k 表示构成短语的词的个数;p 表示句子中含有的短语组合的个数。

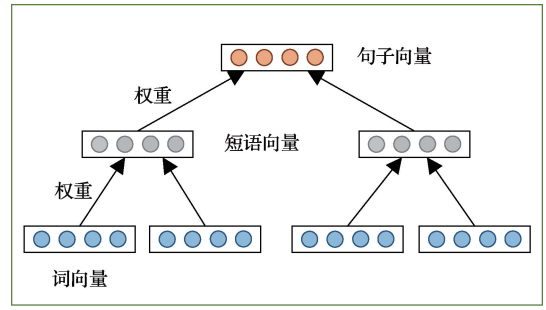

图 2 句法结构分析

Fig. 2 Syntactic structure analysis

2)自适应注意力机制特征提取

考虑到模态间的信息流动性,本文将文本信息作为图片局部关系提取的辅助条件,参考文献[16]提出的 Adaptive-SA 模块对局部关系进行提取。将句法分析得到的句子向量作为视觉特征的自适应条件,根据模态间的关系确认视觉特征的自适应向量,从而计算出视觉特征的局部关系。具体操作如图 3 所示。

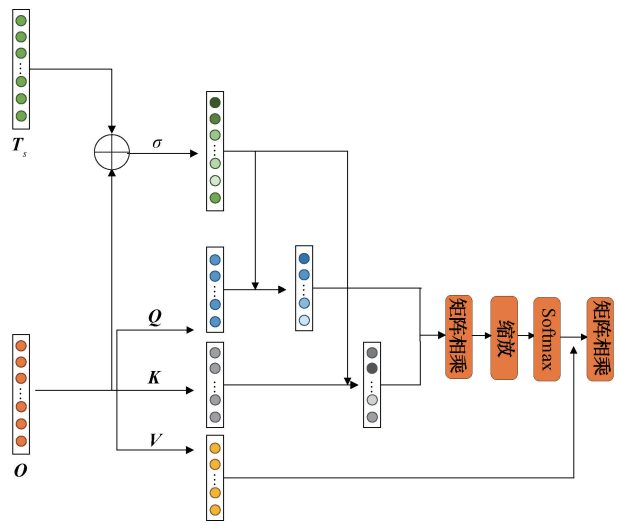

图 3 自适应注意力提取

Fig. 3 Adaptive attention extraction

将句子向量与视觉表示进行融合,通过 Sigmoid 激活函数计算出视觉特征中不同的区域对象权重形成自适应向量:

$$\boldsymbol{G}=\sigma(W_G\cdot[\boldsymbol{T}_s,\boldsymbol{O}]) \tag{6}$$

利用自注意力计算图片内部各区域之间的相关性,从而确定视觉特征的局部关系。由于自适应条件下关键词对视觉特征起了约束作用,因此在对视觉特征进行自注意力计算时,需要通过自适应向量更新视觉特征在不同向量空间的映射:

$$\boldsymbol{O}'_Q=(1+\boldsymbol{G})\cdot\boldsymbol{O}_Q \tag{7}$$

$$\boldsymbol{O}'_K=(1+\boldsymbol{G})\cdot\boldsymbol{O}_K \tag{8}$$

根据自适应条件下的空间映射向量计算视觉特征的局部关系。

$$\begin{cases} atten^{OO} = Softmax\dfrac{\boldsymbol{O'}_Q(\boldsymbol{O'}_K)^{\mathrm{T}}}{\sqrt{d}} \\ O^{updata} = atten^{OO}\cdot\boldsymbol{O}_V \\ O' = Linear(w_o\cdot[O^{update}+O]) \end{cases} \tag{9}$$

式中：d 表示词向量的维度；w_0 表示训练参数；O'表示利用全连接层对拼接后的视觉特征进行维度调整，确保其与句子向量维度相同。

2.3 空间映射

由于不同的特征提取器获得的特征信息是有差别的，所以无法将文本特征和视觉特征进行直接计算。因此，需要将各单模态投影到公共的向量空间中，在向量空间中进行后续的距离计算以及特征融合等操作。

假设文本特征有 N 个 d_1 维向量，视觉特征有 M 个 d_2 维向量，通过如下公式将他们不同的维度转换成 d 维向量表示：

$$o_{n\,n=1}^{N} = O'\boldsymbol{W}_v + b_v \tag{10}$$

$$t_{m\,m=1}^{M} = T_s\boldsymbol{W}_t + b_t \tag{11}$$

式中：$\boldsymbol{W}_v$、$\boldsymbol{W}_t$ 是权重矩阵；b_v、b_t 是偏移量。

2.4 度量学习

鉴于度量学习在图像检索、迁移学习、极端分类等任务的广泛应用，本文使用度量学习中的对比学习和三元组学习分别捕捉模态内和模态间的关系。

1)对比学习

假设 t_p、t_n、o_p、o_n 分别表示文本非谣言、文本谣言、图片非谣言、图片谣言，则 t_p、o_p 组合即(t_p,o_p)表示非谣言对；t_n、o_n 组合即(t_n,o_n)表示谣言对，这两种组合均属于正常情况。当出现 t_p、o_n 组合时，则表明文本和图片标签不匹配，该情况直接判定为谣言情况。综上本文使用余弦相似度[23]计算组合的相似值，并且为同类标签组合赋予高分不同标签组合赋予低分，计算公式如下：

$$sim(\boldsymbol{t}_i,\boldsymbol{o}_j) = \frac{\boldsymbol{t}_i^{\mathrm{T}}\cdot\boldsymbol{o}_j}{\|\boldsymbol{t}_i\|\cdot\|\boldsymbol{o}_j\|} \tag{12}$$

通过最小化对比损失函数最大化不同模态下同类型标签的相似度。经过不断的训练，使模型可以准确区分标签不一致的多模态信息对。

$$L_1 = \sum_{D}\max(0,\alpha_1 - sim(t_p,o_p) + sim(t_p,o_n)) \tag{13}$$

式中：$D=\{(t_p,o_p),(t_n,o_n)\}$表示标签一致的文本图片信息对；$sim()$表示文本图片配对的相似值；$\alpha_1$ 表示同类标签对和异类标签对相似值的最大差值。

2)三元组学习

在对比学习的基础上，三元组学习引进了锚点样本，通过正负样本和锚点之间的距离，进一步确定类内和类间的相对关系。以文本特征为例，假设 t_a 表示锚点样本，t_p 表示和锚点样本标签一致的正样本，t_n 表示和锚点样本标签相反的负样本。通过三元组学习，使得正样本和锚点的距离要小于负样本和锚点的距离，从而完成分类任务。使用经典的马氏距离计算空间中的数据信息：

$$D(t_i,t_j) = (t_i - t_j)^{\mathrm{T}}\boldsymbol{M}(t_i - t_j) \tag{14}$$

式中：$\boldsymbol{M}$ 具有半正定和对称性，$\boldsymbol{M}$ 的特征值要全部非负。

通过最小化三元组损失[24]训练模型参数，使模型对输入信息编码，得到单模态中相同标签信息距离不断靠近，不同标签信息距离不断扩大，从而挖掘出单模态中谣言和非谣言之间的关系。

$$L_2 = \sum_{H}\max(0,\alpha_2 - \Delta^O(D(t_a,t_p) - D(t_a,t_n))) \tag{15}$$

式中：$H=\{(t_a,t_p,t_n)\}$表示同种模态的样本类型；α_2 表示正样本和负样本的间隔参数。在做文本信息分类时，考虑到图片信息对文本内容的约束作用，将图片信息之间的距离作为辅助条件，有助于准确提取模态内关系，即：

$$\Delta^O = Sign(\|o_a - o_p\|_2 - \|o_a - o_n\|_2) \tag{16}$$

式中：o_a、o_p、o_n 分别表示视觉锚点、视觉正样本、视觉负样本，同样的视觉特征的处理流程和文本特征类似。

2.5 特征融合

通过预训练获取文本特征和视觉特征后，分别利用句法分析和自适应注意力获得模态的局部特征关系并对提取的特征进行更新，将更新后的文本特征和视觉特征进行拼接融合。由于不同的特征对模型的贡献程度不同，本文采用两个全连接层神经网络计算每个特征向量的注意力权值，并将文本特征和视觉特征通过注意力权值进行融合，为了避免模态特征的误丢，对注意力权重进行加 1 操作。

$$\alpha = Softmax(\boldsymbol{W}_2^{TO}\tanh(\boldsymbol{W}_1^{TO}[\boldsymbol{T}_s,O'] + b_1^{TO}) + b_2^{TO}) \tag{17}$$

$$F = (1+\alpha)[\boldsymbol{T}_s,O'] \tag{18}$$

式中：$\boldsymbol{W}_1^{TO}$、$\boldsymbol{W}_2^{TO}$ 表示学习参数矩阵；b_1^{TO}、b_2^{TO} 表示偏移量。

2.6 谣言检测

将融合后的多模态特征输入带有 ReLU 和 Softmax 激活函数的两个全连接层进行谣言检测，记谣言检测器为 D：

$$D(T,O) = ReLU(W_2^D\cdot Softmax(W_1^D\cdot F + b_1^D) + b_2^D) \tag{19}$$

最终模型预测信息是谣言的概率：

$$\hat{y} = D(T,O) \tag{20}$$

为提高模型检测准确率，设谣言标签为 1，非谣言标签为 0，使用交叉熵损失函数优化谣言分类器。即：

$$L_3 = \sum -[y\log(\hat{y}) + (1-y)\log(1-\hat{y})] \tag{21}$$

2.7 训练过程

在模型训练过程中，特征提取器和谣言检测器需要协作，以降低损失函数 L_3 的值，从而提高模型检测的性能，

同时考虑到多模态信息之间的约束作用，通过最小化对比损失和三元组损失以准确挖掘模态特征间的关系。因此，本文对最终损失函数进行改进，将对比损失、三元组损失以及谣言分类的交叉熵损失之和作为本模型的最终优化目标，即：

$$L = \alpha_1 L_1 + \alpha_2 L_2 + L_3 \tag{22}$$

式中：α_1、α_2 是控制对比学习和三元组学习与目标函数之间的权重系数。使用反向传播不断优化最终的损失函数，从而更新模型参数以找到模型的最优状态。

3 实 验

3.1 数据集和实验设置

本文实验涉及的数据集分别来自 Weibo 和 Twitter。其中，Weibo 数据集由 Jin 团队[14]提供，具体谣言内容来自微博平台辟谣系统在 2012 年 5 月～2016 年 1 月期间被验证过的虚假谣言帖子，非谣言内容来自新华社认证的帖子；Twitter 数据集来自 MediaEval2015[25]，其中多模态包括帖子的文本信息、图片信息、发布时间以及用户特征等，参考文献[26]的数据处理方法，将其他信息筛除，只保留文本信息和图片信息，剩下的数据按照 4∶1 的比例分成训练集和验证集，数据集统计信息如表 1 所示。

表 1 数据集相关信息统计

Table 1 Data set related information statistics

数据集	Weibo	Twitter
谣言	4 211	7 321
非谣言	3 642	8 630

实验环境：Windows×11_64；GTX3050；Python3.8；Pytorch1.2.0、TensorFlow2.2.0。

对于文本特征，采用 BERT 模型进行特征提取文本长度设置为 50，单词向量为 768。对于视觉特征，使用 VGG-19 倒数第 2 层输出向量，其维度为 4 096。文本和视觉特征的全连接层为 1 024，激活函数为 ReLU 函数。多模态融合维数为 1 024，融合层 $k=2$。

模型的批处理大小为 128，训练次数为 300，学习率为 0.000 5，dropout 为 0.4，设置早停机制，优化器选择 Adam[27]寻找网络最优参数。

通过上述参数设置，模型在训练集上的损失训练变化如图 4 所示。

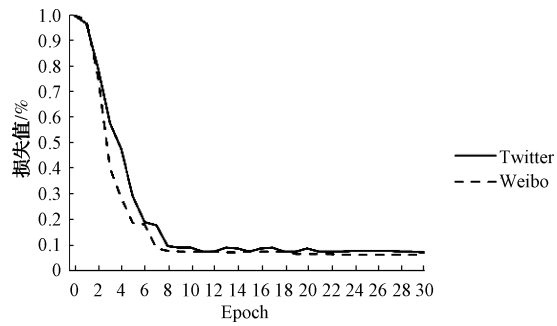

图 4 训练集模型损失变化

Fig. 4 Training set model loss changes

3.2 实验涉及模型

1）BERT

只考虑输入信息的文本内容，使用 BERT 模型进行特征提取，然后将输出的特征向量直接输入分类器进行谣言检测。

2）VGG-19

只考虑输入信息的图片内容，使用 VGG-19 模型进行特征提取，然后将输出的视觉特征输入分类器进行谣言检测。

3）EANN[11]

该模型不仅考虑文本内容也加入了图片信息，首先使用 VGG19 提取视觉特征，Text-CNN 提取文本特征，接着将两个模态特征进行拼接输入事件鉴别器，通过对抗神经网络学习特征之间稳定特征，从而进行谣言检测。

4）att-RNN[12]

同样以文本＋图片的方式进行模型训练，在多模态融合过程中加入注意力机制，通过 LSTM 生成文本特征和上下文特征联合表示的时间步特征作为注意力特征的输入，将注意力特征和视觉特征融合输入到分类器中进行谣言判断。

5）MRML[18]

该模型的输入信息为文本和图片两种模态，采用 BERT 和 VGG-19 特征提取器分别提取文本和视觉特征，使用深度度量学习计算模态内部谣言和非谣言的距离以及多模态间的距离，通过优化损失函数确认模态间的关系，最后将学习到的关系通过分类器判断是否为谣言。

6）MLMRD

本文提出的度量学习的多模态谣言检测方法，对于文本和图片信息，首先使用预训练模型分别提取文本和视觉特征，通过句法分析和注意力机制挖掘局部特征关系，接着使用度量学习确定模态内和模态间谣言和非谣言的关系，然后对模态的表示特征进行融合，最后使用分类器判别输入信息是否为谣言。

7）MLMRD-O

在本模型的基础上，去除视觉特征，将多模态转变为文本单模态特征进行检测，仅考虑文本的谣言和非谣言之间的关系确定特征参数提取特征，然后将特征输入到分类器进行谣言检测。

8）MLMRD-T

同 MLMRD-O 方法类似，在原来模型的基础上去除文本特征，转为视觉特征单模态检测。

9）MLMRD-

表示去除度量学习对模态关系的提取，仅依靠文本特征和视觉特征的简单拼接完成检测任务。

3.3 实验分析

1)对比分析

本文作为二分类任务，采用常用的准确率(accuracy)、精准率(precision)、召回率(recall)和 $F1$ 分值作为分析依据，依次对 3.2 节模型 1)～6)进行对比实验，从多角度分析特征关系对实验结果的影响。

根据预测值和真实值的组合情况，可以将实验结果组合成：预测为真－真实为真(TP)；预测为假－真实为假(TN)；预测为假－真实为真(FN)；预测为真－真实为假(FP)4 种情况。

由上述 4 种情况给出评价指标的计算公式为：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{23}$$

$$Precision = \frac{TP}{TP + FP} \tag{24}$$

$$Recall = \frac{TP}{TP + FN} \tag{25}$$

$$F1 = \frac{2Precision \cdot Recall}{Precision + Recall} \tag{26}$$

本文模型在 Weibo 和 Twitter 两个数据集上的实验结果的混淆矩阵如表 2、3 所示，对比结果如表 4 所示。

表 2 Weibo 数据集结果混淆矩阵

Table 2 Weibo dataset result confusion matrix

	预测值为假	预测值为真
真实值为假	0.87	0.08
真实值为真	0.02	0.89

表 3 Twitter 数据集结果混淆矩阵

Table 3 Twitter dataset result confusion matrix

	预测值为假	预测值为真
真实值为假	0.85	0.22
真实值为真	0.09	0.94

表 4 对比结果

Table 4 Comparison result (%)

模型	Weibo 数据集				Twitter 数据集			
	准确率	精准率	召回率	$F1$	准确率	精准率	召回率	$F1$
BERT	81.0	79.1	60.5	68.6	78.1	77.6	60.2	67.8
VGG-19	59.6	69.5	51.8	59.3	61.5	61.9	62.8	62.4
EANN	86.8	85.2	74.6	80.1	76.2	75.5	71.2	73.3
att-RNN	87.3	72.4	85.9	78.6	80.1	79.2	82.7	81.1
MRML	91.7	90.4	93.9	92.1	84.7	80.3	87.9	83.9
MLMRD	**94.7**	**91.9**	**97.8**	**94.8**	**85.2**	**81.5**	**90.6**	**85.8**

由表 4 可以看出，本文模型的各项评价指标均优于其他模型，表明本模型在多模态条件下能较好地提取文本和图片的有效信息完成谣言检测任务。观察实验数据发现，BERT 和 VGG-19 这两个单模态模型的性能最差，表明多模态模型提取的信息可以有效互补，文本和图片信息的结合可以提高模型的准确率。MLMRD 的性能优于 EANN 和 att-RNN，原因在于 EANN 模型仅采用简单拼接的方法完成特征融合，att-RNN 模型在探索模态的特征关系时遗漏了模态间的联系，而 MLMRD 模型考虑到信息的流动性，将模态间的联系考虑到模型中并对特征进行有效融合，该分析结果表明不同模态间的信息可以相互影响、相互制约，模态间的联系影响模型的检测效果。MLMRD 在 MRML 的基础上加入了局部关系提取操作，根据局部信息之间的联系为输入特征赋予不同的权重，充分发挥关键特征对模型的作用，同时确定模态内和模态间的特征关系并进行多模态特征融合，准确率高于 MRML，表明模态内局部关系影响模态语义的完整表示，多模态信息间存在着信息交互影响各模态的准确表示，将这两个因素引入模型对谣言检测的准确率有明显的提升作用。

综合以上对比结果可知，本文模型加入的度量学习和局部关系提取在获取模态内和模态间的特征关系上均有一定的效果，与最新模型 MRML 相比在精准率、召回率、$F1$ 值和准确率方面分别提高了 1.5%、2.9%、2.7%、1.1%。

2)模型效率分析

由于本文主要针对多模态信息进行谣言检测的研究，因此，关于模型在训练过程中的检测准确率的对比模型主要选择基线模型中涉及到多模态的模型，其对比结果如图 5 所示。从图 5 可以发现，MLMRD 的准确率要高于其

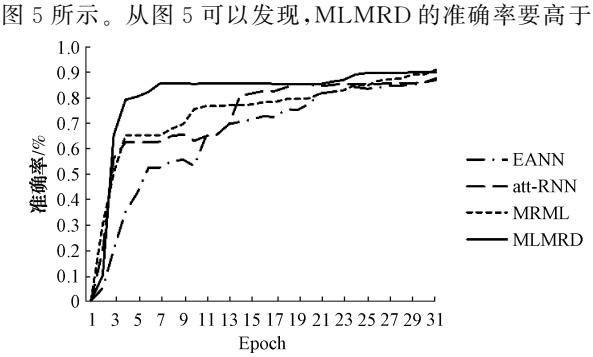

图 5 多模态模型准确率对比结果

Fig. 5 Accuracy comparison results of multimodal models

他基线模型，同时模型达到首次准确率峰值的训练次数最少，由此可以推断 MLMRD 模型的检测效率相对其他模型较好。

3）消融分析

为了验证多模态和模态特征关系对本模型的影响，特别设计了消融实验分析各模块的作用。其中 MLMRD-O 表示仅考虑文本特征；MLMRD-T 表示仅考虑视觉特征；MLMRD－表示去除度量学习对模态关系的提取。

消融实验的各项指标数据如表 5 所示，其对比结果如图 6、7 所示，由图 6、7 可知，单模态模型的准确率低于多模态模型的准确率，表明多模态模型能提取到更多有用的特征，因此多模态的研究对于谣言检测是有意义的。MLMRD-T 的准确率要低于 MLMRD-O 是因为图片提供的有效信息要比文本内容提供的少。与去除局部关系的 MLMRD－实验相比，MLMRD 的准确率更高，表明在图片和文本信息的内部特征之间存在一定的联系，在图片和文本的特征提取过程中，内部信息之间的约束作用确保发挥关键特征的价值，因此 MLMRD 可以提取更精确的表示特征，检测的精度更高。

表 5 消融结果

Table 5 Ablation result (%)

模型	Weibo 数据集				Twitter 数据集			
	准确率	精准率	召回率	*F*1	准确率	精准率	召回率	*F*1
MLMRD-O	87.2	87.6	87.5	87.5	82.6	76.2	63.1	69.0
MLMRD-T	87.1	84.7	64.1	73.0	84.7	80.1	88.7	84.2
MLMRD－	86.9	85.5	78.2	81.7	78.4	75.9	73.3	74.6
MLMRD	**92.8**	**91.9**	**97.8**	**94.8**	**85.2**	**81.5**	**90.6**	**85.8**

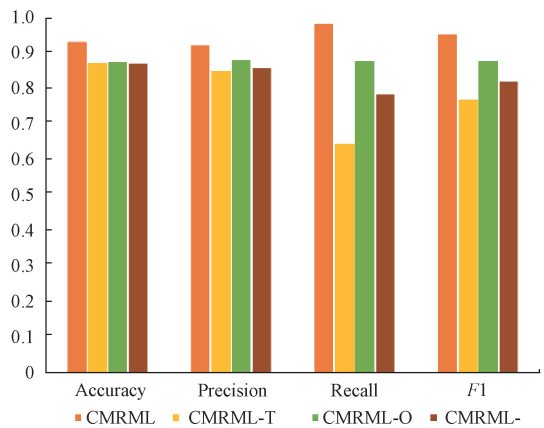

图 6 Weibo 数据集上消融实验结果

Fig. 6 Results of ablation experiments on the Weibo dataset

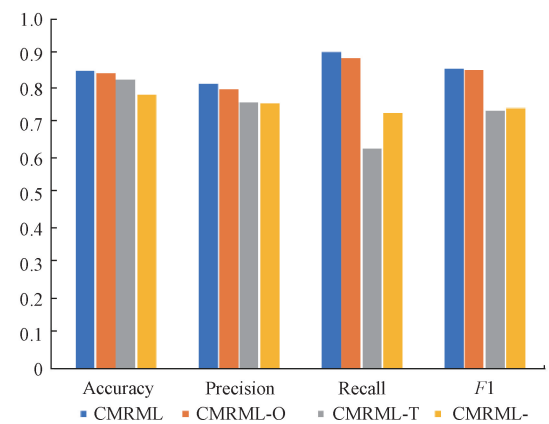

图 7 Twitter 数据集上消融实验结果

Fig. 7 Results of ablation experiments on the Twitter dataset

4 结 论

本文针对模态内局部特征间的联系、模态内和模态间的特征关系对模型准确率的影响问题，提出一种基于度量学习的多模态谣言检测模型。该方法利用句法分析和注意力机制挖掘模态局部关系并动态调节各模态特征的权重，然后引入度量学习确定谣言和非谣言的数据距离，同时将处理后的各模态特征进行融合，用融合特征进行模型训练以提高检测的准确率。在两个公开数据集上进行实验，发现该模型在准确率、精准率、召回率和 *F*1 分值各项指标上均有提高。结果表明模态内的局部特征联系、模态内和模态间的特征交互作用可以进一步提高谣言检测的准确率。

参 考 文 献

[1] 高玉君，梁刚，蒋方婷，等. 社会网络谣言检测综述[J]. 电子学报，2020，48(7)：1421-1435.
GAO Y J, LIANG G, JIANG F T, et al. Social network rumor detection: A survey [J], Acta Electronica Sinica, 2020, 48(7): 1421-1435.

[2] 刘华玲，陈尚辉，曹世杰，等. 基于多模态学习的虚假新闻检测研究[J]. 计算机科学与探索，2023，17(9)：2015-2029.
LIU H L, CHEN SH H, CAO SH J, et al. Survey of fake news detection with multi-model learning[J]. Journal of Frontiers of Computer Science & Technology, 2023, 17(9): 2015-2029.

[3] 陈林威,宋玉蓉,宋波.时序感知的异质图神经谣言检测[J].小型微型计算机系统,2024,45(1):45-51.
CHEN L W, SONG Y R, SONG B. Sequence-aware heterogeneous graph neural rumor detection [J]. Journal of Chinese Computer Systems. 2024, 45(1): 45-51.

[4] 郭秋实,李晨曦,刘金硕.引入知识表示的图卷积网络谣言检测方法[J].计算机应用研究,2022,39(7):2032-2036.
GUO Q SH, LI CH X, LIU J SH. Rumor detection with knowledge representation and graph convolutional network[J]. Application Research of Computers, 2022, 39(7): 2032-2036.

[5] 梁毅,吐尔地·托合提,艾斯卡尔·艾木都拉.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[J].计算机工程与科学,2023,45(6):1087-1096.
LIANG Y, TOHTI T, HAMDULLA A. Multi-modal false information detection via multi-layer CNN-based feature fusion and multi-classifier hybrid prediction[J]. Computer Engineering & Science, 2023, 45(6) :1087-1096.

[6] 孟佳娜,王晓培,李婷,等.基于对抗神经网络的跨模态谣言检测[J].数据分析与知识发现,2023,6(12):32-42.
MENG J N, WANG X P, LI T, et al. Cross-modal rumor detection based on adversarial neural network[J]. Data Analysis and Knowledge Discovery, 2022, 6(12) : 32-42.

[7] 强子珊,顾益军.基于多模态异质图的社交媒体谣言检测模型[J].数据分析与知识发现,2023,7(11):68-78.
QIANG Z SH, GU Y J. Detecting social media rumors based on multimodal heterogeneous graph[J]. Data Analysis and Knowledge Discovery, 2023, 7(11): 68-78.

[8] 高国鹏,房耀东,李彦芳,等.面向虚假新闻检测的社交媒体多模态数据集构建[J].网络与信息安全学报,2023,9(4):144-154.
GAO G P, FANG Y D, LI Y F, et al. Construction of multi-modal social media dataset for fake news detection [J]. Chinese Journal of Network and Information Security, 2023, 9(4): 144-154.

[9] WU K, YANG S, ZHU K Q. False rumors detection on Sina Weibo by propagation structures[C]. IEEE International Conference on Data Engineering, 2015: 651-662.

[10] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]. Proceedings of IJCAI, 2016.

[11] CHEN W L, ZHANG Y, YEO C K, et al. Unsupervised rumor detection based on users' behaviors using neural networks [J]. Pattern Recognition Letters, 2018, 105(C): 226-233.

[12] XU N, CHEN G D, MAO W J. MNRD: A merged neural model for rumor detection in social media[C]. IEEE International Joint Conference on Neural Network, 2018: 1-7.

[13] WANG Y Q, MA F L, JIN Z W, et al. EANN: Event adversarial neural networks for multi-modal fake news detection [C]. ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2018: 849-857.

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

[15] WU Y, ZHAN P W, ZHANG Y J, et al. Multimodal fusion with co-attention networks for fake news detection [C]. Annual Meeting of the Association for Computational Linguistics, 2021: 2560-2569.

[16] 戚力鑫,万书振,唐斌,等.基于注意力机制的多模态融合谣言检测方法[J].计算机工程与应用,2022,58(19):209-217.
QI L X, WAN SH ZH, TANG B, et al. Multimodal fusion rumor detection method based on attention mechanism [J]. Computer Engineering and Applications, 2022, 58(19) :209-217.

[17] SUÁREZ J L, GARCÍA S, HERRERA F. A tutorial on distance metric learning: Mathematical foundations, algorithms and software [C]. Computing Research Repository, 2018.

[18] ZHANG W J, GUI L, HE Y L. Supervised contrastive learning for multimodal unreliable news detection in COVID-19 pandemic[C]. International Conference on Information and Knowledge Management, 2021: 3637-3641.

[19] PENG L, JIAN S, LI D S, et al . MRML: Multimodal rumor detection by deep metric learning[C]. ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023:1-5.

[20] RAN H Y, JIA C Y, YU J. A metric-learning method for few-shot cross-event rumor detection[J]. Neurocomputing, 2023, 533: 72-85.

[21] JACOB D, 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: Human Language Technologies, 2019:4171-4186.

[22] KAREN S, ANDREW Z. Very deep convolutional networks for large-scale image recognition[J]. ArXiv preprint arXiv:1409.1556, 2014.

[23] 王艳新,闫静,王建华,等.基于特征融合度量学习的高压断路器机械故障诊断[J].仪器仪表学报,2022,43(9):98-105.
WANG Y X, YAN J, WANG J H, et al. Mechanical fault diagnosis for high voltage circuit breaker via a novel feature fusion metric learning [J]. Chinese Journal of Scientific Instrument, 2022,43(9):98-105.

[24] 张红颖,田鹏华.结合残差网络与多级分块结构的步态识别方法[J].电子测量与仪器学报,2022,36(6):66-72.
ZHANG H Y, TIAN P H. Gait recognition method combining residual network and multi-level block structure[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(6): 66-72.

[25] CHRISTINA B, KATERINA A, SYMEON P, et al. Verifying multimedia use at MediaEval [C]. MediaEval Benchmarking Initiative for Multimedia Evaluation, 2016.

[26] DHRUV K, JAIPAL S G, MANISH G, et al. MVAE: Multimodal variational autoencoder for fake news detection [C]. The Web Conference, 2019: 2915-2921.

[27] 赵志杰,张艳艳,毛翔宇.基于改进 Adam 优化算法的中文短文本分类方法[J].电子测量技术,2022,45(23):132-138.
ZHAO ZH J, ZHANG Y Y, MAO X Y. Research on Chinese short text classification method based on improved Adam optimization algorithm [J]. Electronic Measurement Technology, 2022, 45(23): 132-138.

作 者 简 介

李娜,硕士研究生,主要研究方向为自然语言处理。

余晓栋,博士,讲师,主要研究方向为自然语言处理。

朱节中(通信作者),硕士,教授,主要研究方向为大数据处理、计算机测控软件。

E-mail:xiaoyuyuann@163.com