

基于矩阵分解双线性池化的多模态 融合虚假新闻检测^{*}

王 婕, 刘 芸, 纪淑娟[†]

(山东科技大学 山东省智慧矿山信息技术重点实验室, 山东 青岛 266590)

摘 要: 现有的大多数虚假新闻检测方法将视觉和文本特征串联拼接, 导致模态信息冗余并且忽略了不同模态信息之间的相关性。为了解决上述问题, 提出一种基于矩阵分解双线性池化的多模态融合虚假新闻检测算法。首先, 该算法将多模态特征提取器捕捉的文本和视觉特征利用矩阵分解双线性池化方法进行有效融合, 然后与虚假新闻检测器合作鉴别虚假新闻; 此外, 在训练阶段加入了事件分类器来预测事件标签并去除事件相关的依赖。在 Twitter 和微博两个多模态谣言数据集上进行了对比实验, 证明了该算法的有效性。实验结果表明提出的模型能够有效地融合多模态数据, 缩小模态间的异质性差异, 从而提高虚假新闻检测的准确性。

关键词: 虚假新闻检测; 社交媒体; 多模态; 双线性池化; 深度学习

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1001-3695(2022)10-012-2968-06

doi: 10.19734/j.issn.1001-3695.2022.03.0139

Matrix factorized bilinear pooling for multi-modal fusion fake news detection

Wang Jie, Liu Yun, Ji Shujuan[†]

(Shandong Key Laboratory of Wisdom Mine Information Technology, Shandong University of Science & Technology, Qingdao Shandong 266590, China)

Abstract: Most of the existing methods only connect multi-modal features, without considering the complex connection between modes. To overcome these limitations, this paper proposed an end-to-end algorithm named matrix factorized bilinear pooling for multi-modal fusion fake news detection. Firstly, the algorithm effectively fused the text and visual features captured by multi-modal feature extractor utilizing matrix factorized bilinear pooling method, and then cooperated with fake news detector to classify posts as fake or not. It added the event classifier to remove the event-specific features. This paper compared with this algorithm and the other algorithms on two multimedia datasets, Twitter and Weibo. Experimental results show that the proposed algorithm can effectively fuse multi-modal data, narrow the heterogeneity difference between modes, so as to improve the accuracy of fake news detection.

Key words: fake news detection; social media; multi-modal; bilinear pooling; deep learning

0 引言

近年来, 人们获取新闻的主要平台已经从报纸等纸质媒介转向以 Twitter、微博等为代表的新兴社交媒体。截至 2020 年, 互联网用户超过 45 亿人, 其中社交媒体用户已超过 38 亿^[1]。然而, 社交媒体的便捷性与低门槛也造成了虚假新闻的泛滥, 可能造成严重的社会问题。虚假新闻是指那些故意捏造而成且可以验证为假的新闻, 例如在 2016 年美国总统大选最后三个月里, 关于两位候选人的虚假新闻在 Facebook 上被分享高达 3 700 万次^[2]。因此, 虚假新闻检测对社交媒体平台的可信环境构造至关重要。由于图像、视频等多模态数据能够吸引更多的关注, 并且视觉信息可以作为辅助信息对新闻的文本内容进行补充, 虚假新闻检测技术的研究重点已经从基于文本内容的检测转向基于多模态内容的检测。例如, Jin 等人^[3]使用循环神经网络(recurrent neural network, RNN)学习文本和社会上下文特征的联合表示, 利用 RNN 的输出对预训练的 VGG-19 提取的视觉特征进行对齐, 两者拼接后放入分类器中进行检测; Wang 等人^[4]通过借鉴对抗思想学习可迁移到不同事件的事

件无关特征。该模型的文本和视觉特征通过卷积神经网络(convolutional neural network, CNN)自动提取, 并与虚假新闻检测器合作完成识别假新闻的主要任务。事件鉴别器学习每个新闻的事件标签; Khattar 等人^[5]通过多模态变分自编码器学习共享(视觉+文本)表示以检测假新闻, 其中, 虚假新闻检测器利用从双模态变分自动编码器获得的多模态表示对新闻进行分类, 解码器从潜在向量重建原始图像和文本。以上方法仅将多模态特征进行简单串联拼接, 没有考虑模态间的相关性。另外, 现有模型高度依赖于事件的特性, 在新事件上的检测效果不佳, 难以实现事件迁移并且会削弱模型的泛化能力。

针对上述研究中缺乏有效的特征融合机制以及无法有效地检测突发事件中的虚假新闻, 本文提出了一种端到端的基于矩阵分解双线性池化的多模态融合虚假新闻检测算法(BPMF)。该算法包括多模态特征提取器、多模态特征融合模块、虚假新闻检测器和事件分类器四个主要模块。首先, 多模态特征提取器采用预训练的 BERT_{base}来捕捉文本特征, 采用预训练的 VGG-19 模型来捕捉视觉特征; 在特征提取之后, 本文使用矩阵分解双线性池化方法融合多模态特征以最大化它们

收稿日期: 2022-03-28; 修回日期: 2022-05-23 基金项目: 国家自然科学基金资助项目(71772107, 62072288); 山东省自然科学基金资助项目(ZR2020MF044, ZR202102230289, ZR2019MF003, ZR2021MF104); 山东省研究生教育质量监督提升计划资助项目(2021); 山东省科技厅科研基金资助项目; 国家教育部人文社科基金资助项目(20YJAZH078, 20YJAZH127)

作者简介: 王婕(1998-), 女, 山东日照人, 硕士研究生, 主要研究方向为人工智能; 刘芸(1992-), 女, 山东青岛人, 博士研究生, 主要研究方向为跨媒体检索; 纪淑娟(1977-), 女(通信作者), 山东青岛人, 教授, 博导, 博士, 主要研究方向为人工智能、智能信息处理(jsjsuzie@sina.com)。

之间的相关性;然后将有效融合后的多模态表示输入虚假新闻检测器以识别虚假新闻;最后,本文加入了梯度反转层和事件分类器,将不同事件的多模态特征映射到同一特征空间,以得到来自多个模态的公共潜在子空间的事件共享特征。

本文的主要贡献有以下两点: a) 提出了一种基于矩阵分解双线性池化的多模态融合虚假新闻检测模型(BPMF),该模型融合了两种模态的特征,并学习更具泛化能力的事件共享特征,去除不可迁移的特定于事件的依赖性; b) 使用矩阵分解双线性池化方法将文本和视觉特征表示有效融合,充分利用多模态特征之间的相关性,缩小模态间的异质性差异。本文在来源于Twitter和微博的多媒体数据集上对提出的方法进行验证。

1 相关工作

1.1 虚假新闻检测

早期社交媒体谣言检测工作主要针对易获取的新闻文字内容,提取其语义或统计特征来检测新闻的可信度。例如,Castillo等人^[6]使用文本中的URL数量、情绪积极/消极词、特殊字符、表情符号、超链接等统计信息检测虚假新闻;Feng等人^[7]用上下文无关语法阶段树规则驱动文本特征进行虚假新闻检测;为了克服早期的社交媒体谣言检测工作主要集中在手动提取特征或规则上的缺陷,Ma等人^[8]利用词嵌入和RNN将事件的社会背景信息建模为可变长度的时间序列,抽取了相关推文组成新闻事件;Chen等人^[9]提出了一种基于RNN的深度注意力模型,通过有选择地学习序列帖子的时间表示识别假新闻,验证了RNN在研究中的有效性。

社交上下文功能是在用户在社交媒体上的社交参与^[10],它可以作为虚假新闻检测的辅助信息。Wu等人^[11]开发了一个基于图核的支持向量机分类器,该分类器结合了来自原始消息主题、响应情感、消息传播模式以及发送消息的用户特征。Jin等人^[3]通过将文本、社会语境和视觉内容与RNN融合来进行谣言检测任务。然而,由于社会背景特征具有嘈杂、无结构等特性,在检测新出现的事件时不能提供足够的信息。

近年来,图像在微博等社交媒体中非常流行,并且对社交媒体新闻传播有很大的影响^[12]。然而,先前工作中提取的图像特征仍然是手工制作的^[13,14],可能很难表示图像复杂的内在嵌入内容。为了解决这个问题,研究人员试图利用深度学习方法捕获更深层次的图像特征实现虚假新闻检测。例如,Jin等人^[3]采用注意力机制学习文本和社会语境中的关键信息,并利用RNN的输出调节VGG-19抽取的图像特征的权重。然而,这些虚假新闻检测模型在新兴社交媒体事件上效果不佳,因为它们高度依赖于从现有数据中捕获特定事件的特征,而这些特征无法传输到新出现的事件。为了解决这个问题,Wang

等人^[4]利用对抗学习的方法引导模型学习到更具泛化性能的事件共享的多模态特征;Khattar等人^[5]提出了一种多模态变分自编码器MVAE来发现各模态之间的相关性,使用编码—解码的方式学习多模态特征表示,并利用重构损失保证编码后的中间表达可以解码回原状态;Zhang等人^[15]提出了模型BDANN(BERT-based domain adaptation neural network),通过添加领域分类器来学习可转移的领域共享特征。然而,一些研究认为背景知识是人们辨别虚假新闻时经常利用的辅助信息。除此之外,Wang等人^[16]将视觉信息转换为图的节点,还从真实世界的知识图谱中提取外部知识作为图的节点补充语义信息来进行虚假新闻检测;Qian等人^[17]提出了一种知识感知多模态自适应图卷积网络KMAGCN,将文本信息、知识概念和视觉信息联合建模到一个统一的虚假新闻检测框架中来捕获语义表示。然而,上述研究仅通过串联新闻的多模态特征进行拼接融合,难以充分捕获模态之间的复杂关联。

1.2 多模态特征融合

通过适当的融合技术有效组合多个来源的信息是研究多模态学习任务的关键问题^[18]。然而,现有的大多数虚假新闻检测方法只是将新闻的视觉和文本特征简单地拼接,以获得新闻的多模态表示,忽略了不同模态之间的潜在联系。因此,一些多模态特征融合的方法被提了出来,例如,Zadeh等人^[19]提出了多模态张量融合网络,通过计算张量外积来获得模态相关性;Hou等人^[20]提出了分层多项式融合网络,建立了多项式张量池块,通过高阶矩阵集成多模态特征;Xu等人^[21]提出了多交互记忆网络,通过递归神经网络对多个交互记忆注意力进行非线性组合以学习全局记忆抽象;Zhou等人^[22]提出了图像字幕模型将图像转换成句子,然后通过测量原文本内容与生成句子的相似度来计算多模态不一致性;Xue等人^[23]利用权值共享将文本和视觉特征转换成一个公共的特征空间,然后计算变换后的多模态特征的余弦相似度。上述方法验证了多模态信息的有效交互和对齐,可以提高预测结果的准确率。

2 基于矩阵分解双线性池化的多模态融合网络

针对现有研究方法的局限性,本文提出了一种端到端的基于矩阵分解双线性池化的多模态融合虚假新闻检测模型(BPMF),模型总体架构如图1所示。该模型集成了四个主要模块,即多模态特征提取器、多模态特征融合模块、虚假新闻检测器和事件分类器。首先,用多模态特征提取器从文本文章和附加图像中提取文本和图像特征,然后通过多模态特征融合模块进行特征融合,并将最终的多模态特征表示输入虚假新闻检测器进行预测,最后事件分类器基于深度融合特征识别新闻的事件标签,并删除特定事件的依赖。

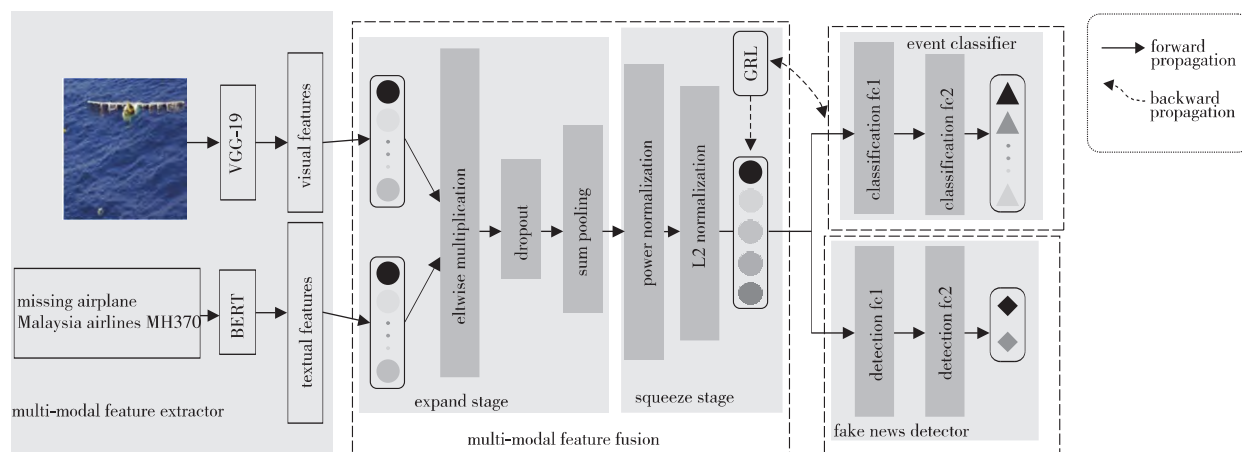


图1 BPMF 框架

Fig.1 Schematic of BPMF framework

2.1 多模态特征提取器

本模块的输入为新闻的文本和图像,输出为从两种模态中学习到特征表示。它可以分为文本特征提取器和视觉特征提取器。

2.1.1 文本特征提取器

为了捕获新闻文本信息中的潜在语义和上下文含义,本文使用以 transformer 为核心的 BERT 模型进行文本特征提取。BERT 本质上是一种基于文献[24]中所述原始实现的多层双向变压器编码器,能捕获底层语义和上下文信息。

首先将输入文本中的单词嵌入到向量中。其中第 i 个词表示为 $W_i \in R^k$, k 为词向量的维度。因此输入句子 i 表示如下:

$$S = [W^0, W^1, \dots, W^n] \quad (1)$$

其中: n 表示句子中的单词数; W^0 是添加在每个输入示例前面的特殊符号 [CLS]^[25]。

将由 12 个编码器层组成的预训练 BERT_{base} 模型记为 R_b 。将 S 输入 R_b 后得到给定句子的特征向量,计算如下:

$$S_f = [W_f^0, W_f^1, \dots, W_f^n] \quad (2)$$

由于每个单词对句子表示的重要性不同,对于每一个特征向量 S_f ,本文根据重要性采用均值池化方法从所有单词中获取文本特征表示 $R_t \in R^{d_t}$,其中 d_t 表示从 BERT 提取的文本特征维度。最后将 R_t 输入到一个全连接层 (full connected layer) 以确保文本特征的最终输出 $R_{tf} \in R^p$ 具有与视觉特征相同的维度 p 。计算如下:

$$R_{tf} = \phi_t(W_{tf} \cdot R_t) \quad (3)$$

其中: $W_{tf} \in R^{d_t \times p}$ 为完全连接层的权重矩阵; ϕ_t 为文本特征提取器中使用的 LeakyReLU 激活函数^[26]。

2.1.2 视觉特征提取器

卷积神经网络 (CNN) 在各种视觉理解任务中非常有效^[27]。为了有效提取视觉特征,本文使用已经在 ImageNet 上预训练过的 VGG-19^[28] 网络提取输入图像 V 中的视觉特征, VGG-19 获得的图像特征 R_v 的维数表示为 d_v 。本文冻结 VGG-19 网络的参数以避免过拟合,并将 R_v 输入全连接层以确保与文本特征 R_{tf} 相同的维度。计算如下:

$$R_{vf} = \phi_v(W_{vf} \cdot R_v) \quad (4)$$

其中: $W_{vf} \in R^{d_v \times p}$ 为权重矩阵; R_{vgg} 为 VGG-19 网络的输出; ϕ_v 为特征提取器中使用的 LeakyReLU 激活函数。

2.2 多模态特征融合模块

现有的大多数虚假新闻检测方法将视觉和文本特征串联拼接不利于发现不同模态之间复杂的潜在联系。因此在获得最终的文本特征表示 R_{tf} 和视觉特征表示 R_{vf} 后,本文使用矩阵分解双线性池化方法将其融合,以最大限度地提高文本和视觉特征表示之间的相关性。计算如下:

$$R_{TV} = R_{tf}^T W_i R_{vf} \quad (5)$$

其中: $W_i \in R^{m \times n}$ 为投影矩阵; R_{TV} 为多模态融合模块的输出。双线性池化在有效地捕获了特征维度之间的成对交互的同时引入了大量的参数,这可能导致较高的计算成本和过拟合的风险。为了减少参数量, W_i 在式 (6) 中被分解为两个低秩矩阵:

$$R_{TV} = R_{tf}^T U_i V_i^T R_{vf} = \sum_{d=1}^k R_{tf}^T U_d V_d^T R_{vf} = \mathbf{1}^T (U_i^T R_{tf} \circ V_i^T R_{vf}) \quad (6)$$

其中: k 为分解矩阵 $U_i = [u_1, \dots, u_k] \in R^{m \times k}$ 和 $V_i = [v_1, \dots, v_k] \in R^{n \times k}$ 的潜在维数; \circ 表示两个向量的元素相乘; $\mathbf{1} \in R_k$ 为一个全 1 向量。为了得到式 (6) 的输出特征 $R_{TV} \in R^o$, 需要学习 $U = [U_1, \dots, U_o] \in R^{m \times k \times o}$ 和 $V = [V_1, \dots, V_d] \in R^{n \times k \times o}$ 两个三阶张量。可以将 U 和 V 进一步转换为二维矩阵 $\hat{U} \in R^{m \times ko}$ 和 $\hat{V} \in R^{n \times ko}$ 。因此,式 (6) 可以改写为

$$R_{TV} = \text{SumPooling}(\hat{U}^T R_{tf} \circ \hat{V} R_{vf}, k) \quad (7)$$

其中: $\text{SumPooling}(x, k)$ 表示使用大小为 k 的一维非重叠窗口

在 x 上执行求和池化。另外,本文将多模态特征的提取和融合过程定义为 B , 计算如下:

$$R_{TV} = B(X; \theta_B) \quad (8)$$

其中: X 代表一组由文本和附加图像组成的新闻; θ_B 为特征提取器和多模态特征融合模块的训练参数集; B 为映射函数; R_{TV} 为融合后的多模态特征表示。

2.3 虚假新闻检测器

本文中检测器包含两个具备相应激活功能的全连接层,并通过 softmax 函数输出预测结果。本文将检测器表示为 $D(R_{TV}; \theta_D)$, 其中 θ_D 表示包含的所有参数, D 表示检测器的映射函数。虚假新闻检测器的输入为融合后的特征表示,输出为该新闻被预测为假的概率,计算如下:

$$\hat{y} = D(R_{TV}; \theta_D) \quad (9)$$

本文用 Y 表示新闻的真实标签,将虚假新闻标记为 1,真实新闻标记为 0,并采用交叉熵计算分类损失,计算如下:

$$L_D(\theta_B, \theta_D) = -E_{(x, y) \sim (X, Y)} [y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \quad (10)$$

为了优化提取、融合、检测时的参数 θ_B 和 θ_D , 最小化分类损失,计算如下:

$$(\hat{\theta}_C, \hat{\theta}_D) = \arg \min_{\theta_B, \theta_D} L_D \quad (11)$$

2.4 事件分类器

事件分类器由两个全连接层和相应的激活函数组成,通过输入的多模态特征表示将新闻分类为 N 个事件之一。本文将事件分类器表示为 $C(R_{TV}; \theta_C)$, 其中 θ_C 表示事件分类器的参数集, C 表示事件分类器的映射函数。将事件标签集记为 Y_e , 并采用交叉熵定义分类器的损失。计算如下:

$$L_C(\theta_B, \theta_C) = -E_{(x, y_e) \sim (X, Y_e)} \left[\sum_{n=1}^N y_e(C(B(x; \theta_B); \theta_C)) \right] \quad (12)$$

为了优化参数 θ_C , 需要最小化损失 $L_C(\theta_B, \theta_C)$ 。计算如下:

$$\hat{\theta}_C = \arg \min_{\theta_C} L_C(\theta_B, \theta_C) \quad (13)$$

根据式 (13), 损失 L_C 可用于估计不同事件分布的差异性,当损失值较大时表示网络学习事件共享的多模态特征。因此需要最大化分类器损失 L_C 来去除事件的特殊性。这种想法促进了多模态特征提取器及融合模块与事件分类器之间的对抗过程。一方面,多模态特征提取器及融合模块倾向于通过最大化事件分类损失 L_C 学习可转移的事件共享特征;另一方面,事件分类器倾向于通过最小化事件分类损失 L_C 从多模态特征表示中发现事件特定信息,以识别事件。

2.5 模型集成

在训练过程中,需要最小化 L_D 以提高模型的性能。为了获得事件共享特征,需要最大化事件分类器损失 L_C ; 同时,事件分类器通过最小化损失 L_C 尝试从多模态特征表示中发现事件特定信息。因此,总损失定义如下:

$$L_C(\theta_B, \theta_D, \theta_C) = L_D - \lambda L_C \quad (14)$$

本文利用系数 $\lambda \in R$ 来平衡虚假新闻检测器和事件分类器的损失函数。为了解决 2.4 节所述的 minimax 博弈,本文采用了梯度反转层 (GRL)^[29] 即在前向传播过程中实现恒等变换,在反向传播过程中, GRL 将梯度乘以 $-\lambda$ 传递到前一层。这避免了一般的对抗模型分步训练的情况,本文将 GRL 层放置于多模态特征融合模块和事件分类器之间。因此,模型参数的优化过程如下:

$$\theta_B \leftarrow \theta_B - \eta \left(\frac{\partial L_D}{\partial \theta_B} - \lambda \frac{\partial L_C}{\partial \theta_B} \right), \theta_D \leftarrow \theta_D - \eta \frac{\partial L_D}{\partial \theta_D}, \theta_C \leftarrow \theta_C - \eta \frac{\partial L_C}{\partial \theta_C} \quad (15)$$

3 实验结果与分析

3.1 数据集

为了评估所构建的模型对虚假新闻的检测效果,本文使用外文 Twitter 和中文微博两个真实的社交媒体数据集进行实

验,两个数据集的统计信息如表 1 所示。

表 1 数据集统计
Tab. 1 Statistics of datasets

统计信息	Twitter	微博
fake news	7 898	4 749
real news	6 026	4 779
all	13 924	9 528

a) Twitter 数据集^[30]用于验证多媒体使用任务。数据集收集了大量发布在 Twitter 上的新闻,每个新闻都包含附加图片/视频、文本以及标签。由于本文的工作只考虑文本和图片信息,删除了带有视频的推文。

b) 微博数据集是 Jin 等人^[3]基于新浪微博平台上构建的多模态的中文虚假新闻数据集。由微博知名用户组成的委员会对平台用户举报的可疑新闻进行核实和判断,微博数据集收集了 2012 年 5 月~2016 年 1 月被微博官方证实的虚假新闻,同时收集了中国权威新闻来源新华社证实的真实新闻。为了确保数据集的完整性,使用近似重复图像检测算法^[31]删除重复的低质量图像,然后删除缺失图像的新闻。

3.2 实验设置

本节基于 Python 环境,利用 numpy、pandas、torch 和 sklearn 库来进行实验。使用准确度、精确度、召回率和 F_1 值等评价标准与基线方法进行比较,从而评估算法的性能。

首先采用标准文本预处理程序^[32]对两个数据集中的文本进行预处理。对于 Twitter 数据集,删除了推文的标点符号、数字、特殊字符和短单词;对于微博数据集,使用 Python 里的 Jieba 模块对新闻进行数据分割。然后使用文本特征提取器 BERT_{base} 提取 768 维文本特征,对于新闻附带的图像,将其大小调整为 $224 \times 224 \times 3$,然后利用在 ImageNet 上预训练的 VGG-19^[26] 提取维数为 4 096 的视觉特征,同时将 BERT_{base} 和 VGG-19 的参数冻结来避免过拟合。最后,设置虚假新闻检测器的全连接层大小为 64,事件分类器为 64 和 32 的两个完全连接的层组成。将本文模型的批处理大小(batch_size)设置为 128,训练 100 次,学习率为 0.001,使用 Adam 优化器^[33]优化模型中的参数。

3.3 比较算法

为了验证本文所提出的模型在虚假新闻检测任务中的性能,根据输入内容将比较算法分为单模态虚假新闻检测模型和多模态虚假新闻检测模型。

3.3.1 单模态虚假新闻检测模型

为了更好地验证多模态信息对虚假新闻检测的有效性,使用如下所述的两个单模态模型进行对比:

a) Textual,该模型的输入仅包含新闻中的文本信息。首先,将新闻的文本信息输入到一个预先训练的 BERT_{base} 模型得到每篇文章的文本特征 R_t ,然后将 R_t 输入到一个具有 softmax 功能的 32 维全连接层来检测虚假新闻。

b) Visual,该模型的输入仅包括新闻中的附加图像。能利用一个预先训练的 VGG-19 和全连接层来提取视觉特征 R_v ,然后将 R_v 输入到一个 32 维的全连接层来检测虚假新闻。

3.3.2 多模态虚假新闻检测模型

为了更好地验证本文所提出的模型在多模态虚假新闻检测任务中的性能,将目前最先进的多模态虚假新闻检测模型及变体与本文的模型进行对比。

a) VQA^[34]。视觉问答(VQA)研究聚焦于回答特定图像的相关问题,该研究利用特征级联来代替原本的文本和图像之间逐元素相乘方法,并用二分类层代替最后的多类层。因此原本适用于多类分类任务的 VQA 模型可以适用于二分类任务。

b) NeuralTalk^[35]。NeuralTalk 旨在生成描述给定图像自然语言句。根据 NeuralTalk 的主要结构,新闻的多模态表示为通过 RNN 每个时间步长输出平均,最后将潜在表示输入到一个 32 维全连接层进行虚假新闻检测。

c) att-RNN^[3]。att-RNN 通过将文本、社会语境和视觉内容与 RNN 融合来进行谣言检测任务。首先通过 LSTM 提取文本和社会背景特征获得文本/社会语境联合表示,然后从预先训练的 VGG-19 模型提取视觉特征,并利用注意力机制对视觉特征与文本/社会语境特征进行特征对齐,最后输入到分类器中。

d) EANN^[4]。通过借鉴对抗思想学习可迁移到不同事件的多模态表示,该模型通过 TextCNN 模型提取文本特征,通过预训练的 VGG-19 模型提取图像特征,将两者拼接作为新闻的多模态表示,并判断新闻的真假以及新闻的来源事件。EANN-模型不包含事件鉴别器。

e) MVAE^[5]。采用一种编码-解码的方法学习共享(视觉+文本)表示,以检测虚假新闻。通过训练多模态变分自编码器将文本和视觉特征拼接得到多模态表示,利用重构损失将多模态表示解码回原始模态,并利用多模态表示鉴别虚假新闻。

f) BDANN^[15]。BDANN 的文本特征通过预训练的 BERT 模型提取,视觉特征通过预训练的 VGG-19 模型提取,通过添加域分类器消除对特定事件的依赖。将不带域分类器的 BDANN 模型记为 BDANN-。

为了分析本文方法中不同模块的影响,还评估了其变体的性能。BPMF-m 表示没有使用多模态特征融合模块的模型,仅将文本特征和图像特征简单拼接作为下游任务的输入;BPMF-c 表示没有使用事件分类器的模型。

3.4 实验结果与分析

3.4.1 对比实验结果分析

为了验证本文提出模型的有效性,将其与其他基线方法在 Twitter 和微博这两个多模态谣言检测数据集上进行比较,并采用分类问题常用的评价指标,即准确率(accuracy)、精确率(precision)、召回率(recall)、 F_1 值(F_1 -score)。实验对比结果如表 2 所示。

表 2 各实验结果对比
Tab. 2 Comparison of experimental results for each method

数据集	算法	accuracy	fake news			real news		
			precision	recall	F_1	precision	recall	F_1
Twitter	Textual	0.706	0.648	0.540	0.589	0.715	0.636	0.673
	Visual	0.596	0.695	0.518	0.593	0.524	0.700	0.599
	VQA	0.631	0.765	0.509	0.611	0.550	0.794	0.650
	NeuralTalk	0.610	0.728	0.504	0.595	0.534	0.752	0.625
	att-RNN	0.664	0.749	0.615	0.676	0.589	0.728	0.651
	EANN-	0.648	0.810	0.498	0.617	0.584	0.759	0.660
	EANN	0.719	0.642	0.474	0.545	0.771	0.870	0.817
	MVAE	0.745	0.801	0.719	0.758	0.689	0.777	0.730
	BDANN-	0.821	0.790	0.610	0.690	0.830	0.920	0.870
	BDANN	0.830	0.810	0.630	0.710	0.830	0.930	0.880
	BPMF-m	0.836	0.820	0.640	0.730	0.840	0.930	0.890
	BPMF-c	0.846	0.740	0.810	0.770	0.910	0.860	0.880
	BPMF	0.884	0.770	0.910	0.830	0.950	0.870	0.910
	Textual	0.804	0.800	0.860	0.830	0.840	0.760	0.800
微博	Visual	0.608	0.610	0.605	0.607	0.607	0.611	0.609
	VQA	0.736	0.797	0.634	0.695	0.695	0.838	0.760
	Neural Talk	0.726	0.794	0.713	0.684	0.684	0.840	0.754
	att-RNN	0.772	0.797	0.713	0.692	0.684	0.840	0.754
	EANN-	0.794	0.790	0.820	0.800	0.800	0.770	0.780
	EANN	0.816	0.820	0.820	0.820	0.810	0.810	0.810
	MVAE	0.824	0.854	0.769	0.802	0.802	0.875	0.837
	BDANN-	0.814	0.800	0.860	0.830	0.840	0.760	0.800
	BDANN	0.842	0.830	0.870	0.850	0.850	0.820	0.830
	BPMF-m	0.847	0.860	0.840	0.850	0.830	0.850	0.840
	BPMF-c	0.852	0.850	0.880	0.860	0.860	0.820	0.840
	BPMF	0.860	0.850	0.890	0.870	0.870	0.830	0.850

从表 2 可以看出,在 Twitter 数据集上,只用新闻的文本信息检测虚假新闻比只用视觉信息进行检测的效果好。Textual 方法用 BERT 模型仅提取文本特征就能达到较高的准确率和 F_1 值,这表明先进的 BERT 模型能较好地捕获文本的双向特

征,并且文本信息在虚假新闻检测中至关重要。同时,可以观察到,基于多模态信息比基于单一模态进行虚假新闻检测的模型表现出更好的性能。在 Twitter 数据集上本文模型准确率和检测假新闻的 F_1 值与基线方法相比提升了 5.4%、12%,在微博数据集上的性能同样优于基线方法(1.8%、2%),证明本文算法能够更准确、高效地检测虚假新闻。

在在微博数据集上可以发现,在单模态模型中,基于文本的 Textual 模型的准确率比基于图像的 Visual 模型高 20%。此外,仅用 BERT 方法提取文本特征的 Textual 模型在中文微博数据集上的准确率比在外文 Twitter 数据集上的准确率高 10%。这是由于 Twitter 数据集包含的语言种类复杂,难以像微博数据集一样高效分词,所以准确率更低。此外,多模态虚假新闻检测方法都优于单模态方法,证明了新闻的附加图片可能提供检测虚假新闻的互补信息。在多模态模型中,本文提出的模型及其变体性能都优于基线方法,验证了本文提出的模型在社交媒体虚假新闻检测中的有效性。

3.4.2 消融实验分析

本文提出的模型及其变体在 Twitter 和微博上的性能如图 2 所示。为了验证事件分类器和多模态融合模块对所提出模型的重要性,分别对这两个模块进行分析。

1) 事件分类器的影响 为了识别新出现的事件,本文使用了事件分类器。为了验证事件分类器的影响,比较了本文提出的模型和变体 BPMF-c,它不包含事件分类器,因此倾向于学习事件特定的特征。与变体 BPMF-c 相比,本文提出的添加了事件分类器的模型准确率在 Twitter 和微博两个数据集上均有提高,可以更好地处理新出现的事件,说明了事件分类器在性能改进方面的有效性。

2) 多模态融合模块的影响 为了解决缺乏有效特征融合机制的问题,本文提出了多模态融合模块。为了评价提出的多模态融合模块,设计提出了模型的变体 BPMF-m,它没有包含多模态特征融合模块,只是将文本特征和视觉特征进行简单拼接输入到下游任务中,忽略了模态之间的交互和对齐。与变体 BPMF-m 相比,在获得最终的文本特征表示 R_{tf} 和视觉特征表示 R_{vf} 后,使用矩阵分解双线性池化方法将其融合,以最大限度提高文本和视觉特征表示之间的相关性,准确率在 Twitter 和微博两个数据集上均有提高。因此,在预测虚假新闻时,只添加多模态特征融合模块(BPMF-c)的预测效果要优于不添加多模态特征融合模块(BPMF-m)的效果,并且整合多模态特征融合模块和事件分类器的模型预测效果最好。

3.4.3 可视化分析

目前的图像处理技术仅仅将文本和视觉特征进行简单拼接来获得联合特征表示,这不利于捕捉不同模态之间的复杂关联,并且可能导致模态信息冗余,误导检测结果。因此,误导性新闻不能被正确分类。为了进一步分析多模态特征融合模块在模型中的有效性,在在微博数据集上用 t-SNE^[36] 方法将模型学习的多模态表示及其变体 BPMF-m 学习的两种模态联合表示可视化,如图 3 所示。

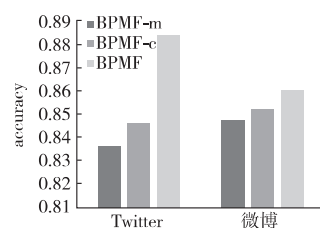


图2 模块分析
Fig. 2 Module analysis

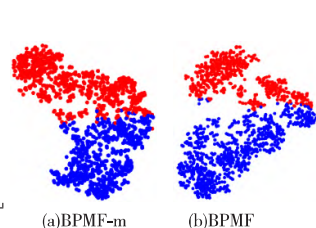


图3 基于微博测试数据的特征表示可视化
Fig. 3 Visualization of feature representation for Weibo data

图3 分别为变体 BPMF-m 和本文模型学习到的特征表示

在二维平面的投影。每条新闻的标签为真或者假。从图 3 可以看出,本文学习到的特征表示的可鉴别性比其变体的效果要好。如图 3(a) 所示,变体 BPMF-m 可以学习可识别的特征,但许多特征仍然被错误分类;相反,如图 3(b) 所示,本文模型学习到的融合特征表示更具有可鉴别性,两种样本之间的间隔更加明显。这是由于模型利用矩阵分解双线性池化方法将文本信息和视觉信息深度融合,最大限度地提高了多模态特征表示之间的相关性,获得了更好的联合表示,有助于模型对多模态新闻进行检测。因此可以得出结论,所提出的模型在多模态特征融合模块的作用下检测虚假新闻效果更好。

3.4.4 参数分析

为了研究所提出模型中参数对模型性能的影响,本文在 Twitter 和微博两个数据集上,通过修改学习率和模态特征 (R_{tf} 、 R_{vf}) 的维度两个参数设置了几组实验进行对比,实验结果如图 4、5 所示。

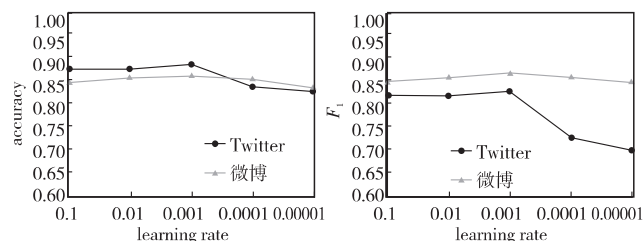


图4 不同学习率对结果的影响
Fig. 4 Influence of different learning rates on results

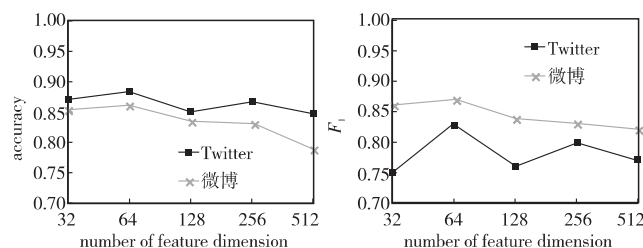


图5 不同特征维度对结果的影响
Fig. 5 Influence of different feature dimensions on results

1) 学习率的影响 图 4 是在 Twitter 和微博两个数据集上,不同大小的学习率对所提出模型性能的影响。从图中可以看出,当学习率为 0.001 时,模型准确率和 F_1 值在两个数据集上达到最高。因此在实验中将学习率设置为 0.001。

2) 特征维度的影响 图 5 是在 Twitter 和微博两个数据集上,不同特征维度对所提出模型性能的影响。从图中可以看出,维度过低或过高有可能造成模型准确率的降低。因此,本文将模态特征维度设置为 64,以实现更好、更稳定的性能。

3.4.5 收敛性分析

为了探究所提出模型的收敛性,图 6 显示了在 Twitter 和微博两个数据集上损失的变化。在训练开始阶段,损失值下降幅度大,说明学习率合适且进入梯度下降过程;在学习到一定阶段后,两个数据集上的损失曲线平稳收敛,表明模型进行了适当的学习。

3.5 样例分析

为了更加直观地证明所提出模型的有效性,本文将 BPMF 模型的结果与单模态模型及多模态模型的结果进行比较。图 7 为两个 Twitter 数据集上的虚假新闻样例被本文提出的模型成功检测到,但被单模态模型错误分类为真实新闻。图 7(a) 中的新闻文本内容正常的但其附加图像看起来相当可疑,可能是伪造或者篡改的,单模态模型的 Textual 模型将该新闻标志为真实新闻;图 7(b) 的图片看起来是正常的,然而带有强烈感情和煽动性文字的文本内容表明这是一个可疑的新闻,单模态

模型的 Visual 模型将其分类为真实新闻。本文提出的模型通过结合新闻的文本和视觉两种模态信息, 将这两篇新闻预测为假新闻。这表明所提出的模型能够有效地提取新闻的文本信息和其附加图像里的视觉信息, 并且附加图像的视觉内容为虚假新闻检测提供了更多的补充信息。图 8 显示了来自 Twitter 数据集的一些真实推文, 这些推文被提出的模型正确地分类了, 而基线模型(如 Textual 和 Visual) 和现有模型(如 EANN 和 BDANN) 对其产生了错误的预测。上述结果证明了所提出的模型在多模态虚假新闻检测工作上的有效性。

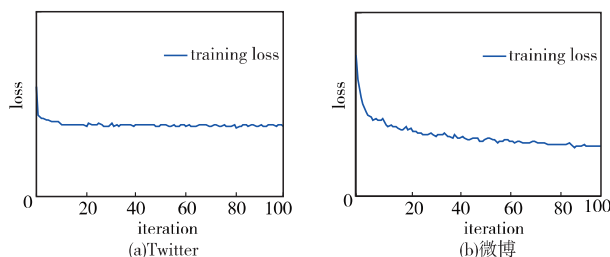


图 6 损失变化
Fig. 6 Change of loss



图 7 假推文被 BPMF 正确分类, 但被单模态模型错误分类
Fig. 7 Fake tweets correctly classified by BPMF but misclassified by single modality model

图 8 一些真实的推文被 BPMF 正确分类, 但被比较方法错误分类
Fig. 8 Some real tweets correctly classified by BPMF but misclassified by existing model

4 结束语

针对真实世界中新兴社交媒体新闻模态的多样性问题, 本文提出了一种基于矩阵分解双线性池化的多模态融合虚假新闻检测模型(BPMF)。通过利用矩阵分解双线性池化的方法更好地融合新闻中的多模态异构信息, 提高检测性能。首先, 新闻中的文本和视觉信息由多模态特征提取器提取, 并通过多模态特征融合模块有效融合, 然后输入到虚假新闻检测器以辨别新闻真假; 事件分类器的作用是去除不可迁移的事件相关依赖, 提升模型对新出现事件的泛化能力。在两个真实多媒体数据集上通过实验证实了本文模型的有效性。未来的工作将进一步研究文本和图像之间的语义对齐以获得更好的融合机制, 以及考虑利用音频和视频等更加新颖的新闻内容检测虚假新闻。

参考文献:

- [1] We Are Social Ltd. Digital around the world in April 2020 [EB/OL]. (2020-04-23). <https://wearesocial.com/uk/blog/2020/04/digital-around-the-world-in-april-2020/>.
- [2] Allcott H, Gentzkow M. Social media and fake news in the 2016 election [J]. *Journal of Economic Perspectives* 2017, 31(2): 211-236.
- [3] Jin Zhiwei, Cao Juan, Guo Han, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs [C]//Proc of the 25th ACM International Conference on Multimedia. New York: ACM Press 2017: 795-816.
- [4] Wang Yaqing, Ma Fenglong, Jin Zhiwei, 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. New York: ACM Press 2018: 849-857.
- [5] Khattar D, Goud J S, Gupta M, et al. MVAE: multimodal variational autoencoder for fake news detection [C]//Proc of World Wide Web Conference. New York: ACM Press 2019: 2915-2921.
- [6] Castillo C, Mendoza M, Poblete B. Information credibility on Twitter [C]//Proc of the 20th International Conference on World Wide Web. New York: ACM Press 2011: 675-684.
- [7] Feng Song, Banerjee R, Choi Y. Syntactic stylometry for deception detection [C]//Proc of the 50th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics 2012: 171-175.
- [8] Ma Jing, Gao Wei, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks [C]//Proc of the 25th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press 2016: 3818-3824.
- [9] Chen Tong, Li Xue, Yin Hongzhi, et al. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection [C]//Proc of Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer 2018: 40-52.
- [10] Shu Kai, Sliva A, Wang Suhang, et al. Fake news detection on social media: a data mining perspective [J]. *ACM SIGKDD Explorations Newsletter* 2017, 19(1): 22-36.
- [11] Wu Ke, Yang Song, Zhu K Q. False rumors detection on Sina Weibo by propagation structures [C]//Proc of the 31st IEEE International Conference on Data Engineering. Piscataway, NJ: IEEE Press 2015: 651-662.
- [12] Jin Zhiwei, Cao Juan, Zhang Yongdong, et al. Novel visual and statistical image features for microblogs news verification [J]. *IEEE Trans on Multimedia* 2016, 19(3): 598-608.
- [13] Gupta M, Zhao P, Han J. Evaluating event credibility on Twitter [C]//Proc of SIAM International Conference on Data Mining. [S. l.]: Society for Industrial and Applied Mathematics 2012: 153-164.
- [14] Tian Dongping. A review on image feature extraction and representation techniques [J]. *International Journal of Multimedia and Ubiquitous Engineering* 2013, 8(4): 385-396.
- [15] Zhang Tong, Wang Di, Chen Huanhuan, et al. BDANN: BERT-based domain adaptation neural network for multi-modal fake news detection [C]//Proc of International Joint Conference on Neural Networks. Piscataway, NJ: IEEE Press 2020: 1-8.
- [16] Wang Youze, Qian Shengsheng, Hu Jun, et al. Fake news detection via knowledge-driven multimodal graph convolutional networks [C]//Proc of International Conference on Multimedia Retrieval. New York: ACM Press 2020: 540-547.
- [17] Qian Shengsheng, Hu Jun, Fang Quan, et al. Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection [J]. *ACM Trans on Multimedia Computing, Communications, and Applications* 2021, 17(3): 1-23.
- [18] 刘建伟, 丁熙浩, 罗雄麟. 多模态深度学习综述 [J]. *计算机应用研究*, 2020, 37(6): 1601-1614. (Liu Jianwei, Ding Xihao, Luo Xionglin. Survey of multimodal deep learning [J]. *Application Research of Computers* 2020, 37(6): 1601-1614.)
- [19] Zadeh A, Chen Minghai, Poria S, et al. Tensor fusion network for multimodal sentiment analysis [C]//Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics 2017: 1103-1114.
- [20] Hou Ming, Tang Jiajia, Zhang Jianhai, et al. Deep multimodal multi-scale fusion with high-order polynomial pooling [C]//Proc of the 33rd International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc. 2019: 12156-12166.
- [21] Xu Nan, Mao Wenji, Chen Guandan. Multi-interactive memory network for aspect based multimodal sentiment analysis [C]//Proc of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press 2019: 371-378.

(下转第 2978 页)

来改进模型,进行相关的交通流预测研究。

参考文献:

- [1] 刘静,关伟.交通流预测方法综述[J].公路交通科技,2004(3):82-85. (Liu Jing,Guan Wei. A review of traffic flow prediction methods[J]. Highway Traffic Technology 2004(3):82-85.)
- [2] Zhang Suqi,Lin Kuoping. Short-term traffic flow forecasting based on data-driven model[J]. Mathematics 2020,8(2):152-169.
- [3] Zhou Teng,Jiang Dazhi,Lin Zhizhe,et al. Hybrid dual Kalman filtering model for short-term traffic flow forecasting[J]. IET Intelligent Transport Systems 2019,13(6):1023-1032.
- [4] Sun Bo,Sun Tuo,Zhang Yujia,et al. Urban traffic flow online prediction based on multi-component attention mechanism[J]. IET Intelligent Transport Systems 2020,14(10):1249-1258.
- [5] 曹阳,茅一波,施旋,等.日交通流预测的编码器-解码器深度学习模型研究[J/OL].计算机工程与应用,(2021-11-22).http://kns.cnki.net/kcms/detail/11.2127.TP.20210913.1903.020.html. (Cao Yang,Mao Yibo,Shi Quan,et al. Encoder-decoder deep learning model study for daily traffic flow prediction[J/OL]. Computer Engineering and Application. (2021-11-22).http://kns.cnki.net/kcms/detail/11.2127.TP.20210913.1903.020.html.)
- [6] 桂智明,李壮壮,郭黎敏,等.基于ACGRU模型的短时交通流预测[J].计算机工程与应用,2020,56(21):260-265. (Gui Zhiming,Li Zhuangzhuang,Guo Limin,et al. Short-time traffic flow prediction based on the ACGRU model[J]. Computer Engineering and Applications 2020,56(21):260-265.)
- [7] Yu Bing,Yin Haoteng,Zhu Zhanxing. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C]//Proc of the 27th International Joint Conference on Artificial Intelligence. 2018:3634-3640.
- [8] Li Yuguang,Yu R,Shahabi C,et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting[EB/OL].https://arxiv.org/abs/1707.01926.
- [9] Yang Gang,Wang Yunpeng,Yu Haiyang,et al. Short-term traffic state prediction based on the spatiotemporal features of critical road sections[J]. Sensors 2018,18(7):2287.
- [10] 姜山,丁治明,徐馨润,等.面向路网交通流态势预测的图神经网络模型[J].计算机科学与探索,2021,15(6):1084-1091. (Jiang Shan,Ding Zhiming,Xu Xinrun,et al. Graph neural network model[J]. Computer Science and Exploration 2021,15(6):1084-1091.)
- [11] 戴俊明,曹阳,沈琴琴,等.基于多时空图卷积网络的交通流预测[J].计算机应用研究,2022,39(3):780-784. (Dai Junming,Cao Yang,Shen Qinqin,et al. Traffic flow prediction based on multi-spatio-temporal graph convolutional networks[J]. Application Research of Computers 2022,39(3):780-784.)
- [12] Zhang Lizong,Alharbe N R,Luo Guangchun,et al. A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction[J]. Tsinghua Science and Technology 2018,23(4):479-492.
- [13] 倪庆剑,彭文强,张志政,等.基于信息增强传输的时空图神经网络交通流预测[J].计算机研究与发展,2022,59(2):282-293. (Ni Qingjian,Peng Wenqiang,Zhang Zhizheng,et al. Spatiotemporal graph neural network traffic flow prediction based on information-enhanced transmission[J]. Computer Research and Development, 2022,59(2):282-293.)
- [14] 张耀方,陈坚.基于GBDT算法的高速公路分车型交通流短时预测模型[J].公路,2022,67(1):221-227. (Zhang Yaofang,Chen Jian. Short-time prediction model of highway subvehicle traffic flow based on GBDT algorithm[J]. Highway 2022,67(1):221-227.)
- [15] Li Ming,Wang Yuewen,Wang Zhaowen,et al. A deep learning method based on an attention mechanism for wireless network traffic prediction[J]. Ad hoc Networks 2020,107:102258.
- [16] Tang Cong,Sun Jingru,Sun Yichuang. Dynamic spatial-temporal graph attention graph convolutional network for short-term traffic flow forecasting[C]//Proc of IEEE International Symposium on Circuits and Systems. Piscataway,NJ: IEEE Press 2020:1-5.
- [17] Zhu Jiawei,Wang Qiongjie,Tao Chao,et al. AST-GCN: attribute-augmented spatiotemporal graph convolutional network for traffic forecasting[J]. IEEE Access 2021,9:35973-35983.
- [18] 薛延明,李光辉,齐涛.融合图小波和注意力机制的交通流预测方法[J/OL].计算机科学与探索,(2022-03-16).https://kns.cnki.net/kcms/detail/11.5602.TP.20220315.1501.002.html. (Xue Yanming,Li Guanghui,Qi Tao. Traffic prediction method integrating graph wavelet and attention mechanism[J/OL]. Journal of Frontiers of Computer Science and Technology. (2022-03-16).https://kns.cnki.net/kcms/detail/11.5602.TP.20220315.1501.002.html.)
- [19] Wu Zonghan,Pan Shirui,Long Guodong,et al. Graph WaveNet for deep spatial-temporal graph modeling[C]//Proc of the 28th International Joint Conference on Artificial Intelligence. 2019:1907-1913.
- [20] Zhao Ling,Song Yujiao,Zhang Chao,et al. T-GCN: a temporal graph convolutional network for traffic prediction[J]. IEEE Trans on Intelligent Transportation Systems 2020,21(9):3848-3858.
- (上接第2973页)
- [22] Zhou Xinyi,Wu Jindi,Zafarani R. SAFE: similarity-aware multi-modal fake news detection[EB/OL].(2020-02-19).http://doi.org/10.48550/arxiv.2003.04981.
- [23] Xue Junxiao,Wang Yabo,Tian Yichen,et al. Detecting fake news by exploring the consistency of multimodal data[J]. Information Processing & Management 2021,58(5):102610.
- [24] Vaswani A,Shazeer N,Parmer N,et al. Attention is all you need[C]//Proc of the 31st International Conference on Neural Information Processing Systems. Red Hook,NY: Curran Associates Inc.,2017:6000-6010.
- [25] Devlin J,Chang M W,Lee K,et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL].(2019-05-24).http://doi.org/10.48550/arxiv.1810.04805.
- [26] Xu Bing,Wang Naiyan,Chen Tianqi,et al. Empirical evaluation of rectified activations in convolutional network[EB/OL].(2015-11-27).http://doi.org/10.48550/arxiv.1505.00853.
- [27] Bengio Y. Learning deep architectures for AI[J]. Foundations and Trends® in Machine Learning 2009,2(1):1-127.
- [28] Simonyan K,Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL].(2015-04-10).http://doi.org/10.48550/arxiv.1409.1556.
- [29] Ganin Y,Lempitsky V. Unsupervised domain adaptation by backpropagation[C]//Proc of the 32nd International Conference on Machine Learning. 2015:1180-1189.
- [30] Boudidou C,Andreadou K,Papadopoulos S,et al. Verifying multimedia use at mediaeval 2015[EB/OL].(2015-09-15).http://ceur-ws.org/Vol-1436/Paper4.pdf.
- [31] Slaney M,Casey M. Locality-sensitive hashing for finding nearest neighbors[J]. IEEE Signal Processing Magazine 2008,25(2):128-131.
- [32] Fortney K. Pre-processing in natural language machine learning[EB/OL].(2017-11-29).https://laptrinhx.com/pre-processing-in-natural-language-machine-learning-2728414203/.
- [33] Kingma D P,Ba J. Adam: a method for stochastic optimization[EB/OL].(2017-01-30).http://doi.org/10.48550/arxiv.1412.6980.
- [34] Agrawal A,Lu Jiasen,Antol S,et al. VQA: visual question answering[J]. International Journal of Computer Vision 2017,123(1):4-31.
- [35] Vinyals O,Toshev A,Bengio S,et al. Show and tell: a neural image caption generator[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2015:3156-3164.
- [36] Hinton G,Van der Maaten L J P. Visualizing data using t-SNE[J]. Journal of Machine Learning Research 2008,9(11):2579-2605.