

基于跨模态交互与特征融合网络的假新闻检测方法

彭广川, 吴飞, 韩璐, 季一木, 荆晓远

引用本文

彭广川, 吴飞, 韩璐, 季一木, 荆晓远. [基于跨模态交互与特征融合网络的假新闻检测方法](#)[J]. 计算机科学, 2024, 51(11): 23-29.

PENG Guangchuan, WU Fei, HAN Lu, JI Yimu, JING Xiaoyuan. [Fake News Detection Based on Cross-modal Interaction and Feature Fusion Network](#) [J]. Computer Science, 2024, 51(11): 23-29.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多模态自适应融合的短视频虚假新闻检测](#)

Multimodal Adaptive Fusion Based Detection of Fake News in Short Videos

计算机科学, 2024, 51(11): 39-46. <https://doi.org/10.11896/jsjcx.240700062>

[多源异构数据渐进式融合的虚假新闻检测](#)

Multi-source Heterogeneous Data Progressive Fusion for Fake News Detection

计算机科学, 2024, 51(11): 30-38. <https://doi.org/10.11896/jsjcx.240700004>

[基于协同网络与度量学习的标签噪声鲁棒联邦学习方法](#)

Collaborative Network and Metric Learning Based Label Noise Robust Federated Learning Method

计算机科学, 2024, 51(10): 391-398. <https://doi.org/10.11896/jsjcx.230900050>

[基于多尺度特征图卷积网络的教学行为识别及分析](#)

Recognition and Analysis of Teaching Behavior Based on Multi-scale GCN

计算机科学, 2024, 51(10): 135-143. <https://doi.org/10.11896/jsjcx.240400089>

[基于鲲鹏处理器的LU并行分解优化算法](#)

LU Parallel Decomposition Optimization Algorithm Based on Kunpeng Processor

计算机科学, 2024, 51(9): 51-58. <https://doi.org/10.11896/jsjcx.230900079>

基于跨模态交互与特征融合网络的假新闻检测方法

彭广川¹ 吴飞¹ 韩璐¹ 季一木² 荆晓远³

1 南京邮电大学自动化学院、人工智能学院 南京 210003

2 南京邮电大学计算机学院 南京 210003

3 武汉大学计算机学院 武汉 430072

(pgcykny@163.com)

摘要 近年来,假新闻的激增对人们的决策过程产生了不利影响。现有的假新闻检测方法大多强调对多模态信息(如文本和图像)的探索和利用。然而,如何为检测任务生成有鉴别性的特征并有效地聚合不同模态的特征以进行假新闻检测,仍然是一个开放性问题。文中提出了一种新颖的假新闻检测模型,即跨模态交互与特征融合网络(Cross-modal Interaction and Feature Fusion Network, CMIFFN)。为了生成有鉴别性的特征,所提方法设计了一个基于监督对比学习的特征学习模块,通过同时进行模态内和模态间的监督对比学习,来确保异类特征相似度更小,同类特征相似度更大。此外,为了挖掘更多有用的多模态信息,所提方法设计了多阶段跨模态交互模块,通过多阶段的跨模态交互,学习带有图结构信息的跨模态交互特征。所提方法引入基于一致性评估的注意力机制,通过学习多模态一致性权重,来有效聚合模态特定特征和跨模态交互特征。在两个基准数据集 Weibo 和 Twitter 上的实验表明,CMIFFN 明显优于现有的多模态假新闻检测方法。

关键词: 假新闻检测; 监督对比学习; 多阶段跨模态交互; 图卷积网络

中图分类号 TP391

Fake News Detection Based on Cross-modal Interaction and Feature Fusion Network

PENG Guangchuan¹, WU Fei¹, HAN Lu¹, JI Yimu² and JING Xiaoyuan³

1 College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

2 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

3 School of Computer Science, Wuhan University, Wuhan 430072, China

Abstract In recent years, the surge in fake news has adversely affected people's decision-making process. Many existing fake news detection methods emphasize the exploration and utilization of multimodal information, such as text and image. However, how to generate discriminative features for the detection task and effectively aggregate features of different modalities for fake news detection remains an open question. In this paper, we propose a novel fake news detection model, i. e., cross-modal interaction and feature fusion network (CMIFFN). To generate discriminant features, a supervised contrastive learning-based feature learning module is designed. By performing intra-modality and inter-modality supervised contrastive learning simultaneously, it ensures that the similarity of heterogeneous features is smaller and the similarity of similar features is greater. In addition, in order to mine more useful multi-modal information, this paper designs a multi-stage cross-modal interaction module to learn cross-modal interaction features with graph structure information. The method introduces consistency evaluation-based attention mechanism to effectively aggregate modality-specific features and cross-modal interaction features by learning multi-modal consistency weight. Experiments on two benchmark datasets Weibo and Twitter show that CMIFFN is significantly superior to the state-of-the-art multimodal fake news detection methods.

Keywords Fake news detection, Supervised contrastive learning, Multi-stage cross-modal interaction, Graph convolutional network

到稿日期:2023-12-27 返修日期:2024-04-26

基金项目:国家自然科学基金(62076139);之江实验室开放课题(2021KF0AB05);未来网络科研基金项目(FNSRFP-2021-YB-15);南京邮电大学 1311 人才计划

This work was supported by the National Natural Science Foundation of China (62076139), Open Research Project of Zhejiang Lab (2021KF0AB05), Future Network Scientific Research Fund Project (FNSRFP-2021-YB-15) and 1311 Talent Program of Nanjing University of Posts and Telecommunications.

通信作者:吴飞(wufei_8888@126.com)

1 引言

随着互联网和社交媒体的普及,一种广泛使用新媒体的社会形式逐渐出现,为人们提供了更快速的信息获取渠道、更迅速的消息传播路径和更广泛的社交领域。然而,在这一现象背后,越来越多的假新闻通过在线平台传播,严重削弱了读者的判断能力和对新闻真实性的信任。假新闻不仅浪费人们的时间和精力,还会给个人、团体甚至整个社会带来负面后果,甚至影响国家和地区的稳定。因此,假新闻检测逐渐成为一个备受关注的研究课题。

传统的假新闻检测方法使用单一模态数据进行检测。然而,随着当代社交媒体平台中模态多样性的增加,例如图像、视频、音频和其他格式,单模态方法无法满足现实需要,同时多模态假新闻检测在实际应用场景中的重要性呈指数级增长。

最近的研究致力于融合文本和图像特征,生成多模态表示以提升虚假新闻检测的性能^[1]。根据 Chen 等的研究^[2],联合使用跨模态融合特征和单模态特征更为有益。因此,通过自适应地聚合跨模态融合特征和单模态特征,可以使多模态分类效果更好。然而,如何在提高假新闻检测任务中特征鉴别性的同时,充分挖掘和利用有用的多模态信息并对模态特定特征和多模态特征实施有效聚合仍然是多模态假新闻检测所面临的挑战。

为解决上述问题,本文提出了一种基于跨模态交互与特征融合网络(Cross-modal Interaction and Feature Fusion Network, CMIFFN)的假新闻检测方法。首先,为了生成具有鉴别性的模态特定特征,本文设计了基于监督对比学习的特征学习模块。具体来说,基于监督对比学习的特征学习模块旨在确保异类的模态间和模态内特征具有更小的相似度,而同类的模态间和模态内特征具有更大的相似度。随后,将模态特定特征输入多阶段跨模态交互模块,以学习带有图结构信息的跨模态交互特征。最后,引入基于一致性评估的注意力机制,该机制评估图像和文本模态表示分布的 Kullback-Leibler(KL)散度,以学习多模态一致性权重,从而能够有效地聚合模态特定特征和跨模态交互特征。

本文的主要贡献如下:

- 1) 提出了一个用于假新闻检测的跨模态交互与特征融合网络 CMIFFN,旨在充分捕捉有用的跨模态和模态特定信息。
- 2) 设计了一个基于监督对比学习的特征学习模块,以增强特征的鉴别能力。此外,本文提出了一个多阶段的跨模态交互模块,以挖掘潜在的跨模态相关性。
- 3) 引入了基于一致性评估的跨模态聚合模块,以有效地整合跨模态交互特征和模态特定特征。
- 4) 在两个广泛使用的数据集 Twitter^[3] 和 Weibo^[4] 上进行实验,结果证明,CMIFFN 取得了最先进的效果。

2 相关工作

2.1 假新闻检测

2.1.1 单模态方法

现有的单模态假新闻检测方法主要依赖于单模态信息分析,如文本或图像。

在文本方面,Qian 等^[5] 构建了一种基于文本的方法,从文章中捕获语义信息,并提出了一种用户响应的生成模型,以辅助假新闻检测。Ajao 等^[6] 探讨了假新闻消息和情绪词之间的关系。Nan 等^[7] 采用了领域门控机制来聚合专家提取的多个表示。对于图像新闻,Qi 等^[8] 研究了空间域的图像特征和频域特征,以进行取证分析。

然而,单模态假新闻检测方法只使用了单个模态的信息,而现在的社交媒体平台拥有丰富的多模态信息,例如文字、图片、视频,多个模态之间相辅相成,单模态方法可能无法满足现实需要。因此,在对假新闻进行检测时,本文同时考虑了文本和图像两种模态。

2.1.2 多模态方法

当下已经出现几种基于跨模态判别模式的方法,以实现在假新闻检测中的卓越性能。Wang 等^[9] 提出了 EANN,利用额外的事件判别器来获得事件不变特征。Khattar 等^[1] 提出了 MVAE,该方法利用多模态变分自编码器来学习图像和文本模态的共享特征。Zhang 等^[10] 提出了 MKEMN,在融合对齐的图像和文本嵌入以检测假新闻时增加了知识级连接。Zhou 等^[11] 提出了 SAFE,它利用预训练模型来衡量图像和文本模态之间的相似性,以检测假新闻。Xue 等^[12] 提出了 MC-NN,该方法融合了图像和文本模态的跨模态相似度和特征。Wu 等^[13] 提出了 MCAN,它堆叠了多个共注意力层,以更好地融合图像和文本模态的特征,用于多模态假新闻检测。Chen 等^[2] 提出了 CAFE,它定义了跨模态模糊性,将其用于评估单模态特征分布之间的 KL 散度,并通过学习得到的模糊得分,在最终的聚合过程中自适应地调整单模态和跨模态融合特征的权重,从而进行更加准确的分类。Singhal 等^[14] 提出了 LIIMR,用于确定每种模态在假新闻检测中的重要性。

然而,大多数现有的多模态方法都集中在融合图像和文本模态的特征上。这些方法中特征的结构和鉴别性并未被充分利用。

2.2 对比学习

目前,对于假新闻检测,对比学习已取得巨大成功。Ma 等^[15] 利用课程学习在对比学习过程中自动选择负样本进行训练。Wang 等^[16] 在对比学习过程中利用辅助任务来最小化负样本的损失。Zhou 等^[17] 融合了由对比语言-图像预训练(Contrastive Language-Image Pre-training, CLIP)提取的单模态特征,并引入了模态注意力模块来衡量单模态特征和多模态特征两者的重要性,以聚合特征并进行分类。

这些方法利用对比学习技术来增强特征的鉴别能力,因为对比学习的目标是学习让相似样本距离更小且不相似样本更大的嵌入空间。而本文使用对比学习技术来从模态间和模态内角度联合增强假新闻检测任务中特征的鉴别性。

2.3 图表示学习

图神经网络(Graph Neural Networks, GNNs) 通过从邻域中收集信息来学习节点表示,即邻域传播/聚合。GNNs 已成功应用于许多机器学习和自然语言处理任务,最近也应用于假新闻检测任务。Xu 等^[18] 提出了一个基于图的模型,通过邻域传播挖掘分散的相关片段之间的长距离语义依赖,并通过执行图结构学习来减少信息的冗余。Wu 等^[19] 提出

了一种基于假新闻传播树构建传播图的方法,并使用门控神经网络为传播图中的每个节点生成表示。Yang等^[20]使用图对抗学习框架来对抗假新闻。Lu等^[21]结合了用户个人资料特征,以预测源推文是否为假,并通过突出显示可疑转发者的证据和他们所关注的词语来生成解释。

上述方法采用图表示技术来更好地保留结构信息,因此本文采用图表示技术来进行多阶段交互模块中的相似性刻画,从而挖掘特征中的结构信息。

3 CMIFFN 方法

给定包括文本和图像的新闻集合 $X = \{x_i\}_{i=1}^n = \{x_i^v, x_i^t\}_{i=1}^n$, 相应的真实标签集合可以表示为 $Y = \{y_i\}_{i=1}^n$ 。新闻集合分为训练集和测试集,本文使用训练集对模型进行训练,使训练后的模型能够对测试集进行真假判断,若结果为

1,则新闻为假;若结果为 0,则新闻为真。

本文提出了一个名为 CMIFFN 的新框架,旨在进行跨模态交互,捕捉潜在的跨模态相关性并有效地聚合多模态特征。该框架由 3 个模块组成,如图 1 所示。

1) 基于监督对比学习的特征学习模块,首先将单模态信息编码为嵌入,将原始的单模态嵌入映射到共享空间,并进一步通过模态间和模态内的监督对比损失增强特征的判别能力;

2) 多阶段跨模态交互模块,通过跨模态注意力融合模态特定特征,并利用图卷积神经网络(Graph Convolutional Network, GCN)进一步学习跨模态交互特征;

3) 跨模态聚合模块,通过基于图像和文本模态之间分布差异的注意力机制,有效聚合图像和文本模态特定特征以及跨模态交互特征。

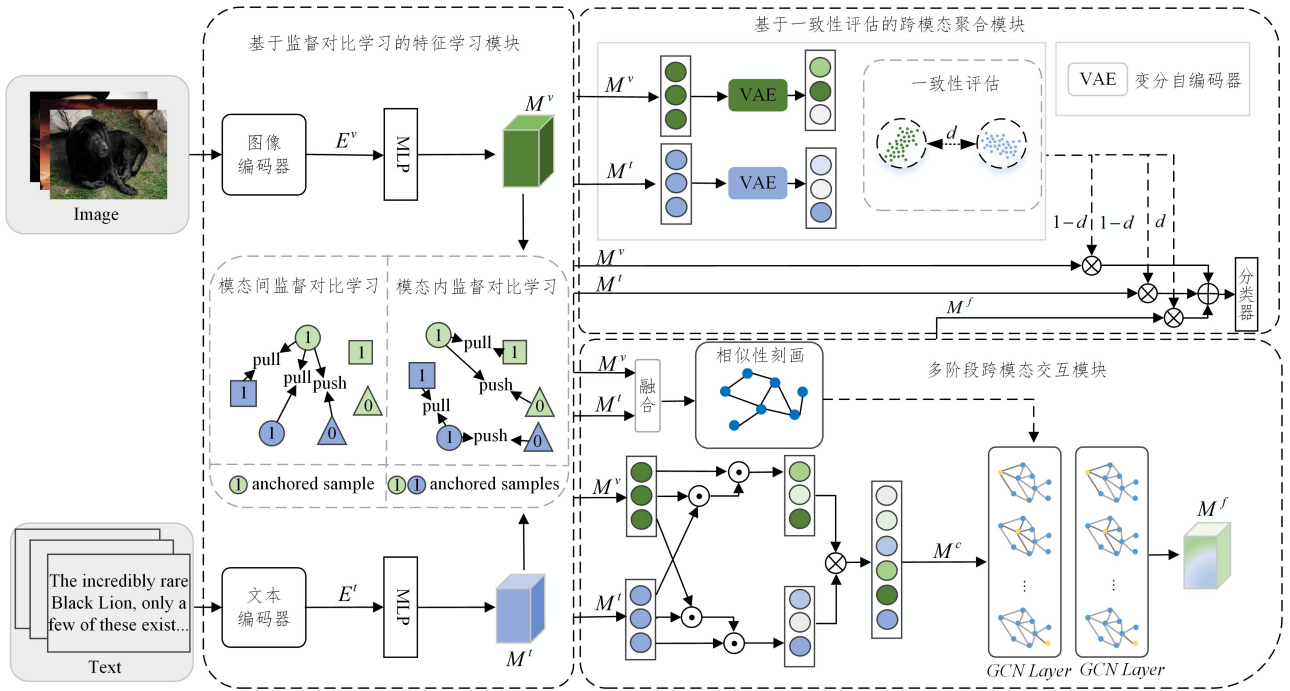


图 1 CMIFFN 的总体框架

Fig. 1 Overall framework of CMIFFN

3.1 基于监督对比学习的特征学习

首先,本文利用模态特定编码器提取原始的单模态特征。然后,设计了基于监督对比学习的特征学习模块,利用模态间和模态内的监督对比损失来增强模态特定特征的判别能力。

3.1.1 特征处理

本文选择了预训练的 BERT 模型作为文本编码器,而图像编码器则利用预训练的 ViT-B/32 模型。由模态特定编码器提取的特征可以表示为 $E^v = \{e_i^v\}_{i=1}^n$ 和 $E^t = \{e_i^t\}_{i=1}^n$ 。

将提取的特征 E^v 和 E^t 输入模态特定的多层感知机(Multilayer Perceptrons, MLPs)中,该 MLP 每层包含 64 个隐藏单元,以得到图像和文本模态的表示 $M^v = \{m_i^v\}_{i=1}^n$ 和 $M^t = \{m_i^t\}_{i=1}^n$ 。最后,本文应用跨模态交互模块,通过跨模态注意力机制和 GCN 来融合模态特定的特征 M^v 和 M^t ,得到联合表示。

3.1.2 模态内监督对比学习

为了增强模态特定特征表示的判别能力,本文设计了模态内监督对比损失 $L_{\text{intra-con}}$ 。

以图像模态为例,将第 i 个样本的图像特征 m_i^v 定义为锚点,图像模态的模态内监督对比损失 $L_{\text{intra-con}}^v$ 可以表示为:

$$L_{\text{intra-con}}^v = \frac{1}{N} \sum_{i \in I} \frac{1}{|P_i|} \sum_{p \in P_i} -\log \frac{\exp(m_i^v \cdot m_p^v / \tau)}{\sum_{k \in \kappa_i} \exp(m_i^v \cdot m_k^v / \tau)} \quad (1)$$

其中, I 是样本的索引集, $P_i = \{p \in \kappa_i; l_p = l_i\}$ 是正图像特征的索引集, $\kappa_i = I \setminus \{i\}$ 表示对比图像特征(索引集 I 除第 i 个样本之外,其他的样本特征)的索引集, $|P_i|$ 表示索引集 P_i 的大小,“ \cdot ”表示内积, τ 是温度系数。文本模态的模内监督对比损失也可以用同样的方式得到。

$$L_{\text{intra-con}}^t = \frac{1}{N} \sum_{i \in I} \frac{1}{|P_i|} \sum_{p \in P_i} -\log \frac{\exp(m_i^t \cdot m_p^t / \tau)}{\sum_{k \in \kappa_i} \exp(m_i^t \cdot m_k^t / \tau)} \quad (2)$$

其中, I 是样本的索引集, $P_i = \{p \in \kappa_i; l_p = l_i\}$ 是正文本特征的

索引集, $\kappa_i = \mathcal{I} \setminus \{i\}$ 表示对比文本特征的索引集, $|P_i|$ 表示索引集 P_i 的大小。

结合 $L_{\text{intra-con}}^v$ 和 $L_{\text{intra-con}}^t$, 总的模态内监督对比损失 L_{con} 可表示为:

$$L_{\text{intra-con}} = L_{\text{intra-con}}^v + L_{\text{intra-con}}^t \quad (3)$$

其中, I 是样本的索引集, $P_i = \{p \in \kappa_i; l_p = l_i\}$ 是正图像特征的索引集, $\kappa_i = \mathcal{I} \setminus \{i\}$ 表示对比图像特征的索引集, $|P_i|$ 表示索引集 P_i 的大小, “ \cdot ” 表示内积, τ 是温度系数。

通过最小化该损失, 将促使类内特征的模态内相似度更大, 而不同类的特征的模态内相似度更小。

3.1.3 模态间监督对比损失

除了提高特征的模态内鉴别性之外, 提高特征的模态间鉴别能力也很重要, 因此本文设计了模态间监督对比损失 $L_{\text{inter-con}}$ 。

将第 i 个样本的图像特征定义为锚点, 图像模态的模态间监督对比损失 $L_{\text{inter-con}}^v$ 可以表示为:

$$L_{\text{inter-con}}^v = \frac{1}{N} \sum_{i \in I} \frac{1}{|P_i|} \sum_{p \in P_i} -\log \frac{\exp(m_i^v \cdot m_p^v / \tau)}{\sum_{k \in \kappa_i} \exp(m_i^v \cdot m_k^v / \tau)} \quad (4)$$

其中, I 是样本的索引集, $P_i = \{p \in \kappa_i; l_p = l_i\}$ 是正图像特征的索引集, $\kappa_i = \mathcal{I} \setminus \{i\}$ 表示对比图像特征的索引集, $|P_i|$ 表示索引集 P_i 的大小。

同样地, 文本模态的模内监督对比损失可以表示为:

$$L_{\text{inter-con}}^t = \frac{1}{N} \sum_{i \in I} \frac{1}{|P_i|} \sum_{p \in P_i} -\log \frac{\exp(m_i^t \cdot m_p^t / \tau)}{\sum_{k \in \kappa_i} \exp(m_i^t \cdot m_k^t / \tau)} \quad (5)$$

其中, I 是样本的索引集, $P_i = \{p \in \kappa_i; l_p = l_i\}$ 是正文本特征的索引集, $\kappa_i = \mathcal{I} \setminus \{i\}$ 表示对比图像特征的索引集, $|P_i|$ 表示索引集 P_i 的大小。

结合 $L_{\text{inter-con}}^v$ 和 $L_{\text{inter-con}}^t$, 总的模态内监督对比损失 $L_{\text{inter-con}}$ 可表示为 $L_{\text{inter-con}}$:

$$L_{\text{inter-con}} = L_{\text{inter-con}}^v + L_{\text{inter-con}}^t \quad (6)$$

3.2 多阶段跨模态交互

为了实现不同模态之间的深层交互并利用潜在的邻域结构, 本文设计了多阶段的跨模态交互模块, 其中包括基于注意力机制的跨模态融合和相似性刻画组件。

3.2.1 基于注意力机制的跨模态融合

本文引入一种跨模态注意力机制来融合模态特定的特征, 以便执行图像和文本模态之间的交互。给定模态特定表示 M^v 和 M^t , 本文通过计算图像和文本模态表示之间的关联权重, 将其作为跨模态注意力权重, 这可以将模态特定信息嵌入到另一模态中。关联权重可以表示为:

$$W_{t2v} = \text{softmax}([M^t][M^v]^T / \sqrt{\text{dim}}) \quad (7)$$

$$W_{v2t} = \text{softmax}([M^v][M^t]^T / \sqrt{\text{dim}})$$

其中, dim 代表 M^v 和 M^t 的维度大小。基于模态间的注意力权重, 跨模态融合表示 M^c 定义如下:

$$M^c = (W_{v2t} \times M^v) \otimes (W_{t2v} \times M^t) \quad (8)$$

其中, \times 表示哈达玛积 (Hadamard), \otimes 表示外积。

3.2.2 相似性刻画

为了进一步利用跨模态相关性并挖掘多模态特征的潜在

结构信息, 本文利用模态特定特征 M^v 和 M^t 来构建相似度矩阵, 并将该矩阵作为图卷积操作的邻接矩阵。具体而言, 首先将模态特定特征 M^v 和 M^t 结合在一起, 多模态综合表示 M^c 定义为:

$$M^c = \{m_i^c\}_{i=1}^{2n} = [\gamma^v M^v, \gamma^t M^t] \quad (9)$$

其中, γ^v 和 γ^t 代表权重参数, 用于调节图像和文本模态的重要性。通过将 M^c 标准化以得到 \tilde{M}^c , 并计算 \tilde{M}^c 的转置与 \tilde{M}^c 的内积, 从而定义融合特征之间的相似性。

$$S_{ij} = (\tilde{m}_i^c)^T (\tilde{m}_j^c) \quad (10)$$

一般来说, S_{ij} 的值越大, m_i^c 和 m_j^c 就越相似。

然而, 由于该方法只考虑 M^c 的共同非零维度, 公式中的内积操作可能忽略了许多有用的信息。考虑到这些因素, 本文引入了一个辅助矩阵 C , 以利用 M^c 中 m_i^c 和 m_j^c 之间的维度差异。具体来说, C 的元素被定义为: $C_{ij} = \exp(-\sqrt{\|m_i^c - m_j^c\|_2 / \mu})$, 其中 μ 是一个缩放参数。在结合 C 和相似性矩阵 S 之后, 邻接矩阵 A 定义如下:

$$\begin{aligned} A &= \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} \\ &= S \times C \\ &= \begin{pmatrix} S_{11}C_{11} & \cdots & S_{1n}C_{1n} \\ \vdots & \ddots & \vdots \\ S_{n1}C_{n1} & \cdots & S_{nn}C_{nn} \end{pmatrix} \end{aligned} \quad (11)$$

然后, 基于注意力机制的跨模态融合模块的输出 M^c 和邻接矩阵 A 被送入 GCN, 以获得更多的结构信息。具体来说, 以 M^c 为 GCN 的输入, 图卷积过程定义如下:

$$H_{(l)} = \text{sigmoid}(D^{-1/2} A D^{-1/2} H_{(l-1)} W_l) \quad (12)$$

其中, $D_{ii} = \sum_j A_{ij}$, $H_{(l)}$ 和 $H_{(l-1)}$ 分别代表 GCN 层的输出和输入, W_l 是 l 层的卷积滤波器, $\text{sigmoid}(\cdot)$ 是第 l 层 GCN 层的激活函数。通过 GCN, 最终得到的跨模态交互特征被定义为 M^f 。

3.3 基于一致性评估的跨模态聚合

由于不同模态可能在分类过程中发挥不同的作用^[14], 本文评估模态特定特征和跨模态交互特征的权重以进行有效的聚合。具体而言, 本文首先从生成的角度对图像和文本模态的特征进行建模, 而特征是通过各向同性高斯先验分布 (Isotropic Gaussian Priors) 采样的。然后, 利用分布差异来表示跨模态信息差距。值得注意的是, 随着多模态一致性的增强, 跨模态信息对于假新闻检测的重要性逐渐降低。

首先通过两个变分自动编码器对输入样本的模态特定分布进行建模, 而 KL 散度被用作分布差异来获取多模态一致性。然后, 学习到的一致性得分被用来自适应调整跨模态交互特征和模态特定特征的贡献。当模态特定特征呈现较强的一致性时, 检测任务应更加关注跨模态交互特征, 反之亦然。

单模态观测的变分后验表示为:

$$q(z|m) = N(z|\mu(m), \sigma(m)) \quad (13)$$

其中, μ 和 σ 分别代表均值和方差, z 是来自输入为 m 的变分

自编码器输出的特征采样。因此,图像和文本模态的所有数据样本的分布可以表示为:

$$q(z_i^* | m_i^*) = N(z_i^* | \mu(m_i^*)), * \in \{v, t\} \quad (14)$$

而模态特定一致性得分 d_i^{v2t} 和 d_i^{t2v} 被定义为:

$$d_i^{v2t} = \left(\frac{D_{KL}(q(z_i^v | m_i^v) \| q(z_i^t | m_i^t))}{D_{KL}(q(z_i^v) \| q(z_i^t))} \right) \quad (15)$$

$$d_i^{t2v} = \left(\frac{D_{KL}(q(z_i^t | m_i^t) \| q(z_i^v | m_i^v))}{D_{KL}(q(z_i^t) \| q(z_i^v))} \right)$$

其中, $D_{KL}(\cdot \| \cdot)$ 代表 KL 散度, $q(z_i^v) = \frac{1}{n} \sum_{i=1}^n q(z_i^v | m_i^v)$, $q(z_i^t) = \frac{1}{n} \sum_{i=1}^n q(z_i^t | m_i^t)$ 。

因此,基于模态特定一致性得分 d_i^{v2t} 和 d_i^{t2v} ,多模态一致性得分可以定义为:

$$d = \{d_i\}_{i=1}^n = \left\{ \text{sigmoid} \left(\frac{1}{2} (d_i^{v2t} + d_i^{t2v}) \right) \right\}_{i=1}^n \quad (16)$$

有了模态特定特征 M^v 和 M^t ,以及跨模态交互特征 M' 和多模态一致性权重 d ,最终的表示如式(17)所示:

$$G = \{g_i\}_{i=1}^n = (d \times M^v) \oplus (d \times M^t) \oplus ((1-d) \times M') \quad (17)$$

其中, \oplus 代表拼接操作。

通过上述方式,可以很好地进行跨模态关联信息和模态特定信息之间的聚合。

然后,将表示 G 输入基于全连接网络的分类器中,以预测标签。

$$\tilde{y}_i = \text{softmax}(MLP(\tilde{x}_i)) \quad (18)$$

本文使用交叉熵损失作为最终的分类损失,它可以表示为:

$$L_{\text{cls}} = -E_{(x,y) \sim (X,Y)} [-y \log(\tilde{y}) + (1-y) \log(1-\tilde{y})] \quad (19)$$

3.4 总损失

本文将 CMIFFN 的总损失定义为:

$$L_{\text{total}} = \alpha L_{\text{intra-con}} + \beta L_{\text{inter-con}} + \eta L_{\text{cls}} \quad (20)$$

其中, α, β 和 η 是超参数。

4 实验

4.1 实验设置

4.1.1 数据集

1) Twitter 数据集由 MediaEval 的验证多媒体使用任务提供。在实验中,本文利用与基准测试相同的数据拆分

方案^[3]。具体而言,训练集包括 6840 条真实推文和 5007 条虚假推文,测试集包含 1406 个帖子。

2) Weibo 数据集^[4] 包括 3749 条假新闻和 3783 条真实新闻用于训练,1000 条虚假新闻和 996 条真实新闻用于测试。在实验中,本文按照之前的研究^[4,12] 中详细介绍的步骤进行操作,以去除重复和低质量的图像,确保数据集的整体质量。

4.1.2 基线

为了评估 CMIFFN 方法的有效性,本文将 CMIFFN 模型与 8 种基准模型进行比较。其中 EANN^[9] 在假新闻检测时引入了事件鉴别器;MVAE^[7] 利用多模态变分自编码器来学习图像和文本模态的共享特征;MKEMN^[10] 在融合对齐的图像和文本嵌入时增加了知识级连接;SAFE^[11] 利用预训练模型来衡量图像和文本模态之间的相似性以检测假新闻;MCNN^[12] 融合了图像和文本模态的跨模态相似度和特征;MCAN^[13] 堆叠了多个共注意力层来融合图像和文本模态的特征;LIIMR^[14] 评估每种模态在假新闻检测中的重要性;CAFE^[2] 学习模糊得分来优化最终的聚合过程。

4.1.3 实施细节

本文使用准确率 (Accuracy, Acc)、精确率 (Precision, P)、召回率 (Recall, R) 和 F1 分数 ($F1$) 来评估 CMIFFN 的性能。批大小设置为 128,并使用初始学习率为 0.001 的 Adam 优化器进行 100 个 epoch 的模型训练。在监督对比学习损失中, τ 被设定为 0.07。用于相似性刻画的权重参数 γ^v 和 γ^t 以及缩放参数 μ 分别设置为 1, 0.8 和 4。总损失 L_{total} 中 α, β 和 η 都被设置为 1。本文采用网格搜索策略来确定模型的最优参数设置, γ^v 和 γ^t 的搜索范围为 $[0.2, 0.4, 0.6, 0.8, 1]$, 而 α, β 和 η 的搜索范围为 $[0.001, 0.1, 1, 10, 100, 1000]$ 。所有代码都是基于 PyTorch 实现的,并在 NVIDIA RTX 1080Ti 上运行。

4.2 与基准方法的比较

表 1 列出了 CMIFFN 模型和 8 种比较方法在两个数据集上的检测性能,最佳结果用粗体标出。表 1 中,就 Acc 和 $F1$ 来说,CMIFFN 方法在每个数据集上的表现都优于其他方法。具体来说,CMIFFN 在两个真实世界数据集上分别实现了 90.4% 和 93.7% 的最高准确度。

表 1 CMIFFN 与其他方法在 Twitter 和 Weibo 数据集上的性能比较

Table 1 Performance comparison between CMIFFN and other methods on Twitter and Weibo datasets

数据集	方法	Acc	Fake News			Real News		
			P	R	F1	P	R	F1
Twitter	EANN	0.648	0.810	0.498	0.617	0.584	0.759	0.660
	MVAE	0.745	0.801	0.719	0.758	0.689	0.777	0.730
	MKEMN	0.715	0.814	0.756	0.708	0.634	0.774	0.660
	SAFE	0.762	0.831	0.724	0.774	0.695	0.811	0.748
	MCNN	0.784	0.778	0.781	0.779	0.790	0.787	0.788
	MCAN	0.809	0.889	0.765	0.822	0.732	0.871	0.795
	CAFE	0.806	0.807	0.799	0.803	0.805	0.813	0.809
	LIIMR	0.831	0.836	0.832	0.830	—	—	—
	CMIFFN	0.904	0.852	0.972	0.908	0.969	0.838	0.899

续表								
数据集	方法	Acc	Fake News			Real News		
			P	R	F1	P	R	F1
Weibo	EANN	0.827	0.847	0.812	0.829	0.807	0.843	0.825
	MVAE	0.824	0.854	0.769	0.809	0.802	0.875	0.837
	MKEMN	0.814	0.823	0.799	0.812	0.723	0.819	0.798
	SAFE	0.816	0.818	0.815	0.817	0.816	0.818	0.817
	MCNN	0.823	0.858	0.801	0.828	0.787	0.848	0.816
	MCAN	0.899	0.913	0.889	0.901	0.884	0.909	0.897
	CAFE	0.840	0.855	0.830	0.842	0.825	0.851	0.837
	LIIMR	0.900	0.882	0.823	0.847	—	—	—
	CMIFFN	0.937	0.931	0.937	0.934	0.942	0.937	0.939

CMIFFN 方法性能改进的原因有以下两点。1) 模态内监督对比学习和模态间监督对比学习的联合使用, 增强了模态特定特征的鉴别能力。此外, 通过跨模态语义交互和基础邻域结构挖掘, 增强了跨模态关联性, 让跨模态交互特征包含更多有用的多模态信息。2) 利用 KL 散度来评估不同模态特征之间的信息差距, 动态学习多模态一致性权重, 可以在假新闻检测聚合过程中自适应地调整跨模态交互特征和模态特定特征的贡献。

4.3 消融实验

本节讨论了 CMIFFN 的细节。将没有监督对比学习模块的 CMIFFN 版本称作 CMIFFN-con, 将没有基于一致性评估的跨模态聚合模块的 CMIFFN 版本称作 CMIFFN-agg, 将没有相似性刻画模块的 CMIFFN 版本称作 CMIFFN-sim。表 2 列出了这些 CMIFFN 变体在两个真实世界数据集上的结果。

由表 2 可知, 各个变体在两个数据集上的表现明显劣于完整的 CMIFFN 版本。这一现象表明, 监督对比学习和基于一致性评估的注意力机制对假新闻检测任务都是有益的, 因为监督对比学习的使用会让检测任务中的特征鉴别性增加, 让异类特征拥有更小的相似度, 让同类特征拥有更大的相似度; 而学习的一致性权重在聚合中的使用也会增强特征的鉴别性, 让模态特定特征和多模态特征更好地聚合。此外, CMIFFN-sim 的结果也劣于 CMIFFN, 这意味着基于跨模态相似性信息融合的 GCN 也有助于提升方法的性能, 因为 GCN 可以利用潜在的邻域结构, 增加有效信息的挖掘, 有助于跨模态特征学习。

表 2 CMIFFN 不同变体的消融研究

Table 2 Ablation study of different variants of CMIFFN

数据集	消融方法	Acc	P	R	F1
Twitter	CMIFFN	0.904	0.852	0.972	0.908
	CMIFFN-con	0.863	0.858	0.887	0.873
	CMIFFN-agg	0.881	0.850	0.918	0.883
	CMIFFN-gra	0.883	0.855	0.907	0.880
Weibo	CMIFFN	0.937	0.931	0.937	0.934
	CMIFFN-con	0.914	0.916	0.882	0.899
	CMIFFN-agg	0.919	0.921	0.912	0.916
	CMIFFN-gra	0.907	0.906	0.902	0.904

4.4 t-SNE 可视化分析

为了进一步分析 CMIFFN 对提升跨模态聚合特征鉴别能力的有效性, 本文使用了 t-SNE 进行可视化分析, 以展示 CMIFFN 在 Weibo 数据集上的特征判别能力。通过图 2 可以看到, 已辨识出的特征展现出了明显的可分离性, 表明它们

在模型的分类过程中可以清晰地被区分开来。这是因为 CMIFFN 中基于监督对比学习的特征学习模块可以让特征的鉴别性从模态间和模态内两个角度联合增强, 而基于一致性评估模块的使用优化了模态特定特征和多模态特征的聚合, 也会让聚合特征的鉴别性更强。

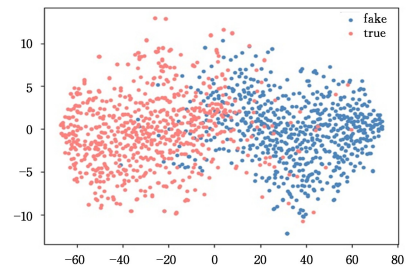


图 2 在 Weibo 数据集上对多模态聚合特征表示的可视化

Fig. 2 Visualization of multimodal aggregation feature representations on Weibo

结束语 本文提出了一种新颖的假新闻检测方法 CMIFFN, 该方法包括基于监督对比学习的特征学习模块、多阶段跨模态交互模块和跨模态聚合模块。对比学习模块能够有效地生成图像和文本模态的鉴别特征。多阶段跨模态交互模块能够消除模态差异, 并通过两种模态之间的深度交互来挖掘潜在的语义和结构信息。

跨模态聚合模块通过学习可评估多模态分布差异的多模态一致性权重, 在跨模态交互特征和模态特定特征之间进行了有效的聚合。实验结果表明, CMIFFN 的性能优于当前最先进的假新闻检测方法。实验结果还验证了所提方法中重要组件的有效性。

在未来工作中, 针对更加通用的假新闻检测场景, 本文将设计面向视频数据的假新闻检测模型。

参考文献

- [1] KHATTAR D, GOUD J S, GUPTA M, et al. Mvae: Multimodal variational autoencoder for fake news detection [C] // The World Wide Web Conference, 2019: 2915-2921.
- [2] CHEN Y, LI D, ZHANG P, et al. Cross-modal ambiguity learning for multimodal fake news detection [C] // The World Wide Web Conference, 2022: 2897-2905.
- [3] BOIDIDOU C, PAPADOPOULOS S, ZAMPOGLOU M, et al. Detection and visualization of misleading content on twitter [J]. International Journal of Multimedia Information Retrieval, 2018, 7(1): 71-86.
- [4] JIN Z, CAO J, GUO H, et al. Multimodal fusion with recurrent

- neural networks for rumor detection on microblogs [C]//International Conference on Multimedia. 2017;795-816.
- [5] QIAN F, GONG C, SHARMA K, et al. Neural user response generator; Fake news detection with collective user intelligence [C]//International Joint Conference on Artificial Intelligence. 2018;3834-3840.
- [6] AJAO O, GARG A, DA COSTA-ABREU M. Exploring content-based and meta-data analysis for detecting fake news infodemic: a case study on covid-19 [C]//International Conference on Pattern Recognition Systems. 2022;134-137.
- [7] NAN Q, CAO J, ZHU Y, et al. Mdfend; Multi-domain fake news detection [C]//ACM International Conference on Information & Knowledge Management. 2021;3343-3347.
- [8] QI P, CAO J, YANG T, et al. Exploiting multi-domain visual information for fake news detection [C]//IEEE International Conference on Data Mining. 2019;518-527.
- [9] WANG Y, MA F, JIN Z, et al. Eann; Event adversarial neural networks for multi-modal fake news detection [C]//ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018;849-857.
- [10] ZHANG H, FANG Q, QIAN S, et al. Multi-modal knowledge-aware event memory network for social media rumor detection [C]//ACM International Conference on Multimedia. 2019;1942-1951.
- [11] ZHOU X, WU J, ZAFARANI R. Safe; Similarity-aware multi-modal fake news detection [C]//Advances in Knowledge Discovery and Data Mining. 2020;354-367.
- [12] XUE J, WANG Y, TIAN Y, et al, WEI L. Detecting fake news by exploring the consistency of multimodal data [J]. Information Processing & Management, 2021, 58(5):102610.
- [13] WU Y, ZHAN P, ZHANG Y, et al. Multimodal fusion with co-attention networks for fake news detection [C]//Findings of the Association for Computational Linguistics. 2021;2560-2569.
- [14] SINGHAL S, PANDEY T, MREIG S, et al. Leveraging intra and inter modality relationship for multimodal fake news detection [C]//The World Wide Web Conference. 2022;726-734.
- [15] MA J, LIU Y, LIU M, et al. Curriculum contrastive learning for fake news detection [C]//ACM International Conference on Information & Knowledge Management. 2022;4309-4313.
- [16] WANG L, ZHANG C, XU H, et al. Cross-modal contrastive learning for multimodal fake news detection [J]. arXiv:2302.14057, 2023.
- [17] ZHOU Y, YING Q, QIAN Z, et al. Multimodal fake news detection via clip-guided learning [J]. arXiv:2022.2205.14304.
- [18] XU W, WU J, LIU Q, et al. Evidence-aware fake news detection with graph neural networks [C]//The World Wide Web Conference. 2022;2501-2510.
- [19] WU Z, PI D, CHEN J, et al. Rumor detection based on propagation graph neural network with attention mechanism [J]. Expert Systems with Applications, 2020, 158:113595.
- [20] YANG X, LYU Y, TIAN T, et al. Rumor detection on social media with graph structured adversarial learning [C]//International Joint Conferences on Artificial Intelligence. 2021;1417-1423.
- [21] LU Y, LI C. GCAN; Graph-aware co-attention networks for explainable fake news detection on social media [J]. arXiv:2004.11648, 2020.



PENG Guangchuan, born in 1999, post-graduate. His main research interests include fake news detection and cross-modal hashing.



WU Fei, born in 1989, Ph.D, professor. His main research interests include pattern recognition and machine learning.

(责任编辑:喻藜)