



计算机科学

COMPUTER SCIENCE

基于主动学习的多模态谣言检测模型

商云娴, 蔡国永, 刘庆华, 蒋艺明

引用本文

商云娴, 蔡国永, 刘庆华, 蒋艺明. [基于主动学习的多模态谣言检测模型](#)[J]. 计算机科学, 2025, 52(12): 391-399.

SHANG Yunxian, CAI Guoyong, LIU Qinghua, JIANG Yiming. [Active Learning-based Multi-modal Fusion Rumor Detection](#) [J]. Computer Science, 2025, 52(12): 391-399.

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)

[基于改进主动学习的入侵检测方法](#)

Intrusion Detection Method Based on Improved Active Learning

计算机科学, 2025, 52(10): 357-365. <https://doi.org/10.11896/jsjcx.240900142>

[一种新的基于凸损失函数的离散扩散文本生成模型](#)

Novel Discrete Diffusion Text Generation Model with Convex Loss Function

计算机科学, 2025, 52(10): 231-238. <https://doi.org/10.11896/jsjcx.240800147>

[基于分步协作融合表示的情感分类方法](#)

Sentiment Classification Method Based on Stepwise Cooperative Fusion Representation

计算机科学, 2025, 52(9): 313-319. <https://doi.org/10.11896/jsjcx.240700161>

[基于雷达和视觉融合的多模态空中手写体识别](#)

Multimodal Air-writing Gesture Recognition Based on Radar-Vision Fusion

计算机科学, 2025, 52(9): 259-268. <https://doi.org/10.11896/jsjcx.240400143>

[基于动态平衡和距离抑制的点云语义分割主动学习](#)

Active Learning for Point Cloud Semantic Segmentation Based on Dynamic Balance and Distance Suppression

计算机科学, 2025, 52(8): 180-187. <https://doi.org/10.11896/jsjcx.240900104>

基于主动学习的多模态谣言检测模型

商云娴 蔡国永 刘庆华 蒋艺明

桂林电子科技大学计算机与信息安全学院 广西 桂林 541004

广西可信软件重点实验室 广西 桂林 541004

(22032202027@mails.guet.edu.cn)

摘要 传统的谣言检测方法仍然有许多不足,如没有充分利用多模态信息,未考虑样本稀缺、标注昂贵、领域变化等实际情况,无法满足需求。为了解决样本稀缺和领域变化的问题,提出了一种新的基于主动学习的多模态谣言检测模型 ALMF。ALMF 设计了一种基于传播结构图增强的不确定性查询策略,使得主动学习筛选的样本更具有学习价值,同时减少了样本标注量;另外,ALMF 使用多模态数据,充分结合文本特征、图像特征以及传播结构特征,通过不同模态特征之间的互补增强,提升了谣言检测能力。在 PHEME 和 WEIBO 两个数据集上进行了实验,结果表明,ALMF 的性能优于对比模型,准确率提高 2%~9%。相比于基于基础查询策略的主动学习,ALMF 仅标注约 5% 的样本取得的性能与全样本时的性能相当。通过使用传播结构图增强的查询策略和跨模态增强的融合方式,ALMF 模型成功应对了新领域事件谣言检测面临的挑战。

关键词: 谣言检测;主动学习;多模态融合;协同注意力;传播图增强

中图分类号 TP391

Active Learning-based Multi-modal Fusion Rumor Detection

SHANG Yunxian, CAI Guoyong, LIU Qinghua and JIANG Yiming

College of Computer and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

Key Laboratory of Guangxi Trusted Software, Guilin, Guangxi 541004, China

Abstract Traditional rumor detection methods still have many shortcomings, such as insufficient utilization of multi-modal information, failure to consider sample scarcity, high labeling costs, and domain shifts. Therefore, it cannot meet the demands. To address the issues of sample scarcity and domain changes, this paper proposes a new Active Learning-based Multi-modal Fusion Rumor detection model, called ALMF. ALMF designs a novel uncertainty query strategy enhanced by the propagation structure graph, which ensures that the samples selected through active learning have greater learning value and reduces the demand for sample labeling. Meanwhile, ALMF employs multi-modal data, fully integrating text features, image features, and propagation structure features. The complementary enhancement between different modal features improves the capability of rumor detection. ALMF is tested on the PHEME and WEIBO datasets. The results show that ALMF outperforms the compared models, achieving an accuracy improvement of 2% to 9%. Compared to active learning based on basic query strategies, ALMF achieves performance that is nearly equivalent to that of full sample utilization with only approximately 5% of the samples labeled. By employing a query strategy enhanced with propagation structure graphs and cross-modal enhancement fusion methods, the ALMF model successfully addresses the challenges associated with rumor detection in new domain events.

Keywords Rumor detection, Active learning, Multi-modal fusion, Co-attention, Propagation graph enhancement

1 引言

随着移动互联网的迅速普及,微博、推特等社交媒体平台已经成为人们获取实时信息的主要途径。然而,网络的广泛使用意味着谣言等虚假信息也会被迅速地广泛传播,可能带来巨大的经济损失和社会影响。例如:2023年,关于“日本排

放核污水会导致中国食盐短缺”的谣言引发了一部分国民抢购食用盐,造成不必要的资源浪费和供应紧张。鉴于谣言在社交平台上的传播会扰乱社会的正常秩序,各种谣言检测的方法相继出现。

社会网络谣言被定义为一种未经验证或已被官方证实为假,并在社会网络中传播的信息^[1]。同时,谣言检测通常被定

到稿日期:2024-10-28 返修日期:2025-03-03

基金项目:国家自然科学基金(62366010);广西自然科学基金(2024GXNSFAA010374)

This work was supported by the National Natural Science Foundation of China (62366010) and Guangxi Natural Science Foundation (2024GXNSFAA010374).

通信作者:蔡国永(ccgycai@guet.edu.cn)

义为一个二分类问题^[2]。

一方面,早期的谣言检测方法利用人工从文本内容或传播结构中提取特征,再采用机器学习的模型学习文本的浅层特征,往往忽视了谣言样本可能包含的其他模态的信息^[3],如图像和语音。因此,一些研究开始关注多模态融合在谣言检测中的应用。Jin等^[4]首次尝试将图像特征用于虚假谣言检测,但该方法依靠人工设置的特征,导致需要复杂的特征工程。随着深度学习的发展,Jin等^[5]提出了一种具有注意力机制的递归神经网络,通过融合图像特征、文本特征和上下文信息实现多模态融合的谣言检测。然而,该方法仅关注发帖用户的信息,未考虑到帖子在网络传播过程中的信息,如评论和转发等互动信息。尽管上述方法在多模态融合的谣言检测上取得了一定进展,但仍未能充分融合文本、图像和传播信息。

另一方面,已有的谣言检测通常假定有充足的标签数据用于建模和训练,并且测试集数据与训练集数据有极强的相关性^[6],但是面对新领域(如新话题)时,社交媒体上的突发事件首先会经历零样本和小样本阶段,此时可供训练的有标签数据极其稀少,现有谣言检测模型将不再能满足需求。为了尽早有效遏制谣言传播带来的恶劣影响,需要对新领域谣言进行及时检测。已有一些工作对此进行了探讨。Ran等^[7]研究了无监督的跨领域谣言检测问题(即新领域中均为无标签样本),提出了基于实例和基于原型的对比学习来对齐特征表示以减少域间的差异,构建了具有相同标签的源数据和目标数据对之间的交叉注意机制来学习领域不变特征,但是该方法仅关注两个不同领域数据集之间的迁移学习,忽略了网络中多个不同领域谣言之间的相互影响。Lin等^[8]提出了一个基于提示学习的少样本谣言检测框架,将社交媒体上传播的谣言表示为不同的传播线程,从不同的传播线程中建模领域不变的结构特征来增强模型的领域自适应性,然而,该方法在新旧领域分布差异较大的场景下适应性较差。为了应对这一挑战,需要引入主动学习,Hasan等^[9]首次提出将主动学习应用于谣言检测任务,使用基于熵的查询策略来训练多模型神经网络集成体系结构,但是该工作未能针对谣言检测这一特定任务优化主动学习的查询策略,且未充分考虑多模态信息元素,因此仍有较大的改进空间。

为了弥补谣言检测在多模态融合和少样本阶段的不足,本文提出一个新的基于主动学习的多模态谣言检测模型ALMF(Active Learning-based Multi-modal Fusion Rumor Detection)。ALMF的主要特点有:1)首次提出了一种使用传播结构图增强的不确定性查询策略,来弥补传统主动学习的不足;2)结合了文本特征、图像特征以及传播结构特征,同时优化了跨模态特征增强及融合。

2 相关工作

根据谣言检测场景的不同,现有的谣言检测方法可分为基于机器学习和基于主动学习两类。其中,基于机器学习的方法主要针对有充足的标签数据用于建模的场景;基于主动学习的方法则主要针对少样本的新领域事件场景。

2.1 基于机器学习的谣言检测

最早的谣言检测方法源于2011年Castillo等^[10]对推特

中信息可信度的检测,该方法使用文本构造特征,采用支持向量机(SVM)对文本进行检测。Yang等^[11]在2012年提出了基于微博的谣言检测方法,利用微博中涉及的地理位置、发文客户端信息等特征,采用SVM构造微博谣言分类器模型。后人在此基础上展开了对推特和微博中谣言检测的研究。然而,这些方法主要针对单模态数据,且依赖特征工程,通常只能提取浅层特征。因此,越来越多的研究者开始转向研究多模态融合和深度学习方法,以期获得更为丰富和深层的特征表示。

针对多模态数据,Jin等^[4]根据视觉特征和统计特征来识别虚假信息。该方法首次尝试将图像特征用于谣言检测任务,但其仅依据人工设置特征,导致需要复杂的特征工程。随着深度学习在文本和图像中的应用越来越广泛,端到端的网络成为提取特征的主流方法。与早期手工设置特征的方法相比,端到端的网络能抽取到更深层次的特征。Yang等^[12]将文本和图像信息与相应的显式特征和潜在特征相结合来进行谣言检测。其中,显式特征是人工设置的特征,潜在特征则通过卷积神经网络(CNN)进行提取,该方法在假新闻检测中表现出良好的性能。Jin等^[5]将微博的多模态数据用于网络谣言检测,提出了一种具有注意力机制的递归神经网络来融合多模态特征。在该网络中,图像特征被结合到文本和上下文的联合特征中,以实现可靠的融合分类。但是上述工作忽略了帖子在网络传播过程中所产生的传播结构信息,未能同时结合文本、图像和传播结构3种信息。Yang等^[13]提出了基于图结构对抗学习的社交媒体谣言检测,构建了异构网络对传播信息进行建模,并提出了一个图对抗学习框架,以增强模型的鲁棒性。该方法将传播结构特征和文本特征进行拼接,最终实现了二分类的谣言检测。然而,该方法忽略了隐私保护或数据抓取限制等因素导致的传播信息缺失的问题,且它仅考虑文本信息和传播信息,并通过简单拼接进行融合,未能充分发挥多模态信息融合的优势。

ALMF同时结合了文本特征、图像特征和传播特征,优化了跨模态特征之间的特征增强和融合,以挖掘出数据更深层次的信息,从而达到更高的准确率。

2.2 基于主动学习的谣言检测

针对新领域少样本谣言的检测方法主要分为基于迁移学习的谣言检测和基于主动学习的谣言检测。

迁移学习被用于提升模型泛化能力,可以提高对新领域无标签谣言的检测能力。Wen等^[14]提出了一种基于跨语言跨平台的社交媒体谣言检测方法,在谣言检测中加入其他平台与该事件相关的信息,来提高检测结果的准确性。Zuo等^[15]使用软提示初始化策略来实现快速泛化,结合任务嵌入的即时生成器网络(TPHNet)完成谣言领域知识的积累,通过更新谣言领域嵌入和软提示库来实现分阶段的持续谣言检测。在迁移学习设置下,虽然新领域的样本不需要进行标注,但在遇到新、旧领域分布变化较大的情况时,很难实现良好的迁移性能。由此,从数据筛选角度进行模型迭代的主动学习算法逐渐被应用于谣言检测任务中。

主动学习是一种通过主动查询策略选取出最有价值的样本,并交给人工标注员进行标注的机器学习或人工智能方法,

目的是减少标注成本,只选择标注少量高质量样本,使模型尽可能接近全样本时的性能^[16]。主动学习的关键是采样查询策略,主要有以下几种^[16]:1)不确定性采样的查询,包括置信度最低、边缘采样、熵方法;2)基于委员会的查询;3)基于模型变化期望的查询;4)基于误差减少的查询;5)基于方差减少的查询;6)基于密度权重的查询。Hasan等^[9]首次提出将主动学习运用于谣言检测任务,提出了一种用于假新闻检测的半监督学习方法,使用基于熵的查询策略来训练多模型神经集成体系结构,其在多个谣言检测数据集下证实了仅需总样本量的4%~28%即可实现接近于全样本下的性能。Sahan等^[17]专注于贝叶斯主动学习方法,研究了在推特数据集下,不同查询策略和不同预训练词嵌入表达对主动学习性能的影响,分析了查询策略和词嵌入表达对模型收敛速度和最终性能的影响。Farinneya等^[18]探索了不同的分类器和预训练词嵌入表达对主动学习的影响。Shao等^[19]提出了一个随领域迭代的谣言检测框架 ACP-RD,该框架使用基本查询策略中的 Margin Sampling 策略,并随着谣言领域的转变实现语言模型及其软提示的动态迭代。尽管上述方法通过主动学习实现了在少样本阶段进行谣言检测,但仅依赖基础的主动学习查询策略难以有效筛选出对特定下游任务(如谣言检测)最具标注价值的样本。为了更精准地筛选出关键样本,必须分析帖子在谣言检测中的重要性。Cui等^[20]提出的基于自适应图对比学习的谣言检测模型 RAGCL 可以很好地解决这个问题。该模型强调了具有密集回复的帖子的重要性,同时考虑到谣言传播树的广泛结构,通过聚合其他节点信息至根节点,实现根节点分析聚焦。

受 Cui等^[20]的启发,本文针对谣言检测这一下游任务,提出了 ALMF 模型。该模型设计了一个结合不确定性查询策略和传播结构特征的主动学习采样策略,弥补了主动学习传统采样策略在谣言检测方面的不足。ALMF 基于 ALiPy 框架^[21],采用 ALiPy 中基于不确定性查询(Query_Instance_Uncertainty)策略,并在此基础上,针对谣言检测这一下游任务,对基础查询策略进行了改进。

3 问题描述

假设社交媒体上的帖子表示为 $P = \{p_1, p_2, \dots, p_n\}$ 。对于每个帖子 $p_i \in P$, $p_i = \{t_i, I_i, u_i, c_i\}$,其中 t_i, I_i, u_i 分别表示该帖子的文本、图像和用户信息, $c_i = \{c_i^1, c_i^2, \dots, c_i^j\}$ 表示 p_i 的评论集,每条评论都由相应的用户 u_i^j 发布。为了表示帖子在社交媒体上的传播信息,构建一个图 $G = (V, A)$,其中 V 是节点集,包括用户节点、评论节点和帖子节点。 $A \in \{0, 1\}^{|V| \times |V|}$ 是节点之间的邻接矩阵,用于描述两个节点之间是否存在发帖、评论和转发关系。

将谣言检测定义为一个二分类任务。 $y \in \{0, 1\}$ 表示标签,其中 $y=1$ 表示谣言,否则 $y=0$ 。目标是学习函数 $F(p_i) = y$ 来预测给定帖子 p_i 的标签。

4 ALMF 模型

ALMF 模型的整体框架如图 1 所示,主要包括特征提取模块、样本筛选模块 ALG、多模态融合检测模块 MF。在每轮

主动学习中,ALMF 模型将无标签数据集随机划分成训练集 Train 和测试集 Test。通过特征提取模块对训练集 Train 的样本进行特征提取后,使用样本筛选模块 ALG 筛选样本,并将选中样本的初始数据交由专家进行标注,标注完成后将样本初始数据更新到有标签数据池中,与前 $t-1$ 个领域的有标签数据集 L_{t-1} 合并成筛选标注后的数据集。随后通过特征提取模块对有标签数据集进行特征提取,再进行模型训练,得到训练后的 MF' 。通过特征提取模块对 Test 数据进行特征提取后,使用训练后的 MF' 进行检测,得到模型准确率 (Accuracy)。

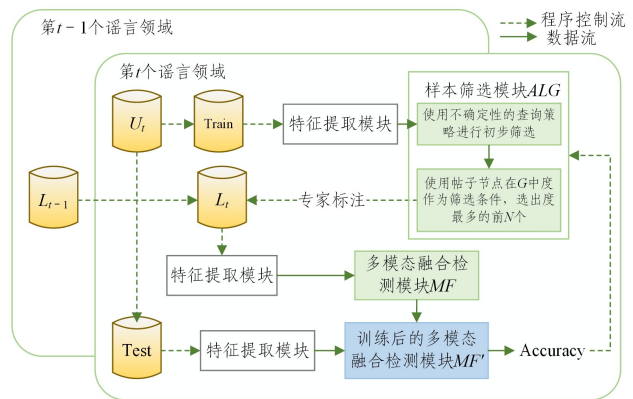


图 1 ALMF 整体结构示意图

Fig. 1 Schematic diagram of ALMF overall structure

4.1 特征提取

特征提取模块用于提取文本、图像、传播结构 3 类特征。其中,采用 CNN 网络提取文本特征;采用在 ImageNet 数据库上预训练的模型 ResNet50^[22] 提取图像特征;采用多头注意力机制从增强后的传播结构图中提取传播结构特征。

4.1.1 文本与视觉图像特征提取

对于文本,使用 CNN 提取文本特征。首先,设词嵌入矩阵 $\mathbf{o}_{1:L}^i = [o_{i1}^1, o_{i1}^2, \dots, o_{i1}^L]^T$,并在词嵌入矩阵 $\mathbf{o}_{j,j+k-1}^i$ 上使用卷积层得到特征图 S_{ij} ,记 $\mathbf{S}^i = [S_{i1}^1, S_{i1}^2, \dots, S_{i(L-k+1)}^1]^T$,其中 $\mathbf{o}_{j,j+k-1}^i \in \mathbf{R}^d$ 表示 t_i 的第 j 个 token, d 为词嵌入向量的维度, L 为 token 的数量, k 为感受野 (Receptive Field) 的大小。然后,经最大池化处理,初步得到文本特征 $\hat{\mathbf{S}}^i = \max(\mathbf{S}^i)$ 。随后,分别使用具有不同感受野 ($K \in \{3, 4, 5\}$) 的滤波器获取多种粒度的文本特征 $\hat{\mathbf{S}}_{K}^i$ 。最终将所有滤波器的输出进行拼接,得到文本特征 $\mathbf{R}_T^i = \text{concat}(\hat{\mathbf{S}}_{K=3}^i, \hat{\mathbf{S}}_{K=4}^i, \hat{\mathbf{S}}_{K=5}^i)$ 。

对于图像,首先提取 ResNet50 最后两层的输出,并将其记为 I_i^j 。然后,通过一个全连接层的处理,得到与文本特征维度相同的图像特征,记为 $\mathbf{R}_I^i = \sigma(\mathbf{W}_I * I_i^j)$,其中 \mathbf{W}_I 为全连接层的参数矩阵, σ 为激活函数 sigmoid。

4.1.2 传播结构图的特征提取

用户在社交媒体上的发帖、评论和转发行为,共同构成帖子在社交媒体上的传播信息。

根据帖子在社交媒体上的传播信息,首先构建传播结构图 $G = (V, A)$ 来表示用户在社交媒体上的发帖、评论、转发等行为,其中 $V = P \cup U \cup C$ 为节点集, P, U, C 依次表示帖子节点、用户节点、评论节点; $A \in \{0, 1\}^{|V| \times |V|}$ 为初始邻接矩阵, $a_{ij} = 1$

表示节点 n_i 和节点 n_j 之间存在发帖或评论等关系。节点的嵌入矩阵记为 $\mathbf{X} = \{x_1, x_2, \dots, x_{|V|}\} \in R^{|V| \times d}$ (d 是维度大小)。其中帖子节点 P 和评论节点 C 的初始嵌入为对应文本的句子向量; 用户节点 U 的初始嵌入为用户发布的帖子节点初始嵌入的平均值。

由于隐私保护或数据抓取的限制, 传播结构图中可能会存在一些边缺失的问题^[23]。根据网络同质性, 相似的节点之间存在边的概率要远高于不相似节点之间。因此通过计算节点之间的特征相似性, 可以推断节点之间是否存在隐藏边。通过在初始传播结构图中添加隐藏边, 得到增强的传播结构图。这些隐藏边的建模, 将有利于更好地抽取帖子传播结构的特征, 并用于谣言检测。图 2 展示了传播结构的增强过程, 左边虚线框内为两个用户 U_1 和 U_2 的初始传播结构示意图, 右边虚线框内为用户 U_1 和 U_2 的增强后的传播结构示意图, 图中的 P, U, C 依次表示帖子节点、用户节点、评论节点。

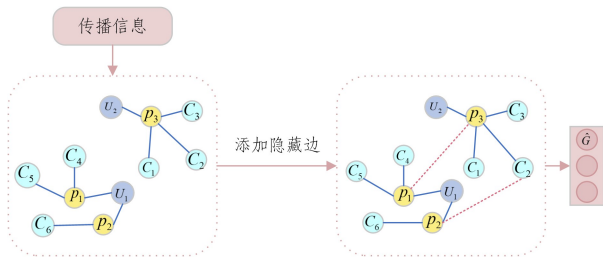


图 2 传播结构的增强过程

Fig. 2 Enhancement process of communication structure

为了推断节点 n_i 和节点 n_j 之间是否存在隐藏边, 首先计算两节点之间的余弦相似度 β_{ij} :

$$\beta_{ij} = \frac{x_i \cdot x_j}{\|x_i\| \cdot \|x_j\|} \quad (1)$$

其中, x_i 和 x_j 分别表示节点 n_i 和节点 n_j 的节点嵌入。如果相似度 β_{ij} 超过 0.5, 则认为两节点之间存在隐藏边 e_{ij} , 其计算式如式(2)所示:

$$e_{ij} = \begin{cases} 0, & \text{if } \beta_{ij} < 0.5 \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

设初始边矩阵为 $\mathbf{A} \in R^{|V| \times |V|}$, 添加隐藏边后的边矩阵 \mathbf{A}' 的元素 a'_{ij} 可表示为:

$$a'_{ij} = \begin{cases} 1, & \text{if } e_{ij} = 1 \text{ or } a_{ij} = 1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

其中, $a_{ij} \in \mathbf{A}$, $a_{ij} = 1$ 表示节点 n_i 和节点 n_j 之间存在初始边。换言之, 当两节点之间存在初始边或者隐藏边时, 即 $a_{ij} = 1$ 或 $e_{ij} = 1$ 时, 则增强后的边矩阵 \mathbf{A}' 中, $a'_{ij} = 1$ 表示两个节点间存在联系。

对于增强后的矩阵, 需要进行节点表示更新。设更新后的节点嵌入矩阵为 $\hat{\mathbf{X}} \in R^{|V| \times d}$, 其中 $|V|$ 是节点的数量, d 是维度大小。嵌入矩阵 $\hat{\mathbf{X}}$ 的元素 \hat{x}_i 表示节点 n_i 的特征向量, $\hat{\mathbf{x}}_i$ 可通过式(4)计算。

$$\hat{\mathbf{x}}_i = \sigma(\mathbf{W}_n * (\mathbf{A}' * \mathbf{X}_i)) \quad (4)$$

其中: \mathbf{W}_n 是全连接层的参数矩阵; $\sigma(\cdot)$ 是激活函数; \mathbf{X}_i 是节点 n_i 的邻居节点集 N_i 的特征矩阵, $N_i = \{n_1', n_2', \dots, n_{|N_i|}'\}$ 。

最后利用多头注意力机制捕捉传播结构特征, 将每个注

意力机制的输出拼接在一起作为增强后的图特征。

$$\hat{\mathbf{G}} = \parallel_{h=1}^H \sigma_h(\hat{\mathbf{X}}) \quad (5)$$

其中, H 表示多头注意力机制中 head 的数量。另外, 在增强后的图特征 $\hat{\mathbf{G}}$ 中, 第 i 个帖子 p_i 的图特征 \mathbf{R}_i^g 对应 $\hat{\mathbf{G}}$ 的第 i 列。

4.2 传播结构图增强的样本筛选

正如文献[20]所指出, 被大量回复的帖子在谣言检测中具有重要意义。因此, 传播结构图增强的样本筛选模块 ALG 采用基于传播结构图的不确定性查询策略, 选取传播结构图 G 中不确定性较强且受关注度较高的帖子, 用于训练模型。假设第 t 个领域为新领域, 它的无标签数据集 U_t 由三元组 (T, I, G) 组成, 前 $t-1$ 个谣言领域的有标签数据集 L_{t-1} 由四元组 (T, I, G, y) 组成。其中 T 是文本, I 是图像, G 是传播信息, y 表示标签。

对训练集 Train 的样本 (T, I, G) , 参照 4.1.1 节, 使用 CNN 和 ResNet50 网络分别进行文本和图像的特征提取; 参照 4.1.2 节, 通过计算余弦相似度并设置阈值来增加隐藏边, 从而构建增强后的传播结构图 $\hat{\mathbf{G}}$, 并利用多头注意力机制进行特征提取。

如果训练集 Train 中样本量大于等于 $2N$, 则利用帖子 p_i 的文本特征 R_T^i , 通过使用不确定性查询策略进行初步筛选。不确定查询策略首先计算每个样本对两个标签(谣言和非谣言)的预测概率之间的绝对差值, 然后选取绝对差值最小的 $2N$ 个样本作为初始筛选样本集 $SelData_{Initial}$; 如果 Train 中样本数少于 $2N$, 则全部选用。不确定性查询策略可形式化为:

$$SelectSamples_{Initial} = \begin{cases} Top_2N_{p_i}(\{|P_0 - P_1|\}), & |Train| \geq 2N \\ Train, & N \leq |Train| < 2N \end{cases} \quad (6)$$

其中, $N = |Train| \times a\%$ 为要筛选样本集 $SelData_{Final}$ 的大小, $a\%$ 表示 N 在 Train 中的占比; $Top_2N_{p_i}(\cdot)$ 函数的输入为样本的二分类预测概率, 输出预测概率差值最小的前 $2N$ 个样本的索引, 表示不确定性最大的样本, \hat{y}_1 代表谣言, \hat{y}_0 代表非谣言; $P_0 = P(\hat{y} = \hat{y}_0 | R_T^i)$ 表示使用文本特征 R_T^i 预测帖子 p_i 为非谣言的概率, 同理 $P_1 = P(\hat{y} = \hat{y}_1 | R_T^i)$ 表示使用文本特征 R_T^i 预测帖子 p_i 为谣言的概率; $|P_0 - P_1|$ 表示帖子的两个标签预测概率的绝对差, 即差值的绝对值; $\{|P_0 - P_1|\}$ 表示所有帖子的预测概率绝对差的集合; $|Train|$ 表示训练集中样本的数量。

计算 $SelData_{Initial}$ 中每个样本在 $\hat{\mathbf{G}}$ 中的度, 记为 $Degree_i = \sum_{j=1}^{|V|} A'_{ij}$ 。根据 $Degree_i$ 将初始筛选样本集中的样本进行降序排序, 筛选出前 N 个样本作为筛选样本集 $SelData_{Final}$ 。该筛选过程可形式化为:

$$SelData_{Final} = Top_N_{Degree}(SelData_{Initial}) \quad (7)$$

其中, $Top_N_{Degree}(\cdot)$ 函数计算返回降序排序后的前 N 个样本的索引, 以此取出度最高的前 N 个样本, 即不确定性较强且传播范围较广的样本。

如果 $SelData_{Initial}$ 中样本量小于 N , 则筛选样本集

$SelData_{Final} = SelData_{Initial}$

最后,由专家对 $SelData_{Final}$ 样本进行标注后,将其更新到 L_t 中并从 Train 中删除已标注的数据。筛选样本模块 ALG 的伪代码如算法 1 所示,因每轮选取的筛选样本集大小为 $N = |Train| \times a\%$,故主动学习的筛选的最大轮数为 $MAX_Q = \lfloor \frac{1}{a\%} \rfloor$ 。

算法 1 筛选样本伪代码

输入:训练集 Train,筛选总轮数 MAX_Q,筛选样本集的大小 N,前

$t-1$ 个有标签数据集 L_{t-1}

输出:有标签数据集 L_t

1. num_Q=1 //查询轮数初始化为 1
2. $L_t = L_{t-1}$ //初始化 L_t
3. while num_Q \leq MAX_Q:
4. if |Train| \geq 2N //训练数据个数大于等于 2N
5. $SelData_{Initial} = Top_2N_{pi}(\{|P_0 - P_1|\})$ // $Top_2N_{pi}()$ 的输入为 Train 中每个样本的二分类预测概率,输出为其差值最小的前 2N 个样本索引
6. 参照 4.1.2 节,为 $SelData_{Initial}$ 中的每个样本构造传播结构图 \hat{G}
7. 计算 $SelData_{Initial}$ 中的每个样本在 \hat{G} 中的度,记为 Degree_i
8. 将初始筛选样本集 $SelData_{Initial}$ 中的样本依据 Degree,降序排序
9. $SelData_{Final} = Top_N_{Degree}(SelData_{Initial})$ // $Top_N_{Degree}()$ 的输入是传播图,输出是度最高的前 N 个样本的索引
10. else if $N \leq |Train| < 2N$ //被筛选的训练数据个数大于等于 N,小于 2N
11. $SelData_{Initial} = Train$ //将初始筛选集设置为训练集
12. 参照 4.1.2 节,为 $SelData_{Initial}$ 中的每个样本构造传播结构图 \hat{G}
13. 计算 $SelData_{Initial}$ 中的每个样本在 \hat{G} 中的度,记为 Degree_i
14. 将初始筛选样本集中样本依据 Degree,降序排序
15. $SelData_{Final} = Top_N_{Degree}(SelData_{Initial})$ // $Top_N_{Degree}()$ 的输入是传播图,输出是度最高的前 N 个样本的索引
16. else if |Train| < N //被筛选的训练数据个数小于 N

17. $SelData_{Final} = Train$ //将筛选集设置为训练集

18. End if

19. 专家对 $SelData_{Final}$ 样本的初始数据进行标注

20. $L_t = L_t + SelData_{Final}$ //将已标注的数据 $SelData_{Final}$ 更新到 L_t 中

21. $MF' = MF(L_t)$ //使用 L_t 训练多模态融合检测模块 MF

22. Train = Train - $SelData_{Final}$ //从 Train 中删除已标注的数据 $SelData_{Final}$

23. num_Q = num_Q + 1

24. end

4.3 多模态融合检测模块

多模态融合检测模块 MF 的结构如图 3 所示。首先, MF 使用多层协同注意力机制实现多模态特征增强,使用自监督损失实现强制跨模态对齐。将增强后的多模态特征融合后输入全连接层,得到帖子 p_i 为谣言的概率 P_i ,随后通过二分类器可以得到帖子 p_i 的预测标签 \hat{y}_i 。

4.3.1 多模态特征融合

为更好地实现跨模态融合, MF 使用多层协同注意力机制来捕获不同模态间的交互信息,通过学习不同模态特征间的注意权重来实现跨模态特征的增强,并使用 Sigmoid 函数和 LeakyReLU 函数得到最终多模态融合的特征。

具体地,在图 3 中,以文本模态和图像模态为例,详细展示了使用协同注意力来实现跨模态特征的增强。其中:文本特征为 R_T ,图像特征为 R_I 。对于模态对 (R_T, R_I) ,使用迭代的协同注意力机制计算图像特征 R_I 增强的文本特征 R_{IT} ,一共迭代 l 次。初始轮次中, Q, K, V 矩阵分别为 $Q_i^l = R_I^l W_i^Q, K_i^l = R_I^l W_i^K, V_i^l = R_I^l W_i^V$,其中, W_i^Q, W_i^K, W_i^V 是线性变换的矩阵。第 j 层注意力机制的输出可表示为:

$$O_j = \left(\prod_{h=1}^H \text{softmax} \left(\frac{Q_i^h K_i^h}{\sqrt{d}} \right) V_i^h \right) W_i^O \quad (8)$$

其中, H 是 head 的数量, W_i^O 是输出层的线性变换的矩阵。第 $j(j > 1)$ 层的 Q_i^j 使用前一层的输出 O_{j-1} 迭代更新。将最后一层协同注意力机制的输出作为增强后的文本特征 $R_{IT} = O_l$ 。

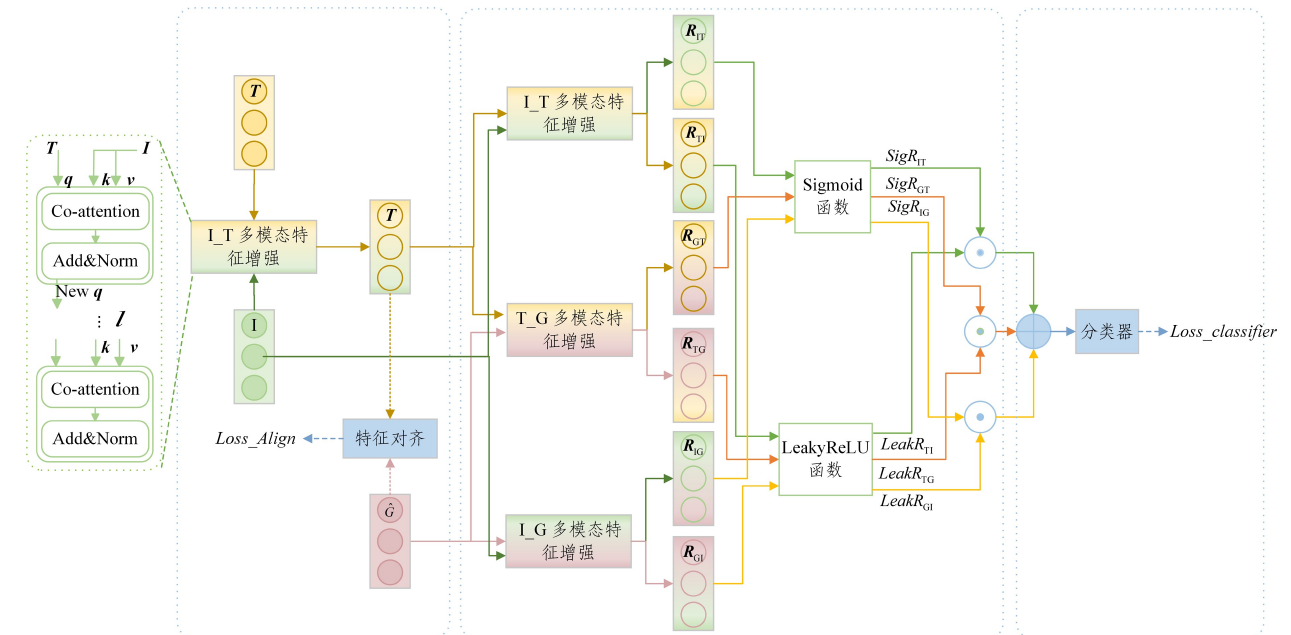


图 3 多模态融合检测模块结构

Fig. 3 Structure of multi-mode fusion detection module

对于模态对 $(\mathbf{R}_T^i, \mathbf{R}_I^i)$, 同样也可以得到 \mathbf{R}_T^i 增强的图像特征 \mathbf{R}_{TI}^i 。同理, 对于其他模态对, 通过上述迭代协同注意力机制的处理, 可以求出 $\mathbf{R}_{TG}^i, \mathbf{R}_{GT}^i, \mathbf{R}_{IT}^i, \mathbf{R}_{TI}^i$ 。

将 $\mathbf{R}_{GT}^i, \mathbf{R}_{IT}^i, \mathbf{R}_{TI}^i$ 分别输入 Sigmoid(), 得到输出 $SigR_{GT}^i, SigR_{IT}^i, SigR_{IG}^i$; 将 $\mathbf{R}_{TG}^i, \mathbf{R}_{TI}^i, \mathbf{R}_{TI}^i$ 分别输入 LeakReLU() 中, 得到输出 $LeakR_{TG}^i, LeakR_{TI}^i, LeakR_{GI}^i$ 。

将相同组合模态对对应的输出相乘后拼接在一起, 作为最终融合后的特征 \mathbf{R}_{final}^i , 该过程可形式化表示为:

$$\mathbf{R}_{final}^i = SigR_{GT}^i \times LeakR_{TG}^i + SigR_{IT}^i \times LeakR_{TI}^i + SigR_{IG}^i \times LeakR_{GI}^i \quad (9)$$

4.3.2 特征对齐

通过使用自监督损失, 确保得到与文本和图像模态一致的传播结构特征, 从而实现跨模态特征对齐。在特征提取和文本-图像特征增强两个关键步骤后, 文本特征和图像特征可成功映射到相同维度的向量空间中。因此, 无论选择文本特征还是图像特征, 都能够有效地实现跨模态对齐。以使用文本特征为例, 具体来说, 对于第 i 个样本 p_i , 首先将 \mathbf{R}_{IT}^i 和 \mathbf{R}_g^i 映射到一个特征空间, 得到 \mathbf{R}_g^i 和 \mathbf{R}_{IT}^i , 该过程可形式化为:

$$\begin{aligned} \mathbf{R}_g^i &= \mathbf{W}_g' \cdot \mathbf{R}_g^i \\ \mathbf{R}_{IT}^i &= \mathbf{W}_T' \cdot \mathbf{R}_{IT}^i \end{aligned} \quad (10)$$

其中, \mathbf{W}_g' 和 \mathbf{W}_T' 是参数矩阵。随后使用均方差误差来表示 \mathbf{R}_{IT}^i 和 \mathbf{R}_g^i 之间的距离 $Loss_Align$, 该过程可形式化为:

$$Loss_Align = \frac{1}{n} \sum_{i=1}^n (\mathbf{R}_g^i - \mathbf{R}_{IT}^i)^2 \quad (11)$$

其中, n 为样本个数, 通过缩小3.3.3节的总损失 $Loss$ 来缩短 \mathbf{R}_{IT}^i 和 \mathbf{R}_g^i 之间的距离。

4.3.3 分类器

将帖子最终融合后的特征 \mathbf{R}_{final}^i 传入全连接层, 可以得到帖子 p_i 为谣言的概率 P_i :

$$P_i = \text{softmax}(\mathbf{W}_c \mathbf{R}_{final}^i + b) \quad (12)$$

如果 $P_i \geq 0.5$, 则认为该帖子 p_i 为谣言, 即预测标签 $\hat{y}_i = 1$ 。然后使用交叉熵损失函数计算分类损失 $Loss_Classifier$, 可表示为:

$$Loss_Classifier = -y \log(\hat{y}_i) - (1-y) \log(1-\hat{y}_i) \quad (13)$$

总损失 $Loss$ 如式(14)所示:

$$Loss = \alpha Loss_Align + \beta Loss_Classifier \quad (14)$$

其中, α 和 β 是用于平衡两个损失的参数, $\alpha + \beta = 1$ 。

5 实验

5.1 数据集

在两个真实数据集微博^[24]和PHEME^[25]上评估了提出的模型。微博数据集来自中国社交媒体平台微博, 使用清华大学THUCTC^[26]提供的预训练模型, 对数据集进行分类, 并选择“娱乐”类作为新出现的领域进行主动学习, 其他类别作为初始有标注的数据集。PHEME是由推特平台上的推文组成的, 以5个领域的突发新闻为基础。PHEME数据集使用文本的话题标签进行分类, 并使用“CharlieHebdo”这一话题作为新出现的领域进行主动学习, 其他类别作为初始有标注的数据集。每个数据集包含文本、图像和评论3种信息。本

实验中使用3种模态特征来检测谣言, 即文本特征、视觉特征和传播结构特征。因此, 对缺失值进行了处理, 剔除了未能同时包含文本、图像和传播结构3种信息的样本。其次, 在文本处理中, 进行了话题分类, 以便更好地组织和分析数据。除此之外, 在使用ALiPy主动学习框架时, 数据划分策略将测试集和训练集的比例设置为3:7。

表1为预处理后两个数据集的统计信息, 其中主动学习中AL_train数据集和test数据集是由ALiPy框架^[21]按照比例根据主动学习数据集的大小进行分割的。筛选样本数量参数 N 为人工设置, 主动学习轮次参数MAX_Q受 N 和无标签样本集大小的影响。针对不同的(N)设置, 进行了敏感性分析, 图4展示了不同设置下初始筛选轮次的准确率。当设置($N=20$)时, 能够较为明显地展现模型在初次筛选后的优良性能。因此, 在PHEME数据集上, 每次筛选样本数量 $N=20$, 且参数 $a=6.76$; 在微博数据集上, 每次筛选样本数量 $N=20$, 且 $a=11.76$ 。硬件环境上, 所有实验均在一块NVIDIA Tesla P100 GPU上进行训练。软件环境上, 使用Python 3.8及PyTorch深度学习框架进行编程。

表1 数据集信息

Table 1 Datasets information

数据集	PHEME	微博	
第 L 个话题数据集	无标签数据集	296	157
	测试集	127	68
	总计	423	225
前 $L-1$ 个话题数据集		1210	947
总计	1633	1172	

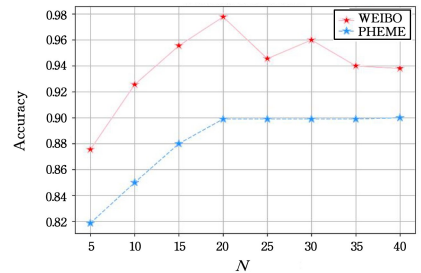


图4 不同大小筛选样本集设置下初始轮次的结果

Fig. 4 Results of initial rounds under different sample set sizes

5.2 对比模型

EANN^[27]: 一个基于GAN的模型。该模型利用文本和图像数据提取事件的不变特征, 适用于新出现的事件。

MVAE^[28]: 使用双模态可变自编码器, 结合二分类器进行多模态假新闻检测。

GLAN^[29]: 共同编码局部语义和全局结构信息, 并应用全局-局部注意力网络进行谣言检测。

MRDMPML^[30]: 仅利用文本和图像特征, 并通过在这两种特征上同时应用提示学习, 实现了这两种模态之间的更好对齐。

REPORT^[31]: 结合用户信息和传播结构两种特征, 通过图神经网络学习用户间的关联关系, 并采用树状结构建模信息传播过程, 提升了模型的准确性和鲁棒性。

MFAN^[23]: 使用多模态特征增强注意力网络进行谣言检测, 利用文本、视觉和社交图特征, 并通过推断隐藏链接来改

进社交图特征学习。

ACP-RD^[19]:基于主动学习、对比学习和提示学习,可随谣言领域变化而迭代更新的检测框架。

SVM(RBF) + Unc(En), SVM(Linear) + QBC, SVM(SGD) + QBC^[32]仅使用文本特征,分别使用结合高斯核函数的支持向量机和基于熵的不确定性查询策略、线性支持向量机和基于委员会的查询策略、结合随机梯度下降的支持向量机和基于委员会的查询策略进行谣言检测。

EANN, MVAE, MRDMPL 使用文本特征和图像特征; GLAN 使用传播结构特征; REPORT 使用用户信息特征和传播结构特征; MFAN 使用文本特征、图像特征和传播结构特征,通过简单的拼接方式进行特征融合; ACP-RD, SVM

(RBF) + Unc(En), SVM(Linear) + QBC, SVM(SGD) + QBC 通过主动学习方法实现少样本谣言检测,但仅依赖文本特征进行检测,未充分利用其他模态的信息。相比之下,ALMF 基于传播结构图增强的主动学习,采用多层迭代注意力机制以实现跨模态特征增强,并优化了多模态融合的方式。

5.3 结果与讨论

表 2 展示了 ALMF 模型与对比模型的性能对比,最优结果和次优结果分别用加粗和下划线进行突出显示。其中,ALMF-w/o G 表示不使用传播结构图增强的查询策略; ALMF-QBC 表示筛选样本模块的查询策略使用基于委员会查询(Query-By-Committee),多模态融合检测模块使用 ALMF 模型的 MF 模块。

表 2 主要实验结果

Table 2 Main experimental results

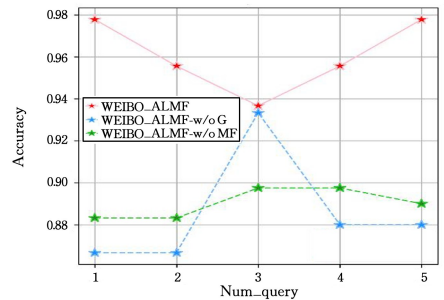
(%)

Method	PHEME				WEIBO			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
EANN	77.13	71.39	70.07	70.44	80.96	80.19	79.68	79.87
MVAE	77.62	73.49	72.25	72.77	71.67	70.52	70.21	70.34
GLAN	83.32	81.25	77.13	78.51	82.44	82.45	80.86	81.26
MRDMPL	86.00	89.60	78.40	83.60	92.90	92.80	93.00	92.90
REPORT	86.00	93.80	82.80	87.70	—	—	—	—
MFAN	89.56	88.48	88.26	87.20	90.38	90.51	89.81	89.86
ACP-RD	82.48	82.32	82.32	82.23	94.35	95.79	95.12	95.45
SVM(RBF) + Unc(En)	71.40	—	—	—	88.40	—	—	—
SVM(Linear) + QBC	71.90	—	—	—	87.90	—	—	—
SVM(SGD) + QBC	68.50	—	—	—	85.40	—	—	—
ALMF-QBC	89.76	89.53	89.79	89.66	88.89	88.99	88.89	88.94
ALMF-w/o G	89.76	89.19	89.76	89.47	93.33	93.44	93.45	93.44
ALMF	92.13	92.13	92.13	92.13	97.78	97.85	97.78	97.81

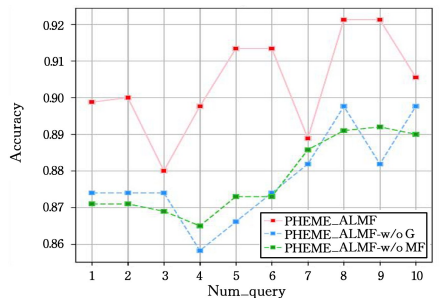
如表 2 所列,在两个数据集上,ALMF 模型性能均优于其他模型,在 PHEME 数据集上准确率为 92.13%,在微博数据集上准确率高达 97.78%。相比较于 ALMF-w/o G 和 ALMF-QBC,ALMF 在 PHEME 数据集上提升了约 2.3 个百分点;在微博数据集上提升了 4~9 个百分点。在 PHEME 数据集上 ALMF-QBC 的表现次优;在微博数据集上 ACP-RD 表现次优。

本文实验还比较了 PHEME 和微博两个数据集上 ALMF,ALMF-w/o G 和 ALMF-w/o MF 的表现,消融实验结果如图 5 所示。其中 ALMF-w/o MF 表示在筛选样本模块 ALG 中使用了传播结构图增强的查询策略,但在检测模块中未使用 ALMF 提出的 MF 模块。比较 ALMF 和 ALMF-w/o G 的结果,在首次数据筛选后,ALMF 的准确率较 ALMF-w/o G 高出 2.5~10 个百分点。特别是在微博数据集上,ALMF 的最低准确率始终高于 ALMF-w/o G 的最高准确率,并且在每次数据筛选后,ALMF 的效果始终不低于同轮次的 ALMF-w/o G。此外,ALMF 仅标注约 5% 的样本便可达到全样本阶段的准确率,验证了所提出的 ALG 模块的有效性。对比 ALMF 和 ALMF-w/o MF 的结果,ALMF 的准确率比 ALMF-w/o MF 的准确率高 2~5 个百分点,且 ALMF 的效果也始终优于同轮次的 ALMF-w/o MF,这表明了所提出的 MF 模块的有效性。总体而言,ALMF 的有效性主要源于 ALG 模块结合不确定性查询策略和传播结构特征的主动学习采样策略,弥补了主动学习传统采样策略在谣言

检测方面的不足;同时 Mf 模块结合了文本特征、图像特征以及传播结构特征,优化了多模态融合的方法,使用多个不同的模态得到了更全面的融合特征。



(a) Accuracy of WEIBO



(b) Accuracy of PHEME

图 5 PHEME 和微博数据集上消融实验的结果
Fig. 5 Results of different query strategy rounds on PHEME and WEIBO datasets

多层协同注意力机制的不同迭代层数 l 对实验结果的影响如图 6 所示。起初,ALMF 模型因出现欠拟合的现象,准确率比较低;随着 l 的增加,ALMF 模型的准确率逐步提升;在 $l=6$ 时,ALMF 在 PHEME 和微博数据集上均取得最好的效果;随后,随着 l 的增加,ALMF 模型出现过拟合的现象,准确率逐步降低。故实验中,ALMF 的多层协同注意力机制的迭代次数为 $l=6$ 。

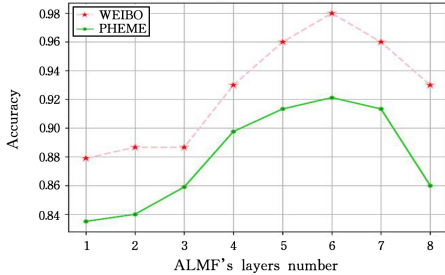


图 6 不同协同注意力迭代层数的准确率

Fig. 6 Accuracy of different levels of cooperative attention iteration

为了验证不同场景下 ALMF 的有效性,从 PHEME 和微博两个数据集的 10 个话题中,选择了 4 个话题(Charlie Hebdo、Sydney Siege、娱乐、社会)分别作为无标签样本集进行实验,实验结果如图 7 所示。在 PHEME 数据集上,Charlie Hebdo 话题指的是法国巴黎的“查理周刊”总部遭遇恐怖袭击的事件,Sydney Siege 则是关于悉尼劫持事件,两个话题属于不同事件的文化背景,但是两个话题在每轮筛选中准确率均值均保持在 3.5% 以内,效果较为稳定,且两个话题经过 4 轮筛选后都能得到比较好的效果,显示出较为一致的性能。在微博数据集中,两个话题的准确率在初始两个轮次中存在较大差距。波动主要源于不同话题的子数据集大小极不平衡。具体而言,微博数据集共包含 1172 个样本,其中“社会”话题约有 800 个样本,而“娱乐”话题仅有约 200 个样本。因此,当“社会”话题作为第 L 个数据集时,初始的有标签数据集 (L_{i-1}) 仅包含约 300 个样本,这导致第一轮筛选后的准确率仅约为 87%。然而,随着多轮筛选的进行,通过逐步筛选有效的有标签数据集,模型的性能逐渐稳定,并在第 5 轮筛选后达到峰值。与表 2 中其他对比模型相比,在 PHEME 数据集上,两个话题的最优准确率分别为 92.13% 和 90.00%,均优于对比模型;在微博数据集中,两个话题的最优准确率分别为 97.78% 和 95.30%,同样超过了其他对比模型。这些结果有效地验证了 ALMF 模型在不同背景主题下性能的稳健性。

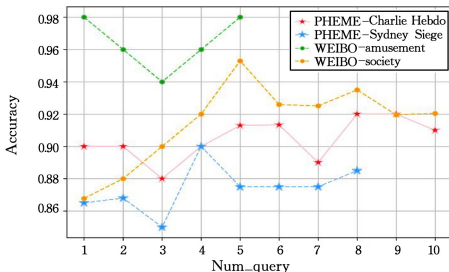


图 7 PHEME 和微博数据集上不同话题的结果

Fig. 7 Results of different topics in PHEME and WEIBO

加筛选样本集的大小(即参数 N),以丰富的有标签数据集来缓解数据不均衡的问题。另外,微博数据集中的数据不均衡现象发生的根本原因在于使用了第三方微博新闻话题的划分模型 THUCTC,而不像推特数据集那样利用自带的话题标签进行划分,这使得某些话题的样本数量较少,从而影响了模型的稳定性和效果。为此,未来将进一步优化新闻话题划分方法,探索更精准的划分策略,以提升数据集的均衡性和模型的整体性能。

结束语 针对目前少样本的谣言检测仍然存在的问题,提出了一个新的基于主动学习的多模态谣言检测模型 ALMF:通过使用传播结构图增强的主动学习和跨模态增强及对齐,实现了少样本的多模态谣言检测。在 PHEME 和微博两个真实数据集上进行实验,结果表明,在第一次筛选样本后,ALMF 在两个数据集上就可以取得较好的效果,证明了 ALMF 在传播早期的谣言检测上具有实际应用的价值。在后续的研究中,将继续探索少样本的谣言检测中存在的问题,进一步研究大模型(LLM)和扩散模型在突发的少样本谣言检测方面应用的可能性和效果。

参考文献

- [1] GAO Y, LIANG G, JIANG F, et al. Overview of social network rumor detection [J]. Acta Electronica Sinica, 2020, 48(7): 1421.
- [2] SHU K, SLIVA A, WANG S, et al. Fake news detection on social media: A data mining perspective [J]. ACM SIGKDD Explorations Newsletter, 2017, 19(1): 22-36.
- [3] LIU H, CHEN S, CAO S, et al. Research on False News Detection based on multi-modal Learning [J]. Exploration of Computer Science and Technology, 2023, 17(9): 2015-2029.
- [4] JIN Z, CAO J, ZHANG Y, et al. Novel visual and statistical image features for microblogs news verification [J]. IEEE Transactions on Multimedia, 2016, 19(3): 598-608.
- [5] JIN Z, CAO J, HAN G, et al. Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs [C] // Proceedings of the 25th ACM International Conference on Multimedia. ACM, 2017: 795-816.
- [6] LU H Y, FAN C Y, WU X J. Small-sample COVID-19 rumor detection for network social media [J]. Journal of Chinese Information Technology, 2022, 36(1): 135-144, 172.
- [7] RAN H, JIA C. Unsupervised cross-domain rumor detection with contrastive learning and cross-attention [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2023: 13510-13518.
- [8] LIN H, YI P, MA J, et al. Zero-shot rumor detection with propagation structure via prompt learning [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2023: 5213-5221.
- [9] HASAN M S, ALAM R, ADNAN M A. Truth or lie: Pre-emptive detection of fake news in different languages through entropy-based active learning and multi-model neural ensemble [C] // 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2020: 55-59.
- [10] CASTILLO C, MENDOZA M, POBLETE B. Information credi-

针对微博样本不均衡的问题,可以在每轮筛选中适当增

- bility on twitter[C] // Proceedings of the 20th International Conference on World Wide Web. 2011:675-684.
- [11] YANG F, LIU Y, YU X, et al. Automatic detection of rumor on sina weibo[C] // Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. 2012:1-7.
- [12] YANG Y, ZHENG L, ZHANG J, et al. TI-CNN: Convolutional Neural Networks for Fake News Detection [J]. arXiv: 1806.00749, 2018.
- [13] YANG X, LYU Y, TIAN T, et al. Rumor detection on social media with graph structured adversarial learning[C] // Proceedings of the Twenty-ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021: 1417-1423.
- [14] WEN W, SU S, YU Z. Cross-lingual cross-platform rumor verification pivoting on multimedia content [J]. arXiv: 1808.04911, 2018.
- [15] ZUO Y, ZHU W, CAI G G. Continually detection, rapidly react: unseen rumors detection based on continual prompt-tuning [C] // Proceedings of the 29th International Conference on Computational Linguistics. 2022:3029-3041.
- [16] SETTLES B. Active learning literature survey; TR1648 [R]. University of Wisconsin-Madison Department of Computer Sciences, 2009.
- [17] SAHAN M, SMIDL V, MARIK R. Active learning for text classification and fake news detection[C] // 2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC). IEEE, 2021: 87-94.
- [18] FARINNEYA P, POUR M M A, HAMIDIAN S, et al. Active learning for rumor identification on social media[C] // Findings of the Association for Computational Linguistics; EMNLP 2021. 2021: 4556-4565.
- [19] SHAO Z, CAI G, LIU Q, et al. An Active Learning Framework for Continuous Rapid Rumor Detection in Evolving Social Media [C] // 2024 International Joint Conference on Neural Networks (IJCNN). IEEE, 2024: 1-8.
- [20] CUI C, JIA C. Propagation Tree Is Not Deep: Adaptive Graph Contrastive Learning Approach for Rumor Detection[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2024: 73-81.
- [21] TANG Y P, LI G X, HUANG S J. ALiPy: Active learning in python[J]. arXiv: 1901.03802, 2019.
- [22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [23] ZHENG J, ZHANG X, GUO S, et al. MFAN: Multi-modal Feature-enhanced Attention Networks for Rumor Detection[C] // IJCAI. 2022: 2413-2419.
- [24] SONG C, YANG C, CHEN H, et al. CED: Credible early detection of social media rumors[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(8): 3035-3047.
- [25] ZUBIAGA A, LIAKATA M, PROCTER R. Exploiting context for rumour detection in social media[C] // Social Informatics; 9th International Conference. Springer, 2017: 109-123.
- [26] SUN M, LI J, GUO Z, et al. Thuctc: an efficient chinese text classifier[R]. GitHub Repository, 2016.
- [27] WANG Y, MA F, JIN Z, 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 & Data Mining. 2018: 849-857.
- [28] 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.
- [29] YUAN C, MA Q, ZHOU W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection [C] // 2019 IEEE International Conference on Data Mining (ICDM). IEEE, 2019: 796-805.
- [30] CHEN F, LI X, LI Z, et al. Multimodal Rumor Detection via Multimodal Prompt Learning [C] // 2024 International Joint Conference on Neural Networks (IJCNN). IEEE, 2024: 1-8.
- [31] LIU T, CAI Q, XU C, et al. Rumor Detection with a novel graph neural network approach[J]. arXiv: 2403.16206, 2024.
- [32] YI F, LIU H, HE H, et al. A Comparative Analysis of Active Learning for Rumor Detection on Social Media Platforms[J]. Applied Sciences, 2023, 13(22): 12098.



SHANG Yunxian, born in 2000, post-graduate, is a member of CCF (No. U6275G). Her main research interests include natural language processing and rumor detection.



CAI Guoyong, born in 1971, Ph.D, professor, Ph. D supervisor, is a distinguished member of CCF (No. 12524D). His main research interests include multi-modal affective computing, trustworthy AI theory and techniques.

(责任编辑:喻黎)