计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(3):264−272 [doi: 10.15888/j.cnki.csa.009439]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于交叉特征感知融合的多领域虚假新闻检测①

王振琦, 陈　涛, 张宝宇, 张明利, 孙晨瑜, 张卫山

(中国石油大学 (华东) 青岛软件学院, 青岛 266580)
通信作者: 王振琦, E-mail: wyx151425@163.com

摘　要: 各领域虚假新闻的传播对社会造成了严重的影响, 不同领域间新闻的领域偏移问题和跨域关联问题也对模型的预测能力造成了极大的挑战. 针对上述问题, 本文提出了一种基于交叉特征感知融合的多领域虚假新闻检测方法. 该方法可以捕捉不同领域间新闻的多种特征差异, 并挖掘新闻之间的关联关系, 从多个维度控制模型在不同领域的特征融合策略. 此外, 本文还提出了一种联合训练框架. 本方法的模型使用本框架进行训练, 在中英文数据集上的预测 *F*1 分数分别达到了 92.84% 和 85.49%, 相较于最先进的模型, 预测效果分别提升了 1.16% 和 1.07%.

关键词: 领域偏移; 跨域关联; 交叉特征感知融合; 多领域虚假新闻检测; 联合训练框架

引用格式: 王振琦,陈涛,张宝宇,张明利,孙晨瑜,张卫山.基于交叉特征感知融合的多领域虚假新闻检测.计算机系统应用,2024,33(3):264–272. http://www.c-s-a.org.cn/1003-3254/9439.html

Multi-domain Fake News Detection Based on Cross-feature Perception Fusion

WANG Zhen-Qi, CHEN Tao, ZHANG Bao-Yu, ZHANG Ming-Li, SUN Chen-Yu, ZHANG Wei-Shan

(Qingdao Institute of Software, China University of Petroleum, Qingdao 266580, China)

Abstract: The dissemination of false news in various domains has a serious impact on society. The problem of domain shift and cross-domain correlation of news between different domains also poses a great challenge to the prediction ability of the model. To address the above problems, this study proposes a multi-domain fake news detection method based on cross-feature perception fusion. This method can capture multiple feature differences in news between different domains, mine the correlations between news, and control the feature fusion strategy of the model in different domains from multiple dimensions. In addition, this study proposes a joint training framework that is adopted to train the proposed model. The model achieves a predictive *F*1 score of 92.84% and 85.49% on the English and Chinese datasets, respectively. Compared to the state-of-the-art model, the prediction results of the proposed model are improved by 1.16% and 1.07%, respectively.

Key words: domain shift; cross-domain correlation; cross-feature perception fusion; multi-domain fake news detection; joint training framework

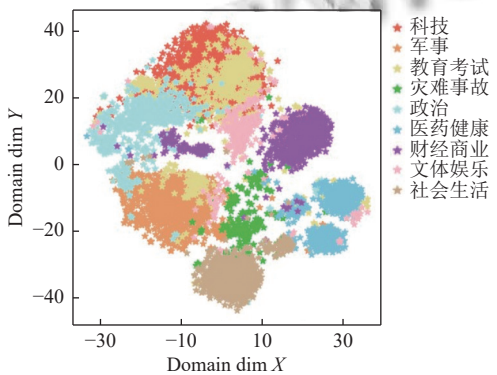
互联网的飞速发展改变了民众获取新闻的方式, 但也加剧了各领域虚假新闻的传播. 每年各个领域层出不穷的虚假新闻不仅对人们的身心健康造成了严重的伤害, 也对国家社会安全造成了巨大的挑战[1]. 当前多领域虚假新闻检测任务面临两个问题, 分别为领域偏移问题和跨域关联问题.

不同领域的新闻在用词、情感表达、写作风格运用上往往存在明显的差别, 该现象被称为领域偏移问题[2]. 由于该问题的存在, 以往使用普通模型不区分领域直接预测的方式, 通常因无法成功捕捉到各领域间

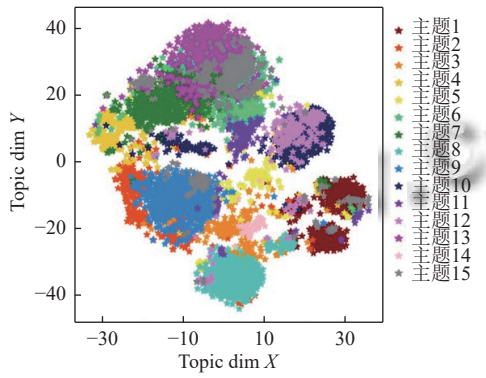
① 基金项目: 国家自然科学基金 (62072469)
收稿时间: 2023-09-20; 修改时间: 2023-10-25; 采用时间: 2023-11-03; csa 在线出版时间: 2024-01-19
CNKI 网络首发时间: 2024-01-22

新闻的差异, 导致模型在各领域上的预测指标差别较大, 从而制约了综合预测指标的提升.

现实中的某个新闻主题往往会引起不同领域内新闻的讨论, 即跨领域主题关联现象, 该现象如图 1 所示. 图 1(a) 是使用基线模型 MFEFND[3]提取新闻特征并使用 t-SNE 对特征降维, 在二维平面上对新闻领域分布情况的展示, 不同颜色表示不同的领域. 图 1(b) 是使用 K-means 算法对新闻特征进行主题聚类后, 对主题划分情况的降维展示. 由图 1(a) 和图 1(b) 相互对比可知, 不同领域的新闻确实存在主题关联现象. 该现象必然造成新闻领域边界划分的困难, 阻碍模型对新闻所属领域的判断, 进而使模型无法准确地运用领域特定的特征提取策略, 进一步制约了模型预测能力的提高. 本文将该现象引起的问题称为跨域关联问题.



(a) 领域分布情况



(b) 主题聚类结果

图 1 新闻主题聚类可视化

针对上述两个问题, 本文提出了一种基于交叉特征感知融合的多领域虚假新闻检测 (multi-domain fake news detection based on cross-feature perception fusion, CFPFND) 方法. 本方法首先从文本意群划分、文本语义表述、情感表达、写作风格这 4 种角度提取新闻特征, 运用交叉注意力机制计算特征之间的交叉注意力权重, 从更深层次上捕捉不同特征之间的运用与表达差异. 本方法构建了 4 种新闻空间, 以新闻空间中的主题特征作为空间特征, 对待检测目标新闻进行空间感知, 通过主题相似性发现待检测新闻的跨领域关联关系. 本方法从新闻内容、新闻所属领域、新闻跨领域关联关系这 3 个维度控制交叉特征的融合过程, 使模型能够针对新闻内容, 从不同领域和不同关联关系中推理出新闻的特征提取策略, 有效应对领域偏移问题和跨域关联问题.

为了进一步提高模型对新闻所属领域的辨别能力, 本方法提出了一种联合训练框架对模型进行训练. 本框架以虚假新闻检测任务为主任务, 以领域分类任务为监督任务, 将两个任务同时训练并回传损失, 对模型参数进行优化, 通过提高模型的领域分类能力, 进一步提高其虚假新闻的检测能力.

综上所述, 本文的创新性如下.

1) 通过对新闻数据进行分析, 本文提出了不同领域间新闻存在的跨域关联问题.

2) 针对领域偏移问题和跨域关联问题, 本文提出了一种基于交叉特征感知融合的多领域虚假新闻检测方法, 捕捉不同领域间新闻的差异, 挖掘新闻之间的关联关系, 从多个维度控制模型在各领域上的特征融合策略.

3) 本文在中英文数据集上使用联合训练框架对本方法的模型进行训练, 模型取得了更优的预测效果. 相较于最先进的模型, 本模型在中英文数据集上的 $F1$ 分数分别达到了 92.84% 和 85.49%, 分别提升 1.16% 和 1.07%, 证明了本文所提出的方法和框架的有效性.

1 相关工作

虚假新闻检测作为一个自然语言处理与社会计算交叉的研究领域, 近年来逐渐受到研究人员的关注. 而深度神经网络具有较强的数据建模能力, 在复杂任务中具有更优的表现, 因此, 深度神经网络一直在虚假新闻检测任务中扮演着重要的角色. Ma 等人[4]首次将循环神经网络应用到虚假新闻检测任务中, 以隐藏层向量作为分类器的输入特征, 取得了较好的预测效果. Yu 等人[5]对新闻文本进行直接建模, 将文本的特征向量拼接成矩阵, 然后利用卷积神经网络提取新闻文本矩阵的特征并进行分类.

随后, 多任务学习, 生成式深度学习, 图神经网络也逐渐被引入到虚假新闻检测任务中. Ma 等人[6]首次将多任务学习方法应用到虚假新闻检测任务中, 以循环神经网络作为主干, 同时训练虚假新闻检测任务与立场检测任务, 通过两种训练任务促进虚假新闻检测模型性能的提升. Cheng 等人[7]利用变分自编码器编码新闻文本信息, 获取新闻文本的嵌入表示, 随后将新闻向量运用于多任务学习. Vaibhav 等人[8]将新闻文章中的句子构建为节点, 按照节点相似度构建节点之间的边, 把虚假新闻检测任务转换为图分类任务. Li 等人[9]利用预训练的事实核查模型在外部知识中查找事实证据, 将事实证据与新闻内容构建为星形图, 使用图卷积网络融合新闻内容和事实证据. Ma 等人[10]利用生成器生成谣言, 将谣言与非谣言相互转化以扩充训练数据, 使用判别器检测虚假新闻, 通过不断训练生成器和判别器, 提升了模型预测的鲁棒性和分类的准确率. Huang 等人[11]提出了一个由自然语言推理指导自我批判序列训练的框架, 用于生成与人类撰写风格策略相似的训练数据, 在生成数据上训练的模型可以在公共数据集上取得较好的效果. Sheng 等人[12]参考了事实核查方法, 使用新闻传播环境中具有时间关联的新闻作为辅助检测的依据, 证明了从宏观和微观两种新闻环境感知检测依据的有效性.

随着虚假新闻检测研究的不断深入, 多领域与跨领域虚假新闻检测逐渐成为研究的热点. Nan 等人[3]首次提出了多领域虚假新闻检测任务, 设计了该任务的基线模型, 并公开了一个多领域数据集. Zhu 等人[2]在新闻文本特征的基础上, 加入了情感和风格特征对新闻真假性进行预测, 使用领域记忆发掘新闻的潜在领域标签. Nan 等人[13]使用元学习方法在多个领域的数据集上进行训练, 获得通用预测模型, 使用困惑度为指标, 研究了通用模型在新的领域上的迁移能力. Yue 等人[14]提出了一种基于元学习的方法, 用于领域自适应的小样本虚假信息检测, 使用多个源任务训练初始模型, 并计算模型与元任务的相似度得分, 以相似度得分重新调整元梯度, 使模型能自适应地从源任务中学习特征. Ran 等人[15]使用实例级对比学习和原型级对比学习方法对不同领域间新闻的领域偏移问题进行了研究, 使用交叉注意力机制增强了模型在跨领域预测中的鲁棒性. Choudhry 等人[16]使用情感分类作为研究跨领域虚假新闻检测的指标, 并通过实验结果验证了情感引导方法在跨领域虚假新闻检测中的有效性.

2 方法

本节对 CFPFND 方法的模型进行介绍, CFPFND 模型的总体结构如图 2 所示. 本模型包含交叉特征融合单元 (cross feature fusion unit, CFFUnit) 和新闻空间感知单元 (news space perception unit, NSPUnit) 两个主体部分. CFFUnit 使用专家网络[17]提取新闻特征, 使用交叉注意力机制计算特征权重, 使用门控网络控制特征融合过程. NSPUnit 用于构建新闻主题空间, 对待检测新闻的主题关联关系进行感知.

本方法使用中文 BERT 模型提取中文新闻文本特征, 使用 RoBERTa 模型提取英文新闻文本特征, 获得每一条待检测新闻的文本所对应的词向量 $E_{\text{txt}} = \{vt_1, vt_2, \cdots, vt_{N_T}\}$, $t_i \in R^{N_{\text{emb}}}$, 其中 N_{emb} 为预训练语言模型的词向量维度. 本方法参考文献[18]中提出的情感特征提取方法, 对新闻的情感极性及用词、情绪分类、表达情感的标点符号、情感极性词、辅助情感特征进行提取, 获得 5 种表示情感的数字特征, 组成新闻情感的表征向量 $r_{\text{emo}} = \{e_1, e_2, \cdots, e_{N_E}\}$. 本方法参考文献[19]中提出的风格特征提取方法, 获得新闻风格的表征向量 $r_{\text{sty}} = \{s_1, s_2, \cdots, s_{N_S}\}$.

CFPFND 模型为每个领域设置了一个领域特征向量 v_{domain}. 具体来说, 当数据集中共有 N_{domain} 个不同的领域时, 模型在初始化时会随机创建 N_{domain} 个 *Embedding* 作为领域特征向量, 每个领域特征向量只与一个领域相对应. 在模型训练过程中, 待检测新闻所属领域的领域特征向量将作为门控网络的输入之一, 参与门控参数的计算. 每个领域特征向量只针对其所在领域的新闻进行训练, 根据损失进行优化, 因此, 该领域特征向量可以被认为只保存了其所在领域中所有新闻的特征.

2.1 交叉特征融合单元 CFFUnit

CFFUnit 包含 3 种特征提取网络及其对应的门控网络, 3 种特征提取网络分别为: 新闻文本特征提取网络 (news text feature extraction network, TxtExtNet)、新闻情感特征提取网络 (news emotion feature extraction network, EmoExtNet)、新闻风格特征提取网络 (news style feature extraction network, StyExtNet), 每个特征提取网络使用多组专家网络提取特征.

2.1.1 新闻文本特征提取网络 TxtExtNet

TxtExtNet 用于提取新闻文本中的意群成分特征

与总体语义特征. 由于一维卷积神经网络的卷积核可以模拟文本词组划分并进行特征提取, 因此, 本方法使用多组 TextCNN[20]作为意群特征专家网络 (sense group feature expert network, SenExpNet) 提取意群特征 $V^{\text{sen}}=\left(\left\{v_i^{\text{sen}}\right\}_{i=1}^{N_{\text{sen}}}\right)$; 同时, 本方法使用多组 TextRNN[21]作为语义特征专家网络 (semantic feature expert network, SemExpNet) 提取语义特征 $V^{\text{sem}}=\left(\left\{v_i^{\text{sem}}\right\}_{i=1}^{N_{\text{sem}}}\right)$, 意群特征和语义特征提取过程分别如式 (1)、式 (2) 所示. 其中, $v_i^{\text{sen}}, v_i^{\text{sem}}\in R^N$, N 表示 CFPFND 模型中特征向量的维度, N_{sen} 表示 SenExpNet 的数量, N_{sem} 表示 SemExpNet 的数量, $(\cdot)$ 表示将特征向量集合转换为特征矩阵.

$$v_i^{\text{sen}}=\text{SenExpNet}_i(E_{\text{txt}}) \tag{1}$$

$$v_i^{\text{sem}}=\text{SemExpNet}_i(E_{\text{txt}}) \tag{2}$$

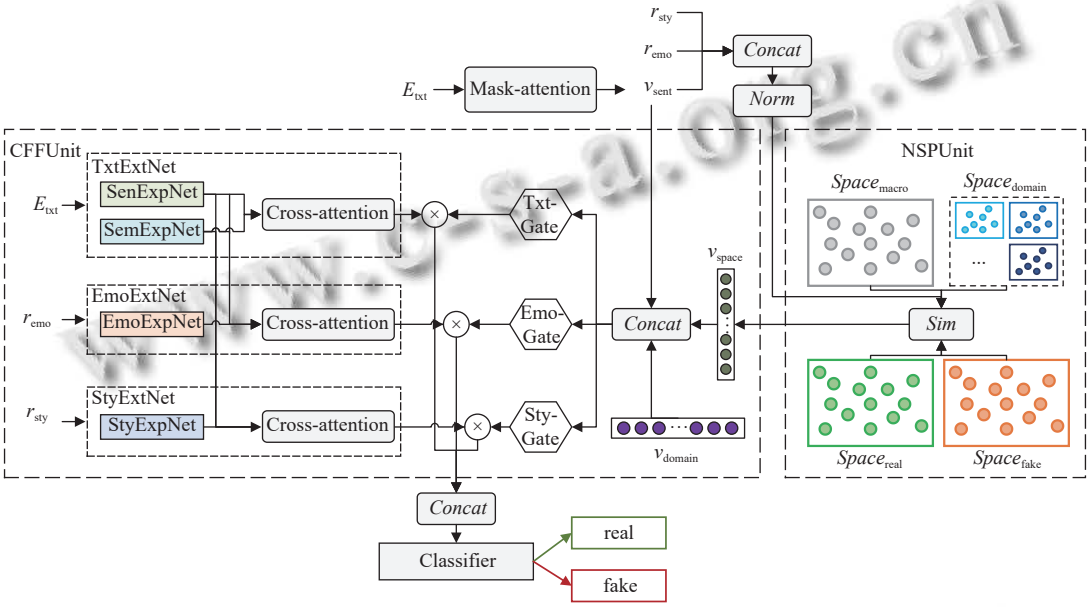


图 2 CFPFND 模型总体结构

TxtExtNet 使用交叉注意力机制 (如式 (3) 所示) 计算意群与语义之间的交叉作用特征, 获得包含语义作用权重的意群交叉特征 $V^{\text{sen}\times\text{sem}}$ 和包含意群作用权重的语义交叉特征 $V^{\text{sem}\times\text{sen}}$, 计算过程分别如式 (4)、式 (5) 所示.

$$f_{\text{att}_{\text{cross}}}\left(Q^s,K^t,V^t\right)=Softmax\left(\frac{Q^s(K^t)^{\text{T}}}{\sqrt{d^k}}\right)V^t \tag{3}$$

$$V^{\text{sen}\times\text{sem}}=f_{\text{att}_{\text{cross}}}\left(V^{\text{sen}},V^{\text{sem}},V^{\text{sem}}\right) \tag{4}$$

$$V^{\text{sem}\times\text{sen}}=f_{\text{att}_{\text{cross}}}\left(V^{\text{sem}},V^{\text{sen}},V^{\text{sen}}\right) \tag{5}$$

意群特征、语义特征及其交叉特征组成新闻文本的交叉特征 $V_{\text{cross}}^{\text{TXT}}$, 如式 (6) 所示, $Concat(\cdot)$ 表示对特征矩阵按最后一维进行拼接.

$$V_{\text{cross}}^{\text{TXT}}=Concat\left(V^{\text{sen}},V^{\text{sem}},V^{\text{sen}\times\text{sem}},V^{\text{sem}\times\text{sen}}\right) \tag{6}$$

2.1.2 新闻情感特征提取网络 EmoExtNet

EmoExtNet 使用多个 MLP 作为情感特征专家网络 (emotion feature expert network, EmoExpNet), 提取不同程度的情感特征 $V^{\text{emo}}=\left(\left\{v_i^{\text{emo}}\right\}_{i=1}^{N_{\text{emo}}}\right), v_i^{\text{emo}}\in R^N$, N_{emo} 表示 EmoExpNet 的数量, 特征提取过程如式 (7) 所示.

$$v_i^{\text{emo}}=EmoExpNet_i(r_{\text{emo}}) \tag{7}$$

EmoExtNet 使用交叉注意力机制分别计算文本中的情感表达与运用特征, 获得情感交叉表达特征 $V^{\text{emo}\times\text{txt}}$ 和情感交叉运用特征 $V^{\text{txt}\times\text{emo}}$, 计算过程分别如式 (8)、式 (9) 所示, 其中 $V^{\text{txt}}=\left(\left\{v_i^{\text{sen}}\right\}_{i=1}^{N_{\text{sen}}},\left\{v_i^{\text{sem}}\right\}_{i=1}^{N_{\text{sem}}}\right)$.

$$V^{\text{emo}\times\text{txt}}=f_{\text{att}_{\text{cross}}}\left(V^{\text{emo}},V^{\text{txt}},V^{\text{txt}}\right) \tag{8}$$

$$V^{\text{txt}\times\text{emo}}=f_{\text{att}_{\text{cross}}}\left(V^{\text{txt}},V^{\text{emo}},V^{\text{emo}}\right) \tag{9}$$

情感特征和情感表达特征、情感运用特征组合为新闻情感的交叉特征 $V_{\text{cross}}^{\text{EMO}}$, 如式 (10) 所示.

$$V_{\text{cross}}^{\text{EMO}}=Concat\left(V^{\text{emo}},V^{\text{emo}\times\text{txt}},V^{\text{txt}\times\text{emo}}\right) \tag{10}$$

2.1.3 新闻风格特征提取网络 StyExtNet

StyExtNet 使用多个 MLP 作为风格特征专家网络

(style feature expert network, StyExpNet), 提取不同角度的风格特征 $V^{\text{sty}}=\left(\left\{v_i^{\text{sty}}\right\}_{i=1}^{N_{\text{sty}}}\right), v_i^{\text{sty}}\in R^N$, N_{sty} 表示 StyExpNet 的数量, 特征提取过程如式 (11) 所示.

$$v_i^{\text{sty}}=\text{StyExpNet}_i\left(r_{\text{sty}}\right) \tag{11}$$

与 EmoExtNet 交叉特征提取过程类似, StyExtNet 使用交叉注意力机制计算获得风格交叉表现特征 $V^{\text{sty}\times\text{txt}}$ 和风格交叉运用特征 $V^{\text{txt}\times\text{sty}}$, 将 3 种特征组合为新闻风格的交叉特征 $V_{\text{cross}}^{\text{STY}}$.

2.1.4 门控网络

CFFUnit 包含 3 种门控网络, 分别为: 文本特征门控网络 (text feature gating network, TxtGate)、情感特征门控网络 (emotion feature gating network, EmoGate)、风格特征门控网络 (style feature gating network, StyGate). 3 种门控网络的计算过程相同, 如图 3 所示.

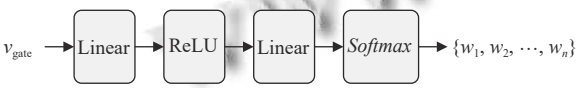


图 3 门控网络的计算过程

门控网络的输入向量 v_{gate} 的计算过程如式 (12) 所示.

$$v_{\text{gate}}=v_{\text{sent}}\oplus v_{\text{domain}}\oplus v_{\text{space}} \tag{12}$$

其中, $\oplus$表示对向量进行简单拼接; v_{sent} 是使用 Mask-attention 机制计算获得的新闻文本句向量; v_{domain} 是待检测新闻所属领域对应的领域特征向量; v_{space} 是新闻空间感知向量, 计算过程如式 (13) 所示, $Norm(\cdot)$ 表示对数据进行标准化处理.

$$v_{\text{space}}=NSPUnit\left(Norm\left(v_{\text{sent}}\oplus v_{\text{emo}}\oplus v_{\text{sty}}\right)\right) \tag{13}$$

3 种门控网络接收 v_{gate} 作为输入, 分别输出 Txt-ExtNet 的交叉特征融合参数 $W^{\text{TXT}}=\left\{w_1,w_2,\cdots,w_{N_{\text{TXT}}}\right\}$、EmoExtNet 的交叉特征融合参数 $W^{\text{EMO}}=\{w_1,w_2,\cdots,w_{N_{\text{EMO}}}\}$ 和 StyExtNet 的交叉特征融合参数 $W^{\text{STY}}=\left\{w_1,w_2,\cdots,w_{N_{\text{STY}}}\right\}$.

2.2 新闻空间感知单元 NSPUnit

NSPUnit 构建了 4 种新闻主题空间, 分别为: 宏观新闻空间 $Space_{\text{macro}}$、现实新闻空间 $Space_{\text{real}}$、谣言新闻空间 $Space_{\text{fake}}$ 和 N_{domain} 个领域新闻空间 $Space_{\text{domain}}$.

2.2.1 新闻空间构建

模型读取数据集中所有的新闻文本数据, 提取特征, 构建新闻组合特征 v_{union}, 如式 (14) 所示.

$$v_{\text{union}}=Norm\left(v_{\text{sent}}\oplus r_{\text{emo}}\oplus r_{\text{sty}}\right) \tag{14}$$

模型将所有新闻的组合特征存入宏观新闻集合, 将真实新闻和虚假新闻的组合特征分别存入真实新闻集合和虚假新闻集合, 按照新闻所属领域的不同, 将其组合特征存入特定的领域新闻集合.

NSPUnit 使用 K-means 聚类算法对 4 种集合内的组合特征进行主题聚类, 将每个新闻主题类簇中心的新闻组合特征 v_{union} 作为该主题的特征 $topic_i$. NSPUnit 以每个空间中的所有主题特征作为该空间的特征, 因此, 宏观新闻空间 $Space_{\text{macro}}=\left\{topic_i^{\text{macro}}\right\}_{i=1}^{N_{t\_\text{macro}}}$, 真实新闻空间 $Space_{\text{real}}=\left\{topic_i^{\text{real}}\right\}_{i=1}^{N_{t\_\text{real}}}$, 虚假新闻空间 $Space_{\text{fake}}=\left\{topic_i^{\text{fake}}\right\}_{i=1}^{N_{t\_\text{fake}}}$, 每个领域的新闻空间 $Space_{domain_n}=\left\{topic_i^{domain_n}\right\}_{i=1}^{N_{t\_\text{domain}}}, n\in\{1,2,\cdots,N_{\text{domain}}\}$, N_{t_macro}、N_{t_real}、N_{t_fake}、N_{t_domain} 表示空间内主题数量. NSPUnit 在每轮训练迭代过程中会使用模型提取的新特征重新构建新闻空间.

2.2.2 新闻空间感知

NSPUnit 提取待检测新闻的组合特征 v_{union}, 以相似度 (计算过程如式 (15) 所示) 作为感知参数在 4 种新闻空间中进行主题感知, 发现每种新闻空间中与待检测新闻组合特征相似度最高的主题特征 $Topic=\left\{topic_{sim}^{\text{macro}},topic_{sim}^{\text{real}},topic_{sim}^{\text{fake}},topic_{sim}^{\text{domain}}\right\}$, 并记录其相似度 $Sim=\left\{sim^{\text{macro}},sim^{\text{real}},sim^{\text{fake}},sim^{\text{domain}}\right\}$, 其中, 领域新闻空间 $Space_{domain_n}$ 所属领域与待检测新闻所属领域相同.

$$Sim(v_{\text{union}},topic_i)=\frac{v_{\text{union}}\cdot topic_i}{\|v_{\text{union}}\|\cdot\|topic_i\|} \tag{15}$$

NSPUnit 以加权相似度 ws^i 对 4 种主题特征进行融合, 得到新闻空间感知向量 v_{space}, 如式 (16) 所示, 加权相似度计算过程如式 (17) 所示. 最后, NSPUnit 将 v_{space} 返回给 *CFFUnit*.

$$v_{\text{space}}=\sum ws^i\cdot topic_{sim}^i,\ i\in[\text{macro},\text{real},\text{fake},\text{domain}] \tag{16}$$

$$ws^i=Softmax(Sim)_i=\frac{sim^i}{sim^{\text{macro}}+sim^{\text{real}}+sim^{\text{fake}}+sim^{\text{domain}}} \tag{17}$$

2.3 预测

模型的分类器使用 MLP 组合 3 种交叉特征, 输出新闻为真或假的概率, 如式 (18) 所示, 其中 $\hat{y}\in[0,1]$. 数

据集中真实新闻的标签为 0, 虚假新闻的标签为 1.

$$\hat{y} = MLP\left(V_{\text{cross}}^{\text{TXT}} W^{\text{TXT}} \oplus V_{\text{cross}}^{\text{EMO}} W^{\text{EMO}} \oplus V_{\text{cross}}^{\text{STY}} W^{\text{STY}}\right) \quad (18)$$

3 实验

3.1 数据集

本实验所使用的中文数据集为 Weibo21 数据集[3], 包含科技、军事、教育考试、灾难事故、政治、医药健康、财经商业、文体娱乐、社会生活 9 个领域, 各领域数据的统计信息如表 1 所示. 英文数据集是由 FakeNewsNet 数据集[22]与 MM-COVID 数据集[23]中的英文文本数据组成的, 包含 Gossipcop、Politifact、COVID 3 个领域的数据集, 各领域数据的统计信息如表 2 所示. 由表 1 和表 2 可知, 两个数据集中不同领域间的数据量存在较大的差异.

表 1 多领域中文数据集统计数据

标签	科技	军事	教育考试	灾难事故	政治	医药健康	财经商业	文体娱乐	社会生活	总数
真	143	121	243	185	306	485	959	1 000	1 198	4 640
假	93	222	248	591	546	515	362	440	1 471	4 488
总数	236	343	491	776	852	1 000	132	1 440	2 669	9 128

表 2 多领域英文数据集统计数据

标签	Gossipcop	Politifact	COVID	总数
真	16 804	447	4 750	2 201
假	5 067	379	1 317	6 763
总数	21 871	826	6 067	28 764

MDEFND 是多领域虚假新闻检测的基线模型, 该模型使用多组 TextCNN 作为专家网络, 提取不同角度的新闻文本语义特征, 使用门控网络融合专家语义特征. M3FEND 在 MDEFND 模型的基础上, 加入了新闻情感与写作风格特征, 使用多头注意力机制融合特征. 但是, 这两个模型都没有考虑跨域关联问题.

3.2 基线模型

本实验选择了 EANN[24]、MMoE[19]、MoSE[25]、EDDFN[26]、MDFEND 和 M3FEND[2], 共 6 个可以应用于多领域虚假新闻检测的模型作为基线模型. EANN 是一个多模态虚假新闻检测模型, 本方法仅使用 EANN 提取新闻文本特征, 单模态 EANN 的结构与 TextCNN 基本一致. MMoE 与 MoSE 是多任务学习模型, MMoE 使用多组专家网络和多组门控网络控制不同任务的特征组合过程. MoSE 是对 MMoE 的改进, 在专家网络之前添加了一个 LSTM 层作为 shared bottom layer. 本实验按领域划分数据集, 将不同领域上的预测任务看作是不同的学习任务, 使 MMoE 和 MoSE 同时对多个领域的新闻做预测. EDDFN 是一种用于跨领域虚假新闻检测的模型, EDDFN 对不同领域进行隐式建模, 同时保留了领域特定的知识和跨领域的知识.

3.3 实验结果

实验使用 6 个基线模型和 CFPFND 模型分别在中文和英文数据集上进行训练, 使用训练好的模型在测试集上进行预测, 统计各项预测结果, 在中文数据集上的预测结果如表 3 所示, 其中单个领域上的预测指标为 *F*1 分数. CFPFND 模型在 7 个领域上的预测效果优于其他基线模型, 在数据量最大的社会生活领域上的 *F*1 分数远高于其他基线模型.

表 3 多领域中文数据集实验结果

模型	科技	军事	教育考试	灾难事故	政治	医药健康	财经商业	文体娱乐	社会生活	预测指标		
										*F*1	Acc	AUC
MMoE	**0.9062**	0.9129	0.8376	**0.9145**	0.8680	0.9350	0.8530	0.8922	0.8753	0.8966	0.8967	0.9551
MoSE	0.8194	0.8779	0.8468	0.8516	0.8879	0.8994	0.8847	0.9138	0.8856	0.8969	0.8969	0.9610
EDDFN	0.8809	0.8982	0.9082	0.8895	0.8615	0.9449	0.8594	0.9069	0.8832	0.9033	0.9033	0.9593
EANN	0.8261	0.9219	0.8679	0.8937	0.8694	0.9350	0.8791	0.9199	0.9040	0.9086	0.9087	0.9623
MDFEND	0.8333	0.9364	0.8987	0.9136	0.8826	0.9450	0.8811	0.9103	0.9019	0.9133	0.9132	0.9708
M3FEND	0.8384	0.9275	0.9091	0.9060	0.8879	0.9550	0.9012	0.9272	0.9094	0.9168	0.9168	0.9723
CFPFND	0.8618	**0.9710**	**0.9185**	0.8965	**0.8973**	**0.9570**	**0.9046**	**0.9276**	**0.9150**	**0.9251**	**0.9251**	**0.9752**

模型在英文数据集上的预测结果如表 4 所示. CFPFND 模型在 Gossipcop 和 COVID 两个领域上的预测结果优于其他基线模型, 在 Politifact 领域上的预测结果与最高预测结果非常接近. Gossipcop 领域的数据量远高于 Politifact 和 COVID 两个领域, 因此在该领域上 *F*1 分数的高低将直接决定 3 项综合预测指标

的高低.

通过表 3 和表 4 各模型间预测结果的对比可知, CFPFND 模型在 $F1$ 分数、准确率、AUC 这 3 项指标上均取得了提升, 预测效果均优于所有的基线模型, 在多个领域上的 $F1$ 分数也优于基线模型. 综上所述, 相较于基线模型, 本文所提出的检测方法及其模型, 具有较优的预测能力.

表 4 多领域英文数据集实验结果

模型	Gossipcop	Politifact	COVID	预测指标		
				$F1$	Acc	AUC
EANN	0.7856	0.8261	0.9016	0.8135	0.8677	0.9078
MoSE	0.7983	0.8564	0.9402	0.8351	0.8767	0.9249
EDDFN	0.8057	0.8436	0.9346	0.8366	0.8810	0.9214
MMoE	0.8050	0.8357	0.9445	0.8396	0.8821	0.9269
MDFEND	0.8057	**0.8609**	0.9421	0.8404	0.8826	0.9179
M3FEND	0.8216	0.8265	0.9341	0.8442	0.8863	0.9214
CFPFND	**0.8263**	0.8601	**0.9596**	**0.8512**	**0.8898**	**0.9261**

3.4 联合训练框架

本框架以 CFPFND 模型的虚假新闻检测任务作为主任务, 以领域分类任务作为监督任务, 将主任务的预测损失与监督任务的预测损失按权重相加, 并将联合损失回传, 对模型参数进行优化, 框架执行过程如图 4 所示.

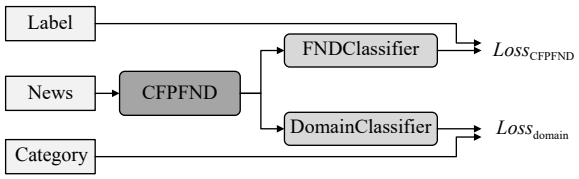


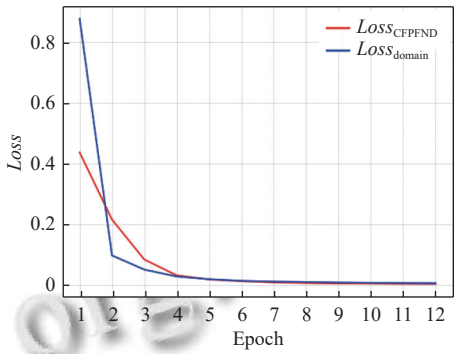
图 4 联合训练框架执行过程

联合损失计算过程如式 (19) 所示, 其中 $Loss_{CFPFND}$ 为 CFPFND 模型的预测损失, $Loss_{domain}$ 为领域分类预测损失, α 为权重参数.

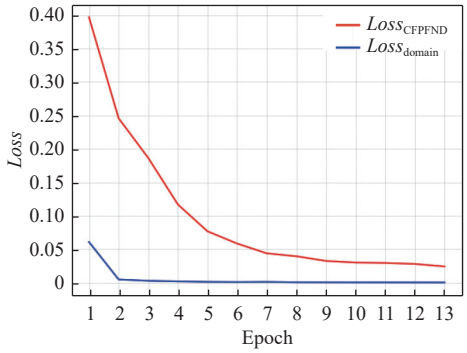
$$Loss = Loss_{CFPFND} + \alpha \times Loss_{domain} \tag{19}$$

具体而言, CFPFND 模型组合 3 种交叉特征之后, 分别送入虚假新闻检测分类器 FNDClassifier 和领域分类器 DomainClassifier, 两个分类器均为 MLP. FND-Classifier 输出新闻真与假的概率, DomainClassifier 输出新闻属于各个领域的概率值. 模型使用二元交叉熵损失函数计算 $Loss_{CFPFND}$, 使用交叉熵损失函数计算 $Loss_{domain}$. 同时, 参数 α 在训练过程中手动调整, 模型以获得最优预测效果的权重数值作为最终 α 取值.

使用联合训练框架对 CFPFND 模型进行训练, 模型在中文和英文数据集上的预测损失变化曲线分别如图 5(a) 和图 5(b) 所示. 由图 5 可知, 虚假新闻检测的预测损失与领域分类预测损失具有相同的变化趋势, 证明两个任务具备相关性.



(a) 中文数据集预测损失变化趋势



(b) 英文数据集预测损失变化趋势

图 5 数据集预测损失变化趋势

3.5 联合训练实验结果

本联合训练实验以 0.1 作为权重取值间隔, 在区间 [0, 1]上取 11 个数值作为权重参数 α 的值. 联合训练实验过程按照联合训练框架, 在每个 α 取值上进行 10 轮以上的训练, 记录每个 α 取值下最优的预测结果, 使用联合训练框架训练的模型在中文和英文测试集上取得的预测结果分别如表 5 和表 6 所示.

由表 5 和表 6 可知, 当中文和英文数据集上 α 的取值分别为 0.1 和 0.5 时, 模型能够获得最优的预测效果. 通过对比可知, 使用联合训练框架后, 取得最优预测效果的 CFPFND 模型不仅在 3 个预测指标上都取得了较大的提升, 还使中文数据集上各领域间 $F1$ 分数的差距获得了缩小, 各领域预测结果较为均衡. 同时, 英文数据集上各领域的 $F1$ 分数均高于其他 α 取值. 上述实验结果证明了 CFPFND 模型和联合训练框架在多领域虚假新闻检测上的有效性, 也证明了区分领域差异可以进一步提升模型的预测效果.

表 5 中文数据集实验结果

α	科技	军事	教育考试	灾难事故	政治	医药健康	财经商业	文体娱乐	社会生活	预测指标		
										*F*1	Acc	AUC
0.0	0.8618	0.9710	0.9185	0.8965	0.8973	0.9570	0.9046	0.9276	0.9150	0.9251	0.9251	0.9752
0.1	**0.8809**	0.9664	**0.9292**	0.8821	**0.8986**	**0.9600**	**0.9055**	**0.9283**	**0.9151**	**0.9284**	**0.9284**	**0.9762**
0.2	0.8799	0.9565	0.9285	0.8817	0.8977	0.9600	0.9043	0.9279	0.9136	0.9266	0.9266	0.9760
0.3	0.8761	0.9516	0.8979	0.8792	0.8974	0.9600	0.9044	0.9258	0.9140	0.9256	0.9256	0.9757
0.4	0.8604	0.9437	0.9078	0.9212	0.8942	0.9550	0.8976	0.9271	0.9134	0.9245	0.9244	0.9738
0.5	0.8618	**0.9714**	0.9185	0.8965	0.8976	0.9570	0.9046	0.9277	0.9148	0.9251	0.9251	0.9753
0.6	0.8622	0.9414	0.9082	0.9207	0.8936	0.9550	0.8967	0.9273	0.9137	0.9245	0.9245	0.9738
0.7	0.8614	0.9418	0.9082	**0.9214**	0.8942	0.9550	0.8971	0.9273	0.9148	0.9245	0.9245	0.9738
0.8	0.8724	0.9568	0.8889	0.9105	0.8745	0.9450	0.9032	0.9239	0.9073	0.9234	0.9229	0.9726
0.9	0.8618	0.9416	0.9078	0.9211	0.8936	0.9550	0.8979	0.9273	0.9139	0.9245	0.9245	0.9739
1.0	0.8712	0.9565	0.8887	0.9117	0.8745	0.9450	0.9024	0.9245	0.9067	0.9234	0.9229	0.9726

表 6 英文数据集实验结果

α	Gossipcop	Politifact	COVID	预测指标		
				*F*1	Acc	AUC
0.0	0.8263	0.8601	0.9596	0.8512	0.8898	0.9261
0.1	0.8249	0.8594	0.9412	0.8506	0.8897	0.9249
0.2	0.8261	0.8582	0.9574	0.8511	0.8978	0.9258
0.3	0.8244	0.8357	0.9445	0.8503	0.8891	0.9269
0.4	0.8266	0.8609	0.9421	0.8522	0.9017	0.9279
0.5	**0.8286**	**0.8663**	**0.9601**	**0.8549**	**0.9044**	**0.9356**
0.6	0.8274	0.8612	0.9577	0.8536	0.9035	0.9351
0.7	0.8269	0.8623	0.9570	0.8526	0.9023	0.9349
0.8	0.8249	0.8592	0.9416	0.8506	0.8898	0.9251
0.9	0.8244	0.8584	0.9477	0.8503	0.8891	0.9247
1.0	0.8240	0.8610	0.9381	0.8498	0.8866	0.9245

由于中文 BERT 模型和 RoBERTa 模型在基本分类任务上的良好性能, 当 α 的取值大于 0.0 时, 模型在领域分类任务上的预测准确率都达到了 99% 以上. 然而, 当领域分类任务损失的权重过大时, 使用联合训练框架得到的模型预测效果却弱于 α=0 时的预测效果. 由此可见, 使用权重参数 α 控制损失的相加是必要的.

4 总结

本文提出了一种基于交叉特征感知融合的多领域虚假新闻检测方法, 该方法的模型通过交叉特征融合单元提取多种新闻特征, 捕捉不同特征之间深层次的运用与表达, 使用新闻空间感知单元提取新闻的跨领域关联关系, 利用门控网络执行不同领域新闻的特征融合策略, 有效地应对了领域偏移问题和跨域关联问题. 本文提出的联合训练框架在提高模型的新闻领域辨别能力的同时, 进一步提高了模型预测的效果.

参考文献

1 Guo B, Ding YS, Yao LN, *et al*. The future of false information detection on social media: New perspectives and trends. ACM Computing Surveys, 2020, 53(4): 68.

2 Zhu YC, Sheng Q, Cao J, *et al*. Memory-guided multi-view multi-domain fake news detection. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(7): 7178–7191.

3 Nan Q, Cao J, Zhu YC, *et al*. MDFEND: Multi-domain fake news detection. Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York: Association for Computing Machinery, 2021. 3343–3347.

4 Ma J, Gao W, Mitra P, *et al*. Detecting rumors from microblogs with recurrent neural networks. Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: AAAI Press, 2016. 3818–3824.

5 Yu F, Liu Q, Wu S, *et al*. A convolutional approach for misinformation identification. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: AAAI Press, 2017. 3901–3907.

6 Ma J, Gao W, Wong KF. Detect rumor and stance jointly by neural multi-task learning. Proceedings of the 2018 Web Conference. Lyon: International World Wide Web Conferences Steering Committee, 2018. 585–593.

7 Cheng MX, Nazarian S, Bogdan P. VRoC: Variational autoencoder-aided multi-task rumor classifier based on text. Proceedings of the 2020 Web Conference. Taipei: Association for Computing Machinery, 2020. 2892–2898.

8 Vaibhav V, Mandyam R, Hovy E. Do sentence interactions matter? Leveraging sentence level representations for fake news classification. Proceedings of the 13th Workshop on Graph-based Methods for Natural Language Processing. Hong Kong: Association for Computational Linguistics, 2019. 134–139.

9 Li JW, Ni SW, Kao HY. Meet the truth: Leverage objective facts and subjective views for interpretable rumor detection. Proceedings of the 2021 Findings of the Association for Computational Linguistics. Association for Computational

Linguistics, 2021. 705–715.
10 Ma J, Gao W, Wong KF. Detect rumors on Twitter by promoting information campaigns with generative adversarial learning. Proceedings of the 2019 World Wide Web Conference. San Francisco: Association for Computing Machinery, 2019. 3049–3055.
11 Huang KH, McKeown K, Nakov P, *et al*. Faking fake news for real fake news detection: Propaganda-loaded training data generation. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto: Association for Computational Linguistics, 2023. 14571–14589.
12 Sheng Q, Cao J, Zhang XY, *et al*. Zoom out and observe: News environment perception for fake news detection. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin: Association for Computational Linguistics, 2022. 4543–4556.
13 Nan Q, Wang DD, Zhu YC, *et al*. Improving fake news detection of influential domain via domain-and instance-level transfer. Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju: International Committee on Computational Linguistics, 2022. 2834–2848.
14 Yue ZR, Zeng HM, Zhang Y, *et al*. MetaAdapt: Domain adaptive few-shot misinformation detection via meta learning. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto: Association for Computational Linguistics, 2023. 5223–5239.
15 Ran HY, Jia CY. Unsupervised cross-domain rumor detection with contrastive learning and cross-attention. Proceedings of the 27th AAAI Conference on Artificial Intelligence. Washington: AAAI Press, 2023. 13510–13518.
16 Choudhry A, Khatri I, Chakraborty A, *et al*. Emotion-guided cross-domain fake news detection using adversarial domain adaptation. Proceedings of the 19th International Conference on Natural Language Processing. New Delhi: Association for Computational Linguistics, 2022. 75–79.
17 Ma JQ, Zhao Z, Yi XY, *et al*. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: Association for Computing Machinery, 2018. 1930–1939.
18 Zhang XY, Cao J, Li XR, *et al*. Mining dual emotion for fake news detection. Proceedings of the 2021 Web Conference. Ljubljana: Association for Computing Machinery, 2021. 3465–3476.
19 Yang YT, Cao J, Lu MY, *et al*. How to write high-quality news on social network? Predicting news quality by mining writing style. arXiv:1902.00750, 2019.
20 Kim Y. Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: Association for Computational Linguistics, 2014. 1746–1751.
21 Liu PF, Qiu XP, Huang XJ. Recurrent neural network for text classification with multi-task learning. Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: AAAI Press, 2016. 2873–2879.
22 Shu K, Mahudeswaran D, Wang SH, *et al*. FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data, 2020, 8(3): 171–188. [doi: 10.1089/big.2020.0062]
23 Li YC, Jiang BH, Shu K, *et al*. Toward a multilingual and multimodal data repository for COVID-19 disinformation. Proceedings of the 2020 IEEE International Conference on Big Data. Atlanta: IEEE, 2020. 4325–4330.
24 Wang YQ, Ma FL, Jin ZW, *et al*. EANN: Event adversarial neural networks for multi-modal fake news detection. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: Association for Computing Machinery, 2018. 849–857.
25 Qin Z, Cheng YC, Zhao Z, *et al*. Multitask mixture of sequential experts for user activity streams. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: Association for Computing Machinery, 2020. 3083–3091.
26 Silva A, Luo L, Karunasekera S, *et al*. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI Press, 2021. 557–565.

(校对责编: 牛欣悦)