

韩晓鸿, 杨港, 郭恒. 基于多源特征领域感知的多领域假新闻检测[J]. 智能计算机与应用, 2026, 16(3): 167-171. DOI: 10.20169/j. issn. 2095-2163. 24050501

## 基于多源特征领域感知的多领域假新闻检测

韩晓鸿, 杨港, 郭恒

(河北工程大学 信息与电气工程学院, 河北 邯郸 056038)

**摘要:** 为解决领域数据不平衡和跨多源域学习域特征不变问题, 本文提出基于多源特征对齐和领域感知的多领域假新闻检测模型。首先, 提取包含领域信息的3个特征: 语义特征、风格特征和情感特征; 其次, 设计领域感知门机制, 对多源特征分成共享特征和私有特征两个子空间。共享空间对齐领域特征, 便于对目标域的数据分类; 私有空间使用恢复器与源特征进行对比; 最后, 使用2个分类器, 联合计算损失。此模型旨在减少各个领域之间的差异性, 实现对多领域假新闻的检测。实验结果证明, 相比于其他方法, 本文方法能够显著提高对多领域假新闻检测的性能。

**关键词:** 多领域假新闻检测; 特征对齐; 多源特征; 领域感知

中图分类号: TP391.1

文献标志码: A

文章编号: 2095-2163(2026)03-0167-05

### Multi-domain fake news detection based on multi-source feature domain perception

HAN Xiaohong, YANG Gang, GUO Heng

(School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, Hebei, China)

**Abstract:** To solve the problem of imbalanced domain data and invariant learning domain features across multiple source domains, this paper proposes a multi-domain fake news detection model based on multi-source feature alignment and domain perception. Firstly, extract three features containing domain information: semantic features, style features, and emotional features; secondly, design a domain perception gate mechanism to divide multi-source features into two subspaces: shared features and private features. For the former, align the shared space with the domain features to facilitate data classification in the target domain. For the latter, use a restorer to compare the private space with the source features; finally, use two classifiers to jointly calculate the loss. This model aims to reduce the differences between different fields and achieve the detection of fake news in multiple fields. The experimental results show that compared to other methods, the proposed method can significantly improve the performance of detecting fake news in multiple fields.

**Key words:** multi-domain fake news detection; feature alignment; multi-source features; domain perception

## 0 引言

在线社交网络中, 假新闻是一种故意且可验证的虚假新闻, 假新闻检测旨在将其与真实的新闻区分开来<sup>[1]</sup>。社交媒体平台每天都会产生大量的新闻信息, 并将其归类为不同领域的新闻<sup>[2]</sup> (例如, 社会、政治和军事等), 在许多领域识别假新闻的任务被称为多域假新闻检测<sup>[3]</sup>。而在多领域假新闻检测中, 一直有着领域差异<sup>[4]</sup>的问题。针对领域差异, 多领域假新闻检测的目标是同时对同一模型中的多个领域进行建模, 而跨领域假新闻检测的目标

是将一般知识转移到目标领域。

随着社交媒体的兴起和用户接触信息来源的多样化, 不同领域的假新闻在网络上不时涌现并迅速传播, 给社会公众的信息获取与判断带来了前所未有的挑战。之前相关问题的解决方案一般采用数据驱动的深度学习方法从用户、社交背景和外部知识中提取特征, 用于假新闻检测<sup>[5-6]</sup>。虽然这些方法在训练阶段得到了令人满意的结果, 但却忽略了领域差异, 即数据在分布上有所不同, 主要表现为单词使用、扩散结构、情感<sup>[7]</sup>、类别分布<sup>[8]</sup>等方面的细微差异。因此在面对新出现的新闻样本时表现不佳。

**作者简介:** 韩晓鸿(1972—), 女, 副教授, 主要研究方向: 人工智能, 数据库。Email: 769762410@qq.com; 杨港(1997—), 男, 硕士, 主要研究方向: 自然语言处理, 假新闻检测; 郭恒(1997—), 男, 硕士, 主要研究方向: 自然语言处理, 假新闻检测。

收稿日期: 2024-05-05

哈尔滨工业大学主办 ◆ 专题设计与应用

在实际应用中,测试数据集(目标域)和训练数据集(源域)通常来自不同的时间段、平台、主题和事件<sup>[9]</sup>,能够从中更加直观地看出不同领域之间的相关性和差异性。新闻主题聚类可视化如图1所示。

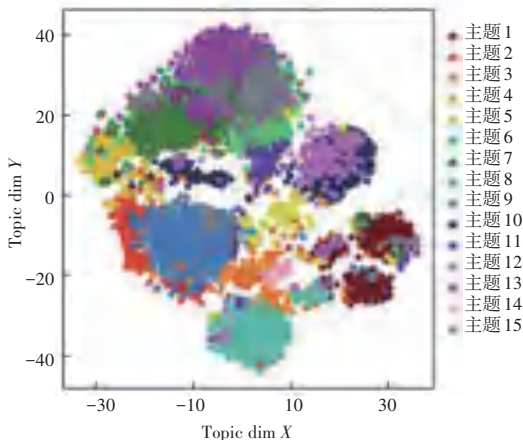


图1 新闻主题聚类可视化

Fig. 1 Visualization of news topic clustering

为解决此问题,采用域自适应方法来学习源域和目标域之间共享的域不变特征(DIF,例如情感语言),以形成通用决策边界<sup>[10]</sup>。虽然泛化能力有所提高,但却需对目标领域样本进行训练或微调,这会引入时间延迟,并失去在早期传播阶段揭穿假新闻的机会。相比之下,域泛化方法努力从多个源域学习DIF以应用于任何目标域<sup>[11]</sup>。综上所述,本文针对领域数据不平衡和跨多个源域学习域不变特征问题,提出基于多源特征领域感知的多领域假新闻检测模型(MFDP)。首先,提取包含领域信息的3个多维特征:语义特征、风格特征和情感特征;其次,设计领域感知门机制,对多源特征分成共享特征和私有特征两个子空间。共享空间对齐领域特征,便于对目标域的数据分类,私有空间使用恢复器与源特征进行对比;最后,使用2个分类器,联合计算损失。

## 1 方法研究

本文模型(MFDP)包含特征提取层、领域感知门、恢复器和分类器四个部分。对于中英文新闻特征,使用BertTokenizer<sup>[12]</sup>对其内容进行标记。模型结构如图2所示。

### 1.1 特征提取层

利用3个深度提取器分别提取包含新闻领域信息的语义特征、情感特征和风格特征。

(1) 语义特征。通过一维卷积神经网络TextCNN<sup>[13]</sup>来提取,用嵌入编码 $\{t_1, \dots, t_{l_{hl}}\}$ 作为语义网络的输入,对此可表示为:

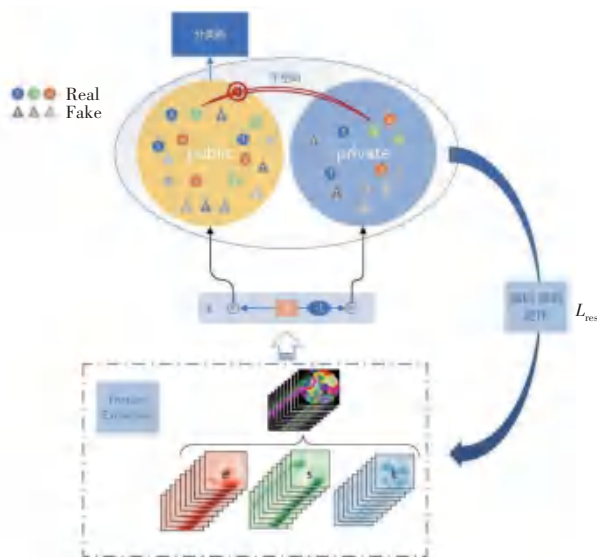


图2 模型结构图

Fig. 2 Model structure diagram

$$\alpha^{\text{sem}} = \text{SemNet}(\{t_1, \dots, t_{l_{hl}}\}) \quad (1)$$

(2) 情感特征。采用多层感知机作为提取网络并引入情感信息流<sup>[14]</sup>的方法提取,将情感嵌入编码 $\{e_1, \dots, e_{l_{el}}\}$ 作为多层感知机的输入,对此可表示为:

$$\beta^{\text{emo}} = \text{EmoNet}(\{e_1, \dots, e_{l_{el}}\}) \quad (2)$$

(3) 风格特征。采用多层感知机作为提取网络,按照文献<sup>[15]</sup>的风格方式进行提取,将风格嵌入编码 $\{s_1, \dots, s_{l_{sl}}\}$ 作为多层感知机的输入,对此可表示为:

$$\gamma^{\text{sty}} = \text{StyNet}(\{s_1, \dots, s_{l_{sl}}\}) \quad (3)$$

使用 $R$ 来表示在一个域中由 $\{(r^{\text{sem}}, r^{\text{emo}}, r^{\text{sty}})\}$ 实例化的随机变量。在多个领域中,用 $R_a$ 表示含有领域信息的变量 $R_a^m = \{(\alpha_n^m, \beta_n^m, \gamma_n^m), y_n^m\}_{n=1}^{N_m}$ ,其中 $N_m$ 是第 $m$ 个领域中的样本数( $1 \leq m \leq M$ ), $y \in \{0, 1\}$ 表示真假性的标签。这里,1表示虚假新闻,0表示真实新闻。

### 1.2 领域感知门

与以往对不同域的原始特征施加对齐约束的研究不同<sup>[16]</sup>,研究中对自适应选择的特征进行对齐,便于分成共享特征和私有特征(即领域特定特征)两个子空间。根据含有领域信息的多维特征,受领域泛化和自适应<sup>[17]</sup>的启发,设计了一个领域感知门,用于将原始特征表示分离成具有范围从0到1的维度信号值的2个互补部分。其公式如下:

$$g = \text{Sigmoid}(R_a W^g + b^g) \quad (4)$$

$$S_{\text{pub}} = g \circ R_a \quad (5)$$

$$P_{\text{pri}} = (1 - g) \circ R_a \quad (6)$$

其中,  $W^s$  和  $b^s$  表示领域感知门的参数;“ $\circ$ ”是矩阵 Hadamard product 符号;  $S_{\text{pub}}$  表示共享特征子空间;  $P_{\text{pri}}$  表示私有特征子空间。

### 1.3 恢复器

在经过领域感知门后,为了防止丢弃尽可能多的私有特征来达到特征对齐的目的,受文献[18]的启发,本模型设计了恢复损失来监督领域感知门的特征提取,防止更多的信息被丢失。首先恢复器使用特征提取器的逆向抽取的方法从分离的特征恢复样本,然后计算原始样本和恢复样本之间的信息损失,最后通过向解码器馈送分离的特征  $S_{\text{pub}}$  和  $P_{\text{pri}}$ , 获得了恢复的样本  $X_{\text{res}}$ , 数学定义公式如下:

$$X_{\text{res}} = D(S_{\text{pub}} + P_{\text{pri}}; W^d) \quad (7)$$

将恢复后的数据与相应的原始数据相比,结果越相似,代表信息损失就越小,推得的公式为:

$$L_{\text{res}} = \frac{1}{N} \sum_{i=1}^N \| X_{\text{res}}^i - X^i \|_1 \quad (8)$$

其中,  $N$  表示样本数量;  $\| \cdot \|_1$  表示  $L1$  损失,使用平均  $L1$  损失的方法,有利于提升模型鲁棒性。

### 1.4 分类器

共享特征子空间是对目标数据的分类,使用 Softmax 的全连接层来预测样本的真假性,公式为:

$$\hat{y}_{\text{pub}} = \text{Sigmoid}(\text{MLP}(S_{\text{pub}})) \quad (9)$$

所有参数都是可学习的,并且可以通过使用反向传播最小化交叉熵损失来优化,具体公式如下:

$$L_{\text{ce\_di}} = - \frac{1}{N} \sum_{i=1}^N y^i \log y_{\text{pub}}^i + (1 - y^i) \log (1 - y_{\text{pub}}^i) \quad (10)$$

同理,在私有特征子空间恢复过程中,需要知道新闻真假性,以便与对应原始数据计算损失。使用 Softmax 的全连接层来预测样本真假性,其公式为:

$$\hat{y}_{\text{pri}} = \text{Sigmoid}(\text{MLP}(P_{\text{pri}})) \quad (11)$$

最后,计算 2 个子空间的联合损失,公式如下:

$$\hat{y}_{\text{all}} = \text{Sigmoid}(\text{MLP}(S_{\text{pub}} + P_{\text{pri}})) \quad (12)$$

$$L_{\text{ce\_all}} = - \frac{1}{N} \sum_{i=1}^N y^i \log y_{\text{all}}^i + (1 - y^i) \log (1 - y_{\text{all}}^i) \quad (13)$$

其中,  $y^i$  和  $y_{\text{all}}^i$  分别表示第  $i$  个样本的真实标签和预测标签;  $N$  表示样本的总数。

## 2 实验与分析

### 2.1 实验环境

实验代码使用的 Pytorch-1.8.0+cu111 版本,

python3.6.13, 所用操作系统为 Windows11。

### 2.2 数据集

本文所使用的中文数据集来自公开的 Weibo21<sup>[19]</sup> 数据集, 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见表 1。表 1 中, “真”与“假”的数据总和分别为 4 640 和 4 488。英文数据集是由 FakeNewsNet 数据集与 MM-COVID 数据集集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见表 2。表 2 中, 真新闻和假新闻的总和分别为 22 001 和 6 763。

表 1 中文数据集

Table 1 Chinese dataset

领域	真	假	领域	真	假
科技	143	93	健康	485	515
军事	121	222	金融	959	362
教育	243	248	娱乐	1 000	440
灾难	185	591	社会	1 198	1 471
政治	306	546			

表 2 英文数据集

Table 2 English dataset

领域	真新闻	假新闻
Gossipcop	16 804	5 067
Politifact	447	379
COVID	4 750	1 317

### 2.3 实验评价指标

现有的大多数假新闻检测任务为二分类,即将新闻分为真或假,评价指标为精确率(Precision)、召回率(Recall)、 $F1$ 值( $F1$ -Score)和准确率(Accuracy)。各指标公式如下:

$$\text{Precision} = \frac{| \text{TP} |}{| \text{TP} | + | \text{FP} |} \quad (14)$$

$$\text{Recall} = \frac{| \text{TP} |}{| \text{TP} | + | \text{FN} |} \quad (15)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (17)$$

### 2.4 实验参数设置

本次实验将数据集按 6 : 2 : 2 的比例随机构建训练集、验证集和测试集,其中测试集中去掉领域标签列,模拟目标域的数据,以此来验证本文模型在检测来自多个未知领域的假新闻时的鲁棒性;模型超参数的设置见表 3。

表3 模型参数

Table 3 Model parameters

参数名称	参数值
batchsize	64
dropout	0.2
lr	5e-4
weight-decay	5e-5

## 2.5 对比实验

为了验证本文模型的性能,本节选取了5个多领域虚假新闻检测的模型作为基线模型,对比结果见表4和表5。图3则是由表4和表5生成的柱状图,能够更加直观地看到模型的对比效果。

表4 英文数据集上的实验结果

Table 4 Experimental results on the English dataset

模型	Gossipcop	Politifact	COVID	overall		
				F1	Acc	AUC
EANN <sup>[20]</sup>	0.793 7	0.755 8	0.883 6	0.812 3	0.874 3	0.905 3
MMoE <sup>[21]</sup>	0.802 2	0.847 7	0.937 9	0.836 1	0.892 0	0.926 5
MoSE <sup>[22]</sup>	0.798 1	0.857 6	0.932 6	0.831 8	0.888 5	0.925 2
EDDFN <sup>[7]</sup>	0.806 7	0.850 5	0.930 6	0.837 8	0.891 2	0.926 3
MDFEND <sup>[19]</sup>	0.808 0	0.847 3	0.933 1	0.839 0	0.893 6	0.923 7
MFDP	<b>0.820 7</b>	<b>0.863 2</b>	<b>0.938 5</b>	<b>0.850 7</b>	<b>0.898 0</b>	<b>0.931 4</b>

表5 在中文数据集上的实验结果

Table 5 Experimental results on the Chinese dataset

模型	科技	军事	教育	灾难	政治	健康	金融	娱乐	社会	overall		
										F1	Acc	AUC
EANN <sup>[20]</sup>	0.822 5	<b>0.927 4</b>	0.862 4	0.866 6	0.870 5	0.910 5	0.871 0	0.895 7	0.887 7	0.897 5	0.897 7	0.961 0
MMoE <sup>[21]</sup>	<b>0.875 5</b>	0.911 2	0.870 6	0.877 0	0.862 0	0.936 4	0.856 7	0.888 6	0.875 0	0.894 7	0.894 8	0.954 7
MoSE <sup>[22]</sup>	0.850 2	0.885 8	0.881 5	0.867 2	0.880 8	0.917 9	0.867 2	0.891 3	0.872 9	0.893 9	0.894 0	0.954 3
EDDFN <sup>[7]</sup>	0.818 6	0.913 7	0.867 6	0.878 6	0.847 8	0.937 9	0.863 6	0.883 2	0.868 9	0.891 9	0.891 9	0.952 8
MDFEND <sup>[19]</sup>	0.830 1	0.938 9	0.891 7	0.900 3	0.886 5	0.940 0	0.895 1	0.906 6	0.898 0	0.913 7	0.913 8	0.970 8
MFDP	0.857 9	0.913 6	<b>0.929 3</b>	<b>0.900 5</b>	<b>0.917 2</b>	<b>0.945 0</b>	<b>0.923 9</b>	<b>0.927 3</b>	<b>0.916 8</b>	<b>0.923 7</b>	<b>0.923 7</b>	<b>0.971 9</b>

(1)EANN<sup>[20]</sup>。是一个多模态的事件对抗神经网络的端到端框架,由3个组件组成:特征提取器、事件过滤器和假新闻检测器。该模型旨在学习事件不变表示,以便能够更好地应对新出现的事件。

(2)MMoE<sup>[21]</sup>。是一种对多任务学习架构的扩展,通过学习不同任务的联系和差异,提高每个任务的学习效率和质量,具体使用 gate 网络来学习融合多位专家对不同任务的权重影响,从而缓解多目标任务相关性低导致的准确率问题。

(3)MoSE<sup>[22]</sup>。是在多任务场景下,显示建模时间维度的信息,对用户的未来行为做出更加准确的预测。与MMoE<sup>[21]</sup>不同的是,此模型使用 LSTM 作为专家网络,捕获各个领域的特征。

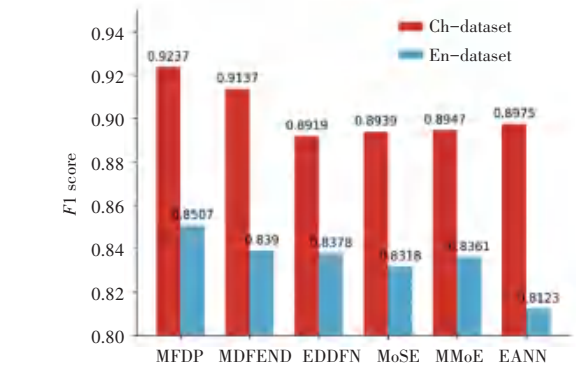


图3 模型对比图

Fig. 3 Model comparison diagram

(4)EDDFN<sup>[7]</sup>。是多领域假新闻检测模型,保留了特定领域和领域共享的知识。所有域私有和域共享的模块都是 MLP。集中的域私有和域共享特征被馈送到分类器中以获得最终预测。

(5)MDFEND<sup>[19]</sup>。公开了中文多领域虚假新闻检测数据集 Weibo21,在每个领域中设置一个专家网络 TextCNN 进行领域特征的抽取,并采用门控机制来聚合不同“专家”网络获取的特征表示。

在与5个多领域假新闻检测模型对比中,在中文数据集上的预测结果的 F1 分数,虽然在科技领域和军事领域中的效果不如 EANN<sup>[20]</sup> 和 MMoE<sup>[21]</sup> 模型,但在其他7个领域上的预测效果优于其他基线模型,这表明在假新闻检测任务中,考虑领域特征

信息是有效的。

在英文数据集中,  $F1$  分数较其他基线模型相比均有提升,说明保留特定领域特征再回传的方式有助于提升模型的检测性能。从2个数据集上的检测效果来看,通过中英文两个数据集的联合训练,能够发挥模型挖掘多源特征的优势,学习多语言间新闻的相关性,能够促进知识迁移,从而提高假新闻检测的性能。

### 3 结束语

本文模型 MFDP 的重点是解决多领域中的数据不平衡性的问题,能够充分利用真假新闻中的领域特征,对假新闻进行检测,通过在真实数据集上进行的对比实验所得数据,表明 MFDP 在二分类模型精度指标  $F1$  上均优于基线方法,验证了模型的有效性。

未来会考虑真假新闻分类的细粒度问题,这有助于更精确地揭示假新闻背后的策略,提高假新闻的可解释性。

### 参考文献

- [1] MA Jing, GAO Wei, WONG K F. Detect rumors in microblog posts using propagation structure via Kernel learning [C]// Proceedings of the 55<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). ACL, 2017: 708-717.
- [2] DAVOUDI M, MOOSAVI M R, SADREDDINI M H. DSS: A hybrid deep model for fake news detection using propagation tree and stance network[J]. Expert Systems with Applications, 2022, 198: 116635.
- [3] SHENG Qiang, CAO Juan, BERNARD H R, et al. Characterizing multi-domain false news and underlying user effects on Chinese Weibo [J]. Information Processing & Management, 2022, 59 (4): 102959.
- [4] ZHU Yongchun, SHENG Qiang, CAO Juan, et al. Memory-guided multi-view multi-domain fake news detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(7): 7178-7191.
- [5] CHEN Yixuan, SUI Jie, HU Liang, et al. Attention-residual network with CNN for rumor detection [C]//Proceedings of the 28<sup>th</sup> ACM International Conference on Information and Knowledge Management. New York: ACM, 2019: 1121-1130.
- [6] KALIYAR R K, GOSWAMI A, NARANG P, et al. FNDNet-A deep convolutional neural network for fake news detection [J]. Cognitive Systems Research, 2020, 61: 32-44.
- [7] SILVA A, LUO Ling, KARUNASEKERA S, et al. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(1): 557-565.
- [8] YUE Zhenrui, ZENG Huimin, KOU Ziyi, et al. Contrastive domain adaptation for early misinformation detection: A case study on COVID-19 [J]. arXiv preprint arXiv, 2208.09578, 2022.
- [9] 王振琦, 陈涛, 张宝宇, 等. 基于交叉特征感知融合的多领域虚假新闻检测 [J]. 计算机系统应用, 2024, 33 (3): 264-272.
- [10] LUO Yongcong, MA Jing, YEO C K. BCMM: A novel post-based augmentation representation for early rumour detection on social media [J]. Pattern Recognition, 2021, 113: 107818.
- [11] ZHANG Huaiwen, QIAN Shengsheng, FANG Quan, et al. Multimodal disentangled domain adaption for social media event rumor detection [J]. IEEE Transactions on Multimedia, 2020, 23: 4441 - 4454.
- [12] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Long and Short Papers). ACL, 2019: 4171-4186.
- [13] KIM Y. Convolutional neural networks for sentence classification [J]. arXiv preprint arXiv, 1408.5882, 2014.
- [14] ZHANG Xueyao, CAO Juan, LI Xirong, et al. Mining dual emotion for fake news detection [C]//Proceedings of the Web Conference 2021. New York: ACM, 2021: 3465-3476.
- [15] YANG Yuting, CAO Juan, LU Mingyan, et al. How to write high-quality news on social network? predicting news quality by mining writing style [J]. arXiv preprint arXiv, 1902.00750, 2021.
- [16] LIN Hongzhan, MA Jing, CHEN Liangliang, et al. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning [J]. arXiv preprint arXiv, 2204.08143, 2022.
- [17] 杨书新, 丁祺伟. 基于局部和全局特征聚合的虚假新闻检测方法 [J]. 计算机工程与应用, 2025, 61(9): 139-147.
- [18] WU Danke, TAN Zhenhua, ZHANG Haoran, et al. LIMFA: Label-irrelevant multi-domain feature alignment-based fake news detection for unseen domain [J]. Neural Computing and Applications, 2024, 36(10): 5197-5215.
- [19] NAN Qiong, CAO Juan, ZHU Yongchun, et al. MDFEND: Multi-domain fake news detection [C]//Proceedings of the 30<sup>th</sup> ACM International Conference on Information & Knowledge Management. New York: ACM, 2021: 3343-3347.
- [20] WANG Yaqing, MA Fenglong, JIN Zhiwei, et al. EANN: Event adversarial neural networks for multi-modal fake news detection [C]//Proceedings of the 24<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 849-857.
- [21] MA Jiaqi, ZHAO Zhe, YI Xinyang, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts [C]//Proceedings of the 24<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1930-1939.
- [22] QIN Zhen, CHENG Yicheng, ZHAO Zhe, et al. Multitask mixture of sequential experts for user activity streams [C]// Proceedings of the 26<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'20). New York: ACM, 2020: 3083-3091.