

用于中英文讽刺检测的动态交互双通道图注意力网络

谭萍萍, 徐计, 李逸骏, 汪海

引用本文

谭萍萍, 徐计, 李逸骏, 汪海. 用于中英文讽刺检测的动态交互双通道图注意力网络[J]. 计算机科学, 2026, 53(2): 300-311.

TAN Pingping, XU Ji, LI Yijun, WANG Hai. [Dynamic Interaction Dual-channel Graph Attention Network for Chinese and English Sarcasm Detection](#) [J]. Computer Science, 2026, 53(2): 300-311.

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)

[基于双层注意力网络的强化学习方法求解柔性作业车间调度问题](#)

Reinforcement Learning Method for Solving Flexible Job Shop Scheduling Problem Based on Double Layer Attention Network

计算机科学, 2026, 53(1): 231-240. <https://doi.org/10.11896/jsjcx.250100088>

[跨模态不一致感知下双视角交互融合的多模态情感分析](#)

Multimodal Sentiment Analysis for Interactive Fusion of Dual Perspectives Under Cross-modal Inconsistent Perception

计算机科学, 2026, 53(1): 187-194. <https://doi.org/10.11896/jsjcx.241100029>

[结合超图学习的多注意力机制新闻推荐方法](#)

Multiple Attention Mechanism News Recommendation Approach with Hypergraph Learning

计算机科学, 2025, 52(11A): 250200067-7. <https://doi.org/10.11896/jsjcx.250200067>

[基于主体注意力与多空间域信息协同的多模态情感分析](#)

Multimodal Sentiment Analysis Based on Dominant Attention and Multi-space Domain Information Collaboration

计算机科学, 2025, 52(11A): 250200022-9. <https://doi.org/10.11896/jsjcx.250200022>

[利用融合2-opt的强化学习算法求解TSP问题](#)

Hybrid Reinforcement Learning Algorithm Combined with 2-opt for Solving Traveling Salesman Problem

计算机科学, 2025, 52(11A): 250200121-8. <https://doi.org/10.11896/jsjcx.250200121>

用于中英文讽刺检测的动态交互双通道图注意力网络

谭萍萍^{1,2} 徐计² 李逸骏³ 汪海³

1 贵州大学计算机科学与技术学院 贵阳 550025

2 省部共建公共大数据国家重点实验室(贵州大学) 贵阳 550025

3 贵州白山云科技股份有限公司 贵阳 550081

(gs.pptan22@std.uestc.edu.cn)

摘要 汉语语义的复杂性和情感的细腻表达,导致中文文本讽刺检测具有挑战性。现有的讽刺检测方法多基于英文开发,难以适应中文的独特表达方式和文化内涵。因此,提出了一种新颖的动态交互双通道图注意力网络(Dynamic Interaction Dual-channel Graph Attention Network, DiDu-GAT),利用独特的双通道结构来分析文本中的句法依赖关系与情感特征。DiDu-GAT设计动态交互机制增强其跨通道学习能力,全面提取情感信息和句法信息,从而显著提升了中文讽刺检测的准确率。在哈工大中文讽刺数据集(GuanSarcasm)和两个公开英文讽刺数据集(IAC-V1和IAC-V2)上的实验结果表明,所提方法在主要性能指标上均显著优于现有基线方法,其中英文讽刺检测任务中的有效性和优越性得到验证。

关键词: 图注意力网络; 中文讽刺检测; 双通道动态交互; 情感分析; 讽刺文本识别

中图分类号 TP311.13; TP3.9

Dynamic Interaction Dual-channel Graph Attention Network for Chinese and English Sarcasm Detection

TAN Pingping^{1,2}, XU Ji², LI Yijun³ and WANG Hai³

1 College of Computer Science and Technology, Guizhou University, Guiyang 550025, China

2 State Key Laboratory of Public Big Data (Guizhou University), Guiyang 550025, China

3 Baishan Cloud Technology Co., Ltd., Guiyang 550081, China

Abstract Due to the complexity of Chinese semantics and the nuanced expression of emotions, Chinese text sarcasm detection presents a challenging task. Existing sarcasm detection methods are predominantly developed for English and struggle to adapt to the unique expressions and cultural connotations of Chinese. Therefore, this paper proposes a novel dynamic interaction dual-channel graph attention network (DiDu-GAT), which utilizes a unique dual-channel structure to analyze syntactic dependencies and emotional features in texts. DiDu-GAT incorporates a dynamic interaction mechanism to enhance its cross-channel learning capabilities, enabling comprehensive extraction of emotional information and syntactic patterns, thereby significantly improving the accuracy of Chinese sarcasm detection. Experimental results on the HIT Chinese sarcasm dataset (GuanSarcasm) and two public English sarcasm datasets (IAC-V1 and IAC-V2) demonstrate that the proposed method significantly outperforms existing baseline methods across key performance metrics, validating its effectiveness and superiority in both Chinese and English sarcasm detection tasks.

Keywords Graph attention network, Chinese sarcasm detection, Dual-channel dynamic interaction, Sentiment analysis, Sarcasm text recognition

1 引言

讽刺是一种通过夸张和反讽挑战传统观念的文学和修辞手法^[1],广泛存在于经典文学、社交媒体和政治漫画中。讽刺

检测极具挑战性,既须解析细腻的文本表达,又要理解文化背景^[2]。尽管讽刺检测在自然语言处理领域已有广泛研究,但是针对中文的研究仍然非常有限,且现有模型通常将情感分析和句法分析分开处理^[3],未能充分整合语义信息,从而影响

到稿日期:2025-05-07 返修日期:2025-07-04

基金项目:国家自然科学基金(62366008,61966005)

This work was supported by the National Natural Science Foundation of China(62366008,61966005).

通信作者:徐计(jixu@gzu.edu.cn)

了检测的准确性。

相比于英文,中文讽刺检测更具挑战性^[4]。首先,中文是一种意合型语言,其讽刺表达往往依赖深层语境和文化背景,不具备英文的形合特性。其次,中文讽刺通常采用更含蓄的表达方式,使用委婉语、双关语和典故等修辞手法,这与英文中更直接的讽刺表达存在显著差异。此外,中文的句法结构复杂多变,词汇的多义性更为普遍,这使得中文讽刺表达形式更具有多样性。

现有讽刺检测方法主要面向英文语料设计,典型模型如基于自匹配机制的 SMSD^[5] (Self-matching Sarcasm Detection)和采用情感依赖图卷积的 ADGCN^[3] (Affective Dependency Graph Convolutional Network),在英文语境下表现出色。然而,由于中英文在语言结构、文化背景和讽刺表达方式上存在显著差异,将这些针对英文设计的模型直接迁移到中文讽刺检测任务时,其性能会出现显著下降。

目前,中文讽刺检测领域的相关研究工作十分有限。近几年较为完善的方法是 Wen 等^[6]提出的 SAAG(Sememe and Auxiliary Enhanced Attention Neural Model for Sarcasm Detection)。该方法首次将义原知识^[7]引入中文讽刺检测任务中,通过义原增强中文词义表示,并利用辅助信息理解中文讽刺表达的语境,从而改善中文讽刺的检测效果。然而,尽管该研究推动了中文讽刺检测的发展,该领域仍然存在研究数量不足等突出问题,亟需更多深入、系统的探索。

在讽刺检测中,句法分析^[8]起着至关重要的作用。例如如图 1 中的示例句子“你真的太棒了,连这么简单的事情都做不好”。表面上,这句话似乎是在称赞,但实际上它通过句法结构的对比和转折表达了讽刺;特别是使用了“连…都…”这种表达方式,强化了句子内部语义的反转,使讽刺意图更为显著。因此,一个有效的讽刺检测模型必须能够识别这种句法特征,以便准确地判断出文本中隐含的讽刺意图。

此外,句子中的情感分析也同样关键。讽刺效果通常依赖于情感表达的并列和反转来实现^[9]。在讽刺文本中,情感的明显错配很常见。例如,在图 1 的句子中,表面的情感表达与真实情感形成了强烈对比。仅分析句子的开头部分可能表示正面情感,但对整句进行情感分析后发现情感从正面转为负面,这正是讽刺意图的典型表现。因此,识别文本中情感的动态变化对于准确检测讽刺至关重要。

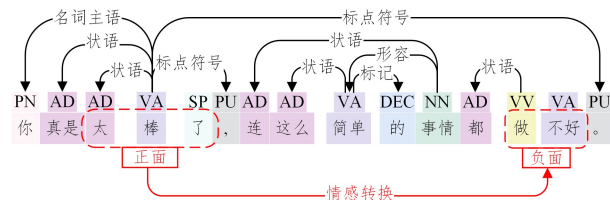


图 1 句法依赖树

Fig. 1 Syntactic dependency tree

针对以上问题,本文提出了一种动态交互双通道图注意力网络(DiDu-GAT)。与依赖静态义原词库的 SAAG 模型不同,DiDu-GAT 通过动态交互机制增强讽刺检测能力。该模

型使情感和句法处理通道之间能够实时交互,显著增强了特征的提取,并增强了对中英文文本中讽刺意图的理解。通过在每个通道内动态调整信息流,模型能够灵活地适应情感和句法的不断变化,这对于分析社交媒体中快速变化且复杂的讽刺内容尤为关键。动态交互的双通道协同工作使模型能更精确地捕捉讽刺表达中的细微情感线索和句法关系,从而显著提升了讽刺检测的精度。

为了更全面地评估 DiDu-GAT 模型的讽刺检测性能,选择了 SMSD 和 ADGCN 这两个基线模型进行特征可视化分析,如图 2 所示。通过利用 t-SNE^[10]技术,从每个数据集中分别随机选取了 300 条测试样本,并将这些模型提取的高维特征降维至二维空间,直观展示了不同模型间区分讽刺与非讽刺样本的能力。这些模型在处理包含复杂句法结构和细微情感变化的讽刺文本时,展现出了差异化的性能表现。

具体而言,SMSD 的特征分布较为混乱,讽刺与非讽刺样本的边界模糊,类别间存在较大交叠,说明其在捕捉复杂句法和情感信息方面存在明显不足。ADGCN 的特征分布较 SMSD 有所改善,类别之间的边界稍显清晰,但仍无法准确划分讽刺与非讽刺样本类别。而 DiDu-GAT 的特征分布表现尤为突出,其类别区分度最高,讽刺与非讽刺样本的边界最为清晰,类间混淆情况显著减少。这一趋势表明,DiDu-GAT 模型在捕捉文本的深层特征方面更具优势。

以具体例子佐证:标题“欧盟史上最严数据保护条例生效影响全球在欧有业务企业”的评论“干得漂亮! 缺钱的时候随便逮一个就能罚款,缺筹码的时候可以随时拉一个出来交易……”中,表面赞美“干得漂亮!”与后文对政策潜在滥用的批评形成鲜明对比。SMSD 和 ADGCN 未能有效捕捉到这种表层称赞与深层批评之间的语义冲突,反映出有模型在处理中文复杂讽刺表达时的根本性不足。另一个例子:标题“某公众人物因争议性行为引发热议”下的评论“感谢某人,让我看到了明星是个啥样子”,表面通过“感谢某人”表达感激,实则通过“让我看到了明星是个啥样子”传达对其行为的批评,SMSD 和 ADGCN 都未能识别其中的讽刺。这表明,现有模型在处理上下文复杂语义关系,捕捉情感反转和句法特征交互方面,仍存在显著不足。

综上所述,本文主要贡献包括 3 个方面。

(1)提出了一种基于情感-句法的双通道动态自适应交互方法,利用情感通道与句法通道的协同学习,挖掘两者之间的深层依赖关系,从而更精准地理解讽刺表达中的隐含语义与情感冲突。

(2)设计了一种适配于情感-句法双通道结构的图注意力网络(DiDu-GAT),包含新型跨通道交互注意力层,以动态调节情感与句法的融合权重,显著提升了讽刺检测的性能表现。

(3)在哈工大中文讽刺数据集 GuanSarcasm 和两个公开英文基准上进行了训练和测试。实验结果表明,DiDu-GAT 模型在主要性能指标上显著优于现有基线方法,其中英文讽刺检测任务中的有效性和优越性得到验证。本文工作填补了中文讽刺检测领域中的不足。

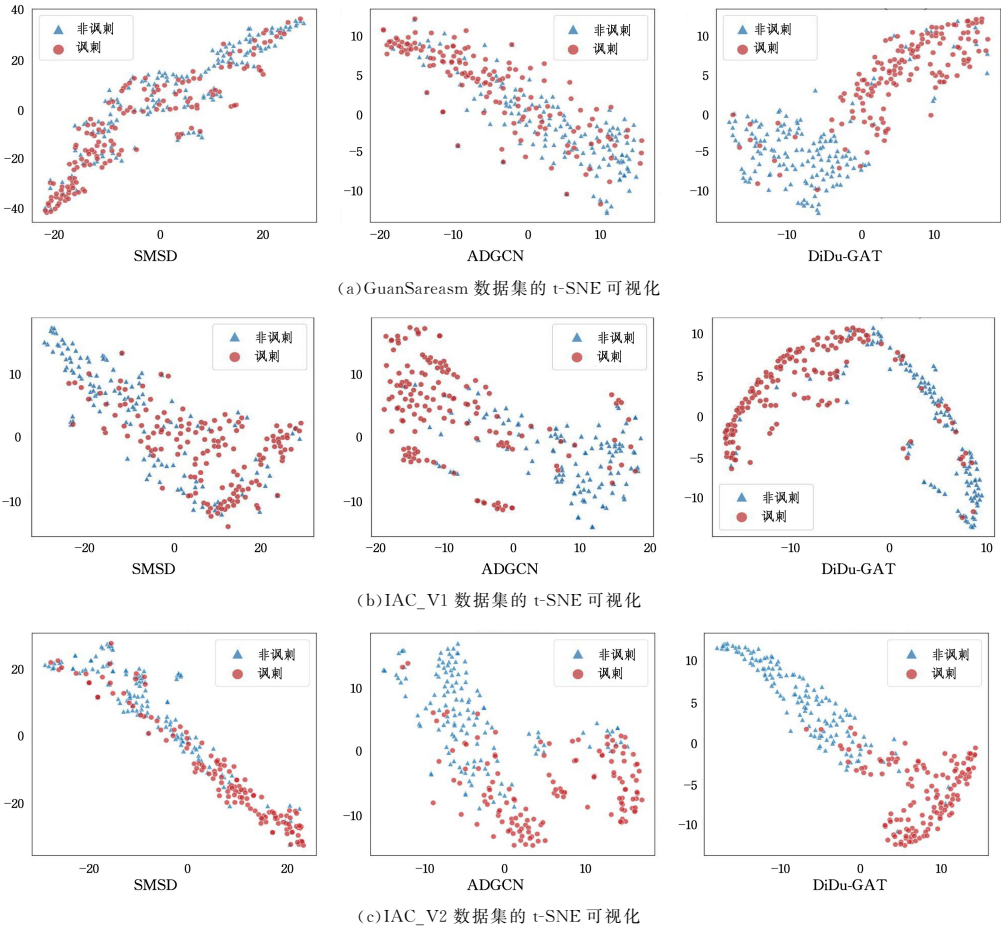


图 2 正负样本分布的 t-SNE 可视化

Fig. 2 t-SNE visualization of positive and negative sample distribution

2 相关工作

讽刺检测是自然语言处理中的关键任务,对于内容审核和公众情绪分析等应用至关重要^[11]。本研究专门针对中英文文本的讽刺检测,旨在应对这一领域中独特的语言和文化复杂性。近年来,深度学习虽然大幅提高了讽刺检测的准确性,但在中文场景中^[12]的应用仍存在挑战,特别是对中文讽刺微妙之处的识别。

2.1 基于语言结构的方法

在讽刺检测中,语言结构的分析通过揭示表面语义与深层语义之间的矛盾^[13],为识别讽刺文本提供了重要依据。Joshi 等^[14]研究了语义的显性和隐性不一致性,通过分析表面语义与深层语义的冲突,提出通过建模语义前后不一致性来识别讽刺文本。针对社交媒体中语法不准和表达多变的问题,Ghosh 等^[15]提出了一种结合 CNN^[16],LSTM^[17]和 DNN^[18]的神经网络语义模型。为提高讽刺识别的准确性,Zhou 等^[19]提出了基于注意力机制^[20]的双向长短期记忆网络(Attention Bidirectional Long Short-Term Memory, Att-BiLSTM),其通过捕捉上下文信息和自适应关注关键语义,显著提高了识别精度。此外,社交对话中的讽刺和恶意语言检测面临语境多样性与对话特征稀疏的挑战。Lukin 等^[21]提出基于自举法^[22]训练分类器,通过结合语法模式和语义特征,其有效解决了多种主观语言任务的分类难题。Xiong 等^[5]提出

的自匹配网络与低秩双线性池化相结合的方法,能有效捕捉句子内部的语义冲突信息。Mandal 等^[23]使用结合 LSTM 和 CNN 的混合神经网络,同时捕捉文本的局部细节和全局语义。针对反讽识别中的潜在语义特征提取问题,Fan 等^[24]提出了基于多语义融合的模型,弥补了传统方法中单一语义信息的不足。针对多语言讽刺检测中的语境适配问题,Krishnan 等^[25]通过 BERT^[26](Bidirectional Encoder Representations from Transformers)捕捉语义上下文信息,并结合高斯模型建模语义不一致性,有效识别了反讽类型。由于反讽文本中字面意义与真实意图之间存在矛盾,Liu 等^[27]提出了一种基于提示调优技术^[28]的方法,以优化模型对语义不一致的理解。针对讽刺表达的复杂性以及孤立句子分析难以全面理解说话者的讽刺意图的问题,Helal 等^[29]提出了一种基于上下文的讽刺检测方法,其通过微调预训练的 Transformer 模型(RoBERTa 和 DistillBERT),解决了在不同情境下的讽刺检测问题。

综上所述,基于语言结构的方法虽然取得了一定进展,但在处理中文特有的语法结构和隐含表达时仍存在局限。特别是,这些方法大多基于英语语料开发,未能充分考虑中文语言的复杂性和含蓄性,导致在中文讽刺检测任务中的适应性不足。

2.2 基于情感分析的方法

讽刺文本常表现为情感极性的冲突,表面情绪与隐含情

绪的不一致性为基于情感分析的讽刺检测提供了关键线索^[30]。Tay 等^[31]在分析推文上下文时引入了用户情感状态信息,解决了讽刺文本常见的语义对比不一致性问题。Zhang 等^[32]进一步提出一种基于情感知识的迁移学习方法,以有效捕捉文本中的隐含不一致性。为了帮助识别讽刺文本中的实际情绪,Bababnejad 等^[33]通过结合情感与上下文特征来扩展 BERT 的架构。Chia 等^[34]针对讽刺和反讽分类方法,融合情感特征(如情感极性、情感强度)和语言特征,改善了网络欺凌检测的效果。为了解决长距离不一致模式的捕捉问题,Lou 等^[3]提出了情感依赖图卷积网络(Affective Dependency Graph Convolutional Network, ADGCN)框架,其结合情感常识和句法信息,建模情感图和依赖图,提升了讽刺检测性能。为了增强情感分类与讽刺检测的能力,Chen 等^[35]开发了一种多任务学习框架,其以软情绪标签纳入情绪线索,有助于减少假阳性问题。同时,Liu 等^[36]提出了双通道网络(Dual-Channel Network, DC-Net),通过分别建模字面情绪和隐含情绪,增强对情感冲突的识别能力。Vitman 等^[37]结合预训练模型与卷积神经网络,帮助模型提取不同的情感与情绪特征。为了能分别捕获文本的语义相关性和情感信息,Huang 等^[38]提出了一种双通道图卷积神经网络(Dual-Channel Graph Convolutional Networks, DC-GCN),提升了方面级情感分析的效果。

总体而言,基于情感分析的讽刺检测方法为识别文本中的情感冲突提供了重要途径,但在处理中文情感表达的细微变化和隐含转折时,仍需更加精细的建模方法和融合策略。

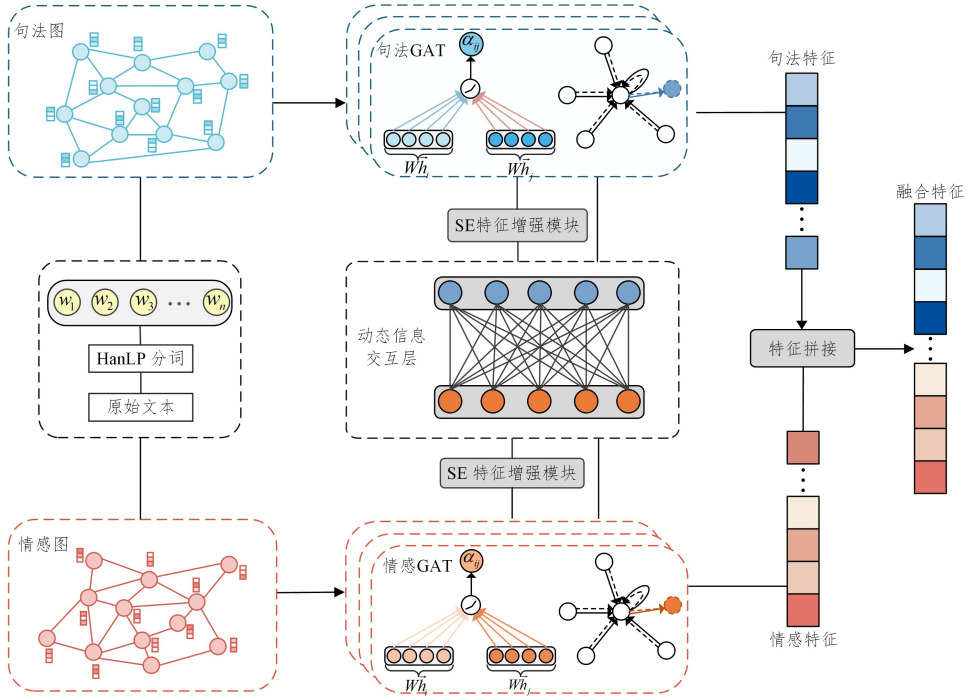


图 3 DiDu-GAT 模型的整体框架

Fig. 3 Overall architecture of DiDu-GAT model

3.1 句法图和情感图构建

(1)句法图:假设输入 $S = \{\tau_1, \tau_2, \dots, \tau_N\}$, τ_i 是句子中的一个词语。对于中文文本,使用 HanLP 工具^[44]对句子 S 进行句法分析,构建句法依赖图 $G = (V, E)$ 。其中, V 是节点集,对应句子中的词汇; E 是有向边集,表示词汇之间的依赖

现有方法在情感与句法特征的协同分析方面存在不足,无法全面捕捉中文讽刺表达的复杂性。

2.3 基于外部知识库的方法

讽刺表达通常依赖外部背景知识,借助外部知识资源(如知识图谱^[39]和结构化知识),增强模型对讽刺语境的理解能力。Liang 等^[40]针对主题讽刺表达,提出了一种结合提示学习^[41]和 BERT 的模型,通过提示学习方法进一步提升了讽刺检测效果。针对中文语义和上下文理解的挑战,Wen 等^[6]引入外部语义知识,利用动态文本表示技术改善中文词汇的表示能力,缓解了中文讽刺检测的困境。考虑到社交媒体讽刺文本中讽刺表达的多样性和复杂性,Min 等^[42]通过引入常识推理工具 COMET^[43]对文本进行增强,推测潜在结果和人类反应,从而优化反讽检测模型。

基于外部知识库的方法尽管丰富了讽刺检测的语义理解,但仍面临知识库覆盖范围有限、更新滞后等问题,特别是对于快速变化的社交媒体语言。此外,如何有效整合外部知识与文本内部特征,构建动态适应的讽刺检测模型,仍是该领域的关键挑战。

3 动态交互双通道 GAT

本章提出了一种动态交互双通道图注意力网络,同时提取文本中的句法结构和情感信息,并设计动态交互机制实现两种信息的深度融合,达成句法与情感的协同优化。图 3 展示了 DiDu-GAT 模型的整体框架。

关系。对于图 G ,其邻接矩阵 A 根据 E 计算出,如果节点 i 指向节点 j ,则 $A_{ij} = 1$,否则 $A_{ij} = 0$ 。

对于英文文本,则采用 Stanza 工具^[45]进行句法分析,以适应英文语料的特点。无论中文还是英文,均会构建词汇之间的句法关系和语义联系,为后续建模提供句法结构信息。

(2)情感图:结合词汇间的动态边权重分布和边类型,增强模型对复杂情感信息的识别能力。情感图中的每一个节点对应文本中的一个词汇,其节点属性包括该词的情感分数和上下文嵌入特征。通过将词汇作为节点,情感图能够直观地刻画文本中各词之间的情感关系。

动态边权重的计算,基于连接节点的情感得分差异。具体而言,给定两个节点 A 与 B 的情感分数分别为 $score(A)$ 和 $score(B)$,其节点间的情感差为 $|score(A) - score(B)|$,该差异反映了两个词汇之间情感强度的对比。

对于权重 $weight$ 的分配,首先计算出两个词汇间的基本权重 w_{base} ,其基于情感分数的绝对差。

$$w_{base} = \alpha \times |score(A) - score(B)| \quad (1)$$

其中, α 是一个调整系数,取值范围为 $[0, 1, 1, 0]$,用于控制情感强度差异对权重的影响程度。较小的 α 值会减弱情感差异的影响,而较大的 α 值则会放大这种影响。

接下来,通过情感极性对比获取 $w_{polarity}$,并调整边权重。如果两个节点 A 与 B 的情感极性相同(即 $sign(score(A)) = sign(score(B))$),那么 $w_{polarity} = 0$;否则(即 $sign(score(A)) \neq sign(score(B))$),通过计算情感分数差异和语义相似度,动态调整边权重,此时权重 $w_{polarity}$ 根据情感极性差异和语义相似度共同决定。其中, $sign(x)$ 为符号函数,用于提取数值的正负性:当 $x > 0$ 时, $sign(x) = 1$;当 $x = 0$ 时, $sign(x) = 0$;当 $x < 0$ 时, $sign(x) = -1$ 。该函数用于判断节点情感极性的一致性,从而决定是否需要额外的权重调整。

最终情感权重表示为:

$$w_{polarity} = \begin{cases} \beta, & \text{if } sign(score(A)) \neq sign(score(B)) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

其中, β 是一个动态系数,其值根据词汇间的情感极性差异和语义关系实时计算得出,而非固定常数。当两个节点情感极性相反时, β 会根据它们的语义相似度和情感强度差异进行自适应调整:语义相似度越高且情感差异越大, β 值越大,从而在边权重中赋予更高的重要性;反之,若语义相似度低或情感差异小, β 值则相应减小。这种动态调整机制,使模型能够精确捕捉文本中情感转折引起的讽刺表达。

综上所述,最后得到一个综合情感权重值 $weight$:

$$weight = w_{base} + w_{polarity} \quad (3)$$

这样的权重不仅反映了词汇间情感强度的差异,还考虑了词汇的上下文关系。

针对文本的特征表示,采用 BERT 预训练模型对文本进行编码,从而获取每个词汇的特征向量。这些向量能够捕捉上下文中的语义信息,形成高质量的词语嵌入表示。最终,词汇的特征矩阵 M_i 由其对应的特征向量 \vec{v}_{ij} 表示为:

$$M_i = [\vec{v}_{i1}, \vec{v}_{i2}, \dots, \vec{v}_{im}] \quad (4)$$

其中, m ($m = 768$) 是 BERT 输出的向量维度。此特征矩阵既保留了局部语义信息,又为情感图和句法图学习提供了丰富的上下文表示。

3.2 双通道 GAT

双通道 GAT(Graph Attention Network)^[46] 是本文方法的核心模块,用于在文本图结构上学习并提取节点间的复杂语义和情感关系。构建情感图和句法图,并将其分别输入两

个独立通道中,每个通道分别专注于情感和句法信息学习,并提取对应的特征表示。每个通道会先通过全连接层对节点特征进行线性变换,以增强模型的表达能力。特征非线性变换函数定义为:

$$h' = Wh + b \quad (5)$$

其中, $W \in \mathbb{R}^{d_{out} \times d_{in}}$ 为权重矩阵, d_{in} 为输入特征的维度, d_{out} 为输出特征的维度; $b \in \mathbb{R}^{d_{out}}$ 为线性变换后的偏置向量。

图注意力层(Graph Attention Layer)用于计算节点间的注意力系数并根据这些注意力系数来聚合邻居节点的特征。每个节点的特征通过图注意力层进行加权聚合,具体的计算过程如下。

对于线性变换后的特征,图注意力层计算节点间的注意力系数,表示为:

$$e_{ij} = LeakYReLU(a^T [Wh_i \parallel Wh_j]) \quad (6)$$

其中,“ \parallel ”表示连接操作, a 是学习的注意力参数。

注意力系数通过 softmax 函数进行归一化,得到最终的注意力权重 α_{ij} :

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})} \quad (7)$$

其中, $N(i)$ 表示节点 i 的邻居节点。

输出特征为:

$$h'_i = \sigma(\sum_{j \in N(i)} \alpha_{ij} Wh_j) \quad (8)$$

其中, σ 是非线性激活函数。

然后, GAT 通过多头注意力机制进一步处理节点特征。多头注意力机制通过多个独立的注意力头对节点特征进行加权聚合,生成每个节点的综合特征表示。每个头的输出特征通过拼接操作合并,再经池化操作生成最终的节点特征表示。

拼接操作表示为:

$$H_{concat} = \text{concat}(head_1, head_2, \dots, head_k) \quad (9)$$

平均池化操作表示为:

$$AvgPool(H) = \frac{1}{N} \sum_{i=1}^N h_i \quad (10)$$

其中, N 表示图中的节点总数。

通过这种方式,双通道 GAT 能够有效处理图结构数据,捕捉节点间复杂的句法关系和情感关系,并提取用于文本分类的深层特征表示。这对于理解和分析结构化的文本数据,特别是在捕捉细微的讽刺语义方面,具有重要作用。

3.3 基于 SE 的特征增强模块

SE(Squeeze-and-Excitation)是一种轻量级的通道权重自适应模块^[47],能够通过学习引导通道的权重分布,对每个特征通道的重要性进行重新标定。

在双通道 GAT 架构中,为了强化通道特征的表达能力,在进入交互学习机制之前,基于 SE 设计了一个通道特征增强模块,用于对情感通道和句法通道的输出特征进行优化。

通道特征增强模块接收来自情感通道输出的特征 h^{sent} 和来自句法通道输出的 h^{syntax} 特征,并对每个通道的特征重要性进行建模和重标定,增强其在后续交互学习中的作用。

对于每个通道的特征输出 $h \in \mathbb{R}^{N \times d}$ (其中 N 为图中节点数, d 为特征维度),首先对其进行全局平均池化,在节点维度上进行统计压缩,生成每个通道的全局描述向量。情感通道

输出的特征 \mathbf{h}^{sent} 的全局描述向量 \mathbf{s}^{sent} 为:

$$\mathbf{s}_k^{\text{sent}} = \frac{1}{N} \sum_{i=1}^N \mathbf{h}_{i,k}^{\text{sent}} \quad (11)$$

句法通道输出特征 $\mathbf{h}^{\text{syntax}}$ 的全局描述向量 $\mathbf{s}^{\text{syntax}}$ 表示为:

$$\mathbf{s}_k^{\text{syntax}} = \frac{1}{N} \sum_{i=1}^N \mathbf{h}_{i,k}^{\text{syntax}} \quad (12)$$

其中, $\mathbf{s}_k \in \mathbb{R}^d$ 表示通道的全局统计信息,用于捕捉每个通道的特征重要性。

对于每个通道的全局描述信息,首先通过全连接层进行非线性映射,学习每个通道特征的重要性权重。

情感通道:

$$\mathbf{z}^{\text{sent}} = \text{ReLU}(\mathbf{W}_1^{\text{sent}} \cdot \mathbf{s}^{\text{sent}} + \mathbf{b}_1^{\text{sent}}) \quad (13)$$

$$\boldsymbol{\sigma}^{\text{sent}} = \text{Sigmoid}(\mathbf{W}_2^{\text{sent}} \cdot \mathbf{z}^{\text{sent}} + \mathbf{b}_2^{\text{sent}}) \quad (14)$$

句法通道:

$$\mathbf{z}^{\text{syntax}} = \text{ReLU}(\mathbf{W}_1^{\text{syntax}} \cdot \mathbf{s}^{\text{syntax}} + \mathbf{b}_1^{\text{syntax}}) \quad (15)$$

$$\boldsymbol{\sigma}^{\text{syntax}} = \text{Sigmoid}(\mathbf{W}_2 \cdot \mathbf{z}^{\text{syntax}} + \mathbf{b}_2^{\text{syntax}}) \quad (16)$$

其中, $\mathbf{W}_1 \in \mathbb{R}^{d_{\text{hidden}} \times d_{\text{in}}}$, $\mathbf{W}_2 \in \mathbb{R}^{d_{\text{out}} \times d_{\text{hidden}}}$ 为全连接层的权重矩阵; $\mathbf{b}_1 \in \mathbb{R}^{d_{\text{hidden}}}$, $\mathbf{b}_2 \in \mathbb{R}^{d_{\text{out}}}$ 为偏置向量。通过 Sigmoid 激活函数,将权重限制在 $[0,1]$ 内,用于表示各特征通道的重要性。

基于学习到的权重 $\boldsymbol{\sigma}$,对输入特征进行重标定,增强重要特征,抑制不相关特征。优化后的特征表示为:

$$\tilde{\mathbf{h}}_{i,k}^{\text{sent}} = \boldsymbol{\sigma}_k^{\text{sent}} \cdot \mathbf{h}_{i,k}^{\text{sent}} \quad (17)$$

$$\tilde{\mathbf{h}}_{i,k}^{\text{syntax}} = \boldsymbol{\sigma}_k^{\text{syntax}} \cdot \mathbf{h}_{i,k}^{\text{syntax}} \quad (18)$$

经过通道特征增强模块处理后,优化后的特征 $\tilde{\mathbf{h}}_{i,k}^{\text{syntax}}$ 和 $\tilde{\mathbf{h}}_{i,k}^{\text{sent}}$ 分别传递给句法通道和情感通道进行动态交互学习。该模块确保每个通道能够更有效地向对方通道提供重要信息,同时避免无关特征的干扰。

3.4 动态信息交互层

情感通道专注于词汇的情感倾向、情感强度等,而句法通道则更关注句子的构造和词汇依存关系等。为了充分利用两个通道的优势并弥补各自的不足,在双通道 GAT 通过各自的 SE 模块增强了特征后,提出了一种动态门控交互方法。该方法允许每个通道在计算注意力系数时,不仅考虑同一通道内的节点信息,还引入另一通道的节点信息作为信息强化的工具。

设定两个通道 SentGAT 和 SynGAT。对于 SentGAT 中的每个节点 i ,在计算注意力系数时考虑 SynGAT 中的节点信息。传统的注意力系数 α_{ij} 通常是基于节点 i 和 j 的特征计算得出的;而针对跨通道的注意力系数 β_{ij} ,被定义为在通道 SentGAT 中的节点 i 的特征和通道 SynGAT 中的节点 j 的特征的函数。

具体地,对于情感通道 SentGAT,其注意力计算式定义为:

$$\beta_{ij}^{\text{sent}} = \text{softmax}(\text{LeakyReLU}(\mathbf{a}_i^{\text{sent}} [\mathbf{W}_A \mathbf{h}_i^{\text{sent}} \parallel \mathbf{W}_B \mathbf{h}_j^{\text{syntax}}])) \quad (19)$$

其中, $\mathbf{h}_i^{\text{sent}}$ 和 $\mathbf{h}_j^{\text{syntax}}$ 分别是情感通道 SentGAT 和句法通道 SynGAT 中节点 i 和 j 经过 SE 增强后的特征, \mathbf{W}_A 是情感分析的变换矩阵, \mathbf{a}_A 是注意力向量。

对于句法通道 SynGAT,其注意力计算式定义为:

$$\beta_{ij}^{\text{syntax}} = \text{softmax}(\text{LeakyReLU}(\mathbf{a}_i^{\text{syntax}} [\mathbf{W}_B \mathbf{h}_j^{\text{syntax}} \parallel \mathbf{W}_A \mathbf{h}_i^{\text{sent}}])) \quad (20)$$

其中, $\mathbf{h}_j^{\text{syntax}}$ 和 $\mathbf{h}_i^{\text{sent}}$ 分别是情感通道 SentGAT 和句法通道 SynGAT 中节点 j 和 i 经过 SE 增强后的特征, \mathbf{W}_B 是句法分析的变换矩阵, \mathbf{a}_B 是注意力向量。

为了增强每个通道对另一个通道的关键信息的参考,同时保持自身特征的独立性,该动态交互方法通过设计门控单元,动态调整信息流动来控制两个通道之间的信息交互。具体表示如下:

$$\mathbf{G}_i^A = \sigma(\mathbf{W}_g^A \mathbf{h}_i^A + \mathbf{b}_g^A) \quad (21)$$

$$\mathbf{G}_j^B = \sigma(\mathbf{W}_g^B \mathbf{h}_j^B + \mathbf{b}_g^B) \quad (22)$$

其中, $\mathbf{G}_i^A \in \mathbb{R}^{d_e}$ 和 $\mathbf{G}_j^B \in \mathbb{R}^{d_e}$ 是门控单元的输出, $\mathbf{h}_i^A \in \mathbb{R}^{d_A}$ 和 $\mathbf{h}_j^B \in \mathbb{R}^{d_B}$ 分别表示两个通道的输入特征, $\mathbf{W}_g^A \in \mathbb{R}^{d_e \times d_A}$ 和 $\mathbf{W}_g^B \in \mathbb{R}^{d_e \times d_B}$ 用于特征变换, $\mathbf{b}_g^A \in \mathbb{R}^{d_e}$ 和 $\mathbf{b}_g^B \in \mathbb{R}^{d_e}$ 是偏置向量。

在更新节点表示时,利用门控单元动态调节来自另一通道的信息权重,以实现跨通道信息的自适应融合对于情感通道 SentGAT 中节点 i 的更新,考虑来自句法通道 SynGAT 的节点 j 的加权特征:

$$\mathbf{h}_i^{A'} = \sigma(\sum_{j \in N_i} \mathbf{G}_j^B \beta_{ij}^{\text{sent}} \mathbf{W}_B \mathbf{h}_j^B) \quad (23)$$

类似地,对于句法通道 SynGAT 中的节点 j ,考虑来自情感通道 SentGAT 的节点 i 的加权特征:

$$\mathbf{h}_j^{B'} = \sigma(\sum_{i \in N_j} \mathbf{G}_i^A \beta_{ij}^{\text{syntax}} \mathbf{W}_A \mathbf{h}_i^A) \quad (24)$$

对于每个文本样本,最终更新后的情感通道 SentGAT 和句法通道 SynGAT 的特征向量分别记为 \mathbf{h}^{sent} 和 $\mathbf{h}^{\text{syntax}}$ 。 $\mathbf{h}^{\text{sent}} \in \mathbb{R}^{d_{\text{sent}}}$ 表示情感通道的输出特征,维度为 d_{sent} ; $\mathbf{h}^{\text{syntax}} \in \mathbb{R}^{d_{\text{syntax}}}$ 表示句法通道的输出特征,维度为 d_{syntax} 。

随后,将这两个特征向量在特征维度上进行拼接:

$$\mathbf{h} = \text{concat}(\mathbf{h}^{\text{sent}}, \mathbf{h}^{\text{syntax}}) \quad (25)$$

拼接后的向量 $\mathbf{h} \in \mathbb{R}^{d_{\text{sent}} + d_{\text{syntax}}}$ 表示情感和句法信息的综合特征。

最后,将拼接后的向量 \mathbf{h} 作为特征输入分类器进行二分类训练。

4 实验与分析

4.1 数据集

(1)哈工大中文讽刺数据集 GuanSarcasm (Guanchazhe Chinses Sarcasm Dataset)^[48]:由哈工大团队于 2020 年创建。该数据集来自中国新闻网站观察者网,涵盖多个新闻领域,包含 2197 条新闻及其 178237 条用户评论,其中 4972 条评论经过专家手工标注。此数据集包括新闻标题、全文和新闻类别等信息。该数据集通过平衡采样策略,为训练讽刺分类器提供了可靠的数据基础,同时评估了多种讽刺分类方法的性能,从而为中文讽刺检测领域提供了重要的基准支持。

(2) IAC (Internet Argument Corpus) 数据集:由 Lukin 等^[21]收集自互联网辩论语料库,用于研究在线社区中的政治辩论。随后他们对该数据集进行了讽刺标注,以便用于讽刺检测的研究。本研究采用了该数据集的两个版本来测试 DiDu-GAT 的性能,分别命名为 IAC-V1 和 IAC-V2。IAC-V1 包含 1996 条样本,IAC-V2 包含 6520 条样本,两个数据集中正负样本的比例均为 1:1。

4.2 实验设置

本研究采用 PyTorch 框架实现双通道动态交互图注意力网络(DiDu-GAT),并在 NVIDIA GeForce RTX 3090 GPU 上进行训练和测试。实验选用的优化器为 Adam,epoch 设置为 50,学习率使用余弦调度器。代码已上传 github¹⁾。

4.3 基线方法

本实验中,选取了以下基线模型作为对比模型。

(1)Att-LSTM^[19]:通过 LSTM 捕获文本的上下文信息,并结合注意力机制聚焦于关键词汇和句式,识别文本中的讽刺意图。

(2)CNN-LSTM-DNN^[15]:CNN 提取文本特征,LSTM 处理序列依赖性,DNN 进行最终分类。

(3)MAIRN^[31]:通过多维度的方式来捕捉每对词之间的多个关联特征,然后对文本进行讽刺分类。

(4)SIARN^[31]:将句子中的每个词转换成词向量后,通过内部注意力机制计算每对词向量之间的相似度得分,从而对整个句子进行讽刺性判断。

(5)SMSD^[5]:通过自匹配网络进行词对词的交互,捕捉句子中的不一致特征传输给分类器。

(6)SMSD-BiLSTM^[5]:利用句子中每对词汇间的相似度寻找情感冲突,然后使用 Bi-LSTM 提取句子的组成信息,最后通过低秩双线性池化(LRBP)组合矛盾信息和组成信息进行分类训练。

(7)ADGCN^[35]:使用 BosonNLP 情感词典^[49]构建情感知识图网络,并挖掘上下文的讽刺表达,最终将隐藏层表示拼接后进行二分类训练。

(8)SAAG^[6]:融合词义素知识与标题辅助信息实现中文讽刺检测的能力,因其设计特性只适合中文语境,故本实验部分仅在 GuanSarcasm 数据集上实现。

4.4 实验结果

为评估本文模型在讽刺识别任务中的有效性和优越性,

分别在哈工大中文讽刺数据集 GuanSarcasm 以及英文讽刺数据集 IAC-V1 和 IAC-V2 上与多个基线模型进行了对比实验。从表 1 和表 2 可以看出,DiDu-GAT 模型在中英文讽刺识别任务的所有评估指标上均显著优于其他基线模型,展现了卓越的性能。

表 1 在 GuanSarcasm 数据集上的实验结果

Table 1 Experimental results on the GuanSarcasm dataset (%)

模型	准确率	精确率	召回率	F1
Att-LSTM ^[19]	70.35	70.41	70.36	70.38
CNN-LSTM-DNN ^[15]	64.52	64.43	64.37	64.40
MAIRN ^[31]	70.01	70.85	70.78	70.81
SIARN ^[31]	70.68	70.58	70.56	70.57
SMSD ^[5]	67.74	67.74	67.60	67.67
SMSD-BiLSTM ^[54]	66.95	66.84	66.77	66.80
ADGCN ^[3]	69.50	67.42	71.05	69.19
SAAG ^[6]	71.95	72.11	72.03	72.07
DiDuGAT	73.69	73.72	73.66	73.69

在 GuanSarcasm 数据集上的实验结果如表 1 所列,DiDu-GAT 在 F1 上达到了 73.69%,显著优于传统模型和最新的基线模型。例如,AttBiLSTM 模型由于仅捕捉序列信息,未能充分利用句法结构和情感特征,F1 得分为 70.38%;CNN-LSTM-DNN 模型虽然结合了多层结构,但对中文语义复杂性的建模能力有限,导致 F1 得分仅为 64.40%。SMSD 模型虽然通过自匹配网络捕捉了句内的冲突信息,但其特征融合是静态的,未能在情感与句法特征之间建立动态交互;而 ADGCN 模型通过构建情感图和依赖图实现了对文本多层特征的捕捉,但其图表示和上下文特征的结合相对独立,导致信息融合深度不足。相比之下,DiDu-GAT 方法中的双通道动态交互机制通过实时共享和整合情感与句法特征,不仅捕捉了中文讽刺文本中的复杂语义关系,还有效建模了句法依赖和情感动态,显著提升了检测性能,模型 F1 指标达到最高 73.69%。

表 2 在 IAC-V1 和 IAC-V2 数据集上的实验结果

Table 2 Experimental results on the IAC-V1 and IAC-V2 datasets (%)

模型	IAC-V1				IAC-V2			
	准确率	精确率	召回率	F1	准确率	精确率	召回率	F1
Att-LSTM ^[19]	71.50	71.55	71.50	71.52	75.23	75.30	75.23	75.26
CNN-LSTM-DNN ^[15]	63.42	63.92	63.42	63.67	73.42	73.62	73.42	73.52
MAIRN ^[31]	64.75	65.72	64.75	65.23	74.36	74.40	74.36	74.38
SIARN ^[31]	71.50	71.58	71.50	71.54	74.95	74.94	74.98	74.96
SMSD ^[5]	61.92	62.13	61.92	62.02	73.08	73.11	73.08	73.09
SMSD-BiLSTM ^[5]	61.17	61.25	61.17	61.21	70.25	70.29	70.25	70.27
ADGCN ^[3]	66.58	66.61	66.58	66.59	75.08	75.12	75.08	75.10
DiDuGAT	73.00	72.85	73.01	72.93	77.51	77.47	77.51	77.49

在 IAC-V1 和 IAC-V2 数据集上的实验结果如表 2 所列,DiDu-GAT 同样表现优越。在 IAC-V1 数据集上,DiDu-GAT 的 F1 得分为 72.93%,显著高于 AttBiLSTM 的 71.52%和 CNN-LSTM-DNN 的 63.67%。在 IAC-V2 数据集上,DiDu-GAT 的 F1 得分达到 77.49%,相比传统模型(如 MAIRN 的

74.38%,SIARN 的 74.96%,以及 SMSD 的 73.09%)均有显著提升。这主要归功于 DiDu-GAT 在情感通道中动态捕捉情感反转特征的能力,以及其句法通道对文本依赖关系的建模能力。例如,IAC 数据集中一些句子通过复杂的句法依赖结构揭示情感转折,这类隐含的情感转折往往难以被传统

¹⁾ https://github.com/Crneyt/chinese_SarcasmDetection

模型捕捉;而 DiDu-GAT 的双通道设计通过动态交互学习,显著提升了对深层特征的理解能力。

综上所述,DiDu-GAT 模型通过动态交互的双通道设计,成功弥补了传统模型在情感与句法特征融合方面的不足,在中文和英文讽刺任务中均展现了卓越性能。特别是在复杂语义和情感反转的文本处理中,DiDu-GAT 的表现尤为突出,其在讽刺检测任务中的有效性和适用性得到验证。

4.5 超参数敏感性分析

本节对 DiDu-GAT 模型的超参数进行了深入分析,以探索学习率、权重衰减、隐藏层大小、注意力头数 nheads、dropout 和 α 超参数对文本讽刺检测性能的影响。选择了两个数据集:GuanSarcasm 中文讽刺数据集和 IAC-V1 英文讽刺数据集。中文讽刺数据集的分析结果如图 4 所示,英文讽刺数据集的分析结果如图 5 所示。

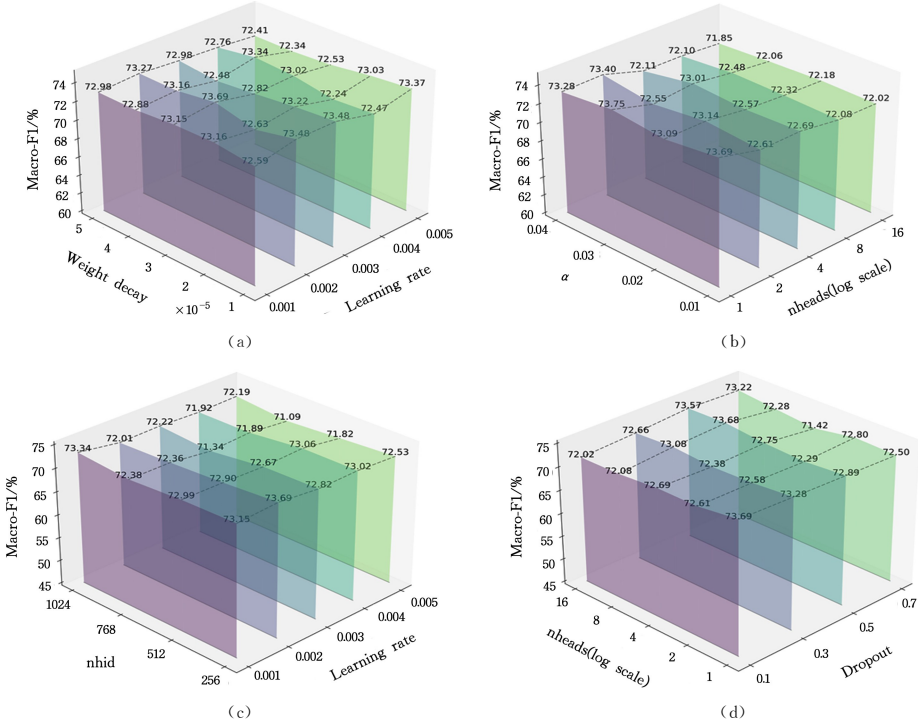


图 4 在 GuanSarcasm 上的超参数敏感性分析

Fig. 4 Hyper-parameter sensitivity analysis on the GuanSarcasm

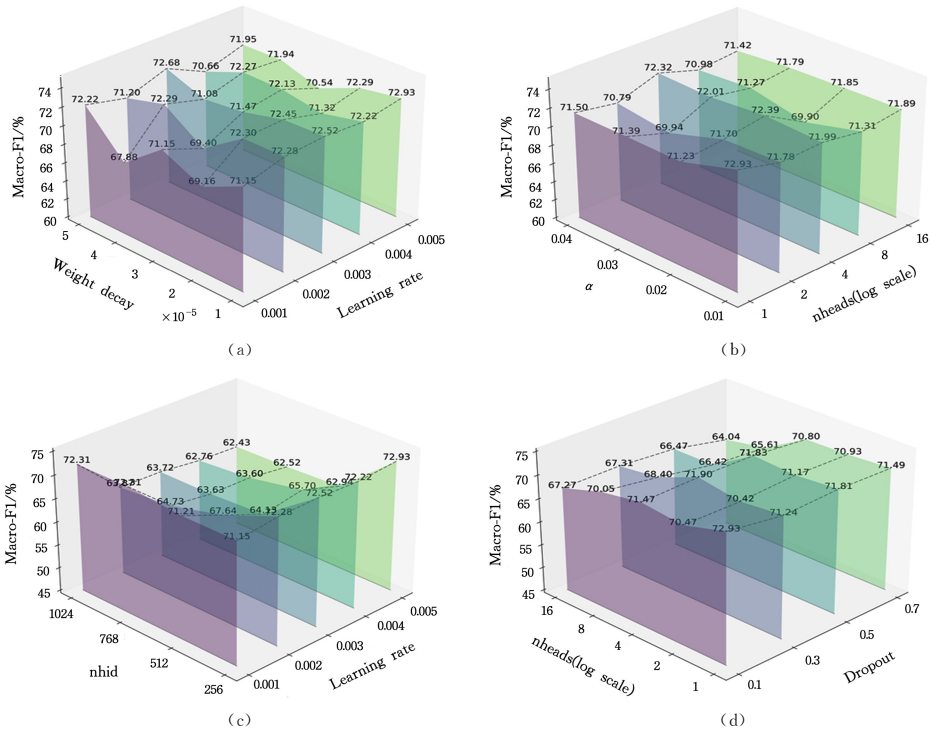


图 5 在 IAC-V1 上的超参数敏感性分析

Fig. 5 Hyper-parameter sensitivity analysis on the IAC-V1

通过对图 4 的分析可知,中文讽刺数据集 GuanSarcasm 上模型性能对超参数变化并不敏感,唯一比较敏感的超参数是 nheads,因为 nheads 直接控制了模型的结构和参数量大小,所以更多地影响了模型的性能。

通过对图 5 的分析可知,英文讽刺数据集 IAC-V1 上超参数的变化对模型性能的影响极其明显,这是因为该数据集在规模上比 GuanSarcasm 更小,数据量更为贫乏,超参数的微小变化可能导致模型学到数据中的噪声或偶然模式,放大超参数的影响。如图 5(d)所示,nheads 注意力头数增大时,模型参数量上升而数据量偏少,导致过拟合,因此模型性能逐渐下降。

通过分析并调整这些参数,模型在捕捉讽刺文本的隐含特征和复杂语义结构方面,性能得到了显著提升,这为模型的进一步优化提供了参考依据。

4.6 消融实验

为全面评估模型各模块对讽刺检测性能的独立贡献,在 GuanSarcasm, IAC-V1 和 IAC-V2 数据集上进行了系统的消融实验。通过逐步移除关键模块,量化分析了本文动态交互方法和双通道架构的有效性。

GuanSarcasm 数据集上的消融实验如表 3 所列。完整的 DiDu-GAT 模型的 F1 达 73.69%;而移除动态交互机制后,性能显著下降至 70.99%(降低 2.7 个百分点)。单独保留情感通道的模型的 F1 为 70.97%,与无交互机制相当。而仅保留句法通道的模型性能较低,F1 仅为 70.16%。这表明在中文讽刺检测中,情感信息略占优势,而动态交互机制能有效整合两种信息,带来了显著的性能提升。特别是对于中文讽刺

表达中常见的情感转折模式,动态交互机制通过协调情感与句法特征,有效捕捉了讽刺的细微表达。值得注意的是,当同时使用两种通道信息却不引入交互机制时,对比单一通道信息模型,性能并没有显著提升,说明两个独立的通道信息在输入模型时未经融合或交互设计而直接并行处理,可能导致特征间的相互抑制和干扰问题,无法带来模型性能的提升。

表 3 DiDuGAT 模型在 GuanSarcasm 上的消融实验结果
Table 3 Ablation study results of the DiDuGAT model on GuanSarcasm

		(%)			
方法	准确率	精确率	召回率	F1	
消融模型	仅情感通道	70.97	71.04	70.90	70.97
	仅句法通道	70.16	70.23	70.09	70.16
	无交互机制	70.99	71.25	70.73	70.99
原始模型	DiDuGAT	73.69	73.72	73.66	73.69

在英文数据集上的实验结果如表 4 所列,动态交互机制同样发挥了关键作用。在 IAC-V1 上,完整模型的 F1 为 72.93%,而无交互机制模型的 F1 为 71.23%,下降了 1.7 个百分点;在 IAC-V2 上,这一差距更为显著,分别为 77.49%和 74.75%,差距达 2.74 个百分点。值得注意的是,在两个英文数据集上,仅包含情感通道均优于仅包含句法通道(IAC-V1 上分别为 70.25%和 70.06%,IAC-V2 上分别为 74.57%和 74.13%)。这一结果表明,英文讽刺表达更依赖于情感特征,而动态交互机制能够根据不同语言特点灵活调整信息流动,有效整合关键信息。消融实验结果充分证明了,动态交互机制是提升模型性能的核心方法,使模型在不同语言的讽刺检测任务中都能够取得优异表现。

表 4 DiDuGAT 模型在 IAC-V1 和 IAC-V2 数据集上的消融实验结果

Table 4 Ablation study results of the DiDuGAT model on IAC-V1 and IAC-V2 datasets

		IAC-V1				IAC-V2			
方法		准确率	精确率	召回率	F1	准确率	精确率	召回率	F1
消融模型	仅情感通道	70.33	70.36	70.14	70.25	74.64	74.70	74.44	74.57
	仅句法通道	70.00	70.12	70.00	70.06	74.18	74.23	74.03	74.13
	无交互机制	71.33	71.40	71.06	71.23	74.69	74.81	74.69	74.75
原始模型	DiDuGAT	73.00	72.85	73.01	72.93	77.51	77.47	77.51	77.49

消融实验结果表明,情感通道、句法通道以及动态交互机制对模型在不同语言数据集上的性能提升均起到关键作用。完整模型的优异表现不仅验证了 DiDu-GAT 模型在中文讽刺检测任务中的设计合理性,还展示了其在英文基准上的广泛适用性。

4.7 注意力可视化

为分析 DiDu-GAT 在讽刺文本识别中的工作机制,本节分别在 GuanSarcasm 和 IAC 数据集中各随机选取了 5 条样本进行注意力分数可视化。通过展示模型对文本中不同词语的注意力分布,直观揭示了模型识别讽刺内容时的关注点和决策依据。具体的注意力可视化如图 6 和图 7 所示。

中文注意力可视化结果表明,模型能够同时捕捉标题与评论之间的词级注意力信息,并通过联合信息识别讽刺表达。具体而言,模型倾向于对具有冲突情感的词汇赋予较高的注意力权重,这些词汇通常同时出现在讽刺表达中,而在非讽刺

表达中较少出现。

例如,在图 6 的第一个示例中,模型对“手机”和“浴霸”的较高关注,表明其能够结合标题语境捕捉到负面情绪及讽刺意图。在第二个示例中,模型对“暴增”和“杀熟”的关注,显示其对暗含讽刺意味词汇的敏感度。类似地,在第三个示例中,模型对“下降”和“锅”的关注,反映了其在理解和处理经济和责任归属讽刺时的关注重点。在第四个示例中,模型对“逃跑”和“组团”的关注,捕捉到评论中对“体制”的讽刺。最后,模型对第五个示例中“套”“壳”的关注,表明其能够识别评论中关于技术创新真实性的讽刺。

综上所述,模型能够有效关注到标题和评论中的关键讽刺词汇,并通过高注意力分数的词汇进一步理解文本的讽刺含义。这种可视化分析,验证了模型在处理复杂讽刺文本时的有效性,并为进一步优化模型的注意力机制提供了重要的参考。

标题:	诺基亚	归来	?	打孔屏	、	五	摄	手机	同	“	台	”	亮相
注意力分数:	0.096	0.073	0.025	0.076	0.026	0.038	0.050	0.110	0.030	0.058	0.033	0.053	0.059
评论:	花	钱	买	浴霸									
注意力分数:	0.035	0.059	0.061	0.120									
标题:	中国	联通	利润	暴增	458%	,	去年	到底	干	了	啥	?	
注意力分数:	0.105	0.077	0.094	0.121	0.099	0.020	0.055	0.028	0.035	0.017	0.053	0.021	
评论:	无他	,	唯	杀熟	耳								
注意力分数:	0.047	0.027	0.057	0.102	0.041								
标题:	三星	预计	盈利	下降	近	40%	,	也	怪	中国	?		
注意力分数:	0.035	0.044	0.096	0.087	0.031	0.078	0.047	0.043	0.051	0.078	0.056		
评论:	反正	都	是	我们	的	锅							
注意力分数:	0.046	0.048	0.043	0.039	0.046	0.133							
标题:	法国	一	监狱	围墙	倒塌	,	223	名	囚犯	无	一	逃跑	
注意力分数:	0.054	0.001	0.006	0.049	0.080	0.024	0.031	0.066	0.022	0.061	0.096	0.062	
评论:	体制	的	胜利	,	公子们	快	组团	住	进去	!			
注意力分数:	0.036	0.018	0.075	0.063	0.041	0.038	0.022	0.045	0.055	0.055			
标题:	中国	首	款	RISC-V	高	性能	家电	芯片	在	青岛	诞生		
注意力分数:	0.069	0.021	0.029	0.040	0.030	0.061	0.027	0.042	0.019	0.023	0.032		
评论:	谁	来	说说	,	这	是	套	了	谁	家	的	壳	。
注意力分数:	0.103	0.038	0.099	0.020	0.031	0.014	0.054	0.036	0.057	0.041	0.013	0.096	0.006

图6 中文注意力可视化

Fig. 6 Attention visualization on Chinese dataset

文本:	And	you'	re	accusing	him	of	not	debating	properly	?	That's	the	best	you
注意力分数:	0.024	0.045	0.103	0.036	0.027	0.054	0.098	0.096	0.029	0.047	0.031	0.102	0.046	
	can	do	?											
	0.049	0.097	0.034											
文本:	Math	is	confusing	too	,	does	n't	mean	we	should	get	rid	of	
注意力分数:	0.122	0.050	0.080	0.071	0.018	0.051	0.077	0.101	0.049	0.077	0.105	0.100	0.041	
	of	.												
	0.045	0.015												
文本:	And	,	I	am	sure	you	ask	yourself	that	very	question	before	you	
注意力分数:	0.012	0.011	0.032	0.026	0.103	0.031	0.065	0.043	0.016	0.015	0.015	0.014	0.022	
	unzip	your	fly	and	whip	out	your	prcker	,	right	?			
	0.109	0.013	0.103	0.013	0.111	0.014	0.012	0.111	0.011	0.088	0.011			
文本:	If	we	were	to	debate	whether	God	made	us	perfect	as	he	could	
注意力分数:	0.010	0.013	0.036	0.046	0.072	0.023	0.094	0.010	0.015	0.084	0.018	0.019	0.040	
	have	,	then	of	course	we	can	make	the	point	that	he	obviously	
	0.034	0.016	0.018	0.033	0.048	0.022	0.042	0.010	0.012	0.049	0.048	0.014	0.052	
	does	n't	exits	.										
	0.013	0.022	0.076	0.011										
文本:	Well	,	you	have	proven	yourself	to	be	a	man	with	no	brain	
注意力分数:	0.016	0.002	0.006	0.019	0.051	0.039	0.015	0.010	0.019	0.037	0.018	0.095	0.009	
	,	that	is	for	sure	.	The	definition	that	was	given	was	the	
	0.002	0.034	0.019	0.018	0.020	0.002	0.046	0.091	0.024	0.010	0.057	0.038	0.007	
	one	that	scientists	use	,	no	the	layperson	.					
	0.019	0.019	0.088	0.042	0.008	0.009	0.015	0.095	0.002					

图7 英文注意力可视化

Fig. 7 Attention visualization on English datasets

结束语 本文提出了一种基于双通道动态交互的图注意力网络(DiDu-GAT),其在中英文讽刺文本检测中取得了显著的检测性能提升。该模型采用双通道架构分别处理句法和情感信息,通过双通道之间的动态交互机制实现信息的深度融合与相互补充,使模型能够更全面地捕捉和理解中英文文本中隐含的讽刺意图。动态的信息交互机制使模型能够灵活适应中文讽刺表达的复杂性。在实验中,DiDu-GAT在准确率、精确率、召回率和F1等多个关键性能指标上均优于现有基线模型,验证了其在中英文讽刺检测任务中的有效性和优越性。

尽管取得了显著的进展,模型仍有进一步完善的空间。在未来工作中,一方面可进一步探索模型在多模态情景(如图文、视频等结合)下对讽刺表达的适应能力;另一方面,还可尝试将该模型应用于情绪分析等其他复杂的中文处理任务,以验证其泛化性能和应用潜力。

参考文献

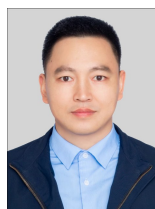
- [1] SINGH B, SHARMA D K. A survey of sarcasm detection techniques in natural language processing[C]// 2023 6th International Conference on Information Systems and Computer Networks(ISCON). IEEE, 2023: 1-6.
- [2] YAO B, ZHANG Y, LI Q, et al. Is sarcasm detection a step-by-step reasoning process in large language models? [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2025: 25651-25659.
- [3] LOU C, LIANG B, GUI L, et al. Affective dependency graph for sarcasm detection[C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 1844-1849.
- [4] DU Y, HE H, CHU Z. Cross-cultural nuances in sarcasm comprehension: a comparative study of Chinese and American per-

- spectives[J]. *Frontiers in Psychology*, 2024, 15:1349002.
- [5] XIONG T, ZHANG P, ZHU H, et al. Sarcasm detection with self-matching networks and low-rank bilinear pooling[C]// *The World Wide Web Conference*. 2019:2115-2124.
- [6] WEN Z, GUI L, WANG Q, et al. Sememe knowledge and auxiliary information enhanced approach for sarcasm detection[J]. *Information Processing & Management*, 2022, 59(3):102883.
- [7] ZHANG X, SHI N, HAUER B, et al. Bridging the Gap Between BabelNet and HowNet: Unsupervised Sense Alignment and Sememe Prediction[C]// *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. 2023:2789-2798.
- [8] YANG L, JIA W, LI S, et al. Deep learning technique for human parsing: A survey and outlook[J]. *International Journal of Computer Vision*, 2024, 132(8):3270-3301.
- [9] WANKHADE M, RAO A C S, KULKARNI C. A survey on sentiment analysis methods, applications, and challenges[J]. *Artificial Intelligence Review*, 2022, 55(7):5731-5780.
- [10] CAI T T, MA R. Theoretical foundations of t-sne for visualizing high-dimensional clustered data[J]. *Journal of Machine Learning Research*, 2022, 23(301):1-54.
- [11] FREDA S, CIGNARELLA A T, BASILE V, et al. The unbearable hurtfulness of sarcasm[J]. *Expert Systems with Applications*, 2022, 193:1.
- [12] REN H, ZHANG J, QUN N, et al. A Method for Chinese Sarcasm Detection Based on Enhanced Cross-Entropy and Regularization[C]// *2024 IEEE 5th International Conference on Pattern Recognition and Machine Learning (PRML)*. IEEE, 2024:133-137.
- [13] LI K, ZHANG Y, LI K, et al. Image-text embedding learning via visual and textual semantic reasoning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(1):641-656.
- [14] JOSHI A, SHARMA V, BHATTACHARYYA P. Harnessing context incongruity for sarcasm detection[C]// *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. 2015:757-762.
- [15] GHOSH A, VEALE T. Fracking sarcasm using neural network[C]// *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 2016:161-169.
- [16] KATTENBORN T, LEITLOFF J, SCHIEFER F, et al. Review on Convolutional Neural Networks(CNN) in vegetation remote sensing[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 173:24-49.
- [17] SHIRI F M, PERUMAL T, MUSTAPHA N, et al. A Comprehensive Overview and Comparative Analysis on Deep Learning Models[J]. *arXiv*:2305.17473, 2023.
- [18] REN W Q, QU Y B, DONG C, et al. A survey on collaborative DNN inference for edge intelligence[J]. *Machine Intelligence Research*, 2023, 20(3):370-395.
- [19] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]// *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. 2016:207-212.
- [20] NIU Z, ZHONG G, YU H. A review on the attention mechanism of deep learning[J]. *Neurocomputing*, 2021, 452:48-62.
- [21] LUKIN S, WALKER M. Really? Well. Apparently Bootstrapping Improves the Performance of Sarcasm and Nastiness Classifiers for Online Dialogue[C]// *Proceedings of the Workshop on Language Analysis in Social Media*. 2013:30-40.
- [22] ZELIKMAN E, WU Y, MU J, et al. Star: Bootstrapping reasoning with reasoning[J]. *Advances in Neural Information Processing Systems*, 2022, 35:15476-15488.
- [23] MANDAL P K, MAHTO R. Deep CNN-LSTM with word embeddings for news headline sarcasm detection[C]// *16th International Conference on Information Technology-New Generations(ITNG 2019)*. Springer, 2019:495-498.
- [24] FAN X C, YANG L, LIN H F, et al. Irony Recognition Based on Multiple Semantic Fusion[J]. *Journal of Chinese Information Processing*, 2021, 35(6):103-111.
- [25] KRISHNAN D, DURAIRAJ T. Getsmartmsec at semeval-2022 task 6: Sarcasm detection using contextual word embedding with gaussian model for irony type identification[C]// *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. 2022:827-833.
- [26] DEVLIN J, CHANG M W, 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*. 2019:4171-4186.
- [27] LIU Y, ZHANG R, FAN Y, et al. Prompt tuning with contradictory intentions for sarcasm recognition[C]// *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. 2023:328-339.
- [28] GU Y, HAN X, LIU Z, et al. PPT: Pre-trained Prompt Tuning for Few-shot Learning[C]// *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. 2022:8410-8423.
- [29] HELAL N A, HASSAN A, BADR N L, et al. A contextual-based approach for sarcasm detection[J]. *Scientific Reports*, 2024, 14(1):15415.
- [30] TAN K L, LEE C P, LIM K M. A survey of sentiment analysis: Approaches, datasets, and future research[J]. *Applied Sciences*, 2023, 13(7):4550.
- [31] TAY Y, LUU A T, HUI S C, et al. Reasoning with Sarcasm by Reading In-Between[C]// *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 2018:1010-1020.
- [32] ZHANG S, ZHANG X, CHAN J, et al. Irony detection via sentiment-based transfer learning[J]. *Information Processing & Management*, 2019, 56(5):1633-1644.
- [33] BABANEJAD N, DAVOUDI H, AN A, et al. Affective and con-

- textual embedding for sarcasm detection[C]// Proceedings of the 28th International Conference on Computational Linguistics. 2020;225-243.
- [34] CHIA Z L,PTASZYNSKI M,MASUI F,et al. Machine Learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection[J]. Information Processing & Management,2021,58(4):102600.
- [35] CHEN W,LIN F,ZHANG X,et al. Jointly learning sentimental clues and context incongruity for sarcasm detection[J]. IEEE Access,2022,10:48292-48300.
- [36] LIU Y,WANG Y,SUN A,et al. A Dual-Channel Framework for Sarcasm Recognition by Detecting Sentiment Conflict[C]// Findings of the Association for Computational Linguistics: NAACL 2022. 2022;1670-1680.
- [37] VITMAN O,KOSTIUK Y,SIDOROV G,et al. Sarcasm detection framework using context, emotion and sentiment features [J]. Expert Systems with Applications,2023,234:121068.
- [38] HUANG J,LIU Y,WANG Q F,et al. Dual-channel graph convolutional network with word-order knowledge for aspect-based sentiment analysis [J]. Application Research of Computers, 2024,41(3):779-785.
- [39] JI S,PAN S,CAMBRIA E,et al. A survey on knowledge graphs:Representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems,2021, 33(2):494-514.
- [40] LIANG B,LIN Z,QIN B,et al. Topic-Oriented Sarcasm Detection:New Task,New Dataset and New Method[C]// Proceedings of the 21st Chinese National Conference on Computational Linguistics. 2022;557-568.
- [41] LIU P,YUAN W,FU J,et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys,2023,55(9):1-35.
- [42] MIN C,LI X,YANG L,et al. Just like a human would,direct access to sarcasm augmented with potential result and reaction [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023;10172-10183.
- [43] BOSSELU T A,RASHKIN H,SAP M,et al. COMET:Commonsense Transformers for Automatic Knowledge Graph Construction[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics,2019.
- [44] HE H,CHOI J D. The Stem Cell Hypothesis:Dilemma behind Multi-Task Learning with Transformer Encoders[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021;5555-5577.
- [45] QI P,ZHANG Y,ZHANG Y,et al. Stanza:A Python Natural Language Processing Toolkit for Many Human Languages [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics:System Demonstrations. Association for Computational Linguistics,2020.
- [46] VRAHATIS A G,LAZAROS K,KOTSIANTIS S. Graph attention networks;a comprehensive review of methods and applications[J]. Future Internet,2024,16(9):318.
- [47] JIN X,XIE Y,WEI X S,et al. Delving deep into spatial pooling for squeeze-and-excitation networks [J]. Pattern Recognition, 2022,121:108159.
- [48] GONG X,ZHAO Q,ZHANG J,et al. The design and construction of a Chinese sarcasm dataset [C]// Proceedings of the Twelfth Language Resources and Evaluation Conference. 2020; 5034-5039.
- [49] MIN K,MA C,ZHAO T,et al. BosonNLP:An ensemble approach for word segmentation and POS tagging[C]// 4th CCF Conference Natural Language Processing and Chinese Computing(NLPCC 2015). Springer,2015:520-526.



TAN Pingping, born in 1999, postgraduate, is a student member of CCF (No. N8697G). Her main research interests include natural language processing and artificial intelligence.



XU Ji, born in 1979, Ph.D, distinguished professor. His main research interests include data mining, granular computing and machine learning.

(责任编辑:柯颖)