

一种融合动态语义与图结构的对比学习谣言检测方法

徐培淇¹, 刘盾¹, 李天瑞²

(1. 西南交通大学经济管理学院, 成都 610031; 2. 西南交通大学计算机与人工智能学院, 成都 611756)

摘要: 随着社交媒体的蓬勃发展, 谣言在大规模用户的线上互动中迅速扩散, 严重干扰公众认知和社会秩序。现有检测方法在建模文本动态语义演化和复杂传播结构方面仍存在局限性, 同时对易混淆类别的判别能力也有待提升。基于此, 本文提出一种融合动态语义与图结构的对比学习谣言检测 (Dynamic semantic and graph feature fusion for contrastive rumor detection, DySGCL) 方法。在文本内容方面, 采用层级 Transformer 提取用户历史帖子的全局语义表示, 并结合时序卷积强化局部语义感知能力, 以识别历史发帖的动态语义演化。在图结构方面, 采用边移除策略模拟对抗性干扰, 并结合图注意力网络 (Graph attention network, GAT) 自适应地突出核心交互关系。最后, 联合自监督与有监督的对比学习机制, 提升模型对易混淆类别的判别能力。实验结果表明, DySGCL 在公开数据集 Twitter15、Twitter16 上的准确率较基准方法分别提升了 1.8% 和 2.0%, 验证了其在谣言检测任务中的有效性。

关键词: 谣言检测; 动态语义表示; 图神经网络; 对比学习; 鲁棒性增强

中图分类号: TP18 **文献标志码:** A

引用格式: 徐培淇, 刘盾, 李天瑞. 一种融合动态语义与图结构的对比学习谣言检测方法[J]. 数据采集与处理, 2026, 41(3): 854-868. XU Peiqi, LIU Dun, LI Tianrui. Contrastive learning approach for rumor detection via fusion of dynamic semantics and graph structure[J]. Journal of Data Acquisition and Processing, 2026, 41(3): 854-868.

引言

在互联网飞速发展的当下, Twitter、微博和 Facebook 等社交媒体已成为信息获取与分享的重要渠道, 也为谣言扩散提供了温床。社会心理学文献将谣言定义为真实性难以验证或故意捏造的事件^[1], 其传播常伴随着误导性内容, 易造成公众认知偏差、社会秩序混乱, 甚至威胁国家安全^[2-3]。例如, 2019 年新冠疫情期间, 各类疫情谣言在网络上迅速扩散, 极大加剧了公众恐慌与公共危机^[4]。然而, 谣言检测往往需要对可疑信息进行深入调查和详细核实, 人工审核不仅耗时费力, 且难以应对社交媒体上指数级增长的信息规模^[5-6]。因此, 实现准确高效的谣言检测已成为亟待解决的现实问题。

在谣言检测任务中, 深入理解文本语义是实现准确检测的关键^[7]。基于文本内容建模的谣言检测方法可分为两类: 传统机器学习方法和深度学习方法, 前者侧重提取词频、情感等浅层特征, 后者则通过神经网络构建文本的深层语义表示, 以挖掘更复杂的语义关系^[8-9]。尽管上述方法在谣言检测方面取得了一定成效, 但多数研究仍将每条帖子视为独立样本, 未考虑用户历史发帖内容中的动态语义特征。在实际应用中, 用户围绕同一主题事件的多次发言往往体现出语义与立场的时间演化特征, 甚至出现表达反转的现象, 这种动态可能预示其存在谣言传播行为^[10]。已有研究表明, 识别用户在时间序列中立场与语义的变化, 有助于提升谣言的检测能力^[11]。因此, 如何捕捉用户语言的动态语义演化, 是提升

检测准确性的关键。

除了文本内容,社交网络中的传播结构同样为谣言检测提供了重要线索^[12]。图结构能够有效刻画用户之间的交互关系与信息扩散路径,已成为建模谣言传播模式的重要手段。与理想的线性传播假设不同,实际的传播图结构通常呈现出高度复杂性,除了受到直接转发行为的影响外,还受到多级转发、用户影响力差异等的间接影响,使得图结构更加难以建模^[13]。此外,传播数据中还常伴随对抗性干扰(如虚假用户操控传播路径)和噪声干扰(如转发记录缺失、评论不完整),均会破坏图结构的完整性,影响模型在复杂场景下的稳定性^[14]。现有基于图结构的谣言检测方法侧重于建模用户之间的直接互动,如通过图卷积网络(Graph convolutional network, GCN)提取局部邻接信息^[15],或借助传播树模拟信息扩散路径^[16]。然而,在面对结构复杂、异常干扰频发的现实传播环境时,现有方法仍存在一定的局限,如高阶语义依赖建模困难、鲁棒性不足等^[15,17]。因此,构建表达充分、结构稳定的图结构,以提升模型在复杂传播环境下的性能,仍是谣言检测任务的重要挑战。

进一步地,谣言文本中常存在模糊措辞的情况,诸如“传言某公司即将破产,股价可能暴跌”,此类表述在语义层面与真实信息高度接近,难以被模型准确分类^[18]。同时,社交媒体中的传播结构往往存在路径重叠、节点属性缺失等问题,使得边界样本更加难以判别^[17]。因此,如何有效识别并建模边界样本的特征,提升模型对细粒度差异的区分能力,是谣言检测的重要研究方向。

为了应对上述挑战,本文提出一种融合动态语义与图结构的对比学习谣言检测(Dynamic semantic and graph feature fusion for contrastive rumor detection, DySGCL)方法。在语义建模方面,构建词级与帖级的Transformer结构以捕捉多条推文的全局语义关联,并结合时序卷积模块(Temporal convolution module, TCM)的滑动窗口,提升对局部异常语义变化的捕捉能力。在图结构建模方面,引入边移除策略模拟社交网络中的对抗性干扰,并通过图注意力网络(Graph attention network, GAT)自适应突出关键交互关系。在特征学习层面,融合自监督与有监督对比学习机制。一方面,自监督分支基于特征聚类构造伪标签,无需标签即可挖掘语义相近样本,增强模型对潜在结构信息的理解。另一方面,有监督分支则通过真实标签推远类间距离,提升对已知类别的判别能力。二者联合优化特征空间,能够有效检测易混淆类别。最后,通过大量实验来验证本文所提方法的有效性和可靠性。

1 相关工作

1.1 谣言检测相关研究

随着人工智能技术的不断发展,谣言检测方法逐步由传统机器学习向深度学习演进,相关研究也从浅层特征提取转向对语义与图结构信息的深度建模。早期方法主要通过构建词频-逆文档频率(Term frequency-inverse document frequency, TF-IDF)、词袋模型(Bag-of-words, BoW)及n-gram等文本统计特征,并结合支持向量机(Support vector machine, SVM)、决策树(Decision tree, DT)等传统分类器实现谣言检测^[19-21]。此类方法在早期实践中因实现便捷性和特征可解释性而被广泛应用,但高度依赖人工标记的特征,受限于先验知识,难以捕捉文本中潜在的复杂语义关联^[22]。为提升模型对复杂语义的建模能力,深度学习方法被广泛应用于谣言检测任务。Ma等^[23]基于循环神经网络(Recurrent neural network, RNN)构建了谣言检测模型,其能够有效捕捉文本序列中的长期依赖,并在建模上下文语义时表现出较强的适应能力。然而,由于RNN递归转换矩阵的权重固定,模型在表达高阶语义关系方面表现不足。针对这一现象,Yu等^[24]采用卷积神经网络(Convolutional neural network, CNN)提取相邻词或短语的局部特征,并构建高级语义交互,从而增强了对短程上下文和细粒度语义的关注。然而,CNN的感受野相对固定,仅能关注局部模式,难以有效建模全局动态演化。为进一步增强语义理解

能力,Cui等^[25]融合实体识别与知识图谱链接,并结合图卷积网络,构建跨节点语义关联,有效丰富了文本的静态语义表示。

除文本特征外,图结构特征也对谣言检测有着至关重要的作用^[12,26]。已有研究发现,谣言与真实信息的传播结构存在明显的差异,为检测提供了新的切入点^[27-28]。Ma等^[29]提出基于树结构的递归神经网络(Recursive neural network, RvNN)模型,通过自底向上和自顶向下两个方向依次整合节点的文本表示,模拟谣言传播过程中的语义扩散结构。Bian等^[2]则构建了双向图卷积网络(Bi-directional graph convolutional network, BiGCN),从传播链上下游双向融合结构信息,提升了对传播特征的表达能力,并取得了更高的检测精度。为了充分利用图结构特征,Guo等^[17]综合考虑了最大扩散、传播深度与全局结构的多维图结构特征,在捕捉复杂传播结构的同时,缓冲了噪声影响。

尽管现有方法在文本与图结构特征建模方面取得了显著进展,但仍存在以下问题:(1)多数方法侧重于单条帖文的静态建模,未能充分利用用户历史发帖记录中的动态语义演化信息;(2)现有方法多采用固定加权或递归形式处理节点关系,难以自适应突出核心联系,从而限制了对复杂谣言传播特征的深层刻画。此外,部分方法对观测到的传播路径往往默认“可信”,导致模型在应对对抗性场景时鲁棒性不足。

1.2 对比学习在谣言检测中的应用

近年来,作为一种有效的特征增强与表示优化策略,对比学习被广泛地应用于自然语言处理、计算机视觉等领域,其核心思想是通过构造正负样本对,引导模型在特征空间中拉近相似样本、推远不相似样本,从而获得更具判别力的特征表示^[30-31]。该策略既适用于自监督学习场景^[32],通过数据增强、特征聚类等方式生成无标签样本对,也可扩展至有监督学习场景^[30],通过标签信息优化类别边界,实现更精准的样本分类。

在谣言检测任务中,对比学习被逐步引入,以缓解标签稀缺、样本混淆和结构干扰等问题。在自监督场景下,乔禹涵等^[32]提出了图自监督对比学习模型,通过扰动传播图结构生成多个视图,构建图级别的对比任务,提升了检测模型的鲁棒性。类似地,Liu等^[28]通过引入边不确定性的图增强策略来模拟真实传播结构中的变动,并借助图对比学习提升模型泛化能力。在有监督场景下,Dong等^[30]将有监督对比学习与对抗训练结合,利用标签信息构造类间约束,通过正样本聚合与负样本分离的方式优化特征空间,使模型更具判别力与鲁棒性。Zhang等^[6]则进一步设计了基于注意力图的对抗双重对比学习(Attention graph adversarial dual contrastive learning, AGAD)模型,通过引入对抗样本增强负样本多样性,有效提高了模型的判别能力。Zhuang等^[33]则结合图对比学习与课程学习,引入时序信息对齐和伪标签动态优化机制,以提升模型在标签不完整与传播信息不完全情境下的稳健性。尽管现有研究在对比学习与谣言检测的结合上取得了积极进展,但大多数方法往往采用单一的对比学习机制,未能充分融合自监督与有监督信号的互补优势,导致模型在面对语义边界模糊或传播结构相似的易混淆类别时仍存在判别能力不足的问题。

2 谣言检测方法

鉴于谣言检测任务面临语义表达随时间演化、传播结构复杂多变以及易混淆样本难以判别的挑战,本文提出DySGCL方法,整体框架如图1所示。相较于将语义与图结构特征独立建模的部分现有方法,DySGCL强调两者的协同表示,并通过伪标签引导的自监督对比与真实标签约束的有监督对比相结合,进而优化特征空间中易混淆样本的分类边界。

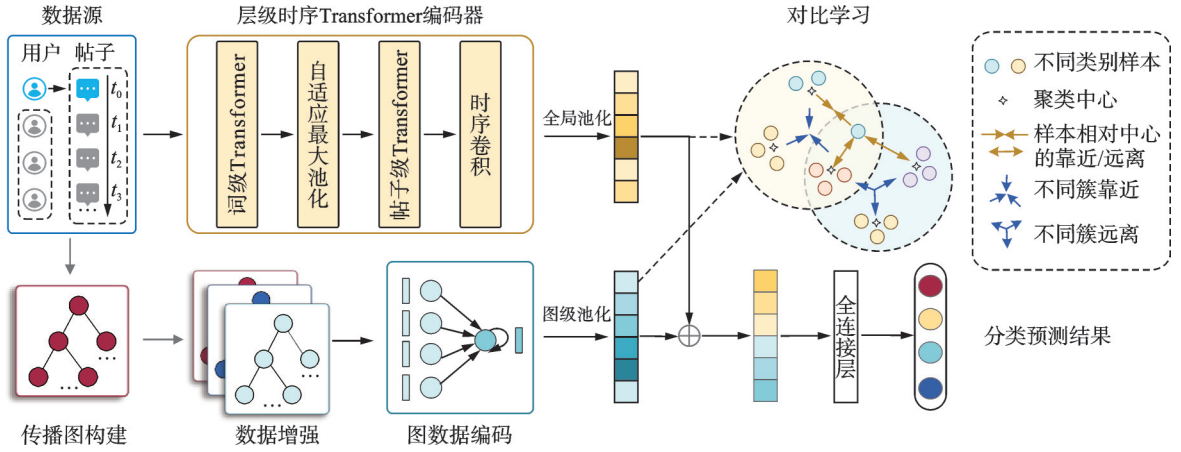


图1 DySGCL模型整体框架

Fig.1 Overall architecture of the proposed DySGCL model

2.1 动态语义建模

为了有效建模用户历史发帖的动态语义演化,本文构建了层级时序Transformer编码器(Hierarchical temporal Transformer encoder, HTTE),由词级Transformer、帖级Transformer和时序卷积模块组成。HTTE以历史发帖序列 P 作为输入,建模文本语义在时间维度上的全局与局部表示。具体而言,本文选取与源帖主题相关的历史发帖中最多 N 条帖子,并按发布时间顺序排列,模拟谣言传播过程中用户语言风格随时间演变的真实语境。历史发帖序列经过分词和嵌入操作,得到四维张量表示 $X \in \mathbb{R}^{B \times N \times T \times d}$,作为编码器输入。其中, B 表示批次大小, T 表示单个帖子的最大长度, d 表示词向量维度。

随后, X 依次通过词级Transformer捕捉单条帖子内部的上下文语义表征,通过帖级Transformer建模多个帖子之间的语义关联,最终输出每个样本的全局语义表示。在词级和帖级Transformer模块的实现中,每一层均采用了残差连接与层归一化,以增强模型训练的稳定性和特征表达能力。各层参数不共享,即每一层Transformer编码器均拥有独立的权重参数。多头自注意力机制各头参数亦互不共享,并在每层后均设置Dropout以提升泛化能力。由于上述Transformer架构主要依赖全局自注意力机制,难以充分建模相邻发帖之间的局部异常语义变化。本文TCM模块通过滑动窗口聚合邻近时刻的语义特征,提升模型对局部动态语义演化的表达能力。TCM采用一维扩张卷积操作,使用滑动窗口在时间维度上聚合相邻帖子的语义表示,具体操作如下

$$y_t = \text{ReLU} \left(\sum_{k=0}^{K_c-1} \mathbf{W}_1^{(k)} * x_{t-r,k} + \mathbf{b}_1 \right) \quad (1)$$

$$y'_t = \sum_{k=0}^{K_c-1} \mathbf{W}_2^{(k)} * y_t + \mathbf{b}_2 \quad (2)$$

式中: K_c 为卷积核的大小, r 表示扩张率, $x_{t-r,k}$ 为卷积感受域内的输入特征, $\mathbf{W}_1^{(k)}$ 、 $\mathbf{W}_2^{(k)}$ 和 \mathbf{b}_1 、 \mathbf{b}_2 为卷积参数。在捕捉文本语义在时间维度上的全局与局部的演化特征后,模型通过自适应池化获取每个样本的整体动态语义表示 h_p ,为后续特征空间的对比优化与分类任务奠定动态语义基础。

2.2 图结构建模

图结构建模结合了边移除策略与图神经网络,能够精准刻画帖子之间的交互关系,提升模型对复

杂传播关系的建模能力。

2.2.1 输入构建与增强

为了有效缓解社交网络中传播路径存在的结构噪声与对抗性干扰问题,本文在图结构建模中引入边移除的数据增强策略。具体而言,将每个谣言事件建模为1个包含传播关系的有向图,以模拟真实社交平台中用户之间的信息扩散过程。给定 n 个谣言事件的集合 $M=\{M_1, M_2, \dots, M_n\}$,第 i 个谣言记为 M_i ,包含1个源帖 $G_i^{(0)}$ 及 $(l-1)$ 个互动帖子 $G_i^{(1)}, G_i^{(2)}, \dots, G_i^{(l-1)}$,并且可以构建一个包含传播关系的有向图 $G_i=(V_i, E_i)$ 。帖子集合 $V_i=\{v_1, v_2, \dots, v_j, \dots, v_k\}$,边集 E_i 表示帖子间的交互关系。若帖子 v_j 与 v_k 存在回复或转发关系,则边连 $(v_j, v_k) \in E_i$,据此构建邻接矩阵 $A_i \in \{0, 1\}^{N \times N}$,其中 $A_i[j, k]=1$ 表示存在有向边。

在训练过程中,为防止模型过拟合,提高泛化能力,本文采用边移除(Edge removing, ER)策略进行结构扰动。具体而言,对每条边 (v_j, v_k) 以概率 ρ 其从边集中移除,得到扰动后边集 $\tilde{E}_i \subseteq E_i$ 。在每个训练周期,可以得到不同扰动图 $G'=(V_i, \tilde{E}_i)$,使得模型可以不断面对不同结构的扰动图样本,能够显著降低对特定边的过拟合风险,进而提升检测的精确度与鲁棒性。

2.2.2 空间编码模块

为了充分挖掘传播图中节点之间的交互结构信息,提升模型对复杂传播关系的建模能力,本文引入基于图注意力网络的空间编码模块,对节点进行特征聚合与表达学习。GAT能够为不同邻接节点分配不同的注意力权重,从而更精准地刻画节点间的交互关系。对于给定的传播图结构 G ,定义第 l 层节点特征矩阵 $H^{(l)} \in \mathbb{R}^{N \times D}$,对于给定节点 v_i ,计算节点 i 与其邻居节点 j 之间的相似度为

$$e_{ij}^{(l,k)} = \text{LeakyReLU}(\mathbf{a}^{(k)\top} [\mathbf{W}^{(l,k)} \mathbf{h}_i^{(l)} // \mathbf{W}^{(l,k)} \mathbf{h}_j^{(l)}]) \quad (3)$$

归一化后得到注意力权重

$$\alpha_{ij}^{(l,k)} = \frac{\exp(e_{ij}^{(l,k)})}{\sum_{m \in N_i} \exp(e_{im}^{(l,k)})} \quad (4)$$

式中: $\mathbf{h}_i^{(l)}$ 和 $\mathbf{h}_j^{(l)}$ 分别是节点 v_i 和节点 v_j 在第 l 层的特征向量, $\mathbf{W}^{(l,k)}$ 表示第 l 层第 k 个注意力头的学习权重矩阵。 $\mathbf{a}^{(k)}$ 表示第 k 个注意力头的注意力向量,LeakyReLU为激活函数, $\alpha_{ij}^{(l,k)}$ 表示节点 v_j 对节点 v_i 的归一化注意力权重, N_i 表示节点 v_i 及其一阶邻居节点集合, $e_{im}^{(l,k)}$ 表示 v_i 与邻居 v_m 之间的注意力得分。最终,聚合邻接节点信息并残差连接形成节点的下一层特征 $\mathbf{h}_i^{(l+1)}$,有

$$\mathbf{h}_i^{(l+1)} = \parallel_{k=1}^{K_h} \left(\sum_{j \in N_i} \alpha_{ij}^{(l,k)} \mathbf{W}^{(l,k)} \mathbf{h}_j^{(l)} \right) \quad (5)$$

$$\mathbf{h}_i^{(f)} = \text{ReLU}(\mathbf{U}^{(l)}[\mathbf{h}_i^{(0)} \parallel \mathbf{h}_i^{(l+1)}]) \quad (6)$$

式中: K_h 表示注意力头数量; $\mathbf{h}_i^{(f)}$ 表示 v_i 的最终特征向量, $\mathbf{h}_i^{(0)}$ 表示节点 v_i 的初始特征向量, $\mathbf{U}^{(l)}$ 表示第 l 层的可学习权重矩阵;“ \parallel ”表示向量拼接。

最后,通过平均池化获得原图 G 与扰动图 G' 的图级表示 $\mathbf{h}_G, \mathbf{h}_{G'}$,为后续特征空间的对比优化与最终分类任务提供结构表达。

2.3 特征融合与对比学习

为了进一步整合动态语义与图结构信息,提升模型对易混淆样本的判别能力,本文将HTTE捕捉到的动态语义特征 \mathbf{h}_p 与GAT提取的图结构特征 $\mathbf{h}_G, \mathbf{h}_{G'}$ 进行拼接融合,得到初步的特征表示 \mathbf{h}_p ,并引入全连接层以获取最终融合特征表示 \mathbf{h}_F ,用于后续特征优化与分类预测。

在此基础上,本文引入联合对比学习机制,结合自监督与有监督分支共同优化特征空间。训练过程中,按照样本与聚类中心的距离由近及远输入训练样本。具体而言,自监督分支通过聚类生成伪标签,以拉近相似样本之间的距离。给定第 i 个样本最终表示为 \mathbf{h}_F^i ,其对应的自监督聚类中心为 \mathbf{c}_i ,自监督对比损失定义为

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \ln \frac{\exp(\sin(\mathbf{h}_F^i, \mathbf{c}_i)/\tau)}{\sum_{j=1}^N \exp(\sin(\mathbf{h}_F^i, \mathbf{c}_j)/\tau)} \quad (7)$$

式中: $\sin(\mathbf{h}_F^i, \mathbf{c}_i)$ 表示 \mathbf{h}_F^i 和 \mathbf{c}_i 的相似度, τ 为温度系数。

与自监督分支不同,有监督分支通过样本标签信息,拉近同类样本、推远异类样本,从而优化特征空间的判别边界,强化模型对类别间微小差异的识别能力,其损失函数为

$$L_2 = -\frac{1}{N} \sum_{i=1}^N \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\sin(\mathbf{h}_F^i, \mathbf{h}_F^p)/\tau)}{\sum_{j=1}^N \exp(\sin(\mathbf{h}_F^i, \mathbf{h}_F^j)/\tau)} \quad (8)$$

式中: $P(i)$ 表示与样本 \mathbf{h}_F^i 属于同一类别的正样本集合, \mathbf{h}_F^p 表示正样本集合中样本 p 的特征。最终对比损失如下

$$L_{\text{con}} = \alpha_1 L_1 + \alpha_2 L_2 \quad (9)$$

式中 α_1, α_2 为两个分支的权重系数,用于平衡两种对比策略在训练中的贡献。

2.4 谣言分类器

通过融合后的特征 \mathbf{h}_F ,得到最终的分类输出 \hat{y}_i 为

$$\hat{y}_i = \text{softmax}(\mathbf{W}_c \mathbf{h}_F + \mathbf{b}_c) \quad (10)$$

式中: \mathbf{W}_c 和 \mathbf{b}_c 为分类器的参数, \hat{y}_i 为第 i 个样本在各类别下的预测概率。

为提升分类性能,模型采用交叉熵损失函数最小化预测分布 \hat{y}_i 与真实标签 y_i 之间的差异,损失函数为

$$L_{\text{ce}} = -\sum_{i=1}^N y_i \log \hat{y}_i \quad (11)$$

式中 y_i 表示第 i 个样本的真实标签。最终训练损失主要由两部分组成,即

$$L_{\text{fin}} = L_{\text{ce}} + \lambda L_{\text{con}} \quad (12)$$

式中 λ 为加权系数,用于平衡交叉熵损失和对比损失。

2.5 算法流程与复杂度分析

基于上述分析,DySGCL的训练流程可分为5个阶段:动态语义建模、图结构建模、特征融合与对比学习、分类预测以及早停机制,整体训练流程如算法1所示。首先,模型提取用户历史发帖序列 P 的动态语义表示 \mathbf{h}_p ,随后通过GAT分别编码原始图 G 和扰动图 G' ,得到相应的结构特征 $\mathbf{h}_G, \mathbf{h}_{G'}$ 。再将语义与图特征融合为最终表示 \mathbf{h}_F ,用于损失计算与分类预测。在训练阶段,联合优化交叉熵损失 L_{ce} 与对比损失 L_{con} ,以提升对易混淆样本的区分能力。训练过程采用早停机制以实现稳定收敛。

算法1 融合动态语义与图结构的对比学习谣言检测方法

输入:用户历史发帖序列 P ,传播图 $G=(V, E)$,自监督伪标签 C ,标签信息 y

输出:预测结果 $\hat{y} \in \{\text{NR}, \text{FR}, \text{TR}, \text{UR}\}$

(1) $\mathbf{h}_p \leftarrow \text{HTTE}(P)$ //动态语义建模,式(1~2)

- (2) $G' \leftarrow \text{StructureAug}(G)$ //扰动图生成
- (3) $h_G \leftarrow \text{GAT}(G)$ //图结构编码,式(3~6)
- (4) $h_{G'} \leftarrow \text{GAT}(G')$
- (5) $h_F \leftarrow \text{Fuse}(h_P, h_G, h_{G'})$ //特征融合
- (6) $L_1 \leftarrow \text{SelfContrast}(h_F, C)$ //式(7)
- (7) $L_2 \leftarrow \text{SupContrast}(h_F, y)$ //式(8)
- (8) $L_{\text{con}} \leftarrow \alpha_1 \cdot L_1 + \alpha_2 \cdot L_2$ //对比学习优化
- (9) $\hat{y} \leftarrow \text{Classifier}(h_F)$ //分类预测
- (10) $L_{\text{ce}} \leftarrow \text{CrossEntropy}(\hat{y}, y)$ //式(11)
- (11) $L_{\text{total}} \leftarrow L_{\text{ce}} + \lambda \cdot L_{\text{con}}$
- (12) 初始化: $\text{patience} \leftarrow 0, \text{bestVal} \leftarrow +\infty$
- (13) for epoch = 1 to maxEpoch do
- (14) 更新参数,计算 L_{val}
- (15) 若 L_{val} 无提升达10次,提早停止训练
- (16) end for
- return \hat{y}

为了进一步验证DySGCL方法的有效性,本文从动态语义建模、图结构建模和对比学习3个核心模块出发,对单个epoch中批处理大小为 B 的样本进行时间复杂度分析。(1)动态语义模块:每个样本包含 M 条历史发帖,每条发帖序列长度为 T ,模型的隐藏维度为 d 。经过多头自注意力、时序卷积等操作后,该部分的复杂度约为 $O(BMT^2d)$ 。(2)图结构模块:单个批次含 $|E_b|$ 条边,该部分的时间复杂度约为 $O(|E_b|d)$ 。(3)对比学习模块:每个epoch对 N 个样本聚类,中心数为 K ,迭代次数为 I ,该部分的时间复杂度约为 $O(NBd) + O(NKId)$ 。考虑到并行化实现和早停机制,本文方法在中等规模的数据集上能够在短时间内收敛,实际运行的时间复杂度往往低于理论上限。

3 实验设计与结果分析

本节旨在评估所提出的DySGCL方法在谣言检测任务中的性能表现。为验证模型的有效性与稳健性,本文从多个角度展开实验分析。首先介绍实验所用的数据集、参数设置与评价指标,随后依次进行基准模型对比、关键模块消融分析、鲁棒性与敏感性分析以及特征可视化分析,以综合验证模型性能。

3.1 实验设置

3.1.1 数据集与实验环境

本实验选取谣言检测领域中常用的数据集: Twitter15和Twitter16^[14]。两个数据集均来自于真实社交平台,事件标签参考权威谣言揭穿网进行标注。数据集的标签包含4类:非谣言(Non-rumor, NR),已证实为假信息(False rumor, FR),已证实为真信息(True rumor, TR)以及未经验证的信息(Unverified rumor, UR)。具体统计信息如表1所示。

表1 数据集统计

Table 1 Dataset statistics			
数据集	Twitter15	Twitter16	
帖子数	331 612	204 820	
用户数	276 663	173 487	
谣言事件数	1 490	818	
非谣言(NR)	374	205	
已证实为假信息(FR)	370	205	
已证实为真信息(TR)	372	205	
未经验证的信息(UR)	374	203	

所有实验均在 Ubuntu 22.04 LTS 系统环境下进行,硬件配置为 Intel Xeon Platinum 8481C (16 vCPU)、RAM80 GB、NVIDIA RTX 4090D(24 GB)。代码基于 Python 3.10 与 PyTorch 2.0.0 实现, CUDA 驱动版本为 11.8。

3.1.2 参数设置与评价指标

在实验参数设置方面,针对文本信息的初始表示,将节点的特征维度固定为 5 000。图神经网络采用多头自注意力机制,隐藏层特征维度设为 64。两个数据集上的边移除率分别为 0.3 和 0.5,以降低过拟合风险,对比学习中的温度系数 $\tau = 0.2$ 。在训练过程中,批处理大小设为 128,学习率为 0.000 5,采用 Adam 优化器结合权重衰减策略更新模型参数,训练轮次上限为 200,在验证集连续 10 轮损失无显著下降时,触发早停机制。此外,自监督对比学习的聚类个数为 4,引入可调超参数 $\lambda = 0.1$,以平衡交叉熵损失和对比损失的贡献,上述实验配置旨在确保模型在收敛速度和泛化性能之间取得有效平衡。

为了有效评估模型的性能表现,实验选取准确率 (Accuracy, Acc)、 F_1 及 Macro- F_1 作为主要评价指标^[17,34-35]。为增强实验结果的稳定性,所有数据集均按照 4:1 的比例划分为训练集和测试集,并采用五折交叉验证策略,以确保实验结果的稳健性。

3.2 对比实验

为了全面评估 DySGCL 在谣言检测任务中的整体性能,本小节将其同多个基准方法进行比较分析。这些方法主要分为两类。

(1) 以内容为主要特征的检测方法。(a) SRLF^[36]:基于强化学习实现自动标注评论立场,并选择高质量数据。(b) KAGN^[25]:将知识图谱同注意力机制与图卷积网络相结合。(c) CE-BERT^[37]:通过减少 BERT 参数规模,在提高检测速度的同时,保持模型精度。(d) Rumor2Topic^[38]:在图注意力网络上融合文本与结构信息进行联合学习。(e) GPT2Ad^[39]:利用 GPT-2 生成类谣言文本的增强数据集。

(2) 以传播结构为主要特征的检测方法。(a) GLAN^[40]:通过异构图建模,融合消息内容与全局传播结构信息,提升识别效果。(b) BiGCN^[2]:提出双向图卷积网络模型,实现更高效的谣言识别。(c) Rumor2vec^[16]:构建谣言传播的联合图,并结合卷积神经网络进行谣言识别。(d) GA-CL^[15]:采用图对抗对比学习方法,有捕获传播网络中的不变特征。(e) MDG^[17]:融合最大扩散、深度传播和全局结构特征进行检测。(f) MPT^[41]:基于传播树提取文本和用户可信度特征。

相关实验结果如表 2 和表 3 所示,粗体数值表示最优性能。由表 2 和表 3 可以看到, DySGCL 在 Twitter15 和 Twitter16 上的准确率分别达到 93.7% 和 93.9%。与以内容为主的最优方法 GPT2Ad 相比, DySGCL 在 Twitter15 和 Twitter16 上的准确率分别提升了 1.8% 和 2.0%。这得益于本方法所采用的 HTTE 模块,从而实现细粒度的动态语义建模,为谣言检测提供更具区分力的语义信息。此外,与以传播结构为主的最优方法 MPT 相比, DySGCL 在 Twitter15 和 Twitter16 上的准确率分别提升了 4.4% 和 2.2%。这表明,尽管侧重于传播结构的方法在捕捉谣言传播路径方面有效,但难以全面捕捉复杂的交互模式。本方法通过融合数据增强与图注意力网络,提升了对关键节点之间影响关系的建模能力,进而有效增强了模型对谣言传播信息的表征能力。

基于上述分析,本文发现在两组数据集上, DySGCL 的准确率 Acc、 F_1 (除 UR 类外) 以及 Macro- F_1 均优于其他模型。虽然 UR 类的 F_1 略逊于最优性能,但仍高于大多数方法。以上结果验证了 DySGCL 在谣言检测任务中的有效性。

表2 各模型在 Twitter15 上的性能

Table 2 Performance of each model on Twitter15

方法类别	方法	Acc	F_1				Macro- F_1
			NR	FR	TR	UR	
基于文本内容	SRLF	0.890	0.890	0.910	0.919	0.842	0.890
	KAGN	0.892	0.868	0.883	0.894	0.927	0.893
	CE-BERT	0.867	0.896	0.806	0.925	0.833	0.865
	Rumor2Topic	0.863	0.889	0.842	0.901	0.843	0.868
	GPT2Ad	0.919	0.883	—	—	—	0.901
基于传播图	GLAN	0.904	0.924	0.917	0.852	0.927	0.905
	BiGCN	0.871	0.860	0.867	0.914	0.854	0.874
	Rumor2vec	0.796	0.883	0.746	0.836	0.723	0.797
	GACL	0.882	0.878	0.853	0.890	0.884	0.877
	MDG	0.802	0.735	0.816	0.878	0.756	0.797
	MPT	0.893	0.880	0.908	0.863	0.919	0.893
本文方法	DySGCL	0.937	0.953 ↑	0.936 ↑	0.953 ↑	0.906 ↓	0.937 ↑

表3 各模型在 Twitter16 上的性能

Table 3 Performance of each model on Twitter16

方法类别	方法	Acc	F_1				Macro- F_1
			NR	FR	TR	UR	
基于文本内容	SRLF	0.886	0.879	0.854	0.909	0.899	0.885
	KAGN	0.901	0.864	0.881	0.946	0.908	0.900
	CE-BERT	0.860	0.776	0.880	0.923	0.870	0.862
	Rumor2Topic	0.886	0.891	0.864	0.925	0.867	0.887
	GPT2Ad	0.919	0.857	—	—	—	0.888
基于传播图	GLAN	0.902	0.921	0.869	0.847	0.968	0.901
	BiGCN	0.885	0.829	0.899	0.932	0.882	0.886
	Rumor2vec	0.852	0.857	0.769	0.927	0.850	0.851
	GACL	0.894	0.906	0.870	0.878	0.932	0.896
	MDG	0.845	0.756	0.827	0.892	0.869	0.838
	MPT	0.917	0.892	0.908	0.925	0.946	0.918
本文方法	DySGCL	0.939	0.948 ↑	0.928 ↑	0.976 ↑	0.907 ↓	0.939 ↑

3.3 消融实验

在验证了 DySGCL 方法的整体性能后,为了进一步探究各关键模块在模型中的作用,本小节聚焦于模块级消融实验。本文构建了 5 个变体模型,具体如下:(1)w/o DAT:取消边移除策略;(2)w/o GAT:以 GCN 代替 GAT;(3)w/o HTTE:仅保留时序词级编码,取消帖子级编码;(4)w/o TCM:在 HTTE 基础上移除时序卷积模块;(5)w/o FIN:去除对比学习(含监督和自监督)模块。本实验通过逐一移除 DySGCL 模型的核心部分,评估各模块对谣言检测性能的贡献。图 2 展示了各变体在 Twitter15 和 Twitter16 数据集上的准确率 Acc 与 Macro- F_1 。

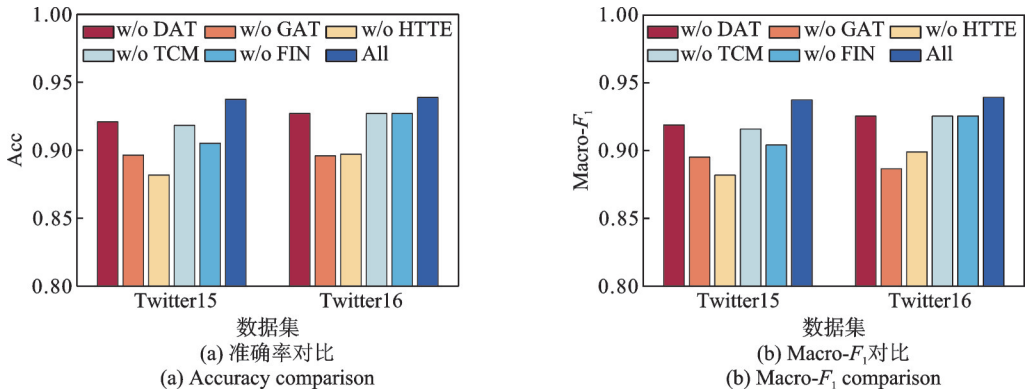


图2 消融实验对比结果

Fig.2 Comparative results of ablation experiments

结果显示,当移除DySGCL中的任意一个模块时,整体性能均有不同程度的下降,这表明各模块对于谣言检测任务都是有效的。具体而言,取消边移除策略使图结构噪声增加,而替换图注意力网络使得关键邻居信息无法得到充分利用,两者都会让整体性能下降。尤其是取消层级编码后,推文间的交互信息与用户历史发帖模式的综合利用受到限制,进而极大降低了模型对谣言的检测能力。移除时序卷积模块后,模型对动态语义特征提取不足,易出现错判。移除对比学习的约束后,模型在判别能力方面大幅削弱。由此可知,DySGCL中各模块均发挥着不可替代的作用,其优异的检测性能归因于各模块之间的协同优化。

3.4 鲁棒性与敏感性分析

在评估DySGCL整体性能及各模块贡献的基础上,为了进一步验证其在复杂环境下的稳定性与适应性,本节进行鲁棒性与参数敏感性分析。

3.4.1 鲁棒性分析

为了模拟社交网络中传播结构不完整的情况,本文对传播图中的边结构进行不同程度的扰动。采用不同的边移除率 $\rho \in \{0.1, 0.2, \dots, 0.8, 0.9\}$,其取值范围一方面参考了近年来图神经网络鲁棒性及谣言检测领域的主流实验惯例^[2,15,32,42],另一方面能够系统覆盖从轻度到极端扰动的多种情形,从而全面考察模型对图结构不完整及噪声的适应能力。

实验结果如图3所示。当移除率 ρ 分别为0.3和0.5时,模型在Twitter15、Twitter16上的表现达到最佳。实验结果表明,DySGCL模型在面对传播图边结构的扰动时表现出一定的鲁棒性。适当的扰动不仅不会显著降低模型的性能,反而可能通过增加数据的多样性来提升模型的泛化能力。然而,当扰动超过一定限度时,模型的性能会受到负面影响。

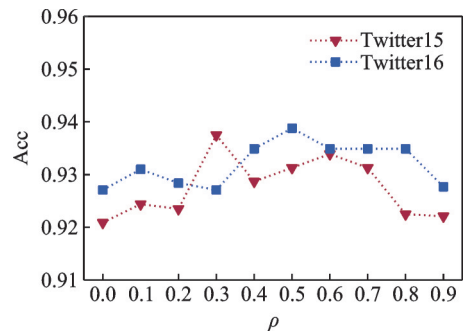


图3 不同边移除率的影响

Fig.3 Effect of different edge removal rates

3.4.2 参数敏感性分析

为了进一步评估DySGCL对关键超参数的敏感性,本节选取以下3个具有代表性的超参数进行实验:对比损失权重 λ 、自监督聚类个数 K 以及学习率 lr 。其中, λ 用于控制对比损失在整体训练目标中的比重; K 决定了聚类中心的数量; lr 控制参数更新步长,直接影响训练过程的收敛速度与稳定性。在实

验过程中,每次仅调整其中一个超参数,其余保持不变,并在 Twitter15 和 Twitter16 两个数据集上分别进行多次重复实验,取平均值以减少随机误差,实验结果如图 4 所示。由图 4 可知,3 个超参数在不同数据集上的变化趋势基本一致,随着超参数取值的逐步增大,模型准确率呈现先升后降的变化趋势。适度增大 λ 和 K 能够提升模型性能,但过大会导致噪声的引入或训练的偏移,反而降低模型性能。学习率 lr 的变化呈现相似趋势,过小导致模型陷入局部最优,过大则导致模型性能不稳定。

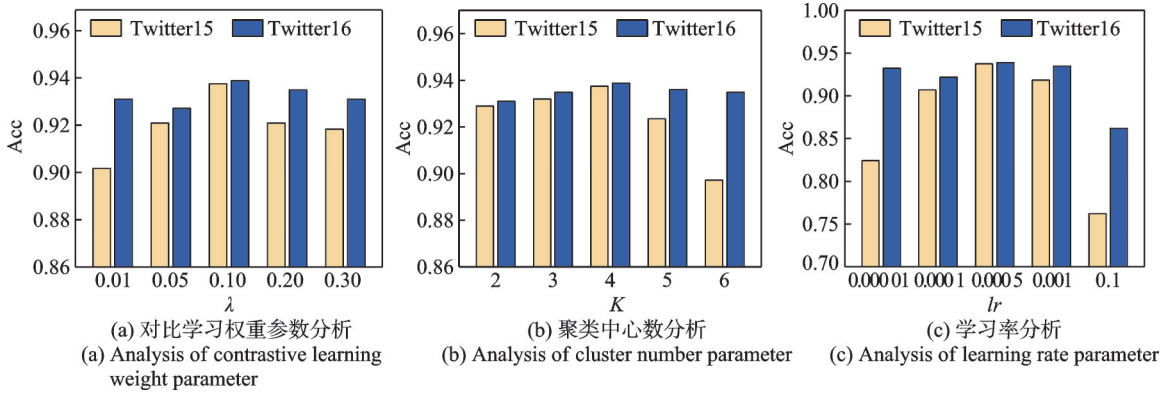


图 4 参数敏感性分析实验

Fig.4 Analysis experiment for parameter sensitivity

综合不同参数的影响,最终确定参数的最优配置如下:对比损失权重为 0.1,自监督聚类个数为 4,学习率为 0.000 5。

3.5 谣言早期检测分析

为系统评估 DySGCL 模型的谣言检测性能,本文设定不同的检测截止时间 $T \in \{0, 3, 6, 12, 24, 36\}$ h,其中 $T=0$ 表示仅包含源帖。对于每个谣言事件,以其发布时间为起点,仅保留 T 小时内的传播帖子,用于构建该阶段的数据集。在每个截止时间点上,本文对比 BiGCN、GACL、GLAN 三种基线方法,从而验证模型的早期检测性能,实验结果如图 5 所示。通过对不同检测截止时间下的实验结果进行对比,可以发现 DySGCL 在谣言传播初期便表现出较强的检测能力,且在信息逐步丰富的各阶段始终优于基线方法。随着信息的逐步丰富,DySGCL 在各阶段始终优于对比方法,充分体现了模型在谣言早期预警场景下的有效性。

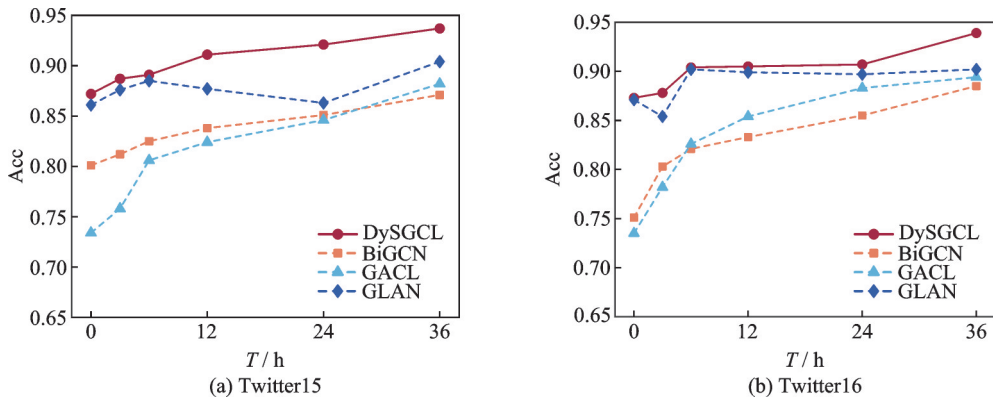


图 5 不同检测截止时间下早期谣言检测的性能表现

Fig.5 Performance of early rumor detection under different detection deadlines

3.6 可视化分析

在上述定量实验验证 DySGCL 有效性的基础上,本文进一步引入 t 分布随机邻域嵌入(t-distributed stochastic neighbor embedding, t-SNE)可视化分析,从定性角度验证模型对不同类别的区分能力,如图 6 所示。由可视化结果可知,多数样本在低维空间中呈现出清晰的分离边界,表明 DySGCL 能够有效提取具有判别性的深层特征。尽管可视化结果显示了 NR 和 UR 类别之间存在极少数重叠,但这一重叠反映了 UR 类别的样本在特征空间中的不确定性。由于 UR 类别通常缺乏明确标签或事实依据,这意味着部分 UR 类样本在经过进一步验证后,存在被归入 NR 类的可能性。因此,该现象可应视为类别标签的模糊性,而非模型特征提取能力的不足。总而言之,t-SNE 可视化结果与定量实验指标表现相一致,从特征空间层面进一步验证了 DySGCL 在谣言检测任务中的有效性。

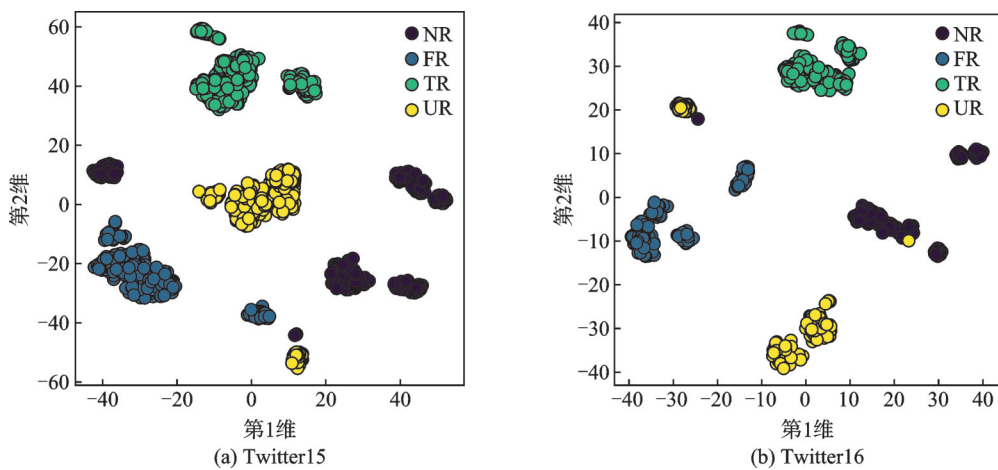


图6 t-SNE降维下的特征分布

Fig.6 Feature distribution under t-SNE dimensionality reduction

4 结束语

本文提出了一种融合动态语义与图结构的对比学习谣言检测方法 DySGCL,用以应对谣言检测中动态语义建模不足、传播结构复杂及类别边界模糊的挑战。DySGCL 通过层级时序 Transformer 建模用户历史发帖中的动态语义演化,并结合边移除策略和图注意力网络提升模型的鲁棒性,最后通过联合对比学习优化特征空间的分布,提升对易混淆类别的判别能力。在 Twitter15 和 Twitter16 两个公开数据集上的实验结果表明,DySGCL 方法的准确率和 Macro- F_1 均优于基准模型,验证了本文方法在谣言检测任务上的有效性。消融实验则进一步验证了各模块在性能提升中的重要性与互补性。

为了进一步提升模型在复杂真实环境下的适应性与实用性,未来研究可从以下两个维度展开:(1) 融合图像、音频等异构模态信息,构建多模态协同建模框架,并完善多语言数据编码方法,以提升模型在新媒体平台和多语言环境下的泛化能力和协同推理能力;(2) 为应对 AIGC 驱动的谣言生成趋势,可结合生成样本增强与对抗学习机制,提升模型对生成式伪信息的检测能力,以适应谣言传播形态的不断演化。

参考文献:

- [1] ZHANG Q, COOK J, YILMAZ E. Detecting and forecasting misinformation via temporal and geometric propagation patterns [C]//Proceedings of Advances Information Retrieval. Cham: Springer, 2021: 455-462.
- [2] BIAN T, XIAO X, XU T, et al. Rumor detection on social media with bi-directional graph convolutional networks[C]//

- Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 549-556.
- [3] REN G, JIANG L, HUANG T, et al. Temporal-spatial hierarchical contrastive learning for misinformation detection: A public-behavior perspective[J]. *Information Processing & Management*, 2025, 62(4): 104108.
- [4] TASNIM S, HOSSAIN M M, MAZUMDER H. Impact of rumors and misinformation on COVID-19 in social media[J]. *Journal of Preventive Medicine and Public Health*, 2020, 53(3): 171.
- [5] PENG X, WU J, LIU R, et al. Rumor detection on social media with temporal propagation structure optimization[C]// *Proceedings of the 31st International Conference on Computational Linguistics*. Abu Dhabi, UAE: Association for Computational Linguistics, 2025: 3865-3878.
- [6] ZHANG B, LIU T, KE Z, et al. Rumor detection based on attention graph adversarial dual contrast learning[J]. *Plos One*, 2024, 19(4): e0290291.
- [7] ZHENG P, HUANG Z, DOU Y, et al. Rumor detection on social media through mining the social circles with high homogeneity[J]. *Information Sciences*, 2023, 642: 119083.
- [8] CHEN X, ZHOU F, TRAJCEVSKI G, et al. Multi-view learning with distinguishable feature fusion for rumor detection[J]. *Knowledge-Based Systems*, 2022, 240: 108085.
- [9] ZHANG X, GAO W. Predicting viral rumors and vulnerable users with graph-based neural multi-task learning for infodemic surveillance[J]. *Information Processing & Management*, 2024, 61(1): 103520.
- [10] ZHANG K, CAO J, PI D. A novel fine-grained rumor detection algorithm with attention mechanism[J]. *Neurocomputing*, 2024, 583: 127595.
- [11] WU Y, SUN J, YUAN X, et al. Dual-channel early rumor detection based on factual evidence[J]. *Expert Systems with Applications*, 2024, 238: 121928.
- [12] ROSENFELD N, SZANTO A, PARKES D C. A kernel of truth: Determining rumor veracity on Twitter by diffusion pattern alone[C]// *Proceedings of The Web Conference 2020*. New York: ACM, 2020: 1018-1028.
- [13] WU Z, PAN S, LONG G, et al. Connecting the dots: Multivariate time series forecasting with graph neural networks[C]// *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 2020: 753-763.
- [14] MA J, GAO W, WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]// *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg, PA: Association for Computational Linguistics, 2017: 708-717.
- [15] SUN T, QIAN Z, DONG S, et al. Rumor detection on social media with graph adversarial contrastive learning[C]// *Proceedings of the ACM Web Conference 2022*. New York, NY: ACM, 2022: 2789-2797.
- [16] TU K, CHEN C, HOU C, et al. Rumor2vec: A rumor detection framework with joint text and propagation structure representation learning[J]. *Information Sciences*, 2021, 560: 137-151.
- [17] GUO Y, JI S, CAO N, et al. MDG: Fusion learning of the maximal diffusion, deep propagation and global structure features of fake news[J]. *Expert Systems with Applications*, 2023, 213: 119291.
- [18] KE Z, LI Z, ZHOU C, et al. Rumor detection on social media via fused semantic information and a propagation heterogeneous graph[J]. *Symmetry*, 2020, 12(11): 1806.
- [19] LIANG G, HE W, XU C, et al. Rumor identification in microblogging systems based on users' behavior[J]. *IEEE Transactions on Computational Social Systems*, 2015, 2(3): 99-108.
- [20] AJAO O, BHOWMIK D, ZARGARI S. Sentiment aware fake news detection on online social networks[C]// *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Piscataway, NJ: IEEE, 2019: 2507-2511.
- [21] MAAN M, JAIN M K, TRIVEDI S, et al. Machine learning based rumor detection on twitter data[C]// *Proceedings of the International Conference on Emerging Technologies in Computer Engineering: Cognitive Computing and Intelligent IoT*. Cham: Springer, 2022: 259-273.
- [22] HUANG X, MA T, JIA L, et al. An effective multimodal representation and fusion method for multimodal intent recognition [J]. *Neurocomputing*, 2023, 548: 126373.
- [23] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]// *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016)*. New York: AAAI, 2016: 3818-3824.
- [24] YU F, LIU Q, WU S, et al. A convolutional approach for misinformation identification[C]// *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. California: IJCAI, 2017: 3901-3907.

- [25] CUI W, SHANG M. KAGN: Knowledge-powered attention and graph convolutional networks for social media rumor detection [J]. Journal of Big Data, 2023, 10(1): 45.
- [26] FENG S, WAN H, WANG N, et al. TwiBot-20: A comprehensive twitter bot detection benchmark[C]//Proceedings of the 30th ACM International Conference on Information and Knowledge Management. New York: ACM, 2021: 4485-4494.
- [27] RAPONI S, KHALIFA Z, OLIGERI G, et al. Fake news propagation: A review of epidemic models, datasets, and insights [J]. ACM Transactions on the Web, 2022, 16(3): 1-34.
- [28] LIU N, ZHANG F, GAO Q, et al. Contrastive learning with edge-wise augmentation for rumor detection[J]. International Journal of Intelligent Systems, 2024, 2024(1): 3858526.
- [29] MA J, GAO W, MITRA P, et al. Rumor detection on twitter with tree-structured recursive neural networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: Association for Computational Linguistics, 2018: 1980-1989.
- [30] DONG S, QIAN Z, LI P. Rumor detection with adversarial training and supervised contrastive learning[C]//Proceedings of 2022 International Joint Conference on Neural Networks (IJCNN). Padua, Italy: IEEE, 2022: 1-8.
- [31] ZHENG P, DOU Y, YAN Y. Sensing the diversity of rumors: Rumor detection with hierarchical prototype contrastive learning [J]. Information Processing & Management, 2024, 61(6): 103832.
- [32] 乔禹涵, 贾彩燕. 基于图自监督对比学习的社交媒体谣言检测[J]. 南京大学学报(自然科学), 2023, 59(5): 823-832.
QIAO Yuhan, JIA Caiyan. Rumor detection on social media based on graph self-supervised contrastive learning[J]. Journal of Nanjing University (Natural Sciences), 2023, 59(5): 823-832.
- [33] ZHUANG W M, CHEN C Y, LI C T. Towards robust rumor detection with graph contrastive and curriculum learning[J]. ACM Transactions on Knowledge Discovery from Data, 2024, 18(7): 1-21.
- [34] 董苏军, 钱忠, 李培峰, 等. 融合时间流特征和传播结构特征的谣言检测[J]. 中文信息学报, 2024, 38(9): 167-176.
DONG Sujun, QIAN Zhong, LI Peifeng, et al. Rumor detection based on fusion of temporal flow features and propagation structure features[J]. Journal of Chinese Information Processing, 2024, 38(9): 167-176.
- [35] CHEUNG T H, LAM K M. Causal diffused graph-transformer network with stacked early classification loss for efficient stream classification of rumours[J]. Knowledge-Based Systems, 2023, 277: 110807.
- [36] YUAN C, QIAN W, MA Q, et al. SRLF: A stance-aware reinforcement learning framework for content-based rumor detection on social media[C]//Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN). Shenzhen, China: IEEE, 2021: 1-8.
- [37] ANGRAININGSIH R, HASSAN G M, DATTA A. CE-BERT: Concise and efficient BERT-based model for detecting rumors on Twitter[J]. IEEE Access, 2023, 11: 80207-80217.
- [38] ZHAO Q, ZHANG Y, FENG X. Joint learning of structural and textual information on propagation network by graph attention networks for rumor detection[J]. Applied Intelligence, 2024, 54(3): 2851-2866.
- [39] ASKARIZADE M. Enhancing rumor detection with data augmentation and generative pre-trained transformer[J]. Expert Systems with Applications, 2025, 262: 125649.
- [40] YUAN C, MA Q, ZHOU W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection [C]//Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM). Piscataway, NJ: IEEE, 2019: 796-805.
- [41] 张鑫昕, 潘善亮, 茅琴娇. 基于传播树的多特征谣言检测方法[J]. 电子学报, 2024, 52(5): 1609-1618.
ZHANG Xinxin, PAN Shanliang, MAO Qinjiao. Multi-feature rumor detection method based on propagation tree[J]. Acta Electronica Sinica, 2024, 52(5): 1609-1618.
- [42] HE Z, LI C, ZHOU F, et al. Rumor detection on social media with event augmentations[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY: ACM, 2021: 2020-2024.

作者简介:



徐培淇(2001-),女,硕士研究生,研究方向:数据分析、深度学习、三支决策与粒计算, E-mail: peggy0081@163.com。



刘盾(1983-),通信作者,男,教授,研究方向:数据挖掘与知识发现、粗糙集与粒计算、决策支持等, E-mail: newton83@163.com。



李天瑞(1969-),男,教授,研究方向:数据挖掘与知识发现、粗糙集与粒计算、云计算与大数据等, E-mail: trli@swjtu.edu.cn。

(编辑:刘彦东)

Contrastive Learning Approach for Rumor Detection via Fusion of Dynamic Semantics and Graph Structure

XU Peiqi¹, LIU Dun^{1*}, LI Tianrui²

(1. School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China; 2. School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China)

Abstract: The rapid growth of social media has enabled rumors to spread swiftly through extensive online interactions, thereby significantly undermining public trust and destabilizing social order. However, existing rumor detection methods face notable limitations in modeling the dynamic semantic evolution of text and accurately capturing complex propagation patterns, and they often struggle to distinguish between ambiguous rumor categories. To address these challenges, we propose DySGCL (Dynamic semantic and graph feature fusion for contrastive rumor detection), a novel contrastive learning framework that fuses dynamic semantic representations with graph-based structural features. Specifically, we employ a hierarchical Transformer to extract global semantic embeddings from users' past posts, while a temporal convolutional module improves sensitivity to fine-grained semantic shifts. For structural modeling, we first simulate adversarial perturbations via edge removal, then leverage a graph attention network (GAT) to highlight critical interaction pathways in the propagation network. Finally, an integrated contrastive objective combining self-supervised and supervised signals further enhances the model's discriminative power. Experiments on the Twitter15 and Twitter16 benchmarks show that DySGCL outperforms state-of-the-art baselines by 1.8% and 2.0% in accuracy, respectively, validating its effectiveness in dynamic and complex rumor detection scenarios.

Highlights:

1. A dynamic semantic and graph-structure fusion framework is proposed for rumor detection.
2. A hierarchical Transformer and temporal convolution are integrated to capture semantic evolution in users' historical posts.
3. Self-supervised and supervised contrastive learning mechanisms are jointly used to improve the discrimination of ambiguous rumor categories.

Key words: rumor detection; dynamic semantic representation; graph neural network; contrastive learning; robustness enhancement

Foundation items: National Natural Science Foundation of China (Nos.62276217, 62402424); Fundamental Research Funds for the Central Universities (Nos.2682024KJ005, 2682024ZTPY021).

Received: 2025-06-15; **Revised:** 2025-07-02

***Corresponding author, E-mail:** newton83@163.com.