Computer Science and Application 计算机科学与应用, 2025, 15(10), 176-188
Published Online October 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.1510259

Hans汉斯

基于多特征图与GCN-GAT的谣言检测模型

黄　媛，苏庆鸥，苏敬科，刘柏霆*

广西民族师范学院数学与计算机科学学院，广西　崇左

收稿日期：2025年9月13日；录用日期：2025年10月16日；发布日期：2025年10月27日

摘　要

社交媒体中的谣言文本通常是非正式且语法不连贯的，这为准确提取语义信息带来了巨大挑战。为解决这一问题，本文提出了RoBERTa-MGAT模型，一种融合多特征图构建、并行图神经网络编码器与基于注意力的融合机制的谣言检测模型。具体而言，该模型利用RoBERTa和Word2Vec生成丰富的词向量表示，并通过构建三种异构图(词性图、词共现图和语义依存图)从不同角度捕捉多样化的语言特征。针对每个图结构，本文并行采用图卷积网络(Graph Convolutional Network, GCN)和图注意力网络(Graph Attention Network, GAT)来联合学习互补的结构化表征，GCN专注于捕获局部平滑特征，而GAT则用于建模特征异构性与重要性。最后通过自注意力机制聚合所有图的输出表示，使模型能够有效整合多视角特征。在Weibo20和Weibo21两个公开谣言检测数据集上的实验结果表明，RoBERTa-MGAT在准确率和F1分数上均优于同类型模型，展现出卓越的性能。

关键词

谣言检测，多特征图，RoBERTa，图卷积网络，图注意力网络

Rumor Detection Model Based on Multi-Feature Graphs and GCN-GAT

Yuan Huang, Qing'ou Su, Jingke Su, Boting Liu*

School of Mathematics and Computer Science, Guangxi Minzu Normal University, Chongzuo Guangxi

Received: September 13, 2025; accepted: October 16, 2025; published: October 27, 2025

Abstract

Rumor texts on social networks are often informal and grammatically incoherent, posing significant challenges for extracting accurate semantic information. To address this issue, we propose RoBERTa-MGAT, a rumor detection model that integrates multi-feature graph construction with parallel graph

*通讯作者。

文章引用: 黄媛, 苏庆鸥, 苏敬科, 刘柏霆. 基于多特征图与 GCN-GAT 的谣言检测模型[J]. 计算机科学与应用, 2025, 15(10): 176-188. DOI: 10.12677/csa.2025.1510259

neural encoders and an attention-based fusion mechanism. Specifically, the model leverages RoBERTa and Word2Vec to generate rich word embeddings and constructs three heterogeneous graphs—a part-of-speech graph, a word co-occurrence graph, and a semantic dependency graph—to capture diverse linguistic features from different perspectives. For each graph, Graph Convolutional Network (GCN) and Graph Attention Network (GAT) are applied in parallel to jointly learn complementary structural representations, with GCN capturing local smoothness and GAT modeling feature heterogeneity and importance. The outputs from all graphs are then aggregated using a self-attention mechanism, allowing the model to effectively integrate multi-view features. Experimental results on two public rumor detection datasets, Weibo20 and Weibo21, demonstrate that RoBERTa-MGAT achieves superior performance, outperforming existing state-of-the-art methods in both accuracy and F1-score.

Keywords

Rumor Detection, Multi-Feature Graph, RoBERTa, Graph Convolutional Network, Graph Attention Network

1. 引言

社交网络的快速发展极大地改变了人们获取与传播信息的方式。然而，这些平台的开放性与即时性也助长了不实信息和谣言的传播，可能对个人造成严重伤害，侵蚀公众信任，并破坏社会稳定。因此，社交媒体上的谣言检测任务越来越受到学术界与产业界的关注。

早期的谣言检测方法主要依赖传统机器学习模型和浅层语言特征，难以有效捕捉社交媒体文本中细腻的语义信息。随着深度学习的发展，Word2Vec 和 FastText 等词嵌入技术使模型能够将文本表示为连续向量空间中的稠密向量，从而捕获更丰富的语义和上下文关系。近年来，诸如 BERT [1] (双向编码器表示模型)和 RoBERTa (优化鲁棒性 BERT 预训练方法) [2]等预训练语言模型的出现显著推动了自然语言理解的发展，在包括谣言检测在内的多项下游任务中取得了显著成果。

尽管这些方法取得了进展，但仅依赖预训练模型仍不足以充分应对社交媒体谣言检测的挑战。社交媒体文本通常具有非正式、口语化、充满俚语、缩写和反讽等特点，这些特征加大了句法和语义关系建模的难度。为解决这一问题，研究者引入了图神经网络(Graph Neural Networks, GNNs)，将文本信息建模为图结构并利用词与词或词与文档之间的关系进行表示学习。例如，TextGCN [3]和 BertGCN [4]将 GCN 与预训练模型结合以提升文本分类性能。然而，现有方法大多依赖于简单的图构建(如共现图)和基础编码方式，难以充分捕捉鲁棒谣言检测所需的多样化语言特征。虽然 GCN 能够通过谱滤波有效捕捉局部邻域信息，但其平等对待所有相邻节点的特性可能在噪声较多的社交媒体数据中稀释重要信号。相比之下，图注意力网络(GAT) [5]引入了注意力机制，使模型能够为不同邻居节点分配差异化权重，实现更细粒度的特征聚合。但仅使用 GAT 又可能忽略 GCN 固有的全局结构模式。为充分发挥两种架构的互补优势，本模型在每个特征图上并行应用 GCN 与 GAT，以同步提取全局特征和上下文敏感表示。这种双路径设计增强了节点嵌入的鲁棒性与表达能力，对噪声环境中的谣言文本细微语言线索捕捉尤为有益。

为克服现有局限性，本文提出 RoBERTa-MGAT，一种新颖的多视图图神经网络模型，有效融合了丰富语义嵌入与结构化语言知识。具体而言，本研究的贡献如下：

构建了三种互补的特征图：词性图、词共现图与语义依存图，从语法、统计和语义三个维度全面建模社交媒体文本中的语言关系；

针对每个特征图，采用 GCN 与 GAT 并行的架构，同时捕捉局部结构平滑性与注意力引导的特征关联性；

引入自注意力机制动态融合多图分支的特征，使模型能够自适应地整合多视角信息。

2. 国内外研究现状

社交媒体平台的蓬勃发展极大地加快了信息传播的速度。然而，这种开放性也助长了谣言与不实信息的散播，可能引发严重的社会混乱和政治后果。因此，社交媒体谣言检测已成为近年来至关重要的研究任务。早期的谣言检测方法主要依赖传统机器学习技术(如支持向量机 SVM 和随机森林)，并采用基于词汇模式、情感极性和传播结构等人工特征进行检测。尽管这些方法在特定场景下有效，但泛化能力不足，且需要花费大量人工进行特征工程。随着深度学习的兴起，研究者开始采用神经网络自动学习文本特征。循环神经网络(Recurrent Neural Network, RNN)和卷积神经网络(Convolutional Neural Network,CNN)被用于捕捉谣言文本中的序列模式与局部特征。例如戚力鑫等人[6]提出了一种基于注意力机制的多模态融合神经网络用于谣言检测，米源等人[7]提出了一种半监督图卷积神经网络，能够有效利用未标记数据，这些方法通过减少对人工特征的依赖提升了性能，但往往难以建模复杂的长程依赖关系和上下文语义，而这对于理解社交媒体中非规范且充满噪声的语言至关重要。

近期研究开始结合用户行为和传播结构进行谣言检测，通过建模信息在用户间的传播来提升检测效果。然而，这类方法存在数据稀疏性问题，且在不同平台和语言间泛化能力较差。此外，基于传播的模型需要时序性或基于图的用户交互数据，而这些数据并非总是可用。这些局限性促使研究者日益关注如何将语义建模与结构信息相融合，以提升对社交媒体中短文本、模糊性和非正式内容的谣言检测性能。

随着 TextGCN [3]在文本分类领域取得成功，图神经网络(GNN)的文本特征提取能力得到验证。越来越多的研究者开始将 GNN 应用于谣言检测任务。例如强子珊[8]等人提出了一种基于动态传播和社区结构的谣言检测模型(Dy_PCRD)。部分研究聚焦于改进 GCN 模型结构，例如 Bian 等人[9]设计了双向图卷积网络(Bi-directional Graph Convolutional Network, Bi-GCN)，从两个方向提取社交媒体谣言文本的结构信息；王昕岩等人[10]提出了一种基于加权图卷积神经网络(Weighted-Graph Convolutional Network, W-GCN)模型的新浪微博谣言检测方法；Sun 等人[11]提出图对抗对比学习(Graph Adversarial Contrastive Learning, GACL)方法，展现出更强的谣言区分能力。Choi 等人[12]则创新性地基于动态图概念构建谣言检测模型，将谣言表示为依时间序列构建的图结构，并采用注意力机制对这些动态图进行表征学习，从而有效捕捉谣言随时间的演化特性。近年来，将 GNN 与 BERT 等大语言模型结合已成为新趋势。Ding 等人[13]采用 BERT 对谣言文本进行编码，并构建多视角图结构，利用 GCN 提取图结构特征；Thirumoorthy 等人[14]开发了融合 BERT 与 GCN 的综合性谣言检测算法。

综上所述，尽管当前研究在 GNN 应用于谣言检测方面取得了显著进展，但仍存在若干局限性。例如部分研究尝试引入注意力机制或将 GNN 与预训练语言模型结合，但少有研究会系统性地探索不同类型 GNN 架构的互补优势。具体地，GCN 通过均匀邻域聚合擅长捕捉全局图结构，而 GAT 通过自适应加权机制突出重要邻居节点，这种特性对处理社交媒体谣言文本中常见的噪声化、非正式化和多样化语言至关重要。然而现有研究大多仅单独使用 GCN，或仅松散地集成注意力机制，未能充分发挥两种方法的协同潜力。其次，现有基于图的谣言检测模型多集中于构建共现图或交互图，但往往忽略其他重要语言关系(如词性标签、语义依存等)。这些句法和语义线索对于理解社交网络谣言中常见的口语化及反讽内容具有重要价值。

为弥补这些不足，本研究提出 GCN-GAT 并行框架，联合处理多类语言学特征图(包括词性图、共现图与语义依存图)，从而在社交媒体平台上实现更强鲁棒性和更细粒度的谣言检测。

3. 方法

3.1. 模型概述

RoBERTa-MGAT 模型的架构(如图 1 所示)包含三个核心组件：嵌入层、多特征图融合层和预测层。

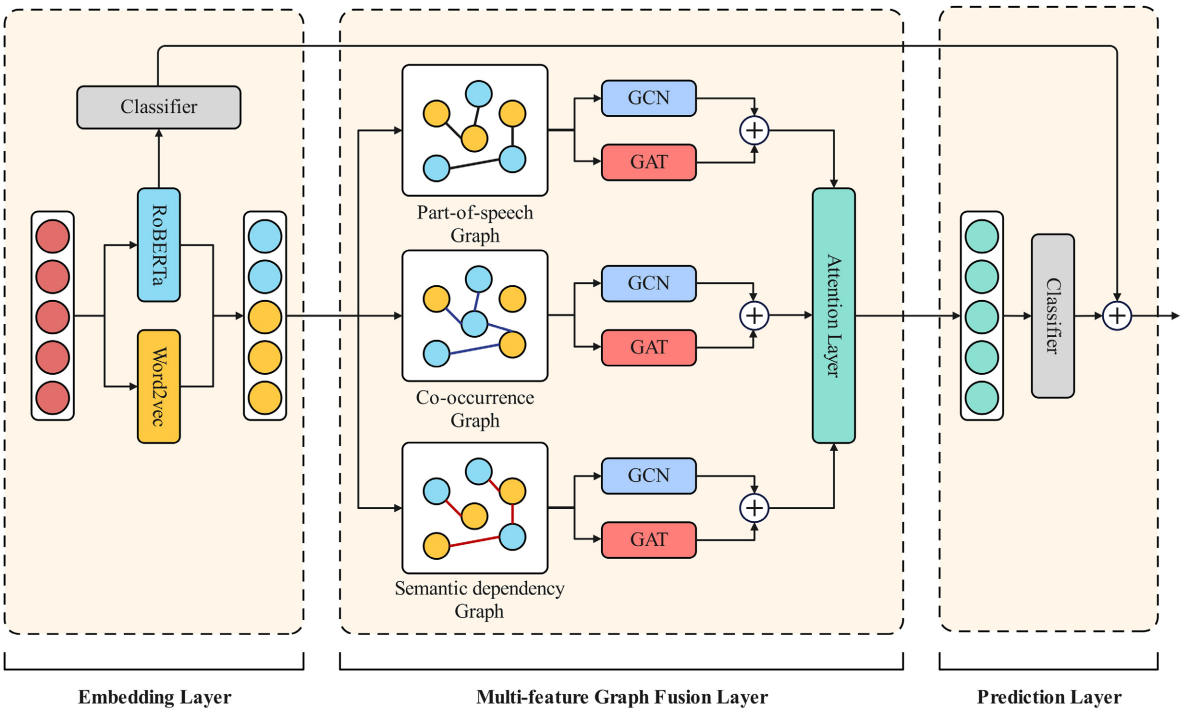

Figure 1. RoBERTa and multi-feature graph fusion for rumor detection on social network
图 1. RoBERTa 和多特征图融合用于社交网络上的谣言检测

使用 RoBERTa-MGAT 进行谣言检测的过程包括以下步骤。

Step 1：数据预处理，包括截断填充、中文分词、词性标注(Part-of-Speech, POS)、语义依存分析(Semantic Dependency, SD)以及词共现统计(Co-Occurrence, CO)。处理后的数据集被转换为可训练的迭代对象，并基于该数据集对 RoBERTa 模型进行领域适应性微调以融入领域知识。利用处理后的语言学特征，构建三种异构图结构：词性图(POS Graph)、词共现图(CO Graph)和语义依存图(SD Graph)。

Step 2：在训练前，从微调后的 RoBERTa 模型获取所有图的初始节点嵌入。训练过程中通过反向传播动态更新这些嵌入。将三种图分别输入并行的 GCN 与 GAT 模块：GCN 分支捕捉局部结构平滑性，GAT 分支学习节点级特征重要性。对每个图的 GCN/GAT 输出进行特征融合后，采用自注意力机制聚合三个图的融合表示，实现多视图特征集成。

Step 3：使用分类层从 RoBERTa 编码器和多特征图模块生成谣言预测分数。通过组合这两个组件的输出获得最终预测，从而为谣言检测生成统一的决策。

3.2. 数据预处理

在社交媒体谣言文本中，存在大量标点符号和表情符号。这些元素对谣言内容至关重要，因其常能

反映内容创作者的情感信息。因此，本研究在数据集中保留了这些符号。

如图 2 所示，本文使用 LTP [15] (Language Technology Platform)工具从谣言文本中提取词语、词性及语义依存信息。随后，采用 Transformers [16]库中的 Tokenizer 工具对谣言文本进行编码，确保其符合 RoBERTa 模型所需的输入格式。此外，将数据按特定批次大小进行分批处理。

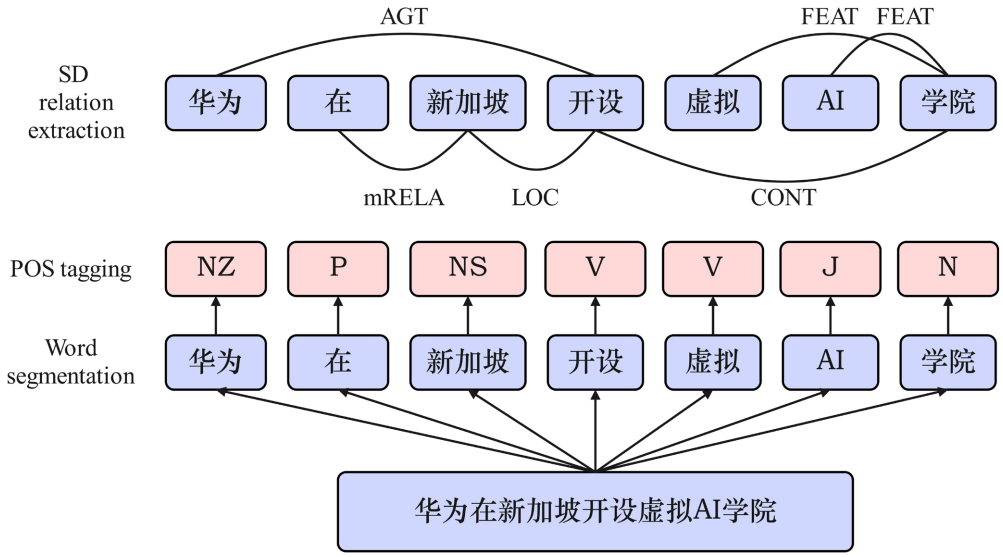

Figure 2. Based on LTP for word segmentation, part-of-speech tagging, and semantic dependency relation extraction
图 2. 基于 LTP 的分词、词性标注和语义依存关系提取

3.3. 嵌入层

传统谣言检测方法仅依赖独热编码(One-Hot Encoding)或随机初始化方式获取图节点的嵌入表示。尽管这些方法简单易实现，但难以有效捕捉词节点与文档节点的语义信息，且无法充分发挥图神经网络在提取图结构信息方面的潜力。

为解决这些局限性并获得更高质量的图嵌入表示，本文采用 Word2Vec [17]和 RoBERTa 分别对词节点与文档节点进行表征。Word2Vec 生成的嵌入能更有效捕捉文本语义信息，从而为后续谣言检测提供更丰富的特征。具体而言，Word2Vec 采用连续词袋(Continuous Bag-of-Words, CBOW)算法学习词向量，其详细计算过程如公式(1)所示：

$$L_{\text{CBOW}} = -\log P\left(w_t \middle| w_{t-c}, \cdots, w_{t+c}\right) \tag{1}$$

其中，t 代表目标词，c 表示该词的上下文窗口。通过最小化 L_{CBOW} 函数，得到该词的语义嵌入。在训练 RoBERTa-MGAT 模型前，需加载预训练的 Word2Vec 词向量来初始化词节点的嵌入表示。

对于文档节点，本文采用 RoBERTa 模型的 CLS (Classification) Token 来生成文档的上下文语义表示。CLS Token 是在 BERT 类模型输入序列首部添加的特殊 Token，作为整个序列的聚合表征。RoBERTa 作为 BERT 的增强版本，通过更大规模的预训练数据和更优化的训练策略，能够捕获更深层的语义信息。然而，由于 RoBERTa 仅基于通用语料进行预训练，其难以有效检测谣言文本的语义细微差异。为使 RoBERTa 能够生成更高质量的谣言文档节点嵌入，本文使用 Transformers 库中的 Trainer 工具对 RoBERTa 模型进行微调。最终，在开始训练流程前，需加载微调后的 RoBERTa 模型以获取文档节点的嵌入表示。

3.4. 多特征图融合层

3.4.1. 多特征图

为了增强模型有效理解谣言文本中常见的口语和网络热梗的能力，模型在词共现图的基础上加入了词性(POS)图和语义依存图。POS 图使模型能够更好地捕捉文本中的语法结构和 POS 信息，从而更准确地理解口语词在句子中的作用和功能。语义依存关系图进一步加强了模型对文本中更深层语义关系的理解。通过分析句子中单词之间的依存关系(如主谓关系、动宾关系等)，模型可以更清楚地了解口语表达中固有的逻辑结构和语义联系。

本研究的组合方法从根本上与 TextGCN 和 BertGCN 一致。采用大小为 20 的窗口来顺序扫描数据集，从而确定词－词和文档－词关系的边权重。但是对词之间边权重的计算方法进行了修改，详细信息如下。POS 图由公式(2)、公式(3)和(4)定义：

$$\text{Distance}(i,j)=\frac{W_{\text{size}}-\left|W_i-W_j\right|}{W_{\text{size}}} \tag{2}$$

$$\text{TF-IDF}=\text{IDF}\times\text{TF} \tag{3}$$

$$EP_{i,j}\begin{cases}\text{Distance}(i,j) & \text{如果 } i,j \text{ 是词},\ \text{POS}_i=\text{POS}_j\\ \text{TF-IDF} & \text{如果 } i \text{ 是文档},\ j \text{ 是词}\\ 1 & \text{如果 } i=j\\ 0 & \text{其他}\end{cases} \tag{4}$$

如公式(2)所示，其中 W_{size} 表示窗口大小，POS_i 表示词 i 的词性类型，W_i 表示词 i 在窗口内的位置。对于词性图，若两个词处于同一窗口且具有相同词性，则通过它们在窗口内的相对距离确定其间边的权重。文本中相邻出现的同词性词语通常具有强语义关联性。该公式的计算相对简单，权重值范围限定在 0 到 1 之间，可有效构建词性图的邻接矩阵。如公式(3)所示，TF-IDF (词频－逆文档频率)能有效衡量词语对文档分类的重要性，因此我们采用 TextGCN 的方法计算词－文档边的权重。

对于词共现图，本文采用与 TextGCN 完全一致的图构建方法，其正式表达式如公式(5)所示：

$$EC_{i,j}\begin{cases}\text{PMI}(i,j) & \text{如果 } i,j \text{ 是词}\\ \text{TF-IDF} & \text{如果 } i \text{ 是文档},\ j \text{ 是词}\\ 1 & \text{如果 } i=j\\ 0 & \text{其他}\end{cases} \tag{5}$$

如公式(5)所示，本文采用点互信息(Pointwise Mutual Information, PMI)来衡量两个节点间的关联强度。当 PMI 值大于 0 时，表明两个词语之间存在显著关联，因此需要在它们之间建立连接边。通过构建词共现图，利用图结构捕捉词语间复杂的全局关联关系。随后，文档节点将采用选择性聚合机制，优先整合那些最具显著性的词节点特征。

对于语义依存图，本文同样采用窗口内的相对距离公式来计算两个词语之间的边权重，具体计算方法如公式(6)所示：

$$ES_{i,j}\begin{cases}\text{Distance}(i,j) & \text{如果 } i,j \text{ 是词}\\ \text{TF-IDF} & \text{如果 } i \text{ 是文档},\ j \text{ 是词}\\ 1 & \text{如果 } i=j\\ 0 & \text{其他}\end{cases} \tag{6}$$

具有语义依存关系的词语通过边相连接。相对距离公式能够有效反映词语在句子中的真实结构关系。

通常情况下，一个词与其相邻词语之间的关系要比与距离较远词语之间的关系更强且更具显著性。随后，将词性图、词共现图和语义依存图分别输入到各自的 GCN 和 GAT 网络中，以提取各图的结构信息。

3.4.2. 并行 GCN 和 GAT

为全面捕捉图中的全局结构一致性与局部特征重要性，本文采用并行图编码器架构，使每个特征图都能被 GCN 和 GAT 独立处理。

GCN 层的定义如下：

$$H_{\mathrm{GCN}}^{l+1}=\sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{7}$$

其中，$H^{(l)}$ 表示第 l 层的节点特征矩阵，$\tilde{A}$ 为添加自环后的邻接矩阵，$\tilde{D}$ 为其度矩阵，$W^{(l)}$ 为可学习权重矩阵，$\sigma(\bullet)$ 为非线性激活函数。GCN 能有效建模节点邻域的局部平滑特性。

同时，GAT 层通过计算学习得到的注意力系数来生成节点表示，使模型能够为相邻节点分配不同的重要性权重。目标节点 i 与其邻居节点 j 之间的注意力系数通过以下方式计算：

$$a_{ij}=\frac{\exp\left(\mathrm{LeakyReLU}\left(a^{\mathrm{T}}\left[Wh_i\,\|\,Wh_j\right]\right)\right)}{\sum_{k\in N_i}\exp\left(\mathrm{LeakyReLU}\left(a^{\mathrm{T}}\left[Wh_i\,\|\,Wh_j\right]\right)\right)} \tag{8}$$

其中，h_i 和 h_j 分别表示节点 i 和节点 j 的输入特征，W 为共享线性变换矩阵，a 为可学习的注意力向量，$\|$ 表示向量拼接操作，N_i 代表节点 i 的邻居集合。

令 H_{GCN}^{l+1} 表示同层级 GAT 层的输出。本文将 GCN 与 GAT 最后一层的输出进行逐元素相加融合：

$$G^{\mathrm{final}}=H_{\mathrm{GCN}}^{(L)}+H_{\mathrm{GAT}}^{(L)} \tag{9}$$

其中，L 表示 GCN/GAT 的层数。该融合策略使模型能够在不增加特征维度的前提下，同时利用 GCN 的结构稳定性和 GAT 的自适应邻居加权优势。融合后的表征将被输入至基于注意力的多图聚合模块。

3.5. 基于注意力机制的多特征图融合

传统方法通常采用简单加法或拼接操作来聚合多特征信息。虽然这些方法简单易实现，但缺乏自适应调整不同特征权重的能力。这种局限性可能导致重要信息丢失或引入无关噪声。为解决该问题，本文提出利用自注意力机制自适应地融合多特征图，具体方法如公式(10)和(11)所示。

$$Q,K,V=\left[G_{pos},G_{co},G_{sd}\right]\times W_{qkv} \tag{10}$$

$$G_{\mathrm{merged}}=\mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{11}$$

Q (查询)、K (键)和 V (值)通过将各图的特征与可学习参数矩阵 W_{qkv} 相乘获得。随后，通过注意力机制聚合的图特征 G_{merged} 经由公式(11)计算得到。

3.6. 预测层

经由多特征图融合得到的最终输出 G_{merged} 将被输入分类器以获得谣言预测结果，具体计算过程如公式(12)和(13)所示：

$$\mathrm{Classifier}(x)=\mathrm{Softmax}\left(\mathrm{linear}(x)\right) \tag{12}$$

$$\text{prediction}_{g} = \text{Classifier}\left(G_{\text{merged}}\right) \tag{13}$$

使用PyTorch实现的线性变换(Linear)和Softmax函数，将特征维度降至类别数量并进行归一化处理。相应地，RoBERTa模型的预测公式如公式(14)所示：

$$\text{prediction}_{\text{roberta}} = \text{Classifier}\left(\text{CLS}\right) \tag{14}$$

基于RoBERTa的CLS Token表示生成谣言检测结果。RoBERTa与GCN各具优势，RoBERTa通过大规模预训练能够捕捉文本中的丰富语义信息，而GCN则利用图结构通过文档间关系促进信息和标签传播。通过融合两者的输出，可以将RoBERTa的语义理解能力与GCN的结构信息处理能力相结合，从而提升模型整体性能。因此，本研究采用公式(15)融合RoBERTa与多特征图的输出：

$$\text{prediction} = \lambda \times \text{prediction}_{g} + \left(1-\lambda\right)\text{prediction}_{\text{roberta}} \tag{15}$$

其中，参数λ用于控制两个组件的权重分配，使模型能够根据任务需求灵活调整RoBERTa和GCN的贡献度。这种平衡机制能有效发挥两种模型的优势，同时缓解依赖单一模型固有的局限性。

4. 结果

4.1. 数据集

本研究采用公开的Weibo20 [18]和Weibo21 [19]数据集作为实验数据，这些数据集均采集自新浪微博平台。数据集的相关信息如表1所示。Weibo20和Weibo21均为二分类数据集，其中谣言文本被分类为真实或虚假两个类别。数据集呈现均衡的类别分布，且谣言文本长度适中，非常适合研究者进行对比实验。对于Weibo20和Weibo21数据集，本文直接采用原论文提供的训练集、测试集和验证集划分方案。在Weibo20数据集中，每个样本均包含原始微博及其对应评论。本文将微博正文内容与其评论内容组合后，作为每个样本的完整谣言文本进行处理。

Table 1. Dataset information
表1. 数据集信息

数据集	真实	虚假	总计	平均文本长度
Weibo20	4640	4488	9128	117
Weibo21	3034	3034	6068	88

4.2. 基线模型

传统方法中，TextGCN模型利用基于图的文本表示方法，同时捕捉词语间的局部和全局上下文关系，以提高谣言检测的准确性。RoBERTaGCN模型将鲁棒优化的BERT变体RoBERTa与GCN相结合，通过融合深度上下文嵌入与图结构信息来增强谣言检测性能。RoBERTa + LSTM (Long Short-Term Memory)模型通过结合RoBERTa生成的上下文词嵌入与LSTM网络，该模型能够捕捉文本中的序列依赖关系，使其在时间敏感数据中有效检测谣言。RoBERTa + CNN模型使用RoBERTa生成深度上下文嵌入，并利用卷积神经网络(CNN)提取局部模式和特征，为谣言检测提供强大框架。

MDFEND [19]：MDFEND是一种多领域虚假新闻检测框架，通过域门控机制自适应聚合专家混合网络提取的多重表征，有效解决领域偏移问题，提升跨领域检测性能。STANKER [18]：该谣言检测模型采用双层级注意力掩码BERT作为基础编码器，并叠加稠密预测层实现分类。PMGM [13]：该模型通过整

合出版商画像与新闻文本风格特征构建多视角图，利用图卷积网络(GCN)与跳跃知识网络(Jumping Knowledge Networks, JK-Nets)生成面向出版商的新闻表征。

4.3. 实验参数

RoBERTa 模型配置为 1024 维度和 24 网络层结构。RoBERTa 的学习率设置为 1e−5，而 GCN 的学习率为 1e−3。RoBERTa-MGAT 模型训练 60 个 epoch，并保存验证集上准确率最高的模型权重用于测试。Dropout 率设为 0.1，并采用 Mish 函数[20]作为激活函数。图节点维度为 1024，GCN 隐藏层尺寸为 512。所有输入的谣言文本将被统一处理为 512 个 Token 的标准长度。

4.4. 实验结果对比

RoBERTa-MGAT 与基线模型在 Weibo20 和 Weibo21 测试集上的性能对比如表 2 所示。

Table 2. Rumor detection on Weibo20 and Weibo21 (%)
表 2. 在 Weibo20 和 Weibo21 上进行谣言检测(%)

	Weibo20		Weibo21	
方法	准确率	F1 分数	准确率	F1 分数
TextGCN	88.91	88.90	87.84	86.55
RoBERTaGCN	96.33	96.27	93.78	93.74
RoBERta + LSTM	96.13	96.13	93.05	92.97
RoBERTa + CNN	96.24	96.15	92.88	92.95
MDFEND	96.45	96.46	91.46	91.37
STANKER	97.17	97.16	93.56	93.64
PMGM	96.66	96.59	93.42	93.15
RoBERTa-MGAT	97.59	97.48	94.83	94.76

表 2 结果表明，RoBERTa-MGAT 在两个数据集上均取得了最先进的性能表现，有效验证了其在社交媒体谣言检测领域的先进性。具体而言，在 Weibo20 数据集上，该模型在准确率和 F1 分数上分别超越基线模型 0.42%和 0.32%。在 Weibo21 数据集上的性能优势更为显著，准确率和 F1 分数分别超出基线模型 1.27%和 1.12%。Weibo21 数据集包含多领域谣言内容，表 3 展示了不同模型在各领域的 F1 分数。

如表 3 所示，本研究提出的方法在全部四个实验领域均取得了最优的 F1 分数，显著超越所有基线模型。此外，本文在 Weibo21 数据集上对比了各模型的 ROC (Receiver Operating Characteristic)曲线，结果如图 3 所示。图 3 表明，RoBERTa-MGAT 获得了最高的曲线下面积(Area Under the Curve, AUC)值，这证明其在 Weibo21 数据集的谣言文本检测中具有超越基线模型的卓越能力。

Table 3. F1-score for multi-domain rumor detection on Weibo21 (%)
表 3. Weibo21 上多领域谣言检测的 F1-score (%)

方法	科学	军事	教育	意外事件	政治
MDFEND	83.01	93.89	89.17	90.03	88.65
STANKER	84.71	94.25	89.94	90.33	89.63
PMGM	84.26	94.37	89.88	90.12	89.47
RoBERTa-MGAT	85.44	95.17	90.34	90.33	89.11

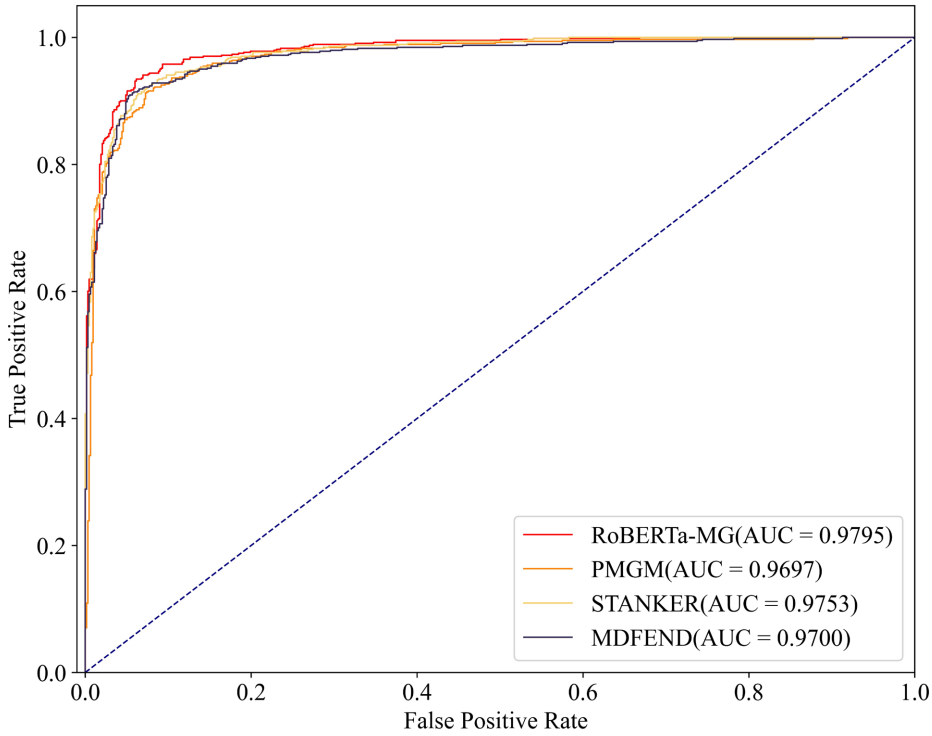

Figure 3. ROC curve on Weibo21
图 3. Weibo21 上的 ROC 曲线

4.5. 消融实验结果

为验证 RoBERTa-MGAT 各模块的有效性，本文开展了消融实验，结果如表 4 所示。表 4 明确显示，从 RoBERTa-MGAT 架构中移除任一模块都会导致 F1 分数出现系统性下降，这为这些模块在提升谣言检测精度方面的有效性提供了实证支持。

Table 4. Ablation experimental F1-score (%)
表 4. 消融实验 F1-score (%)

方法	Weibo20	Weibo21
RoBERTa-MGAT	97.48	94.76
-GAT	97.38	94.65
-ATT	97.13	94.47
-POS, -ATT	96.92	94.11
-POS, -SD, -ATT	96.27	93.74
-POS, -SD, -CO, -ATT	95.88	93.46
-RoBERTa	91.26	88.33

4.6. λ 值实验结果

参数 λ 用于量化 RoBERTa 与多特征图在联合预测中的相对重要性。为确定最佳 λ 值，本文针对不同 λ 取值进行了实验，在 Weibo21 数据集上取得的 F1 分数结果如图 4 所示。

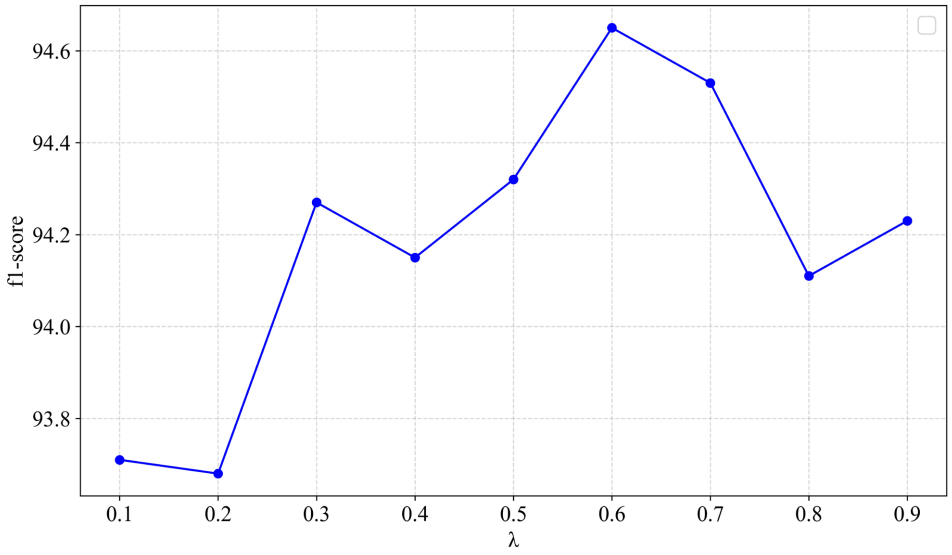

Figure 4. Relationship diagram between λ values and F1-scores on Weibo21 (%)
图 4. Weibo21 上的 λ 值和 F1 分数(%)关系图

如图 4 所示，当参数 λ 取值为 0.6 时，RoBERTa-MGAT 模型达到最佳 F1 分数，这表明模型在训练阶段对多特征图模块(Multi-feature Graph, MG)的依赖程度高于对 RoBERTa 的依赖。

4.7. 多特征图上的注意力

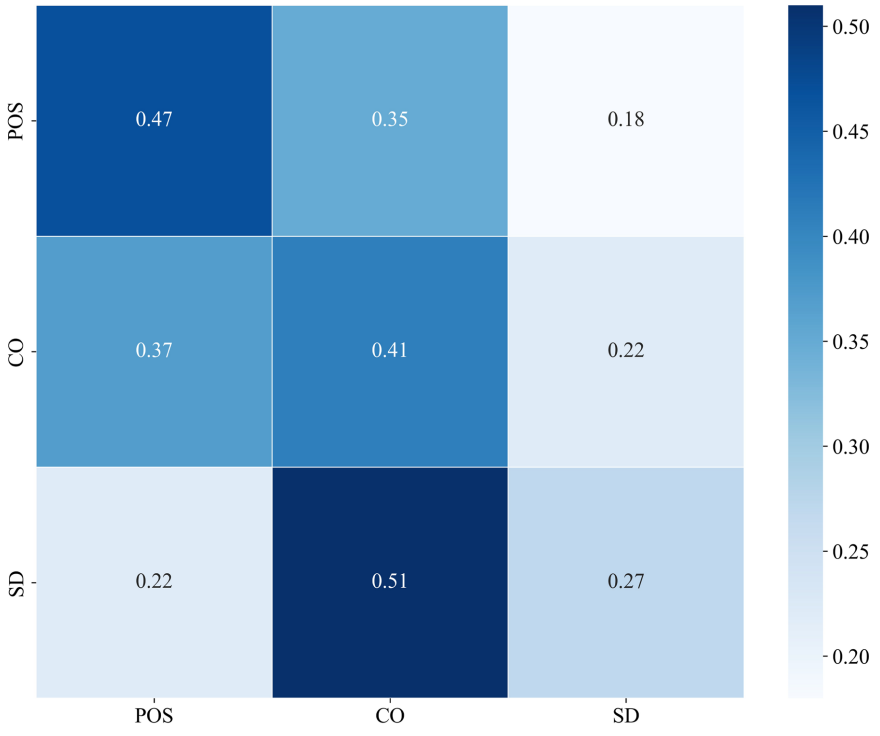

Figure 5. Attention heatmap across multi-features
图 5. 多特征的注意力热图

如图 5 所示，词性图(POS)与词共现图(CO)在谣言检测中发挥着更为重要的作用。它们为模型提供了文本的语言结构信息以及词语间的潜在关联；而语义依存图(SD)则为模型提供语义层面的细节信息。多特征图的融合共同提升了谣言检测任务的性能表现。

5. 结论

本文提出了一种基于 RoBERTa 与多特征图融合的社交媒体谣言检测模型 RoBERTa-MGAT。通过将预训练语言模型与并行图神经网络相结合，该模型显著提升了谣言检测的准确率。具体而言，创新性地构建了词性图、共现图和语义依存图三种语言特征图，分别从语法、统计和语义维度捕捉文本特征。针对每个特征图，并行应用 GCN 和 GAT 以联合学习互补的结构表征：GCN 提取局部拓扑特征，GAT 突出重要节点交互。采用自注意力机制动态融合多图特征，将 RoBERTa 的深度语义表示与全局图结构进行集成。这一设计有效解决了社交媒体文本口语化、结构松散带来的语义理解挑战。

实验结果表明，RoBERTa-MGAT 在公开数据集 Weibo20 和 Weibo21 上均取得最优性能。相较于基线模型，RoBERTa-MGAT 在 Weibo20 上的准确率和 F1 分数分别提升 0.42%和 0.32%，在更具挑战的多领域数据集 Weibo21 上分别提升 1.27%和 1.12%。消融实验验证了各模块的贡献：移除并行 GCN-GAT 编码器、自注意力机制或多特征图均会导致性能持续下降，证实了 RoBERTa、图编码器与融合策略协同作用的重要性。参数敏感性分析表明，模型显著受益于丰富图特征的引入，凸显了多图建模在复杂语义场景中的价值。

本研究的实践价值在于为社交平台提供了高效且适应性强的内容治理方案。通过精准检测误导性内容，该模型有助于维护网络信息生态的完整性。然而仍存在一定局限性：当前实验仅基于中文微博数据，其跨语言迁移能力尚未验证；多图构建与双路径编码架构引入了计算开销，需进一步优化以实现实时部署。未来工作可探索轻量化设计或知识蒸馏技术，以提升模型在大规模应用中的推理效率，并将模型应用于英文谣言检测数据集，以检验模型框架的跨语言泛化能力。

基金项目

广西民族师范学院校级科研基金项目(2024YB124)。

参考文献

[1] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2018) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv: 1810.04805.

[2] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V. (2019) RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv: 1907.11692.

[3] Yao, L., Mao, C. and Luo, Y. (2019) Graph Convolutional Networks for Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 7370-7377. https://doi.org/10.1609/aaai.v33i01.33017370

[4] Lin, Y., Meng, Y., Sun, X., Han, Q., Kuang, K., Li, J. and Wu, F. (2021) BERTGCN: Transductive Text Classification by Combining GCN and BERT. arXiv: 2105.05727.

[5] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y. (2017) Graph Attention Networks. arXiv: 1710.10903.

[6] 戚力鑫, 万书振, 唐斌, 徐义春. 基于注意力机制的多模态融合谣言检测方法[J]. 计算机工程与应用, 2022, 58(19): 209-217.

[7] 米源, 唐恒亮. 基于图卷积网络的谣言鉴别研究[J]. 计算机工程与应用, 2021, 57(13): 161-167.

[8] 强子珊, 顾益军. 融合动态传播和社区结构的社交媒体谣言检测模型[J]. 计算机工程与应用, 2024, 60(18): 198-207.

[9] Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., *et al*. (2020) Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 549-556.

https://doi.org/10.1609/aaai.v34i01.5393

[10] 王昕岩, 宋玉蓉, 宋波. 一种加权图卷积神经网络的新浪微博谣言检测方法[J]. 小型微型计算机系统, 2021, 42(8): 1780-1786.

[11] Sun, T., Qian, Z., Dong, S., Li, P. and Zhu, Q. (2022) Rumor Detection on Social Media with Graph Adversarial Contrastive Learning. *Proceedings of the ACM Web Conference* 2022, Lyon, 25-29 April 2022, 2789-2797. https://doi.org/10.1145/3485447.3511999

[12] Choi, J., Ko, T., Choi, Y., Byun, H. and Kim, C. (2021) Dynamic Graph Convolutional Networks with Attention Mechanism for Rumor Detection on Social Media. *PLOS ONE*, **16**, e0256039. https://doi.org/10.1371/journal.pone.0256039

[13] Ding, X., Teng, C. and Ji, D. (2023) Fake News Detection with Context Awareness of the Publisher. *International Conferences on Software Engineering and Knowledge Engineering*, South San Francisco, 1-10 July 2023, 548-553. https://doi.org/10.18293/seke2023-061

[14] Thirumoorthy, K., Haripriya, R., Shreenee, N., *et al*. (2024) Rumor Detection Using BERT-Based Social Circle and Interaction Network Model. *Social Network Analysis and Mining*, **14**, Article No. 195. https://doi.org/10.1007/s13278-024-01362-2

[15] Che, W., Feng, Y., Qin, L. and Liu, T. (2021) N-LTP: An Open-Source Neural Language Technology Platform for Chinese. *Proceedings of the* 2021 *Conference on Empirical Methods in Natural Language Processing*: *System Demonstrations*, Online, 7-11 November 2021, 42-49. https://doi.org/10.18653/v1/2021.emnlp-demo.6

[16] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., *et al*. (2020) Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the* 2020 *Conference on Empirical Methods in Natural Language Processing*: *System Demonstrations*, Online, 16-20 November 2020, 38-45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

[17] Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013) Efficient Estimation of Word Representations in Vector Space. arXiv: 1301.3781.

[18] Rao, D., Miao, X., Jiang, Z. and Li, R. (2021) Stanker: Stacking Network Based on Level-Grained Attention-Masked BERT for Rumor Detection on Social Media. *Proceedings of the* 2021 *Conference on Empirical Methods in Natural Language Processing*, Punta Cana, 7-11 November 2021, 3347-3363. https://doi.org/10.18653/v1/2021.emnlp-main.269

[19] Nan, Q., Cao, J., Zhu, Y., Wang, Y. and Li, J. (2021) MDFEND: Multi-Domain Fake News Detection. *Proceedings of the* 30*th ACM International Conference on Information & Knowledge Management*, Queensland, 1-5 November 2021, 3343-3347. https://doi.org/10.1145/3459637.3482139

[20] Misra, D. (2019) Mish: A Self Regularized Non-Monotonic Activation Function. arXiv: 1908.08681.