基于改进 GAT 的多特征融合谣言检测模型 MFLAN

马满福，陈嘉豪*，李勇，张聪
（西北师范大学计算机科学与工程学院，甘肃 兰州 730070）

摘 要： 传统的图神经网络（GNN）模型缺乏对节点之间复杂关系的建模能力，且对大规模图的处理能力较弱，无法有效地从大规模图中提取代表性的子图，由此导致训练和推理的精确率不高。为此，提出一种基于改进图注意力网络（GAT）的多特征融合谣言检测模型 MFLAN。首先，MFLAN 通过加入带有注意力机制的特征融合方法，为每个特征赋予不同的权重，对原始特征进行加权求和操作，获得融合后的特征向量；其次，加入正值位置编码，使模型可以获取位置信息表示；然后，引入可学习的参数矩阵，使得模型在训练过程中自动地学习和优化参数值；最后，对注意力分数进行稀疏化，将大规模图中部分不重要节点的注意力置为 0，以此实现 MFLAN 模型。实验结果表明，MFLAN 模型在 Ma-Weibo 和 Weibo23 数据集上的准确率分别达到 97.71%和 97.10%，相较于 Dir-GNN 模型分别提升 1.07%和 1.12%，与其他谣言检测模型相比，MFLAN 各项性能指标均有提升。

关键词： 谣言检测；新浪微博；信息传播；特征融合；图注意力网络
中图分类号： TP18 **文献标志码：** A **DOI：** 10.19678/j.issn.1000-3428.0069383

Multi-Feature Fusion Rumor Detection Model MFLAN Based on Improved Graph Attention Network

MA Manfu, CHEN Jiahao*, LI Yong, ZHANG Cong
(College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, Gansu, China)

【Abstract】 Traditional Graph Neural Network (GNN) models cannot model complex relationships between nodes and exhibit weak performance in handling large-scale graphs. They are unable to effectively extract representative subgraphs from large-scale graphs, which consequently leads to a low accuracy in both training and inference processes. To address these issues, a rumor detection model called MFLAN, based on an improved Graph Attention Network (GAT), is proposed, which incorporates a feature fusion method equipped with an attention mechanism. This approach assigns different weights to each feature and performs a weighted summation operation on the original features to obtain a fused feature vector. Additionally, a positive positional encoding is introduced, which enables the model to capture positional information representations. Subsequently, learnable parameter matrices are incorporated, enabling the model to automatically learn and optimize parameter values during the training process. Finally, the attention scores are sparsified by setting the attention weights of certain unimportant nodes in the large-scale graph to zero, thereby completing the construction of the MFLAN model. Experimental results demonstrate that the proposed model achieves an accuracy of 97.71% on the Ma-Weibo dataset and 97.10% on the Weibo23 dataset. Compared to the accuracies achieved by the Dir-GNN model, these represent improvements of 1.07% and 1.12%, respectively. Compared to other rumor detection models, MFLAN exhibits superior performance across all evaluation metrics.

【Key words】 rumor detection; Sina Weibo; information dissemination; feature fusion; Graph Attention Network (GAT)

0 引言

传统的谣言检测方法主要基于用户特征、文本内容、传播模式等特征来进行训练，如决策树、随机森林、支持向量机等机器学习方法[1]。然而，由于社交媒体上的信息量巨大、内容多样、真假难辨，导致此类检测方法不论在效率还是在效果上都难以满足实际需求[2-3]。为解决该问题，一些深度学习方法通过从谣言的传播方式中发现谣言[4]，其中，图神经网络（GNN）在谣言检测任务中取得了良好的效果。然而，图神经网络也存在一定的局限性：首先，谣言传播图中的节点具有多样化的属性特征，由于缺乏有效的特征提取方法，模型无法获取全面的节点信息表示；其次，传统谣言检测方法通常忽略词向量中

收稿日期： 2024-02-21 **修回日期：** 2024-04-12
基金项目： 国家自然科学基金（72364033）；甘肃省科技计划项目（23JRZA397）；西北师范大学高校重大科研项目培育计划项目（NWNU-LKZD2021-06）。
通信作者 E-mail： * rjcjh1999@163.com

的位置信息,无法充分进行语义理解,导致模型在处理长句子或者复杂语境时效果不佳;最后,模型受限于固定的结构和特征表达方式,难以灵活适应不同数据特点和任务要求,且在谣言检测任务中真实数据通常具有稀疏性,存在大量无关信息,影响计算效率和性能。

针对以上问题,本文提出一种谣言检测模型 MFLAN。通过引入正值位置编码来补充位置信息,提高模型对句子结构和语义的理解能力,加入可学习的参数矩阵,更好地捕捉节点之间的复杂关系,并在注意力分数计算时引入稀疏化技术,降低计算的复杂性,从而更好地处理大规模图。本文主要贡献如下:

1)基于用户个人信息、文本情感特征、文本内容进行多特征融合。

2)提出引入位置编码、可学习参数矩阵和稀疏化技术的谣言检测模型 MFLAN。

3)构建一个新的微博谣言检测数据集 Weibo23。

1 相关工作

1.1 基于特征提取的方法

文献[5]通过关注基于用户、内容和词汇的特征以及 post 序列,提出了一个使用各种基本特征并结合 2 个深度学习模型的框架,将词嵌入与双向长短期记忆(BiLSTM)相结合,并使用多层感知器(MLP)与后置特征相结合,以提高谣言检测任务的准确性。文献[6]通过设计新的特征,研究新特征对任务的影响,从而提升检测性能。文献[7]基于谣言生命周期,提出了使用时间序列来捕捉时间特征的新方法,并通过时间序列建模融合各种社会背景信息。

1.2 基于社交网络结构的方法

文献[8]分析了谣言的传播情况,推断出不同类别和不同价值的谣言的上传和转发率,通过检查包含 Snopes 文章链接的评论的个人转发对级联演变的影响,发现收到这样的评论会增加转发谣言被删除的可能性。文献[9]提出了一个通过分析信息传播路径来进行早期检测的模型,该模型将信息的传播路径建模为多元时间序列,用户的特征为每个元组,通过基于循环和卷积网络的分类器来捕获沿传播路径的用户特征的全局和局部变化,以检测假信息。文献[10]提出了一种基于内核的方法,称为传播树内核,该方法通过评估传播树结构之间的相似性来捕获区分不同类型谣言的高阶模式。文献[11]将谣言的传播过程描述为一个或多个级联,将其定义为谣言传播模式的实例,对谣言级联的传播进行分类。

1.3 基于深度学习的方法

文献[12]使用了图神经网络来捕捉文本内容不同粒度的高级表示,并融合了谣言传播结构。文献[13]提出了一种传播树结构的双注意网络模型 DAN-Tree,该模型设计了节点和路径双注意机制,将谣言传播结构的深度结构和语义信息进行有机融合,并采用路径过采样和结构嵌入来增强深度结构的学习,从而进一步提升模型性能。文献[14]提出了一种基于句法依赖关系和多层交互网络的谣言检测模型。文献[15]提出了一种新颖的双向图模型 Bi-GCN,通过自上而下和自下而上的谣言传播来探索其深度和广度这 2 个特征。文献[16]提出了一种混合深度神经模型 CanarDeep,该模型使用基于上下文(文本+元特征)和基于用户的特征学习,结合分层注意网络(HAN)和 MLP 进行预测。文献[17]提出了一种基于循环神经网络(RNN)的深度关注模型,用于选择性地学习连续帖子的时间隐藏表示,以识别谣言,该研究主要关注发帖的时间序列关系。

综上,基于特征提取的方法侧重于从文本中提取各种特征来提高模型的效果,基于社交网络结构的方法侧重于分析谣言传播的路径、节点的影响力、社区结构等,基于深度学习的方法则使用深度神经网络来学习文本和其他数据的表示,这些都对谣言检测模型的性能提升有所帮助。本文通过添加位置编码,为模型提供位置信息,使其更好地理解文本的语义,通过改进图注意力网络(GAT)模型,引入可学习的参数矩阵,使模型能够通过训练数据学习节点之间的关系,避免过度依赖手工设计的规则或固定的邻接矩阵。此外,在计算注意力分数时,引入稀疏化技术,使模型能充分利用相关信息,提高计算精度和效率,更好地适应复杂的社交网络结构。

2 MFLAN 谣言检测模型

2.1 谣言检测

谣言检测任务可被定义为:设 $P=\{P_1,P_2,\cdots,P_n\}$ 表示数据集中一系列的事件集合,其中 P_i 表示第 i 个事件,$i\in\{1,2,\cdots,n\}$,n 表示数据集中事件总数量。每个社交媒体事件 $P_i\in P$ 都包含一条微博的原帖和若干条转发的帖子,即 $P_i=\{p_i, r_1^i, r_2^i, \cdots, r_n^i, G_i\}$,其中,$p_i$ 表示原帖,r_j^i 表示关于 p_i 的第 j 个转发的帖子,G_i 表示事件 P_i 的传播路

径。每个事件 P_i 都附带一个标签 $y_i \in \{0,1\}$，其中，$y_i=0$ 表示该事件为非谣言，$y_i=1$ 表示该事件为谣言。

定义社交网络传播路径图 $G_i=(V_i,E_i)$，其中，节点集合表示为 $V_i=\{p_i, r_1^i, r_2^i, \cdots, r_n^i\}$，$p_i$ 是图 G_i 的根节点，表示事件 i 的原帖，r^i 为图 G_i 的子节点，表示针对事件 i 的转发帖子。$E_i=\{e_j^i \mid j=0,1,\cdots,n_{i-1}\}$ 是目标节点的边集合，表示帖子之间的转发关系。如果 r_1^i 是 p_i 的一个转发帖子，那么该帖子和原帖之间存在一条有向边 e_{01}^i，由此来定义邻接矩阵 $\boldsymbol{A}_i \in \{0,1\}^{n_i \times n_i}$，表示图中的传播路径，其中元素表示为：

$$a_j^i=\begin{cases}1, & e_{uv}^i \in E_i \\ 0, & \text{其他}\end{cases} \tag{1}$$

谣言检测任务的目标是学习一个分类器，该分类器可以用输入 X_i 到输出空间 Y_i 之间的映射 f 来表示，记作：

$$f: X_i \rightarrow Y_i \tag{2}$$

2.2 模型框架

MFLAN 模型框架如图 1 所示(彩色效果见《计算机工程》官网 HTML 版，下同)，其由 5 个模块组成。

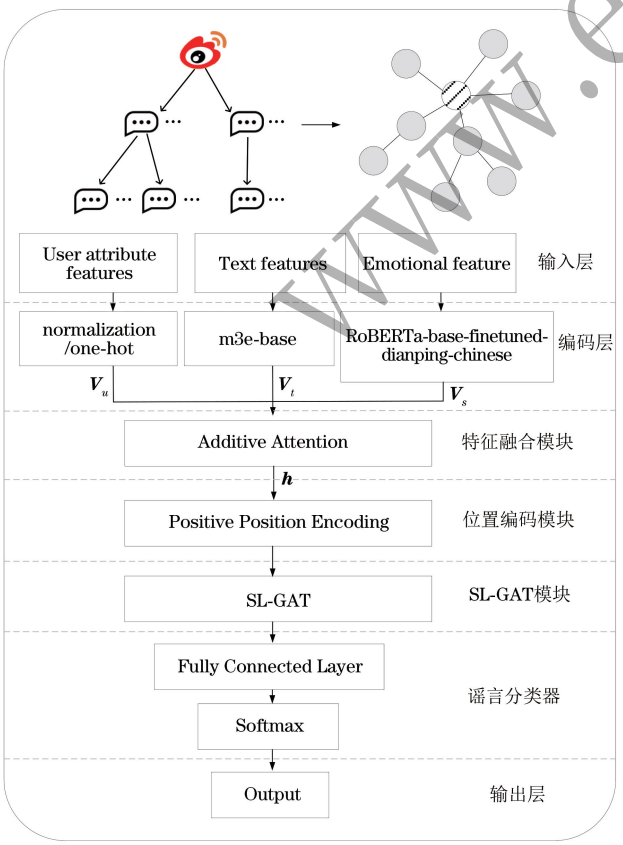

图 1 MFLAN 模型框架

Fig. 1 MFLAN model framework

1)特征提取。使用 m3e-base[18] 模型对帖子文本信息进行文本表示学习，得到文本特征向量表示，使用基于 RoBERTa-base 模型进行微调的中文大语言模型对帖子情感特征进行特征向量表示。

2)特征融合。对不同的输入特征进行加权处理，形成最终的融合输入特征向量。

3)位置编码。为输入序列中的每个位置提供信息，计算位置编码，与节点特征相加，从而将位置信息与节点信息融合到一起。

4)SL-GAT。利用基于改进 GAT 的图神经网络方法对谣言事件进行节点特征的聚合与更新，获取谣言事件图级别的特征表示。

5)谣言分类器。将图级别特征表示输入前馈神经网络中，得到谣言事件的预测标签。

2.3 特征提取

在提取用户属性特征时，选取部分重要的用户属性特征，如性别、用户的互粉数、用户所在省市、关注数、粉丝数等，根据用户属性特征的取值范围构建特征向量空间。对于每个用户的性别及所在省市的属性特征，由于其为离散变量，将其转换为 one-hot 编码。对于其他属性特征，以粉丝数为例，知名博主的粉丝数会远大于一般用户，若只进行简单的最大归一化、最小归一化，可能会造成大量用户在该维度的特征值趋近于 0，为避免这种情况，采取先取对数的方式减小最大值，对最大值、最小值进行归一化处理，如式(3)所示：

$$y_n=\begin{cases}\dfrac{\log_a x-\log_a x_{\min}}{\log_a x_{\max}-\log_a x_{\min}}, & x>0 \\ 0, & x=0\end{cases} \tag{3}$$

式中：x 表示归一化前的用户粉丝数量；y_n 表示归一化后的值；$x_{\max}$ 为用户粉丝数最大值；$x_{\min}$ 为用户粉丝数最小值。当用户的粉丝数为 0 时，$x_{\min}$ 的值为 0。其余连续特征的处理方式相同，最后将所有用户属性特征编码进行合并，形成最终的用户属性特征向量。

在进行文本特征提取时，首先，将输入的文本进行分词处理，将文本切分成词语的序列，为了捕捉每个词语在句子中的位置信息，m3e-base 会为序列中每个词语的位置进行编码；接下来，将每个词语的分词编码转换为密集的词嵌入向量，使用多层编码器处理词嵌入向量序列；然后，该模型会对原始词嵌入向量与编码器输出向量进行相加，获取上下文感知的词嵌入向量并输出，输出的向量维度为 768。与传统词嵌入模型生成的静态词嵌入向量不同，本文模型生成的词嵌入向量包含了上下文信息。

在对微博帖子进行情感特征编码时，使用基于 RoBERTa-base 模型微调的中文大语言模型对帖子文本进行情感分析，该模型在京东完整数据集、大众点评数据集等 5 个中文用户评论数据集上进行预训练，通过学习这些文本中的情感相关特征，包括学习文本的上下文、词语之间的关系，从文本表示中提取情感特征，将提取的情感特征向量作为输入，训练一个情感分类器模型。最后，对输入文本进行预测，为其生成对应的情感特征向量。

2.4 基于文本信息的多特征融合

通过特征提取，将帖子文本内容生成高维向量，但文本情感特征向量维度较低，通过注意力机制来进行加权特征融合，可以将多个特征合并为综合的特征向量，从而减少特征维度，简化模型的训练和推理过程。具体的特征融合步骤如下：

1)定义注意力权重，用于衡量不同特征的重要性。本文使用加性注意力机制来计算注意力权重。

2)计算注意力分数。使用加性注意力机制计算不同特征之间的注意力分数，计算公式如下：

$$A_{\text{score}}^{\text{Add}}=\tanh(\boldsymbol{w}_1\cdot\boldsymbol{V}_u+\boldsymbol{w}_2\cdot\boldsymbol{V}_t+\boldsymbol{w}_3\cdot\boldsymbol{V}_s) \tag{4}$$

式中：$A_{\text{score}}^{\text{Add}}$ 表示融合后的注意力分数；$\boldsymbol{w}_1$、$\boldsymbol{w}_2$、$\boldsymbol{w}_3$ 表示注意力权重的可学习参数矩阵；$\boldsymbol{V}_u$ 表示用户属性特征向量；$\boldsymbol{V}_t$ 表示文本特征向量；$\boldsymbol{V}_s$ 表示情感特征向量。

3)对计算得到的注意力分数进行归一化处理。使用 Softmax 函数对注意力分数进行归一化处理，计算公式如下：

$$w_{\text{add}}=\text{Softmax}(A_{\text{score}}^{\text{Add}}) \tag{5}$$

式中：w_{add} 表示归一化后的注意力分数。

4)根据归一化后的注意力分数，对不同特征的特征向量进行加权融合。每个特征向量乘以对应的注意力权重，并将它们加权求和，得到最终的特征向量 $\boldsymbol{h}$，计算公式如下：

$$\boldsymbol{h}=w_{\text{add}}[0]\cdot\boldsymbol{V}_u+w_{\text{add}}[1]\cdot\boldsymbol{V}_t+w_{\text{add}}[2]\cdot\boldsymbol{V}_s \tag{6}$$

最终得到注意力加权特征融合后的特征向量，作为后续模型的输入，多特征融合的具体过程如图 2 所示。

2.5 正值位置编码

由于注意力机制不受节点顺序的影响，因此它没有任何关于目标节点在序列中位置的概念。为了避免这种情况并使模型可以感知位置信息，本文通过加入正值位置编码，帮助模型更好地理解节点间的空间关系。

正值位置编码是一种绝对位置编码，相较于其

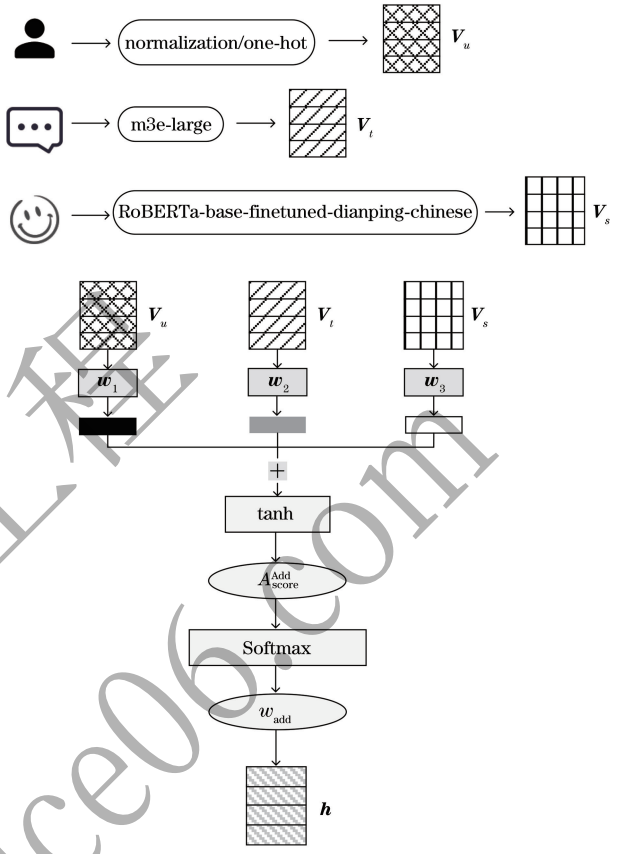

图 2 多特征融合模块

Fig. 2 Multi-feature fusion module

他位置编码方法，其不需要学习参数，因此较为简洁。正值位置编码的计算方式如下：

$$\begin{cases}P_{(\text{pos},2i)}^{\text{PPE}}=\dfrac{1}{2}\times\sin\left(\dfrac{p_{\text{pos}}}{10\ 000^{\frac{2i}{d_{\text{model}}}}}\right)+\dfrac{1}{2}\\[2ex]P_{(\text{pos},2i+1)}^{\text{PPE}}=\dfrac{1}{2}\times\cos\left(\dfrac{p_{\text{pos}}}{10\ 000^{\frac{2i}{d_{\text{model}}}}}\right)+\dfrac{1}{2}\end{cases} \tag{7}$$

式中：$P_{(\text{pos},2i)}^{\text{PPE}}$ 和 $P_{(\text{pos},2i+1)}^{\text{PPE}}$ 分别表示序列中位置为第 p_{pos} 行、第 $2i$ 列和第 p_{pos} 行、第 $2i+1$ 列的正值位置编码值；d_{model} 为维度。通过式(7)为每个行向量生成 d_{model} 维的位置向量。

正值位置编码可以表示序列间各节点的相对位置信息，根据三角函数式(8)：

$$\begin{cases}\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta\\\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta\end{cases} \tag{8}$$

$P_{(\text{pos}+k,2i)}^{\text{PPE}}$ 可由 $P_{(\text{pos},2i)}^{\text{PPE}}$ 线性表示为：

$$\begin{cases}P_{(\text{pos}+k,2i)}^{\text{PPE}}=P_{(\text{pos},2i)}^{\text{PPE}}\times P_{(k,2i+1)}^{\text{PPE}}+\\P_{(\text{pos},2i+1)}^{\text{PPE}}\times P_{(k,2i)}^{\text{PPE}}\\P_{(\text{pos}+k,2i+1)}^{\text{PPE}}=P_{(\text{pos},2i+1)}^{\text{PPE}}\times P_{(k,2i+1)}^{\text{PPE}}-\\P_{(\text{pos},2i)}^{\text{PPE}}\times P_{(k,2i)}^{\text{PPE}}\end{cases} \tag{9}$$

令 $\omega_i=\dfrac{1}{10\ 000^{\frac{2i}{d_{\text{model}}}}}$，则：

$$\begin{pmatrix} P^{\mathrm{PPE}}_{(\mathrm{pos}+k,2i)} \\ P^{\mathrm{PPE}}_{(\mathrm{pos}+k,2i+1)} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\cos(k\omega_i)+\frac{1}{2} & \frac{1}{2}\sin(k\omega_i)+\frac{1}{2} \\ -\frac{1}{2}\sin(k\omega_i)-\frac{1}{2} & \frac{1}{2}\cos(k\omega_i)+\frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} P^{\mathrm{PPE}}_{(\mathrm{pos},2i)} \\ P^{\mathrm{PPE}}_{(\mathrm{pos},2i+1)} \end{pmatrix} \tag{10}$$

由式(10)可知，位于 $p_{\mathrm{pos}+k}$ 的位置编码可以由位于 p_{pos} 的位置编码表示，且其恒为非负的性质与后续模型中的 ReLU 函数相契合，不存在值域的冲突，可以很好地表示节点间的位置关系。正值位置编码的曲线表示如图 3 所示。

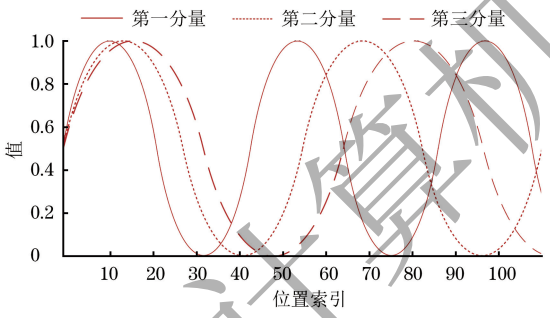

图 3 正值位置编码曲线

Fig. 3 Positive position coding curve

2.6 SL-GAT 模块

GAT 通过注意力机制学习节点的表示，在 GAT 中，每个节点的表示是通过对其邻居节点进行加权求和而得到的，权重根据标签信息进行学习，这使得模型表示受限于标签信息。在大规模图中计算注意力分数有较高的复杂性，且较难捕捉图的高级结构信息。通过引入可学习的参数矩阵，可以更好地捕捉节点之间的复杂关系，并在注意力分数计算时引入稀疏化技术，降低计算的复杂性，从而更好地处理大规模图。该模型的过程如下：

假设一个节点特征矩阵为 $\boldsymbol{V} \in R^{N\times F}$，其中，$N$ 是节点的数量，F 是节点的输入特征数量。令 $\boldsymbol{h}_i$ 为节点 i 的输入向量，本文引入一个可学习的参数矩阵 $\boldsymbol{U} \in R^{F\times F'}$，其中，$F$ 是输入特征的维度，F' 是输出特征的维度。

首先，计算特征变换后的特征向量 $\boldsymbol{h}_i'$，计算公式为 $\boldsymbol{h}_i'=\boldsymbol{U}\boldsymbol{h}_i$，该节点与其邻居节点间的注意力系数使用 ReLU 函数进行计算，计算公式如下：

$$e_{ij}=\mathrm{ReLU}(\vec{\boldsymbol{a}}^{\mathrm{T}}[\boldsymbol{U}\boldsymbol{h}_i \,\|\, \boldsymbol{U}\boldsymbol{h}_j]) \tag{11}$$

式中：$\vec{\boldsymbol{a}} \in R^{2F'}$ 是一个可学习的参数向量；“‖”表示拼接操作。

其次，对其进行稀疏化处理，得到稀疏化后的注意力权重向量集合 α_{ij}，计算公式如下：

$$e'_{ik}=\frac{\max(0,e_{ik}-\tau)}{\sum_{j\in \sup(e_i)}\max(0,e_j-\tau)},k \in N(i)\bigcup\{i\} \tag{12}$$

$$\alpha_{ij}=\{e'_{i1},e'_{i2},\cdots,e'_{ik}\},\ k \in N(i)\bigcup\{i\} \tag{13}$$

式中：e_{ik} 为由节点 i 的特征与邻居节点 k 的特征经过权重矩阵变换后的内积计算得到的未归一化分数；e'_{ik} 为稀疏化后的注意力权重分数；$\sup(e_i)$ 是 e_i 的支持集合，包含所有使 $e_j>\tau$ 的索引 j，其中 τ 为一个阈值；$\max(0,u_i-\tau)$ 为截断项，确保输出非负。$\sum_{j\in \sup(u)}\max(0,u_j-\tau)$ 是对截断项进行归一化的项，确保输出总和为 1，这样将注意力权重中小于阈值的部分置为 0，从而剔除了与节点关联度较低的邻居节点。

然后，计算加权和，得到最终的上下文特征向量 $\boldsymbol{h}_i'$，计算公式如下：

$$\boldsymbol{h}'_i=\sigma\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j\in N(i)\bigcup\{i\}}\alpha_{ij}^{t}\boldsymbol{U}^{t}\boldsymbol{h}_j\right) \tag{14}$$

式中：σ 为激活函数，本文使用 ReLU 函数；T 为注意力头数，本文取值为 4。

最后，对更新后的节点特征向量进行池化处理，利用每个节点计算出的注意力权重分数，使用 Top K 函数选择前 K 个分数最高的节点，保留其节点索引 perm，同时更新节点特征 $\boldsymbol{h}''$，令 $\boldsymbol{h}''=\boldsymbol{h}'_{\mathrm{perm}}\cdot\sigma(w_{\mathrm{perm}})$，其中，$\boldsymbol{h}'_{\mathrm{perm}}$ 表示保留后的节点特征，w_{perm} 表示保留后各节点的权重。根据被保留的节点索引 perm 更新图的边索引 e'_{index}，通过布尔掩码将不在 perm 中的边索引置为 0，确保只有选定的节点之间的边被保留。对更新后的节点特征 $\boldsymbol{h}''$ 进行全局最大池化和全局平均池化处理，对池化结果进行拼接，得出整个图的全局表示 $\boldsymbol{h}_p$：

$$\boldsymbol{h}_p=\mathrm{concat}(\mathrm{GMP}(\boldsymbol{h}''),\mathrm{GAP}(\boldsymbol{h}'')) \tag{15}$$

SL-GAT 模块的具体结构如图 4 所示。

2.7 谣言分类器

将 $\boldsymbol{h}_p$ 输入全连接层和一个 Softmax 层：

$$\hat{y}=\mathrm{Softmax}(\mathrm{FC}(\boldsymbol{h}_p)) \tag{16}$$

式中：$\hat{y} \in R^{1\times C}$ 是预测谣言事件分类标签的概率分布；C 表示谣言类别的数量；FC(•)表示全连接层。

利用数据的真实标签计算预测值，并使用交叉熵损失函数：

$$l_{\mathrm{loss}}=-w_r y_i \log_a \hat{y}_i - w_n(1-y_i)\log_a(1-\hat{y}_i) \tag{17}$$

式中：l_{loss} 表示损失值；w_r 表示谣言样本的权重；w_n 表示非谣言样本的权重。通过对 l_{loss} 最小化来实现目标优化。

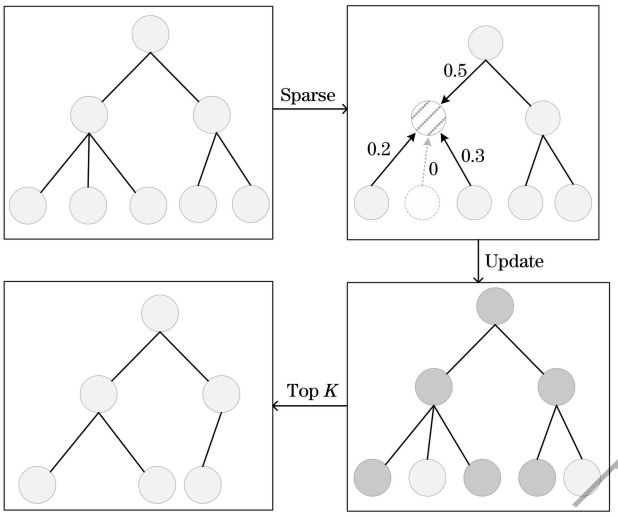

图 4 SL-GAT 模块

Fig. 4 SL-GAT module

3 实验分析

3.1 数据集

Ma-Weibo 数据集是由文献[19]在 2016 年收集并公布的新浪微博数据集，是目前谣言检测领域使用最多的中文谣言数据集。谣言的产生和传播是一个不断演变的过程，受到社会变革和技术发展等因素的影响，其内容、形式和传播渠道会发生变化，开放的数据集通常是在特定时期内收集的，其内容通常受到一定的时效性限制，而最新的数据和趋势对于深入分析谣言现象至关重要。因此，本文爬取了 2023 年最新的微博谣言数据，将其作为 Ma-Weibo 数据集的重要补充，有助于更全面地了解谣言在社交媒体上的表现和传播特征。本文基于新浪微博及微博社区管理平台，爬取了 2023 年的真实微博数据，创建了一个全新的微博数据集 Weibo23，该数据集涵盖了 3 572 个事件，涉及 148 776 名用户。在这些事件中，包括了 1 748 件谣言事件和 1 824 件非谣言事件，以确保数据集在谣言和非谣言之间相对均衡。为使数据更贴近真实场景，本文不仅选取了部分可信度较高的官方媒体微博，还选择了一部分个人微博。数据集构建过程为：

1)从微博社区管理中心获取已被官方平台证实为谣言的微博，作为实验中的谣言数据样本。根据文献[20]的研究，从谣言数据发布的时间分布来看，近 90%的谣言微博在发布一周内会被举报，因此，选取微博发布时间超过一个月且未被举报证实为谣言的微博作为非谣言数据样本。

2)根据谣言微博的 ID 爬取参与评论的用户信息，包括用户的个人信息，如用户所在城市、是否是权威认证用户、用户粉丝数和关注数等，同时，获取该帖子的评论内容和评论时间等重要信息。

3)将上述获取的数据整合构建成一个新的谣言数据集。

本文主要在 Ma-Weibo 和 Weibo23 这 2 个数据集上对社交媒体上传播的谣言展开分析。各数据集的统计信息如表 1 所示。

表 1 各数据集的统计信息

Table 1 Statistical information of each dataset

单位：个

数据集	事件数量	帖子数量	谣言事件数量
Ma-Weibo	4 664	3 805 656	2 351
Weibo23	3 572	183 916	1 748

3.2 评估指标与参数设置

所有数据集按照 4∶1的比例划分训练集和测试集，采用 5 折交叉验证的方法，使用不同的随机种子运行 10 次并计算平均值。本文采用的评价指标包括准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 值。对于参数设置，谣言传播图初始节点的特征采用 817 维的融合特征，每个节点的隐层特征维度为 32，MFLAN 模型的层数为 3，Dropout 设置为 0.5，batch size 为 64，学习率为 0.05。

3.3 对比实验

3.3.1 与词嵌入模型的实验结果对比

本实验采用 m3e-base 获取词向量，并作为 MFLAN 模型的输入向量，比较词嵌入模型对谣言检测任务的效果。将 MFLAN 分别与 doc2vec＋MFLAN、text2vec-large-chinese＋MFLAN、BERT＋MFLAN 进行实验对比，本实验只在 Ma-Weibo 数据集上进行，对比结果如表 2 所示，最优结果加粗标注。

表 2 传统词嵌入模型的实验结果对比

Table 2 Comparison of experimental results of traditional word embedding models

模型	Accuracy	Precision	Recall	F1
doc2vec＋MFLAN	0.913 6	0.921 5	0.916 8	0.914 5
BERT＋MFLAN	0.945 4	0.948 6	0.949 1	0.948 4
text2vec-large-chinese＋MFLAN	0.963 1	0.964 1	0.967 6	0.963 5
m3e-base＋MFLAN	**0.976 4**	**0.973 9**	**0.971 8**	**0.974 5**

由以上结果可以看出，本文采用的 m3e-base 模型在各项指标上的表现明显优于其他词嵌入模型。m3e-base 拥有更强大的语义表示，能捕捉更为丰富的上下文语义信息，通过在大规模的语料库上进行模型预训练，在谣言检测任务中取得了较好的性能。

3.3.2 与基准模型的实验结果对比

为了进一步说明本文所提模型的有效性和合理性，将该模型与 7 种基准模型进行对比，实验对比结果如表 3、表 4 所示，使用的对比模型如下：

1）DTC[21]：一种基于树状结构的分类算法，通过对特征进行逐步分割，形成一个决策树，从而实现样本分类。

2）GRU-2[22]：循环神经网络的一种变体，具有门控机制，用于捕捉文本序列中的长期依赖关系。

3）GraphSAGE[23]：一种用于节点分类的图神经网络模型，通过采样邻居节点并聚合其信息来更新每个节点的表示。

4）GAT[24]：一种具有注意力机制的图神经网络，能够对不同节点分配不同的注意力权重。

5）Bi-GCN[25]：一种适用于二分图结构的图卷积网络。

6）GATv2[26]：GAT 模型的进化版本，具有更高的性能和更好的泛化能力。

7）Dir-GNN[27]：一种新型的面向有向图的图神经网络框架，可以利用边的方向信息来进行节点表征学习。

表 3 各模型在 Ma-Weibo 数据集上的性能指标对比

Table 3 Comparison of performance indicators of various models on the Ma-Weibo dataset

模型	Accuracy	Precision	Recall	F1
DTC	0.848 6	0.816 2	0.828 4	0.829 5
GRU-2	0.882 4	0.894 5	0.885 7	0.897 7
GraphSAGE	0.931 2	0.925 3	0.921 4	0.928 6
GAT	0.945 4	0.941 4	0.939 2	0.943 5
Bi-GCN	0.954 2	0.954 9	0.951 5	0.953 7
GATv2	0.962 4	0.964 1	0.961 7	0.959 3
Dir-GNN	0.966 4	0.968 8	0.957 6	0.961 4
MFLAN	**0.977 1**	**0.974 1**	**0.971 6**	**0.974 9**

表 4 各模型在 Weibo23 数据集上的性能指标对比

Table 4 Comparison of performance indicators of various models on the Weibo23 dataset

模型	Accuracy	Precision	Recall	F1
DTC	0.829 8	0.823 4	0.820 7	0.830 4
GRU-2	0.865 0	0.859 4	0.878 3	0.878 4
GraphSAGE	0.921 7	0.923 3	0.918 7	0.924 5
GAT	0.931 8	0.927 7	0.921 3	0.922 5
Bi-GCN	0.944 3	0.938 7	0.937 5	0.940 5
GATv2	0.952 1	0.948 8	0.951 7	0.942 3
Dir-GNN	0.959 8	0.953 2	0.952 6	0.952 4
MFLAN	**0.971 0**	**0.965 7**	**0.962 2**	**0.964 0**

从表 3、表 4 结果可以看出，在准确率、精确率、召回率、F1 值等评估指标上，本文模型在识别和分类谣言方面展现出卓越的性能，该模型在处理社交网络复杂结构时显示出较强的鲁棒性，能够准确捕捉节点之间的关系，并对谣言传播路径进行全面分析。这一优越的表现可能归因于模型中引入了可学习的参数矩阵和稀疏化技术，可学习的参数矩阵使模型能够灵活地适应不同社交网络的特征，而稀疏化技术有效地提高了模型的计算效率，使其在处理大规模图时能够保持较高的性能。

3.3.3 对 SL-GAT 模块各部分的时间复杂度分析

在 SL-GAT 模块中的特征变换部分，对于每个节点，通过可学习的参数矩阵进行特征变换，时间复杂度为 $O(F\times F')$，其中，F 是输入特征的维度，F' 是输出特征的维度。在计算节点与其邻居节点之间的注意力系数时，时间复杂度为 $O(T\times N\times F)$，其中，T 是注意力头数，N 是节点数量。在稀疏化处理部分，每个节点只与平均 α 个邻居节点进行注意力计算，其中，$\alpha<N$，可以将时间复杂度由 $O(T\times N\times F)$ 变为 $O(T\times\alpha\times F)$，显著减少了内积计算的数量，降低了时间复杂度。加权求和部分的时间复杂度为 $O(N\times F')$，池化处理部分的时间复杂度主要取决于池化过程中选择的前 K 个分数最高的节点，因此为 $O(N\log_{\alpha}K)$。

3.4 消融实验

为了验证 MFLAN 中各模块对于最终结果的影响，本文进行消融实验。在完整 MFLAN 的基础上，设置以下 4 组消融方法：

1）去除可学习的参数矩阵（MFLAN-matrix）。

2）去除稀疏化处理部分，使用原本的注意力权重（MFLAN-sparsemax）。

3）去除正值位置编码，不添加位置信息（MFLAN-PPE）。

4）更改 MFLAN 为单头注意力机制（MFLAN-heads 为 1）。

在 2 个数据集上将 MFLAN 与 4 组消融方法进行性能指标对比，结果如图 5 所示。

在消融实验部分，通过系统性地剔除或调整模型的不同组成部分，验证模型中每个组成部分对于谣言检测性能的具体贡献。由该实验结果可以看出，与各消融方法相比，MFLAN 在 2 个数据集上都取得了更优的效果。

在移除可学习参数矩阵的情况下，模型表现出明显的性能下降，这表明可学习参数矩阵对于模型的学习过程和性能具有关键作用；在移除正值位置

编码的情况下，模型性能出现明显下降，这表明对输入特征加入位置信息可以有效地帮助模型理解语义并提高模型预测效果；在移除对注意力分数稀疏化处理的情况下，模型性能下降，这表明稀疏化注意力分数使模型在处理输入时更加关注特定的信息，适度的稀疏化可以提高模型的泛化性能；通过调整注意力头数研究模型的性能表现，实验结果表明，当注意力头数为 4 时，模型表现出的学习能力和性能最佳。

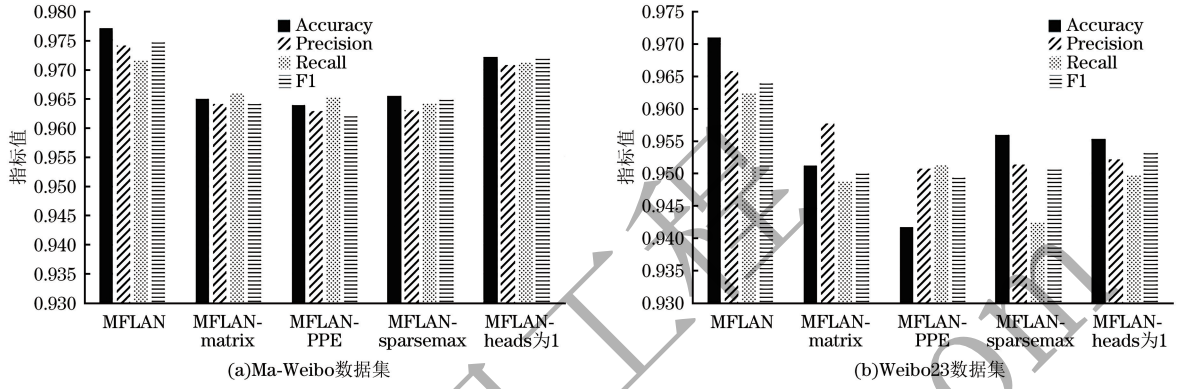

图 5 MFLAN 与 4 种消融方法在 2 个数据集上的性能对比

Fig. 5 Performance comparison of MFLAN and four ablation methods on two datasets

4 结束语

由于社交媒体平台规模庞大，信息传播十分复杂，传统的方法在处理大规模图时面临着精度和效率的挑战。因此，本文提出一种基于改进图注意力网络的多特征融合谣言检测模型 MFLAN，旨在解决大规模图上的谣言检测问题。该模型通过引入可学习的参数矩阵和正值位置编码，能够更好地捕捉节点之间的复杂关系及输入特征的位置信息，并在注意力分数计算时引入稀疏化技术，以降低计算复杂度，从而更好地处理大规模图。在多个真实社交媒体数据集上进行实验，将本文 MFLAN 模型与其他传统的图注意力模型进行比较，实验结果表明，MFLAN 模型在谣言检测任务中取得了显著的性能提升，具有更高的效率和准确性。

社交媒体上的图结构往往是动态变化的，节点和边的出现和消失都可能影响谣言检测任务的结果。接下来的研究将关注如何处理动态图结构，包括节点的新增和删除、边的权重变化等，以应对实际场景中的动态图问题。

参考文献

[1] 朱文龙，陈羽中，饶孟宇. 一种基于动态异构图的谣言检测模型[J]. 小型微型计算机系统，2024，45(2)：319-326.
ZHU W L，CHEN Y Z，RAO M Y. Rumor detection model based on dynamic heterogeneous graph neural network[J]. Journal of Chinese Computer Systems，2024，45(2)：319-326. (in Chinese)

[2] 刘雅辉，靳小龙，沈华伟，等. 社交媒体中的谣言识别研究综述[J]. 计算机学报，2018，41(7)：1536-1558.
LIU Y H，JIN X L，SHEN H W，et al. A survey on rumor identification over social media [J]. Chinese Journal of Computers，2018，41(7)：1536-1558. (in Chinese)

[3] 陈燕方，李志宇，梁循，等. 在线社会网络谣言检测综述[J]. 计算机学报，2018，41(7)：1648-1677.
CHEN Y F，LI Z Y，LIANG X，et al. Review on rumor detection of online social networks[J]. Chinese Journal of Computers，2018，41(7)：1648-1677. (in Chinese)

[4] HUANG Q，ZHOU C，WU J，et al. Deep spatial-temporal structure learning for rumor detection on Twitter[J]. Neural Computing and Applications，2023，35(18)：12995-13005.

[5] SHELKE S，ATTAR V. Rumor detection in social network based on user，content and lexical features[J]. Multimedia Tools and Applications，2022，81(12)：17347-17368.

[6] HAMIDIAN S，DIAB M T. Rumor detection and classification for Twitter data [EB/OL]. [2023-10-05]. https://arxiv.org/abs/1912.08926v1.

[7] MA J，GAO W，WEI Z Y，et al. Detect rumors using time series of social context information on microblogging websites[C]//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. New York，USA：ACM Press，2015：1751-1754.

[8] FRIGGERI A，ADAMIC L，ECKLES D，et al. Rumor cascades [J]. Proceedings of the International AAAI Conference on Web and Social Media，2014，8(1)：101-110.

[9] LIU Y，WU Y F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence，2018，32(1)：14-26.

[10] XU S Z，LIU X D，MA K，et al. Rumor detection on social media using hierarchically aggregated feature via graph neural networks[J]. Applied Intelligence，2023，53(3)：3136-3149.

[11] MA J，GAO W，WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[EB/OL]. [2023-10-05]. https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=5566&context=sis_research.

[12] VOSOUGHI S，ROY D，ARAL S. The spread of true and false news online[J]. Science，2018，359(6380)：1146-1151.

[13] BAI L，HAN X M，JIA C Y. A rumor detection model incorporating propagation path contextual semantics and user information[J]. Neural Processing Letters，2023，55(7)：9831-9850.

[14] CHEN Z D，ZHUANG F Z，LIAO L J，et al. A syntactic multi-level interaction network for rumor detection [J]. Neural Computing and Applications，2024，36(4)：1713-

1726.
[15] BIAN T, XIAO X, XU T Y, et al. Rumor detection on social media with bi-directional graph convolutional networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 549-556.
[16] JAIN D K, KUMAR A, SHRIVASTAVA A. CanarDeep: a hybrid deep neural model with mixed fusion for rumour detection in social data streams[J]. Neural Computing and Applications, 2022, 34(18): 15129-15140.
[17] CHEN T, LI X, YIN H Z, et al. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection[EB/OL]. [2023-10-05]. https://arxiv.org/abs/1704.05973.
[18] WANG Y, SUN Q, HE S. M3E: Moka Massive Mixed Embedding model[EB/OL]. [2023-10-05]. https://github.com/wangyuxinwhy/uniem, 2023.
[19] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[EB/OL]. [2023-10-05]. https://www.ijcai.org/Proceedings/16/Papers/537.pdf.
[20] 刘知远, 张乐, 涂存超, 等. 中文社交媒体谣言统计语义分析[J]. 中国科学: 信息科学, 2015, 45(12): 1536-1546.
LIU Z Y, ZHANG L, TU C C, et al. Statistical and semantic analysis of rumors in Chinese social media[J]. Scientia Sinica (Informationis), 2015, 45(12): 1536-1546. (in Chinese)
[21] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on Twitter[C]// Proceedings of the 20th International Conference on World Wide Web. New York, USA: ACM Press, 2011: 675-684.
[22] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB/OL]. [2023-10-05]. https://arxiv.org/abs/1406.1078v3.
[23] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[EB/OL]. [2023-10-05]. https://arxiv.org/abs/1706.02216v4.
[24] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. [2023-10-05]. https://arxiv.org/abs/1710.10903v3.
[25] MA J, GAO W, WONG K F. Rumor detection on Twitter with tree-structured recursive neural networks[EB/OL]. [2023-10-05]. https://scispace.com/pdf/rumor-detection-on-twitter-with-tree-structured-recursive-2ardit2u80.pdf.
[26] BRODY S, ALON U, YAHAV E. How attentive are graph attention networks?[EB/OL]. [2023-10-05]. https://arxiv.org/abs/2105.14491v3.
[27] ROSSI E, CHARPENTIER B, DI GIOVANNI F, et al. Edge directionality improves learning on heterophilic graphs[EB/OL]. [2023-10-05]. https://arxiv.org/abs/2305.10498v3.

编辑 吴云芳