RAGAT-MIND: 基于 MINDSPORE 的多粒度建模谣言检测方法

Zhenkai Qin[1,2,3]
[1]**School of Information Technology**
[2]**Network Security Research Center**
[3]**Big Data and Policing Technology Laboratory**
Guangxi Police College, Nanning, China
qinzhenkai@gxjcxy.edu.cn

Guifang Yang
School of Information Technology
Guangxi Police College
530028, China
yangguifang@gxjcxy.edu.cn

Dongze Wu
School of Information Technology
Guangxi Police College
530028, China
wudongze@gxjcxy.edu.cn

ABSTRACT

随着虚假信息在社交媒体平台上的持续扩散，有效的谣言检测已成为自然语言处理中的一个紧迫挑战。本文提出了一种用于中文谣言检测的多粒度建模方法 RAGAT-Mind，该方法基于 MindSpore 深度学习框架构建。该模型集成了 TextCNN 用于局部语义抽取，双向 GRU 用于顺序上下文学习，多头自注意力机制用于全局依赖聚焦，以及双向图卷积网络（BiGCN）用于词共现图的结构表示。在 Weibo1-Rumor 数据集上的实验表明，RAGAT-Mind 实现了优越的分类性能，达到了 99.2 % 的准确率和 0.9919 的宏 F1 分数。结果验证了结合层次语言特征和基于图的语义结构的有效性。此外，该模型表现出较强的泛化能力和可解释性，突显了其在实际谣言检测应用中的实用价值。

Keywords Rumor Detection; Multi-Granular Modeling; Graph Attention; TextCNN; BiGCN; MindSpore; Chinese Social Media

1 引言

随着社交媒体的快速发展，微博、推特和脸书等平台已成为重要的信息传播渠道。这些平台为用户提供了便捷的内容发布工具和强大的传播机制，使信息能够在短时间内迅速传播 [**?**] 。然而，这种高度有效的传播也为错误信息、谣言和错误的公共观点的传播提供了肥沃的土壤。在紧急情况、社会热点和公共危机等场景中，虚假内容的广泛传播容易误导公众认知，甚至可能对社会稳定、政策实施和国家安全构成严重威胁 [**?**, **?**] 。

针对这个问题，谣言检测越来越受到学术界和工业界的关注。传统的谣言检测方法主要基于手工特征工程，涉及到提取文本内容、用户行为以及传播图形结构，然后应用支持向量机（SVM）和随机森林等分类器来识别谣言 [**?**] 。虽然这些方法在早期阶段取得了令人鼓舞的结果，但它们严重依赖于领域专业知识，表现出较差的普遍性，并且在社交媒体中流行的非正式和多样化的语言表达上表现不佳。因此，它们的鲁棒性和适应性有限 [**?**] 。

随着深度学习在自然语言处理中的广泛应用，越来越多的研究开始探索使用神经网络进行端到端的谣言检测。卷积神经网络（CNNs）在捕捉局部特征方面表现出色，非常适合于建模文本中的 n-gram 模式；循环神经网络（RNNs）及其变体如门控循环单元（GRUs）能够捕捉上下文依赖关系，从而增强模型对文本序列的理解能力 [**?**, **?**] 。近年来，Transformer 架构和注意力机制的引入，尤其是多头自注意力，显著提高了长文本和复杂语义建模任务的性能，并广泛应用于情感分析、事件检测和相关领域 [**?**] 。

然而，社交媒体文本不仅表现出序列特征，还具有显式的结构关系，例如词语的共现和语义关联。仅仅依赖于序列建模往往难以捕捉词语之间的非线性结构依赖性。为了解决这一限制，近年来图神经网络（GNNs）被引入到文本建模任务中，从而实现结构感知的表示学习。特别是，双向图卷积网络（BiGCNs）利用前向和后向传播从双向语义图中提取深层结构关系。它们在文本分类、情感分析和信息抽取等任务中显示出极大的潜力 [**?**] 。

MindSpore 是由华为独立开发的全场景 AI 框架。它支持动态和静态计算图的混合执行、算子融合优化和自动并行化，从而实现深度神经网络的高效训练和部署。在中文自然语言处理任务中，MindSpore 提供灵活的

图计算能力、异构结构模块的算子支持以及高效的多模块调度机制。这些特性使其特别适合于整合卷积、递归、基于注意力以及基于图的神经结构的复杂建模任务。

在这些基础之上，本文提出了一种用于中文谣言检测的多层次建模框架，名为 RAGAT-Mind（基于 MindSpore 的谣言感知图注意力和 TextCNN 模型），实现了语义和结构表示学习之间的深度融合。该模型集成了四个关键架构模块：用于局部语义特征提取的 TextCNN，用于建模时间依赖性的 GRU，聚焦显著信息的多头注意力（MHA），以及通过词共现图实现结构表示的双向图卷积网络（BiGCN）。通过并行的双路径设计，该模型在处理具有高度语言歧义、结构碎片化及语义不连续性频繁的中文社交媒体谣言时，展示了强大的表现力和稳健的辨别能力。

2 相关工作

随着社交媒体平台的迅速普及，信息传播变得更加高效，但错误信息和网络谣言的传播也愈加严重。因此，自动谣言检测成为了自然语言处理领域的重要任务。现有研究主要分为三种方法：传统的机器学习方法、深度语义建模方法以及基于图的神经网络建模技术。

早期的研究集中在手工特征提取与传统分类器的结合。例如，Rizzo 等人 [?] 从微博帖子中提取统计特征，并结合时间传播线索以实现早期谣言检测。Shi 等人 [?] 应用迁移学习技术以改善不同中文数据集之间的跨域适应性。Zhang 等人 [?] 将用户行为图与文本特征相结合，以评估信息来源的可信度。尽管这些方法提供了有价值的见解，但它们在语义表达、结构建模能力方面较为有限，且在应对社交媒体中普遍存在的非正式和模糊语言表达时表现出较差的鲁棒性。

随着深度学习的进步，该领域已转向端到端的表示学习。卷积神经网络（CNNs）已被广泛采用来捕获局部的 n-gram 语义 [?]，而门控循环单元（GRUs）在建模时间依赖性方面表现出了效果 [?]。多头自注意力（MHA）尤其提高了关注语义显著区域的能力 [?]。Gao 等人 [?] 引入了一种情感感知的 BERT 框架来增强假新闻的分类。然而，这些方法大多忽略了嵌入文本中的结构信息，这对理解词汇关联的基本拓扑结构至关重要。

为了弥补这一缺陷，图神经网络（GNNs）被引入到谣言检测任务中，以捕捉词汇共同出现和拓扑语义。Zichao 等人 [?] 提出了 TextGCN，通过构建异构词-文档图来学习结构感知的表示。Yao 等人 [?] 通过结合边缘加权和位置编码改进了该框架。Bian 等人 [?] 开发了一种双向图卷积网络（BiGCN），用于模拟双向语义传播。在中文谣言检测的背景下，Vu 等人 [?] 提出了一种多模态融合模型，通过 GCNs 结合文本特征与用户传播路径。尽管这些模型在捕捉全局结构上颇为有效，但大多数基于 GNN 的模型依赖于静态拓扑，使得它们在建模真实世界语境中的动态语义变化和语境变异方面显得不足。

为了弥合语义和结构建模之间的差距，已经提出了几种基于融合的架构。Azri 等人 [?] 提出了一种并行的 CNN-GRU 框架，以联合建模局部语义和序列依赖性。Wei 等人 [?] 设计了一种双路径模型，集成了 Transformer 和 GCN 来捕捉全局语义和拓扑结构。Mehta 等人 [?] 将图注意力网络（GATs）纳入多头注意力模块中以执行语义感知的节点级聚合。尽管这些方法展示了有前景的结果，但它们仍然面临以下限制：

- 静态特征融合：大多数融合架构依赖于简单的拼接来自不同模块的输出，缺乏语义和结构分支之间的动态交互或反馈机制。
- 集成深度不足：卷积、顺序和图结构之间的协调通常较浅，这限制了联合表示的表达能力。
- 缺乏对语言不规则性的适应性：这些方法在处理语义跳跃、断裂的句法结构以及非正式语言方面存在困难，而这在中文社交媒体文本中很常见。

为了解决这些挑战，我们提出了一种统一的多粒度架构——TextCNN-GRU-MHA-BiGCN——无缝整合了卷积、递归、基于注意力和基于图的模块。这种混合设计增强了语义表达能力和结构感知能力，为复杂、动态和语言多样的中国社交媒体环境中的谣言检测提供了强有力的解决方案。

3 模型描述

3.1 总体架构

为了有效建模中文社交媒体谣言中的复杂语义和结构信息，我们提出了一种名为 RAGAT-Mind 的多粒度特征建模框架，该框架集成了卷积、递归、基于注意力和基于图的模块。如图 1 所示，整体模型由两条并行路径组成：语义建模路径和结构建模路径。这两条路径旨在捕捉不同层次的语言特征，包括局部语义、序列依赖关系、全局注意力以及词共现结构。

架构中的每个模块在多粒度表示学习过程中扮演特定角色：

- TextCNN：提取局部 n-gram 语义模式并构建初始短语级别的表示，这对于捕捉短文本中的关键信息尤为有效。
- GRU：基于卷积特征对单词之间的时间依赖性进行建模，增强对上下文演变的敏感性。
- MHA：采用多头注意力机制对语义重要区域赋予更高的权重，提升全局语义建模。
- BiGCN: 构建词共现图，并应用双向图卷积以捕获超出线性顺序的结构依赖。
- 融合层：连接语义路径和结构路径的输出，实现语言内容和潜在结构的统一建模。

在语义建模路径中，输入文本首先由 TextCNN 模块处理以提取多尺度的 n-gram 特征，从而构建初步的短语级别表示。卷积滤波器以不同的核大小滑过词嵌入矩阵，捕捉句子中的关键短语模式。然后这些特征被送入门控循环单元（GRU），它以递归方式编码时序上的上下文关系，有效地建模长距离依赖和语义演变。鉴于仅靠 GRU 可能无法充分捕捉不同语义片段的重要性差异，随后应用多头注意力机制（MHA）以自适应加权来聚合标记级别特征。这使得模型能够关注任务相关的语义区域，并生成一个上下文感知的全局表示。

同时，结构建模路径旨在捕捉单词之间的非线性依赖关系。通过使用滑动窗口策略构建词共现图，其中每个词被视为图中的节点，在局部上下文窗口内共现的词之间形成边。这一过程生成了表示本地句法或语义结构的句子级邻接矩阵。然后利用双向图卷积网络（BiGCN）对原始和转置的邻接矩阵进行图卷积，从而建模双向结构传播，并捕捉本地和远距离的共现模式。这提高了模型感知非顺序关系的能力，例如词跳跃、隐含配对和常见于中文社交媒体谣言文本的结构不规则性。

最后，语义路径和结构路径的输出，即全局语义表示 $\boldsymbol{h}$attn 和结构嵌入 $\boldsymbol{h}$gcn ，被连接起来形成统一的表示 $\boldsymbol{h}_{\text{final}}$ 。该向量通过一个 dropout 层进行正则化，然后输入到一个全连接层，接着通过 Softmax 分类器进行最终的二元预测。通过语义和结构模块的联合设计，RAGAT-Mind 将局部上下文建模、长距离依赖捕获、显著特征聚合和结构模式提取整合为一个连贯的、可解释的和健壮的谣言检测框架。

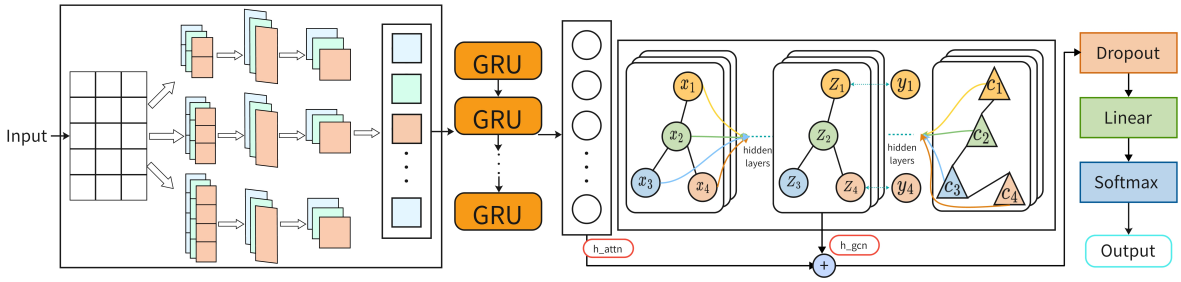

Figure 1: 提出的 TextCNN-GRU-MHA + BiGCN 模型的整体架构。

3.2 文本卷积神经网络

中文社交媒体平台上的谣言文本通常具有句子结构简短、语义密度高、核心信息通常通过局部短语传达的特点。这些特点使卷积结构特别适合于局部语义建模。因此，本研究采用 TextCNN 作为语义建模路径的第一个子模块。通过对词嵌入矩阵应用多个感受野的一维卷积，模型能够在不同粒度上提取 n-gram 语义模式。

如图 2 所示，TextCNN 模块由嵌入层、多个并行一维卷积滤波器、ReLU 激活函数、最大池化层和最终的特征拼接结构组成。这种架构使模型能够并行提取不同长度的语义片段，显著提高了对局部依赖的敏感性，并为下游序列建模模块（例如 GRU）提供高质量的输入表示。

在我们的实现中，我们使用了三组卷积滤波器，它们的核大小分别为 3、4 和 5，分别对应三元语法到五元语法级别的语义特征。每个滤波器捕捉不同跨度上的上下文组合，增强了模型适应各种粒度短语模式的能力。考虑到社交平台上的中文谣言通常包含非正式语言、短句结构和集中的信息，语义通常通过片段传达而不是完整的句子。依赖单一的滤波器大小可能会错过关键组合，限制模型在短文本中的语义覆盖。为了解决这个问题，我们并行应用多个感受野以捕捉多样的片段级语义模式，并增强模型对非标准表达的适应能力。基于这个设计，输入到 TextCNN 模块的文本首先被映射为一个词嵌入矩阵 $\boldsymbol{E} \in \mathbb{R}^{L \times d}$ ，其中 L 是最大句子长度，d 是嵌入维度。对于大小为 k 的一维卷积滤波器 $\boldsymbol{w}_i \in \mathbb{R}^{k \times d}$ ，其特征图的计算如下：

$$\boldsymbol{f}_i = \text{ReLU}(\boldsymbol{w}i * \boldsymbol{E}i : i + k - 1 + b_i) \tag{1}$$

然后，每个卷积输出通过一个最大池化操作，以提取最显著的局部响应：

$$\text{MaxPool}(\boldsymbol{f}_i) = \max(\boldsymbol{f}_i) \tag{2}$$

为了整合来自不同尺度的语义特征，所有卷积滤波器的池化输出被连接成一个统一的语义向量：

$$\boldsymbol{h}_{\text{CNN}} = [\text{MaxPool}(\boldsymbol{f}_1); \text{MaxPool}(\boldsymbol{f}_2); \ldots; \text{MaxPool}(\boldsymbol{f}_K)] \tag{3}$$

所得的向量 $\boldsymbol{h}_{\text{CNN}}$ 封装了跨多个感受野提取的局部语义信号，并作为后续序列建模模块的输入。

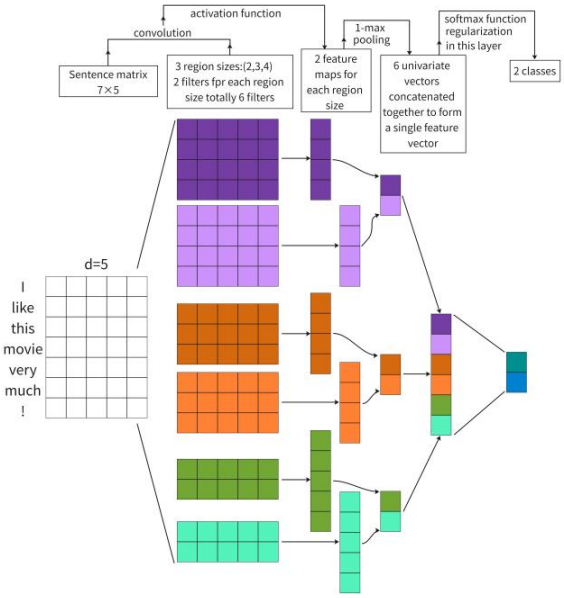

Figure 2: TextCNN 模块的结构

3.3 GRU

尽管卷积结构在捕捉局部 n-gram 语义特征方面表现出色，但其固定的感受野限制了建模词与词之间长距离依赖关系的能力。为了克服这一限制，我们在 TextCNN 模块之后引入了门控循环单元（GRU），以捕捉序列中的时间依赖性，并增强上下文语义演化的动态建模。

在所提出的架构中，GRU 模块作为语义建模路径的核心组件，如图 3 所示。它接收由 TextCNN 模块提取的多尺度局部特征，并沿着时间维度进行递归建模，以捕获令牌之间的动态依赖关系。基于卷积提供的局部表示，GRU 通过门控机制整合历史和当前信息，将静态特征转化为具有上下文连贯性的动态语义表示。这个过程显著增强了模型捕获长距离依赖关系的能力，同时抑制了冗余和噪声输入，从而提高了语义建模的鲁棒性和稳定性。

形式上，设 TextCNN 的输出特征序列为 $\boldsymbol{h}\text{CNN} \in \mathbb{R}^{L\times C}$ ，其中 L 是序列长度，C 是通道数量。在时间步 t 中，GRU 的隐藏状态根据当前输入 $\boldsymbol{h}\text{CNN}, t$ 和前一个隐藏状态 $\boldsymbol{h}_{t-1}$ 更新，如下所示：

$$\boldsymbol{h}t = \text{GRU}(\boldsymbol{h}\text{CNN}, t, \boldsymbol{h}_{t-1}) \tag{4}$$

在内部，GRU 使用更新门和重置门动态调节信息流：

$$z_t = \sigma(W_z \boldsymbol{x}t + U_z \boldsymbol{h}t - 1)\ r_t \qquad = \sigma(W_r \boldsymbol{x}t + U_r \boldsymbol{h}t - 1) \tag{5}$$

在这些门控机制的指导下，候选隐状态计算如下：

$$\tilde{\boldsymbol{h}}_t = \tanh(W_h \boldsymbol{x}t + U_h(r_t \odot \boldsymbol{h}t - 1)) \tag{6}$$

然后通过前一状态和候选状态的门控组合来更新最终的隐藏状态：

$$\boldsymbol{h}t = (1 - z_t) \odot \boldsymbol{h}t - 1 + z_t \odot \tilde{\boldsymbol{h}}_t \tag{7}$$

由 GRU 输出的隐藏状态序列，记作 $\boldsymbol{H}_{\text{GRU}}$，作为输入用于后续的多头注意力（MHA）模块。这个输出提供了一个在语义上不同且时间上连贯的表示，用于建模全局依赖关系。与将卷积特征直接输入注意力机制相比，由 GRU 提供的中间序列建模显著提高了注意力焦点的精准性以及模型对上下文语义的整体感知能力。

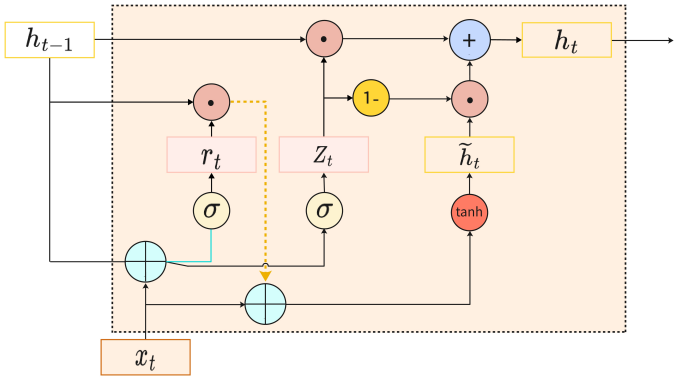

Figure 3: GRU 模块的结构

3.4 多头注意力

尽管门控循环单元（GRU）在建模单词序列中的时间依赖性方面表现出相当大的有效性，但它对顺序的依赖本质上限制了其捕捉非相邻标记之间长距离语义交互的能力。这一局限性在中文社交媒体谣言的背景下尤为明显，那里经常可以观察到非线性语言结构，例如不完整的句子模式、无序的语法和语义不连续。像 GRU 这样的顺序建模机制往往难以适应这些不规则性，从而阻碍了全面的语义理解并降低了分类精度。为了解决这一问题，我们在 GRU 层之后引入了一个多头注意力（MHA）机制。该模块增强了模型捕捉全局语义依赖关系的能力，并有效地编码长距离关系，从而减轻了纯顺序模型在处理非标准文本结构时的局限性。

在所提出的框架中，MHA 模块通过整合全局上下文信息在语义建模路径中起到关键作用。一方面，它继承了 GRU 构建的上下文连贯性；另一方面，它通过多个注意力头同时跨序列建模非局部交互，从而实现多视角语义融合。注意力权重矩阵明确反映了句中每个词的相对重要性，引导模型关注于谣言检测中最具信息量的区域，同时抑制无关或噪声信号。该机制最终有助于提高分类的准确性和鲁棒性。

从技术上讲，MHA 模块将 GRU 生成的隐藏状态序列 $\boldsymbol{H}_{\text{GRU}} \in \mathbb{R}^{L\times d}$ 作为输入，并应用三个并行线性变换来计算查询（$\boldsymbol{Q}$）、键（$\boldsymbol{K}$）和值（$\boldsymbol{V}$）矩阵：

$$\boldsymbol{Q} = \boldsymbol{H}\text{GRU}W^{Q}, \quad \boldsymbol{K} = \boldsymbol{H}\text{GRU}W^{K}, \quad \boldsymbol{V} = \boldsymbol{H}_{\text{GRU}}W^{V} \tag{8}$$

其中 W^{Q}、W^{K} 和 $W^{V} \in \mathbb{R}^{d\times d_k}$ 是可训练的投影矩阵，d_k 表示每个注意力头的维度。接着，使用缩放点积机制计算注意力分布：

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{9}$$

这一操作在 h 个注意力头中并行进行，以捕捉不同子空间中的多样语义关系。所有头的输出被连接，并通过最后的线性变换投影回原始维度：

$$\text{MHA}(\boldsymbol{H}_{\text{GRU}}) = \text{Concat}(\text{head}_1,\ldots,\text{head}_h)W^{O} \tag{10}$$

其中，$W^{O} \in \mathbb{R}^{hd_k\times d}$ 是输出投影矩阵，head_i 表示来自第 i 个注意力头的输出。

3.5 BiGCN

尽管上述的语义建模路径（TextCNN–GRU–MHA）在捕捉文本中的局部语义、时间演变和全局依赖方面表现出很好的能力，其体系结构仍然基本上基于线性顺序结构，这限制了其显式表示非相邻词之间潜在依赖关系的能力。为了解决这一限制，我们引入了一个基于双向图卷积网络（BiGCN）的结构建模分支，旨在通过图神经网络的视角来建模词之间的拓扑依赖关系。该模块提供了一种与顺序建模不同的互补结构表示，增强了模型在结构不规则和语义跳跃的复杂文本中的鲁棒性。

具体来说，针对每个输入句子，使用滑动窗口机制构建一个单词共现图。每个单词被视为图节点，而窗口内的共现单词对则通过边连接，得到一个邻接矩阵 $\boldsymbol{A} \in \mathbb{R}^{L\times L}$ ，其中 L 表示句子的长度。该结构编码了局部的单词级关联，并作为后续图卷积操作的结构先验。

基于构建的图，BiGCN 在原始邻接矩阵 $\boldsymbol{A}$ 及其转置 $\boldsymbol{A}^\top$ 上执行卷积，以建模双向语义传播。设 $\boldsymbol{X} \in \mathbb{R}^{L\times d}$ 表示输入的词嵌入，前向和后向传播可以表示为：

$$\boldsymbol{H}^f = \mathrm{ReLU}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{W}^f), \quad \boldsymbol{H}^b = \mathrm{ReLU}(\boldsymbol{A}^\top\boldsymbol{X}\boldsymbol{W}^b) \tag{11}$$

其中 $\boldsymbol{W}^f$ 和 $\boldsymbol{W}^b$ 分别是用于前向和后向传播的可学习权重矩阵，而 $\boldsymbol{H}^f$ 和 $\boldsymbol{H}^b$ 表示所得的特征。该双向机制允许对称的结构聚合，减轻单向图传播的局限性。

前向和后向特征被连接以形成全面的结构表示：

$$\boldsymbol{H}^{\mathrm{BiGCN}} = [\boldsymbol{H}^f; \boldsymbol{H}^b] \tag{12}$$

为了得到句子级别的表示，在所有节点上应用均值池化：

$$\boldsymbol{h}_{\mathrm{gcn}} = \mathrm{MeanPool}(\boldsymbol{H}^{\mathrm{BiGCN}}) \tag{13}$$

最后，来自结构路径 $\boldsymbol{h}_{\mathrm{gcn}}$ 的输出与后续融合模块中的语义表示 $\boldsymbol{h}_{\mathrm{attn}}$ 融合，提供多层次、结构感知的特征，以支持鲁棒和准确的谣言分类。

4 实验设置与结果

4.1 数据集描述

我们使用公开可用的中文微博谣言检测数据集 Weibo1-Rumor 进行实验，该数据集由 3,387 条真实的社交媒体文本组成，其中包括 1,538 条谣言实例和 1,849 条非谣言。所有文本均收集自实际在线平台，反映了真实的语言特征。为确保可重复性和泛化性，数据集按 8:2 的比例分为训练集和测试集。预处理步骤包括分词、令牌索引和邻接矩阵构建。

4.2 实验环境

本研究中的所有实验都是在使用华为开发的 MindSpore 深度学习框架的 Windows 10 操作系统上进行的。硬件环境配置了 Intel ® Xeon ® CPU、64 GB RAM 和 NVIDIA GeForce RTX 3070 Laptop GPU，提供了充足的计算资源和高吞吐量性能来支持模型训练。MindSpore 支持动态和静态图执行模式，并提供高效的算子融合和异构计算能力，使其非常适合大型文本建模任务。软件环境基于 Python 3.8 构建，并集成了若干必要的库，包括 `pandas` 、`numpy` 、`jieba` 、`scikit-learn` 和 `tqdm` ，分别用于数据预处理、中文分词、评估指标计算和训练进度可视化。

在训练过程中，Adam 优化器与动态学习率衰减策略结合使用，以加速收敛并提高模型稳定性。在各个层引入了 Dropout 正则化以缓解过拟合。所有实验均在统一配置下进行，并设置固定的随机种子以确保公平性和可重复性。此外，启用了提前停止策略，如果验证性能在连续多个周期中未能改善，将自动终止训练，从而增强模型的泛化能力并优化资源利用。

实验中使用的关键超参数设置总结在表格 1 中。

Table 1: 模型超参数设置

Hyperparameter	Value
Max sequence length	128
Embedding dimension	128
Convolution kernel sizes	[3, 4, 5]
GRU hidden size	128
Number of attention heads	4
Dropout rate	0.5
Batch size	32
Epochs	3
Learning rate	0.001
Optimizer	Adam

4.3 评估指标

为了全面评估所提出模型在中文谣言检测任务上的分类性能，采用了四个常用的指标: 准确率、精确率、召回率和 F1-得分。准确率反映了预测的整体正确性；精确率衡量的是所有被预测为正例的实例中正确预测的比例；召回率表示的是实际正例中被正确识别的比例；F1-得分作为精确率和召回率的调和平均，提供了对模型稳健性和判别能力的平衡评估。为了减少由类别不平衡引起的潜在偏差，我们采用了一种宏平均策略，其中精确率、召回率和 F1 得分分别针对每个类别独立计算，然后进行平均以确保每个类别在最终评估中同等贡献。所有指标都使用来自 `scikit-learn` 包的 `classification_report` 函数计算，以确保评估过程的准确性和可重复性。

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

在此，TP 表示真阳性的数量, TN 表示真阴性的数量, FP 表示假阳性的数量, FN 表示假阴性的数量。

4.4 实验结果与分析

为了全面评估所提出的 RAGAT-Mind 模型在中文社交媒体谣言检测中的性能，我们在统一的实验环境和一致的数据集划分下进行了比较实验。选择了几个具有代表性的基线模型进行比较，包括：关注局部语义特征提取的 TextCNN；将序列建模与注意力机制结合的 GRU-ATT；基于词共现图强调结构表示的 TextGCN；建立在大规模预训练基础上的上下文模型 BERT-FT；以及在图结构内引入节点级注意力的 GAT。所有模型均在 Weibo1-Rumor 数据集上进行了训练和评估。表 **??** 总结了每个模型在准确率、精确率、召回率、Macro-F1 和推理延迟（即每样本时间）方面的总体性能。如表中所示，RAGAT-Mind 在准确率和 Macro-F1 方面显著优于所有基线模型，表现出一致且稳健的性能。在训练过程中，模型的准确率从 89.6 % 提升到 99.2 %，而 Macro-F1 得分从 0.8939 提高到 0.9919。同时，训练损失从 0.5896 下降到 0.0447，表明其优秀的收敛行为和泛化能力。与基础模型 TextCNN 相比，RAGAT-Mind 在准确率上提高了约 3.4 %，在 Macro-F1 上提高了约 3.7 %。即使与结构感知模型如 TextGCN 相比，RAGAT-Mind 也在准确率上取得了 2.14 % 的提升，突显其结构语义联合建模策略的有效性。

从建模机制的角度来看，TextCNN 有效地捕捉了局部的 n-gram 特征，但缺乏建模长距离依赖的能力。GRU-ATT 引入注意力机制来增强序列上下文建模，但在结构意识方面仍然有限。TextGCN 和 GAT 通过图结构捕捉全局共现关系，但缺乏语义演变的时间建模能力。相比之下，RAGAT-Mind 采用并行双路径设计，整合了 TextCNN、GRU、多头注意力（MHA）和 BiGCN 模块，共同支持局部语义、序列依赖、显著区域聚焦以及全局结构建模。这种架构协同作用极大地增强了模型对中文社交媒体文本中复杂语义模式的适应能力。

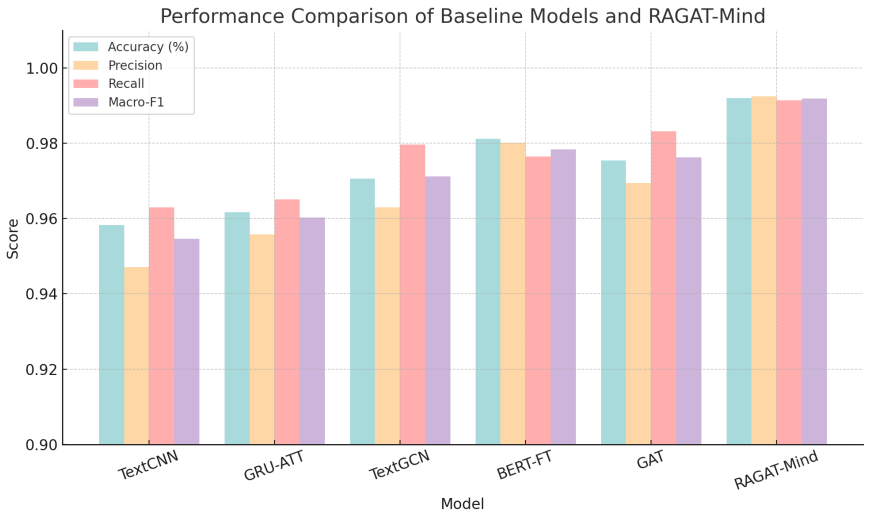

Figure 4: 所有模型的性能比较

为了更清晰地比较所有模型的性能，图 4 提供了一个柱状图，比较了四个关键指标：准确率、精确率、召回率和宏 F1 分数。正如图中所示，传统的序列模型如 TextCNN 和 GRU-ATT 在精确率和召回率方面表现相对较强，反映出它们在捕捉显性语义特征方面的有效性。然而，在不平衡或语言不规则的情况下，它们的宏 F1 分数明显较低，表明其泛化性有限，分类边界模糊。基于图的方法如 TextGCN 和 GAT 在召回率上取得了显著的提升，验证了结构化建模在发现潜在语义结构方面的效用。然而，它们缺乏时间建模，限制了捕捉变化的上下文语义的能力，从而影响了整体的准确性。

进一步分析表明，BERT-FT 在所有指标上表现良好，这得益于其深度上下文嵌入和预训练提供的先验知识。然而，其相对较高的推理延迟（12.34 毫秒）对在对延迟敏感的应用中部署构成了限制。相比之下，RAGAT-Mind 在所有指标上取得了最佳性能——准确率（99.20 %）、精确率（0.9925）、召回率（0.9914）和宏观 F1（0.9919），同时保持较低的推理延迟（5.36 毫秒）。图 4 中相同高度的柱状图反映了模型在处理噪声、碎片和语义多样的谣言文本时的一贯稳健性和判别能力。这些结果进一步验证了该模型混合语义-结构设计的有效性。

总之，RAGAT-Mind 在中文社交媒体谣言检测任务中表现出高稳定性、强辨别准确性和结构敏感性。实验结果全面支持模型的多层次表示学习框架，并为在复杂和对抗性语言环境中的未来研究提供了坚实的技术基础。

5 讨论

为了进一步探索所提出模型在中文谣言检测中的适用性和潜力，本节对实验结果进行了深入讨论。分析从三个角度进行：分类性能和模型局限性、可解释性和结构理解、以及实际适应性和未来研究方向。

5.1 分类效能和模型约束

实验结果表明，所提出的 TextCNN-GRU-MHA-BiGCN 模型实现了强大的分类性能，达到了 99.2% 的准确率和 0.9919 的宏平均 F1 分数，损失值低至 0.0447。这些结果展示了该模型在谣言分类任务中有效拟合和泛化的能力。这一成功得益于四种互补机制的整合：局部语义特征提取（TextCNN）、序列依赖建模（GRU）、基于注意力的关键区域聚焦（MHA）和结构表示学习（BiGCN）。这些组件共同提供了对短文本语义的全面表示，增强了模型对多样性表达的鲁棒性。

然而，仍然存在几个限制。首先，模型性能对训练数据的质量非常敏感。社交媒体数据集中存在的语义模糊和噪声标签可能会干扰学习过程并降低泛化能力。其次，目前的图结构是基于滑动窗口共现机制构建的，未能纳入更高级的语言特征，如句法依赖、语义角色或外部知识。这限制了模型捕捉更深层语言结构的能力。最后，该模型尚未在跨平台或跨领域场景中进行验证。其适应不同社交平台的异构数据的能力仍未测试，需要进一步研究。虽然模型在单一来源数据集上表现良好，但在实际应用中的未来部署需要增强领域适应性和对数据多样性的弹性。

5.2 可解释性和结构意识

除了取得较强的分类结果外，所提出的模型还强调体系结构的可解释性和结构的透明性。多头自注意力机制通过计算词对间的注意力分数，实现了对标记级语义贡献的动态加权。通过可视化这些分数，可以识别影响

分类决策的关键语义区域，从而深入了解模型的决策过程。此外，BiGCN 模块在词图内的前向和后向方向上捕捉结构依赖性。该双向传播机制增强了模型对层级语言结构和词间潜在关系的敏感性。

最终特征融合过程将语义和结构表示连接起来，保留它们各自的贡献，并在内容和结构层面上实现可解释性。该设计促进了事后模型诊断和可解释性分析，使得该框架在实际应用中更加透明和可信。总体而言，该架构不仅提供准确的预测，还为可解释的决策提供了基础，这对于如虚假信息检测等高风险应用越来越重要。

5.3 实用适应性和未来方向

虽然该模型在 Weibo1-Rumor 数据集上表现出色，但更广泛的实际部署需要进一步的泛化和适应性改进。首先，当前的实验设置仅关注于单一平台（微博），未纳入其他流行的中文平台如知乎或微信公众号的数据集。未来的工作应通过采用领域适应和迁移学习策略来应对分布变化，评估模型在跨平台下的稳健性。

其次，目前的图构建仅依赖于基于窗口的共现，这可能不足以捕捉复杂的句法或基于知识的关系。整合外部语言资源——如依存解析、知识图谱或命名实体关系——可以显著增强构建图的表达能力和上下文基础。

最后，在现实世界的情景中，谣言传播通常涉及多模态元素，包括文本图像组合和行为传播路径。未来这项工作的扩展可以包括设计多模态架构，将视觉、行为或时间信号整合到框架中，进一步增强检测的鲁棒性和实际使用性。

6 结论

本研究提出了一种多模块集成模型用于中文谣言检测，该模型结合了卷积、循环、基于注意力和基于图的神经网络架构。具体而言，模型整合了文本卷积神经网络（TextCNN）用于局部 n-gram 模式提取，门控循环单元（GRU）用于序列依赖性建模，多头自注意力（MHA）用于增强对语义重要区域的关注，以及双向图卷积网络（BiGCN）用于通过词共现图捕获结构性依赖性。这些互补组件的整合使模型能够在一个统一的框架内联合建模局部语义、全局上下文和基于图的结构。

在 Weibo1-Rumor 数据集上进行的综合实验证明了该模型的有效性和稳定性，其测试准确率达到了 99.2 %，宏平均 F1 得分为 0.9919。这些结果证实了所提出的混合架构在短文本分类任务中的优越性，并验证了多视角表示学习在复杂社交媒体环境中的价值。此外，该模型的设计保留了结构解释性，提供了性能和可解释性，这对于实际应用中的虚假信息检测至关重要。

尽管模型性能强大，但仍存在进一步改进的机会。未来的工作将集中于通过引入句法依赖关系或外部知识结构（如实体关系或知识图谱）来增强图构建的语义深度。此外，将探索迁移学习和领域适应技术，以改善在不同社交媒体平台上的泛化能力。最后，通过结合视觉和行为线索，将模型扩展到多模态谣言检测，有望进一步提高其稳健性和应用潜力。

除了方法论方面的贡献之外，RAGAT-Mind 在实际应用中的潜力也非常大。该模型可以整合到政府平台、社交媒体门户网站或公共安全仪表板的内容监控系统中，以实现实时谣言筛查和内容可信性评估。其统一且可解释的架构确保了精确性和透明性，使其适用于可靠性和可解释性都至关重要的应用场景。总体而言，所提出的模型为推动谣言检测研究和实际应用的发展贡献了一个可扩展、可解释且高性能的框架。

感谢 MindSpore 社区提供的支持。