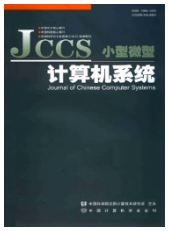

小型微型计算机系统
Journal of Chinese Computer Systems
ISSN 1000-1220,CN 21-1106/TP

《小型微型计算机系统》网络首发论文

题目：背景知识增强的多特征融合谣言检测方法
作者：林兴澎，李家印，徐瑞阳，许力
收稿日期：2023-11-23
网络首发日期：2024-02-29
引用格式：林兴澎，李家印，徐瑞阳，许力．背景知识增强的多特征融合谣言检测方法[J/OL]．小型微型计算机系统.
https://link.cnki.net/urlid/21.1106.tp.20240228.1157.017

小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems

背景知识增强的多特征融合谣言检测方法

林兴澎，李家印，徐瑞阳，许 力

（福建师范大学计算机与网络空间安全学院，福州 350117）

（福建省网络安全与密码技术重点实验室，福州 350117）

E-mail：xuli@fjnu.edu.cn

摘 要：在线社交媒体的普及为人们通信带来便利，但也为谣言滋生创造条件，设计高效的谣言检测方法能保护人民财产和维持社会稳定。已有方法主要集中在利用谣言传播中的丰富信息来检测谣言，这些方法在长期谣言检测具有优越性能，但应对早期谣言检测的效果不佳。针对这些方法无法在谣言传播早期获得丰富信息的问题，本文提出了一种背景知识增强的多特征融合谣言检测方法来提高早期谣言检测性能。首先，从知识图谱和维基百科中挖掘谣言背景知识并建立知识关联图来补充源推文的语义信息；其次，为了解决现有方法难以学习具有不同差异性噪声的谣言传播表示的问题，本文设计了一种基于加性注意力和点积注意力的图神经网络结构对谣言进行插值学习；最后，将知识关联图、谣言传播-扩散图以及社交图的表示进行结合，构建出具有多通道输入的谣言检测器架构，从而实现早期谣言的精准分类。实验结果表明，本文方法在 3 个公开数据集上的准确率分别达到了 87.3、90.4%和 87.0%，与其它对比方法相比，具有更高的早期谣言检测准确率和长期谣言检测准确率。

关键词：社交媒体；谣言检测；图神经网络；知识图谱

中图分类号：TP391 **文献标识码：**A

Multi-feature Fusion for Rumor Detection with Background Knowledge Enhancement

LIN Xingpeng，LI Jiayin，XU Ruiyang，XU Li

（College of Computer and Cyber Security,Fujian Normal University,Fuzhou 350117,China）

（Fujian Provincial Key Laboratory of Network Security and Cryptology,Fuzhou 350117,China）

Abstract: The popularity of online social media brings convenience to people's communication,but it also creates conditions for rumors to breed. Designing efficient rumor detection can protect property and maintain social stability. It has been mainly focused on using the rich information in the spread of rumor to detect rumors,which has superior performance in rumor detection,but the effect of dealing with early rumor detection is not good. To solve these problems,a multi-feature fusion for rumor detection with background knowledge enhancement is proposed to improve the performance of early rumor detection. Firstly,the rumor background knowledge is mined from the knowledge graph and the knowledge association graph is established to supplement the semantic information of the source tweet; Secondly,in order to solve the problem that it is difficult to learn the representation of rumor propagation with different noise,this paper designs a graph neural network structure based on additive attention and dot product attention to learn rumors; Finally,the knowledge association graph,rumor propagation-diffusion graph and social graph are combined to construct a rumor detector architecture with multi-channel input,so as to achieve the accurate classification of early rumors. Experimental results show that the accuracy of this work on three public datasets is 87.3%,90.4%,and 87.0%,respectively. Compared to other comparison methods,it has higher accuracy in early rumor detection and long-term rumor detection.

Keywords: Social media; Rumor detection; Graph neural network; Knowledge graph

0 引 言

随着移动互联网的迅猛发展，微博、Twitter 等社交媒体已成为人们获取和传播信息的主要途径。在社会心理学中谣言被认定为虚假的或是未经证实的陈述，它们常常被设计成一种能够比正常信息引起更多讨论的模式。由于大

收稿日期：2023-11-23 收修改稿日期：2023-12-25 基金项目：国家自然科学基金 NSFC 海峡联合基金项目（U1905211）资助；福建省科技项目（2022G02003，2021L3032）资助.作者简介：**林兴澎，**男，2001 年生，硕士研究生，研究方向为谣言检测、自然语言处理；**李家印，**男，1990 年生，博士，讲师，研究方向为人工智能安全、车联网安全、隐私保护、机器学习；**徐瑞阳，**男，1997 年生，硕士研究生，研究方向为网络与信息安全、社会网络；**许 力**（通信作者），男，1970 年生，博士，教授，博士生导师，CCF 杰出会员，研究方向为网络与信息安全、大数据与信息化、移动社会网络、智能信息处理等.

众言论表达的自由性以及多样性会造成谣言的广泛传播，这会对网络秩序以及社会稳定带来不利影响[1]。例如，在新冠疫情时期社交媒体上出现了许多“国产灭活疫苗根本没用”、“艾灸可以预防新冠”、“一直配戴口罩会导致肺部结节”等谣言，以及 2022 年俄罗斯与乌克兰冲突相关的谣言，它们干扰了大众对事实的了解，甚至造成严重的社会恐慌。因此非常有必要开展对谣言检测的研究，这对保护人民财产以及维持社会稳定具有重要的意义。

谣言检测方法有基于内容、基于传播模式以及基于混合特征的方法。基于内容的方法可以定义为一个短文本分类的任务，但由于短文本存在语义稀疏以及语义模糊的问题，这给谣言检测带来了困难。为了解决谣言内容存在语义缺失的问题，已有的方法主要是通过融合知识图谱进行谣言检测，但由于知识图谱的更新速度缓慢，使这些方法不能够及时地获取到最新、最准确的背景知识，从而无法检测出具有新颖术语的谣言。对于基于传播模式的方法来说，由于不同谣言事件的传播过程中的噪声含量不同，如何自适应学习具有不同差异性噪声的谣言传播表示，仍然是一个待解决的问题。近年来出现了许多基于混合特征的方法，这些方法大多都是通过谣言内容、谣言的传播模式以及用户信息进行谣言检测，这在一定程度上提高了谣言检测的性能。但这些方法大多都需要随着谣言不断传播才能够逐渐获取到更多的信息，因此无法在谣言产生早期阶段具有较好的检测效果。因此如何综合利用现有各种信息来提升早期谣言检测效果仍然是一个具有挑战性的问题。

首先，由于已有方法无法利用谣言传播过程中的丰富信息，这会导致早期谣言检测效果不佳。而维基百科里的内容总是随着谣言中新颖术语变化而同步更新，它相对于知识图谱来说，能够进一步提供更加准确而丰富的背景知识来补充谣言的语义信息。所以本文提出通过融合知识图谱和维基百科中的内容为谣言在早期提供丰富语义信息，以此来解决早期谣言检测性能不佳的问题。另一方面，由于谣言文本通常为短文本，并且每个背景知识的中的单词含量稀疏，直接连接文本和背景知识的词向量表示无法实现语义信息的有效融合。所以本文提出利用图结构对源推文和背景知识进行建模，这使得背景知识可以很容易地与谣言信息中任意单词进行融合，进而有效地补全谣言的语义信息；其次，为了解决谣言传播过程中存在差异性噪声的问题，本文设计了一种具有连接注意力和点积注意力的图神经网络结构来对具有不同差异性噪声的谣言传播表示进行插值学习；最后，由于早期谣言检测阶段无法获取到传播过程中丰富的用户信息，而发布者社交圈信息可以不依赖谣言传播而独立存在。所以本文将谣言背景知识、谣言传播模式与发布者社交圈信息进行融合共同来解决早期谣言检测性能不佳的问题。

综上所述，本文的贡献总结如下：

1)本文提出了一种基于背景知识增强的多特征融合图注意力网络谣言检测方法, 通过融合知识图谱和维基百科的内容来补充源推文的语义信息, 同时结合谣言传播模式以及发布者社交圈以提高早期谣言检测性能。

2)为了解决谣言与背景知识难以有效融合的困难问题，本文设计一种用于结合知识图谱背景知识、维基百科中背景知识以及源推文内容的图结构，它能够将背景知识和源推文有效融合以学习谣言的语义表示。

3)为了解决具有不同差异性噪声的谣言传播表示难以学习的困难问题，本文设计了一种基于加性注意力和点积注意力的图神经网络结构对谣言传播表示进行插值学习。

4)实验结果表明，本文方法在 3 个公开数据集上与一些主流的方法相比，无论是在早期谣言检测上还是在长期谣言检测上都具有更加优秀的谣言检测性能。

1 相关工作

1.1 基于内容的社交媒体谣言检测

基于内容的谣言检测方法把谣言检测任务定义为一个短文本分类的问题，目前主流的方式是通过深度学习的方式挖掘文本的深层语义信息。例如，Ma 等人[2]考虑手动提取特征会耗费大量的时间，他们首次实现深度学习技术在谣言检测任务上的应用，该方法利用循环神经网络(Recurrent Neural Net, RNN)的隐藏向量表示新闻信息。由于主题相关的谣言检测方法存在着过时的问题，于是 Przybyla 等人[3]提出利用双向长短期记忆网络(Bi-directional Long Short-Term Memory, BiLSTM)从内容中提取写作风格进行谣言检测。Vaibhav 等人[4]观察到讽刺类的文章中句子间具有语义连贯性，而可信文章中句子间的语义连贯性没那么强。他们把文章建模成一张以句子为节点，以句子间的相似度作为边的图，并利用图卷积网络(Grpah Convolutional Network, GCN)融合节点之间的信息以获得谣言表示。为了研究情感特征、语法特征以及语言特征等重要因素对微博谣言检测的影响，王等人[5]提出了一种基于事件-词语-特征异质图的微博谣言检测方法。他们通过挖掘微博事件中情感特征、语法特征以及语言特征，并综合微博评论、文本词语以及文本特征来检测谣言。Zhu 等人[6]考虑到不同领域的谣言之间具有相关性，他们从语义、情感、风格等多个视图对谣言进行建模，并设计了一个领域记忆库模块为出现过的谣言样本生成谣言的潜在领域标签，从而实现从多个视图聚合有用的信息来提升谣言检测的效果。由于数据稀缺的问题，少数民族语言中的谣言很难被检测，于是 Lin 等人[7]提出了一种基于提示学习的零样本谣言检测框架，设计了一种层次提示编码的方法来学习与语言无关的谣言表示。然而，由于文本内容能提供的信息有限，并且由于语义稀疏以及语义模糊的问题，基于文本内容的方式不能够取得很好的效果，所以需要考虑非文本内容的其他因素。

1.2 基于传播模式的社交媒体谣言检测

基于传播模式的谣言检测方法通常将谣言传播过程中所涉及的评论、转发以及传播结构等信息作为谣言检测的重要因素。例如，Liu 等人[8]将谣言的传播模式看作一个随时间变化的线性序列，利用 RNN 和卷积神经网络(Convolutional Neural Network, CNN)对进行谣言建模。Ma 等人[9]认为以往的方法的很大程度上都忽略了或过度简化了与消息传播相关的结构信息，于是他们提出了一种基于

内核的谣言检测方法，通过比较不同结构树的相似性来进行谣言检测。Ma 等人[10]发现基于核的方法由于需要对不同的传播树进行比较，这种方法不但存在大量的开销，而且无法自动学习谣言的高级表示。于是他们将谣言的传播模式建模成树形结构，使用递归神经网络(Recursive Neural Network, RvNN)从两个方向对树进行建模来学习谣言的语义信息和传播线索。实验结果显示，该方法能够大大提高谣言检测的性能。由于已有方法大多都只考虑谣言的传播结构而忽略了谣言的扩散结构，于是 Bian 等人[11]将谣言的传播模式建模成图结构。利用自顶向下的信息图表示谣言的传播信息，利用自底向上的信息图表示谣言的扩散信息。他们使用 GCN 学习两个图的表示，并且使用了一种根节点信息增强的方法来增强节点特征。为了充分挖掘谣言评论间的关联特性并充分提取评论的语义相关性，风等人[12]根据转发和回复关系来构建关系特征图，利用余弦相似度来构建评论的语义特征图以此来提高谣言检测性能。Sun 等人[13]认为大多数基于传播的方法不足以描述不同谣言事件传播能力的差异，并且忽略了谣言和用户的全局联系。他们设计了一个基于超图的全局交互学习模块，用来学习用户在谣言传播中的影响。Zhan 等人[14]发现以往使用图神经网络的谣言检测方法缺乏对谣言的解释性，于是他们考虑通过删除不同的子图来引起谣言检测结果的变化并利用子图提供反事实证据。为了提供多视角的解释，该研究还设计了一个受决定性点过程启发的多样性损失函数来鼓励反事实证据的多样性。为了应对谣言检测任务训练过程中需要大量数据的问题，He 等人[15]通过丢弃传播边、删除传播点以及传播子图提取的方式进行数据增强，并且利用对比自监督学习的方式增强模型学习的效果。为了充分挖掘谣言传播过程中的时序信息，陈等人[16]从挖掘事件之间的全局结构关系以及事件内部消息传播的时序关系的角度出发，提出了一种时序感知的异质图神经网络谣言检测模型。他们利用自注意力机制捕获事件内部转发的时序关系，并利用元素级注意力机制捕获事件的全局结构信息来检测谣言。Gao 等人[17]发现有的谣言检测方法对于异构信息的处理主要集中在直接连接它们的表示，但忽略了多个信息源之间的潜在联系。为了揭示不同信息源之间的联系，他们设计一种异构信息源的对比自监督学习方法，通过给定一个谣言事件的社交模式和语义模式的表示来最大化不同视图之间的互信息来区分与其它谣言事件之间的表示。实验结果显示了该方法的有效性。

由于早期谣言检测是衡量谣言检测方法性能的一个重要指标，大多数最先进的方法尝试使用机器学习方法来进行早期谣言检测，由于手工制作的特征提取是手动执行的，因此他们限制了这些模型来捕获谣言中的高级模式。为了充分利用谣言的传播结构特征以提高早期谣言检测性能，Thota 等人[18]利用 GCN 来构建谣言传播图，并根据随时间出现的谣言响应来实现节点更新。他们利用模式匹配算法来检测跨图生成的具有相似性的子图模式来指导谣言传播图的结构重建，以此来实现高效的图结构更新，从而提高早期谣言检测性能。社交网络平台上大规模、快速地传播虚假信息对公共安全和社会治理构成了严重威胁。因此，在社交网络中对谣言进行大规模传播之前，对谣言进行早期准确的检测至关重要。由于信息传播早期阶段的数据稀疏，基于意见相似度的文本注意机制可以聚合和捕获更多的推文传播结构特征，以帮助提高谣言早期检测的效率。Jia 等人[19]从挖掘谣言的传播结构和文本内容的深度表示的角度出发，提出了一种基于推文-词的图注意力网络的谣言检测方法。他们从谣言的全局图中来分解源推文，并在谣言的传播过程中充分捕获源推文的上下文语义相关性。同时他们设计了一种基于意见相似度的文本注意力机制来聚合和捕获更多的谣言传播特征，以提高早期谣言检测的准确率。

1.3 基于混合特征的社交媒体谣言检测

基于混合特征的谣言检测方法一般融合了谣言内容、谣言传播模式以及谣言参与者的用户描述等特征，他们通过融合大量谣言传播过程可能利用到的信息以提高谣言检测性能。例如，Castillo 等人[20]对 Twitter 上的谣言进行了分析，建立了一种融合主题信息、用户信息和传播模式的有监督训练模型。Lu 等人[21]发现已有的方法无法对谣言的真实性做出解释，于是他们提出了一种具有可解释性的谣言检测方法。他们利用注意力机制来学习源推文和转发之间、源推文和用户之间的相关性。实验结果显示，该方法可以展示源推文中单词权重的分布情况，为检测谣言提供证据词。关等人[22]考虑到大部分谣言检测检测工作忽略了源推文和其它特征的有效融合，以及谣言参与者在谣言传播过程中的作用。于是他们通过将多个谣言参与者属性的组合作为传播节点特征，并使用多个图卷积网络来学习具有不同用户属性组合的传播图表达以此来提高谣言检测性能。Chen 等人[23]观察到传统谣言检测方法集中在利用源推文的内容或者社交上下文来检测谣言并很大程度上忽略了推文中富含的话题信息，于是他们通过探索粗粒度的主题信息以及细粒度的主题信号来检测谣言，同时他们开发了一种多任务学习策略用于来联合学习话题可信度以及用户可信度的预测任务来提高检测性能。由于以往的谣言检测方法大多都从内容或者社交上下文中推断线索，忽略了文本后边的背景知识, 于是 Wang 等人[26]发现已有的谣言检测方法大多都是对各种模态的特征进行简单的连接，这些方法本质上不是基于图神经网络的，同时大多都忽略了文本的背景知识。于是他们考虑把文本信息、知识概念以及视觉信息建立成图的形式并用 GCN 学习它们的表示。Cui 等人[27]侧重于如何将知识图谱中的外部知识与谣言进行融合，他们开发了一种知识感知注意力机制来融合局部的知识，通过构建一个由推文文本、实体以及概念组成的图来学习谣言的语义表示。

一些研究集中利用各种特征来提高早期谣言检测性能。例如，由于社交机器人在早期阶段的行为将造成谣言的广泛传播，而现有的方法没有对真实用户和社交机器人进行明确的区分。Zheng 等人[24]通过考虑社交机器人的行为来进行早期谣言检测，他们利用大量社交机器人和真实用户样本训练一个机器人感知模型，并利用该模型计算谣言参与者为社交机器人的可能性，以此来提高早期谣言检测性能。目前的方法在早期检测中面临局限性，因为它们所依

赖的关键信息在信息传播的初始阶段是不可用的。由于难以仅根据内容进行识别谣言并且社交媒体数据是巨大的、多模态的，以及社交媒体平台促进了谣言的廉价和快速传播。针对这些问题，Ghosh 等人[25]提出了一种使用信息传播路径以及文本的语言模式的谣言检测方法，他们开发了一个因果用户属性推断模型，通过将用户标记成潜在的谣言传播者来识别用户对谣言信息的传播和预防的贡献程度。此外，他们还设计了一个动态注意力机制来衡量在谣言传播过程中不同文本的重要性程度。

通过上述研究可以发现：首先，由于社交媒体上的源推文存在着语义稀疏以及语义模糊的问题，这使早期谣言检测性能不佳。现有的一些针对早期谣言检测的方法本质上还是需要利用谣言传播过程中的丰富信息，而这些信息在谣言产生初期难以获取，对谣言引入背景知识可以解决该问题。由于维基百科富含更加丰富而准确的背景知识，所以本文在已有研究的基础上融合维基百科中的背景知识以提高早期谣言检测性能；其次，由于如何学习具有不同差异性噪声的谣言传播表示仍然是一个待解决问题，本文设计了一种图神经网络结构来解决该问题；最后，由于发布者社交圈信息不依赖于谣言传播结构，它可以为谣言提供丰富信息，所以本文进一步融合发布者社交圈来提升早期谣言检测性能。综上所述，本文提出了一种融合源推文的背景知识、谣言传播模式和发布者的社交圈背景信息的谣言检测方案，实验结果显示该方案在早期谣言检测的优越性。

2 问题定义

形式化地，定义$T=\{T_1,T_2,...,T_m\}$是谣言数据集，$T_i=\{G_i^K,G_i^P,G_i^S\}$是与第$i$个谣言事件相关的数据集合，$G_i^K$是源推文的知识关联图，$G_i^P$是推文集合的传播结构，$G_i^S$代表源推文发布者的社交图。具体地，$G_i^K=\{V_i^K,E_i^K\}$被定义成一个由推文单词和背景知识组成的知识关联图，$V_i^K=\{w_1^i,w_2^i,...,w_m^i,c_1^i,c_2^i,...,c_n^i\}$，$w_j^i$代表源推文中的单词，$m$代表源推文中单词节点的数量，$c_j^i$代表从外部获取到的背景知识，$n$代表背景知识节点的数量。$E_i^K=\{e_{i(s,t)}^K|s,t=1,2,...,m+n\}$代表源推文和背景知识对应单词集的边集，该边集通过滑动窗口生成，其邻接矩阵为$\mathbf{A}_i^K\in\{0,1\}^{(m+n)\times(m+n)}$。对于任意$w_j^i-w_{j+1}^i$、$w_j^i-c_{j+1}^i$或$c_j^i-c_{j+1}^i$如果存在连边，那么其对应的邻接矩阵的元素$a_{i(j,j+1)}^K=1$。$\mathbf{X}_i^K$为知识关联图的特征矩阵，对应的特征向量为图中单词的词向量嵌入。G_i^P被定义成$G_i^P=\{V_i^P,E_i^P\}$，其中，$V_i^P=\{r_i,p_1^i,p_2^i,...,p_n^i\}$代表谣言推文集合，$r_i$代表源推文，$p_j^i$代表第$i$个事件的第$j$个转发推文，$n$代表推文的数量，$E_i^P=\{e_{i(s,t)}^P|s,t=1,2,...,n\}$代表响应推文到转发推文或响应推文的边集。如果推文$p_{j+1}^i$转发了推文$p_j^i$，那么它们之间将产生一条连边$p_j^i\rightarrow p_{j+1}^i$，定义邻接矩阵$\mathbf{A}_i^P\in\{0,1\}^{n_i\times n_i}$表示推文的转发关系，矩阵元素的值如式(1)所示。

$$a_{i(s,t)}^P=\begin{cases}1 & e_{i(s,t)}^p\in E_i^P\\ 0 & otherwise\end{cases} \tag{1}$$

对于事件T_i来说，其转发推文的特征矩阵的表示为$\mathbf{X}_i^P=\{x_{i0}^P,x_{i1}^P,...,x_{in}^P\}$，其中$x_{ij}^P$表示第$i$个事件的第$j$个推文的对应的特征向量。

$G_i^S=\{V_i^S,E_i^K\}$被定义为一个由发布者社交圈经过高度同质化的图[36]，其中$V_i^S=\{u_1^i,u_2^i,...,u_n^i\}$，$u_j^i$代表一个与发布者相互关注的用户，$n$代表与源推文发布者相互关注的用户的数量，$E_i^S=\{e_{i(s,t)}^S|s,t=1,2,...,n\}$，$\mathbf{A}_i^S\in\{0,1\}^{n\times n}$为的$E_i^S$邻接矩阵，$\mathbf{X}_i^S$为社交图的特征矩阵，其特征向量为用户的个人信息的嵌入表示。本文把谣言检测任务定义为一个有监督分类问题，每个事件T_i都和一个真实性标签$y_i\in Y$相关联。

3 RDBKE 系统架构

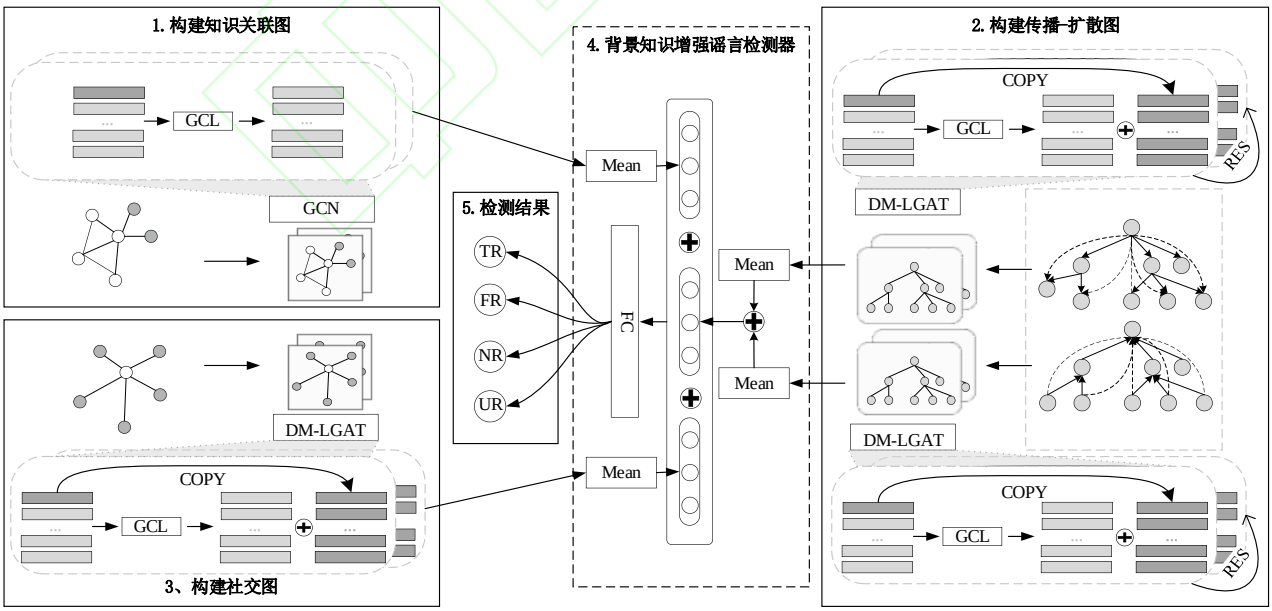

图 1 RDBKE 系统架构

Fig. 1 Architecture of RDBKE

为了表达简便，本文为事件T_i中的有关数据去除下标i。本文提出了一种背景知识增强的多特征融合谣言检测方法

(Multi-feature fusion for rumor detection with background knowledge enhancement, RDBKE)，它融合了谣言的背景知识、谣言传播模式和发布者社交圈。系统架构如图 1 所示，主要包括构建知识关联图、构建传播-扩散图、构建社交图以及背景知识增强谣言检测器 4 部分内容。

3.1 构建知识关联图

在本小节中，首先，介绍了如何从知识图谱以及维基百科中挖掘谣言的背景知识，并使用数据集中的一个谣言事件的源推文来展示不同来源的背景知识的挖掘流程；其次，介绍了如何把源推文构建成图的结构，并展示了背景知识与源推文的融合过程；最后，介绍了如何获取知识关联图的图特征表示。

3.1.1 背景知识挖掘

本文使用实体链接的方式获取源推文的明确实体，该方法能够识别文本中的模糊实体，并将模糊实体与维基百科中的实体进行对齐，从而获取到明确的实体。对于获取到的明确实体，本文使用概念化的方法从知识图谱中提取出每个明确实体对应的概念知识。为了进一步获取到更加丰富、更加准确的背景知识。本文从维基百科中获得每个明确实体对应的定义，并从定义中挖掘谣言的背景知识。

受 Liang 等人[28]的常识性知识补全方法的启发，本文使用了一种知识挖掘的方法来挖掘维基百科定义中与源推文d_k相关的背景知识。首先，对于已获取的明确实体e_i，本文从维基百科中收集了e_i对应的名词、动词以及形容词的定义def_e^i，并过滤了e_i的复数形式、替代形式的定义；其次，本文收集了 ConceptNet[29]知识图谱中具有 isA 关系的知识三元组ck^j。对于收集到的ck^j集合，本文利用 spacy 解析了它们的头部节点ck_{head}^j和尾部节点ck_{tail}^j的词性标注(Part-Of-Speech tagging, POS tagging)，同时统计了 ConceptNet 中每种 POS tagging 模式出现的频率为收集谣言实体中的背景知识做准备。接下来，本文选取了前 k 个频率最高的 POS tagging 模式作为挖掘背景知识词汇的标准。然后，本文从收集到的def_e^i中挖掘这 k 种 POS tagging 模式的对应词汇。直观来说，一个词汇对应 POS tagging 模式的在 ConcepNet 中出现的频率越高，该词汇就越可能成为d_k的背景知识三元组ck^i；最后，本文将def_e^i在d_k中的模糊实体作为头部节点ck_{head}^i，将挖掘到的词汇作为尾部节点ck_{tail}^i，为每个实体的 isA 关系构建了ck^i集合。图 2 展示了背景知识挖掘的流程。

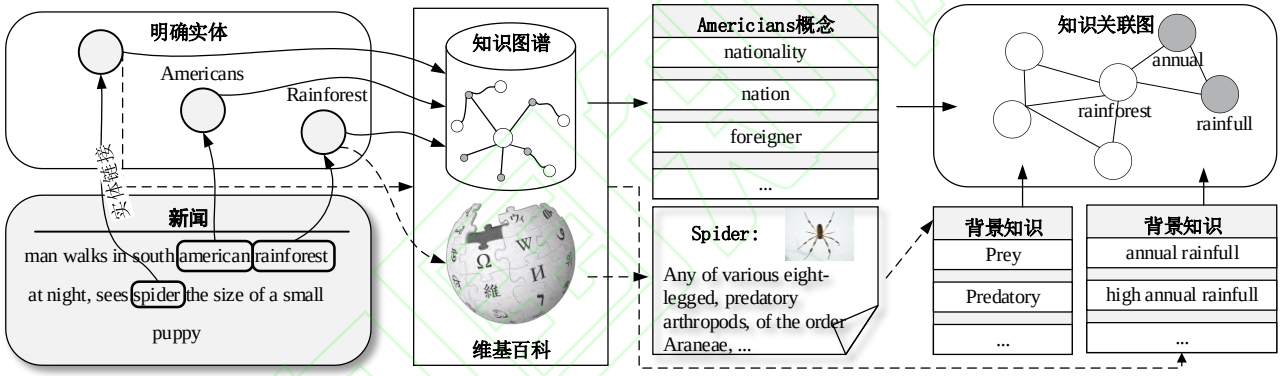

图 2 知识关联图构建过程

Fig. 2 The process of constructing a knowledge association graph

本文使用 Twitter15[2]中的一条源推文d_k：“man walks in south american rainforest at night,sees spider…”展示知识图谱中ck^i的挖掘流程。首先，通过实体链接的方法，本文可以获取到d_k中三个模糊实体：“american”、“rainforest”、“spider”，以及它们的明确实体：e_1：“American”、e_2：“Rainforest”、e_3：“Spider”。然后，本文从知识图谱中寻找与e_i相关的概念知识。以实体e_1为例，本文从知识图谱中获取到概念c_1：“nationality”、c_2：“nation”以及c_3：“foreigner”等概念知识，那么d_k的一个背景知识就可以表述为ck^i={american, isA, nationality}。为展示维基百科中的背景知识的挖掘流程，本文以 spider 在维基百科中的定义def_e^i：“Any of various eight-legged,predatory arthropods,of the order Aranneae,…”为例展示挖掘流程。本文对def_e^i中存在的频繁 POS tagging 模式进行挖掘，从而获取到c_4：“Prey”、c_5：“predatory”等词汇，本文把这些词汇作为背景知识进行引入。与知识图谱类似，spider 对应的对候选背景知识三元组可表述为：ck^i={spider, isA, Prey}、ck^i={spider, isA, predatory}。

3.1.2 知识关联图结构构建

由于源推文大多以短文本的形式存在，这会导致语义稀疏的问题，同时直接连接源推文的词向量表示与背景知识的词向量表示无法有效融合它们之间的语义信息。所以本文把源推文构建成图结构G^D，以便从全局的角度建立单词之间的依赖关系，从而挖掘不同句子之间的潜在联系。G^D中的每个节点为源推文中的单词，本文使用滑动窗口来捕获窗口内的词共现以建立单词与单词之间的联系。

为了建立谣言与背景知识之间的联系，本文把源推文中每个模糊实体与从维基百科中挖掘到的每个背景知识分别进行连接来建立联系。为了进一步加强源推文与背景知识之间的联系，本文把模糊实体与背景知识的连接也看作一条短文本，并同样使用词共现的方式捕获它们之间的联系。由于并不是每个维基百科中的背景知识都与具有很强的语义关联性，为了防止引入过量与谣言无关的背景知识，

本文考虑利用余弦相似度来控制维基百科中背景知识的引入。本文首先获取了模糊实体的词向量表示以及每个背景知识的平均向量的表示。然后计算它们之间的相似度分数，如果该分数大于阈值，则该背景知识与模糊实体的词向量表示表示在语义上相似，就为谣言引入该背景知识，其引入过程可如式(2)和式(3)所示。

$$\overline{x}_c^K = \frac{1}{l}\sum_{j=1}^{l} x_j^K \tag{2}$$

$$a_{(w,c)}^K = \begin{cases} 1 & \cos(\overline{x}_c^K, x_w) > \beta \\ 0 & otherwise \end{cases} \tag{3}$$

其中，x_j^K代表背景知识节点中的一个单词的词向量表示，l代表该知识节点中单词的数量，x_w代表推文单词的词向量表示，β代表相似度阈值。

本文以ck^i={rainforest, isA, Annual rainfull}为例展示知识引入流程。首先，本文获取了“Annual”和“rainfull”的词向量表示x_1^K和x_2^K，并对这两个词向量表示进行相加求和然后求平均，可以获取到ck^i的表示$\overline{x}_c^K$。然后，计算$\overline{x}_c^K$与x_w之间的相似程度，若大于阈值，则为知识关联图引入该知识，同时为知识关联图建立一条“rainforest-Annual-rainfull”的关系连边。最后，本文把该连边看作一个短文本“rainfoest Annual fainfull”，并使用滑动窗口捕获该短文本的词共现。

3.1.3 知识关联图特征学习

本文使用了 GCN 来学习知识关联图的节点表示。针对节点特征x_i^K，GCN 对单词节点表示的变换如式(4)和式(5)所示。

$$x_i^{K'} = \frac{1}{|N_i|}\sum_{j\in N_i} \mathbf{W}_v x_j^K \tag{4}$$

$$\tilde{x}_i^K = f(x_i^{K'}) \tag{5}$$

$x_i^{K'}$代表经过聚合相邻单词节点后的新特征，$|N_i^*|$代表邻居节点的个数，$\mathbf{W}_v$为可学习参数矩阵，f代表激活函数，$\tilde{x}_i^K$代表经过激活函数后的学习到的新特征。最后，本文使用平均池化来获得知识关联图的图特征表示，它可以表述为式(6)的形式。

$$S^{KG} = MEAN\left(\tilde{h}_i\right) \tag{6}$$

3.2 构建传播-扩散图

在本小节中，首先，介绍了如何把谣言的传播模式构建成传播-扩散图；其次，介绍了用于学习具有差异性噪声的谣言传播表示的图神经网络结构；最后，讨论了如何学习传播-扩散图的图特征表示。

3.2.1 传播-扩散图建立

本文基于事件的评论和转发关系，为每个事件构建了自顶向下的传播图G^{TD}以及自底向上的扩散图G^{BU}。在生成图的过程中，本文使用了随机丢弃连边的方法来处理模型过拟合的问题，通过设置丢弃的边的比例来获得新的邻接矩阵$\mathbf{A}^{P'}$，新的邻接矩阵可由式(7)获得。

$$\mathbf{A}^{P'} = \mathbf{A}^P - \mathbf{A}_{drop}^P \tag{7}$$

3.2.2 传播-扩散图特征学习

对于建立完成的谣言传播图以及扩散图，本文通过图神经网络以及特征变换来学习谣言传播模式的节点表示。具体介绍如下：

1)可学习双机制图注意力网络结构

图卷积神经网络和图注意力神经网络已经广泛应用各种与图相关的任务中，已有结论证明不同的图神经网络方法在具有不同性质的数据集上的表现不同[30]，在大多数的情况下，需要采用人工的方法对这两种图神经网络的使用进行交叉验证。由于谣言传播过程中不同事件其中存在噪声节点的含量不同，所以本文引入了可学习图卷积注意力网络(Learnable graph convolutional attention networks, LGAT)[31]，通过在每一层图神经网络中对 GCN 和图注意力网络(Graph Attention Networks, GAT)进行自动插值，能够有效地在模型中对不同图神经网络层进行组合。对于推文节点的特征表示x_i^P，其与 GAT 相邻推文节点的注意力分数计算如式(8)所示。

$$e_{i,j} = Leaky\,\mathrm{Re}lu\left(a^\top[\mathbf{W}_q x_i^P || \mathbf{W}_k x_j^P]\right) \tag{8}$$

其中，$a^\top$、$\mathbf{W}_q$和$\mathbf{W}_k$是可学习参数，x_j^P代表与x_i^P相邻的节点的特征表示。LGAT 的计算方式可由式(9)和式(10)表示。

$$\tilde{x}_i^P = \frac{x_i^P + \lambda_2 \Sigma_{\mathcal{L}\in N_i} x_{\mathcal{L}}^P}{1+\lambda_2|N_i|} \tag{9}$$

$$e_{i,j} = \lambda_1 \cdot \alpha(\mathbf{W}_q \tilde{x}_i^P, \mathbf{W}_k \tilde{x}_j^P) \tag{10}$$

其中，$\tilde{x}_i^P$代表经过插值 GCN 聚合邻居节点后的新特征，N_i代表节点x_i^P的邻居节点集合，λ_1和λ_2是可学习参数，它们的范围在 0-1 之间，用来对 GCN 和 GAT 进行插值学习。α代表式(8)的变换。

为了能够更加有效的捕获节点之间的注意力，受自监督图注意力网络(self-supervised graph attention network, SuperGAT)[32]的启发，本文对 LGAT 的注意力计算方式进行了改进，通过在原有图注意力的基础上添加了点积注意力机制，它能够从链路预测以及捕获节点标签一致性的角度上计算节点注意力，本文将该神经网络结构命名为可学习双机制图注意力网络(Learnable graph attention network with dual mechanisms, DM-LGAT)。其中，两种注意力方式融合后，节点特征的计算方式由式(11)、(12)和(13)所示。

$$e_{i,j} = a^\top([\mathbf{W}_l x_i^P || \mathbf{W}_l x_j^P]) \cdot \sigma((x_i^P)^\top x_j^P) \tag{11}$$

$$a_{i,j} = \frac{\exp(Leaky\,\mathrm{Re}LU(e_{i,j}))}{\sum_{k\in N_i}\exp(Leaky\,\mathrm{Re}LU(e_{i,k}))} \tag{12}$$

$$\tilde{x}_i^P = \sigma\left(\frac{1}{R}\sum_{r=1}^{R}\sum_{j\in N_i}\alpha_{i,j}\mathbf{W}_r x_j^P\right) \tag{13}$$

其中，$a^\top$、$\mathbf{W}_l$和$\mathbf{W}_r$是可学习参数，R代表注意力头数，σ为 Sigmoid 激活函数。

2)传播-扩散图节点特征变换

由于谣言事件的源推文拥有丰富的信息，本文为所有参与讨论的推文都构建一条与源推文相连接的边$r \to p_i$以

此增强转发推文与源推文之间的联系。为了获取谣言传播模式的特征表示，本文以传播图G^{TD}为例展示传播特征学习过程。在图神经网络的第 k 层变换网络中，本文将每个隐藏层节点的节点表示与来自$x_0^{TD(P)}$的源推文表示进行连接得到新的特征向量，连接方法如式(14)、式(15)所示。

$$x_k^{TD(copy)} = copy\left(\left(x_0^{TD(P)}\right)^{source}\right) \tag{14}$$

$$\tilde{x}_k^{TD(P)} = concat\left(x_k^{TD(P)}, x_k^{TD(copy)}\right) \tag{15}$$

对于最后一层图神经网络，本文使用式(16)进行残差连接来防止训练过程中产生的梯度消失问题。

$$h_n^{TD(P)} = x_n^{TD(P)} + \tilde{x}_{n-1}^{TD(P)} \tag{16}$$

最后，本文对传播图G^{TD}以及扩散图G^{BU}使用式(17)和式(18)执行平均池化来获得它们的图特征表示。

$$S^{TD} = MEAN\left(H_n^{TD(P)}\right) \tag{17}$$

$$S^{BU} = MEAN\left(H_n^{BU(P)}\right) \tag{18}$$

3.3 构建社交图

在本小节中，介绍了社交圈的概念以及社交图结构的建立方式，同时给出了学习社交图表示的流程。

3.3.1 社交图建立

社交圈是指兴趣相近、交互频繁的一组具有密切关系的用户。由于同质性的存在，促使这一圈子的形成[33]。由于谣言来源社交圈与非谣言来源的社交圈相比，既有更低的认证率，同时又有更高的主观表达，这种差异为谣言检测提供了可能。对于两个用户节点u_i和u_j，如果两个为相互关注关系，则邻接矩阵$\mathbf{A}_{ij}^{S}$位置的元素为 1，否则为 0。

3.3.2 社交图特征学习

本文中为社交图中的用户节点选取了粉丝数、朋友数、是否认证、创建时间等共 10 个特征。由于用户数据中不同的特征之间的差值具有很大的差异，为了使用户数据更好地适应模型的训练，本文对所有特征进行标准化处理。对于社交图，本文使用 3.2.2 节中的 DM-LGAT 为其主干网络，采用式(14)和式(15)的方式对社交图进行特征变换，并采用平均池化获得最终的社交图表示。

3.4 背景知识增强谣言检测器

本文通过把学习到的谣言事件的知识关联图表示、传播-扩散图表示以及社交图表示进行连接，输入到全连接层中进行分类，其连接过程以及分类过程如式(19)和式(20)所示。

$$F = concat\left(S^{KG}, S^{BU}, S^{TD}, S^{SG}\right) \tag{19}$$

$$y = soft\max\left(FC(F)\right) \tag{20}$$

其中，$FC(\cdot)$为全连接层，$soft\max(\cdot)$函数将输出谣言对应所有类别的概率，其输出为维度为谣言分类的类数。

本文基于最小化交叉熵损失函数对模型的参数进行训练，交叉熵损失函数的计算公式如式(21)所示。

$$L(\theta) = -\sum_{i}\sum_{c}^{C} y_i^c \log y_i + \eta||\theta||_2 \tag{21}$$

其中，θ为整个模型的参数，y_i为样本i的预测标签，y_i^c为样本i的真实值，$\eta||\theta||_2$为L_2正则化项，用来降低模型的过拟合程度。

4 实验及分析

4.1 实验设置

在本小节中，介绍了实验中使用的数据集并列出详细统计信息，同时给出实验中进行对比的基线方法以及本文模型参数细节。

4.1.1 实验数据

实验使用了三个公开的数据集：Twitter15、Twitter16[1]和 Pheme[34]来验证本文方法的有效性，其中 Twitter15 和 Twitter16 分别包含了 1490 和 818 个事件，每个数据集都分为 4 类：虚假谣言(FR)、真实谣言(TR)、非谣言(NR)以及未经验证的谣言(UR)，Pheme 数据集中包含 5802 个事件，数据集中包含真实谣言(TR)和非谣言(NR)，详细的统计数据如表 1 所示。

表 1 数据集信息统计

Table 1 Statistics of dataset information

	Twitter15	Twitter16	Pheme
事件数量	1490	818	5802
真实谣言数量	374	205	3830
虚假谣言数量	370	205	-
非谣言数量	372	205	1972
未经验证谣言数量	374	203	-
节点数量	76351	40867	30376
平均时间跨度（小时）	444	196	18
平均推文数	52	50	6

4.1.2 实验设置

为了验证本文提出方法的有效性，本文选择以下方法作为谣言检测的基线：

DTC[20]: 一种通过特征工程手动提取各种特征，并利用决策树来进行谣言检测的方法。

SVM-TK[9]: 一种使用了时间、结构以及语言特点的随机森林分类器的谣言检测方法。

SVM-TS[35]: 一种基于谣言社会背景随着生命周期变化的线性支持向量机方法。

BU-RVNN[10]: 一种利用递归神经网络学习谣言自底向上扩散结构的谣言检测方法。

TD-RVNN[10]: 一种利用递归神经网络学习谣言自顶向下传播结构的谣言检测方法。

Bi-GCN[11]: 一种利用图神经网络学习谣言传播过程中双向结构信息的谣言检测方法。

RDEA[15]: 一种利用删除连边、删除节点以及子图提取来进行数据增强的对比自监督学习谣言检测方法。

RDMSC[36]: 一种利用融合谣言传播结构和社交圈信息的谣言检测方法。

为了公平比较，本文采用了 K 折交叉验证的方法来进行实验。本文把谣言检测任务为一个分类问题并使用平均准确率(Accuracy, Acc)和 F1-score 作为方法性能的评价指标。

4.1.3 实现细节

在背景知识挖掘中，本文采用了文献[37]提出的实体链接方法，对于获取概念的知识图谱选用了 Probase[38]，POS tagging 模式选择频率最高的前 15 个。构建知识关联图时使用的滑动窗口大小为 3，模型训练时选用的图神经网络层数为 2 层，每层的输出维度为 64 维。使用预训练词向量

Glove 模型来提取单词节点的初始词向量表示。在构建传播-扩散图中，本文使用 TF-IDF 值作为推文的初始特征。模型中图神经网络层数为 2 层，每层的输出维度为 64 维。对于社交图来说，选用的图神经网络层数为 2 层，每层的输出维度为 64 维。使用 Adam 优化器进行训练，学习率为 0.0005，batch_size 为 128，边丢弃率为 0.2，交叉验证折数为 5 折，模型迭代次数为 200，采用早停机制。实验基于 Python3.6 和 Pytorch1.8 实现，服务器配置为 NVIDIA GeForce RTX3070 图形处理器、8 核 CPU、32G 内存。

4.2 实验结果分析

在本小节中，首先，展示了不同谣言检测方法在数据集上的实验结果，并设计了多组消融实验来验证本文方法中不同组件的有效性；其次，对本文方法的早期谣言检测性能进行了分析，并探究没有在谣言传播模式的情况下，背景知识对本文方法产生的影响；最后，探究了背景知识对本文方法的训练时长产生的影响，并列举了一个例子来验证背景知识对提升谣言检测性能的有效性。

4.2.1 实验方案对比

本文对所提出来的方法与对比方法进行了一系列实验，实验结果如表 2 和表 3 所示。

表 2 Twitter15 数据集谣言检测结果

Table 2 Results of rumor detection on Twitter15 dataset

方法	Acc	F1			
		TR	FR	NR	UR
DTC	0.454	0.317	0.355	0.733	0.415
SVM-TS	0.544	0.404	0.472	0.796	0.483
SVM-TK	0.667	0.772	0.669	0.619	0.645
BU-RVNN	0.708	0.759	0.728	0.695	0.653
TD-RVNN	0.723	0.821	0.758	0.682	0.654
Bi-GCN	0.831	0.891	0.847	0.785	0.790
RDEA	0.855	0.903	0.857	0.831	0.816
RDMSC	0.869	0.898	**0.877**	0.863	**0.838**
RDBKE	**0.873**	**0.910**	0.867	**0.886**	0.819

表 3 Twitter16 数据集谣言检测结果

Table 3 Results of rumor detection on Twitter16 dataset

方法	Acc	F1			
		TR	FR	NR	UR
DTC	0.465	0.419	0.393	0.643	0.403
SVM-TS	0.574	0.571	0.420	0.755	0.526
SVM-TK	0.662	0.783	0.623	0.643	0.655
BU-RVNN	0.718	0.779	0.712	0.723	0.659
TD-RVNN	0.737	0.835	0.743	0.662	0.708
Bi-GCN	0.862	0.924	0.851	0.800	0.870
RDEA	0.880	0.937	0.878	0.823	0.875
RDMSC	0.897	0.934	0.870	**0.876**	0.903
RDBKE	**0.904**	**0.946**	**0.887**	0.874	**0.905**

表 4 Pheme 数据集谣言检测结果

Table 4 Results of rumor detection on Pheme dataset

方法	Acc	F1	
		TR	NR
DTC	0.670	0.494	0.755
SVM-TS	0.651	0.639	0.663
SVM-TK	0.785	0.677	0.839
BU-RVNN	0.820	0.731	0.867
TD-RVNN	0.829	0.736	0.873
Bi-GCN	0.835	0.764	0.872
RDEA	0.858	0.785	0.889
RDMSC	0.857	0.786	0.882
RDBKE	**0.870**	**0.803**	**0.902**

由表 2、表 3 和表 4 的实验结果可知，本文提出的 RDBKE 方法在 Twitter15、Twitter16 和 Pheme 数据集上的准确率优于最新对比基线 RDMSC。准确率提高了 0.4%、0.7%和 1.3%。由于 Twitter15 和 Twitter16 数据集富含传播信息，因此本文方法在传播后期提升幅度不是很大，而对于 Pheme 数据集来说，传播信息较少，因此本文方法在传播后期同样具有优越的性能。实验结果表明，本文方法具有更好的谣言检测性能。

具体实验的分析如下：

1)由实验结果观察得到，基于机器学习的谣言检测方法(DTC, SVM-TK, SVM-TS)在各种指标上都低于基于深度学习的方法。其主要原因是基于机器学习的方法需要通过人工方法来提取浅层特征，无法反应出谣言的深层特性，而深度学习方法能够从已有信息挖掘更加丰富的表示。这表明了开展基于深度学习的谣言检测方法的研究的重要性。

2)与基于树的谣言检测方法(BU-RVNN, TD-RVNN)相比，基于图的方法(Bi-GCN, RDEA, RDMSC, RDBKE)具有更高的准确性。其主要原因是基于树的方法注重于谣言传播最终阶段的表现形式，它不能够捕获谣言传播过程中的全局依赖关系，而基于图的谣言检测方法关注谣言传播的整个生命周期，所以能够捕获更多的信息。

3)文本提出的 RDBKE 方法要优于(Bi-GCN, RDEA, RDMSC)，其主要原因是上述谣言检测方法都忽略了谣言背景知识的重要性。去除谣言的背景知识会使得信息缺少与真实性相关的语义信息。所以本文提出利用背景知识补充源推文的语义信息来获得更高的准确率。

4.2.2 消融实验

为了验证 RDBKE 方法中不同组件对谣言检测的有效性，本文设计了以下 5 组消融实验，分别为：去除知识关联图(-KS)、去除社交图(-SG)、去除背景知识(-BK)、DM-LGAT 换为(->GCN)以及 DM-LGAT 换为(->GAT)。每个方法的准确率和 F1-score 的变化如表 4 和表 5 所示。

表 5 Twitter15 数据集消融实验结果

Table 5 Experimental results of ablation on Twitter15

方法	Acc	F1			
		TR	FR	NR	UR
-KS	0.865	0.900	0.860	0.872	0.820
-SG	0.840	0.897	0.849	0.807	0.801
-BK	0.870	0.903	0.865	0.877	0.829
->GCN	0.869	0.902	0.868	0.874	0.827
->GAT	0.871	0.901	0.863	0.871	0.836

表 6 Twitter16 数据集消融实验结果

Table 6 Experimental results of ablation on Twitter16

方法	Acc	F1			
		TR	FR	NR	UR
-KS	0.896	0.944	0.872	0.866	0.893
-SG	0.878	0.928	0.856	0.832	0.883
-BK	0.897	0.940	0.870	0.869	0.901
->GCN	0.895	0.940	0.873	0.826	0.897
->GAT	0.900	0.941	0.885	0.866	0.901

表 7 Pheme 数据集消融实验结果

Table 7 Experimental results of ablation on Pheme

方法	Acc	F1	
		TR	NR
-KS	0.864	0.797	0.896

−SG	0.860	0.797	0.894
−BK	0.862	0.793	0.895
−>GCN	0.862	0.791	0.896
−>GAT	0. 867	0. 796	0. 899

从表 5、表 6 和表 7 中可以发现，去除任何一个组件都会影响 RDBKE 总体的性能。对于知识关联图 KS 来说，由于它自身携带着丰富的源推文信息以及背景知识，去除该组件会导致 RDBKE 丢失大量的语义信息，造成准确率的下降。对于社交图 SK 来说，由于发布者的粉丝、和朋友数和是否认证等信息能够反应发布者的权威性以及可信度，交际活动信息能够反应发布者是否经常与可信用户通信，去除该组件会导致 RDBKE 的性能受到影响。知识图谱以及维基百科中提取的背景知识能够增加源推文的语义信息，所以去除背景知识会影响 RDBKE 的性能。把 DM−LGAT 换成 GCN 或 GAT 会降低模型性能，这是因为不同的谣言事件中的噪声含量不同，GCN 与 GAT 不能够很好地同时学习具有不同差异性噪声的谣言表示。DM−LGAT 能够在一定程度上解决此问题，因此 DM−LGAT 具有更高的准确率。

4. 2. 3 早期谣言检测

早期谣言检测旨在谣言传播的早期阶段检测出谣言，从而实现对谣言的早期干预，这对减小谣言带来的影响具有重要意义。本文分别按照时间、转发推文数量划分，以此来评估各个方法的早期谣言检测效果。时间划分实验结果如图 3 至图 5 所示，转发推文数量划分结果如图 6 至图 8 所示。

1) 按照时间划分

从图 3、图 4 和图 5 中可以观察到，随着截止时间的不断增加，每个方法的性能都会提高。此外，可以观察到 RDBKE 在谣言发布的早期阶段具有更加高的准确率，并且在最后的截止时间内优于其它的基线。这是因为源推文的背景知识以及发布者社交圈信息不依赖于谣言的传播结构。这允许 RDBKE 在早期阶段就拥有更多的信息，同时这些信息会随着谣言传播一直存在，因此 RDBKE 在以时间划分为标准的前提下，具有更加优秀的早期谣言检测性能以及长期谣言检测性能。

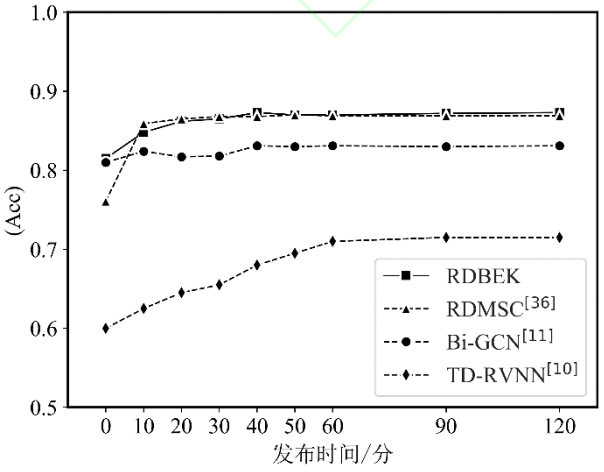

图 3 Twitter15 早期谣言检测时间划分结果

Fig. 3 Results of early rumor detection on Twitter15 by Time division

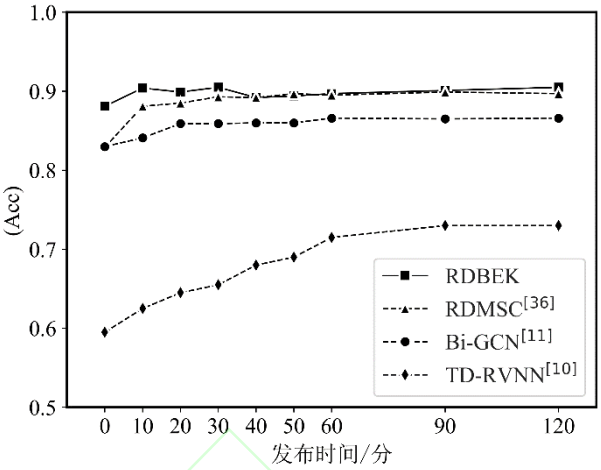

图 4 Twitter16 早期谣言检测实验结果

Fig. 4 Results of early rumor detection on Twitter16 by Time division

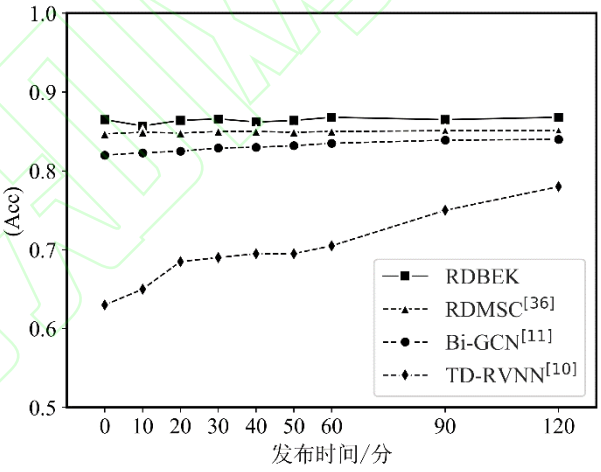

图 5 Pheme 早期谣言检测时间划分结果

Fig. 5 Results of early rumor detection on Pheme by Time division

2) 按照转发推文数量划分

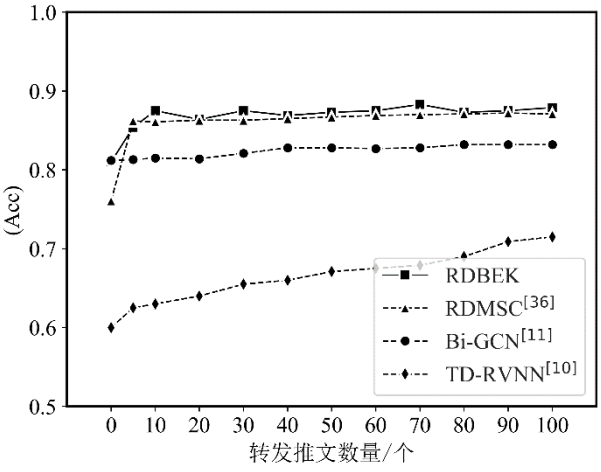

图 6 Twitter15 早期谣言检测数量划分结果

Fig. 6 Results of early rumor detection on Twitter15 by

Quantity division

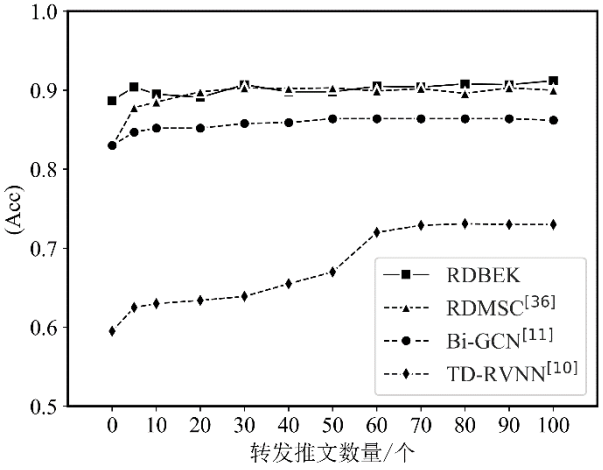

图 7 Twitter16 早期谣言检测数量划分结果

Fig. 7 Results of early rumor detection on Twitter16 by Quantity division

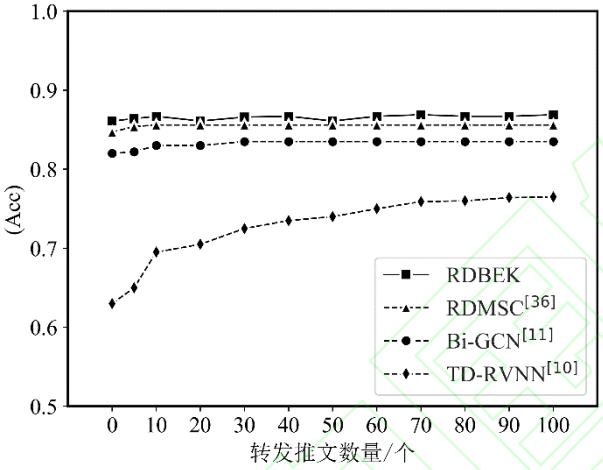

图 8 Pheme 早期谣言检测数量划分结果

Fig. 8 Results of early rumor detection on Pheme by Quantity division

从图 6、图 7 和图 8 中可以观察到，随着源推文对应的转发推文数量不断增加，每个方法的性能都会提高。此外，可以观察到 RDBKE 在谣言发布的初期以及具有少量转发推文的时候具有更加高的准确率，并且随着转发推文的不断增加不断领先其它方法。和按照时间划分的早期谣言检测实验结果一样，因为源推文的背景知识不依赖于谣言的传播结构。这允许 RDBKE 在早期阶段就拥有更多的信息，因此 RDBKE 在以转发推文数量划分为标准的前提下，同样具有更加优秀的早期谣言检测性能以及长期谣言检测性能。

综上所述，由于源推文的背景知识不依赖于谣言的传播结构存在，所以无论是在以时间划分为标准还是以转发推文数量划分为标准的早期谣言检测的情况下，都具有优秀的谣言检测性能。所以本文方法可以在传播信息缺少或者源推文缺乏互动的情况下，依赖于源推文的背景知识来提高谣言检测性能，在传播信息丰富或者源推文富含互动的情况下，谣言的背景知识可以和传播信息联合，进一步提升早期谣言检测性能，这体现了本文方法的优越性。

本文设计多组实验探究在极端情况下，即没有谣言传播模式时不同来源的背景知识对谣言检测的影响。本文设计了纯文本(TEXT)、文本+知识图谱背景知识(+KG)、文本+维基百科背景知识(+WK)、文本+知识图谱背景知识+维基百科背景知识(+KG+WK)几种方案来证明不同知识的有效性，结果如图 9 所示。

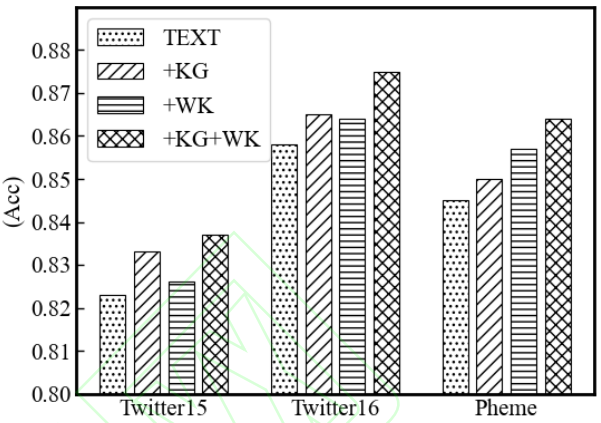

图 9 背景知识消融的结果

Fig. 9 Results of background knowledge ablation

从图 9 中可以看出知识图谱中的背景知识能够补充源推文的语义信息。然而仅仅添加知识图谱中的背景知识的模型的准确性不如添加了知识图谱和维基百科的背景知识的模型，这表明可以维基百科进一步提供更加丰富以及准确的背景知识来辅助谣言检测。同时从图中可以发现在 Twitter15 和 Twitter16 数据集中添加维基百科背景知识的模型的准确率会略低于添加知识图谱背景知识的模型，这其中原因是由于许多实体无法从维基百科中获取到对应的实体定义以及从定义中获取到的背景知识的语义信息不如知识图谱中获取到的语义信息丰富，而对于 Pheme 数据集来说，它从维基百科中的实体定义中获取的背景知识比从知识图谱中获取的背景知识更加丰富。综上所述，知识图谱和维基百科中的背景知识都能够提升谣言检测的效果。

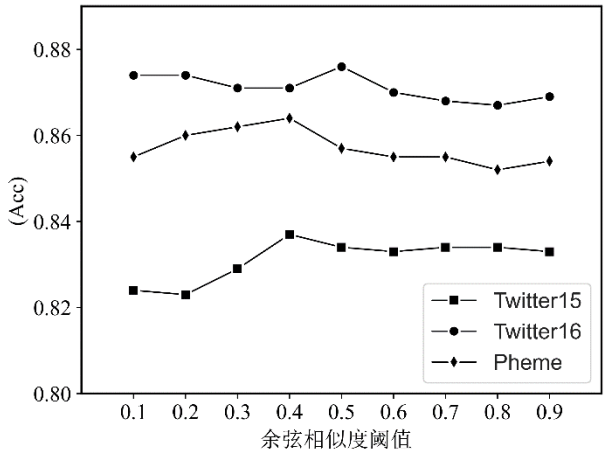

图 10 相似度阈值的影响

Fig. 10 Influence of similarity threshold

为了探究在引入维基百科的背景知识时，不同的阈值

β 对模型性能产生的影响，本文通过将阈值的范围设置为 0.1-0.9 来验证模型的有效性，图 10 显示了不同阈值在所有基准数据集上的影响。

从图 10 中可以看到, 当阈值在 0.1-0.3 的范围时，在 Twitter15 数据集和 Pheme 数据集上，模型的性能低于阈值为 0.4 时的性能。在 Twitter16 数据集上，模型的性能低于阈值为 0.5 时的性能。这其中的原因是当阈值过低的时候，容易引入一些与源推文无关的背景知识，这会导致背景知识无法很好地补充源推文语义信息；当阈值在 0.6-0.9 的时候，在 Twitter15 数据集和 Pheme 数据集上，模型的性能低于阈值为 0.4 时的性能，在 Twitter16 数据集上，模型的性能低于阈值为 0.5 时的性能。这是因为当阈值过高的时候，使与文中实体相似的背景知识过少，不能为源推文提供丰富的语义信息。

综上所述，可以发现谣言的背景知识可以补充源推文的语义信息为谣言检测提供更多的信息，由于纯文本的情况下不需要依赖谣言的传播模式，这也就解释了为什么谣言的背景知识可以提高早期谣言检测的性能。

4. 2. 4 性能分析

为了分析引入背景知识会对模型训练的时间产生的影响，本文对源推文中的实体节点、从知识图谱中的获取到的背景知识节点以及从维基百科中获取到的背景知识数量进行了统计，其结果如表 8 所示。

表 8 实体节点和知识节点数量统计

Table 8 Statistics on the number of entity and knowledge nodes

节点统计	Twitter15	Twitter16	Pheme
实体节点	1368	730	2190
知识图谱知识节点	5422	2735	6570
维基百科知识节点	2452	1401	7763

从表 8 中可以观察到，Twitter15 的实体节点和知识节点的数量几乎都是 Twitter16 的两倍，这其中的原因是 Twitter15 的谣言事件数量几乎是 Twitter16 的两倍，而对于 Pheme 数据集来说，由于源推文中的平均文本内容长度短于 Twitter15 和 Twitter16 中源推文的平均文本内容长度，并且所有事件高度围绕 5 个事件进行讨论，所以该数据集中富含的实体节点数不会远高于 Twitter15 和 Twitter16 数据集。

表 9 背景知识对模型训练时间的影响

Table 9Influence of Background Knowledge on Model Training Time

方法	Twitter15	Twitter16	Pheme
KS 训练时间/秒	10.743	8.347	15.689
KS-BK 训练时间/秒	10.517	7.325	15.162

从表 9 中可以看到加入背景知识前后对 RDBKE 训练时间产生的影响，通过观察可以发现，尽管 Twitter15 数据集中的数据量几乎是 Twitter16 的两倍，但是每个 epoch 的训练时间却不是两倍。这其中的原因可能是因为 Twitter15 相比于 Twitter16 提供了更多的数据给 RDBKE，尽管模型需要训练的数据量变大了，但是数据提供的信息也增加了，这使得 RDBKE 可以加快其训练过程。而对于 Pheme 数据集来说，虽然谣言事件数远大于其它两个数据集，但是由于每个事件的平均转发推文数量小于其它两个数据集，所以其训练时间相比其它两个数据集并不会增加很多。同时训练时间也会受到数据集自身的影响，在数据量相同的情况下，训练时间也可能不同。通过对比同一个数据集添加背景知识前后的每个 epoch 的训练时间可以发现，添加了背景知识的 RDBKE 的训练时间会略大于没添加背景知识的训练时间，这表明背景知识的加入会对 RDBKE 的训练时长产生一定的影响，这与带来的性能提升具有正相关关系。

4. 2. 5 案例分析

本文使用知识图谱和维基百科为源推文提供背景知识，为了验证从维基百科中提取的背景知识能够解决模型在已经引入知识图谱的情况下无法对模糊推文进行判断的问题，本文选取了一个示例并在图 11 中展示了它的有效性。

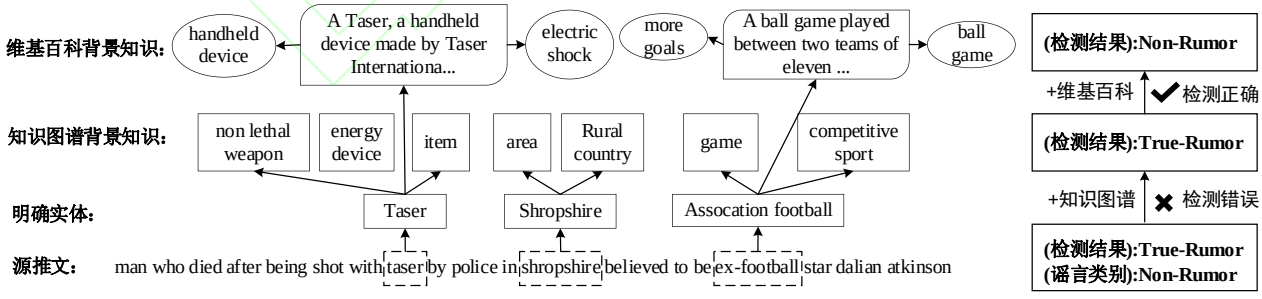

图 11 维基百科知识对谣言预测错误结果的纠正过程

Fig. 11 The process of correcting false predictions based on Wikipedia knowledge

源推文中的模糊实体 taser 在语义上由实体 Taser 进行链接，由于 taser 在训练数据中出现频率较低，通过从知识图谱中获取的知识，例如“non lethal weapon”、“energy device”等，可以丰富该实体的语义信息。而“ex-football”是一个前缀词，同样也属于稀少词汇，能够通过实体链接工具定义到更加明确的实体“Assocation football”，进而从知识图谱中获取更加准确的背景知识。此外，为了获取更多更加准确的背景知识，本文从维基百科中获取了与“Taser”和“Assocation football”有关的知识，其中“handheld device”和“electric shock”补充了知识图谱中

缺失的知识，而“ball game”进一步为“Assocation football”提供了进一步的解释。模型经过两轮的知识融合后，模型最终检测出该谣言的正确类别。

5 结论

由于谣言检测无法在早期阶段获得传播过程中的各种信息，因此会导致早期谣言检测性能不佳的问题。本文提出了一种融合背景知识、谣言传播模式以及社交圈的谣言检测方法。首先，通过从知识图谱和维基百科中获取与谣言相关的背景知识来建立与源推文相关的知识关联图；其次，设计了一种具有双重注意力的图神经网络结构来学习谣言的传播-扩散图表示以解决模型在学习具有差异性噪声的谣言传播表示时存在的困难问题；最后，本文结合谣言发布者社交圈来判断谣言的真实性，这在一定程度上提高了早期谣言检测性能。实验结果表明，本文方法性能优于所有对比的谣言检测方法，并且在早期谣言检测上有着明显的优势，证明了本文方法的有效性。由于社交媒体中富含丰富的对象以及社会关系，今后的研究将考虑利用异质图神经网络对这些对象以及社会关系进行建模来解决谣言检测问题。

References:

[1] MA Y Y.Overview of rumor detection on social platforms[J].Changjiang Information & Communications,2022,35(4):26-29.

[2] 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),2016:3818-3824.

[3] Przybyla P.Capturing the style of fake news[C]//Proceedings of the AAAI conference on artificial intelligence(AAAI),2020:490-497.

[4] Vaibhav V,Mandyam R,Hovy E.Do sentence interactions matter? leveraging sentence level representations for fake news classification[C]//Proceedings of the 13th Workshop on Graph-Based Methods for Natural Language Processing,2019:134-139.

[5] WANG Y W,FENG L Z,WANG W Q,et al.Weibo rumor Detection based on heterogeneous graph of event-word-feature[J].Journal of Chinese Information Processing,2023,37(9):161-174.

[6] Zhu Y,Sheng Q,Cao J,et al.Memory-guided multi-view multi-domain fake news detection[J]. IEEE Transactions on Knowledge and Data Engineering,2022,35(7):7178-7191.

[7] Lin H,Yi P,Ma J,et al.Zero-shot rumor detection with propagation structure via prompt learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence(AAAI),2023:5213-5221.

[8] Liu Y,Wu Y F.Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks[C]//Proceedings of the AAAI conference on artificial intelligence(AAAI),2018:354-361.

[9] 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(ACL),2017:708-717.

[10] Ma J,Gao W,Wong K F.Rumor detection on twitter with tree-structured recursive neural networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(ACL),2018:1980-1989.

[11] Bian T,Xiao X,Xu T,et al.Rumor detection on social media with bi-directional graph convolutional networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence(AAAI),2020:549-556.

[12] FENG L Z,LIU F R,WANG Y W.Rumor detection method based on graph convolution network and attention mechanism [J]. Data Analysis and Knowledge Discovery,2023,doi:10.11925/infotech.2096-3467.2023.0237.

[13] Sun L,Rao Y,Lan Y,et al.HG-SL:jointly learning of global and local user spreading behavior for fake news early detection[C]//Proceedings of the AAAI conference on artificial intelligence(AAAI),2023:5248-5256.

[14] Zhang K,Yu J,Shi H,et al.Rumor detection with diverse counterfactual evidence[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining(SIGKDD),2023:3321-3331.

[15] He Z,Li C,Zhou F,et al.Rumor detection on social media with augmentations[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR),2021:2020-2024.

[16] CHEN L W,SONG Y R,SONG B.Sequence-aware heterogeneous graph neural rumor detection[J]. Journal of Chinese Computer Systems,2024,45(1):45-51.

[17] Gao Y,Wang X,He X,et al.Rumor detection with self-supervised learning on texts and social graph[J]. Frontiers of Computer Science(FCS),2023,doi:10.1007/s11704-022-1531-9.

[18] Thota N R,Sun X,Dai J.Early rumor detection in social media based on graph convolutional networks[C]//International Conference on Computing,Networking and Communications (ICNC),2023:516-522.

[19] Jia H,Wang H,Zhang X.Early detection of rumors based on source tweet-word graph attention networks[J]. PLoS One,2022,doi:10.1371/journal.pone.0271224.

[20] Castillo C,Mendoza M,Poblete B.Information credibility on twitter[C]//Proceedings of the 20th international conference on World wide web(WWW),2011:675-684.

[21] Lu Y J,Li C T.GCAN:graph-aware co-attention Networks for explainable fake news detection on social media[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics(ACL),2020:505-514.

[22] GUAN C S,BING W L,LIU Y H,et al.Multi-feature fusion rumor detection method based on graph convolutional network[J].Journal of Zhengzhou University(Engineering Science),2023,doi:10.13705/j.issn.1671-6833.2024.01.011.

[23] Chen Z,Wang L,Zhu X,et al.TSNN:a topic and structure aware neural network for rumor detection[J]. Neurocomputing,2023,doi:j.neucom.2023.02.016.

[24] Huang Z,Lv Z,Han X,et al.Social bot-aware graph neural network for early rumor detection[C]//Proceedings of the 29th International Conference on Computational Linguistics(COLING),2022:6680-6690.

[25] Ghosh S,Mitra P.How Early Can We Detect? Detecting Misinformation on Social Media Using User Profiling and Network Characteristics[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases(ECML-PKDD),2023:174-189.

[26] Wang Y,Qian S,Hu J,et al.Fake news detection via knowledge-driven multimodal graph convolutional networks[C]//Proceedings of the International Conference on Multimedia Retrieval(ICMR),2020:540-547.

[27] Cui W,Shang M.KAGN:knowledge-powered attention and graph convolutional networks for social media rumor detection[J].Journal of big Data,2023,doi:10.1186/s40537-023-00725-4.

[28] Liang Z,McGuinness D L.Commonsense knowledge mining from term definitions[C]//Commonsense Knowledge Graphs (CSKGs) Workshop of AAAI,2021,doi:10.48550/arXiv.2102.00651.

[29] Speer R,Chin J,Havasi C.ConceptNet 5.5:an open multilingual graph of general knowledge[C]//Proceedings of the 31th AAAI Conference on Artificial Intelligence(AAAI),2017:4444-4451.

[30] Baranwal A,Fountoulakis K,Jagannath A.Effects of graph convolutions in multi-layer networks[C]//11th International Conference on Learning Representations(ICLR),2023,doi:10.48550/arXiv.2204.09297.

[31] Javaloy A,Sanchez-Martin P,Levi A,et al.Learnable graph convolutional attention networks[C]//11th International Conference on Learning Representations(ICLR),2023,doi:10.48550/arXiv.2211.11853.

[32] Kim D,Oh A.How to find your friendly neighborhood:graph attention design with self-supervision[C]//9th International Conference on Learning Representations(ICLR),2021,doi:10.48550/arXiv.2204.04879.

[33] Wang M,Zuo W,Wang Y.An improved density peaks-based clustering method for social circle discovery in social networks[J].Neurocomputing,2016,100(179):219-227.

[34] Zubiaga A,Liakata M,Procter R.Exploiting context for rumour detection in social media[C]//International Conference on Social Informatics,2017:109-123.

[35] Ma J,Gao W,Wei Z,et al.Detect rumors using time series of social context information on microblogging websites[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management(CIKM),2015:1751-1754.

[36] 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,doi:10.1016/j.ins.2023.119083.

[37] Assante M,Candela L,Castelli D,et al.Enacting open science by D4Science[J].Future Generation Computer Systems(FGCS),2019,doi:10.1016/j.future.2019.05.063.

[38] Wu W,Li H,Wang H,et al.Probase:a probabilistic taxonomy for text understanding[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data(SIGMOD),2012:481-492.

附中文参考文献：

[1] 马圆圆.社交平台上的谣言检测综述[J]. 长江信息通信,2022,35(04):26-29.

[5] 王友卫,凤丽洲,王炜琦,等.基于事件-词语-特征异质图的微博谣言检测新方法[J]. 中文信息学报,2023,37(9):161-174.

[12] 凤丽洲,刘馥榕,王友卫.基于图卷积网络和注意力机制的谣言检测方法[J].数据分析与知识发现,2023,doi:10.11925/infotech.2096-3467.2023.0237.

[16]陈林威,宋玉蓉,宋　波.时序感知的异质图神经谣言检测[J].小型微型计算机系统, 2024,45(1):45-51.

[22] 关昌珊,邢万龙,刘雅辉,等.基于图卷积网络的多特征融合谣言检测方法[J].郑州大学学报(工学版),doi:10.13705/j.issn.1671-6833.2024.01.011.