基于多关系图注意力网络的社交机器人检测

孟令君，陈鸿昶，王庚润

引用本文

孟令君，陈鸿昶，王庚润. 基于多关系图注意力网络的社交机器人检测[J]. 计算机科学，2025，52(1)：298-306.

MENG Lingjun, CHEN Hongchang, WANG Gengrun. Social Bots Detection Based on Multi-relationship Graph Attention Network [J]. Computer Science, 2025, 52(1): 298-306.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231100161

基于多关系图注意力网络的社交机器人检测

孟令君[1]　陈鸿昶[2]　王庚润[2]
1 郑州大学网络空间安全学院　郑州 450003
2 信息工程大学国家数字交换系统工程技术研究中心　郑州 450003
(893099552@qq.com)

摘　要　现阶段社交机器人已经广泛存在于社交平台,社交机器人的存在使得网络上的舆论环境可以被人为操纵,这样不仅损害了绿色和谐的网络环境,同时也导致人们正常的网络生活受到极大影响。现有的检测方法可以分为基于特征、基于文本和基于图的方法,其中基于图数据的检测方法大多忽略了图中关系的异质性,并且由于图神经网络存在过渡平滑现象而不能进行深度检测。针对这一问题,提出基于多关系图注意力网络的社交机器人检测方法,在训练时首先将不同关系下的子图抽取出来,然后对子图中的节点采用注意力机制进行聚合,在不同关系下进行节点表示学习并得到节点表示,最后利用通道注意力融合不同关系下的同一节点得到节点表示;同时采用基于LSTM注意力的后连接操作让节点可以自适应地选择邻域进行聚合,以此来缓解过度平滑现象。在Cresci15,Twibot20和MGTAB这3个数据集上的实验结果表明,与11个模型中评价指标的最优值相比,该模型的准确率分别提升了0.47%,1.19%和0.38%,验证了多关系图注意力网络进行社交机器人检测的有效性。

关键词:异质图;图注意力;节点表示学习;LSTM注意力;社交机器人
中图分类号　TP391

Social Bots Detection Based on Multi-relationship Graph Attention Network

MENG Lingjun[1],CHEN Hongchang[2] and WANG Gengrun[2]
1 School of Cyberspace Security,Zhengzhou University,Zhengzhou 450003,China
2 National Digital Switching System Engineering & Technological R&D Center,Information Engineering University,Zhengzhou 450003,China

Abstract　At present,social bots have gained extensive utilization across social platforms and the existence of social bots makes the public opinion environment on the network artificially manipulated. This not only compromises the integrity of a healthy and harmonious online atmosphere but also significantly disrupts people's regular online activities. Existing detection methods can be divided into feature-based,text-based,and graph-based methods. However,graph-based detection methods predominantly ignore the heterogeneous relationships,and cannot perform deep detection due to the transition smoothing phenomenon in graph neural networks. To solve the above problems,a social bots detection method based on a multi-relationship graph attention network is proposed. Firstly,we extract subgraphs with different relationships,then apply the attention mechanism to aggregate the nodes within the subgraph and conduct node representation learning across diverse relationships,resulting in the acquisition of node representations. Finally,we use channel attention to fuse the same node under different relationships to obtain node representation,while using the post-connection operation based on LSTM attention to allow nodes to adaptively select neighborhoods for aggregation,thereby alleviating the over-smoothing phenomenon. Experiments are conducted on three datasets:Cresci15,Twibot20,and MGTAB,and the experimental results show that,compared with the optimal values of the evaluation indicators of 11 models,the accuracy of the model is increased by 0.47%,1.19% and 0.38%,respectively,which demonstrates the effectiveness of the multi-relationship graph attention network for social bots detection.

Keywords　Heterogeneous graph,Graph attention,Nodes representation learning,LSTM attention,Social bots

到稿日期:2023-11-27　返修日期:2024-05-06
基金项目:国家自然科学基金(61803384);嵩山实验室项目(列入河南省科技重大专项)(221100210700-2)
This work was supported by the National Natural Science Foundation of China(61803384) and Program of Song Shan Laboratory(included in the management of Major Science and Technology Program of Henan Province)(221100210700-2).
通信作者:王庚润(wanggengrun@gmail.com)

1 引言

社交机器人（Social Bots）是由人类操纵者所设置的，由自动化算法软件所操纵的社交网络账号集群。其通常通过模仿、模拟和仿真人类在社交网络中的状态和行为，伪装成真实的用户，有组织地与真实用户交互，以达到依照人类操纵者的意图影响目标受众的目的。社交机器人已经广泛存在于各个社交平台，从各个方面影响着我们的生活，例如：购物平台上的“刷单刷评论”的机器人会影响消费者对商品的判断；社交媒体上的社交机器人会宣扬极端意识形态进而影响青少年的认知；有些政治机器人会被用来支持特定的政治家和政治立场；有些舆论引导机器人会引导人们对一些事件的看法，从而消除群体极化。由上述例子可以看出社交机器人包括恶意机器人、引导机器人等。本文检测的社交机器人是针对由人类操纵者所设置的、由自动化算法软件所操纵的社交网络账号集群。

早期的社交机器人检测主要依赖于特征工程，通过提取并评估大量用户特征与传统的机器学习分类器进行结合对社交机器人进行检测。随着深度学习的发展，基于文本和基于图的方法也成为社交机器人检测的主流方法。基于文本的方法分析用户发布内容，通过自然语言处理技术对用户发布的文本进行编解码并获取文本中的语义信息，以此来检测社交机器人。基于图的方法将社交网络中的用户和用户之间的通联关系建模为图并采用图神经网络来识别社交机器人。

然而，随着生成式语言大模型的发展，越来越多的社交机器人开始模仿真实用户的发布内容与语言风格，这使得基于文本的社交机器人检测受到一定的影响。基于图的社交机器人检测大多基于异构图，然而大部分检测方法只关注节点之间的异质性而忽略了节点之间通联关系的异质性。为此，本文提出 MR-GAN（Multi-Relationship Graph Attention Network）模型，该模型根据通联关系的异质性提取不同关系的子图，并引入图注意力机制对图上节点进行表示学习，最后提出后连接机制来使得节点可以自适应地选择邻域，从而提高机器人检测的效率，维护社交网络的绿色健康。

2 相关工作

现有的社交机器人检测技术大多是基于特征、基于文本和基于图的方法。

基于特征的社交机器人检测利用用户特征和用户的推文数据进行检测。2020 年 Maria 等提出从每个用户的元数据和最新的 20 条推文中提取出 36 个特征，并使用随机森林对用户进行分类[1]。2021 年 Wu 等从基于元数据、基于互动、基于内容和基于时间的 4 类数据中提取 30 个特征以区分微博中的社交机器人和正常用户[2]。同年 Abreu 等选择了 5 个基本的推特功能，并使用 4 种机器学习算法对 Twitter 用户进行分类[3]。2023 年 Hu 等建立多维动态特征社交机器人账号检测模型，提出一种改进的基于 AUC 决策树分类评价指标随机森林优化算法[4]，但是其中的操纵者可以通过隐藏或伪造特征对社交机器人的特征进行伪装，以此逃避检测。

基于文本的社交机器人检测应用 NLP 技术对用户描述和用户发布内容信息进行编码检测。2018 年 Sneha 等从用户元数据中提取上下文特征，并将其作为辅助输入提供给处理文本的 LSTM 深度网络，通过文本能检测社交机器人账户[5]。2021 年 Guo 等构造了一个语料库的异构图，将预先训练好的 BERT 嵌入的单词和用户描述作为图节点，将 BERT 和 GCN 输出的结果作为最终预测[6]。2022 年 Hayawi 等使用 LSTM 对用户的文本信息进行编码，并将其与使用密集层从元数据中提取的特征进行融合以检测社交机器人[7]。2023 年 Wu 等通过一个三重网络生成单词、句子和账户嵌入并使用度量学习技术细化原始嵌入，利用用户文本检测机器人[8]，但是操纵者可以窃取真实用户发布的内容，通过仿照真实用户发布内容来逃避检测。

随着图神经网络的出现与发展，人们发现图数据中蕴含着丰富的社交网络结构信息，这种拓扑结构可以作为检测社交机器人的重要特征。目前基于图的方法相较于其他两种方法更加全面，因为基于图的方法不仅考虑节点的特征以及节点文本信息，还关注节点间的拓扑结构，挖掘节点更深层次的特征，所以本文主要研究基于图的社交机器人检测。2018 年 Cornelissen 等利用网络拓扑结构和无监督相学习结合的方法来检测社交机器人，但是该方法依赖一定的先验知识，目前社交机器人网络拓扑结构多样，该方法的效果受到影响。2019 年 Wang 等为了提升检测的准确率，在利用社交网络拓扑结构的基础上，同时使用元路径技术和异构注意力网络模型实现了对社交机器人的检测[9]，但是该方法依赖节点和元路径的特征选择。2021 年 Li 等为了挖掘节点的深层特征，在社交网络中抽取正常用户和异常用户，构建多用户关系图，其次采用相关性感知的 GNN 框架来学习用户的隐藏特征，并对异常用户进行判别[10]，但是该方法在融合多用户关系时忽略了关系间的异质性。2021 年 Feng 等为解决检测的泛化问题提出了 Satar 自监督表示学习框架，对不同的机器人检测任务进行微调[11]，但是该方法需要丰富的用户信息对其进行训练，缺少任何方面的信息都会对模型产生较大影响；同年，他们还提出了 BotRGCN，该方法通过提取丰富的用户特征来构建完善的节点特征并使用关系图卷积网络（RGCN）来检测社交机器人[12]，但是该方法将用户推文和用户简介特征进行压缩，使得推文中的语义模糊。2022 年 Li 等为了优化图神经网络的搭建过程提出了基于强化学习的图神经网络，使用强化学习对图神经网络的框架进行搜索，并进行节点间聚合，在最终分类器中进行机器账号检测[13]，但是该方法搜索得到的模型架构存在层数过深的情况，这会使得模型过平滑进而影响检测效果。2023 年 Xu 等提出了一种结合主动学习与关系图卷积神经网络（RGCN）的检测方法，用主动学习扩充数据集并用 RGCN 获取网络结构信息[14]，但是该方法需要人为对数据集进行标注，人工成本和时间成本较高。

当前，基于图的检测方法虽然比较成熟，但均未考虑不同类型边对节点的影响，并且由于图神经网络存在过平滑现象，因此以上模型无法聚合深层的节点。为此，我们将在已有特征基础上丰富节点特征；然后采用 Transformer 注意力和通道注意力构建一个多关系图注意力网络，利用边的异质性实现不同关系下节点的聚合；最后通过基于 LSTM 注意力的

后连接操作缓解过平滑现象，使得模型可以聚合深层节点，挖掘深层特征。

3 多关系图注意力网络

针对上述问题，本文提出一种多关系图注意力网络(Multi-Relationship Graph Attention Network)，该模型由多关系图注意力网络进行不同节点间聚合并采用 LSTM 注意力实现多层网络间的聚合，其结构如图 1 所示。该模型由节点特征编码、多关系注意力网络和基于 LSTM 注意力的后连接网络等部分构成。

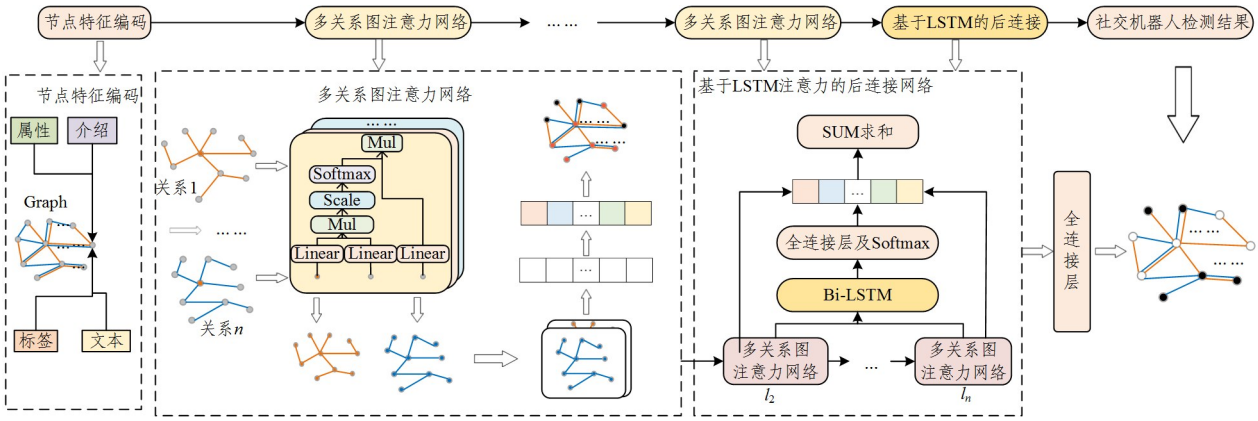


图 1 MR-GAN 模型结构图

Fig. 1 Structure diagram of MR-GAN model

3.1 节点特征编码

本文选取了用户简介、用户推文、用户属性以及用户标签作为节点特征，在用户属性维度中添加了用户推文数量和简介长度两维特征，使得用户属性更加完整，并利用 MLP 对其进行编码。对于用户简介和用户推文，受到文献[15]得到句向量(Sentence Encoder)的启发，我们将 Transformer 模型替换为具有鲁棒性的 RoBERTa，并利用预训练的 RoBERTa[16]网络对每个词进行编码，对得到的词向量进行平均求和进而得到句子向量的表示。对于一个用户的多条推文，我们将每个句子对应的向量进行加权平均得到用户推文特征。最后将这 4 种特征进行维度上的拼接得到节点表示。

$$F_{t,d}=\varphi(\boldsymbol{W}\cdot RoBERTa(f_{t,d})+b) \tag{1}$$

$$Feature=[\boldsymbol{F}_t,\boldsymbol{F}_d,\boldsymbol{F}_n,\boldsymbol{F}_c]\in\mathbb{R}^{D\times 1} \tag{2}$$

其中，F 表示这 4 类特征的向量表示，f 表示特征的原始文本信息，$\boldsymbol{W}$ 为可学习的参数矩阵，b 为神经网络偏置，φ 表示 RELU 激活函数，$Feature$ 表示节点特征，D 表示节点嵌入维度。

3.2 多关系图注意力网络

图是由节点的有穷非空集合 $V(G)$和顶点之间边的集合 $E(G)$组成。根据 3.1 节，在得到节点特征后根据图中边的异质性，将图分解为不同的子图 $Graph_{r1},\cdots,Graph_{rn}$。受到 Transformer 模型的启发，本文不使用传统的 GCN 网络聚合节点信息，而是选择 Transformer 中的多头注意力机制对节点信息进行聚合[17]，中心节点 n_i^r 作为 query，其邻接节点 n_j^r 作为 key 和 value。

$$q_i^r=\boldsymbol{W}_q^r\cdot n_i^r+\boldsymbol{b}_q^r \tag{3}$$

$$k_j^r=\boldsymbol{W}_k^r\cdot n_j^r+\boldsymbol{b}_k^r \tag{4}$$

$$v_j^r=\boldsymbol{W}_v^r\cdot n_j^r+\boldsymbol{b}_v^r \tag{5}$$

其中，$\boldsymbol{W}$ 和 $\boldsymbol{b}$ 表示可学习矩阵。然后通过计算不同节点之间的注意力权重实现节点信息的聚合。

$$N(i)=\{1\in edge(n_i,n_1),\cdots,j\in edge(n_i,n_j)\} \tag{6}$$

$$n_i^r=\beta_i^r\boldsymbol{W}_i^r\cdot n_i^r+(1-\beta_i^r)\sum_{j\in N(i)}\alpha_{i,j}v_j^r \tag{7}$$

$$\alpha_{i,j}=\mathrm{softmax}\left(\frac{(q_i^r)^{\mathrm{T}}(k_j^r)}{\sqrt{d}}\right) \tag{8}$$

$$\beta_i^r=\mathrm{sigmoid}(\boldsymbol{W}_3^r\cdot[\boldsymbol{W}_i^r n_i^r,m_i^r,\boldsymbol{W}_i^r n_i^r-m_i^r]) \tag{9}$$

$$m_i^r=\sum_{j\in N(i)}\alpha_{i,j}\boldsymbol{W}_j^r n_j^r \tag{10}$$

其中，$\boldsymbol{W}$ 为可学习矩阵，$N(i)$为节点 n_i^r 的邻接节点的集合，$\alpha_{i,j}$为中心节点 n_i^r 和其邻接节点 n_j^r 的注意力权重，d 表示 q 的维度。在聚合过程中引入 β 参数，其权重通过中心节点以及邻接节点计算，将中心节点、邻接节点以及它们之间的差值进行拼接，然后通过全连接层进行计算并通过 sigmoid 函数进行归一化。参数 β 考虑不同节点之间影响的异质性，为节点聚合的过程提供一个权重，而不是将所有邻接节点无差别地进行聚合。

RGCN 针对多关系异质图直接将图中节点按照不同关系进行聚合得到节点表示[18]，但是直接按关系聚合的方法忽略了关系的异质性。受到计算机视觉领域中 SElayer 的启发[19]，本文采用通道注意力为跨关系节点分配不同的权重，将得到的不同关系下节点的表示进行拼接，然后通过平均池化操作进行降维并通过全连接层学习不同类型边之间的关系，为其分配权重。在分离不同关系的子图后，需聚合跨关系的节点表示同时保留异质图中关系的异质性，表示如下：

$$Se_r=\mathrm{softmax}(\boldsymbol{W}\cdot[n_{r_1},n_{r_2},\cdots,n_{r_n}]+\boldsymbol{b}) \tag{11}$$

$$n=sum(mat([n_{r_1},n_{r_2},\cdots,n_{r_n}]^{\mathrm{T}},Se_r)) \tag{12}$$

其中，Se_r 表示异质图中不同关系下节点 n 的权重，n_m 为不同关系下的节点表示，n 表示多关系图注意力网络的输出。首先将不同关系下的节点表示进行拼接；然后通过全连接层将特征维度进行降维处理，并计算不同关系所占权重；最后将不同关系下的节点表示相加得到最后的节点输出。

3.3 基于 LSTM 注意力的后连接网络

每层多关系图注意力网络可以聚合其邻接节点的信息，

随着网络层数的增加，会出现过平滑现象[20]，如图 2 所示。

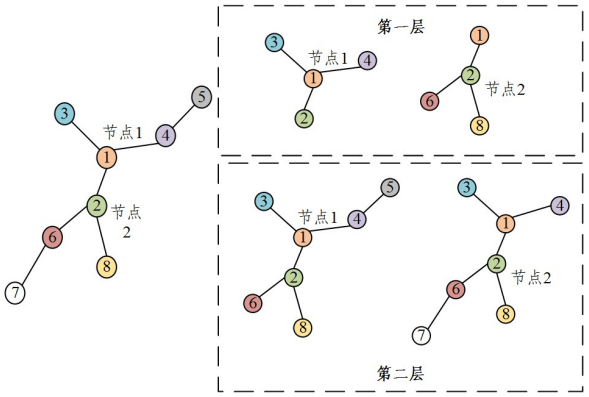


图 2 过平滑现象示意图

Fig. 2 Diagram of over-smoothing

以图中节点 1 和节点 2 为例，经过一层图卷积神经网络(GCN)后所聚合的节点如图中第一层所示，经过两层 GCN 后所聚合的节点如图中第二层所示，可以看出节点 1 和节点 2 所聚合的节点几乎相同，这就会使得两个节点拥有相似的嵌入。随着网络层数的增加，每一个节点的隐层表征会趋近于同一个值，从而使得网络训练困难。为了缓解这一现象，本文使用后连接操作，通过 LSTM[21]注意力使节点有选择性地聚合节点。

$$n'=sum(\alpha^{\mathrm{T}}\cdot[n_{l_1},n_{l_2},\cdots,n_{l_n}]) \tag{13}$$

$$\alpha=\mathrm{softmax}(\boldsymbol{W}\cdot(\overrightarrow{h},\overleftarrow{h})+\boldsymbol{b}) \tag{14}$$

$$\overrightarrow{h}_l=LSTM_{FW}(\overrightarrow{h}_{l-1},n_l);\overleftarrow{h}_l=LSTM_{BW}(\overleftarrow{h}_{l-1},n_l) \tag{15}$$

其中 $\overrightarrow{h}$ 和 $\overleftarrow{h}$ 分别表示正向传播和后向传播 LSTM 网络的输出，α 表示节点 n 在不同邻域范围的注意力权重，n' 为最终的节点表示。其结构图如图 3 所示。

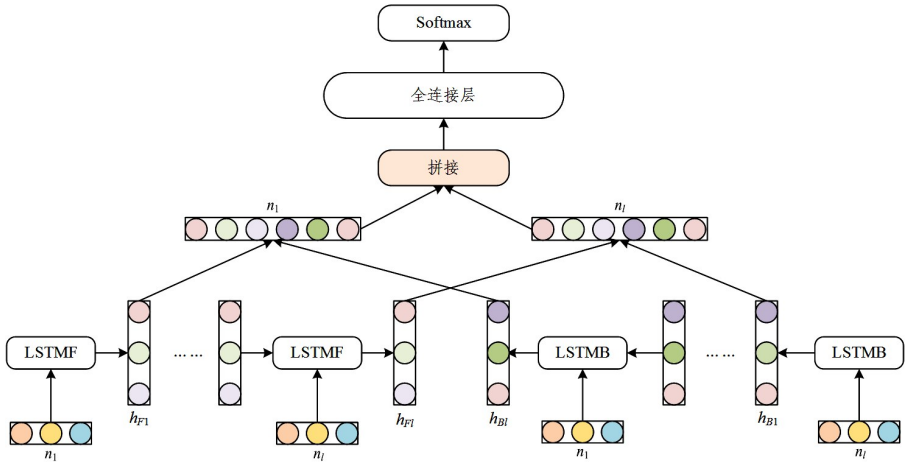


图 3 基于 LSTM 的后连接网络结构图

Fig. 3 Structure diagram of the post-connection network based on LSTM

对于后连接网络，我们将每层的节点表示 n_l 作为 Bi-LSTM 的输入，得到每层的前向输出和后向输出，将两个输出串联后进行线性 mapping 得到一个标量得分，再使用 Softmax 函数得到节点 n 在它不同范围的邻域内的注意力得分。得分 α_l 越高表明在 l 层邻域范围内的节点对 n 越重要，以此来自适应地确定节点 n 所聚合的邻域范围，最后加权平均得到节点 n 的最终表示。

3.4 算法描述

MR-GAN 模型的算法描述如算法 1 所示。

算法 1 MR-GAN 模型算法

输入：异质图 G 中的节点 $n=[n_1,n_2,\cdots,n_3]$，边 edge，以及边类型 edge_type

输出：图 G 中的节点 $[n_1,n_2,\cdots,n_3]$

1. for epoch s←1 to Epoch do
2. for l←1 to num_layer do
3. n_l=Linear(n_l)//节点特征维度变化
4. for j←1 to edge_type do
5. n_l^r=Attention(n_l^r,m_l^r)//对节点进行注意力计算并聚合邻接节点
6. end for
7. n_l=Se($n_l^{r_1}$,$n_l^{r_2}$,…,$n_l^{r_n}$)//对不同关系下的节点 n_i 进行权重计算并聚合
8. end for
9. n=Aggregate(n_{l_1},n_{l_2},…,n_{l_n})//节点进行自适应邻域聚合
10. n_l=Linear(n_l)//对节点特征进行降维处理
11. end for

4 实验与结果分析

4.1 数据集

为了验证本文方法的效果，我们采用 Cresci15，Twibot20 及 MGTAB 这 3 个数据集进行实验。

表 1 数据集统计

Table 1 Dataset statistics

数据集	节点	边	边类型
Cresci15	5301	14220	2
Twibot20	11826	15434	2
MGTAB	10199	1700108	7

Cresci15 数据集包含 5301 个用户组成的节点，这些用户被标记为真实用户或社交机器人，包含 friends 和 follow 两种类型共 14220 条边[22]。

Twibot20 数据集包含 229 580 个用户但是只有 11826 个被标记为真实用户或社交机器人[23]，所以本文选取这 11826 个用户作为节点，包含 friends 和 follow 两种类型共 1434 条边。

针对 Twibot20 和 Cresci15 数据集，我们选用用户简介、用户推文、用户属性及用户标签 4 类特征作为用户特征。

MGTAB 数据集是一个基于机器账户检测领域最大原始数据的数据集，包含 150 多万个用户和 1.3 亿条文本[24]。该数据集提供了这些用户之间 7 种类型关系的信息，但是只有 10199 个用户被专家标注为真实用户或社交机器人。

4.2 评价指标

本文旨在研究社交机器人检测的准确率，故选用 Accuracy，Precision，Recall 和 F1 分数 4 个评价指标来评价模型的效果。

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \tag{16}$$

$$Precision=\frac{TP}{TP+FP} \tag{17}$$

$$Recall=\frac{TP}{TP+FN} \tag{18}$$

$$F1=100\%\times\frac{2\times Precision\cdot Recall}{Precision+Recall+e^{-9}} \tag{19}$$

其中，TP 表示一个节点是真实用户且被判定成真实用户；FN 表示一个节点为真实用户但被判定为社交机器人；FP 表示一个节点为社交机器人但被判定为真实用户；TN 表示节点为社交机器人且被判定成社交机器人。其中 Accuracy 和 F1 分数越高表明模型的效果越好。

4.3 实验设置

实验在 GPU 服务器上进行，模型基于 Pytorch 框架，Python 版本 3.7，Pytorch 版本 1.8.1。按照 7∶1∶2 的比例将数据集划分为训练集、验证集、测试集。选择以下 11 个模型作为对比模型进行对比实验。

机器学习的方法也常用于社交机器人检测，例如 BotOrNot 社交机器人检测系统利用 RF 等机器学习的方法进行社交机器人的检测，本文采用 DT，RF 两种典型机器学习的方法作为社交机器人检测的对比模型。DT[25] 表示决策树，是一种分类规则，其使用递归自上而下的方法从一组训练数据中推断出树形结构。RF[26] 表示随机森林，它将数据集用基础分类器进行分类，然后将其继承到随机森林中来评估和检测社交机器人。

社交机器人检测的本质为节点的分类，故本文选取图神经网络中经典的 4 种节点分类模型 GCN，GAT，GraphSage 和 RGCN 作为社交机器人检测的对比模型。

GCN[27] 是图神经网络的一个基础模型，通过将切比雪夫多项式简化为一阶邻域，得到节点嵌入向量，然后通过 MLP 对社交机器人进行分类。

GAT[28] 是一个半监督图模型，它应用注意力机制来确定节点邻域的权重，通过自适应地将权重分配给不同的邻接节点，然后通过 MLP 对社交机器人进行分类检测。

GraphSage[29] 是一种归纳算法，它学习一种聚合函数，通过聚合节点邻居的特征信息来学习目标节点本身的表达，进而对社交机器人进行分类检测。

RGCN 是 GCN 在多边类型上的应用，符合社交机器人检测的场景，邻居节点的聚合是按照边的类型进行分类，根据边类型的不同进行相应的转换。

BotRGCN 通过构建一个异构图来表示 Twitter 社交网络，并采用 RGCN 进行表示学习和机器人检测。

HGT[30] 是一种异质图转换架构，用于建模 Web-scale 的异质图，利用异构注意力网络实现对社交机器人的检测。

HOFA[31] 通过面向同源性的图增强模块和频率自适应注意力模块来对抗异语言伪装挑战进而检测社交机器人。

RF-GNN[32] 将 GNN 作为基础分类器，将机器学习的方法与图的方法相结合来检测社交机器人。

SHGN[33] 以 GAT 作为主干，融合节点残差模块和边残差模块进行机器人检测。

4.4 实验结果及分析

4.4.1 Cresci15 数据集实验结果

首先在 Cresci15 数据集上进行初步验证，选择 friends 和 follow 两种边类型，训练 100 轮并与上述的 11 个模型进行对比，实验结果如表 2 所列。由表 2 可以看出本文提出的 MR-GAN 模型在 4 个评价指标上均取得了最好的实验结果，初步验证了 MR-GAN 模型在社交机器人检测任务上的有效性。

表 2 Cresci15 数据集上的实验结果

Table 2 Experimental results on Cresci15 dataset

(%)

模型	*Accuracy*	*Precision*	*Recall*	*F1*
DT	97.24	96.95	97.02	96.98
RF	97.74	97.74	98.21	97.46
GCN	97.35	96.56	97.78	97.12
GAT	97.83	97.21	98.14	97.64
RGCN	97.92	97.23	98.34	97.74
GraphSage	97.73	96.92	98.27	97.53
BotRGCN	98.30	97.74	98.63	98.15
HGT	97.44	97.43	96.99	97.20
HOFA	97.72	97.54	98.31	97.92
RF-GCN	96.52	96.21	96.50	96.35
SHGN	98.40	98.23	98.28	98.26
MR-GAN	**98.87**	**98.90**	**98.71**	**98.80**

4.4.2 Twibot20 数据集实验结果

在 Twibot20 数据集上进行实验，实验训练轮数选择 50 轮并与上述 11 个模型进行对比，得到的实验结果如表 3 所列。

表 3 Twibot20 数据集上的实验结果

Table 3 Experimental results on Twibot20 dataset

(%)

模型	*Accuracy*	*Precision*	*Recall*	*F1*
DT	79.49	79.22	79.37	79.27
RF	81.32	81.05	81.15	81.08
GCN	70.92	70.42	70.52	70.45
GAT	71.28	70.19	71.16	70.31
RGCN	82.99	82.42	83.01	82.62
GraphSage	80.32	79.55	80.40	79.81
BotRGCN	83.13	83.03	82.59	82.74
HGT	77.73	77.93	76.57	76.88
HOFA	83.23	82.97	83.55	83.25
RF-GNN	82.72	82.12	82.63	82.36
SHGN	84.36	83.97	84.19	84.07
MR-GAN	**85.55**	**86.01**	**84.71**	**85.11**

由表 3 可以看出本文所提出的 MR-GAN 模型在 4 个评价指标上均取得了最优的实验结果。对比 MR-GAN 的每项评价指标以及 11 种对比模型中该项评价指标的最优值(见图 4)，MR-GAN 在 50 轮时已经取得了相较于其他模型较好

的效果，随着训练轮数的增加，模型逐渐收敛，但是 MR-GAN 仍能取得最好的效果。在训练 50 轮的情况下，Accuracy 提升了 1.19%，Precision 提升了 2.04%，Recall 提升了 0.52%，F1 分数提升了 1.04%。本文所提模型在聚合不同关系的节点时考虑了不同关系下节点的异质性，同时使用后连接使得节点可以自适应地选择邻域。综上所述，MR-GAN 模型在 Twibot20 数据集上的性能获得了较大的提升，同时也验证了其在检测社交机器人时具有一定优势。

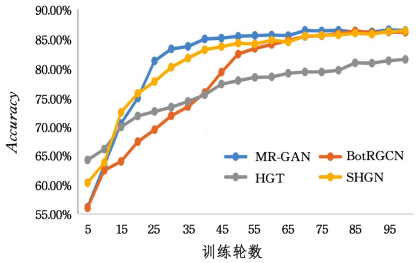


图 4 Accuracy 随着训练轮数增加的变化趋势图

Fig. 4 Variation trend of accuracy with the increase of training epochs

4.4.3 MGTAB 数据集实验结果

在 MGTAB 数据集上进行实验，训练 100 轮并与上述 11 个模型进行对比，实验结果如表 4 所列。由表 4 可以看出本文所提出的 MR-GAN 模型在 Accuracy，Precision 和 F1 分数这 3 个指标上取得了最优的实验结果。对比 11 种模型中该项评价指标的最优值可以看出，*Accuracy* 提升了 0.38%，*Precision* 提升了 1.07%，*F1* 分数提高了 0.13%。本文模型的召回率比 SHGN 低，是因为随着网络层数的增加，图神经网络的特性使得节点出现了拉普拉斯平滑现象，即周围节点特征和中心节点的隐层表示相似，使得节点特征同质化，导致召回率有所下降。MR-GAN 相比 HGT，BotRGCN 和 SHGN 表现更好是因为 MR-GAN 在进行消息传递时针对不同关系子图中的邻接节点进行节点聚合，这使得模型可以学习到不同关系下节点的特征；之后通过不同关系的注意力权重进行节点的表示学习，这使得不同类型节点的特征更加明显，进而提高了社交机器人检测的精度。

表 4 MGTAB 数据集实验结果

Table 4 Experimental results on MGTAB dataset

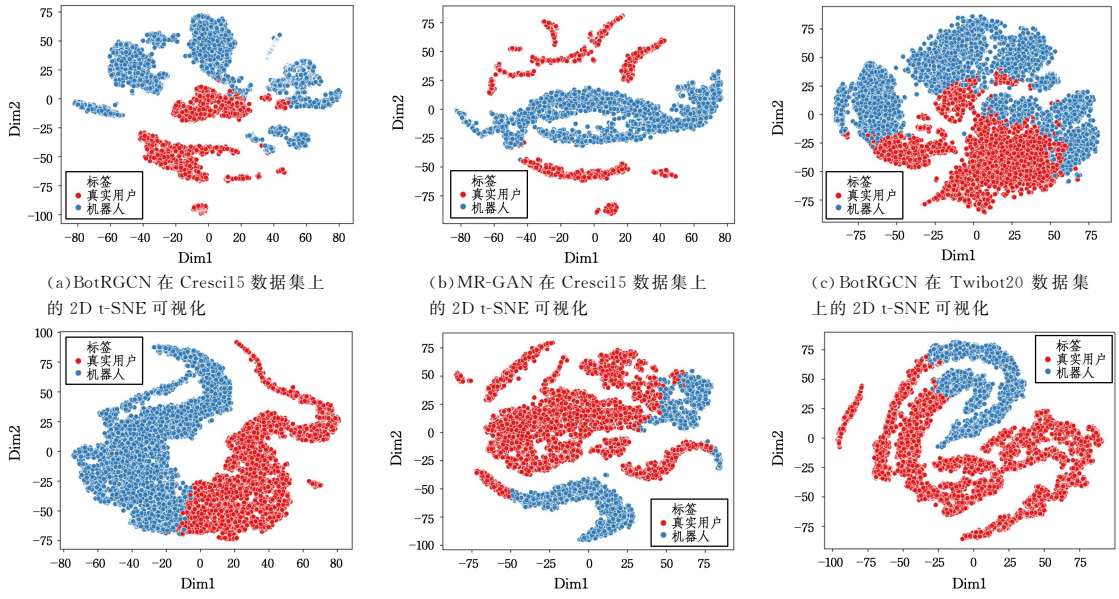
(%)

模型	*Accuracy*	*Precision*	*Recall*	*F1*
DT	87.29	83.81	83.59	83.64
RF	88.17	88.46	82.52	84.16
GCN	85.81	77.29	84.83	79.84
GAT	86.51	79.77	84.41	81.56
RGCN	87.83	83.21	85.08	84.05
GraphSage	88.33	84.12	85.60	84.78
BotRGCN	88.40	83.45	86.18	84.63
HGT	89.25	85.98	87.03	86.46
HOFA	88.68	79.21	80.47	79.81
RF-GCN	87.19	83.37	83.42	83.39
SHGN	89.26	84.67	**87.20**	85.81
MR-GAN	**89.64**	**87.05**	**86.20**	**86.59**

综合上述 3 个实验结果，可以看出本文提出的模型在社交机器人检测任务上的有效性。

4.4.4 t-SNE 可视化

为了更直观地观察实验结果，本文对测试数据进行了 t-SNE 可视化，实验结果如图 5 所示。t-SNE 的主要目的是将多维数据转换为低维数据，并将其映射到 2 维平面进行可视化。在 t-SNE 可视化图中，两类数据距离相距较大、边界明显，则表明分类的效果越好。在图 5 中，左边 3 个图为基线模型中表现优异的模型在 3 个数据集上的 t-SNE 可视化，右边 3 个图为 MR-GAN 在 3 个数据集中的 t-SNE 可视化图。可以看出，右图的分类边界更加明显，由此表明 MR-GAN 在社交机器人检测任务中有更好的效果。



(a) BotRGCN 在 Cresci15 数据集上的 2D t-SNE 可视化

(b) MR-GAN 在 Cresci15 数据集上的 2D t-SNE 可视化

(c) BotRGCN 在 Twibot20 数据集上的 2D t-SNE 可视化

(d) MR-GAN 在 Twibot20 数据集上的 2D t-SNE 可视化

(e) HGT 在 MGTAB 数据集上的 2D t-SNE 可视化

(f) MR-GAN 在 MGTAB 数据集上的 2D t-SNE 可视化

图 5 t-SNE 可视化

Fig. 5 t-SNE visualization

4.4.5 消融实验

为了进一步验证本文模型 MR-GAN 的有效性，进行了两个消融实验。

1）节点聚合方式消融实验：为了验证本文所用图 Transformer 注意力聚合不同关系子图节点的有效性，分别将聚合方式替换为 GCN，GAT 和 GraphSage 进行消融实验，实验结果如表 5 和图 6 所示。

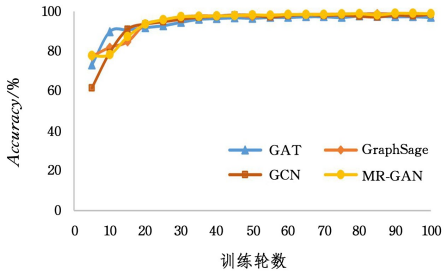
由表 5 可以看出，相比其他聚合方式，本文采用的图 Transformer 注意力在 Cresci15 数据集上比其他聚合方式的最优值仅提升了 0.19%，提升幅度较小可能是由于该数据集本身比较小，模型在该数据集上的检测精度较高。本文方法在 Twibot20 数据集上比其他聚合方式的最优值提升了 1.27%。图 6 可视化了在两个数据集上的消融实验，其中，纵坐标表示模型在数据集上的准确率，根据上述实验可以验证图 Transformer 注意力的有效性。

表 5 聚合方式消融实验

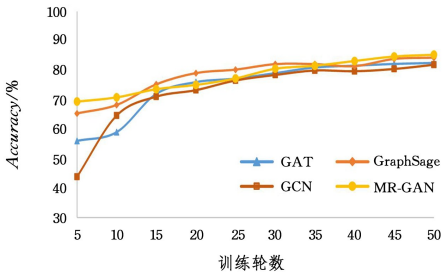
Table 5 Ablation experiments by polymerization

(%)

数据集	聚合方式	*Accuracy*	*F*1
Cresci15	GCN	98.12	98.02
	GAT	98.00	97.83
	GraphSage	98.68	98.61
	MR-GAN	**98.87**	**98.80**
Twibot20	GCN	81.99	81.75
	GAT	82.43	82.16
	GraphSage	84.28	83.93
	MR-GAN	**85.55**	**85.11**



(a) Cresci15 数据集消融实验

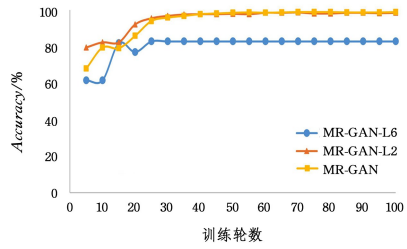


(b) Twibot20 数据集消融实验

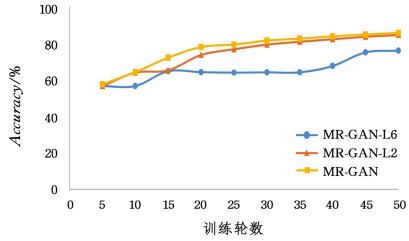
图 6 聚合方式消融实验

Fig. 6 Ablation experiments by polymerization

2）基于 LSTM 注意力的后连接网络的消融实验：为了验证后连接网络的有效性，文本对不使用后连接的 6 层网络模型 MR-GAN-L6 和不使用后连接的 2 层网络模型 MR-GAN-L2 与使用后连接的 6 层 MR-GAN 进行对比实验，实验结果如表 6 和图 7 所示。



(a) Cresci15 数据集消融实验



(b) Twibot20 数据集消融实验

图 7 后连接网络消融实验

Fig. 7 Ablation experiments by post-connection

表 6 后连接消融实验

Table 6 Ablation experiments by post-connection

(%)

数据集	聚合方式	*Accuracy*	*F*1
Cresci15	MR-GAN-L2	98.31	98.21
	MR-GAN-L6	82.86	79.54
	MR-GAN	**98.87**	**98.80**
Twibot20	MR-GAN-L2	84.02	83.83
	MR-GAN-L6	76.92	75.65
	MR-GAN	**85.55**	**85.11**

由图 7 可以看出，随着网络层数的增加会出现过平滑现象。本实验采用 2 层网络的模型是因为大部分图神经网络为了避免过平滑通常选用 2 层图神经网络。由表 6 可以看出，本文提出的 MR-GAN 在 3 个对比模型中取得了最好效果，说明了基于 LSTM 注意力的后连接方式的有效性。

结束语 对于多关系的异质信息网络，本文提出 MR-GAN 模型进行社交机器人检测。该模型完善了用户节点的特征，并在聚合节点时采用注意力机制，考虑了节点之间的差异性；同时，将计算机视觉中的 SElayer 应用于社交机器人的检测，在关系子图聚合时利用通道注意力，结合关系之间的异质性对子图进行聚合；此外，为了缓解过平滑问题，采用后连接机制让节点可以自适应地选择邻域。在 3 个数据集上的实验结果表明，*Accuracy* 分别提升了 0.47%，1.19%和 0.38%，验证了 MR-GAN 模型的有效性。

本文模型虽然在一些数据集上取得了不错的效果，但也存在不足。根据表 4 的实验结果可以看出，MR-GAN 虽然可以缓解过平滑现象但是无法从根本上解决这一问题，这表明层间聚合的方式还有待改进。此外，MR-GAN 在提取节点特征时对语义特征进行压缩使得语义模糊，没有充分利用文本模态的信息。后续工作将从 3 个方面进行：首先继续探索关系子图的聚合方式，例如在聚合时融入用户推文中的语义信息，以进一步提升模型性能，并在更复杂的异质信息网络上进行研究使得模型具有更好的泛化效果[34]；其次将会关注模型

的优化，如多个模块并行以降低模型复杂度；最后针对社交机器人检测任务，将探索如何将多模态信息进行融合，例如文本数据与图数据的有效交互，以及如何有效利用多模态信息进行社交机器人检测。

参 考 文 献

[1] MARIA K，ILIAS D，ATHENA V. Bot-Detective：An explainable Twitter bot detection service with crowdsourcing functionalities[C]// 12th International Conference on Management of Digital Ecosystems. New York：Association for Computing Machinery，2020：55-63.

[2] WU Y H，FANG Y Z，SHANG S K，et al. A novel framework for detecting social bots with deep neural networks and active learning[J]. European Journal of Medicinal Chemistry：Chimie Therapeutique，2021，211(1)：1-16.

[3] ABREU J，GONDIM J，RALHA C. Twitter Bot Detection with Reduced Feature Set[C]// 2020 IEEE International Conference on Intelligence and Security Informatics(ISI). USA：Arlington，2020：1-6.

[4] HU F X，LUO W H. Social robot account detection based on multi-dimensional dynamic feature verification[J]. Journal of Foshan University，2023，41(1)：23-34.

[5] SNEHA K，EMILIO F. Deep Neural Networks for Bot Detection[J]. Information Sciences，2018，467(10)：312-322.

[6] GUO Q，XIE H，LI Y，et al. Social Bots Detection via Fusing BERT and Graph Convolutional Networks[J]. Symmetry，2022，14(1)：1-30.

[7] HAYAWI K，MATHEW S，VENUGOPAL N，et al. DeeProBot：a hybrid deep neural network model for social bot detection based on user profile data[J]. Social Network Analysis and Mining，2022，12(1)：1-19.

[8] WU J，YE X，MAN Y. BotTriNet：A Unified and Efficient Embedding for Social Bots Detection via Metric Learning[C]// 2023 11th International Symposium on Digital Forensics and Security (ISDFS). Turkey：Istanbul，2023：1-6.

[9] WANG X，JI H Y，SHI C，et al. Heterogeneous Graph Attention Network[C]// The World Wide Web Conference(WWW'19). New York：Association for Computing Machinery，2019：2022-2032.

[10] LI Y，JI Y，LI S，et al. Relevance-aware anomalous users' detection in social network via graph neural network[C]// 2021 International Joint Conference on Neural Networks(IJCNN). New York：IEEE，2021：1-8.

[11] FENG S B，WAN H R，WANG N N. SATAR：A Self-supervised Approach to Twitter Account Representation Learning and its Application in Bot Detection[C]// Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York：Association for Computing Machinery，2021：3808-3817.

[12] FENG S B，WAN H R，WANG N N. BotRGCN：Twitter bot detection with relational graph convolutional networks[C]// Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York：Association for Computing Machinery，2021：236-239.

[13] YANG Y G，YANG R Y，LI Y Y，et al. RoSGAS：Adaptive Social Bot Detection with Reinforced Self-supervised GNN Architecture Search[J]. ACM Transactions on the Web，2023，17(3)：1-31.

[14] XU K Y，ZHOU A M，CHEN A L，et al. Social bot detection based on active learning and relational graph convolutional neural networks[J]. Journal of Sichuan University(Natural Science Edition)，2023，60(5)：121-129.

[15] CAI Z J，TAN Z X，LEI Z Y，et al. LMBot：Distilling Graph Knowledge into Language Model for Graph-less Deployment in Twitter Bot Detection[J]. arXiv：2306.17408，2023.

[16] SIRUSSTARA J，ALEXANDER N，ALFARISY A，et al. Clickbait Headline Detection in Indonesian News Sites using Robustly Optimized BERT Pre-training Approach(RoBERTa)[C]// 2022 3rd International Conference on Artificial Intelligence and Data Sciences(AiDAS). IPOH：Malaysia，2022：1-6.

[17] VASWANI A，SHAZEER N，PARMAR N，et al. Attention Is All You Need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems(NIPS'17). San Diego：NIPS，2017：6000-6010.

[18] SCHLICHTKRULL M，KIPF T，BLOEM P，et al. Modeling relational data with graph convolutional networks[C]// European Semantic Web Conference. European：Springer，2018：593-607.

[19] HU J，SHEN L，ALBANIE S，et al. Squeeze-and-Excitation Networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2017：7132-7141.

[20] LI Q M，HAN Z C，WU X M. Deeper Insights Into Graph Convolutional Networks for Semi-Supervised Learning[C]// Thirty-Second AAAI Conference on Artificial Intelligence. New York：AAAI Press，2018：3538-3545.

[21] JIANG Y，ZHAO T，CHAI Y，et al. Bidirectional LSTM-CRF models for keyword extraction in Chinese sport news[J]. MIPPR 2019：Pattern Recognition and Computer Vision，2020，2(1)：11-17.

[22] STEFANO C，ROBERTO D，MARINELLA P，et al. Fame for sale：Efficient detection of fake Twitter followers[J]. Decision Support Systems，2015，80(9)：56-71.

[23] FENG S B，WAN H，WANG N，et al. TwiBot-20：A Comprehensive Twitter Bot Detection Bench-mark[C]// Proceedings of the 30th ACM International Conference on Information & Knowledge Management，New York：ACM，2020：4485-4494.

[24] SHI S H，QIAO K，CHEN J. MGTAB：A Multi-Relational Graph-Based Twitter Account Detection Benchmark[C]// IEEE Conference on Computer Vision and Pattern Recognition. New York：IEEE，2023：1-14.

[25] WEI Y L. Classification and regression trees[J]. Wiley Interdisciplinary Reviews：Data Mining and Knowledge Discovery，2011，1(1)：3-7.

[26] BREIMAN L. Random forests[J]. Machine Learning，2004，45：

5-32.

[27] SHA A, WANG B, WU X, et al. Semi-Supervised Classification for Hyperspectral Images Using Edge-Conditioned Graph Convolutional Networks[C]// IEEE International Geoscience and Remote Sensing Symposium(IGARSS 2019). Japan: Yokohama, 2019: 2690-2693.

[28] PETAR V, GUILLEM C, ARANTXA C, et al. Graph attention networks[C]// International Conference on Learning Representations(ICLR). Washington: ICLR, 2018.

[29] WILL H, YING Z T, JURE L. Inductive representation learning on large graphs[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems(NIPS'17). USA: Curran Associates Inc. 2017: 1025-1035.

[30] HU Z H, DONG Y X, WANG K S, et al. Heterogeneous graph transformer[C]// Proceedings of The Web Conference 2020 (WWW'20). USA: Association for Computing Machinery, 2020: 2704-2710.

[31] YE S, TAN Z, LEI Z, et al. Hofa: Twitter bot detection with homophily-oriented augmentation and frequency adaptive attention [J]. arXiv: 2306. 12870, 2023.

[32] SHI S H, QIAO K, YANG J, et al. RF-GNN: Random Forest boosted graph neural network for social bot detection[J]. arXiv: 2304. 08239, 2023.

[33] LV Q S, DING M, LIU Q, et al. Are we really making much progress? Revisiting, benchmarking and refining heterogeneous graph neural networks[C]// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM, 2021: 1150-1160.

[34] LU H Y, LIU F, WANG Y B. Social Bot Detection for Dynamic Social Networks Based on Link Prediction[J]. Journal of Information Engineering University, 2024, 25(3): 285-291.

MENG Lingjun, born in 1999, postgraduate. His main research interests include data analysis, natural language processing and computer vision.

WANG Gengrun, born in 1987, Ph.D, assistant researcher. His main research interests include telecommunication network security and data processing.

(责任编辑:何杨)