doi:10.6043/j.issn.0438-0479.202405028

基于多门控混合专家网络的社交机器人检测

臧威龙[1,2],余正涛[1,2],高盛祥[1,2*],谭凯文[1,2],张勇丙[1,2]
(1.昆明理工大学信息工程与自动化学院,云南 昆明 650500;2.昆明理工大学云南省人工智能重点实验室,云南 昆明 650500)

摘要:[**目的**]针对社交机器人模仿真实用户特征伪装自身及不同社区用户行为差异显著使得检测难度提升的问题进行研究.[**方法**]提出基于多门控混合专家网络的社交机器人检测模型(multi-gated mixture of experts network bot detection,MGEBot).该方法首先将用户元数据与推文数据编码为序列信息,并对关系数据进行图结构编码,实现多角度用户信息表征.随后,将信息输入多门控混合专家网络,学习不同社区用户群体的独有特征,以应对社区差异性问题.最终,融合 3 种模态的表征进行检测.[**结果**]在 Cresci-15、TwiBot-20 和 TwiBot-22 3 个主流数据集上,MGEBot 在 F_1 等指标上均超越现有基准模型.在泛化性与鲁棒性实验中,MGEBot 表现出更好的稳定性与适应性.分析实验表明门控数量增加可显著提升性能,但存在饱和点;专家数量并非越多越好,需寻求最优配置.[**结论**]MGEBot 能有效应对社区差异性挑战,其多源信息融合与多门控混合专家网络机制提升了检测精度和泛化能力,适用于多样化真实场景的社交机器人检测任务.

关键词:社交机器人检测;社区群体差异性;多门控专家混合网络
中图分类号:TP 183 **文献标志码**:A **文章编号**:0438-0479(2025)04-0629-13

Social bot detection based on multi-gated mixture of experts network

ZANG Weilong[1,2],YU Zhengtao[1,2],GAO Shengxiang[1,2*],
TAN Kaiwen[1,2],ZHANG Yongbing[1,2]
(1. Faculty of Information Engineering and Automation,Kunming University of Science and Technology,Kunming 650500,China;
2. Yunnan Key Laboratory of Artificial Intelligence,Kunming University of Science and Technology,Kunming 650500,China)

Abstract:[**Objective**] This study addresses two major challenges in social bot detection:(1) the increasing ability of social bots to mimic real user characteristics,thus progressively enhancing the difficulty of identification;and (2) significant behavioral disparities among users in different social communities. This difference limits the generalization capability of traditional detection methods. To tackle these two issues,we propose a more adaptable and robust detection model capable of effectively identifying social bots across diverse community environments.[**Methods**] We present MGEBot (multi-gated mixture of experts network bot detection),a novel social bot detection framework that integrates multimodal user information from social media platforms,including user metadata,tweet content,and relational graph data. Specifically,MGEBot first leverages a pre-trained language model to encode user metadata and tweet texts,extracting temporal and semantic features. Simultaneously,it employs graph structural encoding techniques to model relationship data,and preserves user interactions and community structures. These multi-perspective representations are then fed into a multi-gated mixture of experts network,in which each expert specializes in capturing behavioral patterns of distinct user

收稿日期:2024-05-30 **录用日期**:2024-10-01
基金项目:国家自然科学基金(U23A20388,U21B2027,62376111,62266028,62266027);云南省重点研发计划(202303AP140008,202401BC070021,202103AA080015);云南省科技人才与平台计划(202105AC160018);云南省基础研究项目(202301AT070393)
*通信作者:gaoshengxiang.yn@foxmail.com
引文格式:臧威龙,余正涛,高盛祥,等.基于多门控混合专家网络的社交机器人检测[J].厦门大学学报(自然科学版),2025,64(4):629-641.
Citation:ZANG W L,YU Z T,GAO S X,et al. Social bot detection based on multi-gated mixture of experts network[J]. J Xiamen Univ Nat Sci,2025,64(4):629-641. (in Chinese)

communities. Multiple gating mechanisms independently select the most relevant expert outputs for each modality, allowing the model to dynamically adapt to diverse input characteristics. Finally, the fused multimodal features are jointly utilized in a classification layer to achieve comprehensive and accurate social bot detection. [**Results**] MGEBot was thoroughly evaluated on three benchmark datasets: Cresci-15, TwiBot-20, and TwiBot-22. Experimental results demonstrate that MGEBot consistently outperforms all baseline models—including BotRGCN, RGT, and BotMoE—across all datasets. On Cresci-15, MGEBot achieved an accuracy of 98.63% and an F_1 score of 98.91%; on TwiBot-20, 87.10% accuracy and 89.10% F_1 score; and on the most challenging TwiBot-22 dataset, 79.66% accuracy and 57.13% F_1 score. Notably, on TwiBot-22, MGEBot's F_1 score surpassed the second-best model, BotMoE, by 0.74%, highlighting its superior detection capability in complex scenarios. The overall trend across all models reveals that detection difficulty significantly increases with newer datasets: average F_1 scores dropped from 93.49% on Cresci-15 to 82.18% on TwiBot-20, and further down to 47.90% on TwiBot-22. Despite this fact, MGEBot maintained robust performance, underscoring its stability in increasingly adversarial environments. Comparative analysis indicates that graph-based models (e.g., RGT, BotRGCN) generally outperform non-graph-based counterparts (e.g., SGBot, Wei et al.), and models incorporating mixture of experts (e.g., BotMoE, MGEBot) deliver the best overall performance, validating the efficacy of integrating graph learning with expert specialization. Regarding the training efficiency, MGEBot converged faster on TwiBot-20, and achieved an accuracy of 88.52% after 200 epochs—1.45% higher than BotMoE—suggesting that the multi-gating mechanism accelerates adaptation to heterogeneous features. In label scarcity robustness tests, when the label ratio on TwiBot-20 was reduced from 100% to 10%, MGEBot's F_1 score declined by only 1.43%, substantially less than the baseline models. Moreover, MGEBot outperformed most baseline models even when it was trained with full labels. This merit demonstrates its strong generalization under limited supervision. In cross-community generalization experiments—training on TwiBot-20 and testing on TwiBot-22 to simulate emerging bot types—MGEBot's accuracy and F_1 score decreased by 7.44% and 33.21%, respectively. However, it still significantly outperformed all baselines, highlighting its effectiveness in adversarial and evolving scenarios. Parameter sensitivity analyses reveal that increasing the number of gating units yields significant performance gains (e.g., increasing gates from one to two raised average F_1 by 1.56%), whereas simply increasing the number of experts does not always improve results, emphasizing the need for balanced model complexity and effectiveness. [**Conclusions**] Incorporating a multi-source information fusion mechanism and a multi-gated mixture of expert's architecture, MGEBot effectively addresses those detection challenges posed by inter-community behavioral differences. Its design enables the extraction of community-specific behavior features, and enhances the detection accuracy, robustness, and generalizability in complex and diverse social environments. Therefore, MGEBot is well-suited for real-world social bot detection tasks across various application scenarios.

Keywords: social bot detection; community group differences; multi-gated mixture of experts network

在 2024 年第一季度,微博的月活跃用户达到 5.98 亿,Facebook 的用户几乎占世界人口的三分之一,Twitter 的用户数量已增长到 13 亿. 在庞大的社交媒体用户群中,存在大量的社交机器人,这些机器人可被用于散布谣言、诈骗和煽动舆论[1-3]. 准确检测和识别社交机器人不仅有助于维护社交媒体平台的健康生态,还能防止虚假信息的传播,促进公共舆论的真实和可信度[4],这对于维护网络安全和社会稳定至关重要[5-6].

早期的社交机器人仅具备非常简单的属性,几乎没有个人信息和社交关系,因此非常容易区分[7-8],例如:Kudugunta 等[9] 仅使用 AdaBoost 分类器与 SMOTE 过采样技术相结合,就可获得 99.81%的检测精度. 为了增加机器人的真实度,社交机器人的操作员开始丰富机器人的账户信息,并通过多个机器人的互相关注形成自己的关系网络,从而增加了检测的难度[10-11],为此,Feng 等[12] 提出了 BotRGCN,该方法利用账号间的关注关系构建了异构图,并应用关系图卷积网络进行检测,面对这类机器人检测准确度达到 87%. 而随着社交机器人的更新迭代,现有的机器人会从多个角度去伪装人类,他们使用真实用户的名字和头像进行账户伪装,并窃取真实用户的推文信息,以夹杂发布自己的恶意信息[13-15]. 同时,每个社交平台上存在种类繁多的社区,每个社区之间相互独立或存在一定联系. 每个社交用户可能身处一个或多个社区,而不同用户群体之间的社交机器人的行为和特征存在显著差异,这进一步增加了检测的难度. 为此,Liu 等[16] 提出了 BotMoE,该方法利用专家网络分别基于元数据、关系数据和推文数据学习社区的差异表征,然后将 3 类表征融合后再进行社交机器人检测,面对这类机器人准确度达到 56%. 然而,这类方法只通过一次模态特异专家融合来建模不同的组合社群,使得模型不能够适应当前社区的多样性以及复杂性,从而导致模型的鲁棒性和泛化性较差.

针对上述挑战,本文提出了基于多门控混合专家的

社交机器人检测方法(multi-gated mixture of experts network bot detection,MGEBot). 首先,分别使用多层感知机(multilayer perceptron,MLP)、图卷积神经网络和预训练语言模型对社交账户的元数据、关系图数据、推文数据进行编码. 然后,利用多门控混合专家层对上述编码结果进行进一步表征,以缓解社区群体差异性问题. 最后,利用特征融合层进行 3 种表征的融合,并使用多层感知机进行社交机器人检测. 为了验证模型性能,本文在 Cresci-15、TwiBot-20 和 TwiBot-22 数据集上进行了实验,相较于基准模型,MGEBot 的性能有所提升,并具备更好的泛化性和鲁棒性. 本文的贡献总结如下:

1) 提出了一种新的社交机器人检测框架 MGEBot. MGEBot 能够融合用户推文数据、关系数据以及元数据 3 种特征,从多个角度对社交账户进行检测.

2) 将多门控混合专家机制与社交机器人检测任务进行结合,首先用多个专家模拟社交媒体中不同的社区,然后使用多个门控组合不同的专家输出,使得模型获取到来自不同社区群体的独特特征,提高了模型在面对社区群体差异性问题时的表现.

3) 在 3 个广泛使用的数据集 Cresci-15[17]、TwiBot-20[18]和 TwiBot-22[19]上进行了全面的实验来评估 MGEBot 模型以及另外 7 种社交机器人检测模型的准确性、鲁棒性以及泛化能力.

1　相关工作

1.1　社交机器人检测

现有的机器人检测方法大体可以分为基于单一特征检测的方法(图 1(a))与基于混合特征检测的方法(图 1(b)),而按照输入不同基于单一特征检测的方法又可细分为:基于账户元数据的检测、基于账户推文数据的检测、基于关系图的检测.

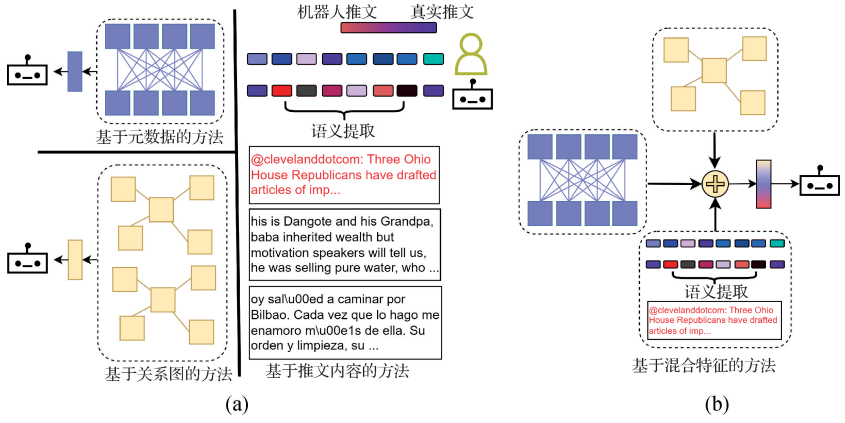

图 1　基于单一特征检测(a)和基于混合特征检测(b)

Fig. 1　Detection based on single feature (a) and mixed feature (b)

基于账户元数据的检测. 早期的许多工作通过分析社交账户的元数据(如个人信息、位置信息、主页和创建日期等)来进行社交机器人检测. 例如:Kudugunta 等[9]选取了 10 个账户元数据作为特征集,检测过程中使用 AdaBoost 分类器与 SMOTE 过采样技术相结合,检测精度达到 99.81%. Yang 等[20]提出的框架使用最小账户元数据通过数据选择来提高社交机器人分类的效率和可扩展性,该方法仅用少量的账户元数据即可实现高效分析并可实时处理 Twitter 的完整公共推文流. Yang 等[21]构造了一千多个基于元数据的特征,基于这些特征实现了快速社交机器人检测,并提供了可大批量检测推特账号是否为社交机器人的检测接口,是社交机器人检测领域中使用频率最高的研究工作之一. 然而有研究指出[8],随着机器人技术的发展,恶意用户可以通过操纵元数据和使用窃取的推文来创建欺骗性账户,以逃避基于账户元数据的检测.

基于账户推文数据的检测. 这类方法常采用 LSTM(long short-term memory network)、注意力机制、词嵌入和预训练语言模型来检测社交机器人的推文内容. 例如:Feng 等[22]使用词向量和双向 LSTM 处理用户推文数据进行社交机器人检测,Dukic 等[23]提出了一种基于 BERT(bidirectional encoder representations from transformers)的社交机器人检测模型,通过分析账户推文数据来鉴别是否为社交机器人. Cai 等[24]首次尝试将 CNN(convolution neural network)和 LSTM 结合,并结合推文间的语义信息和潜在的时态模式进行社交机器人检测. Kudugunta 等[25]使用了 LSTM 捕获社交机器人推文中历史和行为的潜在时

间模式,将推文元数据与推文文本数据的特征结合进行社交机器人检测. 然而,当高级机器人窃取真实用户的推文进行推文发布时,基于文本的方法就变得不那么有效. 这表明,虽然基于推文的数据检测方法在应对传统机器人时表现出色,但面对高级机器人的伪装时仍有局限.

基于关系图的检测. 这类方法常利用社交关系网络图(如关注-追随关系图)进行社交机器人检测. 例如:Feng 等[12]提出的 BotRGCN 利用账号间的关注关系构建了异构图,并应用关系图卷积网络进行社交机器人检测. 该方法利用多模式用户语义和属性信息,避免了繁琐的特征工程,增强了对多样化伪装的社交机器人的检测能力. Shi 等[26]提出了一个基于多关系图的社交机器人检测方法,构建了一个包含 7 种关系类型的数据集. 然而,现有社交机器人会操纵用户关系网络创建僵尸网络,仅利用关系图信息的方法逐渐显得不足.

基于混合特征的检测. 这类方法的主要特点是将用户的 3 种数据特征(即元数据特征、推文数据特征、关系数据特征)融合之后再进行检测,例如:Liu 等[16]提出的 BotMoE 先将元数据、推文数据、关系数据这 3 种特征分别通过混合专家学习用户不同社区的表现,然后将 3 种特征进行融合,最后基于融合表征进行社交机器人检测,该方法能够处理社区差异性问题. Lei 等[27]提出了一个具有文本-图交互和语义一致性的社交机器人检测框架,该框架在融合 3 种特征之前采用了文本-图交互模块进行账户推文特征与关系特征间的信息交换. 该方法整合了多种数据源的信息,能够更全面地分析和检测社交机器人,表现出了更高的检测精度和鲁棒性,但面对复杂多样的社区时仍表现不足.

1.2 多门控混合专家

混合专家网络(mixture of experts)常用于来增加模型的表征能力[28]、推理能力[29]和泛化能力[30],已在社交机器人检测等多个领域得到了应用. 例如:Lynnette[31]提出了一个基于混合专家网络的多平台机器人检测框架 BotBuster,其中每个专家被分配处理特定类型的用户特征,以应对不完整的数据. 然而,BotBuster 没有考虑到社交机器人在不同社区中的表现差异. 为了解决这些问题,Liu 等[10]提出了 BotMoE 模型,为不同的子社区训练不同的专家模型,实现了一个社区感知的专家组合. 然而,BotMoE 中的专家模块只选择了单一的门控网络,这导致模型在学习不同社交群体差异时,只有一个门控网络控制专家的输入输出权重组合,无法更好地适应社区的多样性. 而多门控混合专家网络(multi-gate mixture-of-experts)[30]则是混合专家网络的基础上为每个任务都单独建立一个门控网络. 目前,多门控混合专家网络主要应用在推荐任务上[30],能够让模型更容易捕捉到子任务间的相关性和差异性,但多门控混合专家网络在此前并没有用到社交机器人检测领域上.

综上可见,基于账户元数据的方法在初期阶段表现良好,但面对高级机器人时性能下降;基于推文数据的方法能够捕捉文本信息中的细微差别,但在处理社交机器人窃取的真实用户推文时仍有局限;基于关系图的方法能够充分利用社交网络结构信息,但面对社交机器人创造的僵尸网络是表现明显不足;基于混合特征的方法通过整合多种数据源的信息,实现了更全面的检测,表现出更高的检测精度和鲁棒性. 然而随着社交媒体的逐渐发展,每个社交平台拥有的社区数量种类繁多,每个社区之间相互独立或存在一定联系,并且每个社交用户可能身处一个或多个社区,处于不同用户群体之间的社交机器人的行为和特征存在较大差异,这使得检测难度大大提升. 针对以上问题,本文提出 MGEBot,该方法是一种基于混合特征的方法,同时本方法中还引入多门控混合专家网络,可以利用每个专家模仿一种社区,并利用多个门控学习社区之间的相关性,进而对每个用户进行“个性化”特征表示,以应对当下复杂多样的社区情况.

2 基于多门控混合专家的社交机器人检测

MGEBot 的整体框架如图 2 所示,共包括 3 个核心模块:特征编码层、多门控混合专家层和特征融合层. 在特征编码层,分别对社交账户的 3 种信息进行编码:使用 MLP 对账户元数据进行编码,利用预训练语言模型 RoBERTa[32]对推文数据进行编码,采用图神经网络(graph neural networks,GNNs)对账户关系图数据进行编码. 然后,账户的 3 种编码信息被分别送入对应的多门控混合专家层,以学习社区的差异性并进行融合. 在获取多门控混合专家层的 3 种账户特征输出后,由特征融合层进行融合,并通过 MLP 输出账户的预测结果.

2.1 符号定义

本文所用数学符号定义如表 1 所示.

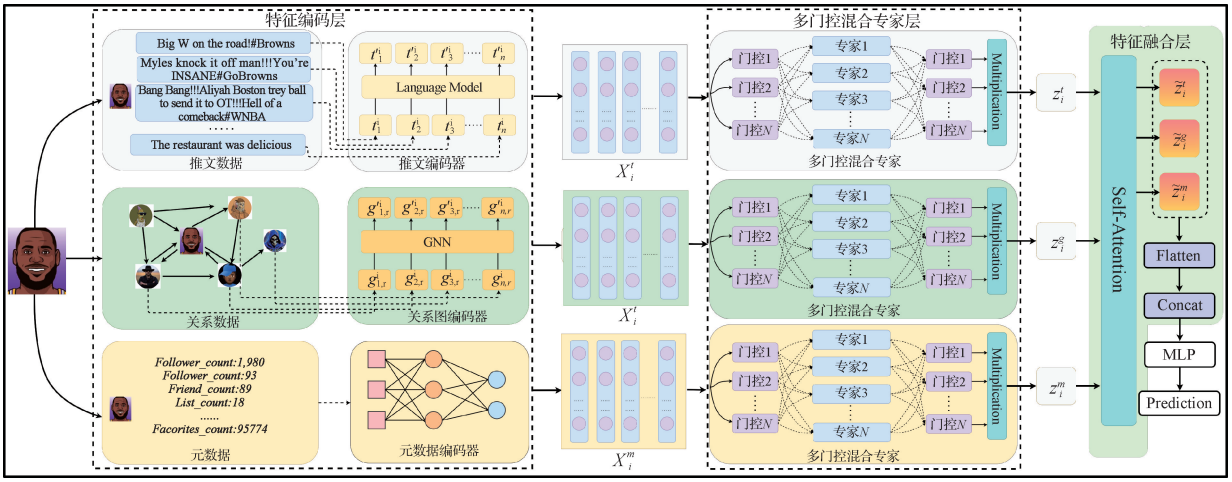

图 2　MGEBot 模型架构图

Fig. 2　MGEBot model architecture diagram

表 1　符号说明

Tab. 1　Symbol explanation

符号	定义
LM	表示预训练语言模型编码器
$t'^{\,i}_n$	$t_n{}^i$ 表示用户 i 的第 n 条推文
$g_i^{(k)}$	表示在 GNN 中第 k 层中用户 i 的隐藏表示
$N_r(i)$	表示在 GNN 中用户 i 的邻域
$a_{i,r}^{(k+1)}$	表示第 k 层中用户 i 的邻域在关系 r 下传递的消息
$\mathrm{AGGREGATE}_r$	表示在图数据处理中关系 r 下的聚合函数
UPDATE	表示在图数据处理中关系 r 下的更新函数
G_h	表示第 h 个门控网络
KeepTop$K$	表示根据输入特征选择前 k 个最高的门控值的函数
$G_h(x_i^{\mathrm{mod}})_j$	表示第 h 个门控网络将用户 i 的推文编码信息、关系编码信息或元数据编码信息分配给第 j 个社群的概率
$E_j(x_i^{\mathrm{mod}})$	表示第 j 个专家网络输出的用户 i 的推文特征、关系编码特征或元数据编码特征
z_i^{mod}	表示所有门控网络输出合并得到用户 i 的推文特征、关系编码特征或元数据编码特征最终输出
TRM	表示 multi-head transformer
x_i^m	表示用户 i 的元数据编码信息
x_i^g	表示用户 i 的关系数据编码信息
x_i^t	表示用户 i 的推文数据编码信息
$\boldsymbol{M}$	表示固定池化矩阵
z_i^g	表示用户 i 关系数据在多门控混合专家层的输出
z_i^t	表示用户 i 推文数据在多门控混合专家层的输出
z_i^m	表示用户 i 元数据在多门控混合专家层的输出

2.2　特征编码层

1) 社交账户元数据处理:由于 MLP 结构简单,能够高效地处理数值型和分类型特征,适用于处理多种类型的输入数据. 因此,本文将使用一个两层的 MLP 来学习社交账户元数据中的 5 个数值型特征(关注者数、关注数、推文数、活跃天数、屏幕名称长度)和 3 个类别型特征(是否保护账号、是否已验证、是否使用默认个人资料图片),最终获得用户 i 的元数据编码为 x_i^m.

2) 社交账户推文数据处理:RoBERTa[32] 能够捕捉文本中的关键信息和语境,同时具有较强的通用性,常被用于短文本表征. 因此,本文用预训练的 RoBERTa[32] 语言模型对社交媒体账户上发布的推文文本内容进行编码,该过程可以定义为:

$$\{t'^{\,i}_1, t'^{\,i}_2, \cdots, t'^{\,i}_n\} = \mathrm{LM}\{t_1^i, t_2^i, \cdots, t_n^i\}, \tag{1}$$

其中,LM 表示预训练语言模型 RoBERTa, t_n^i 表示用户 i 的第 n 条推文. 然后,本文将获取的账户推文编码输入到一个两层 MLP 中,以获得用户推文数据编码信息 x_i^t.

3) 社交账户关系图数据处理:由于在社交媒体上用户与用户之间存在多种关系,例如(关注、粉丝、朋友)等,本文构建了一个社交账户关系图来建模用户与用户之间的关系. 该关系图中,每个节点表示一个

用户,节点的初始特征由该用户的元数据编码拼接推文数据编码组成,如果两个用户之间存在某一种关系,就通过边连接对应节点. 然后,利用 GNNs 的消息传递机制来建模用户之间的关联. 该过程可定义为:

$$a_{i,r}^{(k+1)} = \text{AGGREGATE}_r^{(k+1)}(\{g_j^{(k)}, \forall j \in N_r(i)\}, g_i^{(k)}), \tag{2}$$

$$g_i^{(k+1)} = \text{UPDATE}^{(k+1)}(g_j^{(k)}, a_i^{(k+1)}), \tag{3}$$

其中,$g_i^{(k)}$ 表示第 k 层中用户 i 的隐藏表示,$N_r(i)$ 表示用户 i 的邻域,$a_{i,r}^{(k+1)}$ 表示从用户 i 的邻域在关系 r 下传递的消息,AGGREGATE_r 与 UPDATE 表示关系 r 下的聚合函数和更新函数. 最终经过 l 层 GNN 编码器后,获得用户的关系数据编码信息 x_i^g.

$$h_i^{\text{agg}} = \sum_{u \in N_r(i)} h_u, \tag{4}$$

$$h_i^{\text{new}} = \text{RELU}(W \cdot (h_i + h_i^{\text{agg}}) + b). \tag{5}$$

具体来说,AGGREGATE_r 为关系 r 下的聚合函数,用于从用户节点 i 的邻居节点中收集信息,采用的方式为求和聚合(sum aggregation),具体操作如式(4)所示,其中 h_i^{agg} 是节点 i 的聚合表示,$N_r(i)$ 是节点 i 的邻居集合,h_u 是邻居节点 u 的特征表示. 它的目的是将邻居节点的特征整合成一个综合的表示,以便在更新步骤中使用. UPDATE 表示关系 r 下的更新函数,用于将 AGGREGATE_r 操作得到的信息与用户节点 i 本身的特征结合起来,以更新用户 i 节点的表示,具体操作如公式(5)所示,其中 h_i^{new} 是更新后的节点 i 的特征表示,h_i 是节点 i 当前的特征,h_i^{agg} 是通过 AGGREGATE_r 操作得到的邻居信息.

2.3 多门控混合专家层

在 MGEBot 模型中,专家网络是通过创建一个包含多个专家的模块列表来实现的,每个专家都是一个前馈神经网络. 这些专家网络通过输入大小、输出大小以及隐藏层大小等参数进行配置和初始化. 门控机制在此过程中发挥了关键作用. 输入数据首先通过一个门控网络,计算出每个专家的权重分配. 这些权重决定了每个输入样本应该由哪些专家进行处理.

在前向传播时,输入数据会根据门控网络计算出的权重分配给若干个专家进行处理. 每个专家独立处理自己的输入数据,然后将结果根据门控权重进行加权组合,得到最终的输出. 为了鼓励专家的均匀使用,并减少计算开销,模型还引入了正则化损失. 这种机制不仅提升了模型的适应性和泛化能力,还优化了资源的使用. 多门控混合专家层具体框架如图 3 所示.

多门控混合专家网络在 MGEBot 模型中发挥了关键作用. 其主要功能是通过多个门控网络动态调整

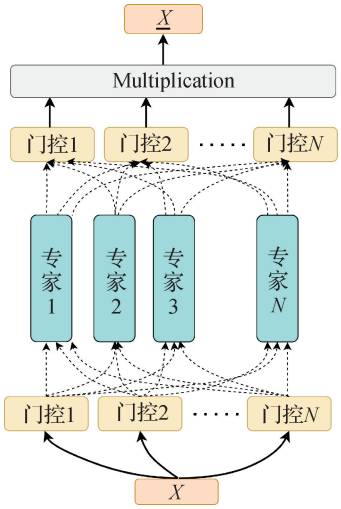

图 3 多门控混合专家

Fig. 3 Multi door control mixed expert

每个专家的权重,使模型能够灵活地选择和组合专家输出,以应对不同社区间的差异性. 具体来说,多门控网络根据输入特征学习到的权重分配,使每个专家在其擅长的领域发挥最大作用,从而优化资源利用和减少计算开销. 这种机制不仅提升了模型的适应性和泛化能力,还增强了其在不同数据集和应用场景中的稳定性和精度. 多门控混合专家网络的学习过程如下:

$$z_{h,i}^{\text{mod}} = \sum_{j=1}^{N} G_h(x_i^{\text{mod}})_j E_j(x_i^{\text{mod}}), \tag{6}$$

$$G_h(x_i^{\text{mod}})_j = \text{soft max}(\text{KeepTop}K(W_g x_i^{\text{mod}}, k))_j, \tag{7}$$

式中 $Z_{h,i}^{\text{mod}}$ 是第 h 个门控网络选择的所有专家的输出的加权和,$\text{mod} \in \{m,t,g\}$ 分别表示元数据、推文数据和图关系. N 表示专家数量. $G_h(x_i^{\text{mod}})_j$ 表示第 h 个门控网络将用户 i 分配给第 j 个社群的概率,$W_g \in \mathbb{R}^{d \times n}$ 表示可学习参数,$\text{KeepTop}K$ 表示一个函数,能够根据输入特征 x_i^{mod} 选择前 k 个最高的门控值. $E_j(x_i^{\text{mod}})$ 表示第 j 个专家网络的输出,是一个两层 MLP. 随后将所有门控网络输出合并得到最终输出 z_i^{mod}:

$$z_i^{\text{mod}} = \text{Multiplication}(z_{1,i}^{\text{mod}}, z_{2,i}^{\text{mod}}, \cdots, z_{h,i}^{\text{mod}}). \tag{8}$$

通过多门控混合专家机制,每个用户被分配到其相应的社群进行处理,使模型能够针对不同社群中的不同用户分布进行调整,从而解决机器人检测中的多样化用户社群挑战.

2.4 特征融合层

在获得经过多门控混合专家层处理过后的 3 种用户表征 z_i^g、z_i^t、z_i^m(分别为关系表征、推文表征、元数据表征)后本文将送入 Transformer 将这些特征组合

起来并评估一致性:

$$\{z_i^{\sim g}, z_i^{\sim t}, z_i^{\sim m}\} = \mathrm{TRM}(\{z_i^g, z_i^t, z_i^m\}), \tag{9}$$

$$z_i^{\mathrm{con}} = \mathrm{flatten}(\mathrm{sample}(\boldsymbol{M}) * w), \tag{10}$$

其中 TRM 表示 Transformer,$\boldsymbol{M}$ 表示注意力矩阵,w 表示滤波器. 滤波器指的是在 2D 卷积操作中使用的卷积核. 它在卷积操作中负责提取特定的特征,通过对下采样后的注意力矩阵进行卷积操作,生成具有特定特征的输出特征图. 这个过程通过滤波器的加权和操作来实现对输入数据的特征提取和转换. 具体而言,我们使用固定池对注意力矩阵 $\boldsymbol{M}$ 执行下采样,使用 ω 执行 2D 卷积,并将输出平坦化为矢量 z_i^{con}.

在经过 Transformer 之后 MGEBot 分别获得了用户 3 种模态的特征. 之后本文将关系表征、推文表征、元数据表征拼接起来,并应用线性变换得到最终 MGEBot 对账户的检测结果 y_i:

$$y_i = W_o \cdot [z_i^{\sim g}, z_i^{\sim t}, z_i^{\sim m}] + b_o, \tag{11}$$

其中,W_o 和 b_o 为可学习的参数,[,]为连接操作.

2.5　损失函数

在 MGEBot 训练中为了防止模型过拟合,本文采用带有 L2 正则化项的交叉熵,并且为了避免某些专家过于主导而其他专家被忽视的情况本文引入了平衡损失函数 $BL(\cdot)$,最终整体的损失函数可以表达为

$$\mathrm{Loss} = -\sum_{i\in Y} t_i \log y_i + \lambda_1 \sum_{\omega\in\theta} \omega^2 + \lambda_2 \sum_{i\in Y} \sum_{\mathrm{mod}}^{\{g,t,m\}} \mathrm{BL}(x_i^{\mathrm{mod}}), \tag{12}$$

其中:y_i 为 MGEBot 对用户 i 的预测输出,t_i 为对应用户 i 的真实标签,θ 表示所有可训练的模型参数,λ_1 和 λ_2 为超参数. 对于平衡损失 $\mathrm{BL}(\cdot)$,可表示为:

$$\mathrm{BL}(x) = w_{\mathrm{imp}} \cdot \mathrm{CV}(G(x))^2 + w_{\mathrm{ld}} \cdot \mathrm{CV}(P(x,i))^2, \tag{13}$$

其中,CV 表示变异系数[33],$G(x)$ 表示门控网络输出,$P(x,i)$ 为 Shazeer 等[34]定义的平滑函数,w_{imp} 和 w_{ld} 为平衡专家重要性和负荷的超参数.

3　实　验

3.1　实验数据集

为了评估 MGEBot 的性能,本文选择在 3 个广泛使用[12,16,27,35]的数据集进行测试:Cresci-15[17]、TwiBot-20[18]和 TwiBot-22[19]. Cresci-15[17]于 2015 年提出,包含 5 301 个社交账户信息,其中机器人账户为 3 351 个,占比为 63%;Twitbot -20[18]提供包含来自体育、经济、娱乐和政治领域的 229 580 名社交账户;Twitbot -22[19]是迄今为止用于社交机器人检测的最广泛的数据集,大小约为 100 G,拥有 Twitter 上 1 000 000用户的各种实体和关系. 其中 # Human 表示真实用户的社交账户数量;# Bot 表示机器人的社交账户数量;# User 表示总计社交账户的数量;# Tweer所有账户总计发布推文的数量;# Human Tweet 表示真实账户发布的推文总数;# Bot Tweet 表示机器人发布的推文总数;# Edge 表示构建关系图中所含边的数量(由于 TwiBot-22 中含有未标记用户所以,# Human + # Bot 不等于 # User). 3 个数据集的情况见表 2.

表 2　数据集

Tab. 2　Data set

指标	数量		
	Cresici-15	TwiBot-20	TwiBot-22
# Human	1 950	5 237	860 057
# Bot	3 351	6 589	139 943
# User	5 301	229 580	1 000 000
# Tweer	2 827 757	33 488 192	88 217 457
# Human Tweet	2 631 730	33 488 192	81 250 102
# Bot Tweet	196 027	33 488 192	6 967 355
# Edge	7 086 134	33 716 171	1 701 859 377

3.2　实验环境及配置

本文实验的神经网络模型是基于 Torch 实现的,Torch 版本为 1.8,编译语言为 Python3.8,在 2 个 NVIDIA 3090 GPU 上进行实验的. 在实验中,学习率设置为 10-4、L2 正则中的 λ 设置为 10-6、BL 中系数 w_{exp} 设置为 10-2、dropout 设置为 0.3、最大 epochs 设置为 400、关系数据专家数量设置为 3、推文数据专家数量设置为 3、元数据专家数量设置为 4、门控数量设置为 3.

3.3　评估指标

本文使用准确率(Accuracy,ACC)与 F_1 作为检测的评估指标,具体计算方法如下所示:

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

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

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

$$F_1 = \frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}, \tag{17}$$

其中 TP、TN、FP、FN、分别表示预测为正确预测社交机器人账户、正确预测真实账户、错误预测社交机器人账户和错误预测真实账户的样本数量. Precision 为预测准确度,表示模型在所有预测为社交机器人账户中正确的预测比例;Recall 为召回率,表示所有社交机器人账户中正确预测的比例;ACC 是预测结果正确的样本数量占总样本的比例;而 F_1 则是综合考虑了模型准确率和召回率,是两者的调和平均数. 因此,两项的评估指标值越大则表示模型的检测效果越好.

3.4 对比实验

本文将用 MGEBot 模型与以下当前主流的几个社交机器人检测模型进行比较:

SGBot[10]:SGBot 主要针对社交账户的账户信息进行检测,SGBot 从用户元数据中提取特征,并将其提供给随机森林分类器,用于可扩展和可推广的机器人识别.

Wei 等[22]:使用入和双向 LSTM 处理用户文本信息进行社交机器人检测.

Varol 等[23]:主要元数据、网络等衍生统计数据进行随机森林社交机器人检测,旨在检测不同类型的 Twitter bot.

BotRGCN[12]:从 Twitter 网络构建了一个异构图,并使用关系图卷积网络进行用户表示学习和 Twitter bot 检测.

RGT[15]:RGT 对 Twittersphere 中固有的异质性进行建模,以改进社交机器人检测.

BotBuster[31]:一种社交机器人检测系统,通过混合专家处理用户元数据和文本信息来增强跨平台机器人检测.

BotMoE[16]:通过混合专家处理用户元数据和文本信息并加上多专家混合技术来提高检测准确度.

MGEBot 在 3 个数据集上与基线模型进行对比,实验结果如表 3 所示.

表 3 对比试验结果

Tab. 3 Comparative test results %

模型	Cresci-15		TwiBot-20		TwiBot-22	
	ACC	F_1	ACC	F_1	ACC	F_1
SGBot	77.11	77.91	81.45	84.86	75.08	36.55
Wei 等	96.09	82.64	71.26	57.31	70.20	53.59
Varol 等	93.18	94.69	78.45	80.99	73.88	27.53
BotRGCN	96.47	97.68	85.66	87.23	78.86	54.98
RGT	97.20	97.77	86.60	88.00	76.50	42.94
BotBuster	96.89	97.51	77.22	81.17	74.01	54.16
BotMoE	98.50	98.82	86.94	88.84	79.11	56.39
MGEBot	**98.63**	**98.91**	**87.10**	**89.10**	**79.66**	**57.13**

从实验结果可以发现:1) 随着数据集从 Cresci-15 到 TwiBot-20,再到 TwiBot-22 的逐年更新,所有模型在 3 个数据集上表现的平均 F_1 分数分别为 93.49%、82.18%和 47.90%,模型的检测性能整体呈现出逐步下降的趋势,这一现象反映了社交机器人技术的不断演变,使得检测难度逐年增加. 2) 基于图学习的模型,例如:RGT、BotRGCN,在检测准确性上普遍优于不使用图学习的模型,例如:SGBot、Wei 等、Varol 等. 以 Twitter-20 数据集上的表现为例,RGT 相较于 SGBot、Wei 等、Varol 等模型 F_1 分数分别提升了 3.14%、30.69%和 7.01%,BotRGCN 则分别提升了 2.37%、29.92%和 6.24. 在基于图学习的模型中,RGT 表现最优,相较于 BotRGCN 在 Twitter-20 数据集上 F_1 分数提升了 0.77%,这可能是得益于 RGT 利用了用户形成的异质图的拓扑结构,并对用户之间的影响强度进行建模所造成. 这表明图学习技术对于捕捉用户之间的关系信息至关重要,能够显著提升社交机器人检测的效果. 3) 基于多专家的模型,例如 BotMoE 和 MGEBot,整体表现最优. 具体而言,BotMoE 和 MGEBot 在 Twitter-20 数据集上相较于 RGT 的 F_1 分数分别提升 0.84%和 1.10%;同时,基于多专家的模型均是在图学习的基础上构建的,说明多专家具有更强的特征提取和复杂模式识别能力,是提升社交机器人检测能力的有效方法. 4) 本文所提出

的 MGEBot 相较于次优方法 BotMoE,在 3 个数据集上 F_1 分数分别提升了 0.09%、0.26%和 0.74,其原因在于 MGEBot 不仅结合了图学习和多专家的能力,还针对社区差异性以及多样性的问题,提出了多门控混合专家机制,能够根据社区的特性动态调整专家的权重组合,进而实现了检测性能的进一步提升.

3.5　收敛速度实验

在社交机器人检测任务中,快速收敛的模型可以更及时地更新和部署,从而提升系统的整体响应速度;并且较快的收敛速度通常也与更好的泛化能力相关,有助于减少模型过拟合的风险[36]. 在本节中,我们将进行 MGEBot 模型与基线模型在训练过程中收敛速度的比较,MGEBot 模型与基线模型将分别在 TwiBot-20 数据集进行训练并统一设置迭代轮次为 200 次并观察每个模型在训练集上的收敛情况. 实验结果如图 4 所示,在收敛速度实验中,表现最差的模型是 Varol 等和 BotBuster. 两者在前期的性能提升缓慢,并且最终性能均低于 80%,这表明它们在训练过程中收敛较慢,难以有效捕捉复杂社交机器人行为. 相对而言,两种基于图学习的方法 RGT 和 BotRGCN 的表现较为接近,这两个模型在训练过程中能够较快地收敛至 85%以上的性能,这表明了图学习的方法在社交机器人检测任务上的有效性. 表现最好的则是两种基于多专家的模型 BotMoE 和 MGEBot,BotMoE 最终在第 200 轮达到 87.07%,而 MGEBot 在第 200 轮达到 88.52%. MGEBot 在第 200 轮的性能相较于 BotMoE 提升了 1.45%,是由于 MGEBot 引入了多门控机制,使得模型能够根据不同社区的特性动态调整专家的权重组合. 这种机制提高了模型对多样化社交机器人特征的适应能力,使得模型能够快速收敛.

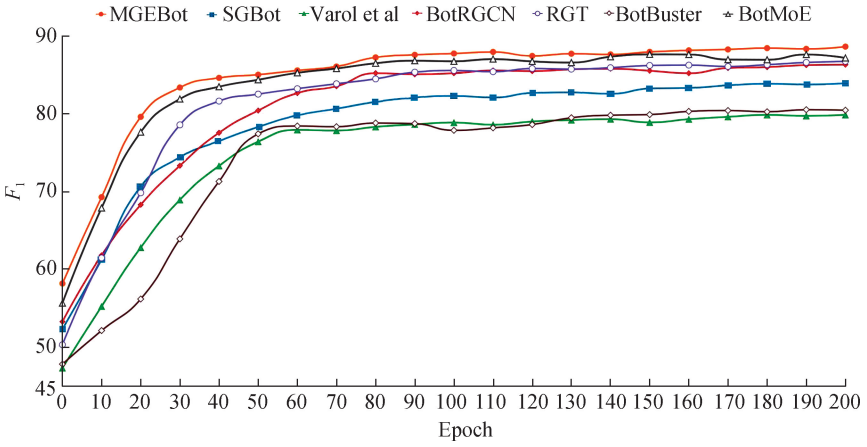

图 4　收敛速度实验

Fig. 4　Convergence speed experiment

3.6　鲁棒性实验

在当前的社交媒体生态系统中,只有少数热门平台提供了标签数据. 然而,为社交机器人获取足够且准确可靠的标签数据,往往需要耗费大量的时间和金钱成本. 现有的社交机器人检测模型在很大程度上依赖这些训练数据,这给社交机器人检测任务带来了显著的挑战. 因此,评估 MGEBot 模型对数据的依赖程度变得尤为重要. 通过对 MGEBot 模型进行测试,我们可以了解其在不同数据可用性条件下的表现. 这不仅有助于优化模型的性能,还能为未来的研究和实践提供宝贵的参考. 深入分析 MGEBot 模型在标签数据缺乏或不完全情况下的鲁棒性和适应能力,可以揭示其在实际应用中的有效性和局限性.

具体来说,本文在 Twitter-20 数据集上随机选择 10%至 100%的标签数量对 MGEBot 进行了测试. 实验结果如图 5 所示(横坐标为数据标签比例,纵坐标为模型测试分数),当标签数量从 100%减少到 10%时,MGEBot 的 F_1 值仅下降了 1.43%,仍高于多数基线模型在 100%标签数量下性能表现. MGEBot 能够在标签数据显著减少的情况下,依然保持较高的识别准确率,这对于实际应用具有重要意义. 在社交媒体平台上,获取全量且高质量的标签数据往往耗时且成本高昂. 同时能够说明,MGEBot 在处理复杂且不均衡的数据集时,仍能维持稳定的性能.

3.7　泛化性实验

在社交机器人检测领域,这一过程犹如“猫捉老鼠”游戏. 随着检测技术的不断进步,社交机器人的操控者们也在不断开发新对策以逃避检测. 这种动态对抗关系要求检测模型必须具备在未来未见过的社交账户或社区中识别机器人的能力. 为应对这一挑战,本

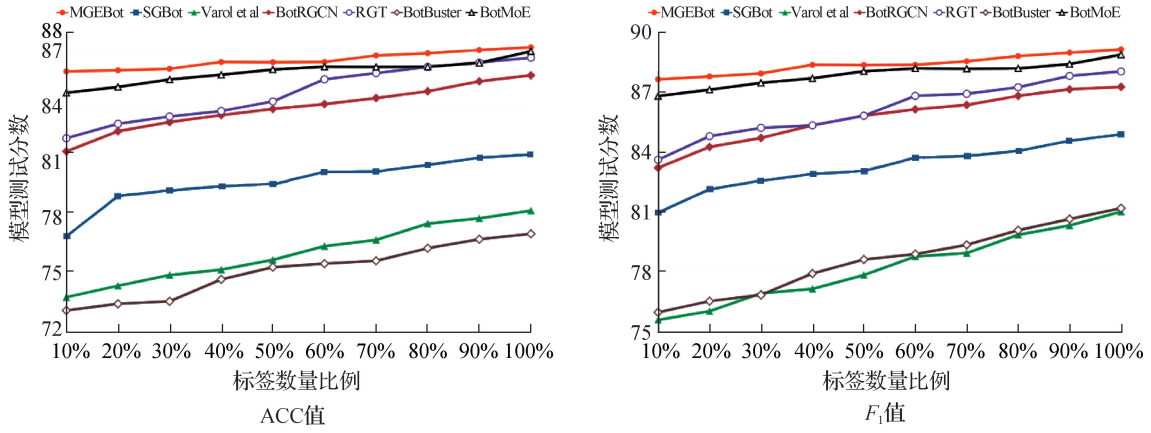

图 5 鲁棒性实验

Fig. 5 Robustness experiment

文在 TwiBot-20 数据集上训练了 MGEBot 和其他基线模型,并在 TwiBot-22 数据集上测试其性能.

实验结果如图 6 所示. 该实验设置很好地模拟了未来可能遇到的更先进的社交机器人群体情境. 具体而言,MGEBot 的 ACC 和 F_1 值分别下降了 7.44%和 33.21%. 尽管这两个指标的下降幅度并非最小,但 MGEBot 在整体性能上仍优于其他基线模型. 这一结果表明,MGEBot 在面对未来更复杂和先进的社交机器人时,依然能够保持较高的检测能力,充分说明了 MGEBot 在动态对抗环境中的有效性.

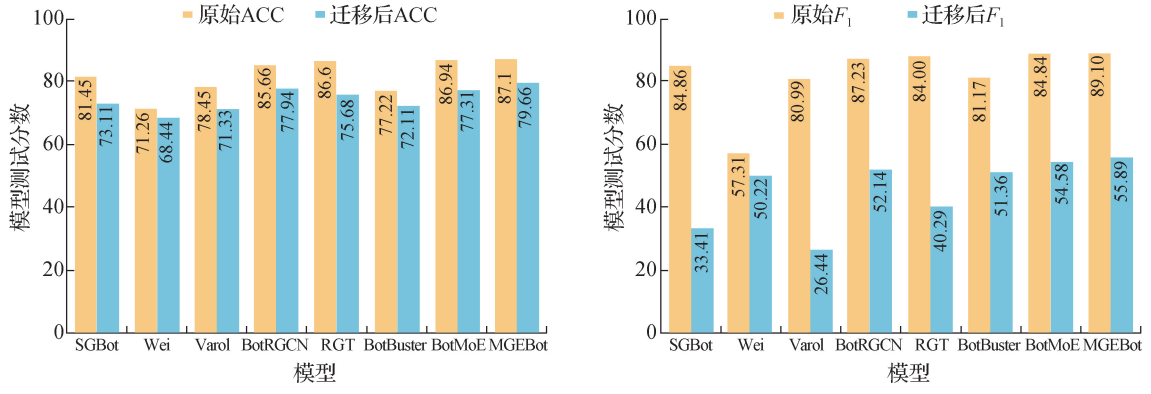

图 6 泛化性实验

Fig. 6 Gereralization experiment

3.8 门控与专家数量分析

在 MGEBot 框架中有 2 个重要的参数,分别为每个模态的专家数量以及门控数量. 为了研究 MGEBot 其对模型的影响,本文进行了网格搜索. 在图 7 中展示了 MGEBot 在 TwiBot-20 上不同专家数量以及门控数量对应的性能表现.

从该实验可分析得到,1) 当门控数量一定时并不是专家数量越多越好,反而有些较大的数量设置并没有体现出最好的性能. 2) 当专家数量一定(设置为 1)时,依次增加门控数量 MGEBot 的 F_1 分数分别为 86.4%、88.4%、88.8%、89.0%,由此可以得出多个门控对模型性能有着十分显著的提升,最高提升了 2.6%. 3) 综合门控与专家数量方面来看,当从只有 1 个门控提升到两个门控时,模型平均 F_1 分数提高了 1.56%,当门控数量从 2 个提升到 3 个时,模型平均 F_1 分数仅提高了 0.02%,然而当提升到 4 个门控数量时,模型的平均 F_1 分数与 3 个门控数量分数一致. 综上所述,当门控数量增加时模型性能有着明显提升,但不能盲目增加门控数量,而是需要寻找一个合适的值;专家数量设置同样也不是越多越好,这样不仅提高了计算成本且对应的表现也并不佳,依然需要寻找一个合适的值.

在实际应用中,对于社交媒体生态系统中社群较多的情况,可以考虑适当增加专家数量,以便模型能

够更好地捕捉不同社群的特征和行为模式,有助于提升模型对多样化社群的理解能力和适应能力. 另一方面,如果社群间的关系极为复杂或者社交媒体平台上的数据结构多样性较大,可以考虑增加门控数量,有助于模型更精确地管理和调整不同模态信息的权重,从而提升模型在处理复杂关系和多样化数据结构时的泛化能力和鲁棒性. 因此,在实际应用中,根据具体的社交媒体平台特性和数据环境,动态调整 MGEBot 的门控和专家数量,可以有效优化模型的性能,提升其在各种复杂应用场景下的效果和可靠性.

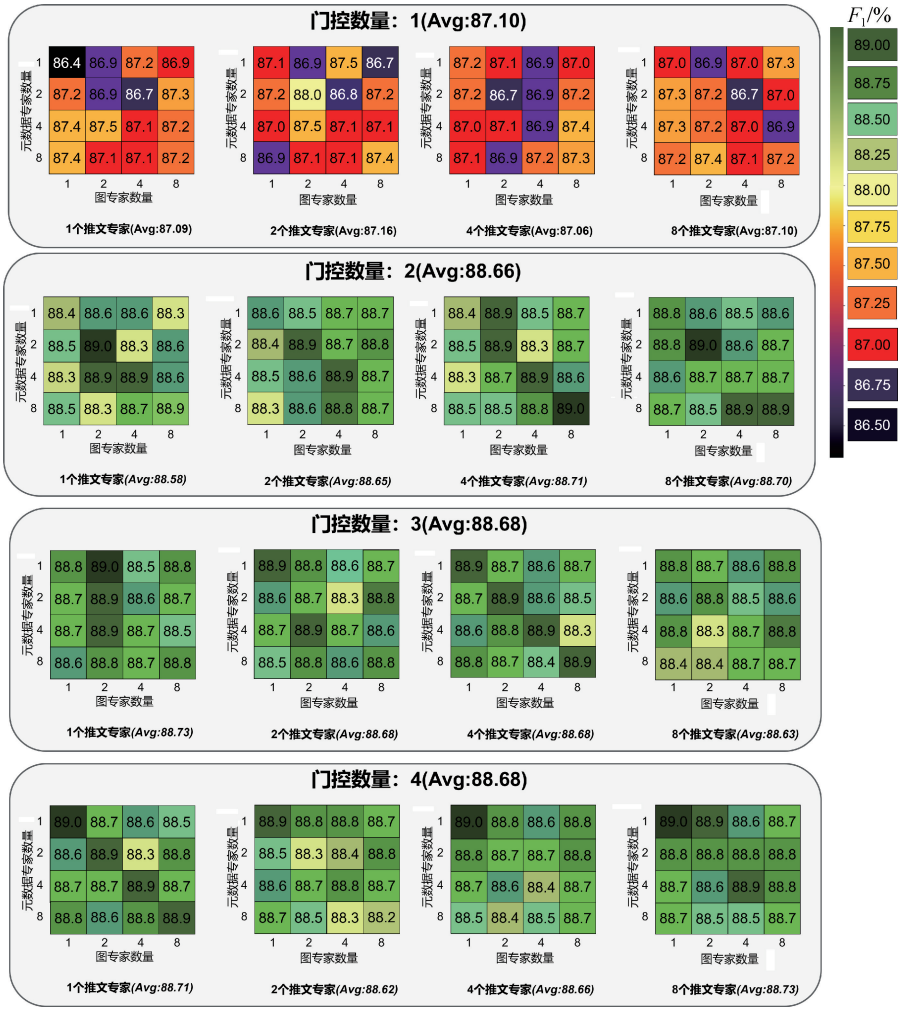

图 7　门控与专家数量分析实验

Fig. 7　Gate control and expert quantity experiment

4　结　论

社交机器人检测任务是针对社交媒体账户的真实性进行区分的关键性任务,本文提出了一种名为 MGEBot 的社交机器人检测框架. MGEBot 融合账户 3 大特征并通过多门控混合专家层学习社区差异性,从而提高了模型对于先进社交机器人的识别能力. 通过在 Cresci-15、TwiBot-20 和 TwiBot-22 数据集上的实验验证,MGEBot 相较于基准模型表现出了更好的效果,展现了其在不同数据集上的泛化能力和鲁棒性. 在鲁棒性实验中,本文测试了 MGEBot 在不同标签数据可用性条件下的性能. 实验表明,即使在标签数据显著减少的情况下,MGEBot 仍然能够保持较高的识别准确率,证明了其在处理复杂且不均衡数据集时的优势. 在泛化性实验中,本文测试了 MGEBot 在未来可能遇到的更先进的社交机器人情境下的性能. 结果显示,虽然在面对更复杂和先进的社交机器人时,MGEBot 的性能略有下降,但仍然优于其他基线模型. 这表明 MGEBot 在动态对抗环境中仍能保持较高的检测能力,充分展现了其在未知情境下的有效性和可靠性.

参考文献:

[1] 于川. 面向社交机器人的检测及差异性研究[D]. 济南:山东大学,2023.

[2] 龙光华. 基于情感和网络特征的社交机器人检测研究[D]. 南昌:南昌大学,2023.

[3] 张玄,李保滨. 微博环境中的机器人账户检测综述[J]. 中文信息学报,2022,36(12):1-15.

[4] FERRARA E. Twitter spam and false accounts prevalence, detection and characterization: a survey[EB/OL]. [2024-05-01]. https://arxiv.org/abs/2211.05913v4.

[5] PENG H L, ZHANG Y J, SUN H, et al. Domain-aware federated social bot detection with multi-relational graph neural networks[C]// 2022 International Joint Conference on Neural Networks (IJCNN). Padua: IEEE, 2022:1-8.

[6] SAYYADIHARIKANDEH M, VAROL O, YANG K C, et al. Detection of novel social bots by ensembles of specialized classifiers[C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Virtual Event Ireland: ACM, 2020: 2725-2732.

[7] CRESCI S, DI PIETRO R, PETROCCHI M, et al. The paradigm-shift of social spambots: evidence, theories, and tools for the arms race[C]// Proceedings of the 26th International Conference on World Wide Web Companion-WWW'17 Companion. Perth: ACM, 2017:963-972.

[8] Cresci S. A decade of social bot detection[J]. Commun ACM, 2020, 63(10):72-83.

[9] KUDUGUNTA S, FERRARA E. Deep neural networks for bot detection[J]. Information Sciences, 2018, 467: 312-322.

[10] YANG K C, VAROL O, HUI P M, et al. Scalable and generalizable social bot detection through data selection [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1):1096-1103.

[11] HU Z N, DONG Y X, WANG K S, et al. Heterogeneous graph transformer[C]// Proceedings of The Web Conference 2020. [S. l.]: WC, 2020:2704-2710.

[12] FENG S B, WAN H R, WANG N N, et al. 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. Virtual Event Ireland: ACM, 2021:236-239.

[13] LEI Z Y, WAN H R, ZHANG W Q, et al. BIC: twitter bot detection with text-graph interaction and semantic consistency[EB/OL]. [2024-05-01]. https://arxiv.org/abs/2208.08320v2.

[14] LI S D, ZHAO C Y, LI Q, et al. BotFinder: a novel framework for social bots detection in online social networks based on graph embedding and community detection[J]. World Wide Web, 2023, 26(4):1793-1809.

[15] FENG S B, TAN Z X, LI R, et al. Heterogeneity-aware twitter bot detection with relational graph transformers [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(4):3977-3985.

[16] LIU Y H, TAN Z X, WANG H, et al. BotMoE: twitter bot detection with community-aware mixtures of modal-specific experts[C]// Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. Taipei: ACM, 2023:485-495.

[17] CRESCI S, DI PIETRO R, PETROCCHI M, et al. Fame for sale: efficient detection of fake Twitter followers[J]. Decision Support Systems, 2015, 80:56-71.

[18] FENG S B, WAN H R, WANG N N, et al. TwiBot-20: a comprehensive twitter bot detection benchmark[C]// Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Virtual Event: ACM, 2021:4485-4494.

[19] FENG S B, TAN Z X, WAN H R, et al. TwiBot-22: towards graph-based Twitter bot detection[EB/OL]. (2023-02-12)[2024-05-01]. https://arxiv.org/PDF/2206.04564.

[20] YANG K C, VAROL O, HUI P M, et al. Scalable and generalizable social bot detection through data selection [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1):1096-1103.

[21] YANG K C, FERRARA E, MENCZER F. Botometer 101: social bot practicum for computational social scientists [J]. Journal of Computational Social Science, 2022, 5 (2):1511-1528.

[22] WEI F, NGUYEN U T. Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings[C]// 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA). Los Angeles: IEEE, 2019:101-109.

[23] DUKIC D, KECA D, STIPIC D. Are you human? detecting bots on Twitter using BERT[C]// 2020 IEEE 7th International Conference on Data Science and Advanced Analytics. Sydney: DSAA, 2020:1-10.

[24] CAI C Y, LI L J, ZENGI D. Behavior enhanced deep bot detection in social media[C]// 2017 IEEE International Conference on Intelligence and Security Informatics

(ISI). Beijing:IEEE,2017:128-130.

[25] KUDUGUNTA S,FERRARA E. Deep neural networks for bot detection[J]. Information Sciences, 2018, 467: 312-322.

[26] SHI S H, QIAO K, CHEN J, et al. MGTAB: a multi-relational graph-based twitter account detection benchmark[EB/OL]. [2024-05-01]. https://arxiv.org/abs/2301.01123v2.

[27] LEI Z Y, WAN H R, ZHANG W Q, et al. BIC: twitter bot detection with text-graph interaction and semantic consistency[EB/OL]. [2024-05-01]. https://arxiv.org/abs/2208.08320v2.

[28] SHAZEER N, MIRHOSEINI A, MAZIARZ K, et al. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer[EB/OL]. [2024-05-01]. https://arxiv.org/abs/1701.06538?context=cs.CL.

[29] MADAAN A, TANDON N, RAJAGOPAL D, et al. Think about it! Improving defeasible reasoning by first modeling the question scenario[EB/OL]. [2024-05-01]. https://arxiv.org/abs/2110.12349v1.

[30] MA J Q, ZHAO Z, YI X Y, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018: 1930-1939.

[31] NG L H X, CARLEY K M. BotBuster: multi-platform bot detection using a mixture of experts[EB/OL]. [2024-05-01]. https://arxiv.org/abs/2207.13658v1.

[32] LIU Y H, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized BERT pretraining approach[EB/OL]. [2024-05-01]. https://arxiv.org/abs/1907.11692.

[33] EVERITT B. The cambridge dictionary of statistics[M]. Cambridge: Cambridge University Press, 1998.

[34] SHAZEER N, MIRHOSEINI A, MAZIARZ K, et al. Outrageously large neural networks: the sparsely-gated mixture-of-experts layer[EB/OL]. [2024-05-01]. https://arxiv.org/abs/1701.06538v1.

[35] VAROL O, FERRARA E, DAVIS C, et al. Online human-bot interactions: detection, estimation, and characterization[J]. Proceedings of the International AAAI Conference on Web and Social Media, 2017, 11(1): 280-289.

[36] ZHANG Y, GAO S, HUANG Y, et al. 3A-COT: an attend-arrange-abstract chain-of-thought for multi-document summarization[J]. International Journal of Machine Learning and Cybernetics, 2024(1): 1-19.

(责任编辑:汪　军)